

ALGEBRA LINEARE E GEOMETRIA

Francesco Bottacin

Padova, 2009

Indice

1 Spazi Vettoriali	1
1.1 Sistemi di equazioni lineari	1
1.2 Vettori geometrici	5
1.3 Spazi vettoriali	8
1.3.1 Sottospazi vettoriali	11
1.3.2 Insiemi di generatori e basi	15
1.3.3 Spazi vettoriali finitamente generati	18
2 Applicazioni Lineari e Matrici	27
2.1 Applicazioni lineari	27
2.1.1 Nucleo e immagine	29
2.1.2 Applicazioni lineari e basi	32
2.2 Matrici	35
2.2.1 Matrici quadrate	43
2.2.2 Cambiamenti di base	46
2.3 Sistemi lineari	50
2.3.1 Risoluzione di un sistema lineare: il metodo dell'eliminazione (o metodo di Gauss)	53
2.3.2 Calcolo del rango di una matrice	58
2.3.3 Calcolo dell'inversa di una matrice	60
3 Determinanti	65
3.1 Permutazioni	65
3.1.1 Cicli e trasposizioni	68
3.2 Il determinante di una matrice quadrata	71
3.2.1 Calcolo del determinante di una matrice mediante l'eliminazione di Gauss	86
3.2.2 Il determinante di un endomorfismo	89
3.3 Determinanti e rango	90
3.4 Orientamenti	93
4 Diagonalizzazione degli Endomorfismi	98
4.1 Autovalori e autovettori	98
4.2 La forma canonica di Jordan	105
5 Spazi Vettoriali Euclidei	120
5.1 Lunghezze e angoli	120
5.1.1 Lunghezza di un vettore	120
5.1.2 Angoli	122

5.1.3	Il prodotto scalare in \mathbb{R}^n	125
5.2	Aree e volumi	127
5.3	Forme bilineari	131
5.4	Forme bilineari e matrici	138
5.4.1	Cambiamenti di base	141
5.4.2	Basi ortogonali e ortonormali	144
5.5	Classificazione delle forme bilineari simmetriche reali	148
5.6	Isometrie	155
5.6.1	Isometrie di \mathbb{R}^2	159
6	Geometria Affine	163
6.1	Spazi affini	163
6.2	Sottospazi affini	166
6.3	Sistemi di riferimento	171
6.4	Equazioni dei sottospazi affini	172
6.5	Alcuni risultati di geometria affine	178
6.6	Applicazioni affini	184
6.6.1	Matrici associate alle applicazioni affini	189
6.7	Spazi affini euclidei	194
6.7.1	Angoli	203
6.8	Isometrie degli spazi affini euclidei	205
Bibliografia		209
Indice analitico		210

Capitolo 1

Spazi Vettoriali

In questo capitolo introdurremo la nozione di spazio vettoriale. Per motivare le definizioni che daremo tratteremo dapprima il caso dei cosiddetti “vettori geometrici.” Daremo poi la definizione generale di uno spazio vettoriale su un campo e studieremo le sue principali proprietà. Introdurremo i concetti fondamentali di vettori linearmente indipendenti, di sistemi di generatori e di basi di uno spazio vettoriale. Studieremo poi in dettaglio le proprietà degli spazi vettoriali di dimensione finita.

Prima di addentrarci nello studio degli spazi vettoriali riteniamo utile richiamare brevemente alcuni fatti fondamentali riguardanti i sistemi di equazioni lineari.

1.1 Sistemi di equazioni lineari

In questa sezione studieremo i sistemi di equazioni lineari a coefficienti in un campo, descrivendo un metodo elementare per determinare le loro soluzioni. La trattazione approfondita della teoria dei sistemi lineari verrà sviluppata in un capitolo successivo, quando avremo a disposizione gli strumenti di algebra lineare necessari.

Definizione 1.1.1. Un *sistema di equazioni lineari* a coefficienti in un campo K è un insieme S di equazioni del tipo

$$S : \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \quad (1.1.1)$$

con $a_{ij}, b_h \in K$. Gli elementi a_{ij} sono detti i *coefficienti* del sistema, mentre b_1, \dots, b_m sono i *termini noti*. Le x_1, \dots, x_n sono le *incognite* del sistema. Risolvere il sistema S significa determinare tutti i valori delle incognite x_1, \dots, x_n

che soddisfano contemporaneamente tutte le equazioni di S . Se tutti i termini noti sono nulli il sistema è detto *omogeneo*.

Desriveremo ora il metodo più semplice che si possa immaginare per risolvere un generico sistema di m equazioni lineari in n incognite. Questo metodo, noto come il “metodo della sostituzione,” può essere sommariamente descritto come segue:

Passo 1. Dato il sistema lineare S , scegliamo una qualunque delle sue equazioni, ad esempio la i -esima,

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{ij}x_j + \cdots + a_{in}x_n = b_i$$

e scegliamo una delle incognite x_j che compaiono effettivamente in tale equazione. Ricaviamo x_j in funzione delle incognite rimanenti:

$$x_j = \frac{b_i - a_{i1}x_1 - \cdots - a_{in}x_n}{a_{ij}}.$$

Passo 2. Sostituiamo l'espressione trovata per x_j nelle rimanenti $m - 1$ equazioni, ottenendo così un nuovo sistema S' composto da $m - 1$ equazioni in $n - 1$ incognite.

Passo 3. Se il sistema S' contiene ancora delle incognite, ritorniamo al Passo 1 con S' al posto di S . In caso contrario le soluzioni del sistema, qualora esistano, possono essere ottenute con una semplice sostituzione all'indietro, partendo dall'ultima incognita che è stata determinata.

Vediamo ora di chiarire l'algoritmo appena descritto analizzando in dettaglio tre esempi concreti, che rappresentano le tre situazioni tipiche che si possono presentare.

Esempio 1. (*Sistema privo di soluzioni*) Consideriamo il seguente sistema di equazioni lineari, a coefficienti nel campo \mathbb{Q} :

$$S : \begin{cases} 2x_1 - x_2 = 1 \\ x_1 + 4x_2 = -2 \\ 3x_1 - 2x_2 = 3 \end{cases}$$

Scegliamo la seconda equazione e da questa ricaviamo x_1 , ottenendo $x_1 = -2 - 4x_2$. Sostituendo questa espressione nella prima e nella terza equazione, ottenendo il sistema

$$S' : \begin{cases} -9x_2 = 5 \\ -14x_2 = 9 \end{cases}$$

Consideriamo ora il sistema S' e ricaviamo x_2 dalla prima equazione: $x_2 = -5/9$. Sostituendo questo valore nella terza equazione si ottiene l'uguaglianza $-14(-5/9) = 9$, che non è verificata. Da ciò si deduce che il sistema S non ammette soluzioni.

Esempio 2. (*Sistema che ammette un'unica soluzione*) Consideriamo ora il seguente sistema lineare, a coefficienti in \mathbb{Q} :

$$S : \begin{cases} 3x_1 - x_2 = 2 \\ 2x_1 + 5x_2 = -3 \end{cases}$$

Dalla prima equazione ricaviamo $x_2 = 3x_1 - 2$. Se sostituiamo questa espressione nella seconda equazione otteniamo il “sistema” (che consiste di una sola equazione)

$$S' : 17x_1 = 7.$$

Da questa equazione si ottiene $x_1 = 7/17$. Ora non resta che “sostituire all’indietro” il valore di x_1 appena trovato nella precedente espressione per x_2 , ottenendo $x_2 = -13/17$. Si conclude pertanto che il sistema S ammette un’unica soluzione data da

$$x_1 = 7/17, \quad x_2 = -13/17.$$

Esempio 3. (*Sistema con infinite soluzioni*) Consideriamo il seguente sistema di equazioni lineari, a coefficienti nel campo \mathbb{Q} :

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

Dalla prima equazione ricaviamo $x_3 = 3x_1 + 2x_2 + x_4 - 2$. Sostituendo questa espressione nella seconda equazione ottenendo il “sistema”

$$S' : 8x_1 + 4x_2 - x_4 = 4.$$

Da questa equazione possiamo ricavare, ad esempio, $x_4 = 8x_1 + 4x_2 - 4$. Arrivati a questo punto ci dobbiamo arrestare, dato che non ci sono altre equazioni che possano essere utilizzate. Abbiamo pertanto determinato il valore di x_4 in funzione delle incognite x_1 e x_2 che rimangono libere di assumere qualsiasi valore. Sostituendo all’indietro l’espressione di x_4 nella precedente espressione per l’incognita x_3 , otteniamo $x_3 = 11x_1 + 6x_2 - 6$.

In conclusione, possiamo affermare che il sistema S ammette infinite soluzioni, le quali dipendono da due parametri¹ liberi di variare:

$$\begin{cases} x_1 \text{ qualsiasi} \\ x_2 \text{ qualsiasi} \\ x_3 = 11x_1 + 6x_2 - 6 \\ x_4 = 8x_1 + 4x_2 - 4 \end{cases}$$

Naturalmente avremmo potuto seguire una strada diversa e ricavare, ad esempio, le incognite x_1 e x_2 in funzione di x_3 e x_4 . In questo modo avremmo ottenuto una diversa espressione per le soluzioni del sistema S . Ovviamente tutte le possibili espressioni per le soluzioni di uno stesso sistema sono tra loro equivalenti.

Esercizi

Esercizio 1.1.1. Risolvere il seguente sistema a coefficienti in \mathbb{Q} :

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

¹Questo fatto viene a volte espresso affermando che il sistema ammette ∞^2 soluzioni.

Esercizio 1.1.2. Risolvere il seguente sistema a coefficienti in \mathbb{Q} :

$$\begin{cases} 2x + 3y + 4z = 3 \\ 4x - y - 2z = 1 \\ 5x - 3y + z = -2. \end{cases}$$

Esercizio 1.1.3. Risolvere il seguente sistema a coefficienti in \mathbb{Q} :

$$\begin{cases} 5x - 2y + z = 1 \\ 3x - 4y - 2z = 3. \end{cases}$$

Esercizio 1.1.4. Risolvere il seguente sistema a coefficienti in $\mathbb{Z}/2\mathbb{Z}$:

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

Esercizio 1.1.5. Risolvere il seguente sistema a coefficienti in $\mathbb{Z}/5\mathbb{Z}$:

$$\begin{cases} x + 3y + 2z = 3 \\ 2x + z = 2 \\ 3x + y + 4z = 3. \end{cases}$$

Esercizio 1.1.6. Risolvere il seguente sistema a coefficienti in $\mathbb{Z}/11\mathbb{Z}$:

$$\begin{cases} 5x + y + 2z + 3w = 1 \\ 2x + 7y + z + 5w = 4 \\ 8x + 2y + 3z + w = 7. \end{cases}$$

Esercizio 1.1.7. Risolvere e discutere in funzione dei valori di $m \in \mathbb{Q}$ il seguente sistema:

$$\begin{cases} x + (m+1)y = m+2 \\ mx + (m+4)y = 3. \end{cases}$$

Esercizio 1.1.8. Risolvere e discutere in funzione dei valori di $m \in \mathbb{R}$ il seguente sistema:

$$\begin{cases} mx + (m-1)y = m+2 \\ (m+1)x - my = 5m+3. \end{cases}$$

Esercizio 1.1.9. Risolvere e discutere in funzione dei valori di $a \in \mathbb{R}$ il seguente sistema:

$$\begin{cases} x + (a-1)y + (2-a)z = a+5 \\ x + ay + 2z = 4 \\ x + (a-2)y + (2-2a^2)z = 6. \end{cases}$$

Esercizio 1.1.10. Al variare di $\lambda \in \mathbb{Q}$ si dica quante soluzioni vi sono per il seguente sistema di equazioni lineari:

$$\begin{cases} (\lambda-1)x_1 + 2x_2 + 3x_4 = 0 \\ \lambda x_2 + (\lambda+1)x_4 = 1 \\ x_1 + \lambda x_3 + x_4 = 0 \\ (\lambda-1)x_1 + x_4 = 0. \end{cases}$$

Esercizio 1.1.11. Dato il sistema di equazioni lineari

$$\begin{cases} \lambda x - \mu y - \mu z = \mu \\ \mu x - \lambda y = \lambda \\ x - y - z = 0 \end{cases}$$

si dica per quali valori di $\lambda, \mu \in \mathbb{R}$ esso è risolubile e per quali la soluzione è unica.

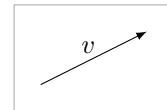
1.2 Vettori geometrici

Il concetto di vettore viene spesso introdotto ricorrendo a delle motivazioni che provengono dalla fisica. In fisica infatti, accanto a grandezze che possono essere adeguatamente espresse con un singolo numero, come ad esempio la temperatura o il tempo, ce ne sono altre la cui descrizione richiede più informazioni. Per descrivere, ad esempio, lo spostamento di un punto, la sola informazione numerica relativa alla “misura” di tale spostamento non basta; è necessario specificare anche la retta lungo la quale avviene lo spostamento e, per finire, occorre specificare anche il verso di percorrenza di tale retta. In modo analogo, per specificare una forza, occorre fornire il valore numerico dell’entità di tale forza (in una qualche unità di misura) assieme alla direzione e al verso di applicazione della forza (in molti casi ciò non è ancora sufficiente, ed occorre specificare anche il punto di applicazione della forza).

Per motivare le considerazioni che faremo nel seguito, ricorreremo al concetto geometrico di “movimento” o, più precisamente, alla nozione di *traslazione* in un piano.

In base a ciò che abbiamo appena detto, per descrivere una traslazione è necessario specificare una retta (la *direzione* in cui avviene lo spostamento), un *verso* di percorrenza di tale retta e, infine, un *numero* che, in qualche modo, misura l’entità di tale traslazione.

Un modo particolarmente comodo per esprimere graficamente tutte queste informazioni è quello di utilizzare un *segmento orientato* v .



La retta su cui giace questo segmento individua la direzione, la freccia posta in una delle due estremità specifica il verso di percorrenza e la lunghezza del segmento stesso (espressa in qualche unità di misura) determina l’entità dello spostamento.

Un oggetto di questo tipo è chiamato *vettore*. Dato un vettore v , rappresentato da un segmento orientato come sopra, la lunghezza di tale segmento è detta il *modulo* (o la *norma*) di v , ed è indicata con $|v|$ (oppure con $\|v\|$).

A questo punto è forse necessaria una precisazione. Un vettore v non descrive lo spostamento di un qualche punto fissato A verso un qualche altro punto B ; esso descrive una traslazione di tutto il piano (o di tutto lo spazio). In tal senso, non ha alcuna importanza “dove” si disegni il segmento che rappresenta graficamente il vettore. In altre parole, due diversi segmenti orientati che abbiamo, tuttavia, la stessa direzione (cioè che si trovino su due rette parallele), lo stesso verso e la stessa lunghezza, sono due rappresentazioni grafiche diverse dello stesso vettore.

Questa idea può essere resa matematicamente precisa nel modo seguente.

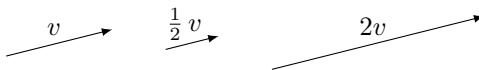
Definizione 1.2.1. Due segmenti orientati sono detti *equipollenti* se hanno la stessa direzione (cioè sono contenuti in due rette parallele), lo stesso verso e la stessa lunghezza.

Si verifica facilmente che la relazione di equipollenza è una relazione di equivalenza nell’insieme di tutti i segmenti orientati. Possiamo quindi dare la seguente definizione di vettore:

Definizione 1.2.2. Un *vettore (geometrico)* è una classe di equipollenza di segmenti orientati.

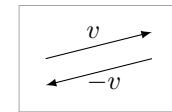
Osservazione 1.2.3. Come abbiamo già fatto notare, a volte è importante specificare anche il punto in cui un vettore si intende “applicato” (come nel caso di una forza). Ciò porta alla definizione della nozione di *vettore applicato*, che deve essere inteso come una coppia (P, v) costituita da un punto P (il punto di applicazione) e da un vettore v .

Nell’insieme dei vettori geometrici sono definite, in modo del tutto naturale, due operazioni. La prima consiste nella moltiplicazione di un vettore v per un numero (reale) λ . Se v rappresenta una determinata traslazione e se $\lambda > 0$, il vettore λv rappresenta una traslazione che avviene nella stessa direzione e nello stesso verso di quella rappresentata da v , ma determina uno spostamento pari a λ volte quello effettuato dalla traslazione rappresentata da v . Il vettore λv è quindi rappresentato da un segmento orientato che ha la stessa direzione e verso di v , ma una lunghezza pari alla lunghezza di v moltiplicata per λ .



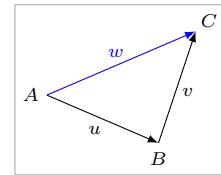
Se $\lambda = 0$ si ottiene un vettore di lunghezza nulla, che corrisponde a una “traslazione nulla.” In questo caso le nozioni di direzione e verso non hanno più alcun significato; il segmento orientato si riduce a un punto, il quale non ha più alcuna direzione e alcun verso.

Se $\lambda < 0$ si intende che il vettore λv rappresenta una traslazione che avviene nella stessa direzione ma nel verso opposto a quella rappresentata da v , per uno spostamento pari al valore assoluto di λ moltiplicato per lo spostamento effettuato dalla traslazione rappresentata da v .



In questo modo si ha che la composizione delle traslazioni corrispondenti ai vettori λv e $-\lambda v$ è la traslazione nulla: $\lambda v + (-\lambda v) = 0$.

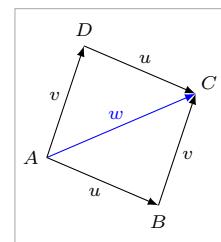
La seconda operazione che consideriamo è la *somma* di due vettori; essa corrisponde alla composizione di due traslazioni. Se u e v sono due vettori, la loro somma $w = u + v$ è, per definizione, il vettore che rappresenta la traslazione che si ottiene effettuando prima la traslazione rappresentata da u e poi quella rappresentata da v . L’effetto di questa composizione di traslazioni è rappresentato nella figura a lato.



Se la traslazione rappresentata da u porta il punto A nel punto B e la traslazione rappresentata da v porta il punto B nel punto C , allora la composizione delle due traslazioni, rappresentata da $w = u + v$, porta il punto A nel punto C .

Si verifica immediatamente che la somma di vettori gode della proprietà commutativa, cioè $u + v = v + u$, come si può vedere nella figura a fianco.

Questa figura illustra la cosiddetta *regola del parallelogramma*: il vettore $w = u + v$ è la diagonale del parallelogramma che ha come lati i vettori u e v .



Se fissiamo un sistema di coordinate cartesiane ortogonali OXY nel piano, ogni vettore v può essere rappresentato da una coppia di numeri reali (v_x, v_y) , che individuano le proiezioni di v sugli assi coordinati (vedi figura 1.1).

Possiamo quindi identificare il vettore v con la coppia $(v_x, v_y) \in \mathbb{R}^2$. In termini di questa identificazione, la somma dei due vettori $u = (u_x, u_y)$ e $v =$

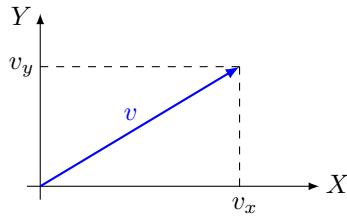


Figura 1.1: Decomposizione di un vettore nelle sue componenti

(v_x, v_y) è data da

$$u + v = (u_x + v_x, u_y + v_y),$$

mentre il prodotto di un numero reale λ per il vettore $v = (v_x, v_y)$ è dato da

$$\lambda v = (\lambda v_x, \lambda v_y).$$

Usando queste formule è ora immediato verificare che la somma di vettori gode delle proprietà associativa e commutativa. Esiste poi un elemento neutro per la somma, il *vettore nullo*, le cui componenti sono tutte nulle, e che indicheremo con $\mathbf{0} = (0, 0)$. Inoltre, per ogni vettore $v = (v_x, v_y)$ esiste il suo opposto $-v = (-v_x, -v_y)$, tale che $v + (-v) = \mathbf{0}$.

Tutto ciò si può riassumere dicendo che l'insieme dei vettori, con l'operazione di somma, forma un *gruppo abeliano*.

Consideriamo ora l'operazione di prodotto tra un numero reale e un vettore. È immediato verificare che questa operazione soddisfa le seguenti proprietà:

- (i) $(\lambda\mu)v = \lambda(\mu v)$,
- (ii) $\lambda(u + v) = \lambda u + \lambda v$,
- (iii) $(\lambda + \mu)v = \lambda v + \mu v$,
- (iv) $1v = v$,

per ogni $\lambda, \mu \in \mathbb{R}$ e per ogni coppia di vettori u e v .

L'insieme dei vettori ha quindi una struttura più ricca di quella di un semplice gruppo abeliano. A questo tipo di struttura daremo il nome di *spazio vettoriale*.

Prima di concludere osserviamo che delle considerazioni del tutto analoghe si possono fare per vettori nell'usuale spazio tridimensionale. Ad ogni tale vettore v si può associare una terna di numeri $(v_x, v_y, v_z) \in \mathbb{R}^3$, i quali rappresentano le proiezioni di v sui tre assi coordinati di un opportuno sistema di riferimento *OXYZ* fissato, come mostrato nella figura 1.2.

Si ottiene in questo modo un'identificazione tra vettori dello spazio tridimensionale e terne di numeri reali, in termini della quale la somma di due vettori $u = (u_x, u_y, u_z)$ e $v = (v_x, v_y, v_z)$ è data da

$$u + v = (u_x + v_x, u_y + v_y, u_z + v_z),$$

e il prodotto di un numero reale λ per un vettore $v = (v_x, v_y, v_z)$ è dato da

$$\lambda v = (\lambda v_x, \lambda v_y, \lambda v_z).$$

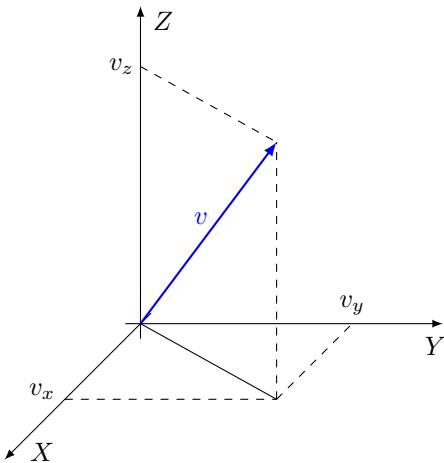


Figura 1.2: Decomposizione di un vettore nelle sue componenti

Osservazione 1.2.4. Prendendo spunto dalle considerazioni precedenti possiamo definire dei vettori a n componenti (vettori di uno spazio n -dimensionale) semplicemente identificandoli con delle n -uple di numeri reali, $v = (a_1, a_2, \dots, a_n) \in \mathbb{R}^n$. Le operazioni di somma di due vettori e di prodotto di un numero reale per un vettore saranno definite in modo analogo a quanto abbiamo già visto nel caso di \mathbb{R}^2 e di \mathbb{R}^3 :

$$(a_1, a_2, \dots, a_n) + (b_1, b_2, \dots, b_n) = (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n)$$

e

$$\lambda (a_1, a_2, \dots, a_n) = (\lambda a_1, \lambda a_2, \dots, \lambda a_n).$$

Questa costruzione verrà ampiamente studiata (e opportunamente generalizzata) nelle prossime sezioni.

1.3 Spazi vettoriali

Nella sezione precedente abbiamo visto che i vettori sono degli “oggetti” che possono essere sommati tra loro e possono anche essere moltiplicati per dei numeri, in modo tale che le operazioni così definite soddisfino tutta una serie di proprietà, essenzialmente analoghe alle usuali proprietà che valgono per la somma e il prodotto tra numeri.

Se ora concentriamo la nostra attenzione non tanto sulla natura di tali “oggetti” (definiti in precedenza come classi di equipotenza di segmenti orientati), quanto piuttosto sull’esistenza di determinate operazioni tra di essi e sulle proprietà che, ragionevolmente, dovrebbero essere soddisfatte da queste operazioni, possiamo fornire una definizione astratta di “insieme di vettori” come un qualche insieme nel quale sono definite un’operazione di somma tra i suoi elementi e un’operazione di prodotto tra un elemento di tale insieme e un numero, in modo tale che siano soddisfatte delle proprietà analoghe a quelle elencate nella sezione precedente.

Cerchiamo ora di rendere precise le idee espresse fin'ora. Sia K un campo² (tanto per fissare le idee, si può supporre che K sia il campo \mathbb{Q} dei numeri razionali, oppure il campo \mathbb{R} dei numeri reali, oppure ancora il campo \mathbb{C} dei numeri complessi).

Definizione 1.3.1. Uno *spazio vettoriale* su K è un insieme non vuoto V dotato di un'operazione $+_V$, detta *somma*,

$$+_V : V \times V \rightarrow V, \quad (v_1, v_2) \mapsto v_1 +_V v_2,$$

e di un'operazione \cdot_V

$$\cdot_V : K \times V \rightarrow V, \quad (\lambda, v) \mapsto \lambda \cdot_V v,$$

detta *prodotto per uno scalare*, che soddisfano le seguenti proprietà: per ogni $\lambda, \lambda_1, \lambda_2 \in K$ e ogni $v, v_1, v_2 \in V$ si ha:

- (1) $(v_1 +_V v_2) +_V v_3 = v_1 +_V (v_2 +_V v_3)$;
- (2) $v_1 +_V v_2 = v_2 +_V v_1$;
- (3) esiste un elemento $\mathbf{0}_V \in V$ tale che $v +_V \mathbf{0}_V = \mathbf{0}_V +_V v = v$;
- (4) per ogni $v \in V$ esiste un elemento $v' \in V$ tale che $v +_V v' = v' +_V v = \mathbf{0}_V$.
Tale elemento v' viene indicato con $-v$ e detto l'opposto di v ;
- (5) $\lambda \cdot_V (v_1 +_V v_2) = (\lambda \cdot_V v_1) +_V (\lambda \cdot_V v_2)$;
- (6) $(\lambda_1 + \lambda_2) \cdot_V v = (\lambda_1 \cdot_V v) +_V (\lambda_2 \cdot_V v)$;
- (7) $(\lambda_1 \lambda_2) \cdot_V v = \lambda_1 \cdot_V (\lambda_2 \cdot_V v)$;
- (8) $1 \cdot_V v = v$.

Gli elementi di uno spazio vettoriale V sono detti *vettori*. Gli elementi del campo K sono detti *scalari*.

Osservazione 1.3.2. Dalle proprietà sopra elencate segue che, in ogni spazio vettoriale V , si ha $0 \cdot_V v = \mathbf{0}_V$, per ogni $v \in V$. Infatti si ha:

$$v + 0 \cdot_V v = 1 \cdot_V v + 0 \cdot_V v = (1 + 0) \cdot_V v = v.$$

Sommando ad ambo i membri di questa uguaglianza l'opposto di v , si ottiene

$$-v + v + 0 \cdot_V v = -v + v = \mathbf{0}_V,$$

da cui segue $0 \cdot_V v = \mathbf{0}_V$. Da ciò possiamo ora dedurre che $(-1) \cdot_V v = -v$. Infatti, si ha:

$$\mathbf{0}_V = 0 \cdot_V v = (1 - 1) \cdot_V v = 1 \cdot_V v + (-1) \cdot_V v = v + (-1) \cdot_V v,$$

da cui segue che il vettore $(-1) \cdot_V v$ è l'opposto di v .

D'ora in poi l'operazione di somma in uno spazio vettoriale V sarà indicata semplicemente con $+$ mentre il simbolo del prodotto per uno scalare sarà omesso: si scriverà quindi $v_1 + v_2$ al posto di $v_1 +_V v_2$ e λv al posto di $\lambda \cdot_V v$.

²Ricordiamo che un campo è un insieme dotato di due operazioni, che indicheremo con $+$ e \cdot , le quali soddisfano delle proprietà del tutto analoghe a quelle della somma e del prodotto di numeri razionali. Più precisamente, un campo è un anello commutativo con unità in cui ogni elemento diverso da 0 ammette un inverso moltiplicativo.

Esempio 1.3.3. Sia $V = K^n$ e definiamo un'operazione di somma tra elementi di V ponendo

$$(a_1, a_2, \dots, a_n) + (b_1, b_2, \dots, b_n) = (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n),$$

e un'operazione di prodotto tra elementi dal campo K ed elementi di V ponendo

$$\lambda(a_1, a_2, \dots, a_n) = (\lambda a_1, \lambda a_2, \dots, \lambda a_n).$$

È immediato verificare che V , con le operazioni appena definite, è uno spazio vettoriale su K .

Esempio 1.3.4. Sia K un campo e indichiamo con $K[X]$ l'insieme dei polinomi a coefficienti in K nell'indeterminata X . Un generico elemento di $K[X]$ si scrive nella forma

$$p(X) = a_0 + a_1 X + a_2 X^2 + \dots + a_n X^n,$$

per qualche $n \geq 0$, ove tutti i coefficienti a_i sono elementi di K .

Rispetto alle operazioni di somma di polinomi e di prodotto di un polinomio per un elemento di K , l'insieme $K[X]$ è uno spazio vettoriale.

Esempio 1.3.5. Sia K un campo e sia S un insieme (non vuoto) qualsiasi. Indichiamo con K^S l'insieme di tutte le funzioni $f : S \rightarrow K$.

Date due funzioni $f, g \in K^S$ possiamo definire la loro somma ponendo

$$(f + g)(s) = f(s) + g(s),$$

e possiamo definire il prodotto di una funzione f per uno scalare $\lambda \in K$ ponendo,

$$(\lambda f)(s) = \lambda(f(s)),$$

per ogni $s \in S$.

Anche in questo caso è immediato verificare che l'insieme K^S , con le operazioni appena definite, è uno spazio vettoriale su K .

Esempio 1.3.6. Sia $K = \mathbb{Q}$ il campo dei numeri razionali e sia $V = \mathbb{R}$. Rispetto alle usuali operazioni di somma e prodotto tra numeri, V risulta essere uno spazio vettoriale su K .

Più in generale, per ogni campo K e ogni estensione di campi $K \subset L$, L risulta essere uno spazio vettoriale su K .

Osservazione 1.3.7. Vogliamo far notare che, nella definizione di spazio vettoriale, la proprietà (8) è necessaria. Infatti l'uguaglianza $1 \cdot_V v = v$ non discende dalle prime sette proprietà, come si può vedere dal seguente esempio.

Sia $V = K^n$. Definiamo la somma di vettori componente per componente (come nell'Esempio 1.3.3) e definiamo il prodotto di un vettore per un elemento di K come segue:

$$\lambda(a_1, a_2, \dots, a_n) = (0, 0, \dots, 0),$$

per ogni $\lambda \in K$ e ogni $(a_1, a_2, \dots, a_n) \in V$.

È immediato verificare che le due operazioni così definite verificano tutte le proprietà elencate nella definizione di spazio vettoriale, ad eccezione della (8).

Osservazione 1.3.8. Si noti che nella definizione di spazio vettoriale non si usa mai il fatto che K sia un campo.

Una definizione del tutto analoga si può dare supponendo solo che K sia un anello commutativo (con unità). L'analogo di uno spazio vettoriale è chiamato, in questo caso, un *modulo* sull'anello K .

Tuttavia, come avremo occasione di osservare in seguito, molti risultati che dimostreremo per gli spazi vettoriali dipendono in modo essenziale dal fatto che K sia un campo e non valgono, invece, per un modulo su un anello. In effetti, la teoria dei moduli risulta essere profondamente diversa dalla teoria degli spazi vettoriali.

Terminiamo questa sezione con la seguente definizione:

Definizione 1.3.9. Sia V uno spazio vettoriale su K . Una *combinazione lineare* di elementi di V è una somma finita del tipo

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n,$$

con $\lambda_1, \dots, \lambda_n \in K$ e $v_1, \dots, v_n \in V$.

1.3.1 Sottospazi vettoriali

Sia V uno spazio vettoriale definito sul campo K .

Definizione 1.3.10. Un *sottospazio vettoriale* W di V è un sottoinsieme non vuoto $W \subseteq V$ tale che la restrizione a W delle operazioni di somma e di prodotto per uno scalare definite su V rende W uno spazio vettoriale sul campo K .

Dalla definizione si deduce che, affinché un sottoinsieme non vuoto W di V sia un sottospazio vettoriale, è necessario e sufficiente che valgano le seguenti proprietà:

- (1) per ogni $w_1, w_2 \in W$, si ha $w_1 + w_2 \in W$;
- (2) per ogni $w \in W$, anche $-w \in W$;
- (3) $\mathbf{0}_V \in W$;
- (4) per ogni $\lambda \in K$ e ogni $w \in W$, si ha $\lambda w \in W$.

In effetti, è sufficiente richiedere che W sia *chiuso* per le operazioni di somma e di prodotto per uno scalare, cioè che si abbia

$$w_1 + w_2 \in W, \quad \forall w_1, w_2 \in W$$

e

$$\lambda w \in W, \quad \forall \lambda \in K, \forall w \in W.$$

Queste due condizioni possono essere raggruppate in una sola:

Proposizione 1.3.11. Un sottoinsieme non vuoto W di uno spazio vettoriale V sul campo K è un sottospazio vettoriale di V se e solo se

$$\lambda_1 w_1 + \lambda_2 w_2 \in W,$$

per ogni $\lambda_1, \lambda_2 \in K$ e ogni $w_1, w_2 \in W$.

