

Appunti di Algebra lineare*

A. Peretti

14 giugno 2005

1 Topologia di \mathbb{R}^n e funzioni di più variabili

1.1 Insiemi in \mathbb{R}^n

Consideriamo dapprima insiemi di punti su un asse cartesiano. Utilizzando la corrispondenza biunivoca che si può stabilire tra i punti di un asse cartesiano e i numeri reali relativi del campo \mathbb{R} assoceremo sempre ad ogni punto la sua ascissa e ad ogni numero reale il punto che può essere considerato la sua immagine geometrica. Così parleremo indifferentemente di punti e di numeri reali relativi.

Insiemi notevoli in \mathbb{R} sono gli intervalli. Dati $a, b \in \mathbb{R}$, con $a < b$, è un intervallo l'insieme

$$[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}.$$

Altri esempi di intervalli sono, come noto,

$$\begin{aligned}(a, b) &= \{x \in \mathbb{R} \mid a < x < b\}, \\ [a, b) &= \{x \in \mathbb{R} \mid a \leq x < b\}, \\ (a, b] &= \{x \in \mathbb{R} \mid a < x \leq b\}.\end{aligned}$$

I numeri a e b sono gli estremi dell'intervallo. Se gli estremi sono *entrambi* compresi nell'intervallo di cui si vuol parlare si dirà che l'intervallo è chiuso; se invece sono *entrambi* esclusi si parlerà di intervallo aperto.

Un tipo particolare di intervalli sono gli *intorni circolari di un punto* x_0 . Sono gli intervalli

$$(x_0 - \delta, x_0 + \delta),$$

dove δ può essere un qualunque numero reale positivo e viene detto il *raggio* dell'intorno. Possono essere aperti o chiusi e hanno come immagine geometrica un segmento del quale il punto di ascissa x_0 è punto medio.

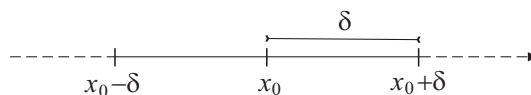
Se, dati due punti x e y in \mathbb{R} , definiamo **distanza di x e y** il numero reale non negativo

$$d(x, y) = |x - y|,$$

allora si ha

$$(x_0 - \delta, x_0 + \delta) = \{z \in \mathbb{R} \mid d(z, x_0) < \delta\} = \{z \in \mathbb{R} \mid |z - x_0| < \delta\} :$$

quindi l'intorno aperto di x_0 di raggio δ è costituito dai punti che distano da x_0 meno di δ .



Avremo occasione di considerare anche intervalli illimitati, che sono insiemi del tipo

$$\begin{aligned}[a, \infty) &= \{x \in \mathbb{R} \mid x \geq a\}, \\ (a, \infty) &= \{x \in \mathbb{R} \mid x > a\}, \\ (-\infty, a] &= \{x \in \mathbb{R} \mid x \leq a\}, \\ (-\infty, a) &= \{x \in \mathbb{R} \mid x < a\},\end{aligned}$$

*In questa dispensa sono trattati gli argomenti che fanno parte del programma del corso di Algebra lineare impartito presso la Facoltà di Scienze Statistiche nell'Anno Accademico 2004/2005.

dove $a \in \mathbb{R}$. Essi hanno come immagine geometrica una delle due semirette dell'asse reale che hanno origine nel punto di ascissa a . Talvolta si scrive anche $(-\infty, \infty)$, intendendo con questo tutto l'insieme \mathbb{R} .

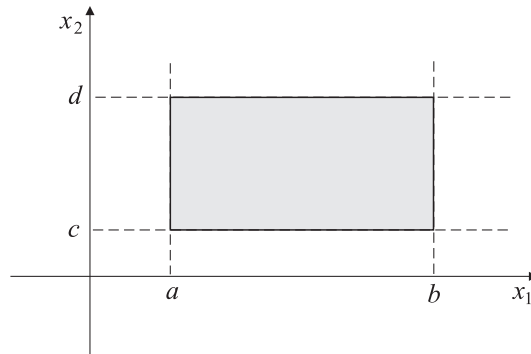
In \mathbb{R} sono noti i concetti di punto *interno* di un insieme, punto *esterno*, punto *di frontiera*, punto *di accumulazione*, punto *isolato*. Sono noti inoltre i concetti di insieme *aperto* e insieme *chiuso*.

Se nel piano consideriamo due assi cartesiani ortogonali con l'origine nel loro punto comune, possiamo stabilire, così come fatto tra i punti di una retta ed \mathbb{R} , una corrispondenza biunivoca tra i punti del piano e l'insieme \mathbb{R}^2 , cioè il prodotto cartesiano $\mathbb{R} \times \mathbb{R}^1$, cioè l'insieme delle coppie ordinate di numeri reali.

Possiamo chiamare *intervallo in \mathbb{R}^2* un sottoinsieme di \mathbb{R}^2 che è prodotto cartesiano di due intervalli di \mathbb{R} . Ad esempio è un intervallo in \mathbb{R}^2 l'insieme

$$A = [a, b] \times [c, d] = \{(x_1, x_2) \in \mathbb{R}^2 \mid a \leq x_1 \leq b; c \leq x_2 \leq d\}, \quad a, b, c, d \in \mathbb{R},$$

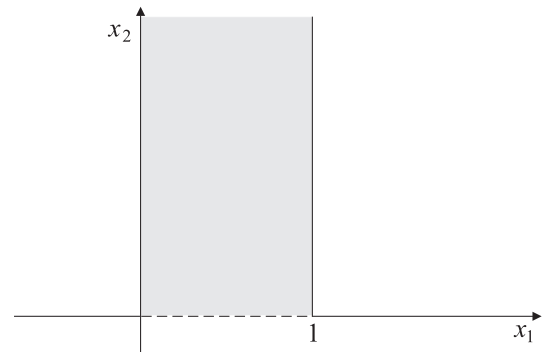
ed ha come immagine geometrica un rettangolo con i lati paralleli agli assi.



L'intervallo è aperto se è prodotto cartesiano di due intervalli aperti ed è chiuso se è prodotto cartesiano di due intervalli chiusi. Le altre possibilità danno origine ad intervalli che non sono né aperti né chiusi.

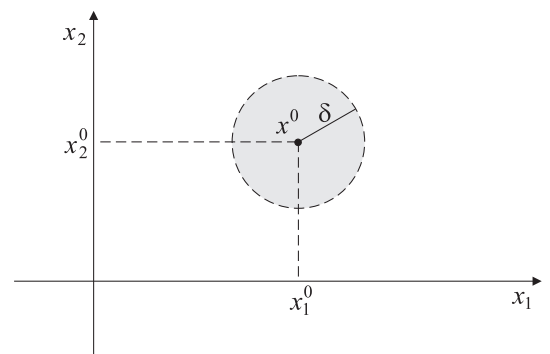
Possiamo anche qui parlare di intervalli illimitati: sono quelli che sono prodotto cartesiano di due intervalli, di cui almeno uno illimitato. È un intervallo illimitato ad esempio l'insieme

$$A = [0, 1] \times (0, \infty) = \{(x_1, x_2) \in \mathbb{R}^2 \mid 0 \leq x_1 \leq 1; x_2 > 0\}.$$



Altri esempi notevoli di sottoinsiemi di \mathbb{R}^2 sono poi anche gli interni circolari di un punto.

Fissato un punto $P(x_1^0, x_2^0)$, si dice intorno circolare di P di raggio $\delta > 0$ l'insieme dei punti del piano che hanno da P distanza (euclidea) minore di δ (o minore o uguale). Essi come è noto riempiono un cerchio di centro P e raggio δ . È da notare che gli interni circolari di P sono infiniti, tutti contengono il punto P e considerato uno di essi, di raggio $\bar{\delta}$, ne esistono infiniti altri di raggio minore tutti contenuti in esso.



Ricordiamo che la *distanza euclidea* tra $P(x_1, x_2)$ e $Q(y_1, y_2)$ è data da

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

Riprendiamo brevemente in \mathbb{R}^2 le nozioni *topologiche*, già note in \mathbb{R} , di punti interni a un insieme, punti esterni a un insieme e punti di frontiera per l'insieme, di punto isolato di un insieme e di punto di accumulazione per l'insieme.

Definizione Dato un insieme $A \subset \mathbb{R}^2$, un punto P si dice

¹ Ricordo che, se A e B sono due insiemi, si definisce prodotto cartesiano di A e B l'insieme $A \times B$ delle coppie ordinate (a, b) di elementi rispettivamente di A e di B .

- *interno* ad A se $P \in A$ ed esiste almeno un suo intorno circolare tutto contenuto in A ;
- *esterno* ad A se esiste almeno un suo intorno circolare tutto costituito da punti non appartenenti ad A ;
- *di frontiera* per A se P non è né interno né esterno ad A . Succederà allora che in qualunque suo intorno circolare cadono almeno un punto di A e almeno un punto non appartenente ad A ;
- *isolato* se $P \in A$ ed esiste almeno un suo intorno circolare che non contiene alcun punto di A eccetto P ;
- *di accumulazione* per A se in qualunque suo intorno circolare cadono infiniti punti di A , indipendentemente dal fatto che esso appartenga o non appartenga ad A .

È importante inoltre l'ulteriore

Definizione Un insieme $A \subset \mathbb{R}^2$ si dice

- *aperto* se ogni punto di A è interno ad A ;
- *chiuso* se il suo insieme complementare² è aperto.

Osservazioni ed esempi.

- ▷ Si può osservare che un punto isolato è certamente di frontiera, un punto interno è certamente di accumulazione, un punto esterno non può essere di accumulazione, un punto isolato di un insieme non può essere di accumulazione per quell'insieme, un punto di frontiera può appartenere o non appartenere all'insieme e può essere o non essere di accumulazione per l'insieme.
- ▷ L'insieme dei punti interni ad un insieme A viene indicato con $\text{int}A$. Possiamo quindi anche dire che un insieme A è aperto se $A = \text{int}A$.
- ▷ \mathbb{R}^2 è aperto e chiuso nello stesso tempo, da cui \emptyset è anch'esso aperto e chiuso.
- ▷ Abbiamo già visto che un intervallo è aperto se e solo se è prodotto cartesiano di due intervalli entrambi aperti ed è chiuso se e solo se è prodotto cartesiano di due intervalli entrambi chiusi.
Se l'intervallo I è prodotto cartesiano di due intervalli, chiamiamoli I_1 e I_2 , allora $\text{int}I = \text{int}I_1 \times \text{int}I_2$. Si tratta del rettangolo privato del suo contorno.
- ▷ In generale i punti esterni sono quelli che risultano interni al complementare dell'insieme in questione.
- ▷ I punti di frontiera di un intervallo sono i punti che stanno sui lati del rettangolo, a prescindere dal fatto che questi facciano parte oppure no del rettangolo stesso.
- ▷ Un intervallo non ha punti isolati. Vediamo tra breve un insieme che ha punti isolati.
- ▷ I punti di accumulazione di un intervallo sono tutti i punti che sono interni oppure di frontiera per l'intervallo. Il loro insieme coincide quindi con l'intervallo chiuso.

Come esempio riassuntivo sugli intervalli, consideriamo $I = [0, 1] \times (-1, 1]$. L'insieme dei suoi punti interni è l'insieme $(0, 1) \times (-1, 1)^3$; l'insieme dei suoi punti di frontiera si può scrivere come

$$\left(\{0\} \times [-1, 1]\right) \cup \left(\{1\} \times [-1, 1]\right) \cup \left([0, 1] \times \{-1\}\right) \cup \left([0, 1] \times \{1\}\right);$$

l'insieme dei punti di accumulazione per I è l'intervallo $I = [0, 1] \times [-1, 1]$.

Altro esempio. Consideriamo l'insieme

$$A = \left((-1, 1) \times \{0\}\right) \cup \{(0, -1)\} \cup \{(0, 1)\}^4.$$

² Ricordiamo che, dato in generale un insieme X ed $A \subseteq X$, l'insieme degli elementi di X che non appartengono ad A si dice complementare di A (rispetto ad X). Nel nostro caso abbiamo $A \subset \mathbb{R}^2$ e sottointendiamo che il complementare sia rispetto a tutto \mathbb{R}^2 .

³ Come lo studente attento avrà notato, la scrittura (a, b) può risultare a questo punto ambigua: infatti, senza ulteriori informazioni, può indicare un intervallo (quindi un sottoinsieme) di \mathbb{R} oppure una coppia di numeri reali, cioè un elemento di \mathbb{R}^2 . Non c'è però ambiguità nella scrittura dell'insieme dei punti interni di I : infatti, trattandosi di un prodotto cartesiano, come noto questa è un'operazione tra insiemi e quindi, nel caso in questione, $(0, 1)$ e $(-1, 1)$ sono da interpretare senz'altro come intervalli di \mathbb{R} .

⁴ Si consideri che anche in questo caso non c'è ambiguità nell'interpretazione della scrittura di tipo (a, b) : $(-1, 1)$ è certamente un intervallo di \mathbb{R} (vedi nota precedente), mentre $(0, -1)$ e $(0, 1)$ sono invece certamente elementi di \mathbb{R}^2 e la scrittura $\{(0, -1)\}$ sta ad indicare quindi l'insieme il cui unico elemento è la coppia $(0, -1)$. Per acquisire dimestichezza con il formalismo, lo studente rifletta sempre con attenzione su questi aspetti.

L'insieme A (lo studente è invitato a disegnarne una rappresentazione grafica) non ha punti interni (si osservi anche che l'intervallo $(-1, 1)$ ha punti interni in \mathbb{R}); l'insieme A ha due punti isolati, e sono $(0, -1)$ e $(0, 1)$; tutti i punti di A sono di frontiera per A ; sono di frontiera anche gli estremi del segmento, cioè $(-1, 0)$ e $(1, 0)$; i punti di accumulazione di A sono dati dall'insieme $[-1, 1] \times \{0\}$.

Gli studenti si costruiscano altri esempi in proposito.

Tutto quanto detto per \mathbb{R}^2 può essere esteso in modo del tutto naturale ad \mathbb{R}^n , l'insieme delle n -ple ordinate di numeri reali. Da notare che, in tal caso, solo per $n = 3$ si può dare ancora una immagine geometrica, che sarà ambientata in quello che si può indicare come lo spazio a tre dimensioni nel senso elementare e intuitivo del termine, attrezzato con una terna di assi cartesiani ortogonali.

Si possono chiamare intervalli in \mathbb{R}^n i prodotti cartesiani di n intervalli in \mathbb{R} . Ad esempio è un intervallo in \mathbb{R}^3 l'insieme

$$A = [0, 1] \times (-1, 0) \times [-1, -1]$$

la cui immagine geometrica è un parallelepipedo con le facce parallele ai piani coordinati.

Si dirà intorno circolare di raggio δ di un punto $P \in \mathbb{R}^n$ l'insieme dei punti di \mathbb{R}^n che hanno distanza euclidea da P minore di δ , ricordando che, se $P(x_1, \dots, x_n)$ e $Q(y_1, \dots, y_n)$, la *distanza euclidea* tra P e Q è data da

$$d(P, Q) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad 5$$

L'intorno circolare di un punto $P \in \mathbb{R}^3$ di raggio δ è rappresentato dalla sfera di raggio δ con centro in P .

Si estendono ai sottoinsiemi di \mathbb{R}^n , in modo del tutto analogo a quanto fatto in \mathbb{R}^2 , tutte le nozioni topologiche.

Come già in \mathbb{R} e in \mathbb{R}^2 , anche in \mathbb{R}^n si dice *aperto* ogni insieme i cui punti siano tutti punti interni. Abbiamo già visto che, ad esempio, in \mathbb{R}^2 un rettangolo privato dei lati è un insieme aperto, e così pure un cerchio privato della circonferenza. In \mathbb{R}^3 sono aperti ad esempio un parallelepipedo privato delle facce oppure una sfera privata della superficie sferica.

Si dirà poi *chiuso* un insieme che sia complementare (rispetto a \mathbb{R}^n) di un insieme aperto.

Si può dimostrare che A è chiuso se e solo se contiene tutti i suoi punti di frontiera, e anche se e solo se contiene tutti i suoi punti di accumulazione.

Un insieme aperto contenente un punto P verrà detto intorno di P (può essere un intorno circolare ma può non esserlo).

1.2 Funzioni reali di n variabili reali

È noto che si dice funzione definita in un insieme A qualunque legge di corrispondenza che ad ogni elemento di A associa *in modo unico* un altro oggetto di un altro insieme. Per il nostro scopo servono le funzioni definite nei sottoinsiemi A di \mathbb{R}^n che associano ad ogni elemento di A uno ed un solo elemento di \mathbb{R} , cioè un numero reale. Per dire che f è una funzione in tal senso possiamo scrivere

$$f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}.$$

Se vogliamo riferirci alla funzione nel suo complesso scriveremo semplicemente f ; se invece, posto $x = (x_1, \dots, x_n)$, vogliamo indicare il numero reale che è immagine di x attraverso la funzione f , scriveremo $f(x)$.

Si dice *campo di esistenza* o *insieme di definizione* della funzione f l'insieme $C.E.(f)$ degli x di cui esiste il corrispondente secondo la legge f .

È immediato che debba essere $A \subseteq C.E.(f)$.

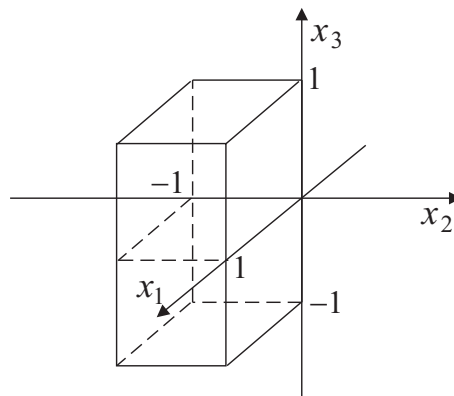
Si indica con il simbolo $f(A)$ l'insieme dei corrispondenti degli elementi di A , che talvolta viene chiamato immagine di A secondo f . In particolare $f(C.E.(f))$ è immagine dell'intero insieme di definizione della funzione e si dice anche "codominio" della funzione.

Chiaramente risulta $f(A) \subseteq f(C.E.(f))$.

▷ **Esempi** Con $n = 2$ sono esempi di funzioni definite in opportuni sottoinsiemi di \mathbb{R}^2 le seguenti funzioni:

$$1. f(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$$

⁵ Ricordo che, se $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, si definisce *norma euclidea* di x il numero reale non negativo $\|x\| = (\sum_{i=1}^n x_i^2)^{1/2}$. Dati allora $x, y \in \mathbb{R}^n$, si può allora ridefinire $d(x, y) = \|x - y\|$.



$$2. f(x_1, x_2) = \frac{\log(x_1 + 3x_2^2)}{3^{x_2}}$$

$$3. f(x_1, x_2) = \max(x_1, x_2).$$

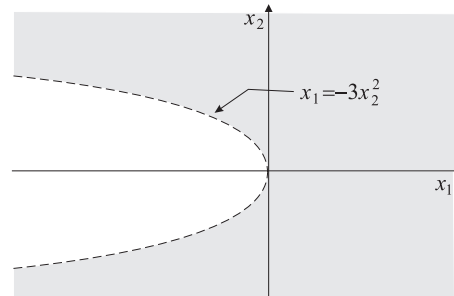
Quando l'insieme di definizione non è indicato si sottointende che esso sia $C.E.(f)$, cioè il più ampio sottoinsieme proprio o improprio di \mathbb{R}^n dotato della proprietà che ogni suo elemento abbia corrispondente secondo la legge f .

Così, nei tre esempi proposti, il campo di esistenza della f in 1 è tutto \mathbb{R}^2 , dato che $x_1^2 + x_2^2 \geq 0$, per ogni $(x_1, x_2) \in \mathbb{R}^2$.

Il campo di esistenza della f in 2 è l'insieme

$$C.E.(f) = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 > -3x_2^2\}$$

e si tratta della parte di piano raffigurata a fianco in grigio.



Il campo di esistenza della f in 3 è ancora tutto \mathbb{R}^2 .

L'insieme delle $(n+1)$ -uple $(x_1, \dots, x_n, f(x_1, \dots, x_n))$, con $(x_1, \dots, x_n) \in A$ costituisce il cosiddetto *grafico di f* , ed è naturalmente ambientato in \mathbb{R}^{n+1} . Solo con $n=1$ oppure $n=2$ possiamo rappresentare il grafico di una funzione. Lo studente ha imparato nel corso di Matematica I ad ottenere il grafico di funzioni $f: A \subset \mathbb{R} \rightarrow \mathbb{R}$ e a rappresentarlo come sottoinsieme di \mathbb{R}^2 .

Il grafico di una $f: A \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ è invece un sottoinsieme di \mathbb{R}^3 . Ad esempio, il grafico della funzione $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ definita da $y = x_1 + x_2$ è un piano nello spazio \mathbb{R}^3 .

Esaminiamo ora un concetto che sarà utile in seguito, quello di restrizione di una funzione. Prima però serve introdurre il concetto di curva in \mathbb{R}^n . Per non appesantire l'esposizione, consideriamo solo curve in \mathbb{R}^2 .

Chiamiamo *curva in \mathbb{R}^2* una funzione continua γ definita in un intervallo I di \mathbb{R} a valori in \mathbb{R}^2 . Pertanto abbiamo

$$\gamma: I \rightarrow \mathbb{R}^2, \quad \text{con } \gamma(t) = (\gamma_1(t), \gamma_2(t)),$$

dove γ_1, γ_2 sono funzioni continue definite nell'intervallo I .

L'immagine della curva γ , cioè il sottoinsieme di \mathbb{R}^2

$$\{(\gamma_1(t), \gamma_2(t)) : t \in I\},$$

si chiama anche *sostegno* della curva γ .

▷ Esempi

1. La funzione $\gamma(t) = (1+t, 1-t)$, con $t \in [-1, 1]$ è una curva in \mathbb{R}^2 . La sua immagine è un segmento nel piano.
2. La funzione $\gamma(t) = (t, t^2)$, con $t \in [0, 1]$ è una curva in \mathbb{R}^2 . La sua immagine è un arco di parabola.
3. La funzione $\gamma(t) = (\cos t, \sin t)$, con $t \in [0, 2\pi]$ è una curva in \mathbb{R}^2 . La sua immagine è una circonferenza con centro nell'origine e raggio 1.⁶

Data una funzione $f: A \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ e data una curva $\gamma: I \rightarrow \mathbb{R}^2$, con immagine contenuta in A , chiamiamo *restrizione di f alla curva γ* la funzione $f \circ \gamma$.

▷ Esempi

1. Data la funzione $f(x_1, x_2) = x_1 - x_2$, la sua restrizione alla curva $\gamma(t) = (t, 0)$, con $t \in \mathbb{R}$, è una restrizione di f all'asse x_1 .⁷ Si tratta della funzione $g(t) = f(t, 0) = t$.
La restrizione della stessa funzione alla curva $\gamma(t) = (t, t)$, con $t \in \mathbb{R}$, è una restrizione di f alla retta di equazione $x_1 = x_2$. Si tratta della funzione $g(t) = f(t, t) = 0$, con $t \in \mathbb{R}$.
2. La restrizione della funzione $f(x_1, x_2) = \frac{x_1 x_2}{x_1^2 + x_2^2}$ alla curva $\gamma(t) = (t, t)$, con $t \in (0, \infty)$ è la funzione $g(t) = f(t, t) = \frac{1}{2}$, con $t \in (0, \infty)$.
Una restrizione della stessa funzione alla parabola di equazione $x_2 = x_1^2$ si può ottenere con la funzione $g(t) = f(t, t^2) = \frac{t}{1+t^2}$, con $t \in (0, \infty)$.

⁶ L'immagine di una curva può quindi non essere il grafico di una funzione.

⁷ Non è unica la restrizione di f all'asse x_1 : ad esempio la funzione $g(t) = f(t^3, 0) = t^3$ è un'altra restrizione di f all'asse x_1 .

Anche se abbiamo presentato i concetti di curva e di restrizione nel caso bidimensionale, tutto si estende in modo ovvio al caso n -dimensionale.

A titolo di esempio, una restrizione della funzione

$$f: \mathbb{R}^3 \rightarrow \mathbb{R}, \quad \text{con } f(x_1, x_2, x_3) = e^{-x_1^2 - x_2^2 - x_3^2}$$

alla retta passante per il punto $(1, 1, 0)$ e parallela all'asse x_3 si ottiene considerando la restrizione di f alla curva $\gamma: \mathbb{R} \rightarrow \mathbb{R}^3$, con

$$\gamma(t) = (1, 1, t);$$

si tratta pertanto della funzione

$$g(t) = f(\gamma(t)) = f(1, 1, t) = e^{-2-t^2}.$$

1.3 Concetti di limite

Prenderemo in considerazione in \mathbb{R}^n soltanto il concetto di limite finito ℓ al tendere di $x = (x_1, \dots, x_n)$ ad $x^0 = (x_1^0, \dots, x_n^0)$ e il concetto di limite ∞ oppure $-\infty$ al tendere di x ad x^0 .

Sia data una funzione f definita in $A \subseteq \mathbb{R}^n$ e sia x^0 un punto di accumulazione per A . Per semplicità supporremo che se $x^0 \in A$ sia punto interno di A , e se $x^0 \notin A$ sia punto isolato del complementare di A rispetto a \mathbb{R}^n .⁸

Definizione Si dice che f *tende al numero reale ℓ per x tendente ad x^0* , e si scrive

$$\lim_{x \rightarrow x^0} f(x) = \ell,$$

se per ogni intorno circolare \mathcal{I} di ℓ esiste un intorno circolare di x^0 la cui immagine è contenuta in \mathcal{I} , con l'eventuale eccezione di x^0 sul quale non si fa alcuna ipotesi.⁹

Si dice poi che f *tende a ∞ per x tendente a x^0* , e si scrive

$$\lim_{x \rightarrow x^0} f(x) = \infty,$$

se per ogni $U = (a, \infty)$ esiste un intorno circolare di x^0 la cui immagine è tutta contenuta in U , eventualmente eccettuato x^0 .¹⁰

Si dice infine che f *tende a $-\infty$ per x tendente ad x^0* , e si scrive

$$\lim_{x \rightarrow x^0} f(x) = -\infty,$$

se per ogni $V = (-\infty, a)$ esiste un intorno circolare di x^0 la cui immagine, eventualmente escluso x^0 , è tutta contenuta in V .¹¹

▷ **Esempi**

1. Consideriamo la funzione $f(x_1, x_2) = x_1 + x_2$, definita in tutto \mathbb{R}^2 . Vogliamo provare che

$$\lim_{(x_1, x_2) \rightarrow (1, 1)} f(x_1, x_2) = 2.$$

Preso un qualunque intorno circolare di 2, indichiamolo con $(2 - \varepsilon, 2 + \varepsilon)$, possiamo trovare le (x_1, x_2) la cui immagine è contenuta in tale intorno risolvendo la disequazione

$$2 - \varepsilon < x_1 + x_2 < 2 + \varepsilon, \quad \text{cioè} \quad 2 - \varepsilon - x_1 < x_2 < 2 + \varepsilon - x_1,$$

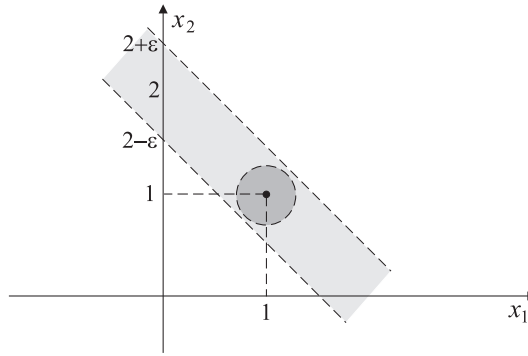
che porta all'insieme raffigurato qui sotto. Tale insieme contiene certamente un intorno circolare di $(1, 1)$, come rappresentato. Pertanto è provato che il limite in questione è 2.

⁸ Possiamo anche dire equivalentemente che esiste un intorno circolare $\mathcal{I}(x^0, \delta)$ di x^0 tale che $\mathcal{I}(x^0, \delta) \setminus \{x^0\} \subset A$.

⁹ Quindi, equivalentemente, se per ogni $\mathcal{I}(\ell, \varepsilon)$ esiste un $\mathcal{I}(x^0, \delta)$ tale che si abbia $f(\mathcal{I}(x^0, \delta) \setminus \{x^0\}) \subset \mathcal{I}(\ell, \varepsilon)$.

¹⁰ Quindi se per ogni (a, ∞) esiste un $\mathcal{I}(x^0, \delta)$ tale che si abbia $f(\mathcal{I}(x^0, \delta) \setminus \{x^0\}) \subset (a, \infty)$.

¹¹ Quindi se per ogni $(-\infty, b)$ esiste un $\mathcal{I}(x^0, \delta)$ tale che si abbia $f(\mathcal{I}(x^0, \delta) \setminus \{x^0\}) \subset (-\infty, b)$.



2. Consideriamo ora la funzione $f(x_1, x_2) = \frac{1}{x_1^2 + x_2^2}$, definita in $\mathbb{R}^2 \setminus \{(0, 0)\}$ ¹². Vogliamo provare che

$$\lim_{(x_1, x_2) \rightarrow (0, 0)} f(x_1, x_2) = \infty.$$

Preso un qualunque $U = (a, +\infty)$, possiamo trovare le (x_1, x_2) la cui immagine sta in U resolvendo la disequazione

$$\frac{1}{x_1^2 + x_2^2} > a.$$

Ora, se $a \leq 0$, la disuguaglianza è vera in tutto $\mathbb{R}^2 \setminus \{(0, 0)\}$. Se invece $a > 0$, la disuguaglianza equivale a

$$x_1^2 + x_2^2 < \frac{1}{a},$$

che risulta soddisfatta da tutti i punti del piano la cui distanza dall'origine è minore di $\sqrt{\frac{1}{a}}$, cioè appunto un intorno circolare dell'origine. Pertanto è provato che il limite in questione è ∞ .

3. Consideriamo infine la funzione $f(x_1, x_2) = \frac{1}{x_1 + x_2}$, definita in \mathbb{R}^2 , ad eccezione dei punti che stanno sulla retta di equazione $x_1 + x_2 = 0$. Osserviamo che, volendo studiare il limite nell'origine, non siamo nella situazione ipotizzata nella definizione, dato che l'origine non appartiene al dominio di f , ma nemmeno è punto isolato del suo complementare. Per rientrare nella prima situazione, estendiamo la funzione f dicendo che sulla retta in questione la f ha, per definizione, valore nullo.

Vogliamo ora provare che la scrittura

$$\lim_{(x_1, x_2) \rightarrow (0, 0)} f(x_1, x_2) = \infty$$

è falsa. Se consideriamo $U = (0, +\infty)$, le (x_1, x_2) la cui immagine sta in U sono le soluzioni della disequazione

$$\frac{1}{x_1 + x_2} > 0, \quad \text{cioè} \quad x_1 + x_2 > 0,$$

semipiano aperto al di sopra della retta di equazione $x_1 + x_2 = 0$. Tale insieme non contiene però un intorno circolare dell'origine.

Lo studente verifichi che sono ugualmente false possibili scritture con limite finito o con limite $-\infty$, e che quindi il limite cercato non esiste.

▷ **Osservazione** Detto in modo molto impreciso ma che rende bene l'idea, il limite di una funzione di più variabili è “più difficile che esista” rispetto al limite di una funzione di una sola variabile.

A titolo di esempio, si consideri la funzione

$$f(x_1, x_2) = \frac{x_1 x_2}{x_1^2 + x_2^2}, \quad \text{con } (x_1, x_2) \neq (0, 0). \quad (1)$$

Si può verificare facilmente che

$$\lim_{(x_1, x_2) \rightarrow (0, 0)} f(x_1, x_2) \quad (2)$$

non esiste, sulla base di un risultato (che enunciamo senza dimostrazione) che appare tuttavia molto plausibile.

Se

$$\lim_{(x_1, x_2) \rightarrow (x_1^0, x_2^0)} f(x_1, x_2) \quad \text{esiste, finito o infinito,}$$

e se γ è una curva in \mathbb{R}^2 con $\gamma(t_0) = (x_1^0, x_2^0)$, allora

$$\lim_{t \rightarrow t_0} (f \circ \gamma)(t) \quad (13)$$

¹² Ricordo che, se A e B sono due insiemi, col simbolo $A \setminus B$ si indica l'insieme degli elementi che appartengono ad A e non appartengono a B .

¹³ La funzione $f \circ \gamma$ è definita per tutti i t tali che $\gamma(t)$ appartiene al campo di esistenza di f : quindi potrebbe non essere definita in t_0 .

esiste e coincide col precedente (finito o infinito rispettivamente).

In sintesi: se il limite esiste, deve essere lo stesso lungo una qualunque restrizione.

Tornando alla nostra funzione (1), consideriamo la famiglia di curve

$$\gamma_m(t) = (t, mt), \quad \text{con } m \text{ parametro e } t \in \mathbb{R}.$$

Si tratta di curve che hanno per immagine rette per l'origine, di pendenza m . Dato che ora

$$\lim_{t \rightarrow 0} (f \circ \gamma_m)(t) = \lim_{t \rightarrow 0} \frac{mt^2}{t^2 + m^2 t^2} = \frac{m}{1 + m^2},$$

il limite evidentemente dipende da m e quindi il limite (2) non può esistere.

1.4 Continuità di una funzione

Vediamo ora un concetto che collega il comportamento di f in un punto x^0 col suo comportamento negli intorno di x^0 . È richiesto ora che x^0 appartenga all'insieme di definizione A della funzione f .

Definizione Si dice che f è *continua in* x^0 se x^0 è punto isolato di A , oppure se, essendo x^0 di accumulazione per A , risulta essere

$$\lim_{x \rightarrow x^0} f(x) = f(x^0).$$

Se poi f è continua in tutti i punti di A si dirà brevemente continua in A .

Per le funzioni continue valgono in \mathbb{R}^n tutti i teoremi già visti in \mathbb{R} . A tale proposito ricordiamo che sono di fondamentale importanza tutti i teoremi che affermano che somme e prodotti di funzioni continue sono funzioni continue. Anche quozienti di funzioni continue, dove il quoziente esiste, sono funzioni continue. Infine funzioni composte di funzioni continue sono funzioni continue.

Si potrebbe dimostrare facilmente che, se $f : \mathbb{R} \rightarrow \mathbb{R}$ è una funzione continua (di una variabile) allora anche ogni funzione $g : \mathbb{R}^n \rightarrow \mathbb{R}$, con

$$g(x_1, \dots, x_n) = f(x_i), \quad 1 \leq i \leq n$$

è una funzione continua. Da quanto appena detto segue allora facilmente che tutte le funzioni “elementari” in n variabili sono continue. Ad esempio sono continue, nei rispettivi campi di esistenza, le funzioni

$$\begin{aligned} f(x_1, x_2) &= 2x_1^2 x_2 + 3x_1 x_2^3, \\ g(x_1, x_2, x_3) &= \log(x_1 - x_2 + x_3) \\ h(x_1, x_2, \dots, x_n) &= e^{-\|(x_1, x_2, \dots, x_n)\|^2}. \end{aligned}$$

Anche la funzione

$$f(x_1, x_2, \dots, x_n) = \frac{1}{\sum_{i=1}^n x_i},$$

dove esiste¹⁴, è continua.

Vale in \mathbb{R}^n un teorema che generalizza il fondamentale Teorema di Weierstrass, che lo studente ha visto per funzioni di una variabile.

Teorema 1 (di Weierstrass) Se f è definita e continua in un sottoinsieme chiuso e limitato (si suole anche dire compatto) di \mathbb{R}^n , allora valgono le seguenti proprietà:

- 1) f è limitata e la sua immagine è un insieme chiuso e limitato;¹⁵
- 2) f assume quindi il valore massimo assoluto e il valore minimo assoluto.

Se in particolare f è definita e continua in un intervallo chiuso e limitato di \mathbb{R}^n , allora

- 3) l'immagine di f è un intervallo chiuso e limitato;
- 4) f assume almeno una volta ciascuno dei valori compresi tra il minimo e il massimo assoluti.

In particolare rientra nella quarta proprietà il seguente

Teorema 2 (degli zeri) Se f è continua in un intervallo chiuso e limitato di \mathbb{R}^n ed assume in esso valori di segno opposto, essa assume almeno una volta anche il valore 0.¹⁶

¹⁴ Chiaramente la funzione esiste nei punti $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ in cui si ha $\sum_{i=1}^n x_i \neq 0$.

¹⁵ In generale non è detto che l'immagine sia un intervallo; vedi il successivo punto 3.

¹⁶ Citiamo a tal proposito il seguente risultato: se f è continua nell'insieme chiuso B , allora $f^{-1}(B)$ è un insieme chiuso. Da questo segue in particolare che l'insieme $\{x : f(x) = 0\} = f^{-1}(\{0\})$ è chiuso e anche $\{x : f(x) \leq 0\} = f^{-1}((-\infty, 0])$ è chiuso.

1.5 Continuità di funzioni definite mediante integrali

Vale la pena di fissare l'attenzione su funzioni di una variabile espresse mediante integrali di funzioni di due variabili. Per ragioni didattiche anziché usare per le due variabili i simboli x_1 e x_2 , useremo x e y .

Sia dunque $I = [a, b]$ un intervallo chiuso e limitato di \mathbb{R} e J un intervallo qualunque, sempre di \mathbb{R} ; siano poi

$$A = I \times J$$

e g una funzione continua in A ¹⁷. Si consideri, per ogni $y \in J$,

$$\int_a^b g(x, y) \, dx.$$

La scrittura non deve confondere inutilmente lo studente, che conosce per ora solo integrali di funzioni di una variabile: per ogni y fissato, la funzione g è funzione della sola variabile x , e quindi l'integrale è dello stesso tipo di quelli già incontrati nel corso di Matematica I.