Dimostrazione. È immediato verificare che, sotto questa ipotesi, le operazioni di somma di vettori e di prodotto di un vettore per uno scalare definite in V rendono W un sottospazio vettoriale. \square

Osservazione 1.3.12. Ogni spazio vettoriale V è, naturalmente, un sottospazio vettoriale di sé stesso. Inoltre $\{\mathbf{0}_V\}$ è banalmente un sottospazio vettoriale di V , detto il *sottospazio nullo*. A volte scriveremo semplicemente 0 per indicare il sottospazio $\{\mathbf{0}_V\}$ (il significato sarà chiaro dal contesto).

Esempio 1.3.13. Sia $V = K^n$ e sia W l'insieme dei vettori $w = (x_1, x_2, \dots, x_n)$ che sono soluzioni di un'equazione lineare del tipo

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = 0,$$

con $a_1, \dots, a_n \in K$ fissati.

È immediato verificare che una combinazione lineare di due elementi di W fornisce ancora una soluzione della precedente equazione, quindi appartiene a W . Ciò significa che W è un sottospazio vettoriale di V .

Al contrario, l'insieme delle soluzioni di un'equazione del tipo

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = k,$$

con $k \neq 0$, non è un sottospazio vettoriale di V , dato che non contiene il vettore nullo $\mathbf{0}$.

Proposizione 1.3.14. *Se $\{W_i\}_{i \in I}$ è una famiglia di sottospazi vettoriali di uno spazio vettoriale V , allora anche la loro intersezione W*

$$W = \bigcap_{i \in I} W_i$$

è un sottospazio vettoriale di V .

Dimostrazione. Siano $\lambda_1, \lambda_2 \in K$ e $w_1, w_2 \in W$. Allora $w_1, w_2 \in W_i$, per ogni $i \in I$, quindi anche $\lambda_1w_1 + \lambda_2w_2 \in W_i$, dato che W_i è un sottospazio vettoriale di V . Da ciò segue che $\lambda_1w_1 + \lambda_2w_2 \in W$. \square

Osservazione 1.3.15. Una proprietà analoga non vale invece per l'unione: se W_1 e W_2 sono due sottospazi vettoriali di V , la loro unione $W_1 \cup W_2$ non è, in generale, un sottospazio vettoriale di V .

A titolo di esempio, consideriamo lo spazio vettoriale $V = K^2$. Poniamo $W_1 = \{(a, 0) \mid a \in K\}$ e $W_2 = \{(0, b) \mid b \in K\}$. È immediato verificare che essi sono due sottospazi vettoriali di V . Si ha $(1, 0) \in W_1$ e $(0, 1) \in W_2$, tuttavia la loro somma $(1, 1)$ non appartiene né a W_1 né a W_2 . Ciò dimostra che l'insieme $W_1 \cup W_2$ non è chiuso per l'operazione di somma, quindi non può essere un sottospazio vettoriale.

Definizione 1.3.16. Sia S un sottoinsieme di uno spazio vettoriale V . Il *sottospazio vettoriale generato da S* , che indicheremo con $L(S)$, è il più piccolo³ sottospazio vettoriale di V contenente S (se S è vuoto si ha $L(S) = \{\mathbf{0}\}$).

³Più piccolo, inteso rispetto alla relazione d'ordine data dall'inclusione.

Dato che l'intersezione di una famiglia di sottospazi vettoriali di V è un sottospazio vettoriale di V , è immediato verificare che $L(S)$ coincide con l'intersezione di tutti i sottospazi vettoriali di V che contengono S .

Un'altra descrizione, ancora più esplicita, di $L(S)$ è data dalla seguente proposizione.

Proposizione 1.3.17. *Il sottospazio vettoriale $L(S)$ generato da S è l'insieme di tutte le combinazioni lineari finite di elementi di S , cioè*

$$L(S) = \left\{ \sum_{i=1}^n \lambda_i v_i \mid n \in \mathbb{N}, \lambda_i \in K, v_i \in S \right\},$$

dove si intende che la combinazione lineare di zero elementi di S è il vettore nullo $\mathbf{0} \in V$.

Dimostrazione. Poniamo

$$\Lambda(S) = \left\{ \sum_{i=1}^n \lambda_i v_i \mid n \in \mathbb{N}, \lambda_i \in K, v_i \in S \right\}.$$

Ovviamente $S \subseteq \Lambda(S)$. Notiamo che ogni sottospazio vettoriale di V contenente S contiene anche tutte le combinazioni lineari finite di elementi di S , quindi contiene $\Lambda(S)$. Da ciò segue che $\Lambda(S)$ è contenuto nell'intersezione di tutti i sottospazi vettoriali di V che contengono S , quindi $\Lambda(S) \subseteq L(S)$.

D'altra parte è evidente che $\Lambda(S)$ è anch'esso un sottospazio vettoriale di V : infatti la combinazione lineare di due combinazioni lineari finite di elementi di S è essa stessa una combinazione lineare finita di elementi di S . Poiché $L(S)$ è il più piccolo sottospazio vettoriale di V contenente S , si ha dunque $L(S) \subseteq \Lambda(S)$, da cui segue che $L(S) = \Lambda(S)$. \square

Osservazione 1.3.18. Se $S = \{v_1, v_2, \dots, v_n\}$, il sottospazio vettoriale $L(S)$ viene anche indicato con $\langle v_1, v_2, \dots, v_n \rangle$.

Come abbiamo già osservato, nel contesto degli spazi vettoriali l'operazione di unione di due sottospazi non ha delle buone proprietà: infatti l'unione di due sottospazi vettoriali non è, in generale, un sottospazio vettoriale (vedi l'Osservazione 1.3.15). Tale operazione viene quindi sostituita dall'operazione di somma:

Definizione 1.3.19. Se W_1 e W_2 sono sottospazi vettoriali di V , la loro *somma* $W_1 + W_2$ è il sottospazio vettoriale $L(W_1 \cup W_2)$ generato da $W_1 \cup W_2$. Tale definizione si generalizza, in modo ovvio, al caso della somma di una famiglia qualsiasi (anche infinita) di sottospazi di V .

Una descrizione esplicita della somma di due sottospazi vettoriali è fornita dalla seguente proposizione:

Proposizione 1.3.20. *Si ha*

$$W_1 + W_2 = \{w_1 + w_2 \mid w_1 \in W_1, w_2 \in W_2\}.$$

Dimostrazione. È immediato verificare che l'insieme

$$\{w_1 + w_2 \mid w_1 \in W_1, w_2 \in W_2\}$$

è un sottospazio vettoriale di V che contiene W_1 e W_2 , quindi contiene anche la loro unione. Poiché $W_1 + W_2$ è, per definizione, il più piccolo sottospazio vettoriale di V contenente $W_1 \cup W_2$, si ha l'inclusione

$$W_1 + W_2 \subseteq \{w_1 + w_2 \mid w_1 \in W_1, w_2 \in W_2\}.$$

D'altra parte, ogni vettore del tipo $w_1 + w_2$ appartiene necessariamente a $W_1 + W_2$. Questo dimostra che vale anche l'inclusione opposta e quindi l'uguaglianza. \square

Osservazione 1.3.21. Un risultato del tutto analogo vale anche per la somma di una famiglia finita di sottospazi vettoriali di V . Si ha cioè

$$W_1 + \cdots + W_n = \{w_1 + \cdots + w_n \mid w_i \in W_i, \text{ per } i = 1, \dots, n\}.$$

Nel caso invece di una famiglia infinita $\{W_i\}_{i \in I}$ di sottospazi vettoriali, è facile verificare che la somma

$$\sum_{i \in I} W_i$$

coincide con l'insieme di tutte le somme *finite* di vettori $w_i \in W_i$.

Definizione 1.3.22. La somma di due sottospazi vettoriali W_1 e W_2 di V si dice *diretta*, e si indica con $W_1 \oplus W_2$, se $W_1 \cap W_2 = \{\mathbf{0}\}$.

Più in generale, la somma di una famiglia qualsiasi $\{W_i\}_{i \in I}$ di sottospazi vettoriali di V si dice diretta se $W_i \cap W_j = \{\mathbf{0}\}$, per ogni $i, j \in I$ con $i \neq j$. La somma diretta di una famiglia $\{W_i\}_{i \in I}$ di sottospazi di V si indica con

$$\bigoplus_{i \in I} W_i.$$

Proposizione 1.3.23. *Ogni vettore $v \in W_1 \oplus W_2$ si scrive in modo unico nella forma $v = w_1 + w_2$, per qualche $w_1 \in W_1$ e qualche $w_2 \in W_2$ (un risultato analogo vale anche per una somma diretta di un numero qualunque di sottospazi di V).*

Dimostrazione. Nella proposizione precedente abbiamo visto che ogni $v \in W_1 \oplus W_2$ si può scrivere nella forma $v = w_1 + w_2$, per qualche $w_1 \in W_1$ e qualche $w_2 \in W_2$. Dobbiamo solo dimostrare che tale scrittura è unica.

Supponiamo che si abbia

$$v = w_1 + w_2 = w'_1 + w'_2,$$

con $w_1, w'_1 \in W_1$ e $w_2, w'_2 \in W_2$. Allora si ha

$$w_1 - w'_1 = w'_2 - w_2 \in W_1 \cap W_2.$$

Poiché la somma di W_1 e W_2 è diretta, si ha $W_1 \cap W_2 = \{\mathbf{0}\}$, quindi $w_1 - w'_1 = w'_2 - w_2 = \mathbf{0}$, da cui si deduce che $w_1 = w'_1$ e $w_2 = w'_2$. \square

Osservazione 1.3.24. La somma diretta di due sottospazi di uno spazio vettoriale V , definita in precedenza, è anche detta *somma diretta interna*. Ora vedremo come sia possibile definire anche la somma diretta di due spazi vettoriali V e W qualunque, in modo tale che V e W si possano poi identificare con due sottospazi vettoriali di $V \oplus W$. Una tale somma è detta *somma diretta esterna*.

Siano dunque V e W due spazi vettoriali sul campo K . Sul prodotto cartesiano $V \times W$ definiamo un'operazione di somma ponendo

$$(v_1, w_1) + (v_2, w_2) = (v_1 + v_2, w_1 + w_2),$$

e un'operazione di prodotto per un elemento di K ponendo

$$\lambda(v_1, w_1) = (\lambda v_1, \lambda w_1),$$

per ogni $(v_1, w_1), (v_2, w_2) \in V \times W$ e ogni $\lambda \in K$. È immediato verificare che queste operazioni definiscono una struttura di spazio vettoriale su $V \times W$. Indichiamo con $V \oplus W$ lo spazio vettoriale così ottenuto.

Le due funzioni $i_V : V \rightarrow V \oplus W$, $v \mapsto (v, \mathbf{0}_W)$ e $i_W : W \rightarrow V \oplus W$, $w \mapsto (\mathbf{0}_V, w)$ sono iniettive e permettono di identificare i due spazi vettoriali V e W con i due sottospazi vettoriali $i_V(V) = V \times \{\mathbf{0}_W\}$ e $i_W(W) = \{\mathbf{0}_V\} \times W$ di $V \oplus W$. Si verifica facilmente che lo spazio vettoriale $V \oplus W$ appena definito coincide con la somma diretta (interna) dei suoi due sottospazi $i_V(V)$ e $i_W(W)$.

Più in generale, se V_1, V_2, \dots, V_n sono una famiglia finita di spazi vettoriali sul campo K è possibile definire, in modo naturale, una struttura di spazio vettoriale sul prodotto cartesiano $V_1 \times V_2 \times \dots \times V_n$, ponendo

$$(v_1, v_2, \dots, v_n) + (w_1, w_2, \dots, w_n) = (v_1 + w_1, v_2 + w_2, \dots, v_n + w_n)$$

e

$$\lambda(v_1, v_2, \dots, v_n) = (\lambda v_1, \lambda v_2, \dots, \lambda v_n),$$

per ogni $\lambda \in K$ e ogni $(v_1, \dots, v_n), (w_1, \dots, w_n) \in V_1 \times \dots \times V_n$. Ogni V_i si identifica in modo naturale con il sottospazio V'_i del prodotto cartesiano $V_1 \times \dots \times V_n$ che consiste di tutti gli elementi del tipo $(0, \dots, 0, v, 0, \dots, 0)$, al variare di $v \in V_i$ (il vettore v si trova nella i -esima posizione). È ora immediato verificare che la somma diretta di tutti questi sottospazi V'_i coincide con il prodotto cartesiano $V_1 \times \dots \times V_n$. Si definisce pertanto la somma diretta esterna della famiglia di spazi vettoriali V_1, V_2, \dots, V_n ponendo

$$\bigoplus_{i=1}^n V_i = \prod_{i=1}^n V_i.$$

In conclusione, possiamo riassumere quanto visto finora, dicendo che, nel caso di una famiglia *finita* di spazi vettoriali, la somma diretta coincide con il prodotto cartesiano. Come vedremo in seguito, tale uguaglianza non vale nel caso della somma diretta di una famiglia di infiniti spazi vettoriali.

1.3.2 Insiemi di generatori e basi

Sia V uno spazio vettoriale su un campo K .

Definizione 1.3.25. Un sottoinsieme $S \subseteq V$ è detto un *insieme di generatori di V* se $L(S) = V$. In tal caso si dice anche che S *genera* V .

Notiamo che ogni spazio vettoriale possiede dei sistemi di generatori: l'intero spazio V è banalmente un insieme di generatori di V .

Dalla Proposizione 1.3.17 segue che, se S è un insieme di generatori di V , ogni vettore $v \in V$ si può scrivere come combinazione lineare finita di elementi di S :

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n,$$

per qualche $v_1, \dots, v_n \in S$ e $\lambda_1, \dots, \lambda_n \in K$. Una tale espressione non è però, in generale, unica.

Definizione 1.3.26. Un sottoinsieme $S \subseteq V$ è detto un *insieme libero* di vettori se esso ha la seguente proprietà: una combinazione lineare finita di elementi di S è il vettore nullo se e solo se tutti i coefficienti λ_i sono nulli. Cioè

$$\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n = \mathbf{0}$$

con $v_1, \dots, v_n \in S$, implica $\lambda_1 = \lambda_2 = \dots = \lambda_n = 0$.

Se $S = \{v_1, v_2, \dots, v_n\}$ è un insieme libero, diremo anche che i vettori v_1, v_2, \dots, v_n sono *linearmente indipendenti*.

Quindi i vettori $v_1, v_2, \dots, v_n \in V$ sono linearmente indipendenti se e solo se l'equazione

$$\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n = \mathbf{0}$$

ha come unica soluzione $\lambda_1 = \lambda_2 = \dots = \lambda_n = 0$.

Osservazione 1.3.27. Se S è l'insieme costituito da un unico vettore v , dire che S è libero equivale a dire che $v \neq \mathbf{0}$.

Analogamente, si trova che se i vettori v_1, v_2, \dots, v_n sono linearmente indipendenti, essi devono essere tutti diversi da zero.

Definizione 1.3.28. I vettori $v_1, v_2, \dots, v_n \in V$ si dicono *linearmente dipendenti* se essi non sono linearmente indipendenti, cioè se esistono degli scalari $\lambda_1, \lambda_2, \dots, \lambda_n \in K$, non tutti nulli, per cui si abbia

$$\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n = \mathbf{0}.$$

Proposizione 1.3.29. I vettori $v_1, v_2, \dots, v_n \in V$ sono linearmente dipendenti se e solo se uno di essi può essere espresso come combinazione lineare dei rimanenti, cioè se e solo se esiste un indice i tale che si abbia

$$v_i = \sum_{j=1, j \neq i}^n \alpha_j v_j,$$

con $\alpha_j \in K$.

Dimostrazione. Se i vettori v_1, v_2, \dots, v_n sono linearmente dipendenti, esiste una combinazione lineare

$$\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n = \mathbf{0}$$

in cui i coefficienti λ_j non sono tutti nulli. Sia dunque i un indice tale che $\lambda_i \neq 0$. Possiamo quindi scrivere

$$\lambda_i v_i = -\lambda_1 v_1 - \dots - \lambda_{i-1} v_{i-1} - \lambda_{i+1} v_{i+1} - \dots - \lambda_n v_n,$$

da cui si ricava

$$v_i = -\frac{\lambda_1}{\lambda_i} v_1 - \dots - \frac{\lambda_{i-1}}{\lambda_i} v_{i-1} - \frac{\lambda_{i+1}}{\lambda_i} v_{i+1} - \dots - \frac{\lambda_n}{\lambda_i} v_n.$$

Viceversa, supponiamo che un vettore v_i sia combinazione lineare dei rimanenti, cioè che si abbia

$$v_i = \alpha_1 v_1 + \dots + \alpha_{i-1} v_{i-1} + \alpha_{i+1} v_{i+1} + \dots + \alpha_n v_n.$$

Allora si ha

$$\alpha_1 v_1 + \cdots + \alpha_{i-1} v_{i-1} - v_i + \alpha_{i+1} v_{i+1} + \cdots + \alpha_n v_n = \mathbf{0},$$

il che dimostra che i vettori v_1, \dots, v_n sono linearmente dipendenti. \square

Osservazione 1.3.30. Notiamo che la dimostrazione della proposizione precedente dipende in modo essenziale dalla possibilità di poter dividere per un elemento non nullo $\lambda_i \in K$; è pertanto indispensabile che K sia un campo. Nel caso in cui V sia un modulo su un anello un analogo risultato non vale, come illustrato dal seguente esempio.

Sia $K = \mathbb{Z}$ e $V = \mathbb{Z}^2$. Consideriamo i tre elementi $u = (1, 2)$, $v = (2, 1)$ e $w = (3, 4)$. Essi sono linearmente dipendenti, infatti

$$5(1, 2) + 2(2, 1) - 3(3, 4) = 0,$$

tuttavia è facile verificare che nessuno di essi può essere espresso come combinazione lineare degli altri due.

Dimostriamo ora che, se un vettore si può scrivere come combinazione lineare di un insieme di vettori linearmente indipendenti, tale espressione è unica.

Proposizione 1.3.31. *Siano $v_1, v_2, \dots, v_n \in V$ dei vettori linearmente indipendenti. Se $v \in V$ si può scrivere come combinazione lineare dei vettori v_1, v_2, \dots, v_n ,*

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n,$$

allora gli scalari $\lambda_1, \lambda_2, \dots, \lambda_n$ sono determinati in modo unico.

Dimostrazione. Supponiamo che sia possibile scrivere v in due modi, come

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n,$$

e come

$$v = \mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_n v_n.$$

Allora si ha

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n = \mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_n v_n,$$

che si può riscrivere come

$$(\lambda_1 - \mu_1) v_1 + (\lambda_2 - \mu_2) v_2 + \cdots + (\lambda_n - \mu_n) v_n = \mathbf{0}.$$

Poiché i vettori v_1, v_2, \dots, v_n sono linearmente indipendenti, si ha dunque

$$\lambda_1 - \mu_1 = 0, \lambda_2 - \mu_2 = 0, \dots, \lambda_n - \mu_n = 0,$$

il che dimostra che $\lambda_i = \mu_i$, per ogni $i = 1, \dots, n$. \square

Dalla proposizione appena dimostrata discende quindi che, se consideriamo un insieme di generatori S di V con la proprietà aggiuntiva che i vettori di S siano linearmente indipendenti, allora ogni vettore di V si può scrivere, in modo *unico*, come combinazione lineare finita di elementi di S .

Definizione 1.3.32. Un insieme libero di generatori di uno spazio vettoriale V è detto una *base* di V . In altri termini, una base di V è un insieme di vettori linearmente indipendenti i quali generano l'intero spazio V .

Osservazione 1.3.33. Solitamente una base di uno spazio vettoriale V viene intesa come un insieme *ordinato* di generatori linearmente indipendenti di V . Ciò significa, ad esempio, che se l'insieme costituito dai vettori v_1 e v_2 è una base di uno spazio vettoriale V , allora $\mathbf{v} = \{v_1, v_2\}$ e $\mathbf{v}' = \{v_2, v_1\}$ sono due basi *diverse* di V .

Da quanto visto in precedenza, si deduce il seguente risultato:

Corollario 1.3.34. *Sia S una base di V . Ogni vettore $v \in V$ si può scrivere, in modo unico, come combinazione lineare finita di elementi di S .*

Osservazione 1.3.35. Supponiamo che $S = \{v_1, v_2, \dots, v_n\}$ sia una base di uno spazio vettoriale V . Allora, per ogni $v \in V$, si ha

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n,$$

e gli scalari $\lambda_i \in K$ sono unicamente determinati da v . Tali scalari sono anche detti le *coordinate* del vettore v rispetto alla base v_1, v_2, \dots, v_n fissata.

1.3.3 Spazi vettoriali finitamente generati

Nella sezione precedente non abbiamo fatto nessuna ipotesi sul numero di generatori di uno spazio vettoriale. Ora ci occuperemo in dettaglio del caso in cui tale numero è finito.

Definizione 1.3.36. Uno spazio vettoriale V è detto *finitamente generato* se esiste un insieme finito di generatori di V .

Cominciamo col dimostrare che ogni spazio vettoriale V finitamente generato ammette una base. Più precisamente, dimostreremo che da ogni insieme di generatori di V si può estrarre una base.

Proposizione 1.3.37. *Sia $S = \{v_1, v_2, \dots, v_n\}$ un insieme di generatori di V . Allora S contiene dei vettori $v_{i_1}, v_{i_2}, \dots, v_{i_r}$, per qualche $r \leq n$, che formano una base di V .*

Dimostrazione. Se i vettori v_1, v_2, \dots, v_n sono linearmente indipendenti, essi sono una base di V e la dimostrazione è così terminata. Se invece essi sono linearmente dipendenti, uno di essi può essere espresso come combinazione lineare dei rimanenti. A meno di rinominarli, possiamo supporre che questo vettore sia v_n . Possiamo quindi scrivere

$$v_n = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_{n-1} v_{n-1}.$$

Da ciò segue che i vettori v_1, v_2, \dots, v_{n-1} generano lo spazio vettoriale V ; infatti ogni vettore che si scrive come combinazione lineare dei vettori v_1, v_2, \dots, v_n si può anche scrivere come combinazione lineare dei soli vettori v_1, v_2, \dots, v_{n-1} . Ora, se i vettori v_1, v_2, \dots, v_{n-1} sono linearmente indipendenti, essi sono una base di V e la dimostrazione è terminata. In caso contrario uno di essi può essere espresso come combinazione lineare dei rimanenti. Anche in questo caso, a meno di riordinare i vettori, possiamo supporre che sia v_{n-1} a potersi scrivere come combinazione lineare dei vettori v_1, v_2, \dots, v_{n-2} . Ma ciò significa che i vettori v_1, v_2, \dots, v_{n-2} sono un insieme di generatori di V .

Ripetendo il ragionamento sopra descritto si arriverà, prima o poi, a un insieme di vettori v_1, v_2, \dots, v_r , per qualche $r \leq n$, che generano tutto lo spazio V e sono linearmente indipendenti. Essi costituiscono quindi una base di V . \square

Il seguente risultato chiarisce le relazioni che esistono tra insiemi di vettori linearmente indipendenti, basi e insiemi di generatori.

Proposizione 1.3.38. *Sia V uno spazio vettoriale finitamente generato. Consideriamo un insieme $\{v_1, v_2, \dots, v_n\}$ di generatori di V e siano w_1, w_2, \dots, w_r dei vettori linearmente indipendenti. Allora $r \leq n$.*

Dimostrazione. Poiché i vettori v_1, v_2, \dots, v_n generano V , il vettore w_1 si può scrivere come una loro combinazione lineare,

$$w_1 = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n.$$

Dato che $w_1 \neq 0$, gli scalari $\lambda_1, \dots, \lambda_n$ non possono essere tutti nulli, quindi esiste un indice i tale che $\lambda_i \neq 0$. Da ciò si deduce che il vettore v_i può essere espresso come combinazione lineare dei vettori $w_1, v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n$.

A meno di rinominare i vettori v_j , possiamo supporre che sia $i = n$, cioè che v_n si possa scrivere come combinazione lineare dei vettori w_1, v_1, \dots, v_{n-1} ; ma ciò significa che anche $\{w_1, v_1, \dots, v_{n-1}\}$ è un insieme di generatori di V . Il vettore w_2 si può quindi scrivere come combinazione lineare dei vettori w_1, v_1, \dots, v_{n-1} :

$$w_2 = \alpha_1 w_1 + \lambda_1 v_1 + \dots + \lambda_{n-1} v_{n-1},$$

e gli scalari $\lambda_1, \dots, \lambda_{n-1}$ non possono essere tutti nulli, perché altrimenti i vettori w_1 e w_2 sarebbero linearmente dipendenti, contro l'ipotesi.

Esiste quindi un indice i per il quale $\lambda_i \neq 0$ e, ancora una volta, possiamo supporre che sia $i = n - 1$ (a meno di riordinare i vettori v_j). Da ciò segue che il vettore v_{n-1} si può scrivere come combinazione lineare dei vettori $w_1, w_2, v_1, \dots, v_{n-2}$, quindi anche $\{w_1, w_2, v_1, \dots, v_{n-2}\}$ è un insieme di generatori di V .

Continuando in questo modo, si dimostra che tutti gli insiemi

$$\{w_1, w_2, \dots, w_h, v_1, \dots, v_{n-h}\}$$

sono insiemi di generatori di V .

Se, per assurdo, fosse $n < r$, ponendo $h = n$ si avrebbe che i vettori w_1, w_2, \dots, w_n generano tutto lo spazio V , quindi il vettore w_{n+1} si potrebbe scrivere come combinazione lineare dei vettori w_1, w_2, \dots, w_n , il che contraddice l'ipotesi che i vettori w_1, w_2, \dots, w_r siano linearmente indipendenti. Deve quindi essere $r \leq n$. \square

Corollario 1.3.39. *Sia V uno spazio vettoriale finitamente generato e sia $\{v_1, v_2, \dots, v_n\}$ una base di V . Allora, per ogni insieme di vettori linearmente indipendenti $\{w_1, w_2, \dots, w_r\}$, si ha $r \leq n$ e, per ogni insieme $\{u_1, u_2, \dots, u_s\}$ di generatori di V , si ha $s \geq n$.*

Dimostrazione. Questo risultato è una conseguenza immediata della proposizione precedente; basta ricordare che i vettori v_1, v_2, \dots, v_n sono linearmente indipendenti e sono anche un insieme di generatori di V . \square

Corollario 1.3.40. *Due basi qualunque di uno spazio vettoriale V (finitamente generato) hanno lo stesso numero di elementi.*

Dimostrazione. Siano $\{v_1, v_2, \dots, v_r\}$ e $\{w_1, w_2, \dots, w_s\}$ due basi di V . Allora, dato che i vettori v_1, v_2, \dots, v_r sono linearmente indipendenti e i vettori w_1, w_2, \dots, w_s sono dei generatori di V , si ha $r \leq s$. Scambiando il ruolo delle due basi, si ottiene anche $s \leq r$, da cui segue l'uguaglianza $r = s$. \square

Il numero di vettori che compongono una base di uno spazio vettoriale finitamente generato V è dunque indipendente dalla base scelta e dipende quindi solo dallo spazio V . Possiamo pertanto dare la seguente definizione:

Definizione 1.3.41. La *dimensione* di uno spazio vettoriale (finitamente generato) V , indicata con $\dim V$, è il numero di elementi di una base di V .

Esempio 1.3.42. Consideriamo lo spazio vettoriale $V = K^n$, definito nell'Esempio 1.3.3. Per ogni $i = 1, \dots, n$, indichiamo con e_i la n -upla di elementi di K le cui componenti sono tutte nulle tranne la i -esima, che è uguale a 1:

$$\begin{aligned} e_1 &= (1, 0, 0, 0, \dots, 0, 0), \\ e_2 &= (0, 1, 0, 0, \dots, 0, 0), \\ e_3 &= (0, 0, 1, 0, \dots, 0, 0), \\ &\quad \dots \quad \dots \quad \dots \\ e_n &= (0, 0, 0, 0, \dots, 0, 1). \end{aligned}$$

Notiamo che, per ogni $\lambda_1, \dots, \lambda_n \in K$, si ha

$$\lambda_1 e_1 + \lambda_2 e_2 + \dots + \lambda_n e_n = (\lambda_1, \lambda_2, \dots, \lambda_n).$$

Da questa uguaglianza si deduce che i vettori e_1, e_2, \dots, e_n sono linearmente indipendenti e generano lo spazio vettoriale V ; essi sono pertanto una base di $V = K^n$. Questa base è detta la *base canonica* di K^n . Si ha pertanto $\dim K^n = n$.

Esempio 1.3.43. Lo spazio vettoriale nullo, $V = \{\mathbf{0}\}$, ha dimensione pari a zero. Esso infatti contiene un solo vettore $v = \mathbf{0}$, ma tale vettore non forma una base di V dato che esso non è linearmente indipendente! Infatti un insieme costituito da un solo vettore v è un insieme libero (cioè v è linearmente indipendente) se e solo se $v \neq \mathbf{0}$.

Dalla Proposizione 1.3.38 derivano anche i prossimi due risultati.

Corollario 1.3.44. *Sia V uno spazio vettoriale finitamente generato. Allora ogni sottospazio vettoriale W di V è finitamente generato e si ha $\dim W \leq \dim V$.*

Dimostrazione. Poniamo $n = \dim V$. Se $\{w_1, \dots, w_r\}$ è un insieme di vettori linearmente indipendenti di W , essi sono anche dei vettori linearmente indipendenti di V ; deve quindi essere $r \leq n$. Se questi vettori non sono un insieme di generatori di W , ciò significa che esiste un vettore $w_{r+1} \in W$ che non può essere espresso come combinazione lineare di w_1, \dots, w_r . Da ciò segue che i vettori w_1, \dots, w_r, w_{r+1} sono linearmente indipendenti. Se essi non sono ancora un insieme di generatori di W , deve esistere un vettore $w_{r+2} \in W$ che non può essere espresso come combinazione lineare di w_1, \dots, w_r, w_{r+1} . Ma allora anche i vettori $w_1, \dots, w_r, w_{r+1}, w_{r+2}$ sono linearmente indipendenti.

Poiché il numero di vettori linearmente indipendenti non può eccedere n , ripetendo il ragionamento precedente si arriva, dopo un numero finito di passi, a costruire un insieme di vettori linearmente indipendenti $\{w_1, \dots, w_s\}$, con $s \leq n$, i quali generano il sottospazio W e sono quindi una base di W . Ciò dimostra che $\dim W \leq \dim V$. \square

Corollario 1.3.45. *Sia V uno spazio vettoriale finitamente generato. Allora ogni insieme di vettori linearmente indipendenti v_1, \dots, v_r può essere completato a una base di V . In altri termini, esistono dei vettori v_{r+1}, \dots, v_n tali che l'insieme $\{v_1, \dots, v_r, v_{r+1}, \dots, v_n\}$ sia una base di V .*

Dimostrazione. La dimostrazione di questo risultato è essenzialmente analoga a quella del corollario precedente. Supponiamo che $v_1, \dots, v_r \in V$ siano dei vettori linearmente indipendenti. Se questi vettori non sono un insieme di generatori di V , ciò significa che esiste un vettore $v_{r+1} \in V$ che non può essere espresso come combinazione lineare di v_1, \dots, v_r . Da ciò segue che i vettori v_1, \dots, v_r, v_{r+1} sono linearmente indipendenti. Se essi non sono ancora un insieme di generatori di V , deve esistere un vettore $v_{r+2} \in V$ che non può essere espresso come combinazione lineare di v_1, \dots, v_r, v_{r+1} . Ma allora anche i vettori $v_1, \dots, v_r, v_{r+1}, v_{r+2}$ sono linearmente indipendenti. Continuando in questo modo, si deve necessariamente ottenere un insieme di vettori linearmente indipendenti $\{v_1, \dots, v_r, v_{r+1}, \dots, v_n\}$ che sono anche un insieme di generatori di V , altrimenti si otterrebbe un insieme infinito di vettori linearmente indipendenti, contro l'ipotesi che V sia finitamente generato. \square

Se la dimensione di uno spazio vettoriale V è nota, la verifica che un determinato insieme di vettori di V forma una base risulta semplificata. Vale infatti il seguente risultato:

Proposizione 1.3.46. *Sia V uno spazio vettoriale di dimensione n e siano v_1, \dots, v_n dei vettori di V .*

- (i) *Se i vettori v_1, \dots, v_n sono linearmente indipendenti, allora essi sono anche un sistema di generatori di V , quindi sono una base di V .*
- (ii) *Se i vettori v_1, \dots, v_n sono un sistema di generatori di V , allora essi sono anche linearmente indipendenti, quindi sono una base di V .*

Dimostrazione. (i) Supponiamo che i vettori v_1, \dots, v_n siano linearmente indipendenti. Per il corollario precedente, essi sono contenuti in una base

$$\{v_1, \dots, v_n, v_{n+1}, \dots, v_{n+r}\}$$

di V . Ma, poiché V ha dimensione n , ogni base di V deve avere n elementi. Da ciò si deduce che $r = 0$ e quindi i vettori v_1, \dots, v_n sono, in effetti, una base di V .

(ii) Supponiamo che i vettori v_1, \dots, v_n siano un insieme di generatori di V . Se, per assurdo, essi fossero linearmente dipendenti, uno di essi sarebbe combinazione lineare dei rimanenti. A meno di riordinare i vettori, non è restrittivo supporre che v_n sia combinazione lineare di v_1, \dots, v_{n-1} . Ma allora i vettori v_1, \dots, v_{n-1} sarebbero anch'essi un insieme di generatori di V . Questo però è assurdo; infatti la cardinalità di un insieme di generatori di V deve essere $\geq \dim V$ (vedi Corollario 1.3.39). Quindi i vettori v_1, \dots, v_n sono linearmente indipendenti, cioè sono una base di V . \square

Corollario 1.3.47. *Sia V uno spazio vettoriale finitamente generato e sia $W \subset V$ un suo sottospazio proprio. Allora $\dim W < \dim V$.*

Dimostrazione. Nel Corollario 1.3.44 abbiamo dimostrato che si ha $\dim W \leq \dim V$. Supponiamo, per assurdo, che si abbia $\dim W = \dim V = n$. Consideriamo quindi una base w_1, \dots, w_n di W . Dato che questi sono n vettori linearmente indipendenti di V , e dato che n è proprio la dimensione di V , per il punto (i) della proposizione precedente essi sono una base di V . Ma da ciò segue che $W = V$, contro l'ipotesi che W sia un sottospazio proprio di V . \square

Osservazione 1.3.48. Se $K \subset L$ è una estensione⁴ di campi, ogni spazio vettoriale V sul campo L può essere anche considerato come spazio vettoriale sul campo K . Se $\{v_1, \dots, v_n\}$ è una base di V in quanto K -spazio vettoriale, questi stessi vettori generano V anche in quanto spazio vettoriale su L , tuttavia, in questo caso, essi potrebbero non essere più linearmente indipendenti, come vedremo nel successivo esempio. In generale, possiamo pertanto affermare che la dimensione di V in quanto L -spazio vettoriale è minore o uguale alla dimensione di V considerato come spazio vettoriale su K , cioè

$$\dim_L V \leq \dim_K V.$$

Illustriamo quanto appena affermato con un esempio concreto. Sia $K = \mathbb{R}$ il campo dei numeri reali e $L = \mathbb{C}$ il campo dei numeri complessi. L'insieme \mathbb{C} è identificato in modo naturale con l'insieme \mathbb{R}^2 , associando ad ogni numero complesso $z = x + iy$ la coppia di numeri reali (x, y) . Poniamo quindi $V = \mathbb{C} \cong \mathbb{R}^2$. Lo spazio vettoriale $V = \mathbb{C}$, in quanto spazio vettoriale sul campo \mathbb{C} , ha naturalmente dimensione 1, e una sua base è costituita dal vettore $v = 1$. Se invece consideriamo \mathbb{C} in quanto spazio vettoriale su \mathbb{R} , esso ha dimensione 2. Una sua base è infatti costituita dai vettori $v_1 = 1$ e $v_2 = i$, i quali corrispondono, nell'identificazione $\mathbb{C} \cong \mathbb{R}^2$ descritta in precedenza, ai due vettori $(1, 0)$ e $(0, 1)$ della base canonica di \mathbb{R}^2 .

Notiamo che i vettori $v_1 = 1$ e $v_2 = i$ sono linearmente indipendenti sul campo \mathbb{R} dei numeri reali; infatti se $\alpha v_1 + \beta v_2 = \alpha + i\beta = 0$, con $\alpha, \beta \in \mathbb{R}$, si deve necessariamente avere $\alpha = \beta = 0$. Essi sono invece linearmente dipendenti sul campo \mathbb{C} : si ha infatti

$$v_1 + iv_2 = 1 + i^2 = 1 - 1 = 0.$$

Più in generale, se $V = \mathbb{C}^n$, si ha $\dim_{\mathbb{C}} V = n$ e $\dim_{\mathbb{R}} V = 2n$.

Terminiamo questa sezione dimostrando un risultato che mette in relazione le dimensioni di due sottospazi vettoriali di V con le dimensioni della loro somma e della loro intersezione:

Proposizione 1.3.49 (FORMULA DI GRASSMANN). *Siano W_1 e W_2 due sottospazi vettoriali di uno spazio vettoriale finitamente generato V . Si ha:*

$$\dim(W_1 + W_2) = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2).$$

Dimostrazione. Poniamo

$$r = \dim W_1, \quad s = \dim W_2, \quad t = \dim(W_1 \cap W_2).$$

⁴Ciò significa semplicemente che L è un campo e K è un suo sottocampo.

Consideriamo una base $\{v_1, \dots, v_t\}$ di $W_1 \cap W_2$. Per il Corollario 1.3.45, questo insieme di vettori può essere completato, tramite aggiunta di altri vettori, in modo da ottenere una base $\{v_1, \dots, v_t, v_{t+1}, \dots, v_r\}$ di W_1 e una base $\{v_1, \dots, v_t, v'_{t+1}, \dots, v'_s\}$ di W_2 .

Dato che ogni vettore di $W_1 + W_2$ può essere scritto come somma di un vettore di W_1 e di uno di W_2 , esso può quindi essere espresso come combinazione lineare dei vettori $v_1, \dots, v_t, v_{t+1}, \dots, v_r, v'_{t+1}, \dots, v'_s$. Vogliamo ora dimostrare che questi vettori, oltre a essere dei generatori di $W_1 + W_2$, sono anche linearmente indipendenti.

Supponiamo quindi che sia

$$\alpha_1 v_1 + \dots + \alpha_t v_t + \beta_1 v_{t+1} + \dots + \beta_{r-t} v_r + \gamma_1 v'_{t+1} + \dots + \gamma_{s-t} v'_s = \mathbf{0}.$$

Da ciò segue che

$$\alpha_1 v_1 + \dots + \alpha_t v_t + \beta_1 v_{t+1} + \dots + \beta_{r-t} v_r = -(\gamma_1 v'_{t+1} + \dots + \gamma_{s-t} v'_s).$$

Se chiamiamo w il vettore precedente, si ha che $w \in W_1 \cap W_2$. Poiché $\{v_1, \dots, v_t\}$ è una base di $W_1 \cap W_2$, il vettore w si può scrivere, in modo unico, nella forma

$$w = \lambda_1 v_1 + \dots + \lambda_t v_t.$$

Si hanno quindi le seguenti uguaglianze:

$$\lambda_1 v_1 + \dots + \lambda_t v_t = \alpha_1 v_1 + \dots + \alpha_t v_t + \beta_1 v_{t+1} + \dots + \beta_{r-t} v_r$$

e

$$\lambda_1 v_1 + \dots + \lambda_t v_t = -(\gamma_1 v'_{t+1} + \dots + \gamma_{s-t} v'_s).$$

Poiché, per ipotesi, i vettori $v_1, \dots, v_t, v_{t+1}, \dots, v_r$ sono una base di W_1 e i vettori $v_1, \dots, v_t, v'_{t+1}, \dots, v'_s$ sono una base di W_2 , dalle uguaglianze precedenti segue che

$$\lambda_1 = \dots = \lambda_t = 0, \quad \alpha_1 = \dots = \alpha_t = 0,$$

$$\beta_1 = \dots = \beta_{r-t} = 0, \quad \gamma_1 = \dots = \gamma_{s-t} = 0.$$

Abbiamo così dimostrato che l'insieme dei vettori

$$\{v_1, \dots, v_t, v_{t+1}, \dots, v_r, v'_{t+1}, \dots, v'_s\}$$

è una base di $W_1 + W_2$. Si ha pertanto

$$\dim(W_1 + W_2) = r + s - t = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2). \quad \square$$

Osservazione 1.3.50. Sia V uno spazio vettoriale finitamente generato. Notiamo che, per ogni sottospazio $W \subseteq V$, è possibile trovare un sottospazio W' di V tale che $V = W \oplus W'$, cioè tale che si abbia $V = W + W'$ e $W \cap W' = \{\mathbf{0}\}$. Un tale W' è detto un *sottospazio complementare* di W .

A tal fine, è sufficiente considerare una base $\{w_1, \dots, w_r\}$ di W e completarla a una base $\{w_1, \dots, w_r, v_{r+1}, \dots, v_n\}$ di V (cf. Corollario 1.3.45). Il sottospazio $W' = \langle v_{r+1}, \dots, v_n \rangle$, generato dai vettori v_{r+1}, \dots, v_n , è il sottospazio cercato.

Notiamo infine che un tale sottospazio W' non è unico. Ad esempio, se $V = K^2$ e se W è il sottospazio generato dal vettore $(1, 0)$, qualunque vettore (a, b) , con $b \neq 0$, genera un sottospazio W' tale che $V = W \oplus W'$.

Esercizi

Esercizio 1.3.1. Si dica se gli insiemi seguenti sono degli spazi vettoriali:

- (1) L'insieme delle funzioni reali definite nell'intervallo $[0, 1]$, continue, positive o nulle, per le operazioni di addizione e di prodotto per un numero reale.
- (2) L'insieme delle funzioni reali f definite in \mathbb{R} , tali che

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

per le operazioni di addizione e di prodotto per un numero reale.

- (3) L'insieme $A = \{x \in \mathbb{R} \mid x > 0\}$, per le operazioni di somma e di prodotto per uno scalare definite rispettivamente da

$$\begin{aligned} x \oplus y &= xy, \quad \forall x, y \in A \\ \lambda \cdot x &= x^\lambda, \quad \forall x \in A, \lambda \in \mathbb{R}. \end{aligned}$$

- (4) L'insieme delle funzioni da \mathbb{R} in \mathbb{R} che si annullano in 1 oppure in 4.
- (5) L'insieme dei polinomi di grado uguale a n (n intero positivo).
- (6) L'insieme delle funzioni da \mathbb{R} in \mathbb{R} , di classe C^2 , tali che

$$f'' + \omega^2 f = 0,$$

con $\omega \in \mathbb{R}$.

- (7) L'insieme delle funzioni reali $f(x)$ definite nell'intervallo $[0, 1]$, continue, tali che

$$\int_0^1 f(x) \sin x \, dx = 0.$$

Esercizio 1.3.2. Sia V lo spazio vettoriale dei polinomi, a coefficienti reali nell'indeterminata x , di grado ≤ 3 . Si verifichi che gli insiemi seguenti sono delle basi di V :

- (1) $\{1, x, x^2, x^3\}$;
- (2) $\{1, 1-x, x-x^2, x^2-x^3\}$;
- (3) $\{1, 1+x, 1+x+x^2, 1+x+x^2+x^3\}$.

Esercizio 1.3.3. Nello spazio vettoriale V dei polinomi di grado ≤ 2 si considerino i polinomi

$$\begin{aligned} p_1(x) &= x^2 + x(1-x) + (1-x)^2 \\ p_2(x) &= x^2 + (1-x)^2 \\ p_3(x) &= x^2 + 1 + (1-x)^2 \\ p_4(x) &= x(1-x). \end{aligned}$$

È possibile estrarre da $\{p_1(x), p_2(x), p_3(x), p_4(x)\}$ delle basi di V ? In caso affermativo, trovarle tutte.

Esercizio 1.3.4. Nello spazio vettoriale delle funzioni continue da \mathbb{R} in \mathbb{R} , si considerino le funzioni $f_1(x) = \sin x$, $f_2(x) = \sin 2x$ e $f_3(x) = \sin 3x$. Si dica se queste funzioni sono linearmente indipendenti.

Esercizio 1.3.5. Si dica se, nei casi seguenti, i vettori v_1 , v_2 e v_3 costituiscono una base di \mathbb{R}^3 . In caso negativo si descriva il sottospazio da essi generato.

- (1) $v_1 = (1, 1, 1)$, $v_2 = (3, 0, -1)$, $v_3 = (-1, 1, -1)$;
- (2) $v_1 = (1, 2, 3)$, $v_2 = (3, 0, -1)$, $v_3 = (1, 8, 13)$;
- (3) $v_1 = (1, 2, -3)$, $v_2 = (1, 0, -1)$, $v_3 = (1, 10, -11)$.

Esercizio 1.3.6. In \mathbb{R}^4 i vettori seguenti formano:

(a) un insieme libero (cioè un insieme di vettori linearmente indipendenti)? In caso affermativo, completarlo per ottenere una base di \mathbb{R}^4 , altrimenti determinare le relazioni di dipendenza lineare tra di loro ed estrarre da questo insieme di vettori almeno un insieme libero.

(b) un insieme di generatori? In caso affermativo, estrarne almeno una base di \mathbb{R}^4 , altrimenti determinare la dimensione del sottospazio da essi generato.

$$(1) \quad v_1 = (1, 1, 1, 1), \quad v_2 = (0, 1, 2, -1), \quad v_3 = (1, 0, -2, 3), \\ v_4 = (2, 1, 0, -1), \quad v_5 = (4, 3, 2, 1);$$

$$(2) \quad v_1 = (1, 2, 3, 4), \quad v_2 = (0, 1, 2, -1), \quad v_3 = (3, 4, 5, 16);$$

$$(3) \quad v_1 = (1, 2, 3, 4), \quad v_2 = (0, 1, 2, -1), \quad v_3 = (2, 1, 0, 11), \\ v_4 = (3, 4, 5, 14).$$

Esercizio 1.3.7. Si determini una base del sottospazio vettoriale V di \mathbb{R}^5 costituito dai vettori (x_1, \dots, x_5) che sono soluzioni del seguente sistema di equazioni lineari:

$$\begin{cases} x_1 - 3x_2 + x_4 = 0 \\ x_2 + 3x_3 - x_5 = 0 \\ x_1 + 2x_2 + x_3 - x_4 = 0. \end{cases}$$

Esercizio 1.3.8. In \mathbb{R}^4 siano $v_1 = (1, 2, 3, 4)$ e $v_2 = (1, -2, 3, -4)$. È possibile determinare due numeri reali x e y in modo tale che $(x, 1, y, 1) \in L\{v_1, v_2\}$? (Ricordiamo che $L\{v_1, v_2\}$ indica il sottospazio generato dai vettori v_1 e v_2 .)

Esercizio 1.3.9. Sia V uno spazio vettoriale. Si dica se le affermazioni seguenti sono vere o false.

- (1) Se i vettori v_1 , v_2 e v_3 sono a due a due non proporzionali allora la famiglia $\{v_1, v_2, v_3\}$ è libera.
- (2) Se nessuno fra i vettori v_1, \dots, v_r è combinazione lineare dei vettori rimanenti allora la famiglia $\{v_1, \dots, v_r\}$ è libera.

Esercizio 1.3.10. In \mathbb{R}^4 siano $v_1 = (0, 1, -2, 1)$, $v_2 = (1, 0, 2, -1)$, $v_3 = (3, 2, 2, -1)$, $v_4 = (0, 0, 1, 0)$, $v_5 = (0, 0, 0, 1)$. Si dica se le affermazioni seguenti sono vere o false.

- (1) $L\{v_1, v_2, v_3\} = L\{(1, 1, 0, 0), (-1, 1, -4, 2)\};$
- (2) $(1, 1, 0, 0) \in L\{v_1, v_2\} \cap L\{v_2, v_3, v_4\};$
- (3) $\dim(L\{v_1, v_2\} \cap L\{v_2, v_3, v_4\}) = 1;$
- (4) $L\{v_1, v_2\} + L\{v_2, v_3, v_4\} = \mathbb{R}^4;$
- (5) $L\{v_1, v_2, v_3\} + L\{v_4, v_5\} = \mathbb{R}^4.$

Esercizio 1.3.11. Si studi la dipendenza o l'indipendenza lineare dei vettori seguenti, e si determini in ogni caso una base del sottospazio da essi generato.

- (1) $(1, 0, 1), (0, 2, 2), (3, 7, 1)$, in \mathbb{R}^3 ;
- (2) $(1, 0, 0), (0, 1, 1), (1, 1, 1)$, in \mathbb{R}^3 ;
- (3) $(1, 2, 1, 2, 1), (2, 1, 2, 1, 2), (1, 0, 1, 1, 0), (0, 1, 0, 0, 1)$, in \mathbb{R}^5 .

Esercizio 1.3.12. Sia V lo spazio vettoriale dei polinomi in x , a coefficienti in \mathbb{R} , di grado $\leq n$, con n intero positivo. Si dimostri che, per ogni $a \in \mathbb{R}$, l'insieme

$$\{1, x - a, (x - a)^2, \dots, (x - a)^n\}$$

è una base di V . Sia poi $f(x) \in V$; si esprima $f(x)$ come combinazione lineare dei precedenti polinomi. Chi sono i coefficienti di tale combinazione lineare?

Esercizio 1.3.13. Siano $U_t = L\{u_1, u_2\}$ e $V_t = L\{v_1, v_2\}$ due sottospazi di \mathbb{R}^4 , con $u_1 = (1, t, 2t, 0)$, $u_2 = (t, t, t, t)$, $v_1 = (t - 2, -t, -3t, t)$ e $v_2 = (2, t, 2t, 0)$.

- (1) Si dica se esiste $t \in \mathbb{R}$ tale che $U_t + V_t = \mathbb{R}^4$.
- (2) Per quali $t \in \mathbb{R}$ si ha $\dim(U_t \cap V_t) = 1$?
- (3) Si determini una base di $U_1 \cap V_1$ e la si estenda a una base di \mathbb{R}^4 .

Esercizio 1.3.14. In \mathbb{R}^4 si considerino i sottospazi $U = L\{v_1, v_2, v_3\}$ e $V = L\{v_4, v_5\}$, dove $v_1 = (1, 2, 3, 4)$, $v_2 = (2, 2, 2, 6)$, $v_3 = (0, 2, 4, 4)$, $v_4 = (1, 0, -1, 2)$ e $v_5 = (2, 3, 0, 1)$. Si determinino delle basi dei sottospazi $U \cap V$, U , V e $U + V$.

Esercizio 1.3.15. Siano U e W due sottospazi vettoriali di uno spazio vettoriale V . Si dimostri che $U \cup W$ è un sottospazio vettoriale di V se e solo se $U \subset W$ oppure $W \subset U$.

Esercizio 1.3.16. Siano U , V e W tre sottospazi di uno stesso spazio vettoriale. Si dica se è vero o falso che

$$U \cap (V + W) = (U \cap V) + (U \cap W).$$

Esercizio 1.3.17. Si dica se è diretta la somma dei due seguenti sottospazi di \mathbb{R}^4 : $U = L\{(1, 0, 1, 0), (1, 2, 3, 4)\}$ e $V = L\{(0, 1, 1, 1), (0, 0, 0, 1)\}$.

Esercizio 1.3.18. Si considerino i seguenti sottospazi di \mathbb{R}^4 :

$$U = L\{(1, 0, 1, 0), (0, 1, 1, 1), (0, 0, 0, 1)\}$$

e

$$V = L\{(1, 0, 1, 0), (0, 1, 1, 0)\}.$$

Si determini un sottospazio $W \subset \mathbb{R}^4$ tale che $U = V \oplus W$, e si dica se tale W è unico.

Esercizio 1.3.19. Dati i seguenti sottospazi di \mathbb{R}^4 ,

$$U = L\{(1, 0, 1, 0), (0, 0, 0, 1)\}$$

e

$$V = L\{(1, 0, 2, 0), (0, 0, 1, 1)\},$$

esiste un sottospazio $W \subset \mathbb{R}^4$ tale che $U \oplus W = V \oplus W = \mathbb{R}^4$?

In caso affermativo si determini W e si dica se è unico.

Esercizio 1.3.20. Nello spazio vettoriale V dei polinomi, nell'indeterminata x a coefficienti reali, di grado ≤ 5 , si considerino i sottospazi seguenti:

$$\begin{aligned} U_1 &= \{p(x) \in V \mid p(0) = 0\}, \\ U_2 &= \{p(x) \in V \mid p'(1) = 0\}, \\ U_3 &= \{p(x) \in V \mid x^2 + 1 \text{ divide } p(x)\}, \\ U_4 &= \{p(x) \in V \mid p(-x) = p(x), \forall x\}, \\ U_5 &= \{p(x) \in V \mid p(x) = xp'(x), \forall x\}. \end{aligned}$$

- (1) Si determinino delle basi dei seguenti sottospazi:

$$U_1, U_2, U_3, U_4, U_5, U_1 \cap U_2, U_1 \cap U_3, U_1 \cap U_2 \cap U_3, U_1 \cap U_2 \cap U_3 \cap U_4.$$

- (2) Si determinino dei sottospazi W_1 e W_2 di V tali che

$$W_1 \oplus U_4 = W_2 \oplus (U_1 \cap U_3) = V.$$

Capitolo 2

Applicazioni Lineari e Matrici

In questo capitolo svilupperemo la teoria delle funzioni lineari tra due spazi vettoriali. Introdurremo il concetto di matrice, descriveremo il legame esistente tra matrici e funzioni lineari e dimostreremo i principali risultati della teoria delle matrici. Infine, utilizzeremo i risultati così ottenuti per descrivere la teoria dei sistemi di equazioni lineari.

2.1 Applicazioni lineari

In questo capitolo ci occuperemo dello studio delle funzioni, definite tra due spazi vettoriali, che “rispettano” la struttura di spazio vettoriale, cioè che sono compatibili con le operazioni di somma di vettori e di prodotto di un vettore per uno scalare.

Definizione 2.1.1. Siano V e W due spazi vettoriali su un campo K . Una funzione $f : V \rightarrow W$ è detta *additiva* se

$$f(v_1 + v_2) = f(v_1) + f(v_2),$$

per ogni $v_1, v_2 \in V$.

Una tale funzione è detta *K -lineare* (o, più semplicemente, *lineare*) se, oltre ad essere additiva, essa soddisfa la seguente uguaglianza:

$$f(\lambda v) = \lambda f(v),$$

per ogni $v \in V$ e ogni $\lambda \in K$.

Una funzione lineare tra due spazi vettoriali è anche detta un *omomorfismo* di spazi vettoriali.

Osservazione 2.1.2. Supponiamo che f sia una funzione additiva tra due spazi vettoriali V e W definiti sul campo K , e supponiamo che $\mathbb{Q} \subseteq K$. Per ogni

intero positivo n ed ogni $v \in V$, si ha

$$f(nv) = f(\underbrace{v + v + \cdots + v}_n) = \underbrace{f(v) + f(v) + \cdots + f(v)}_n = nf(v).$$

Si ha inoltre $f(\mathbf{0}_V) = \mathbf{0}_W$: infatti dall'additività di f si deduce che

$$f(v) = f(v + \mathbf{0}_V) = f(v) + f(\mathbf{0}_V),$$

da cui, sommando ad ambo i membri il vettore $-f(v)$, si conclude.

Utilizzando questo risultato si può dimostrare che $f(-v) = -f(v)$: si ha infatti

$$\mathbf{0}_W = f(\mathbf{0}_V) = f(v + (-v)) = f(v) + f(-v),$$

da cui segue che $f(-v)$ è l'opposto di $f(v)$.

Combinando questi risultati, si conclude che l'uguaglianza $f(nv) = nf(v)$ vale per ogni vettore $v \in V$ ed ogni $n \in \mathbb{Z}$: una funzione additiva è quindi automaticamente \mathbb{Z} -lineare.

In effetti una funzione additiva è anche \mathbb{Q} -lineare. Infatti, per ogni $n \neq 0$, si ha

$$f(v) = f(n \frac{1}{n} v) = nf(\frac{1}{n} v),$$

da cui segue che $f(\frac{1}{n} v) = \frac{1}{n} f(v)$. Infine, per ogni $\frac{m}{n} \in \mathbb{Q}$, si ha

$$f(\frac{m}{n} v) = mf(\frac{1}{n} v) = \frac{m}{n} f(v).$$

Tuttavia, se K contiene propriamente \mathbb{Q} , dall'additività di una funzione non si può dedurre, in generale, la sua K -linearità. Ad esempio, dimostreremo in seguito (vedi Esempio 2.1.19) che esistono delle funzioni additive (e quindi \mathbb{Q} -lineari) che non sono \mathbb{R} -lineari!

Veniamo ora alla definizione di isomorfismo di spazi vettoriali.

Definizione 2.1.3. Sia $f : V \rightarrow W$ un omomorfismo di spazi vettoriali. f è un *isomorfismo* se esiste un omomorfismo $g : W \rightarrow V$ tale che $g \circ f = \text{id}_V$ e $f \circ g = \text{id}_W$.

In altre parole, dire che f è un isomorfismo di spazi vettoriali equivale a dire che f è un omomorfismo invertibile e che la sua funzione inversa è lineare.

Due spazi vettoriali V e W su K si dicono *isomorfi* se esiste un isomorfismo $f : V \rightarrow W$. Quando vorremo indicare che V e W sono isomorfi senza specificare quale sia l'isomorfismo, scriveremo semplicemente $V \cong W$.

Dalla definizione data segue che un isomorfismo di spazi vettoriali è una funzione biiettiva. Dimostriamo ora il viceversa:

Proposizione 2.1.4. *Sia $f : V \rightarrow W$ un omomorfismo di spazi vettoriali. Se la funzione f è biiettiva essa è un isomorfismo.*

Dimostrazione. Poiché f è biiettiva essa è invertibile. Rimane quindi solo da dimostrare che la funzione inversa $f^{-1} : W \rightarrow V$ è lineare.

Siano dunque $w_1, w_2 \in W$ e poniamo $v_1 = f^{-1}(w_1)$ e $v_2 = f^{-1}(w_2)$. Dall'additività di f si deduce che $f(v_1 + v_2) = f(v_1) + f(v_2) = w_1 + w_2$, da cui segue che $f^{-1}(w_1 + w_2) = v_1 + v_2 = f^{-1}(w_1) + f^{-1}(w_2)$; ciò dimostra che f^{-1} è additiva.

Consideriamo ora uno scalare $\lambda \in K$. Dalla linearità di f segue che $f(\lambda v_1) = \lambda f(v_1) = \lambda w_1$, da cui si deduce che $f^{-1}(\lambda w_1) = \lambda v_1 = \lambda f^{-1}(w_1)$. Abbiamo così dimostrato che f^{-1} è lineare. \square

Osservazione 2.1.5. Un omomorfismo iniettivo di spazi vettoriali è anche detto un *monomorfismo*, mentre un omomorfismo suriettivo è chiamato *epimorfismo*. Un monomorfismo che sia anche epimorfismo è dunque un omomorfismo biiettivo e quindi, in base alla proposizione precedente, è un isomorfismo.

L'importanza della nozione di isomorfismo è data dal fatto che esso permette di “identificare” spazi vettoriali diversi, a patto che siano isomorfi. Si può così arrivare a una classificazione degli spazi vettoriali, come descritto nel seguente risultato:

Proposizione 2.1.6. *Sia V uno spazio vettoriale di dimensione n sul campo K . Allora V è isomorfo (non canonicamente) allo spazio vettoriale K^n .*

Dimostrazione. Fissiamo una base $\{v_1, v_2, \dots, v_n\}$ di V . Facciamo notare che ciò è sempre possibile, anche se non c'è, in generale, nessuna scelta “canonica” per una tale base.

Ora possiamo definire una funzione $f : V \rightarrow K^n$ la quale associa a un vettore $v \in V$ l'unica n -upla $(\lambda_1, \dots, \lambda_n) \in K^n$ per cui si ha

$$v = \lambda_1 v_1 + \dots + \lambda_n v_n.$$

È immediato verificare che la funzione f è lineare. Essa è inoltre biiettiva, dato che $\{v_1, v_2, \dots, v_n\}$ è una base di V . Dalla Proposizione 2.1.4 si deduce quindi che f è un isomorfismo.

Vogliamo far notare che la funzione f dipende dalla base di V che è stata scelta. La non esistenza, in generale, di una base canonica ha quindi come conseguenza la non esistenza di una scelta canonica di un isomorfismo tra V e K^n . \square

Corollario 2.1.7. *Due spazi vettoriali di dimensione finita sul campo K sono isomorfi (non in modo canonico) se e solo se hanno la stessa dimensione.*

2.1.1 Nucleo e immagine

Introduciamo ora due sottospazi vettoriali particolarmente importanti associati a una funzione lineare:

Definizione 2.1.8. Sia $f : V \rightarrow W$ una funzione lineare tra due spazi vettoriali. Il *nucleo* di f è l'insieme

$$\text{Ker}(f) = \{v \in V \mid f(v) = \mathbf{0}\}.$$

L'*immagine* di f è l'insieme

$$\text{Im}(f) = \{w \in W \mid w = f(v), \text{ per qualche } v \in V\}.$$

Proposizione 2.1.9. *Il nucleo di una funzione lineare $f : V \rightarrow W$ è un sottospazio vettoriale di V , mentre l'immagine di f è un sottospazio vettoriale di W .*

Dimostrazione. Siano $v_1, v_2 \in \text{Ker}(f)$ e consideriamo una combinazione lineare $\lambda_1 v_1 + \lambda_2 v_2$, con $\lambda_1, \lambda_2 \in K$. Dalla linearità di f segue che

$$f(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 f(v_1) + \lambda_2 f(v_2) = \mathbf{0},$$

quindi $\lambda_1 v_1 + \lambda_2 v_2 \in \text{Ker}(f)$. Questo dimostra che $\text{Ker}(f)$ è un sottospazio vettoriale di V .

Passiamo ora all'immagine di f . Siano $w_1, w_2 \in \text{Im}(f)$ e siano $v_1, v_2 \in V$ tali che $w_1 = f(v_1)$ e $w_2 = f(v_2)$. Dalla linearità di f segue che

$$f(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 f(v_1) + \lambda_2 f(v_2) = \lambda_1 w_1 + \lambda_2 w_2,$$

il che significa che $\lambda_1 w_1 + \lambda_2 w_2 \in \text{Im}(f)$, per ogni $\lambda_1, \lambda_2 \in K$. Ciò dimostra che $\text{Im}(f)$ è un sottospazio vettoriale di W . \square

Il seguente risultato fornisce una caratterizzazione dei monomorfismi in termini di annullamento del nucleo.

Proposizione 2.1.10. *Sia $f : V \rightarrow W$ una funzione lineare. Allora f è iniettiva se e solo se $\text{Ker}(f) = \{\mathbf{0}\}$.*

Dimostrazione. Supponiamo che f sia iniettiva. Sia $v \in \text{Ker}(f)$: si ha quindi $f(v) = \mathbf{0}$. Ricordando che $f(\mathbf{0}) = \mathbf{0}$, dall'iniettività di f si deduce che $v = \mathbf{0}$, il che dimostra che $\text{Ker}(f) = \{\mathbf{0}\}$.

Viceversa, supponiamo che $\text{Ker}(f) = \{\mathbf{0}\}$. Siano $v_1, v_2 \in V$ tali che $f(v_1) = f(v_2)$. Dalla linearità di f si ha

$$f(v_1 - v_2) = f(v_1) - f(v_2) = \mathbf{0},$$

quindi $v_1 - v_2 \in \text{Ker}(f)$. Poiché, per ipotesi, $\text{Ker}(f) = \{\mathbf{0}\}$, si ha $v_1 - v_2 = \mathbf{0}$, cioè $v_1 = v_2$. Questo dimostra che f è iniettiva. \square

Abbiamo visto come il nucleo di un'applicazione lineare $f : V \rightarrow W$ sia sempre un sottospazio vettoriale di V . Ora dimostreremo che, più in generale, l'immagine inversa di un qualsiasi vettore $w \in \text{Im}(f)$ si ottiene semplicemente "traslando" il nucleo di f tramite un qualsiasi vettore $v \in f^{-1}(w)$.

Proposizione 2.1.11. *Sia $f : V \rightarrow W$ una funzione lineare e sia $w \in W$. Se $w \in \text{Im}(f)$ si ha*

$$f^{-1}(w) = v + \text{Ker}(f) = \{v + u \mid u \in \text{Ker}(f)\},$$

ove v è un qualsiasi vettore tale che $f(v) = w$; se invece $w \notin \text{Im}(f)$ si ha $f^{-1}(w) = \emptyset$.

Dimostrazione. Sia $v \in V$ tale che $f(v) = w$. Per ogni $u \in \text{Ker}(f)$, si ha $f(v + u) = f(v) + f(u) = w + \mathbf{0} = w$. Ciò dimostra che $v + \text{Ker}(f) \subseteq f^{-1}(w)$.