È chiaro però che il valore dell'integrale dipenderà a questo punto dal valore di y : detto in altri termini, per ogni y avremo un ben determinato valore dell'integrale. Questo definisce quindi in J una legge di corrispondenza univoca, cioè una funzione in J , di variabile y . Ebbene vale il seguente

Teorema 3 Se g è continua in $A = I \times J$ allora

$$f : J \rightarrow \mathbb{R}, \quad \text{con } f(y) = \int_a^b g(x, y) \, dx$$

è continua in J .

Più generalmente, dette α e β due funzioni definite in J e tali che $\alpha(y) \leq \beta(y)$, $\forall y \in J$, considerato l'insieme

$$A = \{(x, y) \in \mathbb{R}^2 \mid \alpha(y) \leq x \leq \beta(y), y \in J\},$$

si prenda in esame la funzione

$$f(y) = \int_{\alpha(y)}^{\beta(y)} g(x, y) \, dx.$$

Teorema 4 Se α e β sono continue in J e g è continua in A , allora

$$f(y) = \int_{\alpha(y)}^{\beta(y)} g(x, y) \, dx$$

è continua in J .

▷ **Esempi**

1. La scrittura

$$\int_0^1 e^{x+y} \, dx$$

definisce una funzione in tutto \mathbb{R} (della variabile y). Con una semplice integrazione è possibile trovare di tale funzione un'espressione analitica che non faccia uso dell'integrale. Si ha

$$\int_0^1 e^{x+y} \, dx = \int_0^1 e^x e^y \, dx = e^y \int_0^1 e^x \, dx = e^y \cdot [e^x]_0^1 = (e - 1)e^y.$$

La funzione $f(y) = (e - 1)e^y$ è continua in quanto prodotto di funzioni continue. Il teorema enunciato sopra permetteva di stabilirne la continuità anche senza conoscerne l'espressione esplicita, ma solo sulla base della continuità, in \mathbb{R}^2 , della funzione e^{x+y} .

2. La scrittura

$$\int_0^1 e^{xy} \, dx$$

definisce una funzione continua in tutto \mathbb{R} (della variabile y).¹⁸ Anche in questo caso è possibile trovare un'espressione esplicita di tale funzione. Si ha

$$\int_0^1 e^{xy} \, dx = \left[\frac{e^{xy}}{y} \right]_0^1 = \frac{1}{y}(e^y - 1).$$

Da notare che l'espressione ottenuta non definisce la funzione anche in $y = 0$, ma che tale funzione è prolungabile per continuità in $y = 0$, dato che, come noto, $\lim_{y \rightarrow 0} \frac{e^y - 1}{y} = 1$.

¹⁷ g è quindi una funzione definita in un sottoinsieme di \mathbb{R}^2 , quindi una funzione di due variabili.

¹⁸ Si osservi che $f(0) = \int_0^1 dx = 1$.

3. La scrittura

$$\int_0^1 e^{x^2 y} dx$$

definisce come prima una funzione continua in tutto \mathbb{R} (della variabile y). Questa volta però non è possibile trovare, con una semplice integrazione elementare, un'espressione esplicita di tale funzione.

4. Anche la scrittura

$$\int_y^{y^2} \log(x+y) dx, \quad y \in [1, \infty)$$

definisce, in base al teorema enunciato sopra, una funzione continua, definita nell'insieme

$$A = \{(x, y) \in \mathbb{R}^2 \mid y \leq x \leq y^2, y \in [1, \infty)\}.$$

Lo studente determini un'espressione esplicita della funzione f .

1.6 Derivate parziali

Come lo studente ha visto nel corso di Matematica I, il concetto di derivata di una funzione è basato sul concetto di rapporto incrementale. Volendo costruire anche per funzioni di $n > 1$ variabili un rapporto incrementale, cioè un quoziente di variazioni, ci si imbatte in una difficoltà. A partire da una n -upla $x^0 = (x_1^0, \dots, x_n^0)$ ¹⁹ si può sempre darle una variazione passando alla n -pla $x = (x_1, \dots, x_n)$. Corrispondentemente la funzione subisce la variazione $f(x) - f(x^0)$, che è ovviamente un numero reale. Però non si può poi costruire il quoziente tra la variazione della funzione e la variazione della n -upla di variabili, perché quest'ultima non è un numero.

Ci sono alcuni modi per costruire in questa nuova situazione uno strumento simile alla derivata per funzioni di una variabile. Uno di questi è l'introduzione delle cosiddette *derivate parziali*.

Per facilitare la comprensione della definizione riducendo al minimo la pesantezza delle notazioni dovuta al numero di variabili, presentiamo intanto le derivate parziali di funzioni di due variabili.

Sia allora $x^0 = (x_1^0, x_2^0)$. Diamo una variazione alla sola variabile x_1 , considerando, in un intorno di x^0 , il punto (x_1, x_2^0) . Costruiamo il *rapporto incrementale parziale di f rispetto ad x_1*

$$\frac{f(x_1, x_2^0) - f(x_1^0, x_2^0)}{x_1 - x_1^0}.$$

Facciamo infine il limite

$$\lim_{x_1 \rightarrow x_1^0} \frac{f(x_1, x_2^0) - f(x_1^0, x_2^0)}{x_1 - x_1^0}.$$

Se tale limite esiste finito, lo chiamiamo *derivata parziale di f rispetto ad x_1 nel punto x^0* e la indichiamo con uno dei due simboli $\frac{\partial f}{\partial x_1}(x^0)$ oppure $f'_{x_1}(x^0)$. Diciamo anche che f è *derivabile parzialmente rispetto ad x_1 nel punto x^0* .

Analogamente definiamo il

$$\lim_{x_2 \rightarrow x_2^0} \frac{f(x_1^0, x_2) - f(x_1^0, x_2^0)}{x_2 - x_2^0},$$

se esiste finito, la *derivata parziale di f rispetto ad x_2 nel punto x^0* e la indichiamo con $\frac{\partial f}{\partial x_2}(x^0)$ oppure $f'_{x_2}(x^0)$, dicendo che f è *derivabile parzialmente rispetto ad x_2 nel punto x^0* .

Dicendo semplicemente che una funzione f è derivabile in x^0 si intende che essa è derivabile parzialmente rispetto ad x_1 e rispetto ad x_2 in x^0 .

In tal caso si chiama *gradiente di f in x^0* il vettore delle sue derivate parziali. Si scrive

$$\nabla f(x^0) = (f'_{x_1}(x^0), f'_{x_2}(x^0)).$$

In generale, per una funzione di n variabili, avremo n possibili derivate parziali. A partire dal punto $x^0 = (x_1^0, \dots, x_n^0)$ diamo una variazione ad *una sola* delle n variabili, diciamo la x_i , passando dal valore x_i^0 al valore x_i . Questa variazione è $x_i - x_i^0$ ed è un numero, cui corrisponde una variazione della funzione pari a

$$f(x_1^0, \dots, x_{i-1}^0, x_i, x_{i+1}^0, \dots, x_n^0) - f(x_1^0, \dots, x_{i-1}^0, x_i^0, x_{i+1}^0, \dots, x_n^0).$$

Costruiamo quindi il rapporto incrementale

$$\frac{f(x_1^0, \dots, x_{i-1}^0, x_i, x_{i+1}^0, \dots, x_n^0) - f(x_1^0, \dots, x_{i-1}^0, x_i^0, x_{i+1}^0, \dots, x_n^0)}{x_i - x_i^0},$$

¹⁹ Supporremo che il punto sia interno al dominio della funzione.

che viene chiamato rapporto incrementale parziale rispetto ad x_i . Si considera poi il

$$\lim_{x_i \rightarrow x_i^0} \frac{f(x_1^0, \dots, x_{i-1}^0, x_i, x_{i+1}^0, \dots, x_n^0) - f(x_1^0, \dots, x_{i-1}^0, x_i^0, x_{i+1}^0, \dots, x_n^0)}{x_i - x_i^0}.$$

Se esso esiste finito, lo si dice *derivata parziale di f rispetto a x_i nel punto x^0* , e si dirà che f è *derivabile parzialmente rispetto a x_i nel punto x^0* .

Si usa uno dei simboli

$$\frac{\partial f}{\partial x_i}(x^0) \quad \text{oppure} \quad f'_{x_i}(x^0).$$

Se nel punto x^0 la funzione f è derivabile parzialmente rispetto a tutte le sue variabili si dirà brevemente che è derivabile nel punto x^0 .

Se f è una funzione derivabile nel punto x^0 , il vettore delle n derivate parziali in x^0 viene detto *gradiente di f in x^0* . Viene indicato col simbolo $\nabla f(x^0)$. Pertanto

$$\nabla f(x^0) = (f'_{x_1}(x^0), \dots, f'_{x_n}(x^0)).$$

▷ Esempi

1. Consideriamo la funzione

$$f(x_1, x_2) = x_1 + e^{x_2},$$

definita in tutto \mathbb{R}^2 . Consideriamo poi il punto $x^0 = (1, 0)$, in cui la funzione vale 2.

Si ha

$$\frac{\partial f}{\partial x_1}(1, 0) = \lim_{x_1 \rightarrow x_1^0} \frac{f(x_1, x_2^0) - f(x_1^0, x_2^0)}{x_1 - x_1^0} = \lim_{x_1 \rightarrow 1} \frac{f(x_1, 0) - f(1, 0)}{x_1 - 1} = \lim_{x_1 \rightarrow 1} \frac{x_1 + 1 - 2}{x_1 - 1} = 1.$$

Si ha poi

$$\frac{\partial f}{\partial x_2}(1, 0) = \lim_{x_2 \rightarrow x_2^0} \frac{f(x_1^0, x_2) - f(x_1^0, x_2^0)}{x_2 - x_2^0} = \lim_{x_2 \rightarrow 0} \frac{f(1, x_2) - f(1, 0)}{x_2} = \lim_{x_2 \rightarrow 0} \frac{1 + e^{x_2} - 2}{x_2} = 1.$$

Quindi f è derivabile nel punto $(1, 0)$ e $\nabla f(1, 0) = (1, 1)$.

2. Si consideri la funzione

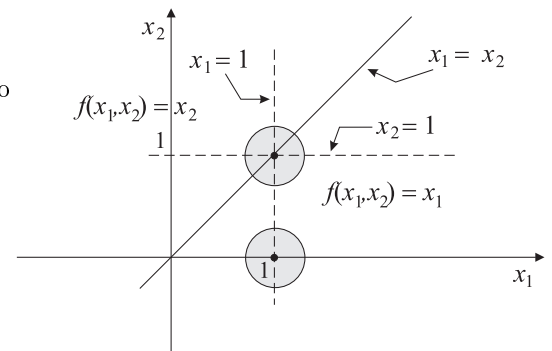
$$f(x_1, x_2) = \max(x_1, x_2),$$

definita in tutto \mathbb{R}^2 .

Per facilitare la comprensione dei calcoli che seguono, possiamo riscrivere la f come

$$f(x_1, x_2) = \begin{cases} x_1 & \text{se } x_1 \geq x_2 \\ x_2 & \text{se } x_2 \geq x_1 \end{cases}$$

e rappresentiamo quanto scritto nella figura a fianco.



Consideriamo ora il punto $x^0 = (1, 1)$, in cui si ha $f(x^0) = 1$. Dare una variazione solo ad x_1 vuol dire muoversi lungo una retta passante per x^0 e parallela all'asse x_1 (tale retta ha equazione $x_2 = 1$). Su di essa, nell'intorno destro di x^0 , la funzione vale x_1 , mentre nell'intorno sinistro vale 1. Si ha perciò

$$\lim_{x_1 \rightarrow 1^+} \frac{x_1 - 1}{x_1 - 1} = \lim_{x_1 \rightarrow 1^+} 1 = 1,$$

$$\lim_{x_1 \rightarrow 1^-} \frac{1 - 1}{x_1 - 1} = \lim_{x_1 \rightarrow 1^-} 0 = 0,$$

e pertanto questa funzione non ammette derivata parziale rispetto ad x_1 nel punto $(1, 1)$. Ciò è sufficiente perchè si possa dire che in quel punto non è derivabile.

Lo studente provi a studiare la derivabilità parziale rispetto ad x_2 .

Si consideri ora il punto $x^0 = (1, 0)$. Facendo variare solo la x_1 e percorrendo quindi l'asse x_1 si trova che sia a destra sia a sinistra, in un intorno sufficientemente piccolo di x^0 , la funzione vale x_1 (si veda ancora la figura sopra). Si ha quindi

$$\lim_{x_1 \rightarrow 1^+} \frac{x_1 - 1}{x_1 - 1} = \lim_{x_1 \rightarrow 1^+} 1 = 1,$$

$$\lim_{x_1 \rightarrow 1^-} \frac{x_1 - 1}{x_1 - 1} = \lim_{x_1 \rightarrow 1^-} 1 = 1,$$

e quindi la funzione è derivabile parzialmente rispetto a x_1 in x^0 . Percorrendo invece una parallela all'asse x_2 si trova ancora che tanto a destra quanto a sinistra di 0 in un intorno sufficientemente piccolo la funzione vale 1. Si ha quindi

$$\lim_{x_2 \rightarrow 0^+} \frac{1 - 1}{x_2 - 0} = 0,$$

$$\lim_{x_2 \rightarrow 0^-} \frac{1 - 1}{x_2 - 0} = 0,$$

e quindi la funzione in $(1, 0)$ è derivabile parzialmente anche rispetto ad x_2 ed è pertanto derivabile. Si ha $\nabla f(1, 0) = (1, 0)$.

▷ **Osservazione** Studiare la derivabilità parziale rispetto ad x_1 di una funzione f in un punto $x^0 = (x_1^0, x_2^0)$ equivale a studiare la derivabilità in $t = 0$ della restrizione di f alla curva $\gamma_1(t) = (x_1^0 + t, x_2^0)$.

Analogamente studiare la derivabilità parziale rispetto ad x_2 in $x^0 = (x_1^0, x_2^0)$ equivale a studiare la derivabilità in $t = 0$ della restrizione di f alla curva $\gamma_2(t) = (x_1^0, x_2^0 + t)$.

Anzi, si può vedere facilmente che, se c'è derivabilità parziale, allora risulta

$$\frac{\partial f}{\partial x_1}(x^0) = (f \circ \gamma_1)'(0) \quad \text{e} \quad \frac{\partial f}{\partial x_2}(x^0) = (f \circ \gamma_2)'(0).$$

Ad esempio, riprendendo la funzione del punto 1 degli esempi precedenti, cioè la

$$f(x_1, x_2) = x_1 + e^{x_2}, \quad \text{nel punto } x^0 = (1, 0),$$

per la derivata parziale rispetto ad x_1 consideriamo la curva $\gamma_1(t) = (1 + t, 0)$ e la restrizione

$$f(1 + t, 0) = 2 + t.$$

Tale funzione è derivabile in $t = 0$ con derivata uguale ad 1.

Per la derivata parziale rispetto ad x_2 consideriamo invece la curva $\gamma_2(t) = (1, t)$ e la restrizione

$$f(1, t) = 1 + e^t,$$

che è derivabile in $t = 0$ con derivata uguale ad 1.

Per la funzione del punto 2, cioè

$$f(x_1, x_2) = \max(x_1, x_2), \quad \text{nel punto } x^0 = (1, 1),$$

si ha invece

$$f(1 + t, 1) = \max(1 + t, 1) = \begin{cases} 1 & \text{se } t \leq 0 \\ 1 + t & \text{se } t > 0 \end{cases}$$

e tale funzione non è derivabile in $t = 0$. Quindi f non è derivabile parzialmente rispetto ad x_1 .

Nemmeno rispetto ad x_2 c'è derivata parziale, dato che la restrizione alla curva $(1, 1 + t)$ coincide con la restrizione alla curva $(1 + t, 1)$.²⁰

Ancora un esempio. Per la funzione

$$f(x_1, x_2) = |x_1 + x_2^2|, \quad \text{nel punto } x^0 = (0, 0),$$

l'esame della derivabilità parziale rispetto ad x_1 porta a considerare la restrizione

$$f(t, 0) = |t|,$$

che non è derivabile in $t = 0$. L'esame della derivabilità parziale rispetto ad x_2 porta invece a considerare la restrizione

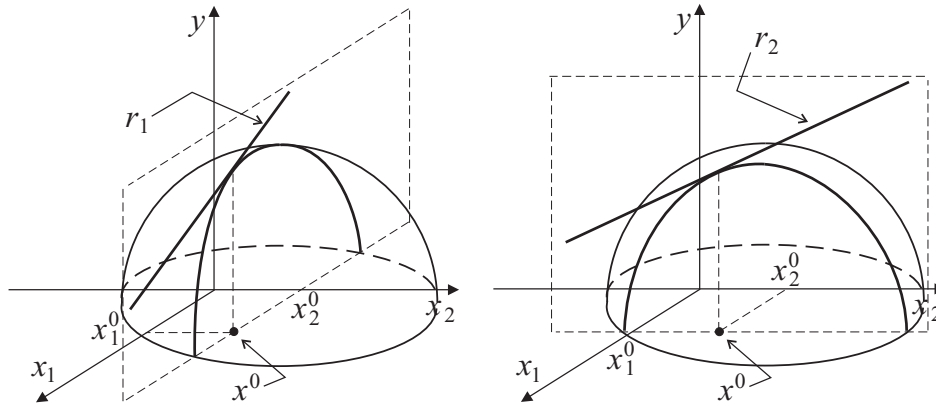
$$f(0, t) = |t^2| = t^2,$$

²⁰ Si osservi che le funzioni componenti sono diverse, ma la funzione composta è la stessa, cioè la funzione, definita su \mathbb{R} , da $g(t) = \max(1 + t, 1)$.

che è derivabile in $t = 0$, con derivata nulla. Pertanto si ha $\frac{\partial f}{\partial x_2}(0, 0) = 0$, mentre la $\frac{\partial f}{\partial x_1}(0, 0)$ non esiste.

Spendiamo ora due parole sull'interpretazione geometrica delle derivate parziali. Tenendo conto di quanto esposto nell'ultima osservazione, che cioè le derivate parziali coincidono con le derivate (monodimensionali) lungo opportune restrizioni, si intuisce facilmente che, se la f è derivabile rispetto ad x_1 nel punto $x^0 = (x_1^0, x_2^0)$, allora nel piano di equazione $x_2 = x_2^0$ ²¹ vi è una (ed una sola) retta tangente al grafico di f (vedi figura sotto a sinistra). La derivata parziale di f rispetto ad x_1 è la pendenza (coefficiente angolare) di tale retta, indicata con r_1 in figura.

Analogamente, se f è derivabile rispetto ad x_2 in x^0 , allora nel piano di equazione $x_1 = x_1^0$ vi è una (ed una sola) retta tangente al grafico di f . La derivata parziale di f rispetto ad x_2 è la pendenza (coefficiente angolare) di questa retta, indicata con r_2 in figura.



Per quanto riguarda il calcolo delle derivate, come avviene per le funzioni di una sola variabile, la definizione (cioè il limite del rapporto incrementale) si utilizza solamente in casi particolari, come ad esempio le funzioni definite a tratti, o perché espressamente richiesta.²² Per il calcolo ci sono metodi più comodi per procedere, quelli che vengono comunemente detti regole di derivazione. Si ottengono comode regole di derivazione parziale semplicemente tenendo conto del fatto che nella derivazione rispetto alla variabile x_i tutte le altre variabili sono da mantenersi costanti: quindi le regole di calcolo sono le consuete regole di derivazione delle funzioni di una sola variabile.²³

▷ Esempi

- Volendo la derivata parziale della funzione $f(x_1, x_2) = x_1^2 x_2 e^{x_1 x_2^2}$ rispetto ad x_1 , basterà ritenere costante x_2 : si ottiene

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = 2x_1 x_2 e^{x_1 x_2^2} + x_1^2 x_2 e^{x_1 x_2^2} \cdot x_2^2 = 2x_1 x_2 e^{x_1 x_2^2} + x_1^2 x_2^3 e^{x_1 x_2^2}.$$

Mantenendo costante invece x_1 si ottiene

$$\frac{\partial f}{\partial x_2}(x_1, x_2) = x_1^2 e^{x_1 x_2^2} + x_1^2 x_2 e^{x_1 x_2^2} \cdot 2x_1 x_2 = x_1^2 e^{x_1 x_2^2} + 2x_1^3 x_2^2 e^{x_1 x_2^2}.$$

- Con la funzione $f(x_1, x_2) = \frac{x_1}{x_2}$, derivando rispetto ad x_1 si ha

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = \frac{1}{x_2}.$$

Derivando rispetto ad x_2 si ha invece

$$\frac{\partial f}{\partial x_2}(x_1, x_2) = -\frac{x_1}{x_2^2}.$$

- Con la funzione $f(x_1, x_2) = \ln \frac{x_1}{\ln x_2}$, derivando rispetto ad x_1 si ha

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = \frac{\ln x_2}{x_1} \cdot \frac{1}{\ln x_2} = \frac{1}{x_1}.$$

Derivando rispetto ad x_2 si ha invece

$$\frac{\partial f}{\partial x_2}(x_1, x_2) = \frac{\ln x_2}{x_1} \cdot \left(-\frac{x_1}{\ln^2 x_2} \right) \cdot \frac{1}{x_2} = -\frac{1}{x_2 \ln x_2}.$$

²¹ Si tratta di un piano "verticale" parallelo al piano x_1, y e che contiene il punto x^0 .

²² Solitamente per capire se lo studente ha studiato bene anche la teoria.

²³ Sarebbe come derivare una funzione di una variabile che dipende anche da alcuni parametri, i quali però sono da ritenersi costanti al momento della derivazione.

- Con la funzione $f(x_1, x_2) = \sin(x_1 \cos x_2)$ si ha

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = \cos(x_1 \cos x_2) \cos x_2$$

e

$$\frac{\partial f}{\partial x_2}(x_1, x_2) = -x_1 \cos(x_1 \cos x_2) \sin x_2.$$

- Con la funzione $f(x_1, x_2, x_3) = \sqrt{x_1 + \sqrt{x_2 + \sqrt{x_3}}}$, si ha

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = \frac{1}{2\sqrt{x_1 + \sqrt{x_2 + \sqrt{x_3}}}},$$

$$\frac{\partial f}{\partial x_2}(x_1, x_2) = \frac{1}{2\sqrt{x_1 + \sqrt{x_2 + \sqrt{x_3}}}} \cdot \frac{1}{2\sqrt{x_2 + \sqrt{x_3}}}$$

e

$$\frac{\partial f}{\partial x_3}(x_1, x_2) = \frac{1}{2\sqrt{x_1 + \sqrt{x_2 + \sqrt{x_3}}}} \cdot \frac{1}{2\sqrt{x_2 + \sqrt{x_3}}} \cdot \frac{1}{2\sqrt{x_3}}.$$

È da mettere in evidenza il fatto che la derivabilità di una funzione nel vettore x^0 non implica la continuità in x^0 . Un esempio in merito è fornito dalla funzione

$$f(x_1, x_2) = \begin{cases} \frac{x_1 x_2}{x_1^2 + x_2^2} & (x_1, x_2) \neq (0, 0) \\ 0 & (x_1, x_2) = (0, 0). \end{cases}$$

Lo studente verifichi che la funzione f è derivabile in $x^0 = (0, 0)$ con derivate nulle, ma non è continua in x^0 . Un altro esempio può essere più semplicemente la funzione

$$f(x_1, x_2) = \begin{cases} 0 & x_1 x_2 = 0 \\ 1 & x_1 x_2 \neq 0. \end{cases}$$

Le restrizioni agli assi x_1 e x_2 della funzione f sono funzioni identicamente nulle, quindi la funzione f è derivabile parzialmente in $x^0 = (0, 0)$ con derivate parziali entrambe nulle. Non c'è però continuità nel punto x^0 , dato che il limite

$$\lim_{(x_1, x_2) \rightarrow (0, 0)} f(x_1, x_2) \quad \text{non esiste.}$$

Infatti ad esempio la restrizione $f(t, 0)$ ha limite 0 per $t \rightarrow 0$, mentre la restrizione $f(t, t)$ ha limite 1 per $t \rightarrow 0$.

Il motivo del fatto che la derivabilità parziale non implichi la continuità²⁴ è facilmente intuibile: l'esistenza delle derivate parziali riguarda il comportamento della funzione soltanto lungo le direzioni parallele agli assi, mentre la continuità in x^0 coinvolge tutto un intorno di x^0 . Quindi la derivabilità lungo tali direzioni particolari non mi dà sufficienti garanzie sul comportamento della funzione lungo altre possibili direzioni.

1.7 Derivate seconde

Se $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ e se D è l'insieme dei punti di A nei quali f è derivabile, ogni f'_{x_i} è una funzione definita in D a valori in \mathbb{R} , che associa ad ogni x appartenente a D il numero reale $f'_{x_i}(x)$. Tali funzioni possono quindi a loro volta essere derivabili parzialmente rispetto alle variabili x_1, \dots, x_n .

Ciascuna f'_{x_i} può avere dunque n derivate parziali seconde; la derivata parziale seconda di f'_{x_i} rispetto ad x_j viene indicata con i simboli

$$f''_{x_i x_j} \quad \text{oppure} \quad \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

Vi sono dunque nel complesso n^2 derivate parziali seconde. Vi sono quelle ottenute derivando due volte rispetto alla stessa variabile, e cioè

$$f''_{x_1 x_1}, \dots, f''_{x_n x_n} \quad \text{oppure} \quad \frac{\partial^2 f}{\partial x_1^2}, \dots, \frac{\partial^2 f}{\partial x_n^2}$$

e quelle ottenute invece derivando rispetto a variabili diverse, e cioè

$$f''_{x_i x_j}, \quad \text{con } i, j = 1, \dots, n \text{ e } i \neq j$$

²⁴ Si ricordi che invece, per funzioni di una variabile, la derivabilità in un punto x^0 implica la continuità in x^0 .

oppure

$$\frac{\partial^2 f}{\partial x_i \partial x_j}, \quad \text{con } i, j = 1, \dots, n \text{ e } i \neq j.$$

Queste ultime si sogliono chiamare *derivate seconde miste*.

▷ *Esempi*

- La funzione $f(x_1, x_2) = x_1^2 + x_2^3$ ha derivate parziali prime

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = 2x_1 \quad \text{e} \quad \frac{\partial f}{\partial x_2}(x_1, x_2) = 3x_2^2$$

e derivate parziali seconde

$$\frac{\partial^2 f}{\partial x_1^2}(x_1, x_2) = 2 \quad , \quad \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1, x_2) = 0 \quad , \quad \frac{\partial^2 f}{\partial x_2 \partial x_1}(x_1, x_2) = 0 \quad , \quad \frac{\partial^2 f}{\partial x_2^2}(x_1, x_2) = 6x_2.$$

- La funzione $f(x_1, x_2) = \frac{x_1}{x_2}$ ha derivate parziali prime

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = \frac{1}{x_2} \quad \text{e} \quad \frac{\partial f}{\partial x_2}(x_1, x_2) = -\frac{x_1}{x_2^2}$$

e derivate parziali seconde

$$\frac{\partial^2 f}{\partial x_1^2}(x_1, x_2) = 0 \quad , \quad \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1, x_2) = -\frac{1}{x_2^2} \quad , \quad \frac{\partial^2 f}{\partial x_2 \partial x_1}(x_1, x_2) = -\frac{1}{x_2^2} \quad , \quad \frac{\partial^2 f}{\partial x_2^2}(x_1, x_2) = \frac{2x_1}{x_2^3}.$$

- La funzione $f(x_1, x_2) = \sin(x_1 \cos x_2)$ ha derivate parziali prime

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = \cos(x_1 \cos x_2) \cos x_2 \quad \text{e} \quad \frac{\partial f}{\partial x_2}(x_1, x_2) = -x_1 \cos(x_1 \cos x_2) \sin x_2$$

e derivate parziali seconde

$$\begin{aligned} \frac{\partial^2 f}{\partial x_1^2}(x_1, x_2) &= \cos x_2 (-\sin(x_1 \cos x_2)) \cos x_2, \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1, x_2) &= -\sin(x_1 \cos x_2) \cdot (-x_1 \sin x_2) \cos x_2 + \cos(x_1 \cos x_2) (-\sin x_2), \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x_1, x_2) &= -\sin x_2 \cos(x_1 \cos x_2) - x_1 \sin x_2 (-\sin(x_1 \cos x_2)) \cos x_2, \\ \frac{\partial^2 f}{\partial x_2^2}(x_1, x_2) &= -x_1 \cos x_2 \cos(x_1 \cos x_2) - x_1 \sin x_2 (-\sin(x_1 \cos x_2)) (x_1 \sin x_2). \end{aligned}$$

Lo studente avrà notato che in tutte le funzioni degli esempi proposti le derivate seconde miste sono uguali. Vale in proposito il seguente importante risultato.

Teorema 5 (di Schwarz) Data la funzione f , definita in $A \subset \mathbb{R}^n$ e derivabile parzialmente due volte nell'aperto $D \subseteq A$, se le derivate seconde miste $f''_{x_i x_j}$ e $f''_{x_j x_i}$ sono continue in D , esse sono uguali tra loro.

Lo studente vedrà nel corso di Matematica 2 come le derivate parziali prime di una funzione possano essere utili nella ricerca dei punti di massimo e di minimo e come invece le derivate parziali seconde siano importanti per la convessità o concavità della funzione.

2 Spazi vettoriali

Nel seguito un \square sul margine destro indica la fine di una dimostrazione.

2.1 Spazi vettoriali

Iniziamo prendendo in considerazione l'insieme i cui elementi sono le n -uple ordinate di numeri reali, dove n è un numero naturale, con $n \geq 1$. Si tratta dell'insieme, che indicherò con \mathbb{R}^n , definito da

$$\mathbb{R}^n = \{(x_1, x_2, \dots, x_n) : x_i \in \mathbb{R}, 1 \leq i \leq n\}.$$

Se (x_1, x_2, \dots, x_n) è un elemento di \mathbb{R}^n , lo si chiama anche *vettore*. I numeri reali x_i , per $i = 1, 2, \dots, n$, si dicono le *componenti* del vettore.

Gli elementi di \mathbb{R}^n , qualora non sia indispensabile fare riferimento alle loro singole componenti, saranno indicati con lettere minuscole: così, scrivendo $x \in \mathbb{R}^n$, intenderò che $x = (x_1, x_2, \dots, x_n)$.

Talvolta nel seguito mi sarà utile considerare un certo numero, diciamo k , di vettori, tutti appartenenti ad \mathbb{R}^n . Li indicherò allora ad esempio con v^1, v^2, \dots, v^k , e metto in guardia lo studente a non confondere le due scritture

$$(x_1, x_2, \dots, x_n) : x_i \in \mathbb{R} \quad \text{e} \quad v^1, v^2, \dots, v^k : v^i \in \mathbb{R}^n.$$

La prima si riferisce ad *un* vettore di n componenti (quindi un elemento di \mathbb{R}^n), mentre la seconda sta ad indicare k vettori di \mathbb{R}^n .

Ovviamente, se ho la necessità di riferirmi alla i -esima componente del j -esimo vettore, userò la scrittura v_i^j .

Nell'insieme \mathbb{R}^n possiamo definire le seguenti due operazioni, che chiamiamo rispettivamente *addizione* e *moltiplicazione per gli scalari* o più semplicemente *moltiplicazione scalare*.

L'addizione viene definita in questo modo: se $x = (x_1, x_2, \dots, x_n)$ e $y = (y_1, y_2, \dots, y_n)$ sono due elementi di \mathbb{R}^n , poniamo

$$x + y = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n)$$

e ovviamente la somma dei due vettori è ancora un vettore di \mathbb{R}^n .

La moltiplicazione per uno scalare, cioè per un numero reale, è definita così: se $x = (x_1, x_2, \dots, x_n)$ è un elemento di \mathbb{R}^n e α è un numero reale, allora

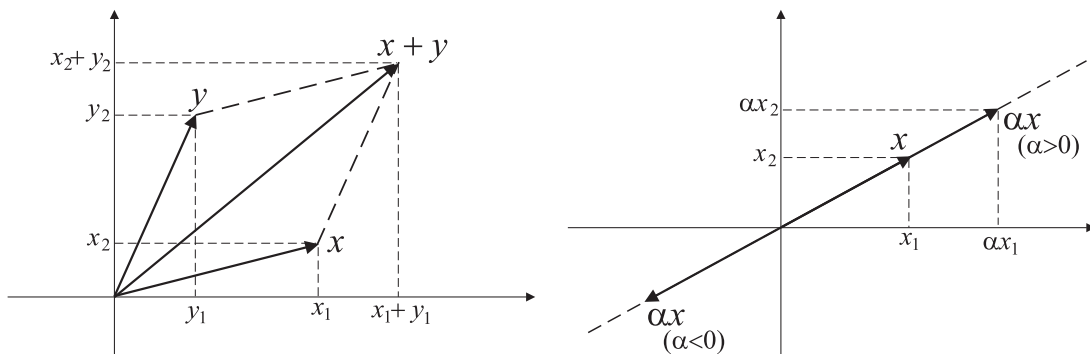
$$\alpha x = (\alpha x_1, \alpha x_2, \dots, \alpha x_n),$$

dove ovviamente αx_i è il prodotto (in \mathbb{R}) dei due numeri reali α e x_i , per ogni $i = 1, 2, \dots, n$.

Si può facilmente verificare che le operazioni appena definite hanno le seguenti proprietà.

- 1a. $x + y = y + x, \forall x, y \in \mathbb{R}^n$ (proprietà commutativa dell'addizione);
- 1b. $(x + y) + z = x + (y + z), \forall x, y, z \in \mathbb{R}^n$ (proprietà associativa dell'addizione);
- 1c. esiste un vettore (è il vettore nullo e si indica con 0) tale che $x + 0 = x, \forall x \in \mathbb{R}^n$;
- 1d. ogni vettore $x \in \mathbb{R}^n$ ha un opposto (si indica con $-x$) tale che $x + (-x) = 0$;
- 2a. $\alpha(\beta x) = (\alpha\beta)x, \forall x \in \mathbb{R}^n$ e $\forall \alpha, \beta \in \mathbb{R}$ (proprietà associativa della moltiplicazione per gli scalari);
- 2b. $1x = x, \forall x \in \mathbb{R}^n$.
- 3a. $\alpha(x + y) = \alpha x + \alpha y, \forall x, y \in \mathbb{R}^n$ e $\forall \alpha \in \mathbb{R}$ (proprietà distributiva della moltiplicazione scalare rispetto all'addizione di \mathbb{R}^n);
- 3b. $(\alpha + \beta)x = \alpha x + \beta x, \forall x \in \mathbb{R}^n$ e $\forall \alpha, \beta \in \mathbb{R}$ (proprietà distributiva della moltiplicazione scalare rispetto all'addizione di \mathbb{R});

▷ **Osservazione** L'interpretazione geometrica, in \mathbb{R}^2 , dell'addizione tra vettori e della moltiplicazione di un vettore per uno scalare è illustrata nelle figure qui sotto.



Diamo ora una definizione generale.

Definizione Si definisce *spazio vettoriale su \mathbb{R}* un insieme \mathbb{V} , i cui elementi sono detti *vettori*, con le seguenti proprietà:

1. c'è in \mathbb{V} un'operazione, detta *addizione vettoriale*, che ad ogni $x, y \in \mathbb{V}$ associa il vettore $x + y \in \mathbb{V}$, in modo che
 - 1a. l'addizione vettoriale è *commutativa*, cioè $x + y = y + x$, $\forall x, y \in \mathbb{V}$;
 - 1b. l'addizione vettoriale è *associativa*, cioè $x + (y + z) = (x + y) + z$, $\forall x, y, z \in \mathbb{R}^n$;
 - 1c. c'è in \mathbb{V} un vettore, indicato con 0 e detto *origine*, tale che $x + 0 = x$, $\forall x \in \mathbb{V}$;²⁵
 - 1d. ogni vettore $x \in \mathbb{V}$ ha un *opposto*, che si indica con $-x$, tale che $x + (-x) = 0$;
2. c'è in \mathbb{V} una seconda operazione, detta *moltiplicazione per gli scalari*, che ad ogni $x \in \mathbb{V}$ e ad ogni $\alpha \in \mathbb{R}$ associa il vettore $\alpha x \in \mathbb{V}$, in modo che
 - 2a. la moltiplicazione per gli scalari è associativa, cioè $\alpha(\beta x) = (\alpha\beta)x$, $\forall \alpha, \beta \in \mathbb{R}$ e $\forall x \in \mathbb{V}$;
 - 2b. $1x = x$, $\forall x \in \mathbb{V}$;
3. inoltre valgono le seguenti proprietà
 - 3a. la moltiplicazione scalare è *distributiva rispetto all'addizione vettoriale*, cioè $\alpha(x + y) = \alpha x + \alpha y$, $\forall \alpha \in \mathbb{R}$ e $\forall x, y \in \mathbb{V}$;
 - 3b. la moltiplicazione scalare è *distributiva rispetto all'addizione in \mathbb{R}* , cioè $(\alpha + \beta)x = \alpha x + \beta x$, $\forall \alpha, \beta \in \mathbb{R}$ e $\forall x \in \mathbb{V}$.

▷ **Osservazione** Si osservi che in ogni spazio vettoriale \mathbb{V} , se $x \in \mathbb{V}$, allora $0x = 0$: infatti $0x = (0 + 0)x = 0x + 0x$ e da questa l'esistenza degli opposti porta a $0x = 0$. Inoltre $\alpha 0 = 0$, dato che $\alpha 0 = \alpha(0 + 0) = \alpha 0 + \alpha 0$ e da questa ancora l'esistenza degli opposti porta a $\alpha 0 = 0$.