Viceversa, per ogni $v' \in f^{-1}(w)$ poniamo $u = v' - v$. Si ha $f(u) = f(v') - f(v) = w - w = \mathbf{0}$, quindi $u \in \text{Ker}(f)$. Da ciò discende che $v' = v + u \in v + \text{Ker}(f)$, quindi vale anche l'inclusione $f^{-1}(w) \subseteq v + \text{Ker}(f)$. L'ultima affermazione è ovvia. \square

Le dimensioni del nucleo e dell'immagine di una funzione lineare sono legate tra loro dalla seguente relazione:

Proposizione 2.1.12. *Sia $f : V \rightarrow W$ una funzione lineare. Se V ha dimensione finita, si ha*

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

Dimostrazione. Poniamo $n = \dim(V)$ e $r = \dim \text{Ker}(f)$; bisogna quindi dimostrare che $\dim \text{Im}(f) = n - r$. Consideriamo una base $\{v_1, v_2, \dots, v_r\}$ di $\text{Ker}(f)$ e completiamola a una base $\{v_1, v_2, \dots, v_r, v_{r+1}, \dots, v_n\}$ di V . Ricordiamo che le immagini tramite f dei vettori di una base di V formano un insieme di generatori di $\text{Im}(f)$. Dato che $f(v_1) = f(v_2) = \dots = f(v_r) = \mathbf{0}$, se ne deduce che i vettori

$$w_1 = f(v_{r+1}), w_2 = f(v_{r+2}), \dots, w_{n-r} = f(v_n)$$

generano l'immagine di f . Dimostriamo ora che tali vettori sono anche linearmente indipendenti. Consideriamo una combinazione lineare

$$\lambda_1 w_1 + \lambda_2 w_2 + \dots + \lambda_{n-r} w_{n-r} = \mathbf{0}.$$

Dalla linearità di f , si ha

$$\begin{aligned} \mathbf{0} &= \lambda_1 w_1 + \lambda_2 w_2 + \dots + \lambda_{n-r} w_{n-r} \\ &= \lambda_1 f(v_{r+1}) + \lambda_2 f(v_{r+2}) + \dots + \lambda_{n-r} f(v_n) \\ &= f(\lambda_1 v_{r+1} + \lambda_2 v_{r+2} + \dots + \lambda_{n-r} v_n) \end{aligned}$$

e pertanto

$$\lambda_1 v_{r+1} + \lambda_2 v_{r+2} + \dots + \lambda_{n-r} v_n \in \text{Ker}(f).$$

Poiché $\{v_1, v_2, \dots, v_r\}$ è una base di $\text{Ker}(f)$, si ha

$$\lambda_1 v_{r+1} + \lambda_2 v_{r+2} + \dots + \lambda_{n-r} v_n = \mu_1 v_1 + \mu_2 v_2 + \dots + \mu_r v_r$$

e quindi

$$\mu_1 v_1 + \mu_2 v_2 + \dots + \mu_r v_r - \lambda_1 v_{r+1} - \lambda_2 v_{r+2} - \dots - \lambda_{n-r} v_n = \mathbf{0}.$$

Dato che, per ipotesi, i vettori v_1, v_2, \dots, v_n sono una base di V , si conclude che

$$\mu_1 = \mu_2 = \dots = \mu_r = 0, \quad \lambda_1 = \lambda_2 = \dots = \lambda_{n-r} = 0,$$

il che dimostra che i vettori w_1, w_2, \dots, w_{n-r} sono linearmente indipendenti. Concludiamo quindi che tali vettori sono una base dell'immagine di f e dunque $\dim \text{Im}(f) = n - r$, come volevasi dimostrare. \square

Osservazione 2.1.13. Sia $f : V \rightarrow W$ una funzione lineare tra due spazi vettoriali. La dimensione dell'immagine di f è detta il *rango* di f ,

$$\text{rk}(f) = \dim \text{Im}(f)$$

mentre la dimensione del nucleo di f è detta la *nullità* di f ,

$$\text{null}(f) = \dim \text{Ker}(f).$$

La proposizione precedente afferma quindi che, per ogni omomorfismo $f : V \rightarrow W$, si ha

$$\text{rk}(f) + \text{null}(f) = \dim(V).$$

Osservazione 2.1.14. Siano V e W due spazi vettoriali su K e indichiamo con $\text{Hom}(V, W)$ l'insieme delle applicazioni lineari da V a W . Definiamo la somma di due applicazioni lineari $f, g \in \text{Hom}(V, W)$ ponendo $(f + g)(v) = f(v) + g(v)$, per ogni $v \in V$; essa è ancora una funzione lineare. Definiamo poi il prodotto di uno scalare $\lambda \in K$ per una funzione lineare $f \in \text{Hom}(V, W)$ ponendo $(\lambda f)(v) = \lambda(f(v))$, per ogni $v \in V$. Si verifica facilmente che l'insieme $\text{Hom}(V, W)$, dotato delle due operazioni appena definite, è uno spazio vettoriale su K .

Notiamo infine che, se $W = V$, la composizione di due applicazioni lineari $f, g : V \rightarrow V$ è ancora una funzione lineare, cioè $g \circ f \in \text{Hom}(V, V)$, per ogni $f, g \in \text{Hom}(V, V)$. L'insieme $\text{Hom}(V, V)$, dotato dell'operazione di somma e dell'operazione di composizione, risulta essere un anello (unitario) non commutativo. Se, in aggiunta a queste due operazioni, consideriamo anche il prodotto di una funzione lineare per uno scalare $\lambda \in K$, si ottiene una struttura nota con il nome di K -algebra.

Osservazione 2.1.15. Un omomorfismo di uno spazio vettoriale V in sé stesso, $f : V \rightarrow V$, è anche detto *endomorfismo*. Se esso è invertibile, si parla allora di *automorfismo*. L'insieme degli endomorfismi di uno spazio vettoriale V è indicato con $\text{End}(V)$ e il sottoinsieme costituito dagli automorfismi è indicato con $\text{Aut}(V)$.

Osservazione 2.1.16. Un *diagramma* costituito da spazi vettoriali e omomorfismi tra di essi è detto *commutativo* se, per ogni coppia di spazi vettoriali, tutte le funzioni tra di essi che si possono ottenere come composizione di omomorfismi del diagramma, sono uguali. A titolo di esempio, consideriamo il seguente diagramma:

$$\begin{array}{ccccc} V_1 & \xrightarrow{f_1} & V_2 & \xrightarrow{f_2} & V_3 \\ f_3 \downarrow & \searrow f_4 & \downarrow f_5 & & \downarrow f_6 \\ V_4 & \xrightarrow{f_7} & V_5 & \xrightarrow{f_8} & V_6 \end{array}$$

Dire che esso è commutativo significa che $f_5 \circ f_1 = f_4$, $f_7 \circ f_3 = f_4$, $f_6 \circ f_2 = f_8 \circ f_5$, etc.

2.1.2 Applicazioni lineari e basi

Ci proponiamo ora di studiare le proprietà di una funzione lineare $f : V \rightarrow W$, in relazione alla scelta di basi per gli spazi vettoriali V e W .

Iniziamo col dimostrare che una funzione lineare $f : V \rightarrow W$ è completamente determinata dalla conoscenza delle immagini dei vettori di una base di V , le quali possono essere scelte arbitrariamente in W .

Proposizione 2.1.17. *Siano V e W due spazi vettoriali sul campo K e sia $\{v_i\}_{i \in I}$ una base (non necessariamente finita) di V .*

- (i) *Un omomorfismo $f : V \rightarrow W$ è determinato, in modo unico, dalle immagini dei vettori v_i , per ogni $i \in I$.*
- (ii) *Scelti arbitrariamente dei vettori $\{w_i\}_{i \in I}$ in W , esiste un unico omomorfismo $f : V \rightarrow W$ tale che $f(v_i) = w_i$, per ogni $i \in I$.*

Dimostrazione. (i) Sia $f : V \rightarrow W$ una funzione lineare e supponiamo di conoscere $f(v_i)$, per ogni $i \in I$. Poiché $\{v_i\}_{i \in I}$ è una base di V , ogni vettore $v \in V$ si può scrivere, in modo unico, come combinazione lineare finita dei vettori v_i :

$$v = \lambda_1 v_{i_1} + \lambda_2 v_{i_2} + \cdots + \lambda_n v_{i_n}.$$

Dalla linearità di f segue che

$$f(v) = \lambda_1 f(v_{i_1}) + \lambda_2 f(v_{i_2}) + \cdots + \lambda_n f(v_{i_n}), \quad (2.1.1)$$

il che dimostra che la conoscenza di $f(v_i)$, per ogni $i \in I$, determina, in modo unico, $f(v)$, per ogni $v \in V$. In altre parole, se $g : V \rightarrow W$ è un omomorfismo tale che $g(v_i) = f(v_i)$, per ogni $i \in I$, da (2.1.1) segue che $g(v) = f(v)$, per ogni $v \in V$.

(ii) Per ogni $i \in I$ scegliamo arbitrariamente un vettore $w_i \in W$. Definiamo una funzione $f : V \rightarrow W$ ponendo $f(v_i) = w_i$, per ogni $i \in I$, ed estendendo f per linearità a tutto V , cioè ponendo

$$f(v) = \lambda_1 f(v_{i_1}) + \lambda_2 f(v_{i_2}) + \cdots + \lambda_n f(v_{i_n}),$$

se $v = \lambda_1 v_{i_1} + \lambda_2 v_{i_2} + \cdots + \lambda_n v_{i_n}$.

Si verifica immediatamente che f è ben definita ed è lineare. L'unicità di una tale f discende dal punto (i). \square

Corollario 2.1.18. *Siano V e W due spazi vettoriali sul campo K , sia $\{v_i\}_{i \in I}$ una base (non necessariamente finita) di V e sia $f : V \rightarrow W$ un'applicazione lineare.*

- (i) f è iniettiva se e solo se $\{f(v_i)\}_{i \in I}$ è un insieme libero;
- (ii) f è suriettiva se e solo se $\{f(v_i)\}_{i \in I}$ è un insieme di generatori di W ;
- (iii) f è un isomorfismo se e solo se $\{f(v_i)\}_{i \in I}$ è una base di W .

Dimostrazione. (i) Ricordiamo, dalla Proposizione 2.1.10, che f è iniettiva se e solo se $\text{Ker}(f) = \{\mathbf{0}\}$. Dimostriamo quindi l'implicazione $\text{Ker}(f) = \{\mathbf{0}\} \Rightarrow \{f(v_i)\}_{i \in I}$ è un insieme libero. Consideriamo una combinazione lineare

$$\lambda_1 f(v_{i_1}) + \lambda_2 f(v_{i_2}) + \cdots + \lambda_n f(v_{i_n}) = \mathbf{0}.$$

Si ha

$$\lambda_1 f(v_{i_1}) + \lambda_2 f(v_{i_2}) + \cdots + \lambda_n f(v_{i_n}) = f(\lambda_1 v_{i_1} + \lambda_2 v_{i_2} + \cdots + \lambda_n v_{i_n}),$$

da cui segue

$$\lambda_1 v_{i_1} + \lambda_2 v_{i_2} + \cdots + \lambda_n v_{i_n} \in \text{Ker}(f).$$

Dato che, per ipotesi, $\text{Ker}(f) = \{\mathbf{0}\}$, si ha

$$\lambda_1 v_{i_1} + \lambda_2 v_{i_2} + \cdots + \lambda_n v_{i_n} = \mathbf{0},$$

da cui segue

$$\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0,$$

perché i vettori $\{v_i\}_{i \in I}$ sono una base di V .

Dimostriamo ora l'implicazione opposta. Sia $v \in \text{Ker}(f)$ e scriviamo

$$v = \lambda_1 v_{i_1} + \lambda_2 v_{i_2} + \cdots + \lambda_n v_{i_n},$$

per qualche n e qualche $\lambda_1, \dots, \lambda_n \in K$. Poiché $f(v) = \mathbf{0}$, dalla linearità di f si deduce che

$$\lambda_1 f(v_{i_1}) + \lambda_2 f(v_{i_2}) + \dots + \lambda_n f(v_{i_n}) = \mathbf{0}.$$

Poiché, per ipotesi, l'insieme $\{f(v_i)\}_{i \in I}$ è libero, si ha

$$\lambda_1 = \lambda_2 = \dots = \lambda_n = 0$$

e dunque $v = \mathbf{0}$, il che dimostra che $\text{Ker}(f) = \{\mathbf{0}\}$.

(ii) Ricordiamo che affermare che f è suriettiva equivale a dire che $\text{Im}(f) = W$. Osserviamo inoltre che $\text{Im}(f)$ è generata dai vettori $f(v_i)$, al variare di $i \in I$. Infatti, per ogni $w \in \text{Im}(f)$ esiste un vettore $v \in V$ tale che $w = f(v)$. Poiché $\{v_i\}_{i \in I}$ è una base di V , è possibile esprimere v come combinazione lineare di un numero finito di v_i ,

$$v = \lambda_1 v_{i_1} + \lambda_2 v_{i_2} + \dots + \lambda_n v_{i_n}.$$

Si ha dunque

$$w = f(v) = \lambda_1 f(v_{i_1}) + \lambda_2 f(v_{i_2}) + \dots + \lambda_n f(v_{i_n}),$$

il che dimostra che l'insieme dei vettori $\{f(v_i)\}_{i \in I}$ genera l'immagine di f .

Da quanto detto segue quindi che $\text{Im}(f) = W$ se e solo se $\{f(v_i)\}_{i \in I}$ è un insieme di generatori di W .

(iii) Poiché f è un isomorfismo se e solo se essa è biettiva (vedi Proposizione 2.1.4), dai punti (i) e (ii) segue che f è un isomorfismo se e solo se $\{f(v_i)\}_{i \in I}$ è un insieme libero di generatori di W , cioè una base di W . \square

Esempio 2.1.19. In questo esempio vedremo come si possa costruire una funzione additiva $f : \mathbb{R} \rightarrow \mathbb{R}$ che non sia \mathbb{R} -lineare.

Consideriamo \mathbb{R} come spazio vettoriale sul campo \mathbb{Q} . I due "vettori" $v_1 = 1$ e $v_2 = \pi$ sono linearmente indipendenti su \mathbb{Q} (cioè deriva dal fatto che π è irrazionale), quindi esiste una base $\{v_i\}_{i \in I}$ di \mathbb{R} su \mathbb{Q} che contiene i numeri 1 e π (osserviamo che una base di \mathbb{R} su \mathbb{Q} non può essere numerabile).

Per la Proposizione 2.1.17 è possibile definire una funzione \mathbb{Q} -lineare $f : \mathbb{R} \rightarrow \mathbb{R}$ fissando arbitrariamente i valori di $f(v_i)$, per ogni $i \in I$. Se poniamo $f(1) = 1$ e $f(\pi) = 2$ (e fissiamo arbitrariamente i rimanenti $f(v_i) \in \mathbb{R}$), otteniamo una funzione additiva (e quindi \mathbb{Q} -lineare) la quale non è \mathbb{R} -lineare. Se lo fosse si avrebbe infatti

$$f(\pi) = f(\pi \cdot 1) = \pi f(1) = \pi,$$

contro l'ipotesi che $f(\pi) = 2$.

Esercizi

Esercizio 2.1.1. Si dica se sono lineari le seguenti funzioni:

- (1) $f : \mathbb{R}^2 \rightarrow \mathbb{R}^3$, $(x, y) \mapsto (x - y, x + y + 1, 0)$;
- (2) $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $(x, y) \mapsto (2x, x + y)$;
- (3) $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, $(x, y) \mapsto \sin(x - y)$.

Esercizio 2.1.2. Si dica per quali valori di $t \in \mathbb{R}$ è lineare la seguente funzione:

$$f : \mathbb{R}^3 \rightarrow \mathbb{R}^2, \quad (x, y, z) \mapsto (x + ty, tyz).$$

Esercizio 2.1.3. Si consideri la funzione tra \mathbb{C} -spazi vettoriali $f : \mathbb{C}^2 \rightarrow \mathbb{C}$ data da $f(x, y) = x + \bar{y}$, ove \bar{y} indica il numero complesso coniugato di y . Si dica se f è lineare (cioè \mathbb{C} -lineare).

Esercizio 2.1.4. Sia $f : V \rightarrow W$ un'applicazione tra due spazi vettoriali. Si dimostri che f è lineare se e solo se il suo grafico è un sottospazio vettoriale di $V \times W$.

2.2 Matrici

Siano V e W due spazi vettoriali sul campo K , di dimensioni n e m , rispettivamente, e fissiamo delle basi $\{v_1, \dots, v_n\}$ di V e $\{w_1, \dots, w_m\}$ di W .

In base alla Proposizione 2.1.17, una funzione lineare $f : V \rightarrow W$ è determinata, in modo unico, dalla conoscenza dei vettori $f(v_j)$, per $j = 1, \dots, n$. Poiché $\{w_1, \dots, w_m\}$ è una base di W , per ogni $j = 1, \dots, n$ possiamo scrivere

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

per degli opportuni $a_{ij} \in K$, con $i = 1, \dots, m$ e $j = 1, \dots, n$.

Da quanto detto si deduce quindi che una funzione lineare $f : V \rightarrow W$ è determinata in modo unico dal dato di mn elementi a_{ij} del campo K . Tali elementi costituiscono ciò che va sotto il nome di *matrice*.

Definizione 2.2.1. Una *matrice*, con m righe e n colonne (o matrice $m \times n$) a coefficienti in K è il dato di mn elementi di K , scritti solitamente sotto forma di tabella rettangolare costituita da m righe e n colonne:

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

Una matrice A di questo tipo sarà spesso indicata semplicemente con la scrittura

$$A = (a_{ij}),$$

dove $i = 1, \dots, m$ è detto *indice di riga* mentre $j = 1, \dots, n$ è detto *indice di colonna*.

Ad ogni funzione lineare $f : V \rightarrow W$ può dunque essere associata una matrice $m \times n$ a coefficienti in K . Naturalmente tale matrice dipende, oltre che dalla funzione f , anche dalla scelta delle basi di V e W .

Osservazione 2.2.2. Ricordiamo che se un vettore $v \in V$ si scrive come combinazione lineare

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n$$

degli elementi di una base $\{v_1, \dots, v_n\}$ di V , i coefficienti $\lambda_1, \dots, \lambda_n$ che compaiono in una tale espressione si dicono le *coordinate* di v rispetto alla base fissata.

Possiamo allora osservare che, dalla definizione della matrice A associata a un omomorfismo $f : V \rightarrow W$, rispetto a delle basi $\{v_1, \dots, v_n\}$ di V e $\{w_1, \dots, w_m\}$ di W , segue che le coordinate del vettore $f(v_j)$, rispetto alla base di W fissata, costituiscono la j -esima colonna della matrice A . Questa osservazione si rivela utile quando è necessario scrivere esplicitamente la matrice associata a una data funzione lineare.

Osservazione 2.2.3. In tutta questa sezione supporremo sempre che gli spazi vettoriali abbiano dimensione finita. Facciamo comunque notare che molti risultati si possono estendere, con opportune modifiche, anche a spazi vettoriali di dimensione infinita.

Consideriamo ora due applicazioni lineari $f, g : V \rightarrow W$ e indichiamo con $A = (a_{ij})$ e $B = (b_{ij})$ le matrici ad esse associate. La somma di f e g è l'applicazione lineare definita da $(f + g)(v) = f(v) + g(v)$, per ogni $v \in V$. In particolare, per ogni vettore v_j della base di V , si ha

$$\begin{aligned} (f + g)(v_j) &= f(v_j) + g(v_j) = \sum_{i=1}^m a_{ij} w_i + \sum_{i=1}^m b_{ij} w_i \\ &= \sum_{i=1}^m (a_{ij} + b_{ij}) w_i. \end{aligned}$$

La matrice associata alla funzione $f + g$ ha quindi come coefficienti le somme $a_{ij} + b_{ij}$ dei coefficienti delle matrici A e B , associate rispettivamente a f e g . Questo risultato motiva la seguente definizione:

Definizione 2.2.4. Siano $A = (a_{ij})$ e $B = (b_{ij})$ due matrici $m \times n$ a coefficienti in K . La loro somma è la matrice

$$A + B = (a_{ij} + b_{ij}),$$

ottenuta sommando i coefficienti di A e B che si trovano nelle stesse posizioni.

Sia ora $\lambda \in K$ e consideriamo la funzione λf definita da $(\lambda f)(v) = \lambda f(v)$, per ogni $v \in V$. Valutando questa funzione sui vettori della base di V , si ha

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

da cui si deduce che la matrice associata alla funzione λf è la matrice i cui coefficienti sono dati dal prodotto di λ per i coefficienti della matrice A di f .

Definizione 2.2.5. Per ogni $\lambda \in K$, il prodotto di λ per una matrice $A = (a_{ij})$ a coefficienti in K è la matrice

$$\lambda A = (\lambda a_{ij}).$$

Indicheremo con $M_{m,n}(K)$ l'insieme delle matrici con m righe e n colonne, a coefficienti in K . Da quanto visto sopra si deduce che esiste una biiezione tra l'insieme $\text{Hom}(V, W)$ e $M_{m,n}(K)$. Dato che $\text{Hom}(V, W)$, con le operazioni di somma di funzioni e di prodotto di una funzione per uno scalare, è uno spazio vettoriale su K , anche l'insieme $M_{m,n}(K)$, con le due operazioni sopra definite, risulta essere un K -spazio vettoriale. Inoltre, i due spazi vettoriali $\text{Hom}(V, W)$ e $M_{m,n}(K)$ sono isomorfi.

Per analogia con la definizione della base canonica di K^n , definiamo delle matrici E_{ij} , con $i = 1, \dots, m$ e $j = 1, \dots, n$, i cui coefficienti sono tutti nulli eccetto quello di posto (i, j) (cioè quello che si trova sulla i -esima riga e sulla j -esima colonna), che è uguale a 1. È immediato verificare che le mn matrici E_{ij} appena definite formano una base dello spazio vettoriale $M_{m,n}(K)$. Ciò è

conseguenza del fatto che ogni matrice $A = (a_{ij})$ si scrive, in modo unico, come segue:

$$A = \sum_{i,j} a_{ij} E_{ij}.$$

Possiamo riassumere quanto appena visto nel seguente risultato:

Proposizione 2.2.6. *$M_{m,n}(K)$ è uno spazio vettoriale di dimensione mn su K . Se V e W sono due K -spazi vettoriali di dimensioni n e m rispettivamente, vi è un isomorfismo $\text{Hom}(V, W) \cong M_{m,n}(K)$. Tale isomorfismo non è canonico, in quanto dipende dalla scelta di una base di V e di una base di W .*

In particolare, se $V = W$ e quindi $m = n$, lo spazio vettoriale $\text{End}(V) = \text{Hom}(V, V)$ è isomorfo a $M_n(K) = M_{n,n}(K)$ e ha dimensione n^2 .

Vediamo ora quale operazione tra matrici corrisponde alla composizione di due funzioni lineari. A tal fine consideriamo tre spazi vettoriali U , V e W , di dimensioni rispettivamente r , n e m , e fissiamo delle loro basi $\{u_1, \dots, u_r\}$, $\{v_1, \dots, v_n\}$ e $\{w_1, \dots, w_m\}$. Siano $f : V \rightarrow W$ e $g : U \rightarrow V$ due applicazioni lineari e indichiamo con A la matrice di f , con B la matrice di g e con C la matrice di $f \circ g : U \rightarrow W$, rispetto alle basi indicate. Ricordiamo che A è una matrice $m \times n$, B è una matrice $n \times r$, mentre C è una matrice $m \times r$.

Per ogni vettore u_j della base di U si ha:

$$\begin{aligned} (f \circ g)(u_j) &= f(g(u_j)) = f\left(\sum_{h=1}^n b_{hj} v_h\right) \\ &= \sum_{h=1}^n b_{hj} f(v_h) = \sum_{h=1}^n b_{hj} \left(\sum_{i=1}^m a_{ih} w_i\right) \\ &= \sum_{i=1}^m \left(\sum_{h=1}^n a_{ih} b_{hj}\right) w_i. \end{aligned}$$

Poiché $C = (c_{ij})$ è la matrice di $f \circ g$, si ha anche

$$(f \circ g)(u_j) = \sum_{i=1}^m c_{ij} w_i.$$

Dall'uguaglianza di queste due ultime espressioni (e dal fatto che i vettori $\{w_1, \dots, w_m\}$ sono una base di W) segue che

$$c_{ij} = \sum_{h=1}^n a_{ih} b_{hj},$$

per ogni $i = 1, \dots, m$ e $j = 1, \dots, r$. Utilizzeremo dunque questa formula per definire un prodotto di matrici, in modo che il prodotto delle matrici A e B associate agli omomorfismi f e g fornisca proprio la matrice associata all'omomorfismo composto $f \circ g$.

Definizione 2.2.7. Date due matrici $A \in M_{m,n}(K)$ e $B \in M_{n,r}(K)$, il loro *prodotto* è la matrice $C \in M_{m,r}(K)$ i cui coefficienti sono dati da

$$c_{ij} = \sum_{h=1}^n a_{ih} b_{hj}, \tag{2.2.1}$$

per ogni $i = 1, \dots, m$ e $j = 1, \dots, r$. Questo prodotto di matrici è anche detto *prodotto righe per colonne*.

Vediamo più in dettaglio come si calcola un tale prodotto di matrici. Siano A e B due matrici come sopra e vogliamo determinare il loro prodotto $C = AB$. Per calcolare l'elemento c_{ij} , che si trova sulla i -esima riga e sulla j -esima colonna della matrice C , dobbiamo selezionare la i -esima riga della matrice A e la j -esima colonna della matrice B :

$$(a_{i1}, a_{i2}, \dots, a_{in}) \begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{pmatrix}$$

dopodiché dobbiamo “moltiplicare” questa riga per questa colonna nel modo indicato dalla formula (2.2.1), cioè dobbiamo effettuare la somma dei prodotti componente per componente dei due vettori indicati:

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{in}b_{nj}.$$

Osserviamo che per fare ciò è indispensabile che la lunghezza delle righe di A coincida con la lunghezza delle colonne di B . La matrice risultante dal prodotto di A per B avrà un numero di righe pari a quello della matrice A e un numero di colonne pari a quello della matrice B .

Un caso particolare di prodotto tra matrici si ha quando la matrice B ha una sola colonna, cioè quando B si riduce a un vettore (scritto in colonna): si ottiene in questo modo il prodotto di una matrice per un vettore, il cui risultato è ancora un vettore. Più precisamente, data una matrice $A \in M_{m,n}(K)$ e un vettore $v = (x_1, x_2, \dots, x_n) \in K^n$ (che scriveremo in colonna), il prodotto Av è un vettore $w = (y_1, y_2, \dots, y_m) \in K^m$ dato da

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Si ottiene in questo modo un'applicazione lineare

$$F_A : K^n \rightarrow K^m, \quad v \mapsto w = F_A(v) = Av.$$

La matrice associata a questa applicazione lineare (rispetto alle basi canoniche di K^n e K^m) è proprio la matrice A .

Osservazione 2.2.8. In modo del tutto equivalente si può considerare il caso particolare del prodotto di A per B , quando la matrice A si riduce a un vettore (questa volta scritto in riga). Consideriamo dunque una matrice $B \in M_{n,r}(K)$ e un vettore $v = (x_1, x_2, \dots, x_n) \in K^n$ (che scriveremo in riga). Il prodotto vB è un vettore $w = (y_1, y_2, \dots, y_r) \in K^r$ dato da

$$(y_1, y_2, \dots, y_r) = (x_1, x_2, \dots, x_n) \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1r} \\ b_{21} & b_{22} & \dots & b_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nr} \end{pmatrix}$$

Anche in questo caso si ottiene un'applicazione lineare

$$G_B : K^n \rightarrow K^r, \quad v \mapsto w = G_B(v) = vB.$$

Si conclude pertanto che un omomorfismo tra due spazi vettoriali quali K^n e K^m può essere descritto sia dal prodotto di un vettore *riga* per una certa matrice, sia dal prodotto di un'altra matrice per un vettore *colonna*. Naturalmente si passa da una descrizione all'altra semplicemente scambiando tra loro i ruoli delle righe con quelli delle colonne. L'operazione che trasforma una matrice $m \times n$ in una matrice $n \times m$ scambiando tra di loro le righe con le colonne si chiama *trasposizione*.

Definizione 2.2.9. Sia $A = (a_{ij}) \in M_{m,n}(K)$. La *trasposta* di A è la matrice ${}^t A \in M_{n,m}(K)$ il cui coefficiente di posto (i, j) è a_{ji} , cioè è il coefficiente di posto (j, i) della matrice A .

Il trasposto di un vettore scritto in colonna è dunque un vettore scritto in riga, e viceversa. Per comodità di notazione, d'ora in poi i vettori di K^n verranno sempre pensati come vettori colonna:

$$v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Per indicare invece un analogo vettore pensato come vettore riga, scriveremo quindi ${}^t v$:

$${}^t v = (x_1, x_2, \dots, x_n).$$

Come ultimo caso particolare del prodotto di due matrici, vediamo cosa succede quando sia A che B si riducono a dei vettori (scritti il primo in riga e il secondo in colonna). In questo caso il risultato del prodotto è uno scalare, cioè un elemento di K :

$$(a_1, a_2, \dots, a_n) \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = a_1 b_1 + a_2 b_2 + \dots + a_n b_n \in K.$$

Si ottiene in questo modo la definizione di un prodotto tra due vettori di K^n , il cui risultato è uno scalare: questo è il cosiddetto *prodotto scalare* di due vettori.

Definizione 2.2.10. Siano $v = {}^t(x_1, x_2, \dots, x_n)$ e $w = {}^t(y_1, y_2, \dots, y_n)$ due elementi di K^n . Il loro *prodotto scalare*, che indicheremo con $v \cdot w$ (o, a volte, con $\langle v, w \rangle$) è definito da

$$v \cdot w = {}^t v w = \sum_{i=1}^n x_i y_i.$$

Di questa nozione di prodotto scalare, e delle sue generalizzazioni, ci occuperemo in seguito. Vediamo ora alcune proprietà dell'operazione di trasposizione.

Proposizione 2.2.11. Siano $A, B \in M_{m,n}(K)$ e sia $\lambda \in K$. Si ha:

- (i) ${}^t({}^tA) = A$;
- (ii) ${}^t(A + B) = {}^tA + {}^tB$;
- (iii) ${}^t(\lambda A) = \lambda {}^tA$.

Se $A \in M_{m,n}(K)$ e $B \in M_{n,r}(K)$, si ha inoltre

- (iv) ${}^t(AB) = {}^tB {}^tA$.

Dimostrazione. Le prime tre proprietà sono ovvie, dimostriamo quindi la quarta. Indichiamo con a_{ij} i coefficienti di A e con \tilde{a}_{ij} i coefficienti di tA : si ha quindi $\tilde{a}_{ij} = a_{ji}$. Analogamente indichiamo con b_{ij} e con \tilde{b}_{ij} i coefficienti di B e tB , rispettivamente. Indichiamo poi con c_{ij} i coefficienti del prodotto AB e con \tilde{c}_{ij} i coefficienti di ${}^t(AB)$. Infine, indichiamo con d_{ij} i coefficienti della matrice ${}^tB {}^tA$. Ricordando la definizione del prodotto di due matrici, si ha:

$$\tilde{c}_{ij} = c_{ji} = \sum_h a_{jh} b_{hi},$$

mentre

$$d_{ij} = \sum_h \tilde{b}_{ih} \tilde{a}_{hj} = \sum_h a_{jh} b_{hi},$$

da cui segue che $d_{ij} = \tilde{c}_{ij}$, per ogni i e j . \square

Ritorniamo ora al prodotto di matrici e studiamo più in dettaglio alcune delle sue proprietà.

Proposizione 2.2.12. *Siano A , B e C tre matrici e siano $\lambda, \mu \in K$. Ogni volta che le somme e i prodotti indicati sono definiti, si ha:*

- (i) $(AB)C = A(BC)$;
- (ii) $(A + B)C = AC + BC$;
- (iii) $A(B + C) = AB + AC$;
- (iv) $\lambda(AB) = (\lambda A)B = A(\lambda B)$;
- (v) $(\lambda + \mu)A = \lambda A + \mu A$;
- (vi) $(\lambda\mu)A = \lambda(\mu A)$.

Dimostrazione. Tutte queste proprietà discendono dalle analoghe proprietà delle operazioni definite sulle funzioni lineari: ad esempio, la proprietà associativa del prodotto di matrici $(AB)C = A(BC)$ equivale alla proprietà associativa del prodotto di composizione $(f \circ g) \circ h = f \circ (g \circ h)$ delle funzioni. In ogni caso, si possono dimostrare direttamente mediante un semplice calcolo. A titolo di esempio, dimostriamo la prima.

Indichiamo con a_{ij} i coefficienti della matrice A , con b_{ij} quelli di B e con c_{ij} i coefficienti di C . Indichiamo inoltre con d_{ij} i coefficienti della matrice prodotto di A per B e con e_{ij} quelli del prodotto $(AB)C$. Dalla definizione del prodotto di due matrici si ha:

$$e_{ij} = \sum_h d_{ih} c_{hj} = \sum_h \left(\sum_k a_{ik} b_{kh} \right) c_{hj} = \sum_{h,k} a_{ik} b_{kh} c_{hj}.$$

Ora basta osservare che se calcoliamo, in modo analogo, i coefficienti del prodotto $A(BC)$, troviamo esattamente la stessa espressione. \square

Sia $f : V \rightarrow W$ un'applicazione lineare tra due spazi vettoriali di dimensioni n e m rispettivamente. Abbiamo già osservato che la scelta di una base $\mathbf{v} = \{v_1, \dots, v_n\}$ di V determina un isomorfismo $\alpha_{\mathbf{v}} : V \xrightarrow{\sim} K^n$ che associa ad ogni vettore $v \in V$ la n -upla $(\lambda_1, \dots, \lambda_n)$ delle sue coordinate rispetto alla base \mathbf{v} . Analogamente la scelta di una base $\mathbf{w} = \{w_1, \dots, w_m\}$ di W determina un isomorfismo $\beta_{\mathbf{w}} : W \xrightarrow{\sim} K^m$ che associa ad ogni vettore $w \in W$ la m -upla (μ_1, \dots, μ_m) delle sue coordinate rispetto alla base \mathbf{w} .

Sia dunque A la matrice di f rispetto alle basi scelte. Essa determina un'applicazione lineare $F : K^n \rightarrow K^m$, definita da

$$F : \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \mapsto \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix} = A \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}.$$

Proposizione 2.2.13. *Con le notazioni precedenti, il diagramma*

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ \alpha_{\mathbf{v}} \downarrow \wr & & \downarrow \wr \beta_{\mathbf{w}} \\ K^n & \xrightarrow{F} & K^m \end{array} \quad (2.2.2)$$

è commutativo

Dimostrazione. Dobbiamo dimostrare che $\beta_{\mathbf{w}} \circ f = F \circ \alpha_{\mathbf{v}}$. Sia dunque $v \in V$ ed esprimiamo v come combinazione lineare dei vettori della base \mathbf{v} :

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n.$$

Per definizione della funzione $\alpha_{\mathbf{v}}$, si ha $\alpha_{\mathbf{v}}(v) = {}^t(\lambda_1, \dots, \lambda_n)$ (si ricordi che abbiamo deciso di scrivere gli elementi di K^n come vettori colonna). Calcolando ora $F(\alpha_{\mathbf{v}}(v))$ si ottiene il vettore

$$A \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix},$$

la cui i -esima componente è

$$a_{i1}\lambda_1 + a_{i2}\lambda_2 + \dots + a_{in}\lambda_n = \sum_{j=1}^n a_{ij}\lambda_j. \quad (2.2.3)$$

Calcoliamo ora $f(v)$. Dalla linearità di f e dalla definizione della matrice $A = (a_{ij})$ associata a f , si ha:

$$\begin{aligned} f(v) &= f(\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n) \\ &= \sum_{j=1}^n \lambda_j f(v_j) \\ &= \sum_{j=1}^n \lambda_j \left(\sum_{i=1}^m a_{ij} w_i \right) \\ &= \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij} \lambda_j \right) w_i. \end{aligned}$$

Poniamo

$$\mu_i = \sum_{j=1}^n a_{ij} \lambda_j, \quad (2.2.4)$$

per $i = 1, \dots, m$, in modo che si abbia

$$f(v) = \sum_{i=1}^m \mu_i w_i.$$

La m -upla (μ_1, \dots, μ_m) rappresenta le coordinate del vettore $f(v)$ rispetto alla base \mathbf{w} e si ha pertanto $\beta_{\mathbf{w}}(f(v)) = {}^t(\mu_1, \dots, \mu_m)$. A questo punto basta osservare che l'espressione di μ_i in (2.2.4) coincide con l'espressione (2.2.3) per la i -esima componente del vettore $F(\alpha_{\mathbf{v}}(v))$. Abbiamo così dimostrato che $F(\alpha_{\mathbf{v}}(v)) = \beta_{\mathbf{w}}(f(v))$, per ogni $v \in V$. \square

Osservazione 2.2.14. Questo risultato fornisce un metodo diretto per calcolare l'immagine tramite $f : V \rightarrow W$ di un qualsiasi vettore $v \in V$, nota la matrice di f rispetto a delle basi prefissate dei due spazi vettoriali V e W .

Dapprima si determinano le coordinate $(\lambda_1, \dots, \lambda_n)$ del vettore v rispetto alla base di V , poi si moltiplica la matrice A associata a f per il vettore $(\lambda_1, \dots, \lambda_n)$, scritto in colonna. Il vettore risultante è costituito dalle coordinate di $f(v)$ rispetto alla base di W .

Dal diagramma commutativo (2.2.2) segue che il nucleo di f e il nucleo di F sono tra loro isomorfi, essendo tale isomorfismo indotto dall'isomorfismo $\alpha_{\mathbf{v}}$. Analogamente, l'isomorfismo $\beta_{\mathbf{w}}$ induce un isomorfismo tra $\text{Im}(f)$ e $\text{Im}(F)$. In particolare questi spazi vettoriali hanno la stessa dimensione. Si ha pertanto

$$\text{null}(f) = \text{null}(F) \quad \text{e} \quad \text{rk}(f) = \text{rk}(F).$$

Dato che l'applicazione lineare $F : K^n \rightarrow K^m$ è data dalla moltiplicazione per la matrice A , diamo la seguente definizione:

Definizione 2.2.15. Sia A una matrice $m \times n$ a coefficienti in K e sia $F : K^n \rightarrow K^m$ l'applicazione lineare data dalla moltiplicazione di un vettore (colonna) per la matrice A (a sinistra). Definiamo il *rango* e la *nullità* della matrice A ponendo

$$\begin{aligned} \text{rk}(A) &= \text{rk}(F) = \dim \text{Im}(F), \\ \text{null}(A) &= \text{null}(F) = \dim \text{Ker}(F). \end{aligned}$$

Osserviamo che il sottospazio $\text{Im}(F)$ di K^m è generato dalle *colonne* di A (possiamo anche osservare che nell'isomorfismo $\beta_{\mathbf{w}} : W \xrightarrow{\sim} K^m$ le colonne della matrice A corrispondono alle immagini, tramite $f : V \rightarrow W$, dei vettori della base di V , le quali generano il sottospazio $\text{Im}(f)$ di W). Pertanto la dimensione di $\text{Im}(F)$, cioè il rango di F , coincide con il massimo numero di colonne linearmente indipendenti della matrice A . Abbiamo così dimostrato il seguente risultato:

Proposizione 2.2.16. *Il rango di una matrice A è il massimo numero di colonne linearmente indipendenti di A .*

Osservazione 2.2.17. Il risultato della proposizione precedente viene spesso usato come definizione del rango di una matrice. Si parla allora di *rango per colonne*, per distinguerlo da un analogo *rango per righe*, definito come il massimo numero di righe linearmente indipendenti. Vedremo in seguito che, in effetti, queste due nozioni di rango coincidono sempre, cioè in ogni matrice il massimo numero di colonne linearmente indipendenti è sempre uguale al massimo numero di righe linearmente indipendenti.

2.2.1 Matrici quadrate

Abbiamo visto che il prodotto di due matrici (così come la composizione di due applicazioni) non è sempre definito: affinché il prodotto AB sia definito è necessario (e sufficiente) che il numero di colonne della matrice A sia uguale al numero di righe di B . Se ci restringiamo a considerare solo matrici di tipo $n \times n$, questi problemi scompaiono e il prodotto di due matrici è sempre definito.

Definizione 2.2.18. Una matrice a coefficienti in K si dice *quadrata di ordine n* se essa ha n righe e n colonne. L'insieme delle matrici quadrate di ordine n è indicato semplicemente con $M_n(K)$, al posto di $M_{n,n}(K)$.

Osservazione 2.2.19. Se V è uno spazio vettoriale di dimensione n su K , e se è stata fissata una base $\{v_1, \dots, v_n\}$ di V , ad ogni endomorfismo $f : V \rightarrow V$ corrisponde una matrice quadrata $A \in M_n(K)$. Questa corrispondenza stabilisce una biiezione tra $\text{End}(V)$ e $M_n(K)$. Poiché $\text{End}(V)$, con le operazioni di somma di funzioni, di prodotto di una funzione per uno scalare e di composizione di due funzioni, è una K -algebra, lo stesso vale per l'insieme delle matrici quadrate $M_n(K)$.

Proposizione 2.2.20. *L'insieme $M_n(K)$ delle matrici quadrate di ordine n a coefficienti in K , dotato delle operazioni di somma e di prodotto di matrici e dell'operazione di prodotto di una matrice per un elemento di K , è una K -algebra.*

Facciamo notare che l'elemento neutro per l'operazione di somma è la *matrice nulla*

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

(la quale corrisponde all'applicazione nulla $f : V \rightarrow V$, $f(v) = \mathbf{0}$, per ogni $v \in V$), mentre l'elemento neutro per l'operazione di prodotto di matrici è la *matrice identica*, definita da

$$\mathbf{1} = \mathbf{1}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

(che corrisponde all'identità $\text{id} : V \rightarrow V$), cioè la matrice avente tutti i coefficienti sulla cosiddetta *diagonale principale* pari a 1, mentre tutti gli altri coefficienti

sono nulli. Infatti è immediato verificare che, per ogni matrice $A \in M_n(K)$, si ha

$$\mathbf{1}_n A = A \mathbf{1}_n = A.$$

Infine notiamo che il prodotto di matrici non gode della proprietà commutativa: se A e B sono due matrici in $M_n(K)$ si ha, in generale,

$$AB \neq BA.$$

Ciò non deve stupire in quanto riflette semplicemente il fatto che la composizione di due funzioni lineari $f, g : V \rightarrow V$ non è, in generale, commutativa, cioè $f \circ g \neq g \circ f$.

Una matrice del tipo $\lambda \mathbf{1}_n$, cioè

$$\begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda \end{pmatrix}$$

con $\lambda \in K$, è detta *matrice scalare*. Essa corrisponde all'omomorfismo $f : V \rightarrow V$ definito da $f(v) = \lambda v$. È immediato verificare che una matrice scalare commuta con ogni altra matrice, cioè

$$(\lambda \mathbf{1}_n)A = A(\lambda \mathbf{1}_n),$$

per ogni $A \in M_n(K)$.

Più in generale, una matrice del tipo

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

cioè una matrice in cui tutti i coefficienti sono nulli, tranne al più quelli sulla diagonale principale, è detta *matrice diagonale*. Si noti che, in generale, una matrice diagonale *non* commuta con un'altra matrice qualsiasi. Tuttavia le matrici diagonali commutano tra loro.

Una matrice *triangolare superiore* è una matrice in cui tutti i coefficienti che si trovano al di sotto della diagonale principale sono nulli, cioè una matrice del tipo

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{pmatrix}$$

Analogamente si definisce una matrice *triangolare inferiore* come una matrice in cui tutti i coefficienti che si trovano al di sopra della diagonale principale sono

nulli, cioè una matrice del tipo

$$\begin{pmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ a_{31} & a_{32} & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix}$$

Si noti che la somma e il prodotto di due matrici triangolari superiori (rispettivamente, inferiori) è ancora una matrice dello stesso tipo.

Osservazione 2.2.21. Consideriamo, a titolo di esempio, il caso di matrici quadrate di ordine 2, a coefficienti razionali. Siano, ad esempio,

$$A = \begin{pmatrix} 2 & -1 \\ -4 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix}.$$

Si verifica immediatamente che il prodotto AB è la matrice nulla, tuttavia né A né B sono nulle! Ciò mostra che, in generale, nell'anello $M_n(K)$ delle matrici quadrate possono esistere degli elementi diversi da zero, con la proprietà che il loro prodotto è uguale a zero (elementi di questo tipo sono detti *divisori di zero*): non vale quindi la cosiddetta “legge di annullamento del prodotto,” secondo la quale il prodotto di due fattori è nullo se e solo se almeno uno dei due fattori è nullo.

Consideriamo ora la matrice

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

La matrice C non è nulla, tuttavia si ha $C^2 = CC = \mathbf{0}$. Più in generale, si può dimostrare che nell'anello $M_n(K)$ esistono delle matrici $C \neq \mathbf{0}$ con la proprietà che $C^r = \mathbf{0}$, per qualche $r > 1$. Tali elementi sono detti *nilpotenti*.

Veniamo ora al problema dell'invertibilità delle matrici di $M_n(K)$. Dato che l'elemento neutro per il prodotto è la matrice identica $\mathbf{1}_n$, l'inversa di una matrice $A \in M_n(K)$ è una matrice $B \in M_n(K)$ tale che si abbia

$$AB = BA = \mathbf{1}_n.$$

Naturalmente l'esistenza in $M_n(K)$ di divisori dello zero impedisce che esistano gli inversi di tutte le matrici non nulle. Infatti, se $A \in M_n(K)$ è un divisore dello zero e se B è una matrice non nulla tale che $AB = \mathbf{0}$, allora, se per assurdo esistesse la matrice A^{-1} inversa di A , si avrebbe

$$B = \mathbf{1}_n B = (A^{-1} A) B = A^{-1} (AB) = A^{-1} \mathbf{0} = \mathbf{0},$$

contro l'ipotesi che $B \neq \mathbf{0}$.

D'altra parte, se ripensiamo all'isomorfismo esistente tra $\text{End}(V)$ e $M_n(K)$, ove V è uno spazio vettoriale di dimensione n con una base fissata, notiamo che affermare che una matrice A sia invertibile equivale ad affermare che la corrispondente funzione lineare $f : V \rightarrow V$ sia invertibile, ma ciò è vero se e solo se f è biiettiva, cioè se e solo se f è un isomorfismo.

Ricordiamo che il sottoinsieme di $\text{End}(V)$ costituito dalle funzioni lineari invertibili (cioè dagli isomorfismi) $f : V \rightarrow V$, è stato indicato con $\text{Aut}(V)$.

Il corrispondente sottoinsieme di $M_n(K)$, costituito dalle matrici associate a elementi di $\text{Aut}(V)$, cioè dalle matrici invertibili, sarà indicato con $\text{GL}(n, K)$, e detto il *gruppo generale lineare* di ordine n a coefficienti in K . Esso è infatti un gruppo (non commutativo), rispetto all'operazione di prodotto tra matrici.

2.2.2 Cambiamenti di base

Abbiamo più volte fatto notare che la matrice associata a una funzione lineare $f : V \rightarrow W$ dipende dalla scelta di una base dello spazio vettoriale V e di una base di W : cambiando scelta delle basi cambia anche la matrice associata a f . In questa sezione ci proponiamo di scoprire in che modo cambia la matrice di f se cambiamo la nostra scelta delle basi di V e W .

Siano dunque V e W due spazi vettoriali su K , di dimensioni rispettivamente n e m e sia $f : V \rightarrow W$ un'applicazione lineare. Siano $\mathbf{v} = \{v_1, \dots, v_n\}$ e $\mathbf{v}' = \{v'_1, \dots, v'_n\}$ due basi di V e siano $\mathbf{w} = \{w_1, \dots, w_m\}$ e $\mathbf{w}' = \{w'_1, \dots, w'_m\}$ due basi di W . Infine, indichiamo con $A = (a_{ij})$ la matrice di f rispetto alle basi \mathbf{v} e \mathbf{w} e con $A' = (a'_{ij})$ la matrice di f rispetto alle basi \mathbf{v}' e \mathbf{w}' . Ricordiamo che ciò significa che

$$f(v_j) = \sum_{i=1}^m a_{ij} w_i, \quad \text{e} \quad f(v'_j) = \sum_{i=1}^m a'_{ij} w'_i,$$

per ogni $j = 1, \dots, n$.

Indichiamo con $\alpha_{\mathbf{v}} : V \xrightarrow{\sim} K^n$ l'isomorfismo che associa ad ogni vettore $v \in V$ la n -upla $(\lambda_1, \dots, \lambda_n)$ delle sue coordinate rispetto alla base \mathbf{v} e con $\alpha_{\mathbf{v}'} : V \xrightarrow{\sim} K^n$ l'isomorfismo che associa ad ogni $v \in V$ la n -upla $(\lambda'_1, \dots, \lambda'_n)$ delle sue coordinate rispetto alla base \mathbf{v}' .

Indichiamo analogamente con $\beta_{\mathbf{w}} : W \xrightarrow{\sim} K^m$ l'isomorfismo che associa ad ogni vettore $w \in W$ la m -upla (μ_1, \dots, μ_m) delle sue coordinate rispetto alla base \mathbf{w} e con $\beta_{\mathbf{w}'} : W \xrightarrow{\sim} K^m$ l'isomorfismo che associa ad ogni $w \in W$ la m -upla (μ'_1, \dots, μ'_m) delle sue coordinate rispetto alla base \mathbf{w}' .

Componendo $\alpha_{\mathbf{v}'}$ con l'inverso dell'isomorfismo $\alpha_{\mathbf{v}}$ otteniamo un isomorfismo di K^n in sé, il quale corrisponde alla moltiplicazione per una qualche matrice $P \in M_n(K)$. Indicheremo questo isomorfismo con $F_P : K^n \xrightarrow{\sim} K^n$. Si ottiene così il seguente diagramma commutativo:

$$\begin{array}{ccc} & V & \\ \alpha_{\mathbf{v}} \swarrow & & \searrow \alpha_{\mathbf{v}'} \\ K^n & \xrightarrow{F_P} & K^n \end{array}$$

Analogamente, componendo $\beta_{\mathbf{w}'}$ con l'inverso dell'isomorfismo $\beta_{\mathbf{w}}$ otteniamo un isomorfismo di K^m in sé, il quale corrisponde alla moltiplicazione per una qualche matrice $Q \in M_m(K)$. Indicheremo questo isomorfismo con $F_Q : K^m \xrightarrow{\sim} K^m$. Si ottiene così il seguente diagramma commutativo:

$$\begin{array}{ccc} & W & \\ \beta_{\mathbf{w}} \swarrow & & \searrow \beta_{\mathbf{w}'} \\ K^m & \xrightarrow{F_Q} & K^m \end{array}$$

Facciamo notare che le due matrici P e Q sono invertibili, dato che le corrispondenti applicazioni lineari F_P e F_Q sono degli isomorfismi.

Vediamo ora di ottenere una descrizione più esplicita delle matrici P e Q . Cominciamo dalla matrice P , la quale corrisponde all'isomorfismo

$$F_P : K^n \xrightarrow{\sim} K^m.$$

Abbiamo già osservato che le colonne di P sono date dalle immagini dei vettori della base canonica di K^n . Sia $e_j = {}^t(0, \dots, 0, 1, 0, \dots, 0)$ il j -esimo vettore della base canonica di K^n (tutte le coordinate sono nulle tranne la j -esima che è uguale a 1). Tramite l'isomorfismo $\alpha_{\mathbf{v}}$, il vettore $e_j \in K^n$ corrisponde al j -esimo vettore v_j della base \mathbf{v} di V . Si ha quindi

$$F_P(e_j) = \alpha_{\mathbf{v}'}(\alpha_{\mathbf{v}}^{-1}(e_j)) = \alpha_{\mathbf{v}'}(v_j),$$

dove ricordiamo che $\alpha_{\mathbf{v}'}(v_j) \in K^n$ è il vettore costituito dalle coordinate del vettore v_j calcolate rispetto alla base \mathbf{v}' ; questo vettore è la j -esima colonna di P .

In conclusione, possiamo affermare che le *colonne* della matrice P non sono altro che le coordinate dei vettori v_1, \dots, v_n della base \mathbf{v} di V calcolate rispetto alla seconda base \mathbf{v}' . Con un analogo ragionamento, scambiando i ruoli delle due basi, si potrebbe dimostrare che le colonne della matrice inversa P^{-1} sono precisamente le coordinate dei vettori v'_1, \dots, v'_n della base \mathbf{v}' di V calcolate rispetto alla prima base \mathbf{v} .

In modo del tutto analogo si dimostra che la j -esima colonna della matrice Q è costituita dal vettore delle coordinate del j -esimo vettore w_j della base \mathbf{w} calcolate rispetto alla base \mathbf{w}' . In altre parole, la matrice Q è la matrice le cui *colonne* sono date dalle coordinate dei vettori w_1, \dots, w_m della base \mathbf{w} di W calcolate rispetto alla seconda base \mathbf{w}' . Analogamente si dimostra che le colonne della matrice inversa Q^{-1} sono le coordinate dei vettori w'_1, \dots, w'_m della base \mathbf{w}' di W calcolate rispetto alla prima base \mathbf{w} .

Ricordando il risultato enunciato nella Proposizione 2.2.13, possiamo riasumere quanto detto finora nel seguente diagramma commutativo

$$\begin{array}{ccccc}
 K^n & \xrightarrow{F_A} & K^m & & \\
 \alpha_{\mathbf{v}} \swarrow & & \nearrow \beta_{\mathbf{w}} & & \\
 F_P \downarrow \wr & V \xrightarrow{f} & W & \downarrow \wr F_Q & \\
 \alpha_{\mathbf{v}'} \swarrow & & \nearrow \beta_{\mathbf{w}'} & & \\
 K^n & \xrightarrow{F_{A'}} & K^m & &
 \end{array}$$

ove F_A e $F_{A'}$ sono le applicazioni lineari date dalla moltiplicazione per A e per A' , rispettivamente.

Dalla commutatività di questo diagramma si deduce che

$$F_{A'} \circ F_P = F_Q \circ F_A,$$

che equivale alla seguente uguaglianza tra matrici

$$A'P = QA.$$

Da ciò segue che

$$A' = QAP^{-1} \quad \text{e} \quad A = Q^{-1}A'P. \quad (2.2.5)$$

Queste due espressioni equivalenti permettono di determinare la matrice A' di un'applicazione lineare $f : V \rightarrow W$ rispetto alle basi \mathbf{v}' di V e \mathbf{w}' di W quando è nota la matrice A di f rispetto a delle basi \mathbf{v} e \mathbf{w} e quando sono note le matrici di cambiamento di base P e Q .

Nel caso particolare in cui $W = V$, cioè quando f è un endomorfismo di uno spazio vettoriale V , il diagramma commutativo precedente si riduce al seguente

$$\begin{array}{ccccc}
 K^n & \xrightarrow{F_A} & K^n & & \\
 \downarrow F_P & \swarrow \alpha_{\mathbf{v}} & \uparrow \alpha_{\mathbf{v}} & & \downarrow F_P \\
 V & \xrightarrow{f} & V & & \\
 \downarrow \iota & \swarrow \alpha_{\mathbf{v}'} & \uparrow \alpha_{\mathbf{v}'} & & \downarrow \iota \\
 K^n & \xrightarrow{F_{A'}} & K^n & &
 \end{array}$$

e le uguaglianze (2.2.5) diventano

$$A' = PAP^{-1} \quad \text{e} \quad A = P^{-1}A'P. \quad (2.2.6)$$

Diamo ora la seguente definizione:

Definizione 2.2.22. Due matrici quadrate A e A' di ordine n a coefficienti in K si dicono *simili* se esiste una matrice invertibile $P \in M_n(K)$ (cioè $P \in \text{GL}_n(K)$) tale che

$$A' = PAP^{-1}$$

o, equivalentemente,

$$A = P^{-1}A'P.$$

Da quanto sopra detto si deduce il seguente risultato:

Corollario 2.2.23. Due matrici $A, A' \in M_n(K)$ rappresentano lo stesso endomorfismo f di uno spazio vettoriale V di dimensione n su K , rispetto a basi diverse, se e solo se sono simili.

Osservazione 2.2.24. Si noti che la relazione di similitudine è una relazione di equivalenza sull'insieme $M_n(K)$ delle matrici quadrate di ordine n a coefficienti in K .

Esercizi

Esercizio 2.2.1. Sia $f : V \rightarrow W$ un'applicazione lineare tra due spazi vettoriali. Siano $\{v_1, v_2, v_3\}$ una base di V e $\{w_1, w_2, w_3, w_4\}$ una base di W , e f sia data da $f(v_1) = 2w_1 - 3w_2 + w_4$, $f(v_2) = w_2 - 2w_3 + 3w_4$ e $f(v_3) = w_1 + w_2 + w_3 - 3w_4$. Si scriva la matrice di f nelle basi date.

Esercizio 2.2.2. Siano V e W due spazi vettoriali di basi rispettivamente $\{v_1, v_2, v_3\}$ e $\{w_1, w_2\}$, e sia $f : V \rightarrow W$ un'applicazione lineare di matrice (rispetto alle basi date)

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

- (1) Si prenda per V la nuova base $v'_1 = v_2 + v_3$, $v'_2 = v_1 + v_3$, $v'_3 = v_1 + v_2$. Qual è la nuova matrice A' di f rispetto alle basi $\{v'_1, v'_2, v'_3\}$ e $\{w_1, w_2\}$?
- (2) Si prenda per W la nuova base $w'_1 = \frac{1}{2}(w_1 + w_2)$ e $w'_2 = \frac{1}{2}(w_1 - w_2)$. Qual è la matrice A'' di f rispetto alle basi $\{v'_1, v'_2, v'_3\}$ e $\{w'_1, w'_2\}$?

Esercizio 2.2.3. Si consideri il sottospazio V di $C^\infty(\mathbb{R})$ generato dalle funzioni $f_1(x) = e^{2x} + \cos x$, $f_2(x) = \cos x + \sin x$ e $f_3(x) = \sin x$. Si dimostri che f_1 , f_2 e f_3 sono linearmente indipendenti e si determini la matrice (rispetto alla base $\{f_1, f_2, f_3\}$) dell'endomorfismo di V che a una funzione associa la sua derivata.

Esercizio 2.2.4. Si determinino le matrici, rispetto alle basi canoniche, di tutte le applicazioni lineari $f : \mathbb{R}^3 \rightarrow \mathbb{R}^4$ tali che $f(1, 2, -1) = (0, 1, 0, 1)$, $f(3, -1, 2) = (1, 2, 0, -1)$ e $f(-1, 5, -4) = (2, 0, 3, 2)$.

Esercizio 2.2.5. Si determinino le matrici, rispetto alle basi canoniche, di tutte le applicazioni lineari $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ tali che $f(0, -2, 1) = (3, -1)$, $f(1, 1, -2) = (1, 2)$ e $f(2, -4, -1) = (11, 1)$.

Esercizio 2.2.6. Sia V l'insieme delle funzioni polinomiali a coefficienti reali di grado ≤ 4 che si annullano in 0 e 1, e sia W l'insieme delle funzioni polinomiali a coefficienti reali di grado ≤ 3 tali che il loro integrale tra 0 e 1 è nullo.

- (1) Si dimostri che V e W sono due spazi vettoriali e se ne determinino delle basi.
- (2) Sia $D : V \rightarrow W$ l'applicazione lineare che associa a una funzione la sua derivata. Si dimostri che D è ben definita e si determini una sua matrice rispetto alle basi precedentemente trovate.

Esercizio 2.2.7. Sia $\phi_\lambda : \mathbb{R}^3 \rightarrow \mathbb{R}^4$ l'omomorfismo di matrice (rispetto alle basi canoniche)

$$A_\lambda = \begin{pmatrix} 1 & \lambda & 0 \\ 0 & \lambda & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

- (1) È vero o falso che, per ogni $\lambda \in \mathbb{R}$, esiste un omomorfismo $\psi : \mathbb{R}^4 \rightarrow \mathbb{R}^3$ tale che $\psi \circ \phi_\lambda$ sia suriettivo?
- (2) Per quali valori di λ esistono $x, y, z \in \mathbb{R}$ tali che, posto

$$B = \begin{pmatrix} 1 & x & 0 & 0 \\ 0 & y & 0 & 0 \\ -1 & z & 1 & 0 \end{pmatrix}$$

si abbia $BA_\lambda = \mathbf{1}$?

Esercizio 2.2.8. Siano V e W due spazi vettoriali, con basi rispettivamente date da $\{v_1, v_2, v_3, v_4\}$ e $\{w_1, w_2, w_3\}$. Si determini la matrice, rispetto alle basi date, dell'applicazione lineare $\phi : V \rightarrow W$ definita da $\phi(v_1) = w_1 - w_2$, $\phi(v_2) = 2w_2 - 6w_3$, $\phi(v_3) = -2w_1 + 2w_2$, $\phi(v_4) = w_2 - 3w_3$. Si determinino inoltre le dimensioni di $\text{Ker } \phi$ e di $\text{Im } \phi$ e si scrivano delle basi di tali sottospazi. Si dica inoltre se $w_1 + w_2 + w_3 \in \text{Im } \phi$.

Esercizio 2.2.9. Si dica se l'endomorfismo di \mathbb{R}^3 definito da

$$f(x, y, z) = (x + 2y, y + z, 2z - x)$$

è iniettivo o suriettivo. Si determinino delle basi di $\text{Ker } f$ e di $\text{Im } f$ e si dica se la somma del nucleo e dell'immagine di f è diretta.

Esercizio 2.2.10. Sia $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ l'endomorfismo definito ponendo

$$\begin{aligned} f(1, 0, 0) &= (2, -1, 0) \\ f(0, 1, 0) &= (1, -1, 1) \\ f(0, 1, -1) &= (0, 2, 2). \end{aligned}$$

Si determini la matrice di f rispetto alla base canonica di \mathbb{R}^3 . Si determinino inoltre le dimensioni del nucleo e dell'immagine di f e delle basi di tali sottospazi.

Esercizio 2.2.11. Sia $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ l'endomorfismo di matrice

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

rispetto alla base canonica. Si determini il rango di f e delle basi di $\text{Ker } f$ e di $\text{Im } f$.

Esercizio 2.2.12. Sia V uno spazio vettoriale di dimensione finita. Si dica sotto quali condizioni su V esiste un endomorfismo $\phi : V \rightarrow V$ tale che $\text{Ker } \phi = \text{Im } \phi$.

Esercizio 2.2.13. Si determini il rango della matrice

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

Esercizio 2.2.14. Si determini, al variare di $a \in \mathbb{R}$, il rango della matrice

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

2.3 Sistemi lineari

Riprendiamo ora lo studio dei sistemi di equazioni lineari, alla luce di ciò che abbiamo appreso al riguardo delle applicazioni lineari e delle matrici.

Dato un sistema di m equazioni lineari in n incognite

$$S : \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

indichiamo con $A = (a_{ij})$ la matrice costituita dei coefficienti del sistema, con $X = {}^t(x_1, \dots, x_n)$ il vettore *colonna* costituito dalle n incognite e con $B = {}^t(b_1, \dots, b_m)$ il vettore *colonna* dei termini noti. Ricordando la definizione del prodotto di una matrice per un vettore, è immediato verificare che il sistema S è equivalente alla seguente equazione:

$$AX = B,$$

cioè

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Se indichiamo con $F_A : K^n \rightarrow K^m$ l'applicazione lineare definita da $F_A(X) = AX$, per ogni $X \in K^n$ (F_A è l'applicazione lineare la cui matrice, rispetto alle basi canoniche di K^n e K^m è A), l'equazione $AX = B$ equivale a $F_A(X) = B$. Da ciò si deduce che l'insieme delle soluzioni del sistema S non è altro che l'antiimmagine tramite F_A del vettore $B \in K^m$:

$$F_A^{-1}(B) = \{X \in K^n \mid AX = B\}.$$

Osservazione 2.3.1. Più in generale, data una funzione lineare $f : V \rightarrow W$ tra due spazi vettoriali V e W , di dimensioni rispettivamente n e m sul campo K , possiamo considerare il seguente problema: dato un vettore $w \in W$, determinare la sua antiimmagine $f^{-1}(w) \subseteq V$, cioè determinare tutti i vettori $v \in V$ tali che $f(v) = w$.

Se scegliamo delle basi $\mathbf{v} = \{v_1, \dots, v_n\}$ di V e $\mathbf{w} = \{w_1, \dots, w_m\}$ di W , all'omomorfismo f risulta associata una matrice $A = (a_{ij})$, con m righe e n colonne, a coefficienti in K , con la proprietà che, per ogni vettore

$$v = \lambda_1 v_1 + \dots + \lambda_n v_n \in V,$$

se esprimiamo $f(v)$ come combinazione lineare dei vettori della base \mathbf{w}

$$f(v) = \mu_1 w_1 + \dots + \mu_m w_m,$$

allora si ha

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{pmatrix}$$

Pertanto, se indichiamo con $(b_1, \dots, b_m) \in K^m$ il vettore delle coordinate di w rispetto alla base \mathbf{w} di W fissata e se indichiamo con $(x_1, \dots, x_n) \in K^n$ le coordinate di un generico vettore $v \in V$ (rispetto alla base \mathbf{v} di V fissata), il problema di determinare i vettori v tali che $f(v) = w$ si traduce nel problema di determinare le soluzioni del seguente sistema di equazioni lineari:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Ricordiamo che un sistema lineare $AX = B$ è detto omogeneo se $B = \mathbf{0}$.

Proposizione 2.3.2. *L'insieme delle soluzioni di un sistema lineare omogeneo di m equazioni in n incognite a coefficienti in K è un sottospazio vettoriale di K^n .*

Dimostrazione. È sufficiente osservare che l'insieme delle soluzioni di un sistema del tipo $AX = \mathbf{0}$ non è altro che il nucleo della funzione lineare $F_A : K^n \rightarrow K^m$, il quale è un sottospazio vettoriale di K^n (vedi Proposizione 2.1.9). Possiamo tuttavia fornire anche una dimostrazione diretta.