Si può anche provare l'unicità dell'opposto e l'unicità dell'origine.

▷ **Esempi**

- Uno spazio vettoriale banale è quello costituito dal solo vettore nullo, cioè $\{0\}$;
- \mathbb{R} è spazio vettoriale su \mathbb{R} , se adottiamo quali addizione vettoriale e moltiplicazione per gli scalari le usuali addizione e moltiplicazione tra numeri reali;
- L'insieme \mathbb{P} di tutti i polinomi in una variabile, a coefficienti reali, è uno spazio vettoriale su \mathbb{R} , con l'addizione vettoriale data dall'usuale addizione tra polinomi e la moltiplicazione per gli scalari data dall'usuale moltiplicazione di un polinomio per una costante reale;
- Come già visto, \mathbb{R}^n è spazio vettoriale su \mathbb{R} (per ogni n), con le operazioni definite sopra;
- Per ogni numero naturale $n \geq 1$, l'insieme \mathbb{P}_n dei polinomi di grado minore o uguale di n in una variabile, a coefficienti reali, è uno spazio vettoriale su \mathbb{R} ;²⁶
- L'insieme delle funzioni reali limitate, definite in un intervallo I di \mathbb{R} è uno spazio vettoriale su \mathbb{R} , se adottiamo quale addizione vettoriale l'usuale addizione tra funzioni reali e quale moltiplicazione per gli scalari l'usuale moltiplicazione tra una funzione reale ed una costante. Ricordo che, se f e g sono due funzioni definite in I e se $\alpha \in \mathbb{R}$, con funzione somma si intende la funzione

$$(f + g)(x) = f(x) + g(x), \quad \text{con } x \in I$$

e con prodotto di f per α si intende la funzione

$$(\alpha f)(x) = \alpha f(x), \quad \text{con } x \in I.$$

▷ **Esercizi**

1. Sia \mathbb{V} uno spazio vettoriale su \mathbb{R} , siano $x, y \in \mathbb{V}$ e sia $\alpha \in \mathbb{R}$. Si provi che

- $-0 = 0$;
- $\alpha 0 = 0$;
- $0x = 0$;
- se $\alpha x = 0$, allora $\alpha = 0$ oppure $x = 0$ (o entrambi);
- $-x = (-1)x$;
- $y + (x + (-y)) = x$.

²⁵ Lo studente non confonda a questo punto lo 0 che è vettore origine di \mathbb{V} con lo 0 che è il numero reale 0 .

²⁶ Intendiamo che tra questi polinomi ci sia anche il polinomio nullo.

2. Sia \mathbb{P} lo spazio vettoriale su \mathbb{R} dei polinomi e sia $\mathbb{S} \subset \mathbb{P}$ formato dai polinomi x tali che

- x ha grado 2;
- $2x(0) = x(1)$;
- $x(t) \geq 0$ se $0 \leq t \leq 1$;
- $x(t) = x(1-t)$, $\forall t$.

In quali di questi casi \mathbb{S} è spazio vettoriale?

Le definizioni che seguono sono fondamentali.

Definizione Sia \mathbb{V} uno spazio vettoriale su \mathbb{R} . Siano dati in \mathbb{V} i vettori

$$v^1, v^2, \dots, v^k \quad \text{con} \quad k \geq 1.$$

Si chiama *combinazione lineare*, d'ora in avanti c.l., dei vettori v^1, v^2, \dots, v^k un qualunque vettore del tipo

$$\alpha_1 v^1 + \alpha_2 v^2 + \dots + \alpha_k v^k,$$

dove gli α_i , per ogni $i = 1, 2, \dots, k$, sono numeri reali. Questi numeri reali si dicono i *coefficienti* della c.l.

Se si esclude lo spazio vettoriale banale $\{0\}$, dati in \mathbb{V} i vettori v^1, v^2, \dots, v^k con $k \geq 1$, di combinazioni lineari ne esistono infinite. Una di queste è quella che si ottiene scegliendo tutti i coefficienti nulli (così facendo ottengo, indipendentemente da quali siano i vettori dati, il vettore nullo).

Ad esempio, in \mathbb{R}^3 , dati i vettori

$$v^1 = (1, 0, -1) \quad , \quad v^2 = (1, -1, 0) \quad , \quad v^3 = (0, 0, 2) \quad , \quad v^4 = (-1, -1, 1) \quad ,$$

una loro c.l. è il vettore $v = (4, -1, -2)$, in quanto $v = v^1 + 2v^2 - v^4$.

2.2 Dipendenza e indipendenza lineare

Definizione Nello spazio vettoriale \mathbb{V} , i vettori v^1, v^2, \dots, v^k si dicono *linearmente dipendenti*, d'ora in avanti l.d., se almeno uno di essi si può scrivere come c.l. dei rimanenti.

Definizione I vettori v^1, v^2, \dots, v^k si dicono *linearmente indipendenti*, d'ora in avanti l.i., se nessuno di essi si può scrivere come c.l. dei rimanenti.

▷ **Esempio** In \mathbb{R}^2 i vettori

$$v^1 = (0, 2) \quad , \quad v^2 = (1, 1) \quad , \quad v^3 = (-2, 0)$$

sono l.d. in quanto $v^3 = v^1 - 2v^2$.

In \mathbb{R}^3 i vettori

$$v^1 = (1, 0, 1) \quad , \quad v^2 = (0, 1, 0) \quad , \quad v^3 = (0, 0, 1)$$

sono l.i. Infatti v^1 non è c.l. degli altri due perchè gli altri due hanno nulla la prima componente, v^2 non è c.l. degli altri due perchè gli altri due hanno nulla la seconda componente, e infine v^3 non è c.l. degli altri due perchè le combinazioni non nulle degli altri due non possono avere nulle la prima e la seconda componente.

▷ **Esempio** Nello spazio \mathbb{P} dei polinomi, i vettori p^1, p^2, p^3 , con

$$p^1(t) = 1 - t, \quad p^2(t) = t(1 - t), \quad p^3(t) = 1 - t^2$$

sono l.d., poichè $p^1 + p^2 - p^3 = 0$.

▷ **Osservazione** In generale, se uno dei vettori v^1, v^2, \dots, v^k è il vettore nullo, i vettori sono l.d., in quanto uno di essi (quello nullo) si può scrivere come c.l. (a coefficienti nulli) dei rimanenti. Alcuni vettori possono però essere l.d. anche se nessuno di essi è il vettore nullo (come visto nell'esempio).

▷ **Esercizio** Se $T = \{v^1, v^2, \dots, v^k\}$ e i vettori di T sono indipendenti, allora ogni sottoinsieme di T è costituito da vettori indipendenti.

Invece, se $T = \{v^1, v^2, \dots, v^k\}$ e i vettori di T sono dipendenti, allora ogni insieme di vettori che contenga T è costituito da vettori dipendenti.

Vale il seguente risultato

Teorema 1

- i) Nello spazio vettoriale \mathbb{V} , i vettori v^1, v^2, \dots, v^k sono l.d. se e solo se esiste una loro c.l. che è uguale al vettore nullo e che non ha tutti i coefficienti nulli (cioè ha almeno un coefficiente non nullo).²⁷
- ii) I vettori v^1, v^2, \dots, v^k sono l.i. se e solo se l'unica loro c.l. uguale al vettore nullo è quella in cui tutti i coefficienti sono nulli.²⁸

Dimostrazione Per quanto riguarda i):

se i vettori v^1, v^2, \dots, v^k sono l.d. allora per definizione ce n'è uno che si può scrivere come c.l. degli altri. Supponiamo (non è restrittivo) che sia

$$v^1 = \alpha_2 v^2 + \alpha_3 v^3 + \dots + \alpha_k v^k .$$

Si può scrivere allora

$$v^1 - \alpha_2 v^2 - \alpha_3 v^3 - \dots - \alpha_k v^k = 0 ,$$

e abbiamo dimostrato la prima implicazione (il primo coefficiente della c.l. è non nullo).

Viceversa, se esiste una c.l. di v^1, v^2, \dots, v^k , uguale al vettore nullo e con coefficienti non tutti nulli, allora possiamo scrivere

$$\alpha_1 v^1 + \alpha_2 v^2 + \alpha_3 v^3 + \dots + \alpha_k v^k = 0 ,$$

e supponiamo che sia $\alpha_1 \neq 0$ (non è restrittivo). Si ottiene allora

$$v^1 = -\frac{\alpha_2}{\alpha_1} v^2 - \frac{\alpha_3}{\alpha_1} v^3 - \dots - \frac{\alpha_k}{\alpha_1} v^k .$$

La proposizione i) è quindi dimostrata.

Per quanto riguarda ii):

la dimostrazione si ottiene immediatamente dal punto i). Infatti, se la condizione espressa nell'enunciato non fosse necessaria, avremmo che esisterebbero c.l. nulle con coefficienti non tutti nulli, e quindi, per la seconda implicazione del punto i), i vettori sarebbero l.d.

Viceversa, la condizione è sufficiente perché, se così non fosse, sarebbe negata questa volta la validità prima implicazione della proposizione i). \square

All'ultimo teorema si può dare una versione leggermente più forte:

Teorema 2 Nello spazio vettoriale \mathbb{V} , i vettori non nulli v^1, v^2, \dots, v^k sono l.d. se e solo se qualche v^i , con $2 \leq i \leq k$, è c.l. dei precedenti.

Dimostrazione Anzitutto è chiaro che, se vale la condizione del teorema, i vettori v^1, v^2, \dots, v^k sono l.d. per definizione. Viceversa, supponiamo che i vettori siano l.d. Sia i il primo intero tra 2 e k per cui v^1, v^2, \dots, v^i sono l.d. Si osservi che $i \geq 2$, dato che i vettori non sono nulli e che, nel caso "peggiore", $i = k$. Allora (teorema 1) $\alpha_1 v^1 + \alpha_2 v^2 + \dots + \alpha_i v^i = 0$ con coefficienti non tutti nulli. Ma possiamo dire che $\alpha_i \neq 0$, per come è stato definito i , e quindi

$$v^i = -\frac{\alpha_1}{\alpha_i} v^1 - \dots - \frac{\alpha_{i-1}}{\alpha_i} v^{i-1}, \quad \text{come volevasi dimostrare.}$$

\square

▷ **Esempio** Lo studente verifichi il teorema precedente in \mathbb{R}^3 , con

$$v^1 = (0, 1, 0), \quad v^2 = (0, 1, 1), \quad v^3 = (1, 0, 0), \quad v^4 = (1, 1, 0), \quad v^5 = (1, 1, 1).$$

Esaminiamo ora un particolare ed importantissimo insieme di vettori di \mathbb{R}^n . Consideriamo in \mathbb{R}^n i vettori

$$u^1, u^2, \dots, u^n ,$$

²⁷ Scritto formalmente, questo risultato dice che i vettori v^1, v^2, \dots, v^k sono l.d. se e solo se

$$\exists \alpha_1, \alpha_2, \dots, \alpha_k : \alpha_1 v^1 + \alpha_2 v^2 + \dots + \alpha_k v^k = 0 \text{ e } \alpha_i \neq 0 \text{ per qualche } i, 1 \leq i \leq k.$$

²⁸ Scritto formalmente, questo risultato invece dice che i vettori v^1, v^2, \dots, v^k sono l.i. se e solo se

$$\alpha_1 v^1 + \alpha_2 v^2 + \dots + \alpha_k v^k = 0 \Rightarrow \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$$

dove il generico vettore u^j ha nulle tutte le sue componenti tranne la j -esima, che è uguale ad 1. Possiamo formalizzare la definizione dicendo che

$$\forall j, \text{ con } 1 \leq j \leq n, u^j \in \mathbb{R}^n \\ \text{e } u_i^j = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j. \end{cases} \quad ^{29}$$

Faccio notare esplicitamente che in \mathbb{R}^n l'insieme in questione è costituito da n elementi.

Ad esempio, in \mathbb{R}^3 , questo insieme è costituito dai vettori

$$u^1 = (1, 0, 0) \quad , \quad u^2 = (0, 1, 0) \quad , \quad u^3 = (0, 0, 1) \quad .$$

I vettori u^1, u^2, \dots, u^n si dicono i *vettori fondamentali* di \mathbb{R}^n . Vediamo perchè questo insieme di vettori è importante.

▷ **Osservazione** Ogni vettore di \mathbb{R}^n si può scrivere (in modo unico) come combinazione lineare dei vettori fondamentali. Basta osservare infatti che, considerato un qualunque vettore $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, si può scrivere

$$x = x_1 u^1 + x_2 u^2 + \dots + x_n u^n \quad ,$$

che è una combinazione lineare dei vettori fondamentali. Inoltre si osserva che i coefficienti di tale combinazione lineare sono le componenti del vettore x . Quando scriviamo un vettore come n -upla di componenti, queste componenti sono quindi i coefficienti della combinazione lineare di vettori fondamentali attraverso la quale si può esprimere il vettore stesso.

▷ **Osservazione** I vettori fondamentali quindi, attraverso loro opportune combinazioni lineari, mi permettono di ottenere tutti i vettori di \mathbb{R}^n . Per questo si dice che i vettori fondamentali sono *generatori* (o un *insieme di generatori*) dello spazio \mathbb{R}^n .

Viene data infatti in generale la seguente

Definizione I vettori v^1, v^2, \dots, v^k si dicono *generatori* dello spazio vettoriale \mathbb{V} se ogni vettore $x \in \mathbb{V}$ si può scrivere come c.l. di v^1, v^2, \dots, v^k , se cioè esistono dei coefficienti $\alpha_1, \alpha_2, \dots, \alpha_k$ tali che

$$x = \alpha_1 v^1 + \alpha_2 v^2 + \dots + \alpha_k v^k, \forall x \in \mathbb{V}.$$

2.3 Base e dimensione di uno spazio vettoriale

Segue una definizione fondamentale.

Definizione Una **base** di uno spazio vettoriale \mathbb{V} è un insieme di vettori di \mathbb{V} che siano l.i. e generatori di \mathbb{V} .³⁰

▷ **Osservazione** I vettori fondamentali di \mathbb{R}^n sono una base di \mathbb{R}^n . Infatti, oltre ad essere, come già visto, generatori di \mathbb{R}^n , è facile dimostrare che sono anche l.i. Lo studente è invitato a dimostrarlo.

▷ **Esercizio** Si verifichi che in \mathbb{R}^n non vi è un'unica base: si costruisca, ad esempio in \mathbb{R}^2 , una base che non sia quella dei vettori fondamentali.

Definizione In \mathbb{R}^n , la base formata dai vettori fondamentali si chiama la *base fondamentale* (o *base canonica*).

Veniamo ora ai risultati fondamentali.

Teorema 3 Se \mathbb{V} è uno spazio vettoriale e se $\{v^1, v^2, \dots, v^k\}$ è una base di \mathbb{V} , allora ogni $x \in \mathbb{V}$ si può scrivere *in modo unico* come c.l. dei vettori della base.

Dimostrazione Che ogni $x \in \mathbb{V}$ si possa scrivere come c.l. dei vettori della base segue dalla definizione di base. Dimostriamo che ciò avviene in modo unico. Abbiamo

$$x = \alpha_1 v^1 + \alpha_2 v^2 + \dots + \alpha_k v^k.$$

Se supponiamo che sia anche

$$x = \beta_1 v^1 + \beta_2 v^2 + \dots + \beta_k v^k,$$

²⁹ La definizione può essere data utilizzando la funzione detta δ di Kronecker: tale funzione viene indicata con δ_{ij} e vale 1 se $i = j$ e vale 0 se $i \neq j$. Pertanto possiamo scrivere $u_i^j = \delta_{ij}$.

³⁰ La definizione data di c.l. prevede che i vettori di cui si fa una combinazione siano in numero finito. Analogamente, vettori indipendenti (o dipendenti) sono sempre in numero finito. Anche il concetto qui fornito di base, quindi, presuppone un numero *finito* di vettori.

Esistono spazi vettoriali che hanno una base formata da infiniti vettori, ma noi non considereremo tale possibilità: pertanto noi parleremo di basi di spazi vettoriali intendendo che tali basi sono finite, cioè fatte da un numero finito di vettori. Tali spazi si dicono *a dimensione finita* e in tutto quello che segue sottintenderemo che gli spazi sono a dimensione finita.

allora

$$(\alpha_1 - \beta_1)v^1 + \dots + (\alpha_k - \beta_k)v^k = 0.$$

Ma quindi, per la lineare indipendenza e per il teorema 1, si ha che $\alpha_i - \beta_i = 0$ per ogni i , $1 \leq i \leq k$, da cui la tesi. \square

Teorema 4 Se \mathbb{V} è uno spazio vettoriale e se $\{y^1, y^2, \dots, y^m\}$ è un insieme di vettori l.i. di \mathbb{V} che non è una base di \mathbb{V} , allora possiamo trovare dei vettori $\{y^{m+1}, \dots, y^{m+p}\}$ tali che $\{y^1, y^2, \dots, y^m, y^{m+1}, \dots, y^{m+p}\}$ sono una base di \mathbb{V} . In altre parole, ogni insieme di vettori l.i. può essere esteso fino ad ottenere una base.

Dimostrazione Supponiamo che $\{x^1, x^2, \dots, x^k\}$ sia una base di \mathbb{V} .³¹ Consideriamo i vettori

$$y^1, y^2, \dots, y^m, x^1, x^2, \dots, x^k.$$

Essi sono l.d. poiché ogni y è c.l. degli x . Allora (teorema 2) ce n'è uno che è c.l. dei precedenti e sia z il primo di questi. Certamente z non è uno degli y (essi sono l.i.), quindi z è uno degli x , diciamo $z = x^i$. Allora consideriamo

$$y^1, y^2, \dots, y^m, x^1, x^2, \dots, x^{i-1}, x^{i+1}, \dots, x^k.$$

Essi sono ancora generatori di \mathbb{V} , dato lo sono x^1, x^2, \dots, x^k e dato che x^i è c.l. di $y^1, y^2, \dots, y^m, x^1, x^2, \dots, x^{i-1}$. Ma ora possiamo riapplicare il teorema 2 a questi vettori, fino a quando otteniamo dei vettori generatori di \mathbb{V} e l.i. (quindi una base), tra i quali ci sono y^1, y^2, \dots, y^m . \square

▷ *Esempio* Come esempio di completamento della base, prendiamo in \mathbb{R}^3 i vettori

$$y^1 = (1, 1, 1), \quad y^2 = (1, 1, 0),$$

che sono l.i. Consideriamo allora i 5 vettori

$$y^1, y^2, u^1 = (1, 0, 0), \quad u^2 = (0, 1, 0), \quad u^3 = (0, 0, 1).$$

Essi sono generatori di \mathbb{R}^3 . Sono l.d. e, applicando una prima volta il teorema 2, possiamo osservare che u^1 non è c.l. dei primi due (lo si verifichi), ma u^2 è c.l. dei primi tre, dato che

$$u^2 = (0, 1, 0) = y^2 - u^1.$$

Quindi “buttiamo via” u^2 e consideriamo

$$y^1, y^2, u^1, u^3.$$

Essi sono ancora generatori di \mathbb{R}^3 . Sono l.d. e applicando una seconda volta il teorema 2 possiamo osservare che u^3 è c.l. dei primi tre, dato che $u^3 = y^1 - y^2$. Quindi buttiamo via u^3 . Rimangono

$$y^1, y^2, u^1.$$

Essi sono generatori di \mathbb{R}^3 e sono l.i., quindi sono una base di \mathbb{R}^3 .

Ora un teorema fondamentale, che giustifica una definizione altrettanto fondamentale.

Teorema 5 Tutte le basi di uno spazio vettoriale \mathbb{V} hanno lo stesso numero di vettori.

Dimostrazione Siano $\mathcal{X} = \{x^1, x^2, \dots, x^n\}$ e $\mathcal{Y} = \{y^1, y^2, \dots, y^m\}$ due insiemi di vettori in \mathbb{V} . Supponiamo che i due insiemi abbiano ciascuno una sola delle due proprietà che caratterizzano una base, e in particolare supponiamo che i vettori di \mathcal{X} siano generatori di \mathbb{V} e che i vettori di \mathcal{Y} siano l.i.

Consideriamo i vettori

$$y^m, x^1, x^2, \dots, x^n.$$

Essi sono generatori di \mathbb{V} , dato che lo sono gli x , e sono l.d., dato che y^m è c.l. degli x . Ragionando come prima, sia x^i il primo che è c.l. dei precedenti. Lo togliamo e quindi aggiungiamo al primo posto y^{m-1} , ottenendo

$$y^{m-1}, y^m, x^1, x^2, \dots, x^{i-1}, x^{i+1}, \dots, x^n.$$

Essi sono ancora generatori di \mathbb{V} . Continuando così e ottenendo ad ogni passo dei generatori di \mathbb{V} , possiamo dire che gli x non possono finire prima degli y (intendo dire che non ci si può trovare ad aver tolto tutti gli x ed avere ancora qualche y da aggiungere) in quanto vorrebbe dire che qualche y è c.l. degli altri y , contro l'ipotesi che gli y sono l.i. Questo significa che $n \geq m$.

³¹ In realtà questo è un punto delicato, dato che, in un generico spazio vettoriale, occorrerebbe prima dimostrare che una base esiste. Questo infatti si può fare, anche se qui non approfondiamo questo punto.

Ma allora, se supponiamo che \mathcal{X} e \mathcal{Y} siano entrambi basi di \mathbb{V} , cioè che abbiano entrambi le due proprietà di una base, segue che $n \geq m$ e nello stesso tempo $m \geq n$, cioè che $n = m$. \square

È giustificata ora la seguente definizione

Definizione La **dimensione** di uno spazio vettoriale \mathbb{V} è il numero di elementi delle basi dello spazio. Essa si indica con $\dim \mathbb{V}$.

Come conseguenza dell'ultimo teorema, si ha che

Teorema 6

- i) Ogni insieme di $n + 1$ vettori in uno spazio vettoriale \mathbb{V} di dimensione n è formato da vettori l.d.
- ii) Un insieme di n vettori di \mathbb{V} è una base di \mathbb{V} se e solo se i vettori sono l.i.
- iii) Un insieme di n vettori di \mathbb{V} è una base di \mathbb{V} se e solo se i vettori sono generatori di \mathbb{V} .

2.4 Sottospazi

Definizione Un sottoinsieme non vuoto \mathbb{S} di uno spazio vettoriale \mathbb{V} si dice *sottospazio di \mathbb{V}* se per ogni $x, y \in \mathbb{S}$ ogni loro c.l. $\alpha x + \beta y$ appartiene a \mathbb{S} .

▷ *Osservazioni*

- Se \mathbb{S} è sottospazio di \mathbb{V} , allora \mathbb{S} contiene l'origine di \mathbb{V} . Infatti, se $x \in \mathbb{S}$, allora anche $x - x = 0 \in \mathbb{S}$.
- Se \mathbb{S} è sottospazio di \mathbb{V} , allora \mathbb{S} è a sua volta spazio vettoriale, con le stesse operazioni definite in \mathbb{V} .
- Sottospazi molto particolari di uno spazio \mathbb{V} sono $\{0\}$ e tutto \mathbb{V} , e sono detti sottospazi banali.

▷ *Esempi* Lo studente verifichi che

- In \mathbb{R}^2 , l'insieme $\mathbb{S} = \{(s, 0) : s \in \mathbb{R}\}$ è sottospazio di \mathbb{R}^2 ;
- In \mathbb{R}^2 , l'insieme $\mathbb{S} = \{(s_1, s_2) : s_1^2 + s_2^2 \leq 1\}$ *non* è sottospazio di \mathbb{R}^2 ;
- In \mathbb{R}^3 , l'insieme $\mathbb{S} = \{(s_1, s_2, s_3) : s_1 + s_2 + s_3 = 0\}$ è sottospazio di \mathbb{R}^3 ;
- In \mathbb{P} , spazio vettoriale dei polinomi, l'insieme $\mathbb{S} = \{p \in \mathbb{P} : p(0) = 0\}$ è sottospazio di \mathbb{P} ;
- In \mathbb{P} l'insieme $\mathbb{S} = \{p \in \mathbb{P} : p(0) = 1\}$ *non* è sottospazio di \mathbb{P} .

Si può dimostrare facilmente il seguente risultato:

Teorema 7 L'intersezione di una qualunque famiglia di sottospazi di uno spazio \mathbb{V} è un sottospazio di \mathbb{V} .

Come applicazione di questo risultato, se T è un qualunque insieme di vettori di \mathbb{V} , possiamo considerare la famiglia di tutti i sottospazi di \mathbb{V} che contengono T e farne l'intersezione. Otteniamo, per il teorema 7, un sottospazio di \mathbb{V} . Questo è il più piccolo sottospazio di \mathbb{V} che contiene T , e si chiama il *sottospazio generato da T* .

▷ *Esempio* Consideriamo in \mathbb{R}^3 due vettori x^1 e x^2 . Vogliamo dimostrare che il sottospazio di \mathbb{R}^3 generato da x^1, x^2 coincide con l'insieme di tutte le c.l. di x^1, x^2 .

Sia \mathbb{S} il sottospazio generato da x^1, x^2 , cioè l'intersezione di tutti i sottospazi che contengono x^1, x^2 .

Sia poi $\mathcal{C} = \{a_1 x^1 + a_2 x^2 : a_1, a_2 \in \mathbb{R}\}$, cioè l'insieme di tutte le c.l. di x^1, x^2 . Intanto possiamo dire che \mathcal{C} è sottospazio di \mathbb{R}^3 , dato che c.l. di c.l. di x^1, x^2 sono certamente c.l. di x^1, x^2 . Inoltre \mathcal{C} contiene x^1, x^2 , dato che $x^1 = 1x^1 + 0x^2$ e $x^2 = 0x^1 + 1x^2$.

Quindi $\mathbb{S} \subset \mathcal{C}$, dato che \mathbb{S} è contenuto in tutti i sottospazi che contengono x^1, x^2 .

Ma si ha anche $\mathcal{C} \subset \mathbb{S}$, dato che \mathbb{S} , essendo sottospazio che contiene x^1, x^2 , contiene anche le c.l. di x^1, x^2 . Dunque $\mathbb{S} = \mathcal{C}$, come si voleva dimostrare. \square

Il risultato si generalizza ad un qualunque insieme finito di vettori di uno spazio vettoriale. Si ha cioè che, se x^1, x^2, \dots, x^k sono vettori di uno spazio \mathbb{V} , il sottospazio generato da x^1, x^2, \dots, x^k è formato da tutte e sole le c.l. di x^1, x^2, \dots, x^k .

Possiamo ripetere per un sottospazio la definizione di vettori generatori e di base.

Definizione Se \mathbb{S} è sottospazio di uno spazio vettoriale \mathbb{V} , si dice che i vettori v^1, v^2, \dots, v^k *generano* (o sono *generatori* di \mathbb{S}) se ogni vettore di \mathbb{S} si può scrivere come c.l. dei vettori v^1, v^2, \dots, v^k .

Definizione I vettori v^1, v^2, \dots, v^k sono una **base** di \mathbb{S} se sono l.i. e nello stesso tempo sono generatori di \mathbb{S} .

Si può dimostrare, come si intuisce facilmente, che se \mathbb{S} è sottospazio di uno spazio vettoriale n -dimensionale \mathbb{V} , allora $\dim \mathbb{S} \leq n$.

Ci si rende conto facilmente che in \mathbb{R}^2 i possibili sottospazi non banali sono tutti e soli gli insiemi la cui rappresentazione nel piano cartesiano è una retta passante per l'origine. Tali sottospazi hanno dimensione 1.

In \mathbb{R}^3 invece i sottospazi non banali sono le rette per l'origine (che hanno dimensione 1) oppure i piani per l'origine (che hanno dimensione 2).

Più in generale è naturale chiedersi quale sia la dimensione del sottospazio generato da alcuni vettori v^1, v^2, \dots, v^k , e come si possa ottenere una base di tale sottospazio. Si può dimostrare che un metodo è quello di eliminare via via i vettori dipendenti. Mi spiego meglio. Se i vettori v^1, v^2, \dots, v^k sono indipendenti, essi sono una base del sottospazio generato e k è la dimensione. Se invece essi sono dipendenti, significa che almeno uno di essi è combinazione lineare degli altri. Se ne sceglie uno di questi e lo si elimina dall'insieme e si ripete il procedimento, fino a quando i vettori rimasti risultano indipendenti.

A titolo di esempio, consideriamo il sottospazio di \mathbb{R}^2 generato dai vettori

$$v^1 = (1, 1) \quad , \quad v^2 = (1, -1) \quad , \quad v^3 = (0, 2) \quad , \quad v^4 = (2, 0) \quad .$$

Possiamo osservare che i vettori sono dipendenti e che, ad esempio, $v^4 = v^1 + v^2$. Eliminiamo v^4 (perchè il sottospazio generato da v^1, v^2, v^3 coincide con quello generato da v^1, v^2, v^3, v^4).

I vettori rimasti sono ancora dipendenti in quanto, ad esempio, $v^3 = v^1 - v^2$. Eliminiamo anche v^3 , per lo stesso motivo di prima. Restano v^1 e v^2 , che sono indipendenti. Essi sono una base per il sottospazio in questione (che è tra l'altro tutto \mathbb{R}^2).

Si intuisce che, in generale, la base trovata con questo procedimento può non essere unica.

2.5 Prodotto interno

Definisco ora una nuova operazione tra vettori di \mathbb{R}^n .

Definizione Siano $x, y \in \mathbb{R}^n$, con $x = (x_1, x_2, \dots, x_n)$ e $y = (y_1, y_2, \dots, y_n)$. Si definisce *prodotto interno* (o *prodotto scalare*) di x e y il numero reale

$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \dots + x_n y_n = \sum_{i=1}^n x_i y_i \quad .$$

▷ **Osservazione** Il nome prodotto scalare fa riferimento al fatto che il risultato di questa operazione tra vettori è uno scalare, cioè un numero reale.

Ad esempio, il prodotto interno di $x = (1, 2, 3)$ e $y = (-3, 2, 1)$ è $\langle x, y \rangle = 4$, il prodotto interno di $x = (0, -1, 1)$ e $y = (1, 1, 0)$ è $\langle x, y \rangle = -1$, il prodotto interno di $x = (0, -1, 1)$ e $y = (1, 1, 1)$ è $\langle x, y \rangle = 0$.

▷ **Osservazione** Possiamo dare al prodotto interno di due vettori in \mathbb{R}^2 (la cosa vale anche in \mathbb{R}^3) un'interpretazione geometrica. Siano $x = (x_1, x_2)$ e $y = (y_1, y_2)$ due vettori di lunghezza unitaria, cioè tali che $x_1^2 + x_2^2 = y_1^2 + y_2^2 = 1$. Allora si può scrivere (vedi figura)

$$x = (\cos \alpha, \sin \alpha) \quad \text{e} \quad y = (\cos \beta, \sin \beta) \quad .$$

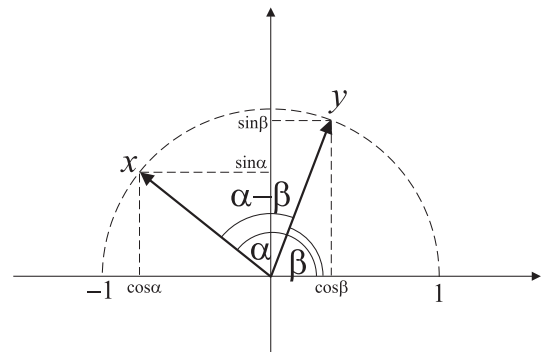
Quindi il prodotto interno di x e y è

$$\cos \alpha \cos \beta + \sin \alpha \sin \beta = \cos(\alpha - \beta) \quad .$$

Pertanto $\langle x, y \rangle$ è il coseno dell'angolo che i due vettori formano tra loro.

▷ **Osservazione** Si verifica facilmente che il prodotto interno tra vettori di \mathbb{R}^n ha queste proprietà:

1. $\langle x, x \rangle \geq 0$, $\forall x \in \mathbb{R}^n$, e $\langle x, x \rangle = 0$ se e solo se $x = 0$;
2. $\langle x, y \rangle = \langle y, x \rangle$, $\forall x, y \in \mathbb{R}^n$;
3. $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$, $\forall \alpha, \beta \in \mathbb{R}$, $\forall x, y, z \in \mathbb{R}^n$.



Definizione In un generico spazio vettoriale \mathbb{V} su \mathbb{R} , si chiama *prodotto interno* una qualunque funzione che associ alla coppia di vettori x, y un numero reale, che indichiamo con $\langle x, y \rangle$, con le proprietà

1. $\langle x, x \rangle \geq 0$, $\forall x \in \mathbb{V}$, e $\langle x, x \rangle = 0$ se e solo se $x = 0$;
2. $\langle x, y \rangle = \langle y, x \rangle$, $\forall x, y \in \mathbb{V}$;
3. $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$, $\forall \alpha, \beta \in \mathbb{R}$, $\forall x, y, z \in \mathbb{V}$.

Se in \mathbb{V} è definito un prodotto interno, \mathbb{V} si dirà *spazio vettoriale con prodotto interno*.

▷ **Esempio** Lo spazio vettoriale delle funzioni continue in un intervallo $[a, b]$ diventa spazio vettoriale con prodotto interno se definiamo prodotto interno di due funzioni f e g , continue in $[a, b]$, il numero reale $\langle f, g \rangle = \int_a^b fg$.

▷ **Esercizio** Lo studente verifichi, in \mathbb{R}^n e nello spazio delle funzioni continue in $[a, b]$, le tre proprietà dei prodotti interni definiti.

▷ **Esercizio** Col prodotto interno definito in \mathbb{R}^n si ha $\langle 0, x \rangle = 0$ per ogni $x \in \mathbb{R}^n$. Si dimostri che tale proprietà vale in un qualunque spazio vettoriale con prodotto interno.

Definizione Si definisce *norma euclidea* del vettore $x \in \mathbb{R}^n$ (e si scrive $\|x\|$) il numero reale

$$\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{\sum_{i=1}^n x_i^2}.$$

▷ **Osservazione** Il concetto di norma è un concetto molto generale. Una *norma* in uno spazio vettoriale \mathbb{V} su \mathbb{R} è in generale una funzione definita in \mathbb{V} a valori in \mathbb{R} che abbia le seguenti proprietà:

1. $\|x\| \geq 0$ $\forall x \in \mathbb{V}$ e $\|x\| = 0$ se e solo se $x = 0$;
2. $\|\alpha x\| = |\alpha| \|x\|$, $\forall x \in \mathbb{V}$, $\forall \alpha \in \mathbb{R}$;
3. $\|x + y\| \leq \|x\| + \|y\|$, $\forall x, y \in \mathbb{V}$.

▷ **Esercizio** Lo studente verifichi tali proprietà sulla norma euclidea in \mathbb{R}^n .

La norma euclidea non è l'unica norma possibile in \mathbb{R}^n .

Altri esempi di norme (non euclidee) in \mathbb{R}^n sono le seguenti, dette rispettivamente *norma uno*, *norma p* e *norma infinito*:

$$\|x\|_1 = \sum_{i=1}^n |x_i|, \quad \|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p} \quad (p > 0), \quad \|x\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$

Si può definire una distanza tra gli elementi di \mathbb{R}^n con la seguente

Definizione Dati $x, y \in \mathbb{R}^n$, si chiama *distanza euclidea* tra x e y il numero reale

$$d(x, y) = \|x - y\|.$$

▷ **Osservazione** Anche il concetto di distanza, come quello di norma, è più generale di quanto la definizione precedente possa far pensare. In generale, in un insieme X ³² si definisce *distanza* una qualunque funzione δ definita in $X \times X$ a valori reali che abbia le seguenti proprietà:

1. $\delta(x, y) \geq 0$ $\forall x, y \in X$ e $\delta(x, y) = 0$ se e solo se $x = y$;
2. $\delta(x, y) = \delta(y, x)$, $\forall x, y \in X$;
3. $\delta(x, y) \leq \delta(x, z) + \delta(y, z)$, $\forall x, y, z \in X$.

▷ **Esercizio** Si determini, in \mathbb{R}^n , la distanza euclidea, la distanza in norma uno, in norma p e in norma infinito dei vettori fondamentali.

▷ **Esercizio** Si verifichi che in \mathbb{R}^n (ma anche in un qualsiasi insieme X) la funzione

$$\delta(x, y) = \begin{cases} 0 & \text{se } x = y \\ 1 & \text{se } x \neq y \end{cases}$$

è una distanza.

▷ **Esercizio** Si dimostri che, se $x \mapsto \|x\|$ è una norma in uno spazio vettoriale, allora $\delta(x, y) = \|x - y\|$ è una distanza.

³² Si noti che viene richiesto solo che X sia un insieme, quindi non uno spazio vettoriale. Questo perché le proprietà della distanza non richiedono nessuna struttura particolare in X . Lo studente è invitato a riconsiderare le proprietà della norma e a constatare che queste ultime richiedono invece una struttura, quella di spazio vettoriale.

▷ **Osservazione** Non tutte le distanze si ricavano da una norma nel modo suggerito nell'esercizio precedente. Ad esempio la distanza $0 - 1$ definita sopra non si può ricavare da una norma. Lo studente cerchi di capire perché.

▷ **Osservazione** Data in uno spazio vettoriale una distanza δ , il definire $\|x\| = \delta(x, 0)$ non porta in generale ad una norma (anche se così accade in \mathbb{R}^n con la distanza euclidea). Lo studente verifichi che ad esempio con la distanza $0 - 1$ la funzione $x \mapsto \delta(x, 0)$ non è una norma.

In uno spazio \mathbb{V} con prodotto interno poniamo $\|x\| = \sqrt{\langle x, x \rangle}$. Si può dimostrare che vale la seguente *disuguaglianza di Cauchy-Schwarz*:

$$|\langle x, y \rangle| \leq \|x\| \|y\|, \quad \forall x, y \in \mathbb{V}.$$

Dimostrazione Consideriamo i vettori $x + ty$ con $t \in \mathbb{R}$: si ha

$$\begin{aligned} 0 &\leq \langle x + ty, x + ty \rangle \\ &= \langle x, x \rangle + \langle x, ty \rangle + \langle ty, x \rangle + \langle ty, ty \rangle \\ &= \langle x, x \rangle + 2t\langle x, y \rangle + t^2\langle y, y \rangle \\ &= \|x\|^2 + 2t\langle x, y \rangle + t^2\|y\|^2. \end{aligned}$$

Quest'ultimo è un polinomio di secondo grado in t , che risulta non negativo per ogni valore di t . Il discriminante dell'equazione deve allora essere ≤ 0 , e quindi

$$\langle x, y \rangle^2 - \|x\|^2 \|y\|^2 \leq 0,$$

da cui

$$\langle x, y \rangle^2 \leq \|x\|^2 \|y\|^2, \quad \text{e quindi} \quad |\langle x, y \rangle| \leq \|x\| \|y\|.$$

□

▷ **Esercizio** La disuguaglianza di Cauchy-Schwarz permette di dimostrare che la funzione $\|x\| = \sqrt{\langle x, x \rangle}$ definisce, in uno spazio con prodotto interno, una norma.

Veniamo ora all'importante concetto di ortogonalità tra vettori in uno spazio vettoriale con prodotto interno.

Definizione Due vettori x e y di uno spazio vettoriale \mathbb{V} con prodotto interno si dicono *ortogonali* se il loro prodotto interno è nullo, cioè se $\langle x, y \rangle = 0$.

▷ **Esempi**

- In \mathbb{R}^3 i vettori $x = (2, -1, 1)$ e $y = (1, 1, -1)$ sono ortogonali.
- Nello spazio vettoriale delle funzioni continue nell'intervallo $[-1, 1]$, con il prodotto interno dato dall'integrale del prodotto³³, le due funzioni $\max(x, 0)$ e $\max(-x, 0)$ sono ortogonali.

▷ **Esercizio** Si dimostri che, se \mathbb{V} è spazio vettoriale e $v \in \mathbb{V}$, allora l'insieme di tutti i vettori di \mathbb{V} ortogonali a v è un sottospazio di \mathbb{V} . Si generalizzi lo stesso risultato ad un qualunque sottoinsieme di \mathbb{V} , cioè: se $T \subset \mathbb{V}$, allora l'insieme dei vettori di \mathbb{V} ortogonali a tutti gli elementi di T è un sottospazio di \mathbb{V} .

▷ **Osservazione** Il sottospazio dei vettori ortogonali ad un insieme T (vedi esercizio precedente) si indica solitamente con T^\perp . Tale insieme è quindi un sottospazio a prescindere dal fatto che T lo sia. Esso però acquista particolare importanza se T è sottospazio di \mathbb{V} . In questo caso T^\perp si chiama il *complemento ortogonale* di T .

Definizione I vettori di un insieme \mathcal{X} si dicono *ortonormali* se, presi x e y in \mathcal{X} , si ha che $\langle x, y \rangle = 0$ se $x \neq y$ e $\langle x, x \rangle = 1$ se $x = y$.

▷ **Osservazione** Se $\mathcal{X} = \{x^1, x^2, \dots, x^k\}$, possiamo dire che i vettori sono ortonormali se si ha

$$\langle x^i, x^j \rangle = \delta_{ij}^{34}, \quad \forall i, j.$$

Sui vettori ortonormali sussistono importanti risultati.

Teorema 8 In uno spazio vettoriale \mathbb{V} con prodotto interno, se i vettori x^1, x^2, \dots, x^k sono ortonormali, allora essi sono l.i.

³³ Vedi l'esempio che segue la definizione di prodotto interno.

³⁴ Il simbolo è il δ di Kronecker.

Dimostrazione Siano x^1, x^2, \dots, x^k vettori ortonormali. Dimostriamo che sono l.i. facendo vedere che l'unica loro c.l. uguale al vettore nullo è quella che ha i coefficienti tutti nulli (vedi teorema 1).

Presa una generica c.l. $\alpha_1 x^1 + \alpha_2 x^2 + \dots + \alpha_k x^k = 0$, avremo, per ogni i , con $1 \leq i \leq k$,

$$0 = \langle \alpha_1 x^1 + \alpha_2 x^2 + \dots + \alpha_k x^k, x^i \rangle = \alpha_1 \langle x^1, x^i \rangle + \dots + \alpha_k \langle x^k, x^i \rangle = \alpha_i \langle x^i, x^i \rangle = \alpha_i.$$

Si ha quindi $\alpha_i = 0$ per ogni i , con $1 \leq i \leq k$. Pertanto tutti i coefficienti devono essere nulli. \square

Teorema 9 Sia \mathbb{V} spazio vettoriale con prodotto interno. Supponiamo che i vettori x^1, x^2, \dots, x^k siano ortonormali e sia $v \in \mathbb{V}$. Poniamo per comodità $\langle v, x^i \rangle = a_i$, con $1 \leq i \leq k$. Allora

1. $\sum_{i=1}^k a_i^2 \leq \|v\|^2$ (Disuguaglianza di Bessel).
2. Il vettore $z = v - \sum_{i=1}^k a_i x^i$ è ortogonale a tutti i vettori x^1, x^2, \dots, x^k .

Dimostrazione Per quanto riguarda il primo punto,

$$\begin{aligned} 0 &\leq \langle z, z \rangle = \langle v - \sum_{i=1}^k a_i x^i, v - \sum_{i=1}^k a_i x^i \rangle \\ &= \langle v, v \rangle - \langle v, \sum_{i=1}^k a_i x^i \rangle - \langle \sum_{i=1}^k a_i x^i, v \rangle + \langle \sum_{i=1}^k a_i x^i, \sum_{i=1}^k a_i x^i \rangle \\ &= \langle v, v \rangle - 2 \sum_{i=1}^k a_i \langle v, x^i \rangle + \sum_{i=1}^k \sum_{j=1}^k a_i a_j \langle x_i, x_j \rangle \\ &= \langle v, v \rangle - 2 \sum_{i=1}^k a_i^2 + \sum_{i=1}^k a_i^2 \\ &= \langle v, v \rangle - \sum_{i=1}^k a_i^2. \end{aligned}$$

Pertanto $\sum_{i=1}^k a_i^2 \leq \|v\|^2$.

Per quanto riguarda il secondo punto, per ogni $1 \leq j \leq k$ si ha

$$\langle z, x^j \rangle = \langle v - \sum_{i=1}^k a_i x^i, x^j \rangle = \langle v, x^j \rangle - \sum_{i=1}^k a_i \langle x^i, x^j \rangle = a_j - a_j = 0.$$

\square

▷ **Osservazione** La disuguaglianza di Bessel permette di ottenere un'altra dimostrazione della disuguaglianza di Cauchy-Schwarz.³⁵ Eccola: se $y = 0$, entrambi i membri della disuguaglianza sono nulli e quindi la disuguaglianza è verificata. Se invece $y \neq 0$, allora il vettore $y/\|y\|$ è ortonormale. Pertanto, per la disuguaglianza di Bessel, possiamo dire che

$$\langle x, y/\|y\| \rangle^2 \leq \|x\|^2,$$

da cui si ricava

$$\langle x, y \rangle^2 \leq \|x\|^2 \|y\|^2 \quad \text{e quindi} \quad |\langle x, y \rangle| \leq \|x\| \|y\|.$$

Particolarmente importanti sono gli insiemi di vettori che sono contemporaneamente ortonormali e basi dello spazio a cui appartengono.

Definizione I vettori x^1, x^2, \dots, x^k sono una *base ortonormale* di uno spazio vettoriale \mathbb{V} con prodotto interno se sono una base di \mathbb{V} e sono ortonormali.

▷ **Esempio** I vettori fondamentali di \mathbb{R}^n sono una base ortonormale di \mathbb{R}^n . Non si tratta dell'unica base ortonormale di \mathbb{R}^n , come vedremo più avanti.

³⁵ Ricordo che la disuguaglianza dice che, dati due vettori x e y , allora

$$|\langle x, y \rangle| \leq \|x\| \|y\|.$$

Sugli insiemi di vettori ortonormali e sulle basi ortonormali sussistono risultati molto generali. Qui ci limitiamo a questi che seguono.

Teorema 10 Se x^1, x^2, \dots, x^k sono una base ortonormale dello spazio vettoriale \mathbb{V} con prodotto interno e se $v \in \mathbb{V}$, allora

1. $v = \sum_{i=1}^k \langle v, x^i \rangle x^i$;
2. $\sum_{i=1}^k \langle v, x^i \rangle^2 = \|v\|^2$.

Dimostrazione Per quanto visto in precedenza, dal fatto che x^1, x^2, \dots, x^k sono una base segue che v si può scrivere in modo unico come c.l. dei vettori di base e sia $v = a_1 x^1 + a_2 x^2 + \dots + a_k x^k$. Ma allora

$$\langle v, x^i \rangle = \langle a_1 x^1 + a_2 x^2 + \dots + a_k x^k, x^i \rangle = a_i, \text{ per ogni } 1 \leq i \leq k.$$

Per il secondo punto abbiamo

$$\|v\|^2 = \langle v, v \rangle = \left\langle \sum_{i=1}^k a_i x^i, \sum_{i=1}^k a_i x^i \right\rangle = \sum_{i=1}^k \sum_{j=1}^k a_i a_j \langle x_i, x_j \rangle = \sum_{i=1}^k a_i^2.$$

□

2.6 Costruzione di una base ortonormale

Concludiamo con un interessante procedimento che permette, data una qualunque base di \mathbb{R}^n , di costruire una base ortonormale. Si tratta del *metodo di ortogonalizzazione di Gram-Schmidt*. Lo vediamo in generale e poi lo vediamo applicato ad un esempio.

Sia $\{v^1, v^2, \dots, v^k\}$ una base di uno spazio vettoriale \mathbb{V} con prodotto interno. Il metodo di Gram-Schmidt costruisce una base ortonormale $\{x^1, x^2, \dots, x^k\}$ con la proprietà che ogni x^j è c.l. di v^1, v^2, \dots, v^j .

PASSO 1. Poniamo $x^1 = v^1 / \|v^1\|$, e così x^1 è un vettore di norma unitaria.

PASSO 2. Sia $z = v^2 - \langle v^2, x^1 \rangle x^1$. Per il teorema 9 z è ortogonale a x^1 .

Inoltre z è c.l. di v^1, v^2 e non è il vettore nullo (se z fosse 0 v^1, v^2 sarebbero l.d.).

Poniamo allora $x^2 = z / \|z\|$.

Quindi a questo punto x^1, x^2 sono ortonormali e x^2 è c.l. di v^1, v^2 .

PASSO 3. Sia $z = v^3 - \langle v^3, x^1 \rangle x^1 - \langle v^3, x^2 \rangle x^2$. Per il teorema 9 z è ortogonale a x^1, x^2 .

Ora z è c.l. di v^1, v^2, v^3 e non è il vettore nullo (se z fosse 0 v^1, v^2, v^3 sarebbero l.d.).

Poniamo allora $x^3 = z / \|z\|$.

Quindi a questo punto x^1, x^2, x^3 sono ortonormali e x^3 è c.l. di v^1, v^2, v^3 .

⋮

PASSO j . Si pone $z = v^j - \langle v^j, x^1 \rangle x^1 - \dots - \langle v^j, x^{j-1} \rangle x^{j-1}$. Per il teorema 9 z è ortogonale a x^1, x^2, \dots, x^{j-1} .

Inoltre z è c.l. di v^1, v^2, \dots, v^j e non è il vettore nullo (se z fosse 0 v^1, v^2, \dots, v^j sarebbero l.d.).

Poniamo allora $x^j = z / \|z\|$.

A questo punto x^1, x^2, \dots, x^j sono ortonormali e x^j è c.l. di v^1, v^2, \dots, v^j .

Si continua così fino ad avere x^1, x^2, \dots, x^k ortonormali, dove ogni x^j è c.l. di v^1, v^2, \dots, v^j .

▷ **Esempio** Siano

$$v^1 = (1, 1, 1) \quad , \quad v^2 = (0, 1, 1) \quad , \quad v^3 = (0, 0, 1).$$

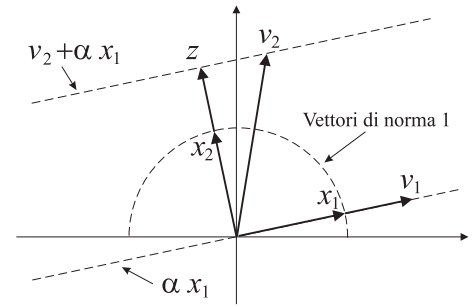
Si tratta di una base di \mathbb{R}^3 . Procedendo con il metodo di Gram-Schmidt, si ottiene

$$\begin{aligned} x^1 &= \frac{1}{\sqrt{3}}(1, 1, 1) \\ z &= v^2 - \langle v^2, x^1 \rangle x^1 = \frac{1}{3}(-2, 1, 1) \\ x^2 &= z / \|z\| = \frac{1}{\sqrt{6}}(-2, 1, 1) \\ z &= v^3 - \langle v^3, x^1 \rangle x^1 - \langle v^3, x^2 \rangle x^2 = \frac{1}{2}(0, -1, 1) \\ x^3 &= z / \|z\| = \frac{1}{\sqrt{2}}(0, -1, 1). \end{aligned}$$

Si verifica facilmente che i vettori $\frac{1}{\sqrt{3}}(1, 1, 1)$, $\frac{1}{\sqrt{6}}(-2, 1, 1)$, $\frac{1}{\sqrt{2}}(0, -1, 1)$ sono ortonormali.

Finiamo con un'interpretazione geometrica, in \mathbb{R}^2 , del metodo di Gram-Schmidt. Si segua il discorso nella figura a fianco.

Supponiamo che v^1, v^2 sia una base. Il primo passo del metodo non fa altro che normalizzare il primo vettore, quindi x^1 è un vettore di norma unitaria nel sottospazio generato da v^1 . Il secondo ed ultimo passo trova anzitutto, nel sottoinsieme dei vettori $v^2 + \alpha x^1$ il vettore $z = v^2 - \langle v^2, x^1 \rangle x^1$, che è ortogonale a x^1 . Si noti che l'insieme $v^2 + \alpha x^1$, al variare di α , è il sottospazio generato da x^1 traslato di v^2 . Il vettore z è c.l. di v^1, v^2 ed è ortogonale a x^1 ; si ottiene infine x^2 normalizzando z .



3 Trasformazioni lineari e matrici

In questa sezione i vettori verranno sempre pensati e rappresentati come *vettori colonna*.

3.1 Trasformazioni lineari

Iniziamo con il concetto di trasformazione lineare, per dare per così dire una motivazione all'altro concetto base di questa dispensa, il concetto di matrice. Consideriamo una funzione f , definita nello spazio vettoriale \mathbb{R}^n , a valori nello spazio vettoriale \mathbb{R}^m . Scriviamo naturalmente

$$f : \mathbb{R}^n \rightarrow \mathbb{R}^m.$$

Qualora se ne presenti l'opportunità, per distinguere una tale funzione dalle usuali funzioni reali di variabile reale, cioè dalle funzioni definite in \mathbb{R} a valori in \mathbb{R} che lo studente ha visto nel corso di Matematica I, si può dire che, se $m > 1$, la funzione è vettoriale (o a valori vettoriali) e, se $n > 1$, che è di variabile vettoriale. Nella prima parte del corso di Algebra lineare abbiamo considerato funzioni che rientrano in questa tipologia, con $m = 1$ (cioè funzioni $f : \mathbb{R}^n \rightarrow \mathbb{R}$, con $n > 1$, quindi funzioni reali di variabile vettoriale³⁶). Possiamo ovviamente pensare anche a funzioni vettoriali di variabile reale ($f : \mathbb{R} \rightarrow \mathbb{R}^m$, con $m > 1$).

Particolare importanza per i nostri scopi hanno le seguenti particolari funzioni:

Definizione Se $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, diciamo che f è una *trasformazione lineare di \mathbb{R}^n in \mathbb{R}^m* se f ha queste due proprietà:

1. $\forall x, y \in \mathbb{R}^n, f(x + y) = f(x) + f(y)$ (proprietà di additività);
2. $\forall x \in \mathbb{R}^n$ e $\forall \alpha \in \mathbb{R}, f(\alpha x) = \alpha f(x)$ (proprietà di omogeneità).

▷ **Osservazione** Si noti che nella definizione di trasformazione lineare si utilizza la struttura (le operazioni) di spazio vettoriale, sia nel dominio sia nel codominio di f .

▷ **Osservazione** Immediata conseguenza della definizione data è che, se f è una trasformazione lineare di \mathbb{R}^n in \mathbb{R}^m e se v^1, v^2, \dots, v^k sono vettori di \mathbb{R}^n , allora per ogni c.l. $\alpha_1 v^1 + \alpha_2 v^2 + \dots + \alpha_k v^k$ dei k vettori si ha

$$f(\alpha_1 v^1 + \alpha_2 v^2 + \dots + \alpha_k v^k) = \alpha_1 f(v^1) + \alpha_2 f(v^2) + \dots + \alpha_k f(v^k).$$

▷ **Esempi**

- Consideriamo la funzione $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ definita da

$$f \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2x_1 - x_2 \\ x_2 + x_3 \end{pmatrix}.$$

Verifichiamo che è una trasformazione lineare. Presi due generici elementi $x = (x_1, x_2, x_3)$ e $y = (y_1, y_2, y_3)$ in \mathbb{R}^3 , abbiamo

$$f(x+y) = f \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{pmatrix} = \begin{pmatrix} 2(x_1 + y_1) - (x_2 + y_2) \\ (x_2 + y_2) + (x_3 + y_3) \end{pmatrix} = \begin{pmatrix} 2x_1 - x_2 + 2y_1 - y_2 \\ x_2 + x_3 + y_2 + y_3 \end{pmatrix} = f \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + f \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = f(x) + f(y).$$

L'additività è dunque dimostrata. Vediamo l'omogeneità. Preso un $x \in \mathbb{R}^3$ e un numero reale α , abbiamo

$$f(\alpha x) = f \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \alpha x_3 \end{pmatrix} = \begin{pmatrix} 2\alpha x_1 - \alpha x_2 \\ \alpha x_2 + \alpha x_3 \end{pmatrix} = \alpha \begin{pmatrix} 2x_1 - x_2 \\ x_2 + x_3 \end{pmatrix} = \alpha f(x).$$

Quindi f è una trasformazione lineare, essendo additiva ed omogenea.

- Consideriamo ora la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ definita da

$$f \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2^2 \end{pmatrix}.$$

Questa non è una trasformazione lineare. Infatti non è additiva. Siano, ad esempio,

$$x = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{e} \quad y = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

³⁶ In quel contesto però non avevamo ancora parlato di \mathbb{R}^n con la struttura di spazio vettoriale.

Risulta allora

$$f(x+y) = f\begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 9 \end{pmatrix} \quad \text{mentre} \quad f(x) + f(y) = f\begin{pmatrix} 1 \\ 1 \end{pmatrix} + f\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \end{pmatrix}.$$

Si vede facilmente che f non è nemmeno omogenea. Si potrebbe dire che f è lineare nella sua prima componente, ma non nella seconda.

- La funzione

$$f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1 - x_2$$

è una trasformazione lineare di \mathbb{R}^2 in \mathbb{R} (lo si verifichi).

- La funzione

$$f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1 x_2$$

non è una trasformazione lineare di \mathbb{R}^2 in \mathbb{R} (lo si verifichi).

- La funzione

$$f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + 1 \\ x_2 \end{pmatrix}$$

non è una trasformazione lineare di \mathbb{R}^2 in \mathbb{R}^2 (lo si verifichi).

▷ **Esercizio** Si dimostri che, se f è lineare, allora $f(0) = 0$.

Consideriamo l'insieme $\mathcal{L}_{n,m}$ di tutte le trasformazioni lineari di \mathbb{R}^n in \mathbb{R}^m . Se definiamo la somma di due trasformazioni lineari e la moltiplicazione di una trasformazione lineare per uno scalare nel modo usuale, e cioè con

$$(f+g)(x) = f(x) + g(x), \quad \forall x \in \mathbb{R}^n$$

e con

$$(\alpha f)(x) = \alpha f(x), \quad \forall x \in \mathbb{R}^n, \forall \alpha \in \mathbb{R},$$

si può verificare facilmente (lo si faccia) che le due operazioni hanno tutte le proprietà che servono per rendere $\mathcal{L}_{n,m}$ uno *spazio vettoriale su* \mathbb{R} . D'ora in avanti, la scrittura $f \in \mathcal{L}_{n,m}$ vorrà dire che f è una trasformazione lineare di \mathbb{R}^n in \mathbb{R}^m .

Sia $f \in \mathcal{L}_{n,m}$. Si può pensare che sia

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix},$$

dove ciascuna f_j è una trasformazione lineare di \mathbb{R}^n in \mathbb{R} (cioè $f_j \in \mathcal{L}_{n,1}$, $\forall j$, $1 \leq j \leq m$).

Si può dimostrare facilmente che f è lineare se e solo se tutte le f_j sono lineari.

Quindi, ad esempio per dimostrare che

$$f\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2x_1 - x_2 \\ x_2 + x_3 \end{pmatrix}$$

è lineare si potrebbe dimostrare che le sue singole componenti lo sono, e cioè che sono lineari le funzioni (a valori reali)

$$f_1\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 2x_1 - x_2 \quad \text{e} \quad f_2\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_2 + x_3.$$

Parliamo ora della rappresentazione di una trasformazione lineare.

Cominciamo con un caso particolare e sia $f \in \mathcal{L}_{n,1}$. Se, dato $x \in \mathbb{R}^n$, scriviamo $x = x_1 u^1 + x_2 u^2 + \dots + x_n u^n$, dove u^1, u^2, \dots, u^n sono i vettori della base canonica di \mathbb{R}^n , la linearità di f permette di scrivere

$$f(x) = f(x_1 u^1 + x_2 u^2 + \dots + x_n u^n) = x_1 f(u^1) + x_2 f(u^2) + \dots + x_n f(u^n).$$

Se poniamo $a = (f(u^1), f(u^2), \dots, f(u^n))$, allora $f(x) = \langle a, x \rangle$.

Il vettore a è sufficiente per rappresentare la trasformazione f .³⁷

³⁷ È però doveroso precisare fin d'ora che la rappresentazione è data da a se esprimi x rispetto alla base canonica.

A questo punto è chiaro che è possibile rappresentare una qualunque trasformazione lineare f attraverso una sorta di “tabella”: basterà rappresentare ogni componente di f con un vettore (riga) come appena visto ed accostare tali vettori riga uno sull'altro.

Consideriamo ad esempio la trasformazione lineare $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$, già esaminata in precedenza, definita da

$$f(x) = f \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2x_1 - x_2 \\ x_2 + x_3 \end{pmatrix}.$$

Se rappresentiamo le due componenti del vettore $f(x)$ rispettivamente con i vettori riga

$$(2 \quad -1 \quad 0) \quad \text{e} \quad (0 \quad 1 \quad 1),$$

e accostiamo poi verticalmente questi due otteniamo

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

Questa si chiama *matrice*. Possiamo scrivere il vettore $f(x)$ come prodotto tra la matrice suddetta e il vettore x , cioè scrivere

$$f \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2x_1 - x_2 \\ x_2 + x_3 \end{pmatrix} = \begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},$$

se pensiamo che tale prodotto matrice \times vettore avvenga calcolando i prodotti interni delle *righe* della matrice per il vettore colonna x .

Saremmo pervenuti allo stesso risultato ragionando anche così: la linearità di f , come prima, porta a dire che

$$f(x) = f \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = f(x_1 u^1 + x_2 u^2 + x_3 u^3) = x_1 f(u^1) + x_2 f(u^2) + x_3 f(u^3) = x_1 \begin{pmatrix} 2 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Quello che abbiamo ottenuto è una c.l. dei vettori $f(u^1), f(u^2), f(u^3)$, che sono le immagini dei vettori fondamentali di \mathbb{R}^3 attraverso la trasformazione f .

Pertanto la f può essere rappresentata da questi soli vettori, questa volta disposti in *colonna*. Se li affianchiamo in una tabella, otteniamo la stessa matrice di prima.

Quindi la scrittura

$$\begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},$$

è un modo di rappresentare un vettore, che si può vedere ottenuto come prodotto tra una matrice ed un vettore facendo i prodotti interni delle righe della matrice per le colonne del vettore, oppure ottenuto come c.l. delle colonne della matrice con coefficienti dati dalle componenti del vettore.

In generale allora, se $f \in \mathcal{L}_{n,m}$, dalla linearità segue che

$$f(x) = f(x_1 u^1 + x_2 u^2 + \dots + x_n u^n) = x_1 f(u^1) + x_2 f(u^2) + \dots + x_n f(u^n) = Ax,$$

intendendo che A è la matrice ottenuta disponendo in colonna le immagini dei vettori fondamentali del dominio di f (cioè \mathbb{R}^n) e che Ax rappresenta una c.l. delle colonne di A con coefficienti nel vettore x .

La matrice A si chiama *matrice di rappresentazione della trasformazione lineare f* .

▷ **Osservazione** Si noti anche che viceversa, data una generica matrice A di m righe ed n colonne, se diamo alla scrittura Ax il significato appena detto, e cioè

$$Ax = x_1 a^1 + x_2 a^2 + \dots + x_n a^n,$$

dove gli a^j sono le colonne della matrice A , allora la scrittura $f(x) = Ax$, per ogni $x \in \mathbb{R}^n$ definisce una funzione di \mathbb{R}^n in \mathbb{R}^m . Tale funzione è una trasformazione lineare.³⁸ Pertanto possiamo dire che c'è una corrispondenza biunivoca

³⁸ Infatti

$$\begin{aligned} \text{(per definizione di } f) \quad f(x+y) &= A(x+y) \\ \text{(per definizione di } Ax) &= (x_1 + y_1)a^1 + \dots + (x_n + y_n)a^n \\ \text{(prop. distributiva della molt. scalare in } \mathbb{R}^m) &= x_1 a^1 + y_1 a^1 + \dots + x_n a^n + y_n a^n \end{aligned}$$

tra l'insieme $\mathcal{L}_{n,m}$ delle trasformazioni lineari di \mathbb{R}^n in \mathbb{R}^m e le matrici di m righe ed n colonne. Questo risultato viene talvolta indicato come *teorema di rappresentazione delle trasformazioni lineari*.

▷ **Osservazione** Si noti che, non appena c'è l'esigenza di dare alla trasformazione lineare una qualche rappresentazione, occorre scrivere i vettori rispetto ad una qualche base dei rispettivi spazi vettoriali. In quanto fatto finora si è implicitamente inteso che tale rappresentazione sia rispetto alle basi fondamentali (di \mathbb{R}^n e di \mathbb{R}^m). Quindi, anche se la trasformazione non dipende dal sistema di riferimento, in quanto stabilisce una corrispondenza tra vettori di due spazi, la matrice che rappresenta la trasformazione dipende dalla base che utilizzo. Possiamo dire che ogni trasformazione lineare può essere rappresentata da molte matrici, a seconda della base, e che una matrice può rappresentare molte trasformazioni lineari, sempre a seconda della base che scelgo. Vedremo più avanti come si modifica la matrice di rappresentazione al variare della base dello spazio.

▷ **Osservazione** Se $f \in \mathcal{L}_{n,m}$, allora la sua matrice di rappresentazione ha n colonne (come la dimensione di \mathbb{R}^n) ed m righe (come la dimensione di \mathbb{R}^m).

▷ **Osservazione** La definizione di matrice di rappresentazione di una trasformazione lineare $f \in \mathcal{L}_{n,m}$ assume implicitamente che, sia in \mathbb{R}^n , sia in \mathbb{R}^m , i vettori vengano scritti rispetto alla base fondamentale. Vedremo più avanti come ottenere la matrice di rappresentazione rispetto ad altre basi. Per il momento quindi, parlando di matrice di rappresentazione, sottintendiamo che sia rispetto alle basi fondamentali.

3.2 Matrici

Per noi, quindi, una matrice non è che un modo per rappresentare una trasformazione lineare.

Solitamente le matrici si indicano con lettere maiuscole. Come già osservato, la matrice che rappresenta una $f \in \mathcal{L}_{n,m}$ ha m righe ed n colonne: diremo che è una matrice $m \times n$.

Indicheremo con \mathcal{M}_{mn} l'insieme di tutte le matrici $m \times n$, cioè delle matrici di rappresentazione delle trasformazioni lineari dello spazio $\mathcal{L}_{n,m}$.

Possiamo vedere un vettore riga come una matrice $1 \times n$, cioè come la rappresentazione di una $f \in \mathcal{L}_{n,1}$; possiamo vedere un vettore colonna come una matrice $m \times 1$, e quindi come la rappresentazione di una $f \in \mathcal{L}_{1,m}$.

In alcuni casi potrà essere utile vedere in una matrice solo l'aspetto generale di rappresentazione di una trasformazione lineare, in altri sarà utile vedere aspetti più particolari, come ad esempio i singoli elementi della matrice. Chiameremo *elemento di posto* (i, j) il numero che si trova nella riga i e colonna j della matrice. Se la matrice viene indicata con A , è consuetudine indicare con a_{ij} il suo elemento di posto (i, j) .

La matrice di rappresentazione della trasformazione nulla è la *matrice nulla* e tutti i suoi elementi sono nulli.

Se A è la matrice di rappresentazione della trasformazione f , allora la matrice che rappresenta $-f$ si indica con $-A$ e si chiama l'opposta di A : il suo elemento di posto (i, j) è l'opposto del corrispondente elemento di A .

Una matrice si dice *quadrata* se rappresenta una trasformazione $f \in \mathcal{L}_{n,n}$. In una matrice quadrata il numero delle righe è uguale al numero delle colonne. Questo numero è detto l'*ordine* della matrice quadrata. Quindi, dicendo ad esempio che A è una matrice quadrata di ordine 3, si intende che A è una matrice 3×3 , che rappresenta cioè una trasformazione lineare $f: \mathbb{R}^3 \rightarrow \mathbb{R}^3$.

In una matrice quadrata gli elementi di posto (i, j) , con $i = j$, formano la *diagonale principale* della matrice.

Definizione Se $A \in \mathcal{M}_{mn}$, si chiama matrice *trasposta* di A , e si indica con A^T , la matrice di n righe ed m colonne il cui elemento di posto (i, j) è a_{ji} .

▷ **Osservazione** Si può anche dire che A^T è la matrice che si ottiene da A scambiando le righe con le colonne. Non si tratta, come forse si potrebbe pensare, della matrice che rappresenta la trasformazione inversa.

Definizione Una matrice quadrata A si dice *simmetrica* se $A^T = A$, e quindi se $a_{ij} = a_{ji}$, per ogni i, j .

▷ **Esempio** E' simmetrica la matrice

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

$$\begin{aligned} \text{(per definizione di } Ax) &= Ax + Ay \\ \text{(per definizione di } f) &= f(x) + f(y). \end{aligned}$$

Così è dimostrata l'additività di f .

Analogamente (lo studente rifletta su quali proprietà vengono utilizzate per ottenere i passaggi indicati)

$$f(\alpha x) = A(\alpha x) = (\alpha x_1)a^1 + \dots + (\alpha x_n)a^n = \alpha(x_1a^1 + \dots + x_na^n) = \alpha Ax,$$

e questo prova l'omogeneità.

Definizione Una matrice quadrata A è *diagonale* se risulta $a_{ij} = 0$ per ogni i, j , con $i \neq j$.

Quindi una matrice è diagonale se sono nulli gli elementi che *non* stanno sulla diagonale principale. Faccio osservare che nella definizione non si fa riferimento agli elementi che stanno sulla diagonale principale, i quali possono essere nulli oppure no.

Una matrice quadrata A è *triangolare superiore* se risulta $a_{ij} = 0$ per ogni i, j , con $i > j$. Significa quindi che sono nulli gli elementi che stanno “al di sotto” della diagonale principale.

Una matrice quadrata A è invece *triangolare inferiore* se risulta $a_{ij} = 0$ per ogni i, j , con $i < j$. Significa quindi che sono nulli gli elementi stanno “al di sopra” della diagonale principale.

Risulta chiaro dalle definizioni che una matrice diagonale è anche triangolare superiore e triangolare inferiore.

▷ **Esempio** Delle seguenti tre matrici quadrate, la prima è diagonale, la seconda è triangolare superiore, la terza è triangolare inferiore:

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

Un particolare ed importante esempio di matrice diagonale è la matrice di rappresentazione della trasformazione lineare identità da \mathbb{R}^n ad \mathbb{R}^n , cioè la funzione $1_n(x) = x$, per ogni $x \in \mathbb{R}^n$. Dato che

$$1_n(x) = x_1 u^1 + x_2 u^2 + \dots + x_n u^n,$$

le colonne della sua matrice di rappresentazione sono i vettori fondamentali. Quindi questa matrice ha 1 sulla diagonale principale e 0 fuori di questa. Tale matrice si chiama ovviamente *matrice identità* e si indica di solito con 1 (oppure I) (o con 1_n (I_n) se si vuole precisare che si tratta dell'identità in \mathbb{R}^n).

È naturale a questo punto definire nell'insieme delle matrici alcune *operazioni*.

Le prime due sono immediate, perché derivano direttamente dalle operazioni definite sulle trasformazioni lineari, di cui le matrici sono rappresentazioni.

Nell'insieme \mathcal{M}_{mn} , se A, B rappresentano rispettivamente due trasformazioni $f, g \in \mathcal{L}_{n,m}$, la matrice che rappresenta $f + g$ deve necessariamente essere quella che si ottiene da A, B sommando gli elementi di posto corrispondente.³⁹ Tale matrice verrà indicata con $A + B$.

Analogamente, se A rappresenta la trasformazione f e $\alpha \in \mathbb{R}$, la matrice che rappresenta la trasformazione αf deve necessariamente essere quella che si ottiene da A moltiplicando tutti i suoi elementi per α . Si indica naturalmente con αA .

Abbiamo quindi definito nell'insieme \mathcal{M}_{mn} una addizione tra matrici e una moltiplicazione di una matrice per uno scalare. Non è difficile intuire che con queste operazioni l'insieme \mathcal{M}_{mn} acquista la struttura di spazio vettoriale su \mathbb{R} .⁴⁰ Lo 0 dello spazio \mathcal{M}_{mn} è chiaramente la matrice nulla $m \times n$.

▷ **Esercizio** Dimostrare che la somma di matrici diagonali è una matrice diagonale, che la somma di matrici triangolari superiori è una matrice triangolare superiore e che la somma di matrici triangolari inferiori è una matrice triangolare inferiore.

A questo punto ci si aspetta un prodotto tra matrici e la prima cosa che viene in mente è la matrice che rappresenta il prodotto delle due trasformazioni. Possiamo per esercizio seguire questa strada, che però non si rivela la più interessante.

Definire il prodotto di due trasformazioni lineari non è così immediato come il prodotto di due funzioni reali di variabile reale. Infatti, scrivendo

$$(fg)(x) = f(x) \cdot g(x)$$

occorre specificare che cosa intendiamo a secondo membro, dato che si tratta in generale di vettori. L'unico prodotto che abbiamo visto in questo ambito è il prodotto interno, quindi possiamo definire

$$(fg)(x) = \langle f(x), g(x) \rangle.$$

La funzione fg è ovviamente una funzione di \mathbb{R}^n in \mathbb{R} . Il problema è che non è una trasformazione lineare. Infatti si ha

$$(fg)(\alpha x) = \langle f(\alpha x), g(\alpha x) \rangle = \langle \alpha f(x), \alpha g(x) \rangle = \alpha^2 \langle f(x), g(x) \rangle = \alpha^2 (fg)(x).$$

³⁹ Se indichiamo con a^j e b^j le colonne di A e B , allora $f(x) = Ax = \sum_{j=1}^n x_j a^j$ e $g(x) = Bx = \sum_{j=1}^n x_j b^j$. Quindi

$$(f+g)(x) = f(x) + g(x) = Ax + Bx = \sum_{j=1}^n x_j a^j + \sum_{j=1}^n x_j b^j = \sum_{j=1}^n x_j (a^j + b^j).$$

⁴⁰ Si intuisce anche che che gli spazi vettoriali \mathcal{M}_{mn} e $\mathcal{L}_{n,m}$ hanno “forti somiglianze” e che entrambi hanno forti somiglianze con \mathbb{R}^{mn} , cioè lo spazio euclideo di dimensione mn .

La via da seguire è un'altra, quella di pensare al prodotto di due trasformazioni in termini di composizione delle due.

Ricordo intanto che, se $f, g \in \mathcal{L}_{n,m}$, allora non si può in generale parlare di funzione composta $f \circ g$ (o $g \circ f$) a meno che non sia $m = n$.⁴¹

Il caso più generale è quindi il seguente: sia $f \in \mathcal{L}_{n,m}$ e $g \in \mathcal{L}_{m,p}$. Allora esiste certamente la funzione composta $g \circ f : \mathbb{R}^n \rightarrow \mathbb{R}^p$. Vale il seguente risultato:

Teorema 1 Se $f \in \mathcal{L}_{n,m}$ e $g \in \mathcal{L}_{m,p}$, allora $g \circ f \in \mathcal{L}_{n,p}$.