Siano X_1 e X_2 due soluzioni del sistema $AX = \mathbf{0}$. Per ogni $\lambda_1, \lambda_2 \in K$, si ha

$$A(\lambda_1 X_1 + \lambda_2 X_2) = \lambda_1 AX_1 + \lambda_2 AX_2 = \lambda_1 \mathbf{0} + \lambda_2 \mathbf{0} = \mathbf{0},$$

quindi anche $\lambda_1 X_1 + \lambda_2 X_2$ è una soluzione del sistema in questione. Ciò significa precisamente che l'insieme delle soluzioni del sistema $AX = \mathbf{0}$ è un sottospazio vettoriale di K^n . \square

Nel caso di sistemi non omogenei, si ha:

Proposizione 2.3.3. *Ogni soluzione del sistema lineare non omogeneo*

$$S : AX = B$$

può essere espressa come somma di una soluzione particolare di S con una soluzione del sistema omogeneo associato. In altri termini, se indichiamo con Σ_B l'insieme delle soluzioni di S e con Σ_0 l'insieme delle soluzioni del sistema omogeneo associato $AX = \mathbf{0}$, si ha $\Sigma_B = \emptyset$ (se S non ammette soluzioni), oppure

$$\Sigma_B = \bar{X} + \Sigma_0 = \{\bar{X} + Y \mid Y \in \Sigma_0\}, \quad (2.3.1)$$

ove \bar{X} è una qualsiasi soluzione di S .

Dimostrazione. Sia $F_A : K^n \rightarrow K^m$ l'applicazione lineare di matrice A (rispetto alle basi canoniche di K^n e K^m). Allora si ha $\Sigma_B = F_A^{-1}(B)$ e $\Sigma_0 = F_A^{-1}(\mathbf{0}) = \text{Ker } F_A$. L'uguaglianza (2.3.1) discende allora dalla Proposizione 2.1.11. \square

Siamo ora in grado di determinare delle condizioni che garantiscono l'esistenza di soluzioni di un sistema di equazioni lineari.

Proposizione 2.3.4. *Sia $S : AX = B$ un sistema di m equazioni lineari in n incognite. Sia $F_A : K^n \rightarrow K^m$ la funzione lineare data da $F_A(X) = AX$, per ogni $X \in K^n$. Le condizioni seguenti sono equivalenti:*

- (i) Il sistema S ammette soluzioni;
- (ii) $B \in \text{Im}(F_A)$;
- (iii) Il vettore B è combinazione lineare delle colonne di A ;
- (iv) Il rango della matrice A è uguale al rango della matrice completa¹ $(A|B)$, ove quest'ultima è la matrice ottenuta aggiungendo ad A la colonna B dei termini noti.

Dimostrazione. Dato che l'insieme delle soluzioni di S coincide con $F_A^{-1}(B)$, l'equivalenza di (i) e (ii) discende dal fatto che $F_A^{-1}(B) \neq \emptyset$ se e solo se $B \in \text{Im}(F_A)$.

¹La matrice A è spesso chiamata la *matrice incompleta* del sistema lineare, mentre la matrice $(A|B)$ ottenuta aggiungendo alla matrice A la colonna B dei termini noti è detta la *matrice completa* del sistema.

Per dimostrare l'equivalenza di (ii) e (iii) è sufficiente ricordare che il sottospazio $\text{Im}(F_A)$ di K^m è generato dalle colonne di A . Pertanto $B \in \text{Im}(F_A)$ se e solo se B è combinazione lineare delle colonne di A .

Dimostriamo ora che (iii) \Rightarrow (iv). A tal fine basta ricordare che il rango della matrice A (che coincide con la dimensione di $\text{Im}(F_A)$) è il numero massimo di colonne linearmente indipendenti di A . Pertanto, se la colonna B è combinazione lineare delle colonne di A , l'aggiunta di B alla matrice A non ne altera il rango. Si ha quindi $\text{rk}(A) = \text{rk}(A|B)$.

Viceversa, il fatto che le matrici A e $(A|B)$ abbiano lo stesso rango significa che l'aggiunta della colonna B alla matrice A non ne ha modificato il rango, quindi B deve essere combinazione lineare delle colonne di A . Ciò dimostra che (iv) \Rightarrow (iii). \square

Quanto visto finora ci consente di dimostrare il seguente teorema:

Teorema 2.3.5 (TEOREMA DI ROUCHÉ-CAPELLI). *Sia $S : AX = B$ un sistema di m equazioni lineari in n incognite. S ammette soluzioni se e solo se $\text{rk}(A) = \text{rk}(A|B)$. In tal caso, se indichiamo con r il valore comune dei ranghi delle due matrici, si ha:*

- (i) se $r = n$ il sistema ammette un'unica soluzione;
- (ii) se $r < n$ il sistema ammette infinite soluzioni, le quali dipendono da $n - r$ parametri liberi di variare (si suole anche dire che S ammette ∞^{n-r} soluzioni).

Dimostrazione. Che l'uguaglianza tra i ranghi delle matrici A e $(A|B)$ sia una condizione necessaria e sufficiente per la risolubilità del sistema S è stato dimostrato nella proposizione precedente. Inoltre, nella Proposizione 2.3.3 abbiamo visto che l'insieme delle soluzioni di S è dato da

$$\bar{X} + \Sigma_0 = \{\bar{X} + Y \mid Y \in \Sigma_0\},$$

ove \bar{X} è una soluzione particolare di S e Σ_0 è l'insieme delle soluzioni del sistema omogeneo $AX = \mathbf{0}$, associato a S . Si ha dunque $\Sigma_0 = \text{Ker}(F_A)$, ove $F_A : K^n \rightarrow K^m$ è la funzione lineare definita da $F_A(X) = AX$. Poiché $r = \text{rk}(A) = \dim(\text{Im } F_A)$, dalla Proposizione 2.1.12 segue che

$$\dim \Sigma_0 = n - r.$$

Se $r = n$ si ha dunque $\Sigma_0 = \{\mathbf{0}\}$, quindi S possiede l'unica soluzione \bar{X} . Se invece $r < n$, il sottospazio vettoriale Σ_0 ha dimensione positiva, pari a $n - r$. Ciò significa che i suoi elementi possono essere descritti come combinazioni lineari di $n - r$ vettori di base. In una tale combinazione lineare compaiono quindi $n - r$ coefficienti i quali possono assumere qualunque valore nel campo K . \square

2.3.1 Risoluzione di un sistema lineare: il metodo dell'eliminazione (o metodo di Gauss)

Il metodo di risoluzione di un sistema lineare che ora descriveremo, noto come *metodo dell'eliminazione di Gauss*, si basa sull'osservazione che determinate manipolazioni algebriche, quali scambiare tra loro due equazioni, moltiplicare

entrambi i membri di un'equazione per una stessa costante diversa da zero, sommare o sottrarre membro a membro due equazioni o, più in generale, sommare a un'equazione un multiplo di un'altra, trasformano un dato sistema lineare in uno ad esso equivalente, cioè in un nuovo sistema avente le stesse soluzioni di quello precedente.

L'idea è dunque quella di utilizzare le operazioni sopra descritte (note anche col nome di *operazioni elementari*) per trasformare un sistema di equazioni lineari in sistemi, via via più semplici, ad esso equivalenti. Cercheremo ora di descrivere sommariamente come questa idea possa essere effettivamente realizzata.

Consideriamo un sistema di m equazioni lineari in n incognite

$$S : \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \quad \vdots \quad \ddots \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

Indichiamo con $A = (a_{ij})$ la matrice dei coefficienti e con $(A|B)$ la matrice completa del sistema, ottenuta aggiungendo ad A la colonna B dei termini noti.

Se la prima colonna della matrice A è interamente nulla, l'incognita x_1 non compare effettivamente nel sistema S . In tal caso passiamo alla colonna (cioè all'incognita) successiva. In caso contrario scegliamo una riga di A in cui il coefficiente dell'incognita x_1 sia diverso da zero. Supponiamo si tratti della riga i -esima: si ha dunque $a_{i1} \neq 0$. Possiamo quindi dividere ambo i membri della i -esima equazione per a_{i1} , e successivamente scambiare la i -esima equazione con la prima. Si ottiene così un nuovo sistema, equivalente a quello dato, in cui la matrice completa è del tipo

$$\left(\begin{array}{cccc|c} 1 & a'_{12} & \cdots & a'_{1n} & b'_1 \\ a'_{21} & a'_{22} & \cdots & a'_{2n} & b'_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a'_{m1} & a'_{m2} & \cdots & a'_{mn} & b'_m \end{array} \right)$$

A questo punto, per ogni $i \geq 2$, sostituiamo la i -esima riga di questa matrice (cioè la i -esima equazione del sistema) con la somma della riga in questione e della prima riga moltiplicata per $-a'_{i1}$, ottenendo così un nuovo sistema la cui matrice completa è del tipo

$$\left(\begin{array}{cccc|c} 1 & a'_{12} & \cdots & a'_{1n} & b'_1 \\ 0 & a''_{22} & \cdots & a''_{2n} & b''_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & a''_{m2} & \cdots & a''_{mn} & b''_m \end{array} \right)$$

Indichiamo con S' il sottosistema ottenuto trascurando la prima equazione. Nel sistema S' non compare più l'incognita x_1 . Ora possiamo ripetere la procedura sopra descritta al sistema S' .

Alla fine di questo procedimento otterremo un sistema, equivalente a quello iniziale, la cui matrice completa è nella cosiddetta "forma a scala," cioè in una forma in cui, in ciascuna riga, il primo coefficiente diverso da zero (nel

nostro caso specifico tale coefficiente è uguale a 1) si trova alla destra del primo coefficiente non nullo della riga precedente.

A questo punto il sistema può essere facilmente risolto partendo dall'ultima equazione tramite una semplice sostituzione all'indietro, analoga a quella impiegata nel metodo della sostituzione.

Vediamo ora di chiarire l'algoritmo appena descritto per mezzo di alcuni esempi concreti.

Esempio 1. (Sistema privo di soluzioni) Consideriamo il seguente sistema di equazioni lineari, a coefficienti nel campo \mathbb{Q} .

$$S : \begin{cases} 2x_1 - 4x_2 = -4 \\ 3x_1 - 6x_2 + 3x_3 = -3 \\ x_1 - 2x_2 - x_3 = -2 \end{cases}$$

La matrice completa di questo sistema è

$$\left(\begin{array}{ccc|c} 2 & -4 & 0 & -4 \\ 3 & -6 & 3 & -3 \\ 1 & -2 & -1 & -2 \end{array} \right)$$

Per far comparire un coefficiente uguale a 1 nella posizione (1, 1) della matrice possiamo operare in tre modi diversi: dividere la prima riga per 2, dividere la seconda riga per 3 e scambiarla con la prima, oppure scambiare tra loro la prima e la terza riga. Scegliamo quest'ultima possibilità, ottenendo la matrice

$$\left(\begin{array}{ccc|c} 1 & -2 & -1 & -2 \\ 3 & -6 & 3 & -3 \\ 2 & -4 & 0 & -4 \end{array} \right)$$

Ora dobbiamo far comparire degli zeri nella prima colonna, al di sotto del primo coefficiente. Per fare ciò sommiamo alla seconda riga la prima moltiplicata per -3 , e poi sommiamo alla terza riga la prima moltiplicata per -2 . Si ottiene così

$$\left(\begin{array}{ccc|c} 1 & -2 & -1 & -2 \\ 0 & 0 & 6 & 3 \\ 0 & 0 & 2 & 0 \end{array} \right)$$

Ora ricominciamo dalla seconda riga, dividendola per 6 in modo tale che il suo primo coefficiente non nullo sia 1.

$$\left(\begin{array}{ccc|c} 1 & -2 & -1 & -2 \\ 0 & 0 & 1 & 1/2 \\ 0 & 0 & 2 & 0 \end{array} \right)$$

Alla terza riga sommiamo quindi la seconda moltiplicata per -2 , ottenendo la matrice

$$\left(\begin{array}{ccc|c} 1 & -2 & -1 & -2 \\ 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & -1 \end{array} \right)$$

Questa matrice è finalmente nella forma che vogliamo (la forma a scala). Essa corrisponde al sistema

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

il quale è equivalente al sistema originario. Dato che la terza equazione è stata ridotta all'uguaglianza $0 = -1$, che non è verificata, concludiamo che tale sistema non ammette soluzioni.

Esempio 2. (Sistema che ammette un'unica soluzione) Consideriamo il seguente sistema di equazioni lineari, a coefficienti nel campo \mathbb{Q} .

$$S : \begin{cases} 2x_1 - 2x_2 + 8x_3 = 5 \\ 2x_2 + 6x_3 = 1 \\ x_1 - 2x_2 + 4x_3 = -1 \\ x_1 + 10x_3 = 0 \end{cases}$$

La matrice completa di questo sistema è

$$\left(\begin{array}{ccc|c} 2 & -2 & 8 & 5 \\ 0 & 2 & 6 & 1 \\ 1 & -2 & 4 & -1 \\ 1 & 0 & 10 & 0 \end{array} \right)$$

Scambiamo tra loro la prima e la terza riga:

$$\left(\begin{array}{ccc|c} 1 & -2 & 4 & -1 \\ 0 & 2 & 6 & 1 \\ 2 & -2 & 8 & 5 \\ 1 & 0 & 10 & 0 \end{array} \right)$$

Alla terza riga sommiamo la prima moltiplicata per -2 e alla quarta riga sottraiamo la prima:

$$\left(\begin{array}{ccc|c} 1 & -2 & 4 & -1 \\ 0 & 2 & 6 & 1 \\ 0 & 2 & 0 & 7 \\ 0 & 2 & 6 & 1 \end{array} \right)$$

Ora dividiamo per 2 la seconda riga:

$$\left(\begin{array}{ccc|c} 1 & -2 & 4 & -1 \\ 0 & 1 & 3 & 1/2 \\ 0 & 2 & 0 & 7 \\ 0 & 2 & 6 & 1 \end{array} \right)$$

Continuiamo sommando alla terza riga la seconda moltiplicata per -2 e sommando alla quarta riga la seconda moltiplicata per -2 . Si ottiene così:

$$\left(\begin{array}{ccc|c} 1 & -2 & 4 & -1 \\ 0 & 1 & 3 & 1/2 \\ 0 & 0 & -6 & 6 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

Ora dividiamo la terza riga per -6 :

$$\left(\begin{array}{ccc|c} 1 & -2 & 4 & -1 \\ 0 & 1 & 3 & 1/2 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

Questa matrice è nella forma che vogliamo. Essa corrisponde al sistema

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

il quale è equivalente al sistema originario. La soluzione di questo sistema si può ottenere facilmente, partendo dall'ultima equazione, tramite una sostituzione all'indietro: dalla terza equazione ricaviamo $x_3 = -1$ che, sostituito nella seconda, fornisce $x_2 = 1/2 - 3x_3 = 7/2$. Infine, sostituendo nella prima equazione i valori appena trovati, si ottiene $x_1 = -1 + 2x_2 - 4x_3 = 10$. In conclusione, il sistema dato ammette un'unica soluzione:

$$\begin{cases} x_1 = 10 \\ x_2 = 7/2 \\ x_3 = -1 \end{cases}$$

Esempio 3. (Sistema che ammette infinite soluzioni) Consideriamo il seguente sistema di equazioni lineari, a coefficienti nel campo \mathbb{Q} .

$$S : \begin{cases} 2x_1 + 5x_3 = 1 \\ 4x_1 - 3x_2 + 4x_3 = 5 \\ 2x_1 - x_2 + 3x_3 = 2 \end{cases}$$

La matrice completa di questo sistema è

$$\left(\begin{array}{ccc|c} 2 & 0 & 5 & 1 \\ 4 & -3 & 4 & 5 \\ 2 & -1 & 3 & 2 \end{array} \right)$$

Dividiamo la prima riga per 2:

$$\left(\begin{array}{ccc|c} 1 & 0 & 5/2 & 1/2 \\ 4 & -3 & 4 & 5 \\ 2 & -1 & 3 & 2 \end{array} \right)$$

Alla seconda riga sommiamo la prima moltiplicata per -4 e alla terza riga sommiamo la prima moltiplicata per -2 :

$$\left(\begin{array}{ccc|c} 1 & 0 & 5/2 & 1/2 \\ 0 & -3 & -6 & 3 \\ 0 & -1 & -2 & 1 \end{array} \right)$$

Ora dividiamo la seconda riga per -3 :

$$\left(\begin{array}{ccc|c} 1 & 0 & 5/2 & 1/2 \\ 0 & 1 & 2 & -1 \\ 0 & -1 & -2 & 1 \end{array} \right)$$

Infine, sommando alla terza riga la seconda, si ottiene:

$$\left(\begin{array}{ccc|c} 1 & 0 & 5/2 & 1/2 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

Questa matrice è nella forma che vogliamo. Essa corrisponde al sistema

$$S' : \begin{cases} x_1 + 5x_3/2 = 1/2 \\ x_2 + 2x_3 = -1 \\ 0 = 0 \end{cases}$$

il quale è equivalente al sistema originario. Si può così notare che il sistema ammette infinite soluzioni. Infatti, dalla seconda equazione si ricava $x_2 = -1 - 2x_3$ mentre dalla prima si ottiene $x_1 = \frac{1}{2} - \frac{5}{2}x_3$. L'incognita x_3 rimane indeterminata e può dunque assumere qualsiasi valore. In conclusione, il sistema dato ammette infinite soluzioni, dipendenti da un parametro (∞^1 soluzioni), che possono essere espresse nella forma

$$\begin{cases} x_1 = 1/2 - 5x_3/2 \\ x_2 = -1 - 2x_3 \\ x_3 \text{ qualsiasi.} \end{cases}$$

2.3.2 Calcolo del rango di una matrice

Sia A una matrice $m \times n$ a coefficienti nel campo K . Ricordiamo che il rango di A è stato definito come il massimo numero di colonne linearmente indipendenti della matrice A (vedi Definizione 2.2.15). Tale numero coincide con la dimensione dell'immagine della funzione lineare $F_A : K^n \rightarrow K^m$ la cui matrice, rispetto alle basi canoniche di K^n e K^m , è A .

In modo del tutto analogo, possiamo definire il *rango per righe* di A , come il massimo numero di righe linearmente indipendenti, cioè come la dimensione del sottospazio vettoriale di K^n generato dalle righe di A . A prima vista potrebbe sembrare che non vi sia alcun motivo per cui il rango per colonne di una matrice debba coincidere con il suo rango per righe. Tuttavia questi due numeri risultano essere sempre uguali, come dimostreremo in seguito (vedi Capitolo 3, Teorema 3.3.1).

Vedremo ora come il metodo di eliminazione di Gauss fornisca uno strumento molto utile per il calcolo del rango per righe di una matrice. Ciò discende dal fatto che le operazioni elementari sulle righe di una matrice utilizzate nel metodo di Gauss non alterano il numero di righe linearmente indipendenti. Di conseguenza, se applichiamo l'eliminazione di Gauss per trasformare una matrice A in una matrice A' che si trovi nella forma a scala, il rango per righe di A' sarà necessariamente uguale al rango per righe di A . Arrivati a questo punto, il calcolo del rango è immediato. Infatti vale il seguente risultato:

Proposizione 2.3.6. *Se una matrice A è nella forma a scala, il suo rango per righe coincide con il numero di righe non nulle.*

Dimostrazione. Supponiamo che la matrice A abbia m righe, di cui le prime r sono non nulle. Indichiamo con v_1, v_2, \dots, v_r i vettori riga non nulli di A , ove $v_i = (a_{i1}, a_{i2}, \dots, a_{in})$. Poiché A è nella forma a scala, il primo coefficiente non nullo di ciascuna riga v_i si trova alla destra del primo coefficiente non nullo della riga precedente v_{i-1} . Pertanto, se indichiamo con a_{1h_1} il primo coefficiente non nullo della riga v_1 , tutti gli altri elementi della colonna h_1 della matrice A , cioè gli elementi a_{jh_1} con $j \geq 2$, sono nulli. Consideriamo ora una combinazione

lineare dei vettori v_1, v_2, \dots, v_r :

$$\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_r v_r = \mathbf{0}.$$

Se concentriamo la nostra attenzione sulle componenti di posto h_1 , si trova

$$\lambda_1 a_{1h_1} + \lambda_2 0 + \dots + \lambda_r 0 = 0$$

da cui segue $\lambda_1 = 0$.

In modo analogo, indichiamo con a_{2h_2} il primo coefficiente non nullo della riga v_2 . Allora tutti gli elementi della colonna h_2 , dalla terza riga in poi, cioè gli elementi a_{jh_2} con $j \geq 3$, sono nulli. Dall'uguaglianza

$$\lambda_2 v_2 + \dots + \lambda_r v_r = \mathbf{0}$$

si ottiene, considerando solo le componenti di posto h_2 ,

$$\lambda_2 a_{2h_2} + \lambda_3 0 + \dots + \lambda_r 0 = 0$$

da cui segue $\lambda_2 = 0$.

Ripetendo il ragionamento sopra descritto si dimostra così che tutti i coefficienti $\lambda_1, \lambda_2, \dots, \lambda_r$ sono nulli. Questo prova che i vettori v_1, v_2, \dots, v_r sono linearmente indipendenti, quindi il numero r di righe non nulle della matrice A coincide con il numero di righe linearmente indipendenti, cioè con il rango per righe di A . \square

Esempio 2.3.7. Illustriamo su un esempio concreto il metodo appena descritto. Vogliamo calcolare il rango (per righe) della seguente matrice:

$$A = \begin{pmatrix} 0 & 2 & -1 & 6 & 1 \\ 3 & -4 & 14 & 3 & 2 \\ 1 & -2 & 5 & 3 & 3 \\ 1 & 0 & 4 & 3 & 0 \end{pmatrix}$$

Scambiamo la prima riga con la terza:

$$\begin{pmatrix} 1 & -2 & 5 & 3 & 3 \\ 3 & -4 & 14 & 3 & 2 \\ 0 & 2 & -1 & 6 & 1 \\ 1 & 0 & 4 & 3 & 0 \end{pmatrix}$$

Alla seconda riga sommiamo la prima moltiplicata per -3 e alla quarta riga sottraiamo la prima:

$$\begin{pmatrix} 1 & -2 & 5 & 3 & 3 \\ 0 & 2 & -1 & -6 & -7 \\ 0 & 2 & -1 & 6 & 1 \\ 0 & 2 & -1 & 0 & -3 \end{pmatrix}$$

Alla terza riga sottraiamo la seconda e alla quarta riga sottraiamo la seconda:

$$\begin{pmatrix} 1 & -2 & 5 & 3 & 3 \\ 0 & 2 & -1 & -6 & -7 \\ 0 & 0 & 0 & 12 & 8 \\ 0 & 0 & 0 & 6 & 4 \end{pmatrix}$$

Infine, alla quarta riga sommiamo la terza moltiplicata per $-1/2$:

$$\begin{pmatrix} 1 & -2 & 5 & 3 & 3 \\ 0 & 2 & -1 & -6 & -7 \\ 0 & 0 & 0 & 12 & 8 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Questa matrice è nella forma a scala. Poiché ci sono tre righe non nulle, si conclude che il suo rango (per righe) è 3. Questo è allora anche il rango della matrice A da cui eravamo partiti.

Come ulteriore esercizio si consiglia al lettore di applicare l'algoritmo appena descritto alla matrice trasposta di A , al fine di verificare che anche il rango per colonne è uguale a 3.

Osservazione 2.3.8. Volendo calcolare il rango per colonne di una matrice A è sufficiente applicare il metodo dell'eliminazione di Gauss alla matrice trasposta di A . In alternativa si può modificare il metodo di Gauss in modo da effettuare le operazioni elementari sulle colonne di A piuttosto che sulle sue righe. L'obiettivo è quello di portare la matrice data in un'appropriata "forma a scala," che non è altro che la trasposta della forma a scala descritta in precedenza. Si tratta cioè di una forma a scala in cui il primo elemento non nullo di ciascuna *colonna* si trova al di sotto del primo elemento non nullo della colonna precedente.

2.3.3 Calcolo dell'inversa di una matrice

Vogliamo ora descrivere un algoritmo, derivato dal metodo di eliminazione di Gauss, per il calcolo dell'inversa di una matrice quadrata. Tale algoritmo si basa sull'osservazione che effettuare delle operazioni elementari sulle righe di una matrice A equivale a moltiplicare A , a sinistra, per un'opportuna matrice invertibile.

Consideriamo, ad esempio, l'operazione elementare che consiste nello scambio di due righe della matrice A . Per ogni $i, j = 1, \dots, n$, con $i \neq j$, indichiamo con $P(i, j)$ la matrice i cui elementi p_{hk} (con $1 \leq h, k \leq n$) sono dati da:

$$p_{hk} = \begin{cases} 1 & \text{se } h = k, h \neq i, h \neq j, \\ 1 & \text{se } h = i \text{ e } k = j \text{ oppure se } h = j \text{ e } k = i, \\ 0 & \text{altrimenti.} \end{cases}$$

Se A è una matrice $n \times n$ e se poniamo $A' = P(i, j)A$, si verifica facilmente che la matrice A' è ottenuta dalla matrice A semplicemente scambiando tra loro le righe i -esima e j -esima. Ad esempio, se $n = 4$, $i = 1$, $j = 3$, si ha

$$P(1, 3) = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

e

$$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} = \begin{pmatrix} a_{31} & a_{32} & a_{33} & a_{34} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{11} & a_{12} & a_{13} & a_{14} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$$

Da quanto appena detto segue immediatamente che

$$P(i, j)^2 = P(i, j)P(i, j) = \mathbf{1},$$

quindi le matrici $P(i, j)$ sono invertibili.

Consideriamo ora la matrice $M(i; \lambda) = (m_{hk})$ definita, per ogni $i = 1, \dots, n$ e ogni $\lambda \in K$, $\lambda \neq 0$, ponendo

$$m_{hk} = \begin{cases} 1 & \text{se } h = k, h \neq i, \\ \lambda & \text{se } h = k = i, \\ 0 & \text{altrimenti.} \end{cases}$$

La moltiplicazione a sinistra di $M(i; \lambda)$ per una matrice A ha come effetto quello di moltiplicare la i -esima riga di A per λ . Ad esempio, se $n = 4$ e $i = 2$, si ha

$$M(2; \lambda) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

e

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ \lambda a_{21} & \lambda a_{22} & \lambda a_{23} & \lambda a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$$

Notiamo che, poiché si è supposto $\lambda \neq 0$, si ha

$$M(i; \lambda)M(i; \lambda^{-1}) = \mathbf{1},$$

quindi le matrici $M(i; \lambda)$ sono invertibili.

Infine, per ogni $i, j = 1, \dots, n$, con $i \neq j$, e ogni $\alpha \in K$, definiamo una matrice $S(i, j; \alpha) = (s_{hk})$ ponendo

$$s_{hk} = \begin{cases} 1 & \text{se } h = k, \\ \alpha & \text{se } h = i \text{ e } k = j, \\ 0 & \text{altrimenti.} \end{cases}$$

Se A è una matrice $n \times n$ e se poniamo $A' = S(i, j; \alpha)A$, è immediato verificare che la matrice A' è ottenuta dalla matrice A sommando alla i -esima riga la j -esima moltiplicata per α . Ad esempio, se $n = 4$, $i = 2$, $j = 4$, si ha

$$S(2, 4; \alpha) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \alpha \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

e

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \alpha \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} + \alpha a_{41} & a_{22} + \alpha a_{42} & a_{23} + \alpha a_{43} & a_{24} + \alpha a_{44} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$$

Si ha pertanto

$$S(i, j; \alpha)S(i, j; -\alpha) = \mathbf{1},$$

quindi le matrici $S(i, j; \alpha)$ sono invertibili.

Una successione di operazioni elementari sulle righe di una matrice A equivale dunque a una successione di moltiplicazioni, a sinistra, per delle matrici invertibili del tipo descritto in precedenza. Poiché il prodotto di un numero qualsiasi di matrici invertibili è ancora una matrice invertibile, concludiamo che l'effetto di un numero qualunque di operazioni elementari sulle righe di una matrice A può essere ottenuto semplicemente moltiplicando la matrice A , a sinistra, per un'opportuna matrice invertibile.

Osservazione 2.3.9. Si verifica facilmente che moltiplicare una matrice A a destra per le matrici $P(i, j)$, $M(i; \lambda)$ e $S(i, j; \alpha)$ descritte in precedenza equivale ad effettuare delle operazioni elementari sulle colonne di A . Più precisamente, se poniamo $A' = AP(i, j)$, la matrice A' è ottenuta scambiando tra loro la i -esima e la j -esima colonna di A , se $A' = AM(i; \lambda)$ allora A' è ottenuta moltiplicando per λ la i -esima colonna di A , e se $A' = AS(i, j; \alpha)$ allora A' è ottenuta dalla matrice A sommando alla i -esima colonna la j -esima colonna moltiplicata per α . Pertanto, l'effetto di un numero qualunque di operazioni elementari sulle colonne di A può essere ottenuto moltiplicando la matrice A , a destra, per un'opportuna matrice invertibile.

Supponiamo ora che la matrice quadrata A , di ordine n , sia invertibile. Tramite operazioni elementari sulle righe è possibile trasformare A in una matrice nella forma a scala, in cui il primo coefficiente non nullo di ciascuna riga può essere reso uguale a 1. Poiché A è invertibile, il suo rango deve essere n ,² quindi nella forma a scala non ci devono essere righe interamente nulle. Poiché A è una matrice quadrata di ordine n ciò equivale a dire che la forma a scala che otteniamo dopo l'applicazione dell'algoritmo di eliminazione di Gauss è una matrice A' , triangolare superiore, con tutti gli elementi sulla diagonale principale uguali a 1.

Arrivati a questo punto è facile convincersi che, mediante opportune operazioni elementari sulle righe della matrice A' , è possibile trasformare quest'ultima nella matrice identica $\mathbf{1}$. Se indichiamo con B la matrice che rappresenta l'effetto di tutte le operazioni elementari sulle righe che abbiamo eseguito per trasformare la matrice A nella matrice identica, si ha dunque $BA = \mathbf{1}$. La matrice B è pertanto l'inversa della matrice A . Essa può quindi essere determinata tenendo scrupolosamente conto di tutte le matrici corrispondenti alle operazioni elementari sulle righe che sono state effettuate.

Un metodo molto più efficace per determinare la matrice B è il seguente. Scriviamo a fianco della matrice A la matrice identica $\mathbf{1}$, in modo da ottenere una matrice con n righe e $2n$ colonne che indicheremo con $(A | \mathbf{1})$. In questo modo tutte le operazioni che eseguiremo sulle righe di A dovranno essere effettuate anche sulle righe della matrice $\mathbf{1}$. L'effetto di queste operazioni elementari è equivalente alla moltiplicazione a sinistra per la matrice (incognita) B . Si ha pertanto:

$$B(A | \mathbf{1}) = (BA | B\mathbf{1}) = (\mathbf{1} | B).$$

²Dire che A è invertibile equivale a dire che la corrispondente applicazione lineare $F_A : K^n \rightarrow K^n$ è un isomorfismo, il che significa che $\text{Im } F_A = K^n$, cioè che $\text{rk}(A) = \text{rk}(F_A) = n$.

Ciò significa che quando avremo trasformato la matrice A nella matrice identica, la matrice $\mathbf{1}$ scritta a destra di A sarà stata automaticamente trasformata nella matrice B , la quale non è altro che l'inversa di A .

A titolo di esempio, applichiamo l'algoritmo appena descritto per determinare l'inversa della matrice

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

Per prima cosa affianchiamo alla matrice A la matrice identica, ottenendo

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 & 1 & 0 \\ 1 & 2 & 2 & 0 & 0 & 1 \end{array} \right)$$

A questo punto, utilizzando operazioni elementari sulle righe, cerchiamo di trasformare la matrice A nella matrice identica. Se riusciamo a fare ciò, la matrice che troveremo a destra sarà la matrice inversa di A .

In dettaglio le operazioni da fare sono, ad esempio, le seguenti: sottraiamo alla seconda riga la prima, e alla terza riga la prima, ottenendo

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{array} \right)$$

Ora sommiamo alla seconda riga la terza, mentre alla prima sottraiamo la terza, ottenendo

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 2 & 0 & -1 \\ 0 & 1 & 0 & -2 & 1 & 1 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{array} \right)$$

Ora sommiamo alla prima riga la seconda moltiplicata per -2 , ottenendo

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 6 & -2 & -3 \\ 0 & 1 & 0 & -2 & 1 & 1 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{array} \right)$$

Si ha pertanto

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

Come esercizio si verifichi che la matrice appena trovata è effettivamente l'inversa di A , cioè che $AA^{-1} = \mathbf{1}$.

Osservazione 2.3.10. Se la matrice A di cui si cerca l'inversa non fosse invertibile essa avrebbe rango strettamente minore di n , quindi la sua forma a scala A' , ottenuta nella prima parte dell'algoritmo precedentemente descritto, avrebbe almeno una riga interamente nulla. A questo punto sapremmo che A non è invertibile.

Esercizi

Esercizio 2.3.1. Si risolvano gli esercizi proposti alla fine della Sezione 1.1 del Cap. 1 utilizzando il metodo dell'eliminazione di Gauss.

Esercizio 2.3.2. Si calcoli l'inversa della matrice

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

Capitolo 3

Determinanti

In questo capitolo definiremo la nozione di determinante di una matrice quadrata a coefficienti in un campo e studieremo le sue principali proprietà. Prima però, avremo bisogno di richiamare alcune proprietà elementari delle permutazioni di un insieme finito di elementi.