Dimostrazione Occorre dimostrare l'additività e l'omogeneità della funzione composta. Si ha

$$(g \circ f)(x + y) = g(f(x + y)) = g(f(x) + f(y)) = g(f(x)) + g(f(y)) = (g \circ f)(x) + (g \circ f)(y)$$

e

$$(g \circ f)(\alpha x) = g(f(\alpha x)) = g(\alpha f(x)) = \alpha g(f(x)) = \alpha (g \circ f)(x).$$

Quindi la linearità è dimostrata. □

Vediamo ora come si ottiene la matrice di rappresentazione della trasformazione composta.

Siano dunque $f \in \mathcal{L}_{n,m}$ e $g \in \mathcal{L}_{m,p}$ e siano A e B le rispettive matrici di rappresentazione, quindi abbiamo $f(x) = Ax$, con $x \in \mathbb{R}^n$ e $g(y) = By$, con $y \in \mathbb{R}^m$.⁴² Siano poi

a^j le colonne di A , e quindi $a^j = f(u^j)$, con $j = 1, \dots, n$

b^i le colonne di B , con $i = 1, \dots, m$.

Sia ora C la matrice di rappresentazione della trasformazione composta $g \circ f$, e siano c^j le sue colonne, con $j = 1, \dots, n$. Si ha

$$\begin{aligned} c^j &= (g \circ f)(u^j) = g(f(u^j)) = g(a^j) = Ba^j \\ &\quad (\text{cioè una c.l. delle colonne di } B \text{ con coefficienti in } a^j) \\ &= b^1 a_1^j + b^2 a_2^j + \dots + b^m a_m^j. \end{aligned}$$

La matrice C si indica con BA e si chiama *prodotto righe per colonne delle matrici B e A* .

▷ **Osservazione** L'elemento di posto (i, j) della matrice $C = BA$ si ottiene ovviamente considerando l' i -esima componente del vettore c^j , e cioè

$$c_{ij}^j = c_{ij} = b_i^1 a_1^j + b_i^2 a_2^j + \dots + b_i^m a_m^j.$$

Se, come consuetudine, anziché scrivere l'indice di colonna ad esponente lo scriviamo a fianco dell'indice di riga (come indicato per l'elemento della matrice C), riferendoci agli elementi delle matrici, si ottiene

$$\begin{aligned} c_{ij} &= b_{i1} a_{1j} + b_{i2} a_{2j} + \dots + b_{im} a_{mj} \\ &= \sum_{k=1}^m b_{ik} a_{kj} \\ &= \langle i\text{-esima riga di } B, j\text{-esima colonna di } A \rangle. \end{aligned}$$

Pertanto possiamo dire a conclusione che la matrice prodotto righe per colonne di due matrici date è la matrice che ha come elemento di posto (i, j) il prodotto interno della i -esima riga della prima per la j -esima colonna della seconda.

▷ **Esempio** Siano

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

Risulta

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

Seguono ora alcune osservazioni.

⁴¹ Si ricordi che $(f \circ g)(x)$ significa $f(g(x))$ e che quindi occorre che $g(x)$ appartenga al dominio di f .

⁴² Ricordo che questo significa che $f(x)$ è c.l. delle colonne di A con coefficienti in x e che $g(y)$ è c.l. delle colonne di B con coefficienti in y .

▷ Osservazioni

- Date due matrici A e B , per poter effettuare il prodotto righe per colonne è necessario e sufficiente che il numero di colonne di A sia uguale al numero di righe di B .
- Il prodotto righe per colonne AB (quando si può fare) è una matrice che ha tante righe quante ne ha A e tante colonne quante ne ha B (si dia una motivazione a questa regola pensando al significato delle due matrici).
- Il prodotto tra matrici ha come caso particolare il prodotto di una matrice per un vettore, o di un vettore per una matrice, o anche di un vettore per un vettore.

Ad esempio, se

$$A = \begin{pmatrix} 0 & 1 & -1 \\ 1 & 2 & 0 \end{pmatrix} \quad \text{e} \quad x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix},$$

possiamo scrivere

$$Ax = \begin{pmatrix} -1 \\ 5 \end{pmatrix}.^{43}$$

- Un caso particolare in cui il prodotto tra due matrici si può fare è quando le matrici sono quadrate (ovviamente devono essere anche dello stesso ordine). Questo perché la composizione di due trasformazioni lineari di uno spazio in sé ovviamente è sempre possibile. Si verifica facilmente su qualche esempio che in questo caso il prodotto *non è commutativo*. Quindi, se A e B sono matrici quadrate, può succedere che AB sia diverso da BA .

Consideriamo il seguente esempio:

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \quad \text{e} \quad B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Risulta

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

mentre

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

Nel prodotto tra matrici occorre dunque fare attenzione all'ordine in cui si effettua l'operazione: quindi moltiplicare A *a destra* per B non equivale in genere a moltiplicare A *a sinistra* per B .⁴⁴

- Non vale, nel prodotto tra matrici, la *legge di annullamento del prodotto*, alla quale siamo abituati in \mathbb{R} . Può cioè succedere che il prodotto di due matrici non nulle sia la matrice nulla.

Ad esempio, questo succede con

$$A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad \text{e} \quad B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

- Nell'insieme delle matrici quadrate $n \times n$ la matrice identità 1_n è elemento neutro del prodotto. È facile dimostrare infatti che si ha $A1_n = 1_n A = A$, qualunque sia la matrice $A \in \mathcal{M}_{nn}$.

▷ **Esercizio** Si dimostri che il prodotto di due matrici diagonali è una matrice diagonale, che il prodotto di matrici triangolari superiori è una matrice triangolare superiore e che il prodotto di matrici triangolari inferiori è una matrice triangolare inferiore. Il prodotto tra matrici diagonali è commutativo? E il prodotto tra matrici triangolari?

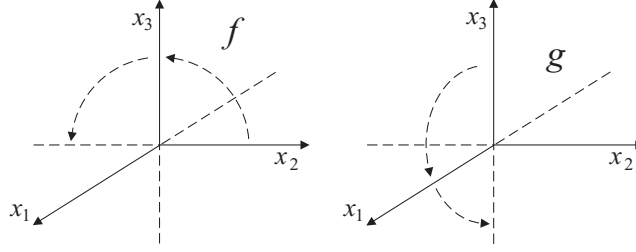
▷ **Esempio** La non commutatività del prodotto tra matrici non è che la conseguenza della non commutatività della composizione di trasformazioni lineari. Questo significa ad esempio che se noi operiamo due rotazioni dello spazio \mathbb{R}^3 in sé, la trasformazione risultante dipende dall'ordine con cui operiamo le due trasformazioni. Verifichiamo per esercizio il tutto su di un esempio particolare.

Consideriamo la trasformazione lineare f nello spazio tridimensionale (vedi figura sotto) che mantiene fisso l'asse x_1 e fa ruotare in senso antiorario gli assi x_2 e x_3 di un angolo di $\pi/2$.⁴⁵ Consideriamo poi la trasformazione g che mantiene fisso l'asse x_2 e fa ruotare in senso antiorario gli assi x_1 e x_3 di un angolo di $\pi/2$.

⁴³ Quindi la scrittura Ax , utilizzata già da un po' per indicare una c.l. delle colonne di A con coefficienti in x , non è altro che un prodotto righe per colonne di due matrici, delle quali la seconda è un vettore colonna.

⁴⁴ Naturalmente, in base a quanto detto prima, può succedere che sia possibile moltiplicare A a destra per B ma non a sinistra.

⁴⁵ Si rifletta sul fatto che tale trasformazione lineare è univocamente individuata da queste condizioni.



Indicando come sempre con u^1, u^2, u^3 i vettori fondamentali di \mathbb{R}^3 , possiamo dire che le due trasformazioni sono univocamente individuate dalle seguenti condizioni:

$$\begin{aligned} f(u^1) &= u^1 \\ f(u^2) &= u^3 \\ f(u^3) &= -u^2 \end{aligned} \quad \text{da cui la matrice di rappresentazione è } A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$\begin{aligned} g(u^1) &= -u^3 \\ g(u^2) &= u^2 \\ g(u^3) &= u^1 \end{aligned} \quad \text{da cui la matrice di rappresentazione è } B = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}$$

Costruiamo ora le trasformazioni composte: con $f \circ g$ si ha

$$\begin{aligned} (f \circ g)(u^1) &= f(g(u^1)) = f(-u^3) = u^2 \\ (f \circ g)(u^2) &= f(g(u^2)) = f(u^2) = u^3 \\ (f \circ g)(u^3) &= f(g(u^3)) = f(u^1) = u^1 \end{aligned} \quad \text{con matrice } AB = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Con $g \circ f$ invece si ha

$$\begin{aligned} (g \circ f)(u^1) &= g(f(u^1)) = g(u^1) = -u^3 \\ (g \circ f)(u^2) &= g(f(u^2)) = g(u^3) = u^1 \\ (g \circ f)(u^3) &= g(f(u^3)) = g(-u^2) = -u^2 \end{aligned} \quad \text{con matrice } BA = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{pmatrix}$$

Lo studente verifichi che le matrici prodotte sono quelle che si ottengono eseguendo il prodotto righe per colonne.

3.3 Immagine e nucleo di una trasformazione lineare

Sia $f \in \mathcal{L}_{n,m}$.

Definizione Definiamo *immagine* di f l'insieme

$$\text{Im } f = \{f(x) : x \in \mathbb{R}^n\}.$$
⁴⁶

Definiamo inoltre *nucleo* di f l'insieme

$$\text{Ker } f = \{x \in \mathbb{R}^n : f(x) = 0\}.$$

▷ **Osservazione** Ovviamente risulta $\text{Im } f \subseteq \mathbb{R}^m$ e $\text{Ker } f \subseteq \mathbb{R}^n$ e il vettore nullo che compare nella definizione di $\text{Ker } f$ è il vettore nullo dello spazio \mathbb{R}^m . Inoltre, dal fatto che $f(0) = 0$ segue che nessuno dei due insiemi ora definiti è vuoto.

⁴⁶ Si noti che si può scrivere in modo equivalente

$$\text{Im } f = \{y \in \mathbb{R}^m : \exists x \in \mathbb{R}^n \text{ tale che } y = f(x)\}.$$

Si ha anche, con altra notazione, $\text{Im } f = f(\mathbb{R}^n)$.

Teorema 2

- i) $\text{Im } f$ è un sottospazio di \mathbb{R}^n ;
 ii) $\text{Ker } f$ è un sottospazio di \mathbb{R}^n .

Dimostrazione Riguardo al primo punto, se $y^1, y^2 \in \text{Im } f$, significa che esistono $x^1, x^2 \in \mathbb{R}^n$ tali che $y^1 = f(x^1)$ e $y^2 = f(x^2)$. Allora, per la linearità di f , se $\alpha_1, \alpha_2 \in \mathbb{R}$, si ottiene

$$\alpha_1 y^1 + \alpha_2 y^2 = \alpha_1 f(x^1) + \alpha_2 f(x^2) = f(\alpha_1 x^1 + \alpha_2 x^2),$$

e questo significa che $\alpha_1 y^1 + \alpha_2 y^2$ è immagine di un vettore di \mathbb{R}^n e cioè appartiene a $\text{Im } f$.

Riguardo al secondo punto, se $x^1, x^2 \in \text{Ker } f$, significa che $f(x^1) = f(x^2) = 0$. Ma allora per $\alpha_1 x^1 + \alpha_2 x^2$, sfruttando ancora la linearità di f , vale

$$f(\alpha_1 x^1 + \alpha_2 x^2) = \alpha_1 f(x^1) + \alpha_2 f(x^2) = 0,$$

e quindi $\alpha_1 x^1 + \alpha_2 x^2 \in \text{Ker } f$. □

▷ **Osservazione** Sia $f \in \mathcal{L}_{n,m}$. Dato che ogni $y \in \text{Im } f$ è c.l. dei vettori $f(u^j)$, cioè i trasformati dei vettori fondamentali di \mathbb{R}^n , allora il sottospazio $\text{Im } f$ è generato dai vettori $f(u^j)$. E dato che tali vettori sono le colonne della matrice di rappresentazione di f , possiamo anche dire che $\text{Im } f$ è generato dalle colonne di tale matrice.

Definizione Se $f \in \mathcal{L}_{n,m}$, si definisce *rango* di f e si indica con $r(f)$ la dimensione dello spazio $\text{Im } f$. Si definisce *nullità* di f e si indica con $\nu(f)$ la dimensione dello spazio $\text{Ker } f$.

Enunciamo, senza dimostrazione, un risultato fondamentale:

Teorema 3 (nullità + rango) Se f è una trasformazione lineare definita in \mathbb{R}^n , allora

$$\nu(f) + r(f) = n.$$

▷ **Esempio** Data la trasformazione lineare $f: \mathbb{R}^4 \rightarrow \mathbb{R}^3$ definita da

$$f \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} = \begin{pmatrix} x - y + t \\ 0 \\ y + z - t \end{pmatrix},$$

determiniamo la dimensione di $\text{Im } f$ e di $\text{Ker } f$. Troviamo poi una base di $\text{Im } f$ e di $\text{Ker } f$. Dall'identità

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

si ricava che i vettori

$$v^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad v^2 = \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad v^3 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad v^4 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$$

sono generatori di $\text{Im } f$. Tali vettori sono linearmente dipendenti, in quanto ad esempio $v^3 = v^1 + v^2$. Quindi il massimo numero di vettori indipendenti è al più 3. Si può osservare poi che $v^4 = -v^2$ e quindi restano v^1 e v^2 , che sono indipendenti. Quindi $\dim \text{Im } f = 2$ e una base di $\text{Im } f$ è formata appunto da v^1 e v^2 .

Dal teorema nullità + rango segue che allora $\dim \text{Ker } f = 2$. Per trovare una base del nucleo possiamo procedere in questo modo. Dall'equazione vettoriale

$$\begin{pmatrix} x - y + t \\ 0 \\ y + z - t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

che caratterizza gli elementi del nucleo, ricaviamo il sistema

$$\begin{cases} x - y + t = 0 \\ y + z - t = 0 \end{cases} \quad \text{che equivale a} \quad \begin{cases} x = y - t \\ z = t - y. \end{cases}$$

I vettori del nucleo sono quindi del tipo $\begin{pmatrix} y-t \\ y \\ t-y \\ t \end{pmatrix}$. Si può chiaramente scrivere

$$\begin{pmatrix} y-t \\ y \\ t-y \\ t \end{pmatrix} = y \begin{pmatrix} 1 \\ 1 \\ -1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 1 \\ 1 \end{pmatrix},$$

e quindi abbiamo trovato due generatori del nucleo.

Si verifica facilmente che due generatori sono indipendenti e che quindi costituiscono una base di $\text{Ker } f$.

La nozione di rango si estende alle matrici con la seguente

Definizione Data una matrice $A \in \mathcal{M}_{mn}$, che rappresenta una trasformazione $f \in \mathcal{L}_{n,m}$, si definisce *rango* di A e si indica con rA il rango della trasformazione f .

▷ **Osservazione** Dato che il rango della matrice A è la dimensione di $\text{Im } f$ e che questo spazio è generato dalle colonne di A , allora rA è il massimo numero di colonne l.i. di A .

Concludiamo il paragrafo con un altro risultato importante, di cui non diamo dimostrazione.

Teorema 4 Data una qualunque matrice A , risulta

$$rA = rA^T.$$

▷ **Osservazione** Mettendo insieme i risultati e le osservazioni precedenti, possiamo allora dire che in una qualunque matrice il massimo numero di colonne indipendenti coincide con il massimo numero di righe indipendenti. Si noti ancora una volta che le righe e le colonne sono vettori appartenenti in genere a spazi diversi.

Non disponiamo ancora di un metodo “comodo” per il calcolo del rango di una matrice. Arriveremo a questo tra un po’.

3.4 Inversione di una trasformazione lineare

In questo paragrafo ci poniamo il problema di studiare l’invertibilità di una trasformazione lineare. Dato che, per motivi che via via risulteranno chiari, l’unico caso significativo a questo proposito è quello delle trasformazioni lineari di uno spazio in sé, per la definizione seguente consideriamo una trasformazione di questo tipo.

Definizione La trasformazione $f \in \mathcal{L}_{n,n}$ si dice *invertibile* se esiste $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ tale che

$$f \circ g = g \circ f = 1_n,$$

avendo indicato con 1_n la trasformazione identità su \mathbb{R}^n .

Se f è invertibile la trasformazione g si chiama *inversa* di f e si indica con f^{-1} .

▷ **Esercizio** Si dimostri che, se f è invertibile, allora anche f^{-1} lo è ed f è la sua trasformazione inversa.

Teorema 5 Se $f \in \mathcal{L}_{n,n}$ ed f è invertibile, allora $f^{-1} \in \mathcal{L}_{n,n}$.

Dimostrazione Presi $x, y \in \mathbb{R}^n$, per ipotesi abbiamo che $f(f^{-1}(x)) = x$ e $f(f^{-1}(y)) = y$. Ma allora

$$f(f^{-1}(x) + f^{-1}(y)) = x + y,$$

e quindi, applicando ad entrambi la funzione f^{-1} ,

$$f^{-1}(x + y) = f^{-1}(x) + f^{-1}(y),$$

e abbiamo dimostrato l’additività della trasformazione inversa.

Analogamente, se $\alpha \in \mathbb{R}$, allora $f(\alpha f^{-1}(x)) = \alpha x$, e quindi $f^{-1}(\alpha x) = \alpha f^{-1}(x)$, e abbiamo dimostrato anche l’omogeneità. \square

Torniamo ora a considerare una generica $f \in \mathcal{L}_{n,m}$. La trasformazione f può avere queste due proprietà:

- i) Per ogni $x \in \mathbb{R}^n$, se $f(x) = 0$, allora $x = 0$;
- ii) Per ogni $y \in \mathbb{R}^m$ esiste un $x \in \mathbb{R}^n$ tale che $f(x) = y$.

▷ **Esercizio** Si dimostri che la proprietà i) equivale a dire che per ogni $x, y \in \mathbb{R}^n$, se $f(x) = f(y)$, allora $x = y$ e che questa a sua volta equivale a dire che per ogni $x, y \in \mathbb{R}^n$, se $x \neq y$, allora $f(x) \neq f(y)$.

▷ **Osservazione** La proprietà i) si esprime dicendo che f è *iniettiva*. La proprietà ii) invece si esprime dicendo che f è *suriettiva*. Possiamo dire inoltre che f è iniettiva se e solo se $\text{Ker } f = \{0\}$ e che f è suriettiva se e solo se $\text{Im } f = \mathbb{R}^m$ (mentre in generale si ha $\text{Im } f = f(\mathbb{R}^n) \subset \mathbb{R}^m$).

Sull'invertibilità di una trasformazione sussiste questo risultato:

Teorema 6 Una trasformazione $f \in \mathcal{L}_{n,n}$ è invertibile se e solo se valgono le proprietà i) e ii) (contemporaneamente), cioè f è invertibile se e solo se è iniettiva e suriettiva.

Dimostrazione Supponiamo che esista la trasformazione inversa f^{-1} . Se $f(x) = 0$, allora $f^{-1}(f(x)) = f^{-1}(0)$, da cui segue che $x = 0$, dato che per il teorema precedente f^{-1} è lineare. Inoltre, per ogni y si ha $y = f(f^{-1}(y))$, e quindi la ii) vale con $x = f^{-1}(y)$.

Viceversa, supponiamo che valgano la i) e la ii). Allora, dato y , per la ii) esiste x tale che $f(x) = y$ e, per la i), tale x è unico.⁴⁷ Sia allora g la funzione che, dato y , associa ad y l'unico x tale che $f(x) = y$. Dimostriamo ora che g è l'inversa di f .

$$\forall y \quad (f \circ g)(y) = f(g(y)) = y,$$

dato che per costruzione $g(y)$ ha per immagine y attraverso f . Analogamente

$$\forall x \quad (g \circ f)(x) = g(f(x)) = x.$$

□

▷ **Esempi** Una trasformazione $f \in \mathcal{L}_{n,m}$ può essere iniettiva ma non suriettiva, oppure suriettiva ma non iniettiva.

- Si consideri ad esempio la trasformazione lineare $f \in \mathcal{L}_{1,2}$, definita da

$$f(x) = \begin{pmatrix} x \\ 2x \end{pmatrix}.$$

Essa è iniettiva, dato che $\begin{pmatrix} x \\ 2x \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ può aversi solo con $x = 0$. Essa non è suriettiva, dato che ad esempio il vettore $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ non è immagine di nessun x .

- Si consideri ora la trasformazione lineare $f \in \mathcal{L}_{2,1}$, definita da

$$f \begin{pmatrix} x \\ y \end{pmatrix} = x.$$

Essa non è iniettiva, dato che ad esempio i vettori $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ e $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ hanno la stessa immagine. Essa è invece suriettiva, dato che ogni $x \in \mathbb{R}$ è immagine, ad esempio, del vettore $\begin{pmatrix} x \\ 0 \end{pmatrix}$.

Ci sono anche esempi di trasformazioni che non sono né iniettive né suriettive:

- consideriamo la $f \in \mathcal{L}_{2,2}$, definita da

$$f \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 0 \end{pmatrix}.$$

Essa non è iniettiva, dato che ad esempio i vettori $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ e $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ hanno la stessa immagine. Essa non è nemmeno suriettiva, dato che ad esempio il vettore $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ non è immagine di nessun vettore.

▷ **Esercizio** Si provi che in realtà è sufficiente una sola delle due proprietà (iniettività o suriettività) per garantire l'invertibilità di una trasformazione lineare $f \in \mathcal{L}_{n,n}$.

▷ **Osservazione** Per capire meglio come opera una trasformazione invertibile, si possono dimostrare due risultati interessanti.

⁴⁷ Se fosse anche $f(z) = y$, allora avremmo $f(x - z) = 0$ e per la i) questo significherebbe $x = z$.

- i) Se f è iniettiva, allora trasforma vettori l.i. in vettori l.i.;
- ii) Se f è suriettiva, allora trasforma generatori in generatori;
- iii) Se f è invertibile, allora trasforma basi in basi.

Dimostrazione

- i) Sia f iniettiva e siano v^1, v^2, \dots, v^k vettori l.i. di \mathbb{R}^n . Consideriamo i vettori $f(v^1), f(v^2), \dots, f(v^k)$. Presa una loro c.l. nulla,

$$\alpha_1 f(v^1) + \alpha_2 f(v^2) + \dots + \alpha_k f(v^k) = 0$$

segue che

$$f(\alpha_1 v^1 + \alpha_2 v^2 + \dots + \alpha_k v^k) = 0,$$

quindi $\alpha_1 v^1 + \alpha_2 v^2 + \dots + \alpha_k v^k = 0$ e quindi (i v^i sono l.i.) tutti gli α_i sono nulli.

Questo prova che i vettori $f(v^1), f(v^2), \dots, f(v^k)$ sono l.i.

- ii) Sia f suriettiva e siano v^1, v^2, \dots, v^k generatori di \mathbb{R}^n . Consideriamo i vettori $f(v^1), f(v^2), \dots, f(v^k)$. Preso un qualunque $y \in \mathbb{R}^m$, dato che f è suriettiva, avremo che esiste un $x \in \mathbb{R}^n$ tale che $y = f(x)$. Ma, essendo v^1, v^2, \dots, v^k generatori, si potrà scrivere $x = \alpha_1 v^1 + \alpha_2 v^2 + \dots + \alpha_k v^k$. Quindi sarà

$$y = f(x) = f(\alpha_1 v^1 + \alpha_2 v^2 + \dots + \alpha_k v^k) = \alpha_1 f(v^1) + \alpha_2 f(v^2) + \dots + \alpha_k f(v^k).$$

Questo prova che i vettori $f(v^1), f(v^2), \dots, f(v^k)$ sono generatori (di \mathbb{R}^m).

- iii) Segue immediatamente da i) e ii), dato che, se f è invertibile, allora è iniettiva e suriettiva. Quindi se v^1, v^2, \dots, v^n sono una base di \mathbb{R}^n , allora $f(v^1), f(v^2), \dots, f(v^n)$ sono l.i. (per la i)) e generatori (per la ii)), quindi una base (sempre di \mathbb{R}^n).

□

▷ **Esercizio** Si dimostri che in realtà le condizioni viste sono anche sufficienti per le relative proprietà, che cioè

- i) f è iniettiva se e solo se trasforma vettori l.i. in vettori l.i.;
- ii) f è suriettiva se e solo se trasforma generatori in generatori;
- iii) f è invertibile se e solo se trasforma basi in basi.

▷ **Osservazione** Come caso particolare di quanto appena visto, possiamo dire che, se f è invertibile, allora i vettori $f(u^1), f(u^2), \dots, f(u^n)$ sono l.i. e viceversa.

Visti i risultati sulle trasformazioni invertibili, possiamo dare la seguente

Definizione Se A è la matrice di rappresentazione di $f \in \mathcal{L}_{n,n}$ e se f è invertibile, allora si chiama *matrice inversa* di A la matrice di rappresentazione di f^{-1} .

Questo equivale a chiamare *invertibile* una matrice $A \in \mathcal{M}_{nn}$ se esiste un'altra matrice $B \in \mathcal{M}_{nn}$ tale che valga

$$AB = BA = 1_n,$$

dove 1_n è la matrice identità in \mathcal{M}_{nn} .

▷ **Osservazione** Si osservi che, se la matrice A è invertibile, allora le sue colonne a^j sono l.i. (vedi osservazione precedente). Da $AB = 1_n$ si ha inoltre che

$$a^1 b_1^j + a^2 b_2^j + \dots + a^n b_n^j = u^j$$

e cioè b^j , la j -esima colonna di A^{-1} , è la scrittura di u^j nella base $\{a^1, a^2, \dots, a^n\}$.

Si tratta ora di trovare un procedimento di calcolo che, data una matrice, mi dica intanto se essa è invertibile e, in caso affermativo, mi permetta poi di trovare la matrice inversa. Arriveremo a tale procedimento tra un po'.

3.5 Cambiamento di base nello spazio \mathbb{R}^n

In \mathbb{R}^n , sia $\mathcal{U} = \{u^1, u^2, \dots, u^n\}$ la base fondamentale e sia $\mathcal{V} = \{v^1, v^2, \dots, v^n\}$ un'altra base.

Dato un vettore x , avremo $x = x_1 u^1 + x_2 u^2 + \dots + x_n u^n$ e $x = y_1 v^1 + y_2 v^2 + \dots + y_n v^n$. Si pone il seguente problema: se conosco la scrittura di x rispetto alla base fondamentale, come posso ottenere la scrittura di x rispetto alla base \mathcal{V} ?

Consideriamo la trasformazione lineare $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ tale che $f(u^j) = v^j$, per ogni $j = 1, \dots, n$.⁴⁸

Avremo allora

$$f\left(\sum_j x_j u^j\right) = \sum_j y_j v^j.$$

Sia A la matrice di rappresentazione della trasformazione f . Ovviamente le colonne di A sono i vettori della base \mathcal{V} . La matrice A realizza un cambio di base, precisamente dalla base \mathcal{V} alla base fondamentale, nel senso che, data la rappresentazione (y_1, y_2, \dots, y_n) di un vettore nella base \mathcal{V} , Ay fornisce la rappresentazione (x_1, x_2, \dots, x_n) del vettore nella base fondamentale. Infatti

$$Ay = y_1 a^1 + y_2 a^2 + \dots + y_n a^n = x_1 u^1 + x_2 u^2 + \dots + x_n u^n.$$

Tutto questo giustifica la seguente

Definizione Date in \mathbb{R}^n due basi $\mathcal{V} = \{v^1, v^2, \dots, v^n\}$ e $\mathcal{W} = \{w^1, w^2, \dots, w^n\}$, chiamiamo *matrice di cambio di base* da \mathcal{V} a \mathcal{W} la matrice ${}_{\mathcal{W}}B_{\mathcal{V}}$ ⁴⁹ tale che, se y è la rappresentazione di un qualunque vettore nella base \mathcal{V} , allora ${}_{\mathcal{W}}B_{\mathcal{V}}y$ è la rappresentazione del vettore nella base \mathcal{W} .

Pertanto la matrice A , ottenuta in precedenza disponendo in colonna i vettori della base \mathcal{V} , non è altro che ${}_{\mathcal{U}}B_{\mathcal{V}}$, cioè la matrice di cambio di base dalla base \mathcal{V} alla base fondamentale.

La matrice ${}_{\mathcal{U}}B_{\mathcal{V}}$ è invertibile, dato che le sue colonne sono l.i. (sono i vettori di una base). Senza ulteriori dettagli, diciamo che ${}_{\mathcal{V}}B_{\mathcal{U}} = {}_{\mathcal{U}}B_{\mathcal{V}}^{-1}$, è cioè che la matrice di cambio di base dalla base fondamentale alla base \mathcal{V} è la matrice inversa dell'altra.

▷ **Osservazione** Da quanto enunciato, per trovare ${}_{\mathcal{W}}B_{\mathcal{V}}$, basta fare

$${}_{\mathcal{W}}B_{\mathcal{V}} = {}_{\mathcal{W}}B_{\mathcal{U}} {}_{\mathcal{U}}B_{\mathcal{V}} = {}_{\mathcal{U}}B_{\mathcal{W}}^{-1} {}_{\mathcal{U}}B_{\mathcal{V}}.$$

▷ **Esempio** ⁵⁰

Consideriamo, in \mathbb{R}^2 , la base fondamentale $\{u^1, u^2\}$ e la base $\mathcal{V} = \{v^1, v^2\}$ indicata in figura. Si ha $v^1 = (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ e $v^2 = (-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$. Consideriamo poi il vettore x indicato sempre in figura.

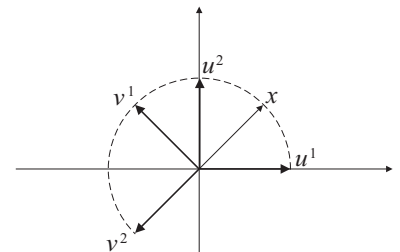
Rispetto alla base \mathcal{V} il vettore x è indicato dalla coppia $(0, -1)$. Rispetto alla base fondamentale è invece indicato dalla coppia $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$.

La matrice di cambio di base è

$${}_{\mathcal{U}}B_{\mathcal{V}} = \begin{pmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

Si ha infatti

$${}_{\mathcal{U}}B_{\mathcal{V}} \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}.$$



3.6 Determinante di una matrice

Veniamo ora alla definizione di determinante di una matrice quadrata.

Quella che presentiamo è una definizione “ricorsiva”. Per avere un esempio (più semplice) di definizione ricorsiva, lo studente consideri quanto segue.

Il fattoriale di un numero naturale $n \geq 1$ si può definire in modo diretto come

$$n! = 1 \cdot 2 \cdot 3 \cdots n.$$

⁴⁸ Si noti che la condizione è sufficiente per definire in modo unico la trasformazione.

⁴⁹ La notazione va letta, per così dire, da destra a sinistra, un po' come nel caso della composizione tra funzioni.

⁵⁰ Non possiamo fornire un esempio di cambio di base che non coinvolga la base fondamentale in quanto non abbiamo ancora detto nulla sul calcolo della matrice inversa.

Allo stesso risultato si perviene però anche definendo

$$\begin{aligned} 0! &= 1 \\ \text{e } n! &= n(n-1)! \quad , \text{ per } n \geq 1. \end{aligned}$$

Questa è una definizione ricorsiva, in quanto nel definire il fattoriale di n si utilizza il fattoriale di $n-1$. La definizione non avrebbe senso se non ci fosse quella che si chiama la base della definizione ricorsiva, e cioè la definizione di $0!$.

La definizione ricorsiva di determinante di una matrice quadrata, che ora presentiamo, procede in modo analogo, definendo il determinante di una matrice di ordine n in funzione del determinante di una matrice di ordine $n-1$.

Sono necessarie alcune definizioni preliminari.

Definizione Se A è una matrice (anche non quadrata), si chiama *sottomatrice* di A una qualunque matrice che si ottiene da A eliminando alcune righe e alcune colonne.

▷ *Esempio* Se

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

una sottomatrice di A è la matrice

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

ottenuta eliminando la seconda e la terza colonna. Un'altra sottomatrice di A è la matrice

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

ottenuta eliminando la terza riga e la prima e terza colonna.

▷ *Osservazione* Le sottomatrici di una matrice data possono essere quadrate oppure no. È evidente che, per ottenere una sottomatrice quadrata da una matrice che non lo è, dovrò eliminare un numero di righe diverso dal numero di colonne, mentre le sottomatrici quadrate di una matrice quadrata si ottengono eliminando un numero di righe uguale al numero di colonne.

Inizia ora la definizione ricorsiva di determinante. Qui parliamo solo di matrici quadrate.

Sia dunque $A \in \mathcal{M}_{nn}$, cioè una matrice quadrata di ordine n . Se $n = 1$, allora il determinante di A è l'unico elemento che la costituisce (un numero reale).⁵¹

Sia ora $n > 1$. Se a_{ij} è l'elemento di posto (i, j) della matrice A , si definisce *minore complementare* di a_{ij} il numero

M_{ij} = determinante della sottomatrice di A ottenuta eliminando la i -esima riga e la j -esima colonna.

È ovvio che la sottomatrice di cui si parla nella definizione di minore complementare è una sottomatrice quadrata di ordine $n-1$. Faccio esplicitamente notare che il minore complementare di a_{ij} non dipende dal valore di a_{ij} .

Si definisce *complemento algebrico* di a_{ij} il numero

$$A_{ij} = (-1)^{i+j} M_{ij}.$$

In pratica il complemento algebrico di a_{ij} coincide col suo minore complementare se la somma degli indici di riga e di colonna è pari, ed è invece l'opposto del minore complementare se tale somma è dispari.

Si definisce infine *determinante* di A il numero reale

$$\det A = a_{11}A_{11} + a_{12}A_{12} + \dots + a_{1n}A_{1n} = \sum_{i=1}^n a_{1i}A_{1i}.$$

Il determinante di una matrice di ordine n viene quindi definito attraverso i determinanti di matrici di ordine $n-1$ (quelli che figurano nei complementi algebrici).

Vediamo qualche esempio. Per le matrici di ordine 1 il calcolo del determinante non crea problemi. Consideriamo una generica matrice quadrata di ordine 2:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

⁵¹ Questa è la base della definizione ricorsiva di determinante.

Risulta

$$\begin{aligned}\det A &= a_{11} \cdot A_{11} + a_{12} \cdot A_{12} \\ &= a_{11} \cdot (-1)^{1+1} M_{11} + a_{12} \cdot (-1)^{1+2} M_{12} \\ &= a_{11} a_{22} - a_{12} a_{21}.\end{aligned}$$

Abbiamo ora una comoda regola per il calcolo del determinante di tutte le matrici quadrate di ordine 2. Consideriamo ora una generica matrice quadrata di ordine 3:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

Procedendo con la definizione abbiamo

$$\begin{aligned}\det A &= a_{11} \cdot A_{11} + a_{12} \cdot A_{12} + a_{13} \cdot A_{13} \\ &= a_{11} \cdot (-1)^{1+1} M_{11} + a_{12} \cdot (-1)^{1+2} M_{12} + a_{13} \cdot (-1)^{1+3} M_{13} \\ &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}.\end{aligned}$$

La formula che risulta non è evidentemente così semplice da ricordare. Esiste però un metodo pratico che aiuta a ricostruire la formula (*regola di Sarrus*): si affiancano alla matrice data nuovamente la prima e la seconda colonna, ottenendo così la seguente matrice 3×5 :

$$\begin{array}{cccccc} \searrow & \searrow & \searrow & \swarrow & \swarrow & \swarrow \\ \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{pmatrix} \end{array}.$$

Ora basta sommare i prodotti degli elementi che figurano nelle diagonal indicate dalle frecce che puntano a SE (\searrow) e sottrarre i prodotti degli elementi delle diagonal indicate dalle frecce che puntano a SO (\swarrow).

Consideriamo ad esempio la matrice

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

Con la definizione di determinante otteniamo

$$\begin{aligned}\det A &= 1 \cdot A_{11} + 1 \cdot A_{12} + 2 \cdot A_{13} \\ &= 1 \cdot (-1)^{1+1} M_{11} + 1 \cdot (-1)^{1+2} M_{12} + 2 \cdot (-1)^{1+3} M_{13} \\ &= M_{11} - M_{12} + 2M_{13} \\ &= -1 - 1 + 4 = 2.\end{aligned}$$

Con la regola di Sarrus, dalla matrice

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

otteniamo

$$\det A = (-1 + 0 + 4) - (0 + 0 + 1) = 2.$$

La regola di Sarrus vale per le matrici di ordine 3. Non c'è una regola analoga per matrici di ordine maggiore di 3. Consideriamo, a titolo di esempio conclusivo, una matrice quadrata di ordine 4:

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

Risulta, con la definizione,

$$\begin{aligned}\det A &= 1 \cdot A_{11} + 0 \cdot A_{12} + (-1) \cdot A_{13} + 1 \cdot A_{14} \\ &= M_{11} - M_{13} - M_{14}.\end{aligned}$$

Procedendo ora sulle sottomatrici di ordine 3 abbiamo

$$\begin{aligned}M_{11} &= 1 \cdot 1 + (-1) \cdot (-1) = 2 ; \\ M_{13} &= 1 \cdot (-1) - 1 \cdot (-1) - 1 \cdot 0 = 0 ; \\ M_{14} &= 1 \cdot (-1) - 1 \cdot (-1) = 0\end{aligned}$$

e quindi $\det A = 2$.

▷ **Osservazione** Nella definizione di determinante, cioè nella formula

$$\det A = a_{11}A_{11} + a_{12}A_{12} + \dots + a_{1n}A_{1n}$$

i calcoli sono svolti con riferimento alla prima riga: praticamente la definizione dice che il determinante di una matrice A è il numero reale che si ottiene facendo la somma dei prodotti degli elementi della prima riga di A per i rispettivi complementi algebrici.

Si può dimostrare un risultato interessante e per nulla prevedibile secondo il quale, se noi effettuiamo lo stesso tipo di calcolo con riferimento ad una *qualunque riga* o *qualunque colonna*, il risultato è sempre lo stesso.

In altre parole, se fissiamo una qualunque riga (o colonna) e facciamo la somma dei prodotti degli elementi di quella riga (o colonna) per i rispettivi complementi algebrici, otteniamo sempre $\det A$. Quindi

$$\begin{aligned}\det A &= a_{i1}A_{i1} + a_{i2}A_{i2} + \dots + a_{in}A_{in} \quad (\text{per qualunque } i = 1, \dots, n) \\ &= a_{1j}A_{1j} + a_{2j}A_{2j} + \dots + a_{nj}A_{nj} \quad (\text{per qualunque } j = 1, \dots, n).\end{aligned}$$

Questo risultato va sotto il nome di *primo teorema di Laplace*.

Rispetto a quale riga o colonna calcolare il determinante di una matrice è quindi una scelta di convenienza. Ovviamente conviene operare rispetto alla riga o colonna che contiene il maggior numero di zeri.

▷ **Osservazione** Si può altresì dimostrare che, sommando i prodotti degli elementi di una certa riga (o colonna) per i complementi algebrici dei corrispondenti elementi di *un'altra* riga (o colonna), si ottiene sempre zero (e questo risultato va sotto il nome di *secondo teorema di Laplace*). Lo studente è invitato a verificare questo risultato su di una matrice a sua scelta.

▷ **Esercizio** Dimostrare che il determinante di una matrice diagonale è il prodotto degli elementi che stanno sulla diagonale principale. Dimostrare poi che lo stesso avviene con una matrice triangolare superiore o triangolare inferiore.

Elenco qui di seguito alcune tra le principali *proprietà del determinante*.

Siano a^1, a^2, \dots, a^n vettori colonna di \mathbb{R}^n .⁵² Indico con $(a^1 \ a^2 \ \dots \ a^n)$ la matrice che si ottiene affiancando tali vettori. Valgono le proprietà:

i) Proprietà di *linearità* del determinante.

a. (*additività* del determinante): se per la generica colonna a^k si ha $a^k = b^k + c^k$, allora

$$\det(a^1 \ \dots \ a^k \ \dots \ a^n) = \det(a^1 \ \dots \ b^k + c^k \ \dots \ a^n) = \det(a^1 \ \dots \ b^k \ \dots \ a^n) + \det(a^1 \ \dots \ c^k \ \dots \ a^n).$$

b. (*omogeneità* del determinante): se per la generica colonna a^k si ha $a^k = \alpha b^k$, allora

$$\det(a^1 \ \dots \ a^k \ \dots \ a^n) = \det(a^1 \ \dots \ \alpha b^k \ \dots \ a^n) = \alpha \det(a^1 \ \dots \ b^k \ \dots \ a^n).$$

ii) $\det(\alpha A) = \alpha^n \det A$: immediata conseguenza dell'omogeneità del determinante (si noti che moltiplicare per α la matrice vuol dire moltiplicare per α tutti i suoi elementi).

iii) È nullo il determinante di una matrice che ha una riga (o una colonna) nulla: la cosa è ovvia se si pensa di calcolare il determinante rispetto a quella riga (o colonna), ma si può anche ottenere come conseguenza della proprietà di linearità del determinante.

⁵² Lo stesso se si trattasse di vettori riga.

- iv) Scambiando tra loro due righe (o due colonne) il determinante cambia segno.
- v) Il determinante è nullo se vi sono due righe (o due colonne) uguali: immediata conseguenza della precedente.
- vi) Il determinante non cambia se si aggiunge ad una riga un'altra riga moltiplicata per una costante (lo stesso con le colonne): immediata conseguenza della linearità e della precedente.
- vii) Il determinante è nullo se le righe (o le colonne) sono linearmente dipendenti.

▷ **Osservazione** Faccio notare che i punti 3,5 e 7 forniscono condizioni sufficienti per l'annullarsi del determinante. Si può dimostrare che solo la condizione espressa al punto 7 è anche necessaria, e ovviamente raccoglie in sé le altre come casi particolari.

Risulta utile il seguente risultato, di cui non diamo dimostrazione.

Teorema 7 (di Binet) Se A e B sono matrici quadrate dello stesso ordine, vale l'uguaglianza

$$\det(AB) = \det A \cdot \det B.$$

▷ **Osservazione** Si noti che, come conseguenza del teorema di Binet, le matrici AB e BA possono essere diverse ma hanno lo stesso determinante.

3.7 Calcolo della matrice inversa

Torniamo a parlare ora di matrice inversa. Abbiamo già visto in precedenza che cosa significa che una matrice è invertibile. Non abbiamo ancora però a disposizione un procedimento di calcolo che consenta di dire se una matrice è invertibile e di calcolare la matrice inversa.

Un metodo generale, ma di non facile applicazione se l'ordine della matrice è maggiore di 2, è quello che vediamo ora su questo esempio.

▷ **Esempio** Consideriamo la matrice

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

Cerchiamo una matrice $B = \begin{pmatrix} x & y \\ z & t \end{pmatrix}$ tale che $AB = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Si ottiene

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

pertanto deve essere $B = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$ e si vede subito che anche $BA = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Pertanto A è invertibile e B è la sua matrice inversa.

Consideriamo ora la matrice

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

Procedendo come prima, calcoliamo

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

È evidente che la matrice a secondo membro non può essere la matrice identità per nessun valore delle incognite. In questo caso la matrice A non è invertibile. Si osservi che la matrice che risultava prima invertibile ha determinante 1 e quest'ultima invece ha determinante 0.

Ecco ora il risultato fondamentale.

Teorema 8 Una matrice A è invertibile se e solo se il suo determinante non è zero.

Dimostrazione Se A è invertibile, allora esiste la sua matrice inversa A^{-1} ed è tale che $AA^{-1} = 1$.⁵³

⁵³ Ricordo che con 1 (o con 1_n se c'è la necessità di specificare l'ordine) indico la matrice identità.

Allora, ricordando il teorema di Binet, si ha

$$\det(AA^{-1}) = \det A \cdot \det A^{-1} = \det 1 = 1.^{54}$$

Dunque non può essere $\det A = 0$.

Viceversa, supponiamo che sia $\det A \neq 0$.

Considero la matrice B in cui $b_{ij} = A_{ji}$, cioè la matrice il cui elemento di posto (i, j) è il complemento algebrico dell'elemento di posto (j, i) della matrice A .

Ora considero la matrice prodotto AB . La sua j -esima colonna è

$$a^1 b_{1j} + a^2 b_{2j} + \dots + a^n b_{nj} = a^1 A_{j1} + a^2 A_{j2} + \dots + a^n A_{jn}.$$

Il generico i -esimo elemento di tale vettore colonna è dunque

$$a_{i1} A_{j1} + a_{i2} A_{j2} + \dots + a_{in} A_{jn}.$$

Ora abbiamo che, se $i = j$, tale elemento è

$$a_{i1} A_{i1} + a_{i2} A_{i2} + \dots + a_{in} A_{in} = \det A,$$

dato che si tratta della somma dei prodotti della i -esima riga di A per i rispettivi complementi algebrici (primo teorema di Laplace).

Se invece $i \neq j$, allora tale elemento è

$$a_{i1} A_{j1} + a_{i2} A_{j2} + \dots + a_{in} A_{jn} = 0,$$

dato che si tratta della somma dei prodotti degli elementi di una certa riga di A per i complementi algebrici degli elementi di un'altra riga (secondo teorema di Laplace).

Ma allora la j -esima colonna della matrice AB è un vettore che ha un unico elemento non nullo, uguale a $\det A$, al posto j . Si tratta cioè del vettore $\det A \cdot u^j$, dove come sempre u^j è il j -esimo vettore fondamentale. Quindi possiamo dire che

$$AB = \det A \cdot 1$$

e quindi, dividendo per $\det A$ (che non è nullo), possiamo dire che

$$A \cdot \frac{B}{\det A} = 1.$$

Si dimostra nello stesso modo che anche $\frac{B}{\det A} \cdot A = 1^{55}$ e che quindi la matrice

$$\frac{B}{\det A} = \frac{1}{\det A} \cdot B$$

è la matrice inversa di A . □

▷ **Osservazione** Il teorema appena dimostrato fornisce un metodo per il calcolo della matrice inversa, quando questa esiste. La matrice B che si incontra nella dimostrazione si chiama *matrice aggiunta di A* e si indica di solito con A^* . La matrice aggiunta di A è quindi la matrice *trasposta*⁵⁶ della matrice dei complementi algebrici di A .

Pertanto, se A è invertibile, si ha

$$A^{-1} = \frac{1}{\det A} \cdot A^*.$$

Definizione Si dicono *singolari* le matrici con determinante nullo (e quindi non invertibili). Quelle invertibili si dicono anche *non singolari*.

▷ **Esempi**

Vediamo un paio di esempi di calcolo della matrice inversa.

⁵⁴ Il determinante della matrice identità è 1 dato che questa è una matrice diagonale e gli elementi della diagonale principale sono tutti uguali a 1.

⁵⁵ La j -esima colonna di BA è

$$b^1 a_{1j} + b^2 a_{2j} + \dots + b^n a_{nj}$$

e l' i -esimo elemento di tale vettore è

$$b_{i1} a_{1j} + b_{i2} a_{2j} + \dots + b_{in} a_{nj}$$

che, come prima, è $\det A$ se $i = j$ ed è nullo se $i \neq j$. La conclusione segue come prima.

⁵⁶ Si ricordi che $b_{ij} = A_{ji}$.

La matrice

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

è invertibile in quanto $\det A = -2$. La matrice dei complementi algebrici è

$$\begin{pmatrix} 4 & -3 \\ -2 & 1 \end{pmatrix}. \quad \text{La matrice aggiunta è } A^* = \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix}$$

e la matrice inversa è quindi

$$A^{-1} = -\frac{1}{2} \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix}.$$

Consideriamo ora la matrice

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

Essa è invertibile dato che $\det A = -1$. La matrice dei complementi algebrici è

$$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad \text{che, essendo simmetrica, coincide con } A^* = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Pertanto la matrice inversa è

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

3.8 Calcolo del rango

Abbiamo visto in precedenza che il rango rA di una matrice $A \in \mathcal{M}_{mn}$ è il rango della trasformazione lineare che A rappresenta (cioè la dimensione dell'immagine di questa) e che questo coincide con il massimo numero di colonne (o righe) l.i. di A . Ricordando che in uno spazio non ci possono essere più vettori l.i. di quella che è la dimensione dello spazio, si ha subito che risulta sempre

$$0 \leq rA \leq \min\{m, n\}.$$
⁵⁷

▷ **Osservazione** La definizione di rango non ha carattere “operativo”, cioè non suggerisce come si possa determinare il rango nei casi concreti, a meno che non si trovi un modo “operativo” per decidere se dei vettori assegnati sono o non sono indipendenti.

Esiste un tale metodo operativo e mi limito ad enunciare il risultato teorico sul quale tale metodo si fonda. Anzitutto diamo la seguente

Definizione Data la matrice $A \in \mathcal{M}_{mn}$, si chiamano *minori di A di ordine k* i determinanti delle sottomatrici quadrate di A di ordine k .

▷ **Osservazione** Ogni minore di A è un numero reale e non una sottomatrice. L'ordine di ogni minore di una matrice $A \in \mathcal{M}_{mn}$ (con $m, n \geq 1$) è sicuramente (un numero naturale) compreso tra 1 e il minimo tra m ed n .

Vale ora il seguente risultato:

Teorema 9 Sono dati i vettori $v^1, v^2, \dots, v^k \in \mathbb{R}^n$ e sia V la matrice che si ottiene disponendo tali vettori in colonna (o in riga, è lo stesso). I vettori sono indipendenti se e solo se nella matrice V esiste un minore non nullo di ordine k .

▷ **Osservazione** Dal teorema segue immediatamente che, dati i vettori $v^1, v^2, \dots, v^k \in \mathbb{R}^n$, se $k > n$, allora i vettori sono sicuramente linearmente dipendenti. Infatti non può esserci nella matrice V un minore di ordine k in quanto, come già osservato, l'ordine dei minori di V non può superare il minimo tra n e k , che in questo caso è certamente n .

Valendo il risultato dell'ultimo teorema, si può dire allora che il rango di una matrice coincide col massimo ordine dei suoi minori non nulli oppure, equivalentemente, col massimo ordine delle sue sottomatrici quadrate non singolari.

In molti testi lo studente può trovare questa quale definizione di rango di una matrice.

⁵⁷ L'unico caso in cui può risultare $rA = 0$ è quando si tratta della matrice nulla, quindi tipicamente sarà $1 \leq rA \leq \min\{m, n\}$.

Vediamo ora un paio di esempi. Consideriamo la matrice

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

È evidente che A ha sottomatrici quadrate di ordine 1 e 2.

I minori di ordine 1 sono: $2, -2, 0, -1, 1$.

I minori di ordine 2 sono: $0, 2, -2$. Il rango di A è quindi 2.

La matrice

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

ha invece rango 1, dato che tutti i suoi minori di ordine 2 sono nulli.

▷ **Osservazione** È chiaro che in generale non serve trovare tutti i minori della matrice. Se ne trovo uno di ordine k diverso da zero, non serve trovare gli altri minori di ordine k , dato che il rango è sicuramente almeno k .

La matrice

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

può avere al più rango 3. Dato che la terza colonna è nulla, ogni sottomatrice di ordine 3 che contiene la terza colonna ha sicuramente determinante nullo. L'unica sottomatrice di ordine 3 non singolare può essere quella che non contiene la terza colonna. Tale sottomatrice ha determinante uguale a -1 . Quindi il rango di A è 3.

▷ **Osservazione** Si dice che una matrice $A \in \mathcal{M}_{mn}$ ha *rango massimo* (o *rango pieno*) se risulta $rA = \min\{m, n\}$. Se la matrice è quadrata di ordine n , essa ha un solo minore di ordine n , che coincide con $\det A$. Allora una matrice quadrata ha rango massimo se e solo se il suo determinante non è zero, il che equivale al fatto che le righe (e le colonne) sono indipendenti, oppure al fatto che la matrice è invertibile.

Lo studente presti attenzione al fatto che è falso affermare in generale che in una matrice a rango pieno sono linearmente indipendenti sia le righe sia le colonne, a meno che la matrice non sia quadrata.

▷ **Osservazione** Sia $A \in \mathcal{M}_{mn}$. Possiamo affermare che, affinché il rango di A sia k , è necessario e sufficiente che valgano le seguenti due proprietà:

1. esiste un minore non nullo di A di ordine k ;
2. non esistono minori di A di ordine $k+1$ oppure, se ne esistono, sono tutti nulli.⁵⁸

Relativamente al punto 2, è evidente che esistono minori di ordine $k+1$ se e solo se $k < \min\{m, n\}$.

▷ **Esempio** Talvolta la matrice può dipendere da uno (o più) parametri reali. In questi casi è ovvio che il rango della matrice dipende anch'esso dai valori che si attribuiscono ai parametri.

Consideriamo ad esempio la matrice

$$A_x = \begin{pmatrix} x & 0 & 1 & 1 \\ -1 & x & 0 & -1 \\ 1 & 2 & x & 1 \end{pmatrix}.$$

A priori (cioè per il solo fatto che la matrice è 3×4) possiamo dire che i valori possibili del rango della matrice sono 1, 2 o 3.

Da un esame un po' più dettagliato risulta che rA è almeno 2, qualunque sia il valore di x . Esiste infatti, qualunque sia x , un minore non nullo di ordine 2 (quale?).

Ora vediamo se il rango è 3 oppure no. Consideriamo un minore di ordine 3. In questi casi occorre un po' di "scaltrezza". Un minore di una matrice che dipende da un parametro x è un polinomio nella variabile x . Anche se in generale non è vero che il numero di x presenti nella sottomatrice coincida con il grado del polinomio, è comunque bene fare in modo che non ci siano troppe x nella sottomatrice che vado a scegliere.

Quindi la scaltrezza consiste nello scegliere, possibilmente, una sottomatrice che contenga "tanti zeri e poche x ".

La cosa è particolarmente apprezzabile nell'esempio che stiamo considerando. Se scegliamo la sottomatrice di ordine 3 costituita dalle prime 3 colonne di A , il minore corrispondente è

$$x \cdot x^2 + 1 \cdot (-2 - x) = x^3 - x - 2.$$

Possiamo dire che esiste uno zero reale di questo polinomio, ma non siamo in grado di trovarlo in quanto il polinomio non si fattorizza in polinomi di grado inferiore a coefficienti razionali.

⁵⁸ Si noti che, se i minori di ordine $k+1$ sono tutti nulli ed esistono minori di ordine $> k+1$, allora questi sono certamente tutti nulli.

Se scegliamo invece la sottomatrice costituita dalle ultime 3 colonne, il minore corrispondente è

$$-1 \cdot (x+2) + 1 \cdot x^2 = x^2 - x - 2 = (x+1)(x-2).$$

Possiamo ora affermare che, se $x \neq -1$ e $x \neq 2$, il rango è 3, poiché questo minore di ordine 3 è non nullo.

Ci restano da studiare due casi: $x = -1$ e $x = 2$.

Lo studente stia attento a non cadere nella tentazione di affermare che per tali valori di x il rango è sicuramente 2: prima di arrivare a questa conclusione occorre esaminare i minori di ordine 3 non ancora considerati. Di solito conviene sostituire i valori nella matrice e poi calcolare tutti i minori non ancora esaminati (nel nostro caso ce ne sono 3 oltre a quello già considerato): se ne troviamo uno non nullo, possiamo concludere che il rango è 3; il rango è invece 2 se tutti e 3 i minori di ordine 3 sono nulli.

Nel nostro caso, dato che il minore relativo alla sottomatrice formata dalle prime tre colonne risulta essere il polinomio $P(x) = x^3 - x - 2$, possiamo intanto calcolare tale polinomio per i valori $x = -1$ e $x = 2$. Risulta $P(-1) = -2$ e $P(2) = 4$. Entrambi i valori sono non nulli e quindi concludiamo che il rango di A è 3, qualunque sia $x \in \mathbb{R}$.

Consideriamo ora la matrice

$$A_x = \begin{pmatrix} x & 0 & 1 & 1 \\ -1 & x & -1 & -1 \\ 1 & 2 & x & 1 \end{pmatrix}.$$

Possiamo subito affermare che il rango di A è almeno 2 in quanto il determinante della sottomatrice di ordine 2 formata da 2^a e 3^a riga e 2^a e 3^a colonna è non nullo qualunque sia x .

Poi, con procedimento analogo a quello usato nell'esempio precedente, il minore relativo alla sottomatrice formata dalle ultime 3 colonne è

$$-1 \cdot (x+2) + 1 \cdot (x^2+2) = x^2 - x = x(x-1).$$

Possiamo dire allora che, se $x \neq 0$ e $x \neq 1$, il rango è 3.

Se $x = 0$ oppure se $x = 1$ si ottengono rispettivamente le due matrici

$$A_0 = \begin{pmatrix} 0 & 0 & 1 & 1 \\ -1 & 0 & -1 & -1 \\ 1 & 2 & 0 & 1 \end{pmatrix} \quad \text{e} \quad A_1 = \begin{pmatrix} 1 & 0 & 1 & 1 \\ -1 & 1 & -1 & -1 \\ 1 & 2 & 1 & 1 \end{pmatrix}.$$

Il rango di A_0 è 3 in quanto la sottomatrice formata dalle prime tre colonne è non singolare.

Il rango di A_1 è invece 2. Si può infatti osservare che ci sono tre colonne uguali. Pertanto ogni sottomatrice quadrata di ordine 3 è singolare, avendo sicuramente almeno due colonne uguali.

Ora torniamo con un esempio conclusivo alla questione del cambiamento di base.

Abbiamo visto in precedenza che, se

$$\mathcal{V} = \{v^1, v^2, \dots, v^n\} \quad \text{e} \quad \mathcal{W} = \{w^1, w^2, \dots, w^n\}$$

sono due basi di \mathbb{R}^n , allora la matrice che realizza il cambio di base da \mathcal{V} a \mathcal{W} è la matrice⁵⁹

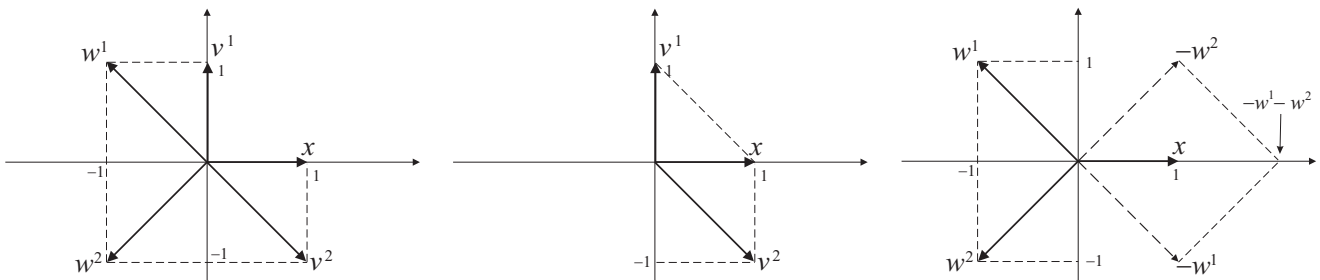
$${}_W B_V = {}_W B_U \cdot {}_U B_V = ({}_U B_W)^{-1} \cdot {}_U B_V,$$

dove la matrice ${}_U B_V$ è la matrice del cambio di base dalla base \mathcal{V} alla base fondamentale e ${}_U B_W$ è la matrice del cambio di base dalla base \mathcal{W} alla base fondamentale. Ricordo anche che queste ultime due si ottengono semplicemente scrivendo in colonna i vettori delle due basi, cioè

$${}_U B_V = (v^1 \ v^2 \ \dots \ v^n) \quad \text{e} \quad {}_U B_W = (w^1 \ w^2 \ \dots \ w^n).$$

Facciamo allora un esempio. Siano

$$v^1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad v^2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \text{e} \quad w^1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad w^2 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}.$$



⁵⁹ Ricordo che ${}_W B_V$ è la matrice che realizza il cambio di base dalla base \mathcal{V} a \mathcal{W} se, date le componenti y di un generico vettore rispetto alla base \mathcal{V} , con ${}_W B_V \cdot y$ trovo le componenti di tale vettore rispetto alla base \mathcal{W} .

Consideriamo il vettore indicato con x . Dalla regola del parallelogramma (figure al centro e a destra) si vede facilmente che

$$x = v^1 + v^2 = \frac{1}{2}(-w^1 - w^2) = -\frac{1}{2}w^1 - \frac{1}{2}w^2,$$

e quindi le componenti di x sono $(1, 1)$ rispetto alla base \mathcal{V} e $(-\frac{1}{2}, -\frac{1}{2})$ rispetto alla base \mathcal{W} .

Usando ora le matrici, con la formula per il cambio di base si ha

$${}_w B_v = \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix}^{-1} \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -1 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & -2 \\ -1 & 0 \end{pmatrix}$$

e, operando sulle componenti del vettore x , si ottiene infatti

$${}_w B_v \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & -2 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}.$$

È chiaro che si ha infine

$${}_v B_w = ({}_w B_v)^{-1} = \begin{pmatrix} 0 & -2 \\ -1 & -1 \end{pmatrix}$$

e questa realizza il cambio di base dalla base \mathcal{W} alla base \mathcal{V} .

4 Sistemi di equazioni lineari

Lo studente ha forse già incontrato i sistemi di equazioni lineari alla scuola secondaria. Con il termine *equazione lineare in n incognite* si intende l'equazione

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = b,$$

dove a_1, a_2, \dots, a_n e b sono numeri reali fissati e x_1, x_2, \dots, x_n sono le incognite.

Con *sistema di m equazioni lineari in n incognite* si intende la scrittura

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m. \end{cases}$$

Ciascuna riga del sistema è ovviamente una equazione lineare nelle incognite x_1, x_2, \dots, x_n .

Solitamente i numeri a_{ij} , con $1 \leq i \leq m$ e $1 \leq j \leq n$, si dicono i *coefficienti* del sistema, i b_i , con $1 \leq i \leq m$, sono i *termini noti* e le x_j , con $1 \leq j \leq n$, sono appunto le *incognite* del sistema. Talvolta, anziché dire più correttamente *sistema di equazioni lineari*, diremo più sinteticamente *sistema lineare* o semplicemente *sistema*.

È immediato osservare che il sistema si può scrivere, in forma matriciale, con $Ax = b$, avendo posto

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{e} \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

(A si dice la matrice del sistema, x il vettore delle incognite e b il vettore dei termini noti).

A e b si presumono fissati, cioè sono rispettivamente una matrice ed un vettore di numeri reali assegnati. Con il termine *soluzione* del sistema intendiamo ogni vettore x tale che $Ax = b$.

Risolvere il sistema significa trovare *tutte* le sue soluzioni. Un sistema si dice *possibile* se ha almeno una soluzione. Si dice *impossibile* se non ha soluzioni. Si è soliti anche chiamare *determinato* un sistema possibile che abbia un'unica soluzione, *indeterminato* un sistema possibile che abbia più di una soluzione. Un sistema si dice *quadrato* se la matrice A è quadrata.

▷ **Osservazione** Dato il sistema $Ax = b$, è chiaro che risolvere il sistema significa trovare la controimmagine di b attraverso la trasformazione lineare f rappresentata dalla matrice A ($f \in \mathcal{L}_{n,m}$, con $f(x) = Ax$)⁶⁰, cioè l'insieme dei vettori di \mathbb{R}^n che la f trasforma nel vettore b . È evidente che il sistema è possibile se e solo se $b \in \text{Im } f$ o, se si preferisce, se e solo se b si può scrivere, in \mathbb{R}^m , come combinazione lineare delle colonne di A .

Definizione Un sistema si dice *omogeneo* se $b = 0$.

▷ **Osservazione** Si osservi che risolvere il sistema omogeneo $Ax = 0$ equivale a determinare il nucleo della trasformazione lineare rappresentata da A . Un sistema omogeneo $Ax = 0$ è pertanto sempre possibile, avendo sicuramente almeno la soluzione nulla $x = 0$. Quindi la domanda interessante nel caso di sistema omogeneo è se esso ha altre soluzioni oltre a quella nulla, oppure no.

▷ **Osservazione** L'insieme delle soluzioni di un sistema omogeneo $Ax = 0$, essendo il nucleo della trasformazione rappresentata da A , è un sottospazio di \mathbb{R}^n .

Si noti che invece, se il sistema non è omogeneo, l'insieme delle sue soluzioni non è un sottospazio di \mathbb{R}^n (non contiene infatti il vettore nullo). Il teorema che segue illustra quale è in generale la struttura dell'insieme delle soluzioni di un sistema lineare.

Teorema 1 Se il sistema $Ax = b$ ha la soluzione \bar{x} , allora l'insieme delle sue soluzioni è

$$\mathcal{S} = \{\bar{x} + y : Ay = 0\} = \{\bar{x}\} + \text{Ker } f,$$

dove f è la trasformazione lineare rappresentata da A .

⁶⁰ Ricordo che in generale, data una funzione $g : A \rightarrow B$, se $C \subset B$, allora la controimmagine di C attraverso la funzione g è l'insieme

$$g^{-1}(C) = \{x \in A : g(x) \in C\}.$$

Quindi la controimmagine di b attraverso f è

$$f^{-1}(\{b\}) = \{x \in \mathbb{R}^n : f(x) = b\}.$$

▷ **Osservazione** Dato un sistema $Ax = b$, si usa chiamare *sistema omogeneo associato* il sistema $Ax = 0$. Il teorema dice quindi che (tutte) le soluzioni di un sistema si ottengono sommando ad una soluzione particolare del sistema stesso il nucleo di f , ossia sommando ad una soluzione particolare le soluzioni del sistema omogeneo associato.

Dimostrazione Dobbiamo dimostrare due cose: che ogni soluzione del sistema può essere scritta come somma di \bar{x} e di un elemento del nucleo di f e che, viceversa, ogni vettore di questo tipo è soluzione.

Per ipotesi $A\bar{x} = b$. Sia v una generica soluzione del sistema. Scriviamo $v = \bar{x} + v - \bar{x}$. Ora il vettore $v - \bar{x}$ è soluzione del sistema omogeneo associato, infatti $A(v - \bar{x}) = Av - A\bar{x} = b - b = 0$ e la prima parte è dimostrata.

Viceversa, dato un vettore del tipo $\bar{x} + y$, con $Ay = 0$, chiaramente si ha $A(\bar{x} + y) = A\bar{x} + Ay = b + 0 = b$ e quindi un tale vettore è soluzione. \square

▷ **Osservazione** L'insieme delle soluzioni di un sistema non omogeneo, pur non essendo un sottospazio di \mathbb{R}^n , tuttavia gli assomiglia molto, essendo sostanzialmente una *traslazione* di un sottospazio propriamente detto: viene detto sottospazio *affine*. Si può parlare di dimensione del sottospazio affine $\{\bar{x}\} + \text{Ker } f$, riferendosi alla dimensione di $\text{Ker } f$.

Il seguente è un risultato molto generale. Conviene prima dare questa

Definizione Dato un sistema $Ax = b$, la matrice, che si indica con $A|b$, ottenuta affiancando ad A il vettore b , quale ulteriore colonna, si chiama *matrice completa* del sistema. Solitamente A viene detta allora *matrice incompleta*.

Teorema 2 (di Rouché – Capelli) Un sistema $Ax = b$ ha almeno una soluzione se e solo se $rA = r(A|b)$.

Dimostrazione Ovviamente si ha in generale $r(A|b) \geq rA$.

Se il sistema è possibile significa, come già osservato, che è possibile scrivere b come combinazione lineare delle colonne di A . È chiaro a questo punto che il massimo numero di colonne indipendenti di $A|b$ non può essere maggiore del massimo numero di colonne indipendenti di A , dato che il vettore b dipende linearmente dalle altre colonne. Quindi $rA = r(A|b)$.

Viceversa, sia $rA = r(A|b)$ e supponiamo che tale rango sia k . Non è restrittivo supporre che i vettori a^1, a^2, \dots, a^k siano l.i. Dall'ipotesi sappiamo che i vettori a^1, a^2, \dots, a^k, b sono l.d. Significa che esiste una loro combinazione lineare $x_1 a^1 + x_2 a^2 + \dots + x_k a^k + yb = 0$ con coefficienti non tutti nulli. Se fosse $y = 0$, avremmo che anche $x_i = 0 \forall i : 1 \leq i \leq k$ (in quanto i vettori a^1, a^2, \dots, a^k sono indipendenti), contro il fatto che almeno un coefficiente è non nullo. Ma allora $y \neq 0$ e quindi possiamo scrivere b come combinazione lineare di a^1, a^2, \dots, a^k e di conseguenza come combinazione lineare di a^1, a^2, \dots, a^n . Se b si può scrivere come combinazione lineare delle colonne di A , il sistema è possibile. \square

▷ **Esercizio** Quando un sistema $Ax = b$ è possibile, con A matrice $m \times n$, la dimensione dello spazio delle sue soluzioni è $n - rA$.

Grazie al teorema di Rouché – Capelli è relativamente semplice sapere se un sistema ha soluzioni oppure no. Come trovare le soluzioni (quando esistono) lo vediamo tra breve.

Vediamo intanto un altro importante risultato.

Teorema 3 (di Cramer) Un sistema quadrato $Ax = b$ ha una ed una sola soluzione se e solo se $\det A \neq 0$.

Dimostrazione Se $\det A \neq 0$, allora A è invertibile ed esiste A^{-1} . Quindi, se $Ax = b$, allora $x = A^{-1}b$ e quindi vi è una ed una sola soluzione.

Viceversa, se $Ax = b$ ha una ed una sola soluzione, allora $\text{Ker } f = \{0\}$. Quindi f è invertibile e A è non singolare, cioè $\det A \neq 0$. \square

▷ **Osservazione** Relativamente al calcolo della soluzione di un sistema quadrato $Ax = b$, con $\det A \neq 0$, il modo più naturale per trovare la soluzione è sicuramente $x = A^{-1}b$. Esiste un metodo equivalente, che consente di trovare la soluzione componente per componente, evitando così il calcolo della matrice inversa. Si tratta della cosiddetta *regola di Cramer*: dato il sistema $Ax = b$, con A quadrata di ordine n non singolare, l'unica soluzione del sistema è il vettore $x \in \mathbb{R}^n$ la cui i -esima componente è

$$x_i = \frac{\det A_i}{\det A},$$

dove A_i è la matrice che si ottiene da A , sostituendo alla i -esima colonna il vettore b .

La regola di Cramer trova giustificazione osservando che

$$x = A^{-1}b = \frac{1}{\det A} A^* b = \frac{1}{\det A} (b_1 \bar{a}^1 + b_2 \bar{a}^2 + \dots + b_n \bar{a}^n),$$

avendo indicato con \bar{a}^j le colonne della matrice aggiunta A^* . Pertanto la i -esima componente della soluzione x è

$$x_i = \frac{1}{\det A} (b_1 \bar{a}_i^1 + b_2 \bar{a}_i^2 + \dots + b_n \bar{a}_i^n).$$

Ma

$$\begin{aligned}\bar{a}_i^j &= \text{elemento di posto } (i, j) \text{ di } A^* \\ &= \text{compl. alg. dell'elemento di posto } (j, i) \text{ di } A \\ &= A_{ji}.\end{aligned}$$

Quindi

$$\begin{aligned}x_i &= \frac{1}{\det A}(b_1 A_{1i} + b_2 A_{2i} + \dots + b_n A_{ni}) \\ &= \frac{1}{\det A} \det A_i,\end{aligned}$$

dove appunto A_i è la matrice che ottengo da A sostituendo con b la i -esima colonna.

▷ **Esempio** Risolvere il sistema

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

Dopo aver osservato che $\det A = 3$, utilizzando la regola di Cramer si ottiene

$$\begin{aligned}x_1 &= \frac{1}{3} \det \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \frac{2}{3}, \\ x_2 &= \frac{1}{3} \det \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} = -\frac{1}{3}.\end{aligned}$$

L'unica soluzione è quindi il vettore $(\frac{2}{3}, -\frac{1}{3})$.

▷ **Esempio** Risolvere il sistema

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

Risulta $\det A = 3$ e pertanto anche in questo caso il sistema ha una sola soluzione. Applicando la regola di Cramer si ottiene

$$\begin{aligned}x_1 &= \frac{1}{3} \det \begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & 1 \\ 2 & -1 & 0 \end{pmatrix} = 1; \\ x_2 &= \frac{1}{3} \det \begin{pmatrix} 1 & 0 & -1 \\ 1 & 1 & 1 \\ 1 & 2 & 0 \end{pmatrix} = -1; \\ x_3 &= \frac{1}{3} \det \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 2 \end{pmatrix} = 0.\end{aligned}$$

L'unica soluzione è quindi il vettore $(1, -1, 0)$.

▷ **Osservazione** Qualche ulteriore considerazione sui sistemi omogenei. Come già osservato in precedenza, essi hanno sempre almeno la soluzione nulla. Sull'esistenza di altre soluzioni, distinguiamo i due casi.

- Se il sistema $Ax = 0$ è quadrato, con A matrice $n \times n$, in base al teorema di Cramer possiamo dire che esso ha soluzioni non nulle se e solo se A è singolare, cioè $\det A = 0$. Condizioni equivalenti per l'esistenza di soluzioni non nulle sono: $rA < n$ oppure $\text{Ker } f \neq \{0\}$, dove come sempre f è la trasformazione lineare rappresentata da A .
- Se il sistema $Ax = 0$ non è quadrato, con A matrice $m \times n$ e $m \neq n$, il sistema ha soluzioni non nulle se e solo se $\text{Ker } f \neq \{0\}$, e quindi se e solo se $rA < n$. A tale proposito si osservi che la condizione è sicuramente verificata se $m > n$, cioè se il sistema ha più equazioni che incognite.

Vediamo ora come si possano calcolare le soluzioni di un sistema in tutti i casi diversi da quello di sistema quadrato con $\det A \neq 0$. Esiste un metodo generale che porta ad esprimere le soluzioni in funzione di un certo numero di parametri arbitrari. Vediamo questo metodo e poi lo applicheremo ad un paio di esempi conclusivi.

Dato il sistema $Ax = b$, supponiamo di aver trovato che $rA = r(A|b) = r$ e indichiamo con \overline{A} una sottomatrice di A , quadrata, non singolare di ordine r (potrebbe ovviamente non essere unica tale sottomatrice). Riscriviamo il sistema eliminando le eventuali equazioni corrispondenti a righe di A che non figurano in \overline{A} e “portando a secondo membro” le eventuali incognite relative a colonne di A che non figurano in \overline{A} .⁶¹

Si può dimostrare che in questo modo otteniamo un sistema equivalente a quello dato (cioè con lo stesso insieme di soluzioni).⁶² In tale sistema le r incognite relative alle colonne di A che figurano in \overline{A} vengono espresse in funzione delle altre $n - r$, che a questo punto diventano parametri arbitrari. Infatti è chiaro che, fissati in modo arbitrario i valori di questi $n - r$ parametri, e cioè per ogni scelta di questi, il sistema ha una ed una sola soluzione poiché è un sistema quadrato e il determinante della matrice di questo sistema (cioè $\det \overline{A}$) è non nullo.