3.1 Permutazioni

Definizione 3.1.1. Sia $A = \{a_1, a_2, \dots, a_n\}$ un insieme di n elementi. Una *permutazione* degli elementi di A è una funzione biiettiva $\sigma : A \rightarrow A$.

Una tale permutazione σ può essere convenientemente rappresentata mediante una tabella del tipo

$$\begin{pmatrix} a_1 & a_2 & a_3 & \dots & a_n \\ a_{j_1} & a_{j_2} & a_{j_3} & \dots & a_{j_n} \end{pmatrix}$$

ove si conviene che $\sigma(a_i) = a_{j_i}$, per $i = 1, \dots, n$.

Ad esempio, se $n = 4$ e σ è la permutazione definita da $\sigma(a_1) = a_3$, $\sigma(a_2) = a_2$, $\sigma(a_3) = a_4$ e $\sigma(a_4) = a_1$, la tabella corrispondente sarà

$$\begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ a_3 & a_2 & a_4 & a_1 \end{pmatrix}$$

In una tale rappresentazione la presenza del simbolo a è, in effetti, del tutto superflua. È più conveniente rappresentare σ mediante la tabella

$$\begin{pmatrix} 1 & 2 & 3 & \dots & n \\ j_1 & j_2 & j_3 & \dots & j_n \end{pmatrix}$$

il che equivale a interpretare σ come una permutazione dell'insieme $\{1, 2, \dots, n\}$.

Indicheremo con \mathfrak{S}_n l'insieme delle permutazioni di n oggetti. Osserviamo che la composizione di due permutazioni è ancora una permutazione e lo stesso

vale per l'inversa di una permutazione. È immediato verificare che la composizione delle permutazioni definisce su \mathfrak{S}_n una struttura di gruppo. Notiamo inoltre che, se $n \geq 3$ e $\sigma, \tau \in \mathfrak{S}_n$, si ha, in generale $\sigma \circ \tau \neq \tau \circ \sigma$, quindi \mathfrak{S}_n è un gruppo non abeliano.

Proposizione 3.1.2. *La cardinalità di \mathfrak{S}_n è $n!$.*

Dimostrazione. Per definire una permutazione $\sigma \in \mathfrak{S}_n$ dobbiamo specificare $\sigma(i)$ per ogni $i = 1, \dots, n$. Per $\sigma(1)$ ci sono n possibili valori, mentre per $\sigma(2)$ i valori possibili sono solo $n-1$, dato che deve essere $\sigma(2) \neq \sigma(1)$. Analogamente, per $\sigma(3)$ i possibili valori sono solo $n-2$ (deve essere $\sigma(3) \neq \sigma(1)$ e $\sigma(3) \neq \sigma(2)$). Continuando in questo modo si conclude che i possibili valori che possiamo attribuire a $\sigma(j)$ sono $n-j+1$, per ogni $j = 1, \dots, n$. Il numero di possibili permutazioni è quindi dato dal prodotto $n(n-1)(n-2) \cdots 2 \cdot 1 = n!$. \square

Definizione 3.1.3. Consideriamo una permutazione $\sigma \in \mathfrak{S}_n$. Diremo che in σ è presente una *inversione* ogni volta si ha $i < j$ ma $\sigma(i) > \sigma(j)$.

Ad esempio, nella permutazione

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 5 & 1 & 4 \end{pmatrix}$$

ci sono 5 inversioni: si ha infatti $1 < 2$ ma $\sigma(1) = 3 > \sigma(2) = 2$, $1 < 4$ ma $\sigma(1) = 3 > \sigma(4) = 1$, $2 < 4$ ma $\sigma(2) = 2 > \sigma(4) = 1$, $3 < 4$ ma $\sigma(3) = 5 > \sigma(4) = 1$ e $3 < 5$ ma $\sigma(3) = 5 > \sigma(5) = 4$. Si noti che per contare il numero di inversioni di σ basta osservare che nella seconda riga della tabella il numero 3 compare prima dei numeri 2 e 1 (due inversioni), 2 viene prima di 1 (un'altra inversione) e infine il 5 precede i numeri 1 e 4 (altre due inversioni).

Definizione 3.1.4. Diremo che $\sigma \in \mathfrak{S}_n$ è una permutazione *pari* (risp. *dispari*) se il numero di inversioni presenti in σ è pari (risp. dispari). Definiamo inoltre il *segno* di σ , indicato con $\text{sgn}(\sigma)$, ponendo

$$\text{sgn}(\sigma) = \begin{cases} 1 & \text{se } \sigma \text{ è pari,} \\ -1 & \text{se } \sigma \text{ è dispari.} \end{cases}$$

In altri termini, si ha

$$\text{sgn}(\sigma) = (-1)^{i(\sigma)},$$

ove $i(\sigma)$ è il numero di inversioni presenti in σ .

Consideriamo ora il polinomio

$$P = \prod_{1 \leq i < j \leq n} (x_i - x_j)$$

nelle n indeterminate x_1, \dots, x_n . Una permutazione $\sigma \in \mathfrak{S}_n$ agisce su P trasformandolo nel polinomio

$$\sigma(P) = \prod_{1 \leq i < j \leq n} (x_{\sigma(i)} - x_{\sigma(j)}).$$

Ad esempio, se $n = 4$ si ha

$$P = (x_1 - x_2)(x_1 - x_3)(x_1 - x_4)(x_2 - x_3)(x_2 - x_4)(x_3 - x_4).$$

La permutazione

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix}$$

trasforma P nel polinomio

$$\sigma(P) = (x_2 - x_4)(x_2 - x_3)(x_2 - x_1)(x_4 - x_3)(x_4 - x_1)(x_3 - x_1).$$

Poiché in P compaiono tutti i fattori del tipo $x_i - x_j$, per ogni $i < j$, e poiché una permutazione σ scambia le indeterminate x_i tra loro, ogni fattore $x_i - x_j$ viene trasformato in $x_{\sigma(i)} - x_{\sigma(j)}$ il quale coincide, a meno del segno, con uno dei fattori presenti in P . Più precisamente, se $\sigma(i) < \sigma(j)$, il fattore $x_{\sigma(i)} - x_{\sigma(j)}$ compare in P , se invece $\sigma(i) > \sigma(j)$ allora è il fattore $x_{\sigma(j)} - x_{\sigma(i)} = -(x_{\sigma(i)} - x_{\sigma(j)})$ che compare in P . Si conclude pertanto che, per ogni permutazione σ , si ha $\sigma(P) = \pm P$. Più precisamente, quanto sopra detto mostra che ogni inversione presente in σ corrisponde a un fattore -1 che moltiplica P , quindi si ha

$$\sigma(P) = (-1)^{i(\sigma)} P,$$

ove $i(\sigma)$ denota, come sopra detto, il numero di inversioni presenti in σ . Poiché $(-1)^{i(\sigma)}$ è precisamente il segno di σ , si ha:

$$\sigma(P) = \operatorname{sgn}(\sigma) P. \quad (3.1.1)$$

Siamo ora in grado di dimostrare il seguente risultato:

Proposizione 3.1.5. *Siano $\sigma, \tau \in \mathfrak{S}_n$ e consideriamo la loro composizione $\sigma \circ \tau$. Si ha*

$$\operatorname{sgn}(\sigma \circ \tau) = \operatorname{sgn}(\sigma) \operatorname{sgn}(\tau).$$

Si ha inoltre

$$\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma),$$

per ogni $\sigma \in \mathfrak{S}_n$.

Dimostrazione. Applichiamo prima la permutazione τ e poi σ al polinomio P . Si ottiene:

$$\tau(P) = \prod_{1 \leq i < j \leq n} (x_{\tau(i)} - x_{\tau(j)}),$$

quindi

$$\begin{aligned} \sigma(\tau(P)) &= \prod_{1 \leq i < j \leq n} (x_{\sigma(\tau(i))} - x_{\sigma(\tau(j))}) \\ &= \prod_{1 \leq i < j \leq n} (x_{(\sigma \circ \tau)(i)} - x_{(\sigma \circ \tau)(j)}) \\ &= (\sigma \circ \tau)(P). \end{aligned}$$

Dalla formula (3.1.1) segue che

$$\operatorname{sgn}(\sigma \circ \tau)P = (\sigma \circ \tau)(P) = \sigma(\tau(P)) = \operatorname{sgn}(\sigma) \operatorname{sgn}(\tau)P,$$

da cui si deduce che

$$\operatorname{sgn}(\sigma \circ \tau) = \operatorname{sgn}(\sigma) \operatorname{sgn}(\tau).$$

Dato che $\sigma^{-1} \circ \sigma$ è la permutazione identica, il cui segno è 1, si ha

$$\operatorname{sgn}(\sigma^{-1}) \operatorname{sgn}(\sigma) = 1,$$

e pertanto

$$\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma)^{-1} = \operatorname{sgn}(\sigma). \quad \square$$

3.1.1 Cicli e trasposizioni

Definizione 3.1.6. Consideriamo r elementi $i_1, i_2, \dots, i_r \in \{1, 2, \dots, n\}$. Una permutazione $\sigma \in \mathfrak{S}_n$ tale che $\sigma(i_1) = i_2, \sigma(i_2) = i_3, \dots, \sigma(i_{r-1}) = i_r, \sigma(i_r) = i_1$, e $\sigma(h) = h$ se $h \neq i_1, i_2, \dots, i_r$, è detta un *ciclo* di lunghezza r , o r -ciclo; essa permuta ciclicamente gli elementi i_1, i_2, \dots, i_r e lascia fissi tutti gli altri. Questa permutazione verrà indicata nel modo seguente:

$$\sigma = (i_1 i_2 \dots i_r).$$

Due cicli sono detti *disgiunti* se nella loro rappresentazione non compaiono simboli comuni.

Ad esempio, i cicli $(1\ 4\ 2)$ e $(3\ 5)$ sono disgiunti, mentre non lo sono i cicli $(2\ 1\ 3)$ e $(1\ 4)$.

È un fatto del tutto elementare che ogni permutazione $\sigma \in \mathfrak{S}_n$ può essere scritta come prodotto (cioè come composizione) di un numero finito di cicli disgiunti. Ad esempio, la permutazione

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

si scrive come segue:

$$\sigma = (1\ 5\ 3\ 8)(2\ 6\ 7)(4).$$

Per esprimere una data permutazione σ come prodotto di cicli disgiunti basta vedere dove σ manda il numero 1, poi dove manda $i_1 = \sigma(1)$, poi dove manda $i_2 = \sigma(i_1)$, etc.

Si noti che un ciclo di lunghezza 1 (come il (4) nell'esempio precedente) corrisponde alla permutazione identica, quindi può essere omesso. Si conclude quindi che ogni permutazione può essere scritta come prodotto di cicli disgiunti di lunghezza ≥ 2 .

Definizione 3.1.7. Un ciclo di lunghezza 2 (o 2-ciclo) è detto una *trasposizione*.

Si verifica facilmente che ogni r -ciclo può essere scritto (non necessariamente in modo unico) come prodotto di $r - 1$ trasposizioni. Si ha infatti

$$(i_1 i_2 \dots i_r) = (i_1 i_r)(i_1 i_{r-1}) \cdots (i_1 i_3)(i_1 i_2), \quad (3.1.2)$$

oppure anche

$$(i_1 i_2 \dots i_r) = (i_{r-1} i_r)(i_{r-2} i_r) \cdots (i_2 i_r)(i_1 i_r). \quad (3.1.3)$$

Conviene qui ricordare che, nel prodotto di composizione di due o più permutazioni si usa la seguente convenzione: $(\sigma \circ \tau)(i) = \sigma(\tau(i))$.

Dato che ogni permutazione è prodotto di un numero finito di cicli disgiunti, e poiché ogni ciclo è un prodotto finito di trasposizioni, ne segue che ogni

permutazione può sempre essere scritta come prodotto di un numero finito di trasposizioni.

Consideriamo ora una trasposizione $\sigma = (i\ j) \in \mathfrak{S}_n$, con $i < j$. Si verifica facilmente che in questa permutazione sono presenti $2(j-i)-1$ inversioni (farlo per esercizio), quindi il suo segno è

$$\text{sgn}(\sigma) = (-1)^{2(j-i)-1} = -1.$$

Ogni trasposizione ha dunque segno -1 , pertanto se una permutazione si esprime come prodotto di r trasposizioni, il suo segno è $(-1)^r$. Dalle formule (3.1.2) e (3.1.3) deriva quindi il seguente risultato:

Proposizione 3.1.8. *Se σ è un r -ciclo, si ha $\text{sgn}(\sigma) = (-1)^{r-1}$.*

Dimostrazione. Abbiamo visto infatti che un r -ciclo $(i_1\ i_2 \dots i_r)$ si può scrivere come prodotto di $r-1$ trasposizioni. \square

Corollario 3.1.9. *Se una permutazione σ si scrive come prodotto di s cicli disgiunti di lunghezze rispettivamente r_1, r_2, \dots, r_s , il suo segno è dato da*

$$\text{sgn}(\sigma) = (-1)^{r_1+r_2+\dots+r_s-s}.$$

Dimostrazione. Poiché il segno di un ciclo di lunghezza r è $(-1)^{r-1}$, il segno di σ è dato dal seguente prodotto:

$$(-1)^{r_1-1}(-1)^{r_2-1} \dots (-1)^{r_s-1} = (-1)^{r_1+r_2+\dots+r_s-s}.$$

\square

Esempio 3.1.10. Consideriamo la seguente permutazione $\xi \in \mathfrak{S}_n$:

$$\xi = \begin{pmatrix} 1 & 2 & 3 & \dots & h & h+1 & \dots & n \\ h & 1 & 2 & \dots & h-1 & h+1 & \dots & n \end{pmatrix}$$

Questa permutazione è un ciclo di lunghezza h , infatti si ha

$$\xi = (h\ h-1\ h-2\ \dots\ 2\ 1),$$

da cui segue che $\text{sgn}(\xi) = (-1)^{h-1}$.

Consideriamo ora un particolare tipo di trasposizioni.

Definizione 3.1.11. Diremo che una permutazione $\sigma \in \mathfrak{S}_n$ è uno *scambio di elementi contigui* se si ha $\sigma(i) = i+1$ e $\sigma(i+1) = i$ per qualche $i = 1, \dots, n-1$, mentre $\sigma(j) = j$, per ogni $j \neq i, i+1$. Una tale permutazione scambia tra loro due elementi contigui e lascia fissi tutti gli altri.

Concludiamo questa sezione dimostrando il seguente risultato:

Proposizione 3.1.12. *Ogni permutazione $\sigma \in \mathfrak{S}_n$ può essere ottenuta come composizione di un numero finito (eventualmente nullo) di scambi di elementi contigui.*

Dimostrazione. Sia σ la permutazione

$$\sigma = \begin{pmatrix} 1 & 2 & \dots & n-1 & n \\ j_1 & j_2 & \dots & j_{n-1} & j_n \end{pmatrix}$$

Partendo dalla permutazione identica

$$\begin{pmatrix} 1 & 2 & \dots & n-1 & n \\ 1 & 2 & \dots & n-1 & n \end{pmatrix}$$

e scambiando ripetutamente il numero j_n con quello successivo, è possibile portare j_n nell'ultima posizione, ottenendo la permutazione

$$\begin{pmatrix} 1 & 2 & \dots & n-1 & n \\ j'_1 & j'_2 & \dots & j'_{n-1} & j_n \end{pmatrix}$$

Ora che l'ultimo elemento è stato sistemato, possiamo scambiare ripetutamente il numero j_{n-1} con quello successivo in modo da ottenere la permutazione

$$\begin{pmatrix} 1 & 2 & \dots & n-1 & n \\ j''_1 & j''_2 & \dots & j_{n-1} & j_n \end{pmatrix}$$

Continuando in questo modo è possibile sistemare nella posizione corretta il numero j_{n-2} , etc., fino al numero j_1 , ottenendo così la permutazione σ . \square

Osservazione 3.1.13. Poiché il segno di uno scambio di elementi contigui è -1 , se una permutazione σ è ottenuta come composizione di k scambi, si ha

$$\text{sgn}(\sigma) = (-1)^k.$$

Quindi una permutazione è pari (risp. dispari) se e solo se può essere ottenuta come composizione di un numero pari (risp. dispari) di scambi di elementi contigui.

Esercizi

Esercizio 3.1.1. Consideriamo le permutazioni

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 1 & 4 & 2 \end{pmatrix}, \quad \tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 2 & 1 & 4 & 3 \end{pmatrix}$$

Si determinino $\sigma \circ \tau$, $\tau \circ \sigma$, σ^{-1} e τ^{-1} .

Esercizio 3.1.2. Si determini il segno delle seguenti permutazioni:

$$\sigma_1 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 1 & 3 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 5 & 4 & 1 \end{pmatrix}$$

Esercizio 3.1.3. Si scriva la permutazione

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

come prodotto di cicli disgiunti e come prodotto di trasposizioni. Se ne determini poi il segno.

Esercizio 3.1.4. Si scrivano le seguenti permutazioni come prodotto di cicli disgiunti:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 6 & 5 & 1 & 3 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 3 & 2 & 6 & 4 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 5 & 6 & 4 & 1 & 2 \end{pmatrix}$$

Esercizio 3.1.5. Si considerino le permutazioni rappresentate dai seguenti prodotti di cicli:

$$\sigma_1 = (1\ 2\ 3\ 4)(5\ 6\ 7)(2\ 6\ 1)(4\ 7)$$

$$\sigma_2 = (1\ 2\ 3\ 4\ 5)(6\ 7)(1\ 3\ 5\ 7)(1\ 6\ 3)$$

$$\sigma_3 = (1\ 4)(1\ 2\ 3)(4\ 5)(1\ 4)$$

Si scrivano σ_1 , σ_2 e σ_3 come prodotto di cicli disgiunti.

Esercizio 3.1.6. Si esprima la permutazione

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}$$

come composizione di scambi di elementi contigui e se ne determini il segno.

3.2 Il determinante di una matrice quadrata

In questo paragrafo definiremo il determinante di una matrice quadrata a coefficienti in un campo K e studieremo le sue principali proprietà.

Definizione 3.2.1. Sia $A = (a_{ij}) \in M_n(K)$ una matrice quadrata di ordine n a coefficienti in un campo K . Il *determinante* di A è definito da

$$\det A = \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)},$$

ove la somma è estesa a tutte le permutazioni di n elementi $\sigma \in \mathfrak{S}_n$. Il determinante di una matrice A viene spesso indicato anche con il simbolo $|A|$.

Osservazione 3.2.2. Notiamo che la definizione di determinante si applica anche al caso in cui A è una matrice quadrata a coefficienti in un anello commutativo R ; in tal caso il determinante di A è un elemento dell'anello R .

A titolo di esempio, applichiamo questa definizione per calcolare esplicitamente il determinante di una matrice per piccoli valori di n .

Se $n = 1$, cioè se la matrice A è costituita da un solo elemento, $A = a_{11}$, si ha $\det A = a_{11}$ (l'unica permutazione di un solo elemento è la permutazione identica, che ha segno 1).

Se $n = 2$, cioè se

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

ci sono due permutazioni di due elementi,

$$\sigma_1 = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

Si ha $\operatorname{sgn}(\sigma_1) = 1$ e $\operatorname{sgn}(\sigma_2) = -1$, da cui si ottiene

$$\det A = a_{11}a_{22} - a_{12}a_{21}.$$

Consideriamo infine il caso $n = 3$. Ci sono sei permutazioni di tre elementi, e precisamente

$$\begin{aligned} \sigma_1 &= \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, & \sigma_2 &= \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, & \sigma_3 &= \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \\ \sigma_4 &= \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, & \sigma_5 &= \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, & \sigma_6 &= \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}. \end{aligned}$$

Si verifica facilmente che $\operatorname{sgn}(\sigma_1) = \operatorname{sgn}(\sigma_3) = \operatorname{sgn}(\sigma_5) = 1$, mentre $\operatorname{sgn}(\sigma_2) = \operatorname{sgn}(\sigma_4) = \operatorname{sgn}(\sigma_6) = -1$, quindi

$$\begin{aligned} \det A &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} \\ &\quad - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}. \end{aligned}$$

Osservazione 3.2.3. Un metodo pratico per ricordare la formula precedente per il calcolo del determinante di una matrice di ordine 3 è la cosiddetta *regola di Sarrus*. Si tratta di ricopiare, a destra dell'ultima colonna della matrice A , le sue prime due colonne, come qui indicato:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{pmatrix}$$

Ora bisogna considerare la somma dei prodotti degli elementi situati lungo le tre diagonali orientate da nord-ovest a sud-est

$$a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}$$

e a questa sottrarre la somma dei prodotti degli elementi situati lungo le tre diagonali orientate da nord-est a sud-ovest

$$a_{13}a_{22}a_{31} + a_{11}a_{23}a_{32} + a_{12}a_{21}a_{33}.$$

Si ottiene così

$$a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33},$$

che è proprio il determinante di A .

Conviene far notare che non esiste un analogo della *regola di Sarrus* per il calcolo del determinante di una matrice di ordine maggiore di 3.

In generale, ricordando che le permutazioni di n elementi sono $n!$, il calcolo del determinante di una matrice quadrata A di ordine n consiste in una somma di $n!$ addendi, ciascuno dei quali è un prodotto di n elementi di A , presi uno per ogni riga e ogni colonna, con un segno dato dal segno della permutazione corrispondente.

Esempio 3.2.4. Cerchiamo di determinare quante “operazioni” sono necessarie per calcolare il determinante di una matrice di ordine 50. Lo sviluppo di un tale determinante consiste in una somma di $50!$ addendi, ciascuno dei quali è un prodotto di 50 elementi della matrice (preso con il segno opportuno). Ci sono pertanto $50 \times 50! = 50! (50+1) = 51!$ “operazioni” (e dove abbiamo trascurato tutte le operazioni necessarie a determinare il segno di ciascuna delle $50!$ permutazioni). Notiamo che $51! \approx 1.55 \times 10^{66}$.

Se disponessimo di un calcolatore in grado di effettuare mille miliardi di tali “operazioni” al secondo, il tempo necessario a calcolare un tale determinante sarebbe all’incirca 1.55×10^{54} secondi, che equivale a circa 4.9×10^{46} anni, il che corrisponde a più di 3×10^{36} volte la vita dell’universo!

Nel caso di una matrice di ordine 30, lo stesso ragionamento porta a un tempo necessario per il calcolo del determinante pari a circa 17300 volte la vita dell’universo attuale.

Questo esempio mostra come la formula usata per definire il determinante sia sostanzialmente inutilizzabile per il calcolo effettivo, tranne nei casi in cui n è molto piccolo. Tuttavia ciò non significa affatto che una tale formula sia “inutile.” Essa permette infatti di dimostrare molte proprietà notevoli dei determinanti.

Proposizione 3.2.5. *Sia*

$$A = \begin{pmatrix} a_{11} & 0 & 0 & \dots & 0 \\ 0 & a_{22} & 0 & \dots & 0 \\ 0 & 0 & a_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{pmatrix}$$

una matrice diagonale. Si ha $\det A = a_{11}a_{22} \cdots a_{nn}$.

Dimostrazione. Poiché tutti gli elementi al di fuori della diagonale principale di A sono nulli l'unico prodotto non nullo che si trova nello sviluppo del determinante è $a_{11}a_{22} \cdots a_{nn}$, il quale corrisponde alla permutazione identica che ha segno 1. \square

Più in generale, vale il seguente risultato:

Proposizione 3.2.6. *Sia*

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{pmatrix}$$

una matrice triangolare superiore. Si ha $\det A = a_{11}a_{22} \cdots a_{nn}$.

Dimostrazione. Dato che $A = (a_{ij})$ è una matrice triangolare superiore, si ha $a_{ij} = 0$ se $i > j$. Osserviamo che per ogni permutazione σ di $\{1, 2, \dots, n\}$, diversa dalla permutazione identica, esiste almeno un indice i tale che $i > \sigma(i)$. Il prodotto $a_{1\sigma(1)}a_{2\sigma(2)} \cdots a_{n\sigma(n)}$ è quindi nullo, dato che almeno uno dei suoi fattori è zero. Pertanto l'unica permutazione che fornisce un contributo non nullo al calcolo del determinante è la permutazione identica, quindi si ha $\det A = a_{11}a_{22} \cdots a_{nn}$. \square

Naturalmente un risultato analogo (con un'analogia dimostrazione) vale anche per matrici triangolari inferiori. Ciò deriva anche dal risultato seguente:

Proposizione 3.2.7. *Sia $A \in M_n(K)$ e sia ${}^t A$ la sua trasposta. Si ha*

$$\det({}^t A) = \det(A).$$

Dimostrazione. Sia $\sigma \in \mathfrak{S}_n$ e sia σ^{-1} la sua inversa. Notiamo che quando σ varia tra tutti gli elementi di \mathfrak{S}_n lo stesso accade anche per la sua inversa. Indichiamo con a_{ij} gli elementi della matrice A e con \tilde{a}_{ij} gli elementi di ${}^t A$. Ricordiamo che $\tilde{a}_{ij} = a_{ji}$. Dalla definizione di determinante, si ha

$$\begin{aligned} \det({}^t A) &= \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) \tilde{a}_{1\sigma(1)} \tilde{a}_{2\sigma(2)} \cdots \tilde{a}_{n\sigma(n)} \\ &= \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n}. \end{aligned}$$

Ora osserviamo che in ciascun prodotto $a_{\sigma(1)1}a_{\sigma(2)2}\cdots a_{\sigma(n)n}$ compaiono n elementi, presi uno per ciascuna riga e per ciascuna colonna di A . Da ciò segue che tale prodotto si può anche scrivere nella forma $a_{1\sigma^{-1}(1)}a_{2\sigma^{-1}(2)}\cdots a_{n\sigma^{-1}(n)}$, cioè si ha

$$a_{\sigma(1)1}a_{\sigma(2)2}\cdots a_{\sigma(n)n} = a_{1\sigma^{-1}(1)}a_{2\sigma^{-1}(2)}\cdots a_{n\sigma^{-1}(n)},$$

per ogni $\sigma \in \mathfrak{S}_n$.

Un esempio può servire a chiarire quanto appena affermato. Sia $n = 4$ e consideriamo la permutazione

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}$$

la cui inversa è

$$\sigma^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}$$

Allora si ha

$$a_{\sigma(1)1}a_{\sigma(2)2}a_{\sigma(3)3}a_{\sigma(4)4} = a_{31}a_{12}a_{43}a_{24}$$

e

$$a_{1\sigma^{-1}(1)}a_{2\sigma^{-1}(2)}a_{3\sigma^{-1}(3)}a_{4\sigma^{-1}(4)} = a_{12}a_{24}a_{31}a_{43}$$

i quali sono evidentemente uguali.

Se ricordiamo inoltre che $\text{sgn}(\sigma^{-1}) = \text{sgn}(\sigma)$, possiamo scrivere:

$$\begin{aligned} \det({}^t A) &= \sum_{\sigma \in \mathfrak{S}_n} \text{sgn}(\sigma) a_{\sigma(1)1}a_{\sigma(2)2}\cdots a_{\sigma(n)n} \\ &= \sum_{\sigma \in \mathfrak{S}_n} \text{sgn}(\sigma) a_{1\sigma^{-1}(1)}a_{2\sigma^{-1}(2)}\cdots a_{n\sigma^{-1}(n)} \\ &= \sum_{\sigma \in \mathfrak{S}_n} \text{sgn}(\sigma^{-1}) a_{1\sigma^{-1}(1)}a_{2\sigma^{-1}(2)}\cdots a_{n\sigma^{-1}(n)} \\ &= \sum_{\tau \in \mathfrak{S}_n} \text{sgn}(\tau) a_{1\tau(1)}a_{2\tau(2)}\cdots a_{n\tau(n)} \\ &= \det A, \end{aligned}$$

ove abbiamo posto $\tau = \sigma^{-1}$. □

Data una matrice $A \in M_n(K)$ indicheremo con $A^{(1)}, A^{(2)}, \dots, A^{(n)}$ le sue righe

$$A^{(i)} = (a_{i1}, a_{i2}, \dots, a_{in}),$$

e con $A_{(1)}, A_{(2)}, \dots, A_{(n)}$ le sue colonne

$$A_{(j)} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix}$$

Potremo quindi scrivere

$$A = (A_{(1)}, A_{(2)}, \dots, A_{(n)}) = \begin{pmatrix} A^{(1)} \\ A^{(2)} \\ \vdots \\ A^{(n)} \end{pmatrix}$$

Possiamo ora enunciare e dimostrare la seguente proprietà:

Proposizione 3.2.8. *Se la i -esima riga di una matrice A è combinazione lineare di due vettori riga*

$$A^{(i)} = \alpha V + \alpha' V',$$

con $V = (v_1, v_2, \dots, v_n)$ e $V' = (v'_1, v'_2, \dots, v'_n)$, allora si ha

$$\det A = \alpha \det \begin{pmatrix} A^{(1)} \\ \vdots \\ A^{(i-1)} \\ V \\ A^{(i+1)} \\ \vdots \\ A^{(n)} \end{pmatrix} + \alpha' \det \begin{pmatrix} A^{(1)} \\ \vdots \\ A^{(i-1)} \\ V' \\ A^{(i+1)} \\ \vdots \\ A^{(n)} \end{pmatrix}$$

Analogamente, se la j -esima colonna di A è combinazione lineare di due vettori colonna

$$A_{(j)} = \beta W + \beta' W',$$

con $W = {}^t(w_1, w_2, \dots, w_n)$ e $W' = {}^t(w'_1, w'_2, \dots, w'_n)$, allora si ha

$$\begin{aligned} \det A &= \beta \det(A_{(1)}, \dots, A_{(j-1)}, W, A_{(j+1)}, \dots, A_{(n)}) \\ &\quad + \beta' \det(A_{(1)}, \dots, A_{(j-1)}, W', A_{(j+1)}, \dots, A_{(n)}). \end{aligned}$$

Dimostrazione. Dato che il determinante di una matrice coincide con quello della sua trasposta, è sufficiente dimostrare l'asserto riguardante le righe. Supponiamo dunque che sia $A^{(i)} = \alpha V + \alpha' V'$, cioè $a_{ij} = \alpha v_j + \alpha' v'_j$, per $j = 1, \dots, n$. Si ha quindi:

$$\begin{aligned} \det A &= \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdots a_{i\sigma(i)} \cdots a_{n\sigma(n)} \\ &= \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdots (\alpha v_{\sigma(i)} + \alpha' v'_{\sigma(i)}) \cdots a_{n\sigma(n)} \\ &= \alpha \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdots v_{\sigma(i)} \cdots a_{n\sigma(n)} \\ &\quad + \alpha' \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdots v'_{\sigma(i)} \cdots a_{n\sigma(n)} \\ &= \alpha \det \begin{pmatrix} A^{(1)} \\ \vdots \\ V \\ \vdots \\ A^{(n)} \end{pmatrix} + \alpha' \det \begin{pmatrix} A^{(1)} \\ \vdots \\ V' \\ \vdots \\ A^{(n)} \end{pmatrix} \end{aligned}$$

□

Proposizione 3.2.9. *Sia A' la matrice ottenuta scambiando tra loro due righe (oppure due colonne) di A . Allora è*

$$\det A' = -\det A.$$

Dimostrazione. Dato che $\det({}^t A) = \det(A)$, è sufficiente dimostrare l'affermazione riguardante le righe. Sia dunque $A' = (a'_{ij})$ la matrice ottenuta scambiando tra loro le righe h -esima e k -esima di A , con $1 \leq h < k \leq n$. Si ha dunque:

$$\begin{aligned}\det A' &= \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) a'_{1\sigma(1)} \cdots a'_{h\sigma(h)} \cdots a'_{k\sigma(k)} \cdots a'_{n\sigma(n)} \\ &= \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdots a_{k\sigma(h)} \cdots a_{h\sigma(k)} \cdots a_{n\sigma(n)}.\end{aligned}$$

Indichiamo con τ la permutazione che scambia h con k , lasciando invariati tutti gli altri elementi, e poniamo $\eta = \sigma \circ \tau$. Si ha così

$$\det A' = \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) a_{1\eta(1)} \cdots a_{h\eta(h)} \cdots a_{k\eta(k)} \cdots a_{n\eta(n)}.$$

Ora basta osservare che quando σ percorre tutto l'insieme \mathfrak{S}_n lo stesso accade anche per η , e che $\operatorname{sgn}(\eta) = \operatorname{sgn}(\sigma) \operatorname{sgn}(\tau) = -\operatorname{sgn}(\sigma)$, in quanto il segno di una trasposizione è -1 . Si ha pertanto

$$\begin{aligned}\det A' &= \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) a_{1\eta(1)} \cdots a_{h\eta(h)} \cdots a_{k\eta(k)} \cdots a_{n\eta(n)} \\ &= \sum_{\eta \in \mathfrak{S}_n} -\operatorname{sgn}(\eta) a_{1\eta(1)} \cdots a_{h\eta(h)} \cdots a_{k\eta(k)} \cdots a_{n\eta(n)} \\ &= -\det A.\end{aligned}$$

□

Osservazione 3.2.10. Il determinante di una matrice può essere considerato come una funzione delle sue n righe

$$\det : K^n \times \cdots \times K^n \rightarrow K, \quad (A^{(1)}, \dots, A^{(n)}) \mapsto \det \begin{pmatrix} A^{(1)} \\ \vdots \\ A^{(n)} \end{pmatrix}$$

oppure come una funzione delle sue n colonne

$$\det : K^n \times \cdots \times K^n \rightarrow K, \quad (A_{(1)}, \dots, A_{(n)}) \mapsto \det(A_{(1)}, \dots, A_{(n)}).$$

La Proposizione 3.2.8 afferma che entrambe queste funzioni sono *multilinear*, cioè sono lineari in ciascuna delle loro n variabili. La Proposizione 3.2.9 afferma poi che queste due funzioni sono *alternanti*, cioè cambiano di segno ogni volta che due delle loro variabili vengono scambiate tra loro. Il determinante fornisce quindi un esempio di applicazione multilineare alternante.