▷ **Osservazione** Il fatto che tutte le soluzioni si possano esprimere al variare di $n - r$ parametri arbitrari è una sorta di modo “operativo” di affermare che la dimensione dello spazio delle soluzioni è $n - r$.

È evidente che, nel caso sia $r = n$, avremo necessariamente $m > n$ (altrimenti il sistema è quadrato con matrice non singolare e si ricade nel teorema di Cramer) e il procedimento descritto sopra consiste nella sola eliminazione delle equazioni “superflue”: una volta eliminate queste equazioni, il sistema è quadrato ed ha una sola soluzione per il teorema di Cramer.

Per il calcolo esplicito delle soluzioni basta risolvere il sistema ottenuto con la regola di Cramer, considerando parametri, come già detto, le incognite che figurano a destra. Concludiamo allora con un certo numero di esempi che illustrano i vari casi possibili.

▷ **Esempio** Risolvere il sistema

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

Poniamo

$$A|b = \left(\begin{array}{ccc|c} 1 & -1 & 1 & 2 \\ -1 & 1 & 1 & 1 \end{array} \right). \text{ Risulta } rA = r(A|b) = 2.$$

Il sistema è quindi possibile e lo spazio delle sue soluzioni ha dimensione $3 - 2 = 1$.

Quale sottomatrice di ordine massimo non singolare possiamo prendere quella formata dalla 1ª e dalla 3ª colonna di A , e cioè $\overline{A} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. Si noti che invece, ad esempio, la sottomatrice formata dalle prime due colonne è singolare.

Riscriviamo allora il sistema dato nel sistema equivalente

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

Con la regola di Cramer si ottiene

$$\begin{aligned} x &= \frac{1}{2} \det \begin{pmatrix} 2+y & 1 \\ 1-y & 1 \end{pmatrix} = \frac{1}{2}(2+y-1+y) = \frac{1}{2} + y, \\ z &= \frac{1}{2} \det \begin{pmatrix} 1 & 2+y \\ 1 & 1-y \end{pmatrix} = \frac{1}{2}(1-y+2+y) = \frac{3}{2}. \end{aligned}$$

Pertanto le soluzioni si possono scrivere come i vettori

$$\left(\frac{1}{2} + y, y, \frac{3}{2} \right),$$

dove y è un numero reale arbitrario. Si osservi che le soluzioni trovate si possono anche esprimere con $(\frac{1}{2}, 0, \frac{3}{2}) + y(1, 1, 0)$. Si tratta, al variare di $y \in \mathbb{R}$, del sottospazio generato dal vettore $(1, 1, 0)$, traslato del vettore $(\frac{1}{2}, 0, \frac{3}{2})$. Il vettore $(1, 1, 0)$ è chiaramente una base dello spazio delle soluzioni del sistema omogeneo associato al sistema dato.

▷ **Osservazione** Quello seguito non era l'unico modo di procedere. Si poteva anche porre

$$\overline{A} = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \quad (2^{\text{a}} \text{ e } 3^{\text{a}} \text{ colonna di } A),$$

⁶¹ Si ottiene così un sistema di r equazioni, con r incognite “a sinistra” e $n - r$ incognite “a destra”.

⁶² Il portare a destra alcune incognite è chiaro che non fa cambiare le soluzioni. L'eliminazione delle equazioni è un'azione più pericolosa ovviamente, ma in questo caso non muta l'insieme delle soluzioni poiché si tratta di equazioni dipendenti dalle altre.

riscrivere il sistema come

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

trovare, con la regola di Cramer,

$$\begin{aligned} y &= -\frac{1}{2} \det \begin{pmatrix} 2-x & 1 \\ 1+x & 1 \end{pmatrix} = -\frac{1}{2}(2-x-1-x) = -\frac{1}{2} + x, \\ z &= -\frac{1}{2} \det \begin{pmatrix} -1 & 2-x \\ 1 & 1+x \end{pmatrix} = -\frac{1}{2}(-1-x-2+x) = \frac{3}{2}, \end{aligned}$$

ed esprimere quindi le soluzioni con i vettori

$$(x, -\frac{1}{2} + x, \frac{3}{2}) = (0, -\frac{1}{2}, \frac{3}{2}) + x(1, 1, 0), \quad \text{con } x \in \mathbb{R}.$$

È chiaro che questo è un modo solo formalmente diverso di scrivere le soluzioni trovate prima: al variare dei parametri in tutto l'insieme \mathbb{R} si ottiene lo stesso insieme di vettori in \mathbb{R}^3 .

Si osservi che non era invece possibile esprimere le soluzioni in funzione di z , in quanto la sottomatrice di A formata dalla 1^a e 2^a colonna è singolare.

▷ **Esempio** Risolvere il sistema

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

Poniamo

$$A|b = \left(\begin{array}{cccc|c} 1 & -1 & 0 & 1 & 0 \\ -1 & 1 & 1 & 0 & 1 \end{array} \right). \quad \text{Risulta } rA = r(A|b) = 2.$$

Il sistema è possibile e lo spazio delle soluzioni ha dimensione $4 - 2 = 2$.

Quale sottomatrice di ordine massimo non singolare possiamo prendere quella formata dalla 1^a e dalla 3^a colonna di A , e cioè $\overline{A} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$.

Riscriviamo allora il sistema dato nel sistema equivalente

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

Con la regola di Cramer (ma è sicuramente più semplice in questo caso sostituire ad x nella seconda equazione $y - t$) si ottiene

$$x = y - t, \quad z = 1 - t.$$

Pertanto le soluzioni si possono scrivere come i vettori

$$(y - t, y, 1 - t, t) = (0, 0, 1, 0) + y(1, 1, 0, 0) + t(-1, 0, -1, 1),$$

dove y e z sono numeri reali arbitrari. I vettori $(1, 1, 0, 0)$ e $(-1, 0, -1, 1)$, essendo l.i., sono una base dello spazio delle soluzioni del sistema omogeneo associato.

▷ **Osservazione** Anche in questo caso non era l'unico modo di procedere. Si poteva anche porre

$$\overline{A} = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix} \quad (1^{\text{a}} \text{ e } 4^{\text{a}} \text{ colonna di } A),$$

riscrivere il sistema come

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

esprimere le soluzioni come i vettori

$$(-1 + y + z, y, z, 1 - z), \quad \text{con } y, z \in \mathbb{R}.$$

Non era invece possibile esprimere le soluzioni in funzione di x e y , in quanto la sottomatrice di A formata dalla 1^a e 2^a colonna è singolare.

▷ **Esempio** Risolvere il sistema

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

Poniamo

$$A|b = \left(\begin{array}{cccc|c} 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 \end{array} \right). \text{ Risulta } rA = r(A|b) = 2.$$

Il sistema è possibile e lo spazio delle soluzioni ha dimensione $4 - 2 = 2$.

Quale sottomatrice di ordine massimo non singolare possiamo prendere quella formata dalla 2^a e dalla 3^a colonna di A , e cioè $\bar{A} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$.

Riscriviamo allora il sistema dato nel sistema equivalente

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

Con la regola di Cramer si ottiene

$$y = \frac{1}{2} \det \begin{pmatrix} 1 - x + t & -1 \\ 1 - x + t & 1 \end{pmatrix} = 1 - x + t \quad \text{e} \quad z = \frac{1}{2} \det \begin{pmatrix} 1 & 1 - x + t \\ 1 & 1 - x + t \end{pmatrix} = 0.$$

Pertanto le soluzioni si possono scrivere come i vettori

$$(x, 1 - x + t, 0, t),$$

dove x e t sono numeri reali arbitrari.

▷ **Esempio** Risolvere il sistema

$$\begin{cases} x - 2y = 1 \\ -x + 3y = 1 \\ x - y = 3. \end{cases}$$

Poniamo

$$A|b = \left(\begin{array}{cc|c} 1 & -2 & 1 \\ -1 & 3 & 1 \\ 1 & -1 & 3 \end{array} \right). \text{ Risulta } rA = r(A|b) = 2.^{63}$$

Il sistema è possibile e lo spazio delle soluzioni ha dimensione $3 - 2 = 1$, cioè la soluzione è unica.

Quale sottomatrice di ordine massimo non singolare possiamo ad esempio prendere quella formata dalle prime due righe e dalle prime due colonne di A , e cioè $\bar{A} = \begin{pmatrix} 1 & -2 \\ -1 & 3 \end{pmatrix}$.

Possiamo quindi eliminare la terza equazione, perché dipendente dalle altre due. Il sistema si riduce allora al sistema equivalente

$$\begin{cases} x - 2y = 1 \\ -x + 3y = 1, \end{cases}$$

che è un sistema quadrato con determinante della matrice diverso da zero. L'unica soluzione è il vettore di componenti

$$x = \det \begin{pmatrix} 1 & -2 \\ 1 & 3 \end{pmatrix} = 5 \quad \text{e} \quad y = \det \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = 2.$$

Pertanto la soluzione è $(5, 2)$.

▷ **Esempio** Risolvere il sistema

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

Poniamo

$$A|b = \left(\begin{array}{ccc|c} 1 & -2 & 1 & 1 \\ -1 & 3 & 1 & 2 \\ 0 & 1 & 2 & 3 \end{array} \right).$$

⁶³ Si noti che l'annullarsi del determinante della matrice $A|b$ è in questo caso condizione necessaria (ma non sufficiente) affinché esista almeno una soluzione.

Il sistema è quadrato, ma il determinante di A è nullo. Il rango di A è 2 e anche il rango di $A|b$ è 2. Quindi il sistema è possibile e lo spazio delle soluzioni ha dimensione $3 - 2 = 1$.

Quale sottomatrice di ordine massimo non singolare possiamo ad esempio prendere quella formata dalle prime due righe e dalle prime due colonne di A , e cioè $\overline{A} = \begin{pmatrix} 1 & -2 \\ -1 & 3 \end{pmatrix}$.

Possiamo eliminare la terza equazione, perché dipendente dalle altre due. Il sistema si riduce allora al sistema equivalente

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

Si può far diventare parametro la z e, con il metodo già visto prima, si trovano le soluzioni

$$(7 - 5z, 3 - 2z, z), \quad \text{con } z \in \mathbb{R}.$$

▷ **Esempio** Risolvere il sistema

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

Poniamo

$$A|b = \left(\begin{array}{ccc|c} 1 & -1 & 1 & 1 \\ 2 & 0 & -1 & 0 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 2 & 1 \end{array} \right).$$

La matrice completa $A|b$ è quadrata e risulta $\det A|b = 0$ (la prima e la terza riga sono opposte). Si noti, come già osservato in precedenza su di un caso analogo, che l'annullarsi del determinante di $A|b$ è una condizione necessaria per l'esistenza di soluzioni: infatti se fosse $\det(A|b) \neq 0$, avremmo che $r(A|b) = 4$, mentre sicuramente $rA < 4$.⁶⁴

Risulta (lo studente faccia la verifica) $rA = r(A|b) = 2$. Per il teorema di Rouché – Capelli lo spazio delle soluzioni ha dimensione 1. Possiamo scegliere quale sottomatrice di ordine massimo non singolare quella formata da 1^a e 2^a riga, 2^a e 3^a colonna, e cioè $\overline{A} = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$. Il sistema equivalente è

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

Con la regola di Cramer (ma qui è più veloce una semplice sostituzione) si trova

$$\begin{cases} x_3 = 2x_1 \\ x_2 = 3x_1 - 1. \end{cases}$$

Le soluzioni sono i vettori $(x_1, 3x_1 - 1, 2x_1)$, al variare di x_1 in \mathbb{R} .

▷ **Osservazione** Anche in questo caso si potevano considerare altre sottomatrici non singolari, ad esempio

$$\overline{A} = \begin{pmatrix} 2 & 0 \\ -1 & 1 \end{pmatrix} \quad (2^{\text{a}} \text{ e } 3^{\text{a}} \text{ riga, } 1^{\text{a}} \text{ e } 2^{\text{a}} \text{ colonna}).$$

Avremmo ottenuto il sistema equivalente

$$\begin{cases} 2x_1 = x_3 \\ -x_1 + x_2 = x_3 - 1 \end{cases} \quad \text{e quindi} \quad \begin{cases} x_1 = \frac{1}{2}x_3 \\ x_2 = \frac{3}{2}x_3 - 1, \end{cases}$$

da cui le soluzioni

$$\left(\frac{1}{2}x_3, \frac{3}{2}x_3 - 1, x_3\right) \quad \text{con } x_3 \in \mathbb{R} \text{ arbitrario.}$$

È ovviamente lo stesso insieme di prima: basta pensare che, in entrambi i casi, si tratta dei vettori in cui la terza componente è “doppia della prima” e la seconda è “3 volte la prima meno 1”.

⁶⁴ Non è ovviamente una condizione sufficiente per l'esistenza di soluzioni, dato che, con $\det A = 0$, si potrebbe comunque avere ad esempio $r(A|b) = 3$ e $rA = 2$.

5 Matrici simili, autovalori e diagonalizzabilità

5.1 Matrici simili

Torniamo su di un aspetto precedentemente solo sfiorato (si veda una delle osservazioni che seguono la definizione della matrice di rappresentazione di una trasformazione lineare).

La matrice di rappresentazione di una $f \in \mathcal{L}_{n,m}$ è stata introdotta come la matrice le cui colonne sono le immagini dei vettori fondamentali di \mathbb{R}^n attraverso la trasformazione f . Dalla scrittura

$$f(x) = f(x_1 u^1 + x_2 u^2 + \dots + x_n u^n) = x_1 f(u^1) + x_2 f(u^2) + \dots + x_n f(u^n)$$

si ottiene appunto che, detta A la matrice $(f(u^1) \ f(u^2) \ \dots \ f(u^n))$, possiamo scrivere $f(x) = Ax$.

La matrice A rappresenta quindi la trasformazione f , ma questa rappresentazione è rispetto alle basi fondamentali nei due spazi \mathbb{R}^n ed \mathbb{R}^m , dato che x è scritto nella base fondamentale di \mathbb{R}^n (e questo è evidente) e $f(x)$ è scritto nella base fondamentale di \mathbb{R}^m (e questo diciamo che è assunto implicitamente quando, non specificando quale è la base in cui esprimo $f(u^1), f(u^2), \dots, f(u^n)$, assumo che sia la base fondamentale). Si pone ora il seguente problema: come cambia la matrice A se vogliamo rappresentare i vettori in basi diverse da quelle fondamentali?

Anche se il problema può essere affrontato e risolto nel caso generale di una $f \in \mathcal{L}_{n,m}$ ⁶⁵, consideriamo il caso particolarmente rilevante di una trasformazione lineare dello spazio \mathbb{R}^n in sé, quindi consideriamo una $f \in \mathcal{L}_{n,n}$.

Se v è la scrittura di un vettore x in una data base \mathcal{V} , allora ${}_{\mathcal{U}}B_{\mathcal{V}}v$ è la scrittura di x nella base fondamentale.⁶⁶ Quindi $A \cdot {}_{\mathcal{U}}B_{\mathcal{V}}v$ è l'immagine di x attraverso la trasformazione f , scritta in base fondamentale. Infine ${}_{\mathcal{V}}B_{\mathcal{U}} \cdot A \cdot {}_{\mathcal{U}}B_{\mathcal{V}}v$ è l'immagine di x attraverso la trasformazione f , scritta in base \mathcal{V} .

Pertanto la matrice di rappresentazione di f nella base \mathcal{V} è la matrice

$${}_{\mathcal{V}}B_{\mathcal{U}} \cdot A \cdot {}_{\mathcal{U}}B_{\mathcal{V}} = ({}_{\mathcal{U}}B_{\mathcal{V}})^{-1} A \cdot {}_{\mathcal{U}}B_{\mathcal{V}}.$$

Definizione Date due matrici (quadrate) A e B , diciamo che B è *simile* ad A se esiste una matrice invertibile S tale che si abbia

$$B = S^{-1}AS.$$

▷ **Osservazione** La relazione di similitudine è una relazione di *equivalenza*. Infatti A è simile ad A (proprietà *riflessiva*); se A è simile a B , anche B è simile ad A (proprietà *simmetrica*); se A è simile a B e B è simile a C , allora A è simile a C (proprietà *transitiva*). Si provi a dimostrare per esercizio le tre proprietà.

Per quanto appena detto nell'osservazione, d'ora in poi, anziché dire che A è simile a B (o che B è simile ad A), diremo che A e B sono simili.

Dimostriamo ora un risultato generale.

Teorema 1 Due matrici sono simili se e solo se rappresentano la stessa trasformazione lineare.

Dimostrazione Quanto visto nella premessa alla definizione di similitudine prova sostanzialmente che, se due matrici rappresentano la stessa trasformazione, allora sono simili. Vediamo allora il viceversa e dimostriamo cioè che, se due matrici A e B sono simili, allora rappresentano la stessa trasformazione. Supponiamo che A e B siano simili, cioè che $B = S^{-1}AS$ per qualche matrice non singolare S .

Se A rappresenta una trasformazione lineare f rispetto alla base fondamentale \mathcal{U} , allora il risultato è immediato: infatti, detta S la base formata dalle colonne della matrice S ⁶⁷, per quanto visto in precedenza abbiamo

$$S = {}_{\mathcal{U}}B_{\mathcal{S}},$$

cioè S è la matrice del cambio dalla base \mathcal{S} alla base fondamentale \mathcal{U} . Quindi

$$B = S^{-1}AS = ({}_{\mathcal{U}}B_{\mathcal{S}})^{-1} A \cdot {}_{\mathcal{U}}B_{\mathcal{S}} = {}_{\mathcal{S}}B_{\mathcal{U}} \cdot A \cdot {}_{\mathcal{U}}B_{\mathcal{S}}$$

è la matrice che rappresenta f nella base \mathcal{S} .

Ora supponiamo che invece A rappresenti f in una qualunque base \mathcal{V} . Premettiamo un lemma.⁶⁸

Se S è una matrice non singolare e se $\mathcal{V} = \{v^1, v^2, \dots, v^n\}$ è una base di \mathbb{R}^n , allora esiste un'altra base $\mathcal{W} = \{w^1, w^2, \dots, w^n\}$ tale che $S = {}_{\mathcal{V}}B_{\mathcal{W}}$, cioè tale per cui S è la matrice che realizza il cambio dalla base \mathcal{W} alla base \mathcal{V} .

⁶⁵Più avanti c'è un esercizio a tale riguardo.

⁶⁶Le notazioni sono quelle già utilizzate in precedenza. Ricordo che ${}_{\mathcal{U}}B_{\mathcal{V}}$ è la matrice che realizza il cambio dalla base \mathcal{V} alla base fondamentale, che viene qui indicata con \mathcal{U} . Ricordo anche che per ottenere ${}_{\mathcal{U}}B_{\mathcal{V}}$ basta disporre in colonna i vettori della base \mathcal{V} .

⁶⁷Si ricordi che S è non singolare, e che quindi le sue colonne sono l.i.

⁶⁸Lemma è un risultato intermedio che serve a dimostrare un risultato successivo, che è generalmente il vero risultato a cui si vuole arrivare.

Per dimostrare il lemma, sia \mathcal{W} la base formata dalle colonne della matrice

$${}_{\mathcal{U}}B_{\mathcal{V}} \cdot S.^{69}$$

Avremo quindi ${}_{\mathcal{U}}B_{\mathcal{V}} \cdot S = {}_{\mathcal{U}}B_{\mathcal{W}}$. Ma allora

$$S = ({}_{\mathcal{U}}B_{\mathcal{V}})^{-1} {}_{\mathcal{U}}B_{\mathcal{W}} = {}_{\mathcal{V}}B_{\mathcal{U}} \cdot {}_{\mathcal{U}}B_{\mathcal{W}} = {}_{\mathcal{V}}B_{\mathcal{W}},$$

e dunque S è la matrice del cambio dalla base \mathcal{W} alla base \mathcal{V} , come volevamo dimostrare.

Tornando allora alla questione originaria, se A rappresenta la trasformazione f nella base \mathcal{V} , per il lemma precedente si ha allora

$$B = S^{-1}AS = {}_{\mathcal{W}}B_{\mathcal{V}} \cdot A \cdot {}_{\mathcal{V}}B_{\mathcal{W}},$$

e quindi B rappresenta f nella base \mathcal{W} . □

▷ **Esercizio** Si consideri una trasformazione $f \in \mathcal{L}_{n,m}$ e si supponga che la matrice A rappresenti f rispetto alle basi canoniche in \mathbb{R}^n e in \mathbb{R}^m . Si supponga ora che $\mathcal{V} = \{v^1, v^2, \dots, v^n\}$ sia una nuova base in \mathbb{R}^n e che $\mathcal{W} = \{w^1, w^2, \dots, w^m\}$ sia una nuova base in \mathbb{R}^m . Si dica come si può ottenere la matrice di rappresentazione di f rispetto alle basi \mathcal{V} e \mathcal{W} .

Prima di studiare altre proprietà interessanti delle matrici simili, vediamo alcune definizioni fondamentali.

5.2 Autovalori e autovettori

Consideriamo una trasformazione lineare $f \in \mathcal{L}_{n,n}$, rappresentata da una matrice A .

Definizione Uno scalare λ è un *autovalore* di f (o di A) se esiste un vettore *non nullo* v tale che

$$f(v) = \lambda v, \quad \text{ossia} \quad Av = \lambda v.$$

In tal caso v si dice *autovettore di f (o di A) associato a λ* .

▷ **Osservazione** Il senso della definizione è chiaro: se λ è un autovalore di f e v è un autovettore, significa che la f trasforma il vettore v in un vettore proporzionale a v , e il coefficiente di proporzionalità è λ .

▷ **Osservazione** Nella definizione si chiede espressamente che l'autovettore v non sia il vettore nullo, dato che con $v = 0$ l'equazione sarebbe banalmente soddisfatta.

▷ **Osservazione** Può invece essere $\lambda = 0$, e 0 è autovalore di f se e solo se ci sono vettori non nulli che vengono trasformati nel vettore nullo, e cioè se e solo se il $\text{Ker } f$ non è banale.

▷ **Osservazione** Sia λ un autovalore di f e sia \mathcal{S}_{λ} l'insieme di tutti gli autovettori di f associati a λ . Certamente \mathcal{S}_{λ} non contiene il vettore nullo, per definizione di autovettore, e quindi \mathcal{S}_{λ} non è un sottospazio di \mathbb{R}^n . Però possiamo dimostrare facilmente che $\mathcal{S}_{\lambda} \cup \{0\}$ lo è. Infatti, se $x, y \in \mathcal{S}_{\lambda} \cup \{0\}$, allora

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) = \alpha \cdot \lambda x + \beta \cdot \lambda y = \lambda(\alpha x + \beta y)$$

(è chiaro che, se ad esempio fosse $x = 0$, allora $\alpha x = 0$ e $f(x) = 0$).

Nel seguito chiameremo *autospatio associato all'autovalore λ* il sottospazio $\mathcal{S}_{\lambda} \cup \{0\}$.

Definizione Chiamiamo *molteplicità geometrica* dell'autovalore λ , e la indichiamo con $m_g(\lambda)$, la dimensione del suo autospatio associato.

Diremo che λ è un autovalore *semplice* se la sua molteplicità è 1, quindi se $\dim(\mathcal{S}_{\lambda} \cup \{0\}) = 1$; diremo che λ è un autovalore *doppio* se la sua molteplicità è 2, e così via.

Definizione L'insieme degli autovalori di f (di A) è detto lo *spettro* di f (di A).

▷ **Osservazione** Lo spettro di f (di A) è l'insieme degli scalari λ per cui $f - \lambda$ ($A - \lambda$) *non* è invertibile. ⁷⁰

Dimostriamo dunque che

$$\text{spettro di } f = \{\lambda : f - \lambda \text{ non è invertibile}\}.$$

Se λ appartiene allo spettro di f , cioè se λ è autovalore di f , allora $f(v) = \lambda v$ per qualche $v \neq 0$; quindi $(f - \lambda)(v) = 0$, e dunque $\text{Ker}(f - \lambda) \neq \{0\}$. Pertanto $f - \lambda$ non è invertibile.

⁶⁹Le colonne di tale matrice sono certamente l.i., dato che la matrice ${}_{\mathcal{U}}B_{\mathcal{V}} \cdot S$ è non singolare, essendo prodotto di matrici non singolari.

⁷⁰Scrivendo $f - \lambda$ (si noti che f è una trasformazione e λ è un numero reale) intendiamo la trasformazione $f - \lambda \cdot 1$, dove 1, come già indicato in precedenza, è la trasformazione identità (in \mathbb{R}^n). Analogamente, $A - \lambda$ è la matrice $A - \lambda \cdot 1$, dove anche qui 1 è la matrice identità, quella che rappresenta la trasformazione identità.

Viceversa, se $f - \lambda$ non è invertibile, allora il suo nucleo non è banale ed esiste $v \neq 0$ tale che $(f - \lambda)(v) = 0$. Ma quindi $f(v) = \lambda v$, e quindi λ è autovalore di f .

Si può anche dire che la molteplicità geometrica $m_g(\lambda)$ di un autovalore λ è uguale alla nullità di $f - \lambda$ ⁷¹, o anche che è uguale a $n - r(f - \lambda)$.⁷²

▷ **Osservazione** Ricordando che condizione necessaria e sufficiente affinché una matrice sia invertibile è che il suo determinante non si annulli, possiamo anche dire che λ è un autovalore della matrice A se e solo se

$$\det(A - \lambda) = 0.$$

Definizione Data una matrice A (quadrata di ordine n), chiamiamo *polinomio caratteristico di A* il polinomio

$$p(\lambda) = \det(A - \lambda) = \det \begin{pmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{pmatrix}.$$

Chiamiamo inoltre *equazione caratteristica di A* l'equazione $p(\lambda) = 0$.

▷ **Osservazione** Quindi possiamo anche dire che λ è un autovalore della matrice A se e solo se è una soluzione dell'equazione caratteristica di A . Questo risultato ha un'importanza fondamentale. Ragionando da un punto di vista geometrico sarebbe difficile trovare gli autovalori di A . Il risultato permette di ricondurre la ricerca degli autovalori ad un problema algebrico di risoluzione di un'equazione. Si noti che comunque in generale la risoluzione di un'equazione algebrica non è un problema "facile".

Si può dimostrare che il polinomio caratteristico di A è un polinomio di grado n e che il coefficiente di λ^n è $(-1)^n$.

Definizione Chiamiamo *molteplicità algebrica* dell'autovalore λ , e la indichiamo con $m_a(\lambda)$, la molteplicità di λ quale soluzione dell'equazione caratteristica.⁷³

▷ **Osservazione** Per quanto detto, l'equazione caratteristica è un'equazione algebrica di grado n . Nel campo reale, cioè se i coefficienti dell'equazione sono reali e cerchiamo le soluzioni in \mathbb{R} , non è detto che un'equazione algebrica abbia soluzioni. Si pensi ad esempio all'equazione di secondo grado $\lambda^2 + 1 = 0$. È per questo motivo che una trattazione più generale di questi argomenti viene condotta in un campo numerico in cui questo problema non si presenta: il campo dei numeri complessi è preferibile a quello dei numeri reali a tale proposito, dato che un'equazione algebrica a coefficienti complessi ha sempre almeno una soluzione (eventualmente complessa). Si può anzi dimostrare che un'equazione algebrica di grado n a coefficienti complessi ha sempre n soluzioni, se queste vengono contate con le rispettive molteplicità.

Noi non vogliamo qui svolgere una trattazione generale. Restiamo nel campo reale e ci limitiamo quindi a considerare come sempre matrici ad elementi reali e quindi equazioni caratteristiche a coefficienti reali, che potranno avere oppure no soluzioni (reali).

Possiamo dire che il numero di soluzioni (reali) delle nostre equazioni caratteristiche di grado n è compreso tra 0 ed n , contando sempre le soluzioni con le rispettive molteplicità.

▷ **Esempio** Consideriamo la matrice 1_n , cioè la matrice identità $n \times n$. Dato che $1_n - 1$ è la matrice nulla, allora $\lambda = 1$ è autovalore di 1_n di molteplicità geometrica n . L'equazione caratteristica di 1_n è chiaramente $(1 - \lambda)^n = 0$, che ha la soluzione $\lambda = 1$, di molteplicità n . Quindi la molteplicità algebrica dell'autovalore è n .

▷ **Esempio**

$$A = \begin{pmatrix} 1 & 3 \\ 1 & -1 \end{pmatrix} \quad ; \quad A - \lambda = \begin{pmatrix} 1 - \lambda & 3 \\ 1 & -1 - \lambda \end{pmatrix} \quad ; \quad \det(A - \lambda) = \lambda^2 - 4.$$

L'equazione caratteristica è $\lambda^2 - 4 = 0$. Gli autovalori sono $\lambda_1 = +2, \lambda_2 = -2$, entrambi di molteplicità algebrica 1, cioè

$$m_a(2) = m_a(-2) = 1.$$

Gli autovettori associati a $\lambda_1 = 2$ sono le soluzioni del sistema

$$\begin{cases} x_1 + 3x_2 = 2x_1 \\ x_1 - x_2 = 2x_2 \end{cases} \quad \text{ossia} \quad \begin{cases} -x_1 + 3x_2 = 0 \\ x_1 - 3x_2 = 0, \end{cases}$$

⁷¹Ricordo che la nullità di una trasformazione lineare è la dimensione del suo nucleo.

⁷²Dal teorema nullità + rango segue che la dimensione del nucleo di $f - \lambda$ è uguale ad n meno il rango di $f - \lambda$.

⁷³Ad esempio, l'equazione $(x-1)(x+2) = 0$ ha due soluzioni ($x = 1$ e $x = -2$), entrambe di molteplicità 1; l'equazione $(x^2-1)(x+2)^2 = 0$ ha tre soluzioni, di cui due ($x = 1$ e $x = -1$) di molteplicità 1 e una ($x = -2$) di molteplicità 2.

cioè lo spazio

$$\mathcal{S}_2 = \{(3\alpha, \alpha) : \alpha \in \mathbb{R}\},$$

che ha dimensione 1: quindi $m_g(2) = \dim \mathcal{S}_2 = 1$.

Gli autovettori associati a $\lambda_2 = -2$ sono le soluzioni del sistema

$$\begin{cases} x_1 + 3x_2 = -2x_1 \\ x_1 - x_2 = -2x_2 \end{cases} \quad \text{ossia} \quad \begin{cases} 3x_1 + 3x_2 = 0 \\ x_1 + x_2 = 0, \end{cases}$$

cioè lo spazio

$$\mathcal{S}_{-2} = \{(\alpha, -\alpha) : \alpha \in \mathbb{R}\},$$

che ha dimensione 1: quindi $m_g(-2) = \dim \mathcal{S}_{-2} = 1$.

▷ **Esempio** Consideriamo

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ -1 & 0 & 3 \end{pmatrix} \quad ; \quad A - \lambda = \begin{pmatrix} 1-\lambda & 0 & 1 \\ 0 & 2-\lambda & 0 \\ -1 & 0 & 3-\lambda \end{pmatrix} \quad ; \quad \det(A - \lambda) = (2 - \lambda)^3.$$

L'equazione caratteristica è $(2 - \lambda)^3 = 0$. C'è una sola soluzione $\lambda = 2$, con $m_a(2) = 3$. Gli autovettori associati a $\lambda = 2$ sono le soluzioni del sistema

$$Ax = 2x, \quad \text{cioè} \quad (A - 2)x = 0,$$

che si riduce all'unica equazione

$$-x_1 + x_3 = 0.$$

L'autospazio associato all'autovalore $\lambda = 2$ è quindi

$$\mathcal{S}_2 = \{(\alpha, \beta, \alpha) : \alpha, \beta \in \mathbb{R}\}.$$

In questo caso si ha pertanto $m_g(2) = 2$, che non coincide con la molteplicità algebrica trovata in precedenza.

▷ **Osservazione** L'ultimo esempio prova che molteplicità geometrica e molteplicità algebrica possono non essere uguali. In generale si può dimostrare che, per ogni autovalore λ , si ha

$$m_g(\lambda) \leq m_a(\lambda).$$

▷ **Esercizio** Gli autovalori di una matrice triangolare, in particolare di una matrice diagonale, sono gli elementi che si trovano sulla diagonale principale.

▷ **Esercizio** Si provi che $\lambda = 0$ è autovalore di A se e solo se A è singolare.

▷ **Esercizio** Si provi che, se A è invertibile, allora

$$\text{spettro di } A^{-1} = \left\{ \frac{1}{\lambda} : \lambda \in \text{spettro di } A \right\}.$$

▷ **Esercizio** Si provi che, se 0 è autovalore di A^2 , allora 0 è anche autovalore di A (il viceversa è banale).

▷ **Esercizio** Si provi che, se λ è autovalore di A , allora λ^2 è autovalore di A^2 . Si verifichi che il viceversa non è vero, e cioè che se λ è autovalore di A^2 , non necessariamente λ è il quadrato di un autovalore di A , sull'esempio

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

▷ **Esercizio** Si provi che, se $\lambda > 0$ è autovalore di A^2 , allora $\lambda = \alpha^2$, dove α è un autovalore di A .

▷ **Esercizio** Si provi che, se λ è autovalore di A , allora λ^k è autovalore di A^k , dove k è un qualunque numero naturale.

▷ **Osservazione** Un caso particolare in cui possiamo affermare che le molteplicità geometriche degli autovalori coincidono con le molteplicità algebriche è il seguente: se $\lambda_1, \lambda_2, \dots, \lambda_p$ sono autovalori *distinti* di $A_{n \times n}$, con rispettive molteplicità geometriche $m_g(\lambda_1), m_g(\lambda_2), \dots, m_g(\lambda_p)$, e si ha $m_g(\lambda_1) + m_g(\lambda_2) + \dots + m_g(\lambda_p) = n$, allora $m_a(\lambda_j) = m_g(\lambda_j)$ per ogni j .

▷ **Osservazione** Ricordiamo che, data un'equazione algebrica di grado n

$$\alpha_0 + \alpha_1 \lambda + \alpha_2 \lambda^2 + \dots + \alpha_n \lambda^n = 0,$$

il prodotto delle radici dell'equazione è uguale a $(-1)^n \frac{\alpha_0}{\alpha_n}$.

Ora supponiamo che $\lambda_1, \lambda_2, \dots, \lambda_p$ siano gli autovalori *distinti* di $A_{n \times n}$, e che m_1, m_2, \dots, m_p siano le rispettive molteplicità algebriche, e si abbia $m_1 + m_2 + \dots + m_p = n$. Se consideriamo l'equazione caratteristica di A , abbiamo che: le sue radici sono gli autovalori di A , $\alpha_n = (-1)^n$, e infine α_0 non è altro che il determinante di A ⁷⁴.

Da tutto questo si deduce allora che

$$\det A = \prod_{j=1}^p \lambda_j^{m_j}.$$

Più avanti ci occuperemo di *diagonalizzabilità* di una matrice, cioè del fatto che una matrice sia simile ad una matrice diagonale. Come vedremo non è detto che in generale una matrice sia simile ad una matrice diagonale, in altre parole alcune matrici sono diagonalizzabili e altre no.

Si può dimostrare che invece ogni matrice è simile ad una matrice triangolare, cioè che data una qualunque matrice A , esiste una matrice invertibile S tale che la matrice $S^{-1}AS$ è triangolare.

Vediamo ora alcune proprietà delle matrici simili.

Teorema 2 Se A e B sono simili, allora:

- i) A e B hanno lo stesso determinante.
- ii) A e B hanno lo stesso polinomio caratteristico.
- iii) A e B hanno gli stessi autovalori.

▷ **Osservazione** Si dice che il determinante, il polinomio caratteristico, gli autovalori sono *invarianti* della trasformazione. Mentre le matrici che la rappresentano possono variare, e abbiamo visto il perché, le entità indicate invece sono proprie della trasformazione e non della sua rappresentazione. Potremmo dire che tali entità sono più profondamente legate alla trasformazione e non cambiano se cambia solo il modo di rappresentare la trasformazione stessa.

Dimostrazione Per quanto riguarda i), se $B = S^{-1}AS$, allora

$$\det B = \det(S^{-1}AS) = \det S^{-1} \det A \det S = \det A.⁷⁵$$

Riguardo a ii), si ha

$$\begin{aligned} \det(B - \lambda) &= \det(S^{-1}AS - \lambda) \\ &= \det(S^{-1}AS - \lambda S^{-1}1S) \\ &= \det(S^{-1}(A - \lambda)S) \\ &= \det S^{-1} \det(A - \lambda) \det S \\ &= \det(A - \lambda). \end{aligned}$$

Riguardo infine a iii), si tratta di un'immediata conseguenza di ii), dato che gli autovalori sono le radici del polinomio caratteristico. □

▷ **Osservazione** Matrici simili hanno, come appena visto, gli stessi autovalori.

Per quanto riguarda gli autovettori, o se si preferisce gli autospazi, c'è un importante punto da osservare. Anche l'autospazio associato ad un dato autovalore è invariante, essendo un sottospazio che la funzione lineare trasforma in se stesso. Questo aspetto si coglie meglio se si pensa all'autovettore della trasformazione lineare, cioè al vettore tale che $f(v) = \lambda v$.

Non appena però tiriamo in ballo la matrice, ecco che si pone il problema della rappresentazione, dato che A è una delle possibili rappresentazioni di f .

Pertanto, se utilizziamo una matrice B simile ad A per rappresentare f , è ragionevole che gli autovettori, pur restando gli stessi, abbiano una diversa rappresentazione, quella relativa alla base in cui B rappresenta f .

Infatti, se $B = S^{-1}AS$, se λ è autovalore (di A e di B) e se v è autovettore di A associato a λ , si può dimostrare (farlo per esercizio) che $t = S^{-1}v$ è autovettore di B associato a λ . Il risultato come detto non deve sorprendere: se

⁷⁴Infatti, se

$$p(\lambda) = \det(A - \lambda) = \alpha_0 + \alpha_1 \lambda + \dots + (-1)^n \lambda^n,$$

allora $p(0) = \det A = \alpha_0$.

⁷⁵Si è fatto uso del teorema di Binet.

la matrice B rappresenta la stessa trasformazione in una base diversa e se la matrice S realizza il cambio di base, è chiaro che, mentre l'autovalore e l'autospazio associato non dipendono dalla scelta della base, la rappresentazione di quest'ultimo, cioè dei vettori che costituiscono l'autospazio, dipende invece dalla base e quindi le scritture di tali vettori risentiranno di un cambio di base.

Si noti anche che, se per fissare le idee diciamo che la matrice A rappresenta una f nella base \mathcal{V} e la matrice B rappresenta f nella base \mathcal{W} , allora la matrice S è appunto la ${}_V B_W$, cioè la matrice del cambio di base da \mathcal{W} a \mathcal{V} . Quindi per avere gli autovettori di B basterà scrivere gli autovettori di A in base \mathcal{W} , cioè moltiplicarli per ${}_W B_V$, che è S^{-1} .

Vale ora il seguente risultato:

Teorema 3 Autovettori associati ad autovalori distinti sono linearmente indipendenti.

Dimostrazione Supponiamo che $\lambda_1, \lambda_2, \dots, \lambda_k$ siano autovalori distinti di A e che x^1, x^2, \dots, x^k siano k autovettori associati. Vogliamo dimostrare che x^1, x^2, \dots, x^k sono l.i.

Supponiamo che essi siano l.d.: allora tra loro ce n'è almeno uno che è c.l. dei precedenti. Sia x^j il primo di questi. Quindi avremo che

$$x^j = \alpha_1 x^1 + \alpha_2 x^2 + \dots + \alpha_{j-1} x^{j-1}$$

e x^1, x^2, \dots, x^{j-1} sono l.i.

Moltiplicando a sinistra per A , si ottiene

$$Ax^j = \alpha_1 Ax^1 + \alpha_2 Ax^2 + \dots + \alpha_{j-1} Ax^{j-1}$$

e quindi

$$\lambda_j x^j = \alpha_1 \lambda_1 x^1 + \alpha_2 \lambda_2 x^2 + \dots + \alpha_{j-1} \lambda_{j-1} x^{j-1}.$$

Sostituendo ad x^j la sua espressione come c.l. degli altri vettori, si ottiene

$$\lambda_j (\alpha_1 x^1 + \alpha_2 x^2 + \dots + \alpha_{j-1} x^{j-1}) = \alpha_1 \lambda_1 x^1 + \alpha_2 \lambda_2 x^2 + \dots + \alpha_{j-1} \lambda_{j-1} x^{j-1}$$

cioè

$$\alpha_1 x^1 (\lambda_j - \lambda_1) + \alpha_2 x^2 (\lambda_j - \lambda_2) + \dots + \alpha_{j-1} x^{j-1} (\lambda_j - \lambda_{j-1}) = 0.$$

Ora, ricordando che i vettori x^1, x^2, \dots, x^{j-1} sono l.i. e che gli autovalori sono distinti, si deduce che deve essere

$$\alpha_1 = \alpha_2 = \dots = \alpha_{j-1} = 0.$$

Questo però significa $x^j = 0$, contro l'ipotesi che sia un autovettore. □

5.3 Diagonalizzabilità

Definizione Una matrice A è *diagonalizzabile* se è simile ad una matrice diagonale, cioè se esiste una matrice invertibile S tale che la matrice $S^{-1}AS$ è diagonale.

Definizione Chiamiamo *matrice modale* di A una qualunque matrice che abbia per colonne n autovettori di A .

▷ **Osservazione** Ovviamente la matrice modale non è unica. Inoltre può essere singolare.⁷⁶

Nel seguito indicheremo con $\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ la matrice diagonale che ha sulla diagonale principale gli scalari $\lambda_1, \lambda_2, \dots, \lambda_n$.

Teorema 4 Se x^1, x^2, \dots, x^n sono n autovettori associati rispettivamente agli autovalori $\lambda_1, \lambda_2, \dots, \lambda_n$, vale l'uguaglianza

$$AV = V \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n),$$

dove V è la matrice modale degli x^i .

Dimostrazione Si ha direttamente

$$\begin{aligned} AV &= A(x^1 \ x^2 \ \dots \ x^n) \\ &= (Ax^1 \ Ax^2 \ \dots \ Ax^n) \\ &= (\lambda_1 x^1 \ \lambda_2 x^2 \ \dots \ \lambda_n x^n) \\ &= (x^1 \ x^2 \ \dots \ x^n) \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) \\ &= V \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n). \end{aligned}$$

⁷⁶Banalmente, per avere una matrice modale singolare, basta ad esempio ripetere almeno due volte lo stesso autovettore, oppure scrivere un autovettore insieme ad un suo multiplo. Si noti che nella definizione di matrice modale non si richiede che gli autovettori siano distinti o che siano associati ad autovalori distinti. Quindi si possono ottenere matrici modali "molto banali".

□

▷ **Osservazione** Si noti che nel teorema non si fa l'ipotesi che gli autovalori siano distinti e non si fa nemmeno alcuna ipotesi sugli autovettori associati.

▷ **Osservazione** Si noti che nella tesi del teorema 4 siamo “molto vicini alla diagonalizzabilità” della matrice A . Mi spiego meglio: consideriamo la

$$AV = V \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n).$$

Se noi sapessimo che V è invertibile, moltiplicando a sinistra per V^{-1} otterremmo

$$V^{-1}AV = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n),$$

il che vorrebbe dire che A è diagonalizzabile. Quindi, se noi potessimo affermare che V è non singolare, avremmo che ogni matrice è simile ad una matrice diagonale. Questo purtroppo in generale non è vero.

A tale proposito cade il teorema che segue.

Teorema 5 Se una matrice ha n autovalori distinti, allora è diagonalizzabile.

Dimostrazione Se $\lambda_1, \lambda_2, \dots, \lambda_n$ sono gli n autovalori distinti di A e se x^1, x^2, \dots, x^n sono autovettori di A associati agli autovalori, in base al Teorema 3 possiamo affermare che x^1, x^2, \dots, x^n sono l.i. e che quindi la matrice modale V costruita con tali vettori è invertibile. Quindi, come conseguenza del Teorema 4, si ottiene

$$V^{-1}AV = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n),$$

e cioè A è diagonalizzabile. □

▷ **Osservazione** Il teorema appena dimostrato fornisce soltanto una condizione *sufficiente* per la diagonalizzabilità di una matrice. In altre parole il teorema non dice che una matrice, per essere diagonalizzabile, debba necessariamente avere n autovalori distinti.

Il teorema che segue fornisce invece una condizione necessaria e sufficiente per la diagonalizzabilità di una matrice.

Teorema 6 Una matrice A è diagonalizzabile se e solo se ha n autovettori linearmente indipendenti.

Dimostrazione Se A è diagonalizzabile, allora è simile ad una matrice diagonale, cioè esiste una matrice invertibile S tale che la matrice

$$D = S^{-1}AS$$

è diagonale. Ma allora $AS = SD$ e quindi, se s^1, s^2, \dots, s^n sono le colonne di S e d_{jj} sono gli elementi sulla diagonale principale di D , avremo

$$As^j = d_{jj}s^j \quad \text{per ogni } j = 1, 2, \dots, n.$$

Quindi i d_{jj} sono autovalori di A e gli s^j sono autovettori associati a questi. Dato che S è invertibile, possiamo concludere che s^1, s^2, \dots, s^n sono l.i.

Viceversa, se A ha n autovettori l.i. x^1, x^2, \dots, x^n , essi saranno associati agli autovalori $\lambda_1, \lambda_2, \dots, \lambda_n$, non necessariamente distinti. Costruiamo ora con gli autovettori la matrice modale V . Essa è invertibile, e quindi per il Teorema 4, possiamo scrivere $AV = V \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ e quindi la matrice $V^{-1}AV$ è diagonale. □

▷ **Osservazione** Come abbiamo visto, se la matrice A ha n autovalori distinti, allora ci sono senz'altro n autovettori l.i. e la matrice è diagonalizzabile. In questo caso abbiamo n autospazi, tutti di dimensione 1, per così dire indipendenti tra loro.

Se invece gli autovalori *non* sono distinti, e quindi c'è qualche autovalore λ_j di molteplicità $m_a(\lambda_j) \geq 2$, allora possono accadere due cose:

- l'autospazio associato ha dimensione $m_g(\lambda_j) = m_a(\lambda_j)$ e quindi in esso ci sono $m_g(\lambda_j)$ autovettori l.i.
- l'autospazio associato ha dimensione $m_g(\lambda_j) < m_a(\lambda_j)$ e quindi in esso *non* ci sono abbastanza autovettori l.i.

La diagonalizzabilità di A dipende da quale delle due situazioni si verifica. È chiaro che, per essere A diagonalizzabile, deve presentarsi la prima situazione per ogni autovalore di A . Basta che per un particolare autovalore non ci sia coincidenza tra le molteplicità, e questo fa sì che la matrice non sia diagonalizzabile.

Nel seguito vediamo alcune proprietà che coinvolgono il prodotto interno di \mathbb{R}^n . Ricordiamo che, se x e y sono vettori,

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

Osserviamo che se A è una matrice (quadrata)⁷⁷, allora

$$\langle Ax, y \rangle = (Ax)^T y = x^T A^T y = \langle x, A^T y \rangle.$$

Infine, se A è una matrice simmetrica, cioè se $A^T = A$, allora la precedente diventa

$$\langle Ax, y \rangle = \langle x, Ay \rangle.$$

Un paio di risultati che ci saranno utili in seguito.

Teorema 7 $Ax = 0$ per ogni x se e solo se $\langle Ax, y \rangle = 0$ per ogni x, y .⁷⁸

Dimostrazione Ovviamente, se $Ax = 0$ per ogni x , allora $\langle Ax, y \rangle = \langle 0, y \rangle = 0$ per ogni x, y .

Viceversa, se $\langle Ax, y \rangle = 0$ per ogni x, y , allora $\langle Ax, Ax \rangle = 0$ per ogni x . Quindi $\|Ax\| = 0$ per ogni x e dunque $Ax = 0$ per ogni x . \square

Teorema 8 Se A è simmetrica, allora $Ax = 0$ per ogni x se e solo se $\langle Ax, x \rangle = 0$ per ogni x .

Dimostrazione Per il teorema 7, se $Ax = 0$ per ogni x , allora $\langle Ax, x \rangle = 0$ per ogni x .

Viceversa, supponiamo che $\langle Ax, x \rangle = 0$ per ogni x . Allora per ogni x, y si ha

$$\begin{aligned} 0 &= \langle A(x+y), x+y \rangle - \langle Ax, x \rangle - \langle Ay, y \rangle \\ (\text{il prodotto interno è distributivo}) &= \langle Ax, y \rangle + \langle Ay, x \rangle \\ (A \text{ è simmetrica}) &= \langle Ax, y \rangle + \langle y, Ax \rangle \\ (\text{il prodotto interno è commutativo}) &= \langle Ax, y \rangle + \langle Ax, y \rangle \\ &= 2\langle Ax, y \rangle. \end{aligned}$$

Quindi per il Teorema 7 si ha che $Ax = 0$ per ogni x . \square

Definizione Diciamo che una matrice U (o una trasformazione lineare U) è *ortogonale* se $U^{-1} = U^T$. Si può anche dire equivalentemente che U è ortogonale se $U^T U = U U^T = 1$.

▷ *Osservazione* Se U è ortogonale e indicando con x^j le sue colonne, dalla $U^T U = 1$ segue che $\langle x^i, x^j \rangle$ è nullo se $i \neq j$, mentre vale 1 se $i = j$. Quindi possiamo dire che le colonne di U sono una base ortonormale. Lo stesso si può dire per le righe di U , ragionando sulla $U U^T = 1$.

Teorema 9 Sia U una matrice quadrata. Sono equivalenti le seguenti proprietà:

- i) $U^T U = 1$;
- ii) $\langle Ux, Uy \rangle = \langle x, y \rangle$, per ogni x, y ;
- iii) $\|Ux\| = \|x\|$, per ogni x .

Dimostrazione Dimostriamo che i) implica ii). Se vale i), allora

$$\langle Ux, Uy \rangle = \langle x, U^T U y \rangle = \langle x, y \rangle.$$

Dimostriamo che ii) implica iii). Se vale ii), allora come caso particolare

$$\|Ux\|^2 = \langle Ux, Ux \rangle = \langle x, x \rangle = \|x\|^2,$$

da cui segue la iii).

Dimostriamo infine che iii) implica i). Se vale iii), allora per ogni x

$$\begin{aligned} \langle (U^T U - 1)x, x \rangle &= \langle U^T U x, x \rangle - \langle x, x \rangle \\ &= \langle Ux, Ux \rangle - \langle x, x \rangle \\ &= 0, \end{aligned}$$

e quindi, per il teorema 8, $U^T U - 1 = 0$ ⁷⁹, cioè $U^T U = 1$. \square

⁷⁷Possiamo anche considerare A come una trasformazione lineare, valendo le ben note relazioni tra i due concetti, e scrivere quindi Ax per l'immagine di x attraverso la trasformazione A .

⁷⁸Se vediamo A come trasformazione lineare, la proprietà che $Ax = 0$ per ogni x significa naturalmente che la trasformazione è identicamente nulla.

⁷⁹Si osservi che la matrice $U^T U - 1$ è simmetrica dato che è somma di matrici simmetriche.

▷ **Osservazione** La proprietà iii) implica che

$$\text{iv)} \quad \|Ux - Uy\| = \|x - y\|, \quad \text{per ogni } x, y.$$

Vale ovviamente anche il viceversa, cioè iv) implica iii) (si dimostrino entrambe per esercizio).

Quindi le trasformazioni di cui parla il teorema 9, e che hanno tutte le proprietà i), ii), iii) e iv), in particolare conservano le distanze e per questo si chiamano *isometrie*.

Sulle isometrie e sulle matrici simmetriche esistono molti risultati interessanti. Ne citiamo alcuni.

Teorema 10 Se $\mathcal{X} = \{x^1, x^2, \dots, x^n\}$ è una base ortonormale e U è un'isometria, allora anche $\{Ux^1, Ux^2, \dots, Ux^n\}$ è una base ortonormale. Viceversa, se $\{Ux^1, Ux^2, \dots, Ux^n\}$ è una base ortonormale, allora U è un'isometria.

Dimostrazione Supponiamo che $\mathcal{X} = \{x^1, x^2, \dots, x^n\}$ sia una base ortonormale e U sia un'isometria. Allora, per la proprietà ii),

$$\langle Ux^i, Ux^j \rangle = \langle x^i, x^j \rangle, \quad \text{per ogni } x^i, x^j.$$

Il risultato segue dal fatto che gli x sono una base ortonormale.

Viceversa, se entrambe $\{x^1, x^2, \dots, x^n\}$ e $\{Ux^1, Ux^2, \dots, Ux^n\}$ sono basi ortonormali, allora arriviamo al risultato dimostrando che vale la proprietà ii). Anzitutto è immediato che

$$\langle Ux^i, Ux^j \rangle = \langle x^i, x^j \rangle, \quad \text{per ogni } x^i, x^j,$$

dato che entrambi i prodotti interni sono nulli se $i \neq j$ e valgono 1 se $i = j$.

Quindi, se a, b sono due vettori qualunque, allora

$$\begin{aligned} \langle Ua, Ub \rangle &= \langle U(a_1x^1 + a_2x^2 + \dots + a_nx^n), U(b_1x^1 + b_2x^2 + \dots + b_nx^n) \rangle \\ &= \sum_{i,j} a_i b_j \langle Ux^i, Ux^j \rangle \\ &= \sum_{i,j} a_i b_j \langle x^i, x^j \rangle \\ &= \langle a, b \rangle \end{aligned}$$

e la proprietà ii) è dimostrata. □

Teorema 11 Ogni autovalore di un'isometria ha modulo unitario.

Dimostrazione Dai risultati precedenti segue che, se x è un autovettore associato all'autovalore λ , allora

$$\|Ux\| = |\lambda| \|x\| = \|x\|,$$

e quindi (gli autovettori non sono nulli) $|\lambda| = 1$. □

Le matrici simmetriche, dal punto di vista degli autovalori, sono matrici “speciali”. Infatti valgono i seguenti risultati:

Teorema 12 Se A è simmetrica, allora i suoi autovalori sono reali.

▷ **Osservazione** Detto così il risultato può sembrare banale. In effetti nella nostra trattazione gli autovalori (se esistono) sono sempre numeri reali. Quindi il teorema 12, che non afferma che gli autovalori esistono, sostanzialmente dice che, se esistono, sono reali, cosa ovvia nella nostra trattazione.

Lo stesso enunciato avrebbe tutto un altro significato in una trattazione più generale, alla quale si accennava qualche pagina fa. Trattando di matrici in campo complesso, dove gli autovalori sono in genere complessi, risulta molto significativo infatti che, una matrice di un certo tipo⁸⁰ abbia sempre autovalori reali. Il risultato del teorema 12, nella nostra trattazione in campo reale, viene meglio espresso nel modo seguente:⁸¹

Teorema 13 Ogni radice dell'equazione caratteristica di una matrice simmetrica è reale.

▷ **Osservazione** Il teorema 13 afferma che l'equazione caratteristica di una matrice simmetrica $n \times n$ ha n radici reali, se le contiamo con le rispettive molteplicità. Non è detto ovviamente in generale che le radici dell'equazione caratteristica siano distinte.

⁸⁰Non dico espressamente simmetrica perché in quel contesto la simmetria è leggermente diversa: non consiste infatti nella uguaglianza tra la matrice e la sua trasposta, ma tra la matrice e la trasposta della coniugata. A quel punto la si chiama anche in modo diverso, non più simmetrica ma *autoaggiunta*.

⁸¹Non forniamo una dimostrazione del risultato, dato che per ottenere la tesi in modo abbastanza semplice occorre comunque fare ricorso al campo dei numeri complessi, e mi si perdoni il gioco di parole.

▷ *Osservazione* Il risultato del teorema precedente non vale per le isometrie. Infatti, la matrice

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

rappresenta un'isometria, dato che $A^T A = 1$. Essa però non ha autovalori reali, dato che l'equazione caratteristica è $\lambda^2 + 1 = 0$.

Il seguente teorema rafforza, nel caso di matrici simmetriche o di isometrie, il teorema 3.

Teorema 14 Se A è simmetrica oppure un'isometria, allora gli autovettori di A associati ad autovalori distinti sono ortogonali.

Dimostrazione Supponiamo che λ_1 e λ_2 siano autovalori di A e che $\lambda_1 \neq \lambda_2$. Supponiamo inoltre che x e y siano autovettori associati rispettivamente a λ_1 e λ_2 . Vogliamo dimostrare che allora x e y sono ortogonali.

Se A è simmetrica si ha

$$\begin{aligned} \lambda_1 \langle x, y \rangle &= \langle \lambda_1 x, y \rangle \\ &= \langle Ax, y \rangle \\ &= \langle x, Ay \rangle \\ &= \langle x, \lambda_2 y \rangle \\ &= \lambda_2 \langle x, y \rangle. \end{aligned}$$

Quindi

$$(\lambda_1 - \lambda_2) \langle x, y \rangle = 0.$$

Dato che $\lambda_1 - \lambda_2 \neq 0$, deve necessariamente essere $\langle x, y \rangle = 0$.

Vediamo ora la seconda parte, e cioè che lo stesso succede se A è un'isometria. In tale caso abbiamo

$$\begin{aligned} \langle x, y \rangle &= \langle Ax, Ay \rangle \quad (\text{proprietà ii}) \\ &= \langle \lambda_1 x, \lambda_2 y \rangle \\ &= \lambda_1 \lambda_2 \langle x, y \rangle. \end{aligned}$$

Quindi

$$(\lambda_1 \lambda_2 - 1) \langle x, y \rangle = 0.$$

Ricordando il risultato del teorema 11 e ricordando che gli autovalori sono distinti, deve necessariamente essere $\lambda_1 \lambda_2 = -1$, da cui segue ancora che $\langle x, y \rangle = 0$. \square

A questo punto resta ancora aperta una questione: nel caso di matrici simmetriche, o di isometrie, può succedere quello che talvolta succede con matrici qualunque, e cioè che ad un autovalore di molteplicità algebrica maggiore di 2 sia associato un autospazio troppo piccolo per consentire in esso la presenza di abbastanza autovettori indipendenti?

In altre parole la questione è: una matrice simmetrica, o un'isometria, può essere non diagonalizzabile?

Il teorema che segue chiarisce questo aspetto e dà alla domanda una risposta negativa.

Teorema 15 Se A è simmetrica oppure un'isometria, allora la molteplicità algebrica di ogni autovalore è uguale alla molteplicità geometrica di questo.

▷ *Osservazione* Il teorema 15 dice quindi che le matrici simmetriche e le isometrie sono sempre diagonalizzabili.⁸² Conseguenza dei risultati precedenti, in particolare del teorema 14, è anche che la diagonalizzazione di una matrice simmetrica può avvenire attraverso una matrice ortogonale. Quindi, se A è simmetrica, esiste una matrice *ortogonale* U tale che

$$U^T A U = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n),$$

dove $\lambda_1, \lambda_2, \dots, \lambda_n$ sono gli autovalori di A , non tutti necessariamente distinti.

⁸²Per le isometrie la diagonalizzazione potrebbe richiedere il campo complesso in quanto come visto gli autovalori non sono necessariamente reali.

6 Forme quadratiche

Si dice forma quadratica (scriveremo per comodità f.q.) nelle indeterminate x_1, x_2, \dots, x_n un polinomio omogeneo di secondo grado nelle variabili x_1, x_2, \dots, x_n .

Posto $x = (x_1, x_2, \dots, x_n)$, la f.q. è

$$Q(x) = \sum_{i,j=1}^n a_{ij} x_i x_j.$$

Come subito si verifica, usando la scrittura matriciale si può più sinteticamente scrivere

$$Q(x) = \langle Ax, x \rangle = x^T A x,$$

dove A è la matrice dei coefficienti a_{ij} .⁸³

Si verifica facilmente che si può sempre scrivere la f.q. mediante una matrice *simmetrica*. Infatti per ogni i, j possiamo scrivere

$$a_{ij} x_i x_j + a_{ji} x_j x_i = \frac{a_{ij} + a_{ji}}{2} x_i x_j + \frac{a_{ij} + a_{ji}}{2} x_j x_i.$$

D'ora in avanti sottointenderemo che la matrice della f.q. sia simmetrica.

Qualunque f.q. assume il valore 0 in corrispondenza del vettore nullo.

Con riguardo al segno dei valori assunti in corrispondenza di vettori non nulli si possono classificare le f.q. come segue.

Definizione

i) Una f.q. si dice *definita positiva* se

$$Q(x) > 0 \quad \forall x \neq 0;$$

ii) Una f.q. si dice *definita negativa* se

$$Q(x) < 0 \quad \forall x \neq 0;$$

iii) una f.q. si dice *semidefinita positiva* se

$$Q(x) \geq 0 \quad \forall x \quad \text{ed esiste } x \neq 0 \text{ tale che } Q(x) = 0;$$

iv) una f.q. si dice *semidefinita negativa* se

$$Q(x) \leq 0 \quad \forall x \quad \text{ed esiste } x \neq 0 \text{ tale che } Q(x) = 0;$$

v) una f.q. si dice *indefinita* se

$$\text{esistono } x, y \neq 0 \text{ tali che } Q(x) > 0 \text{ e } Q(y) < 0.$$

Per il riconoscimento di una f.q. dal punto di vista del suo segno esistono alcuni metodi pratici, che ora esamineremo. Vedremo in particolare due metodi: uno coinvolge l'uso degli autovalori della matrice della f.q., l'altro coinvolge il segno di alcuni minori della matrice stessa.

Vediamo prima il metodo con gli autovalori. Sia dunque $Q(x) = \langle Ax, x \rangle$ una f.q. Poiché A è simmetrica, essa ha n autovettori ortogonali, che possono essere normalizzati. Supponiamo dunque che x^1, x^2, \dots, x^n sia una base ortonormale formata da autovettori della matrice A .

Con questi vettori possiamo costruire una matrice modale V , che è ortogonale. Abbiamo allora

$$\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = V^T A V, \quad \text{ossia} \quad A = V \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) V^T,$$

dove $\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ è la matrice diagonale degli autovalori di A : chiamiamola per comodità D .

Possiamo quindi scrivere

$$\begin{aligned} Q(x) &= \langle Ax, x \rangle \\ &= \langle V D V^T x, x \rangle \\ &= \langle D V^T x, V^T x \rangle \\ (\text{ponendo } V^T x = y) &= \langle D y, y \rangle. \end{aligned}$$

⁸³Una f.q. non è altro quindi che una particolare funzione di n variabili a valori reali (o una funzione scalare di variabile vettoriale se si preferisce).

Quest'ultima esprime chiaramente la f.q. rispetto alla base ortonormale formata dagli autovettori di A . Il cambio di base è particolarmente conveniente, dato che rispetto a tale base ortonormale la matrice della f.q. è diagonale.

Si noti che l'equazione vettoriale $V^T x = y$ stabilisce una trasformazione lineare invertibile di \mathbb{R}^n in sé. L'ultimo risultato è quindi da intendere nel senso che i valori della f.q. $Q(x)$, al variare di x in tutto \mathbb{R}^n si ottengono attraverso i valori di $\langle Dy, y \rangle$ al variare di y in tutto \mathbb{R}^n .

Possiamo anche dire che $x = 0$ se e solo se $y = 0$, e cioè $x \neq 0$ se e solo se $y \neq 0$. Quindi ai fini del riconoscimento del segno della f.q., lo studio di $\langle Ax, x \rangle$ equivale allo studio di $\langle Dy, y \rangle$, che risulta di molto semplificato, dato che la matrice D è diagonale.

Risulta infatti

$$\langle Dy, y \rangle = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2. \quad {}^{84}$$

A questo punto è facile classificare la f.q. Si può dire che:

- la f.q. è definita positiva se e solo se $\lambda_i > 0 \forall i = 1, \dots, n$;
- la f.q. è definita negativa se e solo se $\lambda_i < 0 \forall i = 1, \dots, n$;
- la f.q. è semidefinita positiva se e solo se $\lambda_i = 0$ per qualche i e $\lambda_i > 0$ per i rimanenti;
- la f.q. è semidefinita negativa se e solo se $\lambda_i = 0$ per qualche i e $\lambda_i < 0$ per i rimanenti;
- la f.q. è indefinita se e solo se ci sono autovalori positivi e autovalori negativi.

▷ **Osservazione** Dimostriamo ad esempio la prima doppia implicazione (lo studente faccia per esercizio le altre). Se la f.q. è definita positiva, significa che $\langle Dy, y \rangle$ è strettamente positivo su tutti i vettori y non nulli. Se per assurdo ci fosse un $\lambda_i \leq 0$, allora avremmo $\langle Du^i, u^i \rangle = \lambda_i \leq 0$, contro l'ipotesi. Se viceversa tutti gli autovalori sono positivi, allora, preso un $y \neq 0$, ci sarà almeno una sua componente $y_i \neq 0$; ma allora $\langle Dy, y \rangle = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2 \geq \lambda_i y_i^2 > 0$.

Vediamo ora alcuni esempi di studio del segno di una f.q. con il metodo degli autovalori.

▷ **Esempio** Consideriamo la f.q.

$$Q(x_1, x_2) = 9x_1^2 + 12x_1x_2 + 4x_2^2.$$

Possiamo scrivere $Q(x) = \langle Ax, x \rangle$, dove

$$A = \begin{pmatrix} 9 & 6 \\ 6 & 4 \end{pmatrix}.$$

Il calcolo degli autovalori di A porta a trovare $\lambda_1 = 0$ e $\lambda_2 = 13$. Quindi la f.q. è semidefinita positiva.

Verifichiamolo in base al segno: abbiamo $9x_1^2 + 12x_1x_2 + 4x_2^2 = (3x_1 + 2x_2)^2$, e quindi si ha chiaramente $Q(x_1, x_2) \geq 0$ ovunque e $Q(x_1, x_2) = 0$ sul sottospazio $\{\alpha(2, -3) : \alpha \in \mathbb{R}\}$.

▷ **Esempio** Consideriamo la f.q.

$$Q(x_1, x_2, x_3) = x_1^2 - 2x_1x_2 + 2x_1x_3 + 2x_2^2 - 2x_2x_3 + 2x_3^2.$$

Possiamo scrivere $Q(x) = \langle Ax, x \rangle$, dove

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

Il calcolo degli autovalori di A porta a trovare $\lambda_1 = 1$, $\lambda_2 = 2 + \sqrt{3}$ e $\lambda_3 = 2 - \sqrt{3}$. Quindi la f.q. è definita positiva. Verifichiamolo in base al segno: si può scrivere infatti

$$x_1^2 - 2x_1x_2 + 2x_1x_3 + 2x_2^2 - 2x_2x_3 + 2x_3^2 = (x_1 - x_2)^2 + (x_2 + x_3)^2 + x_3^2,$$

⁸⁴Basta pensare che $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ è la matrice le cui colonne sono i vettori $\lambda_1 u^1, \lambda_2 u^2, \dots, \lambda_n u^n$ e quindi

$$\begin{aligned} \langle \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)y, y \rangle &= \langle (\lambda_1 u^1 \ \lambda_2 u^2 \ \dots \ \lambda_n u^n), y \rangle \\ &= \left\langle \sum_i \lambda_i u^i y_i, y \right\rangle \\ &= \sum_i \lambda_i y_i \langle u^i, y \rangle \\ &= \sum_i \lambda_i y_i^2. \end{aligned}$$

Si noti anche che gli autovalori non sono necessariamente tutti distinti.

e si tratta intanto di una quantità sicuramente non negativa. Può essere nulla solo se i tre termini quadratici sono contemporaneamente nulli, cioè se

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

sistema omogeneo che ha come unica soluzione la soluzione nulla. Pertanto per definizione la f.q. è definita positiva.

▷ **Esempio** Consideriamo la f.q.

$$Q(x_1, x_2, x_3, x_4) = (x_1 - x_2)^2 + (x_3 + x_4)^2.$$

Si tratta chiaramente di una f.q. semidefinita positiva, dato che è non negativa e si annulla sul sottospazio $\{\alpha(1, 1, 0, 0) + \beta(0, 0, 1, -1) : \alpha, \beta \in \mathbb{R}\}$.

Verifichiamolo con gli autovalori. La matrice della f.q. è

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

Il calcolo degli autovalori di A porta a trovare $\lambda_1 = 0$ (di molteplicità 2) e $\lambda_2 = 2$ (di molteplicità 2). Questo conferma che la f.q. è semidefinita positiva.

▷ **Esempio** Consideriamo la f.q.

$$Q(x_1, x_2, x_3, x_4) = 2x_1x_2 + 2x_2^2 + 2x_2x_3 + x_3^2 + 2x_3x_4 + 2x_4^2.$$

La matrice della f.q. è

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

Il calcolo degli autovalori di A porta a trovare $\lambda_1 = 0$, $\lambda_2 = 3$, $\lambda_3 = 1 + \sqrt{3}$ e $\lambda_4 = 1 - \sqrt{3}$. La f.q. è pertanto indefinita.

▷ **Esercizio** Sia $Q(x) = \langle Ax, x \rangle$ e sia $\lambda = 0$ autovalore di A . Si dimostri che Q si annulla sull'autospazio associato all'autovalore nullo.

▷ **Esercizio** Sia $Q(x) = \langle Ax, x \rangle$. Si dimostri che $Q(x) = \sum_i \lambda_i x_i^2$, dove $x = (x_1, x_2, \dots, x_n)$ e $\lambda_1, \lambda_2, \dots, \lambda_n$ sono gli autovalori di A .

Un altro criterio utile per classificare una f.q. è quello che prende in esame il segno di alcuni particolari minori della matrice della forma quadratica. Occorrono intanto un paio di definizioni.

Definizione Sia A una matrice simmetrica.

- Si chiamano *minori principali di ordine k* di A i minori⁸⁵ che si ottengono considerando sottomatrici formate da k righe e dalle corrispondenti colonne.
- Si chiamano *minori principali di Nord-Ovest (NO) di ordine k* di una matrice simmetrica A i minori principali che si ottengono considerando le prime k righe (e quindi di conseguenza le prime k colonne).

Ad esempio, in una matrice A di ordine 4, vi sono;

- ▷ 4 minori principali di ordine 1⁸⁶,
- ▷ 6 minori principali di ordine 2⁸⁷,
- ▷ 4 minori principali di ordine 3⁸⁸,

⁸⁵Ricordo che *minore* di una matrice A è in generale ogni determinante di una sottomatrice (quadrata) di A . Un minore è di *ordine k* se la sottomatrice da cui proviene è di ordine k , cioè è una sottomatrice $k \times k$.

⁸⁶Sono gli elementi della diagonale principale di A .

⁸⁷Si ottengono scegliendo in tutti i modi possibili 2 righe su 4 nella matrice: quindi 1^a e 2^a, 1^a e 3^a, 1^a e 4^a, 2^a e 3^a, 2^a e 4^a, 3^a e 4^a. Non c'è libertà di scelta sulle colonne, dato che devono essere quelle corrispondenti alle righe. Come noto il numero di scelte possibili di due elementi in un insieme di 4 è $\binom{4}{2}$.

⁸⁸Si ottengono scegliendo in tutti i modi possibili 3 righe su 4 nella matrice e quindi 1^a, 2^a e 3^a, 1^a, 2^a e 4^a, 1^a, 3^a e 4^a e 2^a, 3^a e 4^a.

▷ 1 minore principale di ordine 4⁸⁹

Vi sono invece soltanto 4 minori principali di NO, uno per ogni ordine. Naturalmente i minori principali di NO sono un sottoinsieme dei minori principali.

▷ **Esercizio** Quanti sono i minori principali di NO di una matrice di ordine n ? Quanti sono i minori principali? E infine quanti sono i minori?

Qui di seguito forniamo, senza dimostrazione, le regole per decidere il segno di una forma quadratica in base ai minori principali della sua matrice.

- La f.q. $\langle Ax, x \rangle$ è definita positiva se e solo se tutti i *minori principali di NO* di A sono positivi;
- La f.q. $\langle Ax, x \rangle$ è definita negativa se e solo se tutti i *minori principali di NO* di A di ordine pari sono positivi e tutti i minori principali di NO di A di ordine dispari sono negativi;
- La f.q. $\langle Ax, x \rangle$ è semidefinita positiva se e solo se tutti i *minori principali* di A sono non negativi e $\det A = 0$;
- La f.q. $\langle Ax, x \rangle$ è semidefinita negativa se e solo se tutti i *minori principali* di A di ordine pari sono non negativi, tutti i minori principali di A di ordine dispari sono non positivi e $\det A = 0$;
- la f.q. è indefinita se e solo se non si verifica nessuna delle situazioni precedenti.

Vediamo alcuni esempi di studio del segno di una f.q. con i minori principali, riprendendo anche alcuni degli esempi già proposti in precedenza e già studiati con il metodo degli autovalori.

▷ **Esempio** Consideriamo la f.q.

$$Q(x_1, x_2) = 9x_1^2 + 12x_1x_2 + 4x_2^2.$$

La matrice è

$$A = \begin{pmatrix} 9 & 6 \\ 6 & 4 \end{pmatrix}.$$

Il determinante è nullo e i minori principali di ordine 1 sono positivi: quindi la f.q. è semidefinita positiva.

▷ **Esempio** Consideriamo la f.q.

$$Q(x_1, x_2, x_3) = x_1^2 - 2x_1x_2 + 2x_1x_3 + 2x_2^2 - 2x_2x_3 + 2x_3^2.$$

La matrice è

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

I tre minori principali di NO sono tutti uguali ad 1, quindi la f.q. è definita positiva.

▷ **Esempio** Consideriamo la f.q.

$$Q(x_1, x_2, x_3) = -2x_1^2 + 4x_1x_2 - 4x_2^2 + 5x_2x_3 - 2x_3^2.$$

La matrice è

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

I minori principali di ordine 1 sono negativi; i minori principali di ordine 2 sono positivi; il determinante di A (è un minore principale di ordine 3) è però positivo, e quindi la f.q. è indefinita.

▷ **Esempio** Consideriamo la f.q.

$$Q(x_1, x_2, x_3, x_4) = (x_1 - x_2)^2 + (x_3 + x_4)^2.$$

La matrice è

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

⁸⁹È evidentemente il determinante di A .

Il determinante è nullo e gli altri minori principali sono tutti uguali a 0 oppure a 1. La f.q. è quindi semidefinita positiva.

▷ *Esempio* Consideriamo la f.q.

$$Q(x_1, x_2, x_3, x_4) = 2x_1x_2 + 2x_2^2 + 2x_2x_3 + x_3^2 + 2x_3x_4 + 2x_4^2.$$

La matrice è

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

La f.q. non è definita, dato che il primo minore principale di NO è nullo. Il secondo minore principale di NO è negativo e, essendo di ordine pari, rende la f.q. indefinita. Come si vede il metodo dei minori può rendere lo studio del segno molto più veloce del metodo degli autovalori.