Corollario 3.2.11. *Sia $A = (a_{ij})$ una matrice quadrata di ordine n . Per ogni permutazione $\sigma \in \mathfrak{S}_n$ si ha:*

$$\det \begin{pmatrix} A^{(\sigma(1))} \\ A^{(\sigma(2))} \\ \vdots \\ A^{(\sigma(n))} \end{pmatrix} = \operatorname{sgn}(\sigma) \det \begin{pmatrix} A^{(1)} \\ A^{(2)} \\ \vdots \\ A^{(n)} \end{pmatrix} = \operatorname{sgn}(\sigma) \det A$$

e

$$\det(A_{(\sigma(1))}, A_{(\sigma(2))}, \dots, A_{(\sigma(n))}) = \operatorname{sgn}(\sigma) \det(A_{(1)}, A_{(2)}, \dots, A_{(n)}) = \det A.$$

Dimostrazione. Basta ricordare che ogni permutazione si può scrivere come prodotto di un numero finito di trasposizioni e che, se σ è un prodotto di r trasposizioni, si ha $\text{sgn}(\sigma) = (-1)^r$. \square

Corollario 3.2.12. *Se una matrice quadrata A ha due righe (oppure due colonne) uguali, allora $\det A = 0$.*

Dimostrazione. Scambiando tra loro le due righe uguali (oppure le due colonne uguali) la matrice A non viene alterata, ma il suo determinante deve cambiare di segno. Si ha pertanto $\det A = -\det A$, cioè $2\det A = 0$. Se la caratteristica del campo K è diversa da 2, si conclude che $\det A = 0$, come volevasi dimostrare. Per trattare il caso in cui K è un campo di caratteristica 2, utilizziamo la definizione di determinante. Supponiamo dunque che le righe h -esima e k -esima di A siano uguali, cioè che $a_{hj} = a_{kj}$ per $j = 1, \dots, n$. Poiché $\text{char}(K) = 2$, si ha $\text{sgn}(\sigma) = 1$ per ogni $\sigma \in \mathfrak{S}_n$, quindi

$$\det A = \sum_{\sigma \in \mathfrak{S}_n} a_{1\sigma(1)} \cdots a_{h\sigma(h)} \cdots a_{k\sigma(k)} \cdots a_{n\sigma(n)}.$$

Per ogni permutazione $\sigma \in \mathfrak{S}_n$ indichiamo con σ' la composizione di σ con la trasposizione che scambia tra loro h e k . Nello sviluppo del determinante di A , l'addendo relativo a σ' coincide con quello corrispondente a σ , infatti:

$$\begin{aligned} a_{1\sigma'(1)} \cdots a_{h\sigma'(h)} \cdots a_{k\sigma'(k)} \cdots a_{n\sigma'(n)} &= a_{1\sigma(1)} \cdots a_{h\sigma(k)} \cdots a_{k\sigma(h)} \cdots a_{n\sigma(n)} \\ &= a_{1\sigma(1)} \cdots a_{k\sigma(k)} \cdots a_{h\sigma(h)} \cdots a_{n\sigma(n)} \end{aligned}$$

ove nell'ultima uguaglianza abbiamo usato il fatto che le righe di indici h e k sono uguali. Ciò significa che gli addendi che compaiono nella sommatoria su tutte le permutazioni $\sigma \in \mathfrak{S}_n$ sono a due a due uguali. Poiché la caratteristica di K è 2, ciò implica che $\det A = 0$. \square

Osservazione 3.2.13. Abbiamo già osservato che il determinante può essere pensato come un'applicazione multilinearare alternante delle righe o delle colonne di una matrice quadrata. In realtà si tratta proprio dell'unica applicazione multilinearare alternante che, valutata sulla matrice identica, è uguale a 1. Infatti supponiamo che

$$F : K^n \times \cdots \times K^n \rightarrow K$$

sia un'applicazione multilinearare alternante tale che $F(e_1, e_2, \dots, e_n) = 1$, ove e_1, e_2, \dots, e_n sono i vettori della base canonica di K^n . Sia $A = (a_{ij})$ una matrice quadrata di ordine n e siano $A^{(1)}, \dots, A^{(n)}$ le sue righe. Per ogni $i = 1, \dots, n$, il vettore $A^{(i)} \in K^n$ si può scrivere come combinazione lineare dei vettori della base canonica come segue:

$$A^{(i)} = a_{i1}e_1 + a_{i2}e_2 + \cdots + a_{in}e_n.$$

Dato che F è multilineare alternante, si ha:

$$\begin{aligned}
 F(A^{(1)}, \dots, A^{(n)}) &= F\left(\sum_{j_1=1}^n a_{1j_1} e_{j_1}, \dots, \sum_{j_n=1}^n a_{nj_n} e_{j_n}\right) \\
 &= \sum_{j_1, \dots, j_n} a_{1j_1} \cdots a_{nj_n} F(e_{j_1}, \dots, e_{j_n}) \\
 &= \sum_{\sigma \in \mathfrak{S}_n} a_{1\sigma(1)} \cdots a_{n\sigma(n)} F(e_{\sigma(1)}, \dots, e_{\sigma(n)}) \\
 &= \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)} F(e_1, \dots, e_n) \\
 &= \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)} \\
 &= \det A.
 \end{aligned}$$

Proposizione 3.2.14. *Se una matrice $A \in M_n(K)$ ha una riga (oppure una colonna) nulla, il suo determinante è nullo.*

Dimostrazione. Dato che $\det(A) = \det({}^t A)$, è sufficiente considerare il caso in cui A ha una riga nulla. Supponiamo dunque che la i -esima riga di A sia nulla, cioè che $a_{ij} = 0$ per $j = 1, \dots, n$. Per ogni permutazione $\sigma \in \mathfrak{S}_n$, nel prodotto

$$a_{1\sigma(1)} \cdots a_{i\sigma(i)} \cdots a_{n\sigma(n)}$$

compare il fattore $a_{i\sigma(i)} = 0$, quindi tale prodotto è nullo. Dalla definizione di determinante, segue che $\det A = 0$. \square

Proposizione 3.2.15. *Il determinante di una matrice $A \in M_n(K)$ non cambia se a una riga (risp. a una colonna) di A si somma una combinazione lineare delle righe (risp. delle colonne) rimanenti.*

Dimostrazione. Anche in questo caso è sufficiente dimostrare l'affermazione riguardante le righe. Sia dunque B la matrice ottenuta da A sostituendo la sua i -esima riga $A^{(i)}$ con la riga

$$B^{(i)} = A^{(i)} + \sum_{j \neq i} \alpha_j A^{(j)},$$

con $\alpha_j \in K$. Ricordando la multilinearità del determinante si ha:

$$\det B = \det \begin{pmatrix} A^{(1)} \\ \vdots \\ A^{(i-1)} \\ A^{(i)} \\ A^{(i+1)} \\ \vdots \\ A^{(n)} \end{pmatrix} + \sum_{j \neq i} \alpha_j \det \begin{pmatrix} A^{(1)} \\ \vdots \\ A^{(i-1)} \\ A^{(j)} \\ A^{(i+1)} \\ \vdots \\ A^{(n)} \end{pmatrix}$$

Poiché j è diverso da i , nella matrice

$$\begin{pmatrix} A^{(1)} \\ \vdots \\ A^{(i-1)} \\ A^{(j)} \\ A^{(i+1)} \\ \vdots \\ A^{(n)} \end{pmatrix}$$

ci sono due righe uguali, quindi il determinante di tale matrice è nullo. Si ha pertanto $\det B = \det A$. \square

Proposizione 3.2.16. *Se $A \in M_n(K)$ e $\alpha \in K$, si ha*

$$\det(\alpha A) = \alpha^n \det(A).$$

Dimostrazione. Dalla multilinearità del determinante segue che

$$\det(\alpha A) = \det \begin{pmatrix} \alpha A^{(1)} \\ \alpha A^{(2)} \\ \vdots \\ \alpha A^{(n)} \end{pmatrix} = \alpha \det \begin{pmatrix} A^{(1)} \\ \alpha A^{(2)} \\ \vdots \\ \alpha A^{(n)} \end{pmatrix} = \cdots = \alpha^n \det \begin{pmatrix} A^{(1)} \\ A^{(2)} \\ \vdots \\ A^{(n)} \end{pmatrix} = \alpha^n \det A. \quad \square$$

Siamo ora in grado di dimostrare il seguente risultato:

Teorema 3.2.17 (TEOREMA DI BINET). *Date due matrici $A, B \in M_n(K)$, si ha:*

$$\det(AB) = \det(A) \det(B).$$

Dimostrazione. Siano $A = (a_{ij})$ e $B = (b_{ij})$ e scriviamo B come segue:

$$B = \begin{pmatrix} B^{(1)} \\ \vdots \\ B^{(n)} \end{pmatrix}$$

Sviluppando il prodotto righe per colonne di A per B , si ottiene

$$\det(AB) = \det \begin{pmatrix} a_{11}B^{(1)} + \cdots + a_{1n}B^{(n)} \\ a_{21}B^{(1)} + \cdots + a_{2n}B^{(n)} \\ \vdots \\ a_{n1}B^{(1)} + \cdots + a_{nn}B^{(n)} \end{pmatrix}$$

Ricordando ora la multilinearità del determinante, si ha

$$\det(AB) = \sum_{j_1} \sum_{j_2} \cdots \sum_{j_n} a_{1j_1} a_{2j_2} \cdots a_{nj_n} \det \begin{pmatrix} B^{(j_1)} \\ \vdots \\ B^{(j_n)} \end{pmatrix}$$

Dato che il determinante di una matrice avente due righe uguali è nullo, l'espressione precedente si riduce a:

$$\det(AB) = \sum_{\sigma \in \mathfrak{S}_n} a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \det \begin{pmatrix} B^{(\sigma(1))} \\ \vdots \\ B^{(\sigma(n))} \end{pmatrix}$$

Infine, ricordando che il determinante è una funzione alternante (vedi Corollario 3.2.11), si ha:

$$\begin{aligned} \det(AB) &= \sum_{\sigma \in \mathfrak{S}_n} a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \operatorname{sgn}(\sigma) \det \begin{pmatrix} B^{(1)} \\ \vdots \\ B^{(n)} \end{pmatrix} \\ &= \left(\sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \right) \det B \\ &= \det(A) \det(B). \end{aligned}$$

□

Corollario 3.2.18. *Se $A \in M_n(K)$ è una matrice invertibile, si ha*

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

Dimostrazione. Se A è invertibile, si ha $AA^{-1} = \mathbf{1}_n$. Applicando il Teorema di Binet, si ottiene:

$$1 = \det(\mathbf{1}_n) = \det(AA^{-1}) = \det(A) \det(A^{-1}),$$

quindi $\det(A^{-1}) = (\det A)^{-1}$.

□

Definizione 3.2.19. Data una matrice $A = (a_{ij}) \in M_n(K)$, per ogni $i, j \in \{1, \dots, n\}$ indicheremo con A_{ij} la matrice di ordine $n-1$ ottenuta da A cancellando la sua i -esima riga e la sua j -esima colonna. Il determinante della matrice A_{ij} è detto il *minore* di indici i e j della matrice A . La quantità

$$a_{ij}^* = (-1)^{i+j} |A_{ij}|$$

è detta il *complemento algebrico* (o *cofattore*) dell'elemento a_{ij} di A . La trasposta della matrice costituita dai complementi algebrici degli elementi di A è detta la *matrice aggiunta* (o *matrice cofattore*) di A , e sarà indicata con $\operatorname{adj}(A)$ o, più semplicemente, con A^* :

$$A^* = \operatorname{adj}(A) = {}^t(a_{ij}^*) \in M_n(K).$$

Possiamo ora dimostrare il seguente risultato, che fornisce un metodo molto utile per il calcolo del determinante di una matrice.

Proposizione 3.2.20 (FORMULA DI LAPLACE). *Sia $A \in M_n(K)$. Per ogni indice di riga i si ha:*

$$\det A = \sum_{h=1}^n (-1)^{i+h} a_{ih} |A_{ih}|. \quad (3.2.1)$$

Analogamente, per ogni indice di colonna j si ha:

$$\det A = \sum_{k=1}^n (-1)^{k+j} a_{kj} |A_{kj}|. \quad (3.2.2)$$

La prima formula è detta *sviluppo del determinante di A secondo la i -esima riga*, mentre la seconda è lo *sviluppo del determinante di A secondo la j -esima colonna*.

Dimostrazione. Poiché il determinante di una matrice coincide con quello della sua trasposta, scambiando i ruoli delle righe e delle colonne della matrice A , la formula (3.2.1) si riduce alla (3.2.2). Pertanto è sufficiente dimostrarne una delle due, ad esempio la (3.2.1).

Ci proponiamo ora di mostrare che, in effetti, è sufficiente dimostrare la formula (3.2.1) per $i = 1$. Supponiamo dunque che la formula di Laplace valga per lo sviluppo del determinante di A secondo la prima riga. Fissiamo un indice di riga $i > 1$ e scriviamo la matrice A nella forma

$$A = \begin{pmatrix} A^{(1)} \\ \vdots \\ A^{(i-1)} \\ A^{(i)} \\ A^{(i+1)} \\ \vdots \\ A^{(n)} \end{pmatrix}$$

Mediante $i - 1$ scambi di righe contigue è possibile portare la i -esima riga di A al posto della prima riga, ottenendo così la matrice

$$A' = \begin{pmatrix} A^{(i)} \\ A^{(1)} \\ \vdots \\ A^{(i-1)} \\ A^{(i+1)} \\ \vdots \\ A^{(n)} \end{pmatrix}$$

Poiché ad ogni scambio di due righe il determinante cambia di segno, si ha

$$\det A = (-1)^{i-1} \det A'.$$

Possiamo ora sviluppare il determinante di $A' = (a'_{ij})$ secondo la prima riga, ottenendo

$$\det A' = \sum_{h=1}^n (-1)^{1+h} a'_{1h} |A'_{1h}|.$$

Ma, dalla definizione di A' , si vede che $a'_{1h} = a_{ih}$ e $A'_{1h} = A_{ih}$, quindi si ha

$$\det A' = \sum_{h=1}^n (-1)^{1+h} a_{ih} |A_{ih}|.$$

Si conclude pertanto che

$$\det A = (-1)^{i-1} \det A' = \sum_{h=1}^n (-1)^{i+h} a_{ih} |A_{ih}|,$$

che è precisamente lo sviluppo del determinante di A secondo la i -esima riga.

Non rimane altro che dimostrare la formula (3.2.1) nel caso in cui $i = 1$. Dobbiamo quindi dimostrare che

$$\det A = \sum_{h=1}^n (-1)^{1+h} a_{1h} |A_{1h}|.$$

Notiamo che A_{1h} è la seguente matrice quadrata di ordine $n-1$

$$A_{1h} = \begin{pmatrix} a_{21} & \dots & a_{2,h-1} & a_{2,h+1} & \dots & a_{2n} \\ a_{31} & \dots & a_{3,h-1} & a_{3,h+1} & \dots & a_{3n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n1} & \dots & a_{n,h-1} & a_{n,h+1} & \dots & a_{nn} \end{pmatrix}$$

Dalla definizione di determinante, si ha

$$|A_{1h}| = \sum_{\tau \in \mathfrak{S}_{n-1}^{(h)}} \operatorname{sgn}(\tau) a_{2,\tau(1)} \cdots a_{h,\tau(h-1)} a_{h+1,\tau(h+1)} \cdots a_{n,\tau(n)},$$

ove la sommatoria è estesa a tutte le permutazioni τ dell'insieme di $n-1$ elementi

$$\{1, 2, \dots, h-1, h+1, \dots, n\}$$

(il simbolo $\mathfrak{S}_{n-1}^{(h)}$ indica proprio l'insieme di tali permutazioni).

Si ottiene così la seguente espressione:

$$\begin{aligned} \det A &= \sum_{h=1}^n (-1)^{1+h} a_{1h} \sum_{\tau \in \mathfrak{S}_{n-1}^{(h)}} \operatorname{sgn}(\tau) a_{2,\tau(1)} \cdots a_{h,\tau(h-1)} a_{h+1,\tau(h+1)} \cdots a_{n,\tau(n)} \\ &= \sum_{h=1}^n \sum_{\tau \in \mathfrak{S}_{n-1}^{(h)}} (-1)^{1+h} \operatorname{sgn}(\tau) a_{1h} a_{2,\tau(1)} \cdots a_{h,\tau(h-1)} a_{h+1,\tau(h+1)} \cdots a_{n,\tau(n)}. \end{aligned}$$

Ora alla permutazione $\tau \in \mathfrak{S}_{n-1}^{(h)}$

$$\tau = \begin{pmatrix} 1 & 2 & \dots & h-1 & h+1 & \dots & n \\ \tau(1) & \tau(2) & \dots & \tau(h-1) & \tau(h+1) & \dots & \tau(n) \end{pmatrix}$$

associamo la permutazione $\sigma \in \mathfrak{S}_n$ definita da

$$\sigma = \begin{pmatrix} 1 & 2 & \dots & h-1 & h & h+1 & \dots & n \\ h & \tau(1) & \dots & \tau(h-2) & \tau(h-1) & \tau(h+1) & \dots & \tau(n) \end{pmatrix}$$

Precisamente, σ è definita ponendo:

$$\sigma(1) = h \quad \text{e} \quad \sigma(k) = \begin{cases} \tau(k-1) & \text{se } 2 \leq k \leq h, \\ \tau(k) & \text{se } k > h. \end{cases}$$

Osserviamo che, al variare dell'indice h da 1 a n e di τ nell'insieme $\mathfrak{S}_{n-1}^{(h)}$, le corrispondenti permutazioni σ descrivono tutto l'insieme \mathfrak{S}_n delle permutazioni degli n elementi $\{1, 2, \dots, n\}$.

L'ultima cosa che rimane da capire a questo punto è quale sia la relazione tra il segno di τ e quello della corrispondente permutazione σ .

Possiamo notare che σ è ottenuta componendo la permutazione

$$\xi = \begin{pmatrix} 1 & 2 & 3 & \dots & h & h+1 & \dots & n \\ h & 1 & 2 & \dots & h-1 & h+1 & \dots & n \end{pmatrix}$$

con la permutazione τ' definita da

$$\tau' = \begin{pmatrix} 1 & 2 & \dots & h-1 & h & h+1 & \dots & n \\ \tau(1) & \tau(2) & \dots & \tau(h-1) & h & \tau(h+1) & \dots & \tau(n) \end{pmatrix}$$

Si ha infatti $\sigma = \tau' \circ \xi$, come si può facilmente verificare. Da ciò segue che $\text{sgn}(\sigma) = \text{sgn}(\tau') \text{sgn}(\xi)$. Ora osserviamo che le due permutazioni τ e τ' hanno la stessa rappresentazione come prodotto di cicli disgiunti (nella rappresentazione di τ' comparirebbe il ciclo di lunghezza uno (h) la cui presenza è irrilevante, dato che esso rappresenta la permutazione identica), quindi $\text{sgn}(\tau) = \text{sgn}(\tau')$. Si ha poi $\text{sgn}(\xi) = (-1)^{h-1} = (-1)^{h+1}$, come già visto nell'Esempio 3.1.10. Si ottiene così $\text{sgn}(\sigma) = (-1)^{h+1} \text{sgn}(\tau)$ e lo sviluppo di Laplace può dunque essere riscritto come segue:

$$\begin{aligned} \det A &= \sum_{h=1}^n \sum_{\tau \in \mathfrak{S}_{n-1}^{(h)}} (-1)^{1+h} \text{sgn}(\tau) a_{1h} a_{2,\tau(1)} \cdots a_{h,\tau(h-1)} a_{h+1,\tau(h+1)} \cdots a_{n,\tau(n)} \\ &= \sum_{\sigma \in \mathfrak{S}_n} \text{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{h\sigma(h)} a_{h+1,\sigma(h+1)} \cdots a_{n\sigma(n)}. \end{aligned}$$

Ma quest'ultima è precisamente la definizione del determinante di A . \square

Esempio 3.2.21. Utilizziamo la formula di Laplace per calcolare il determinante della seguente matrice:

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

Possiamo sviluppare questo determinante secondo una riga o una colonna qualsiasi ma, ovviamente, converrà scegliere una riga (o una colonna) tra quelle che contengono il maggior numero di zeri. Scegliendo, ad esempio, la terza riga, si ottiene:

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

Consideriamo il primo di questi due determinanti di ordine tre e sviluppiamolo secondo la seconda colonna:

$$\begin{vmatrix} 2 & 1 & 1 \\ -1 & 0 & 2 \\ 1 & -1 & 0 \end{vmatrix} = - \begin{vmatrix} -1 & 2 \\ 1 & 0 \end{vmatrix} + \begin{vmatrix} 2 & 1 \\ -1 & 2 \end{vmatrix} = -(-2) + 5 = 7.$$

Consideriamo ora il secondo determinante di ordine tre e sviluppiamolo secondo la prima riga:

$$\begin{vmatrix} 2 & 0 & 1 \\ -1 & 3 & 2 \\ 1 & 4 & 0 \end{vmatrix} = 2 \begin{vmatrix} 3 & 2 \\ 4 & 0 \end{vmatrix} + \begin{vmatrix} -1 & 3 \\ 1 & 4 \end{vmatrix} = -16 - 7 = -23.$$

In conclusione, si trova $\det A = -14 + 69 = 55$.

Come corollario del Teorema di Laplace otteniamo il seguente utile risultato, che fornisce una formula esplicita per calcolare l'inversa di una matrice quadrata.

Corollario 3.2.22. *Sia $A \in M_n(K)$ e indichiamo con A^* la matrice aggiunta di A . Sussiste la seguente identità:*

$$A A^* = \det(A) \mathbf{1}_n.$$

Di conseguenza, se $\det(A)$ è invertibile, si ha

$$A^{-1} = (\det A)^{-1} A^*.$$

Dimostrazione. L'identità $A A^* = \det(A) \mathbf{1}_n$ equivale alle seguenti uguaglianze:

$$\sum_{h=1}^n (-1)^{j+h} a_{ih} |A_{jh}| = \begin{cases} \det A & \text{se } i = j, \\ 0 & \text{se } i \neq j. \end{cases}$$

Se $i = j$, l'espressione precedente si riduce a

$$\sum_{h=1}^n (-1)^{i+h} a_{ih} |A_{ih}|,$$

la quale non è altro che lo sviluppo di $\det A$ secondo la i -esima riga di A .

Consideriamo ora il caso $i \neq j$. L'espressione

$$\sum_{h=1}^n (-1)^{j+h} a_{ih} |A_{jh}|$$

può ora essere interpretata come lo sviluppo, secondo la j -esima riga, del determinante della matrice B ottenuta da A sostituendo la sua j -esima riga con una copia della i -esima:

$$B = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{j-1,1} & a_{j-1,2} & \dots & a_{j-1,n} \\ a_{i1} & a_{i2} & \dots & a_{in} \\ a_{j+1,1} & a_{j+1,2} & \dots & a_{j+1,n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

Poiché questa matrice ha due righe uguali, il suo determinante è nullo, pertanto si ha

$$\sum_{h=1}^n (-1)^{j+h} a_{ih} |A_{jh}| = 0,$$

per ogni $i \neq j$. \square

Osservazione 3.2.23. Come abbiamo visto nel Corollario 3.2.18, dal Teorema di Binet si deduce che l'invertibilità del determinante di una matrice $A \in M_n(K)$ è una condizione necessaria per l'invertibilità di A . Il risultato precedente dimostra che questa condizione è anche sufficiente. Si conclude pertanto che una matrice $A \in M_n(K)$ è invertibile se e solo se il suo determinante è invertibile. Possiamo notare che questo risultato vale anche, più in generale, per matrici a coefficienti in un anello commutativo unitario. Ad esempio, una matrice $A \in M_n(\mathbb{Z})$ è invertibile se e solo se $\det A = \pm 1$ (la condizione $\det A \neq 0$ garantisce infatti l'esistenza di A^{-1} nell'anello $M_n(\mathbb{Q})$ e non in $M_n(\mathbb{Z})$).

Descriviamo ora un'applicazione di questi risultati alla teoria dei sistemi lineari; si tratta di una formula, nota come *regola di Cramer*, che permette di esprimere la soluzione di un sistema di n equazioni lineari in n incognite.

Teorema 3.2.24 (REGOLA DI CRAMER). *Sia $AX = B$ un sistema di n equazioni lineari in n incognite, ove*

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \quad X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

Se $\det A$ è invertibile, il sistema ammette un'unica soluzione data, per ogni $i = 1, \dots, n$, da

$$x_i = \Delta^{-1} \Delta_i,$$

ove $\Delta = \det A$ e Δ_i è il determinante della matrice ottenuta da A sostituendo la sua i -esima colonna con la colonna B dei termini noti.

Dimostrazione. Se $\det A$ è invertibile, la matrice A è invertibile, pertanto il sistema $AX = B$ ha un'unica soluzione data da $X = A^{-1}B$. Ricordando che l'inversa di A è data dalla formula $A^{-1} = (\det A)^{-1} A^*$, sviluppando il prodotto righe per colonne di A^{-1} per B si ottiene

$$x_i = \Delta^{-1} \sum_{h=1}^n (-1)^{i+h} |A_{hi}| b_h.$$

Consideriamo ora il determinante Δ_i della matrice ottenuta da A sostituendo la sua i -esima colonna con la colonna B dei termini noti. Sviluppando Δ_i secondo la i -esima colonna, si trova

$$\Delta_i = \sum_{h=1}^n (-1)^{i+h} b_h |A_{hi}|,$$

da cui si deduce che $x_i = \Delta^{-1} \Delta_i$, come volevasi dimostrare. \square

3.2.1 Calcolo del determinante di una matrice mediante l'eliminazione di Gauss

Tutti i metodi di calcolo del determinante di una matrice quadrata di ordine n che abbiamo visto finora coinvolgono, in generale, una somma di $n!$ termini, ciascuno dei quali è, a sua volta, un prodotto di n elementi di A , preso con un opportuno segno. L'utilizzo della formula di Laplace semplifica il calcolo del determinante nel caso in cui la matrice contenga molti zeri; se invece la matrice è priva di zeri, il numero di termini da sommare rimane sempre $n!$.

Vedremo ora come, sfruttando le proprietà di multilinearità e di alternanza del determinante, sia possibile trovare un algoritmo di calcolo molto più efficiente.

Nel Capitolo 2, al paragrafo 2.3.2, abbiamo descritto un metodo per il calcolo del rango di una matrice basato sulla cosiddetta “eliminazione di Gauss.” Si trattava di ridurre una matrice a una forma a scala tramite una successione di opportune operazioni elementari sulle righe. Nel caso di una matrice quadrata, la forma finale a scala non è altro che una matrice triangolare superiore. Ciò significa che, se $A = (a_{ij}) \in M_n(K)$ è una matrice quadrata di ordine n , mediante una successione di operazioni elementari sulle righe è possibile trasformarla in una matrice triangolare superiore

$$A' = \begin{pmatrix} a'_{11} & a'_{12} & a'_{13} & \dots & a'_{1n} \\ 0 & a'_{22} & a'_{23} & \dots & a'_{2n} \\ 0 & 0 & a'_{33} & \dots & a'_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a'_{nn} \end{pmatrix}$$

Il calcolo del determinante di A' è immediato:

$$\det A' = a'_{11} a'_{22} \cdots a'_{nn}.$$

Poiché l'effetto sul determinante di una matrice di ciascuna operazione elementare sulle righe è noto, è possibile tenerne conto per trovare la relazione esistente tra il determinante della matrice A e quello di A' . Vale la pena ricordare i tre tipi di operazioni elementari sulle righe e i loro effetti sul determinante di una matrice:

- (1) scambiare due righe tra loro: in tal caso il determinante cambia di segno;
- (2) moltiplicare una riga per uno scalare λ : in tal caso il determinante risulta moltiplicato per λ ;
- (3) sommare a una riga un multiplo di un'altra riga: in tal caso il determinante non cambia.

Questo metodo permette dunque di calcolare il determinante di una matrice A effettuando solamente delle operazioni elementari sulle sue righe, al fine di ridurla a una forma triangolare superiore (naturalmente si potrebbero anche effettuare operazioni elementari sulle colonne, oppure si potrebbe ridurre la matrice a una forma triangolare inferiore).

Possiamo fare una stima grossolana del numero massimo di operazioni necessarie per calcolare, nel modo appena descritto, il determinante di una matrice

quadrata di ordine n . Gli elementi al di sotto della diagonale principale sono

$$1 + 2 + \cdots + (n-1) = \frac{(n-1)n}{2}$$

e tutti questi devono essere trasformati in zeri. La produzione di uno di questi zeri avviene sommando a una riga un multiplo di un'altra riga, e la somma di due vettori riga equivale a n somme di numeri (in realtà, più si avanza nell'algoritmo, meno somme sono necessarie, perché molti elementi delle righe da sommare sono degli zeri). Pertanto il numero di “operazioni” necessarie (trascorrendo eventuali scambi di righe) è circa pari a

$$n \frac{(n-1)n}{2} = \frac{n^3 - n^2}{2}.$$

Poiché, per n grande, il numero $\frac{n^3 - n^2}{2}$ è molto più piccolo di $n!$, questo algoritmo per il calcolo del determinante risulta di gran lunga più efficiente di quelli descritti in precedenza.

Esempio 3.2.25. Descriviamo ora il calcolo del determinante della seguente matrice utilizzando il metodo dell'eliminazione di Gauss.

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

Scambiando la prima con la seconda riga, si ha:

$$\det A = - \begin{vmatrix} -1 & 3 & 1 & 2 \\ 2 & 1 & -1 & 3 \\ 2 & -1 & 4 & 1 \\ 3 & -2 & 1 & 4 \end{vmatrix}$$

Ora alla seconda riga sommiamo il doppio della prima, alla terza riga sommiamo il doppio della prima e alla quarta riga sommiamo la prima moltiplicata per 3:

$$\det A = - \begin{vmatrix} -1 & 3 & 1 & 2 \\ 0 & 7 & 1 & 7 \\ 0 & 5 & 6 & 5 \\ 0 & 7 & 4 & 10 \end{vmatrix}$$

Ora scambiamo la seconda e la terza colonna, ottenendo:

$$\det A = \begin{vmatrix} -1 & 1 & 3 & 2 \\ 0 & 1 & 7 & 7 \\ 0 & 6 & 5 & 5 \\ 0 & 4 & 7 & 10 \end{vmatrix}$$

A questo punto, alla terza riga sommiamo la seconda moltiplicata per -6 e alla quarta riga sommiamo la seconda moltiplicata per -4 :

$$\det A = \begin{vmatrix} -1 & 1 & 3 & 2 \\ 0 & 1 & 7 & 7 \\ 0 & 0 & -37 & -37 \\ 0 & 0 & -21 & -18 \end{vmatrix}$$

Ora raccogliamo -37 dalla terza riga:

$$\det A = -37 \begin{vmatrix} -1 & 1 & 3 & 2 \\ 0 & 1 & 7 & 7 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -21 & -18 \end{vmatrix}$$

Infine, alla quarta riga sommiamo la terza moltiplicata per 21 :

$$\det A = -37 \begin{vmatrix} -1 & 1 & 3 & 2 \\ 0 & 1 & 7 & 7 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 3 \end{vmatrix}$$

Si ottiene così

$$\det A = -37 \cdot (-3) = 111.$$

Esempio 3.2.26 (IL DETERMINANTE DI VANDERMONDE). Siano $x_1, \dots, x_n \in K$, con $n \geq 2$, e consideriamo la seguente matrice, detta *matrice di Vandermonde*:

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ \vdots & \vdots & & \vdots \\ x_1^{n-1} & x_2^{n-1} & \dots & x_n^{n-1} \end{pmatrix}$$

Il determinante di questa matrice è detto *determinante di Vandermonde* e verrà indicato con $V(x_1, x_2, \dots, x_n)$. Vogliamo dimostrare che si ha

$$V(x_1, x_2, \dots, x_n) = \prod_{1 \leq i < j \leq n} (x_j - x_i). \quad (3.2.3)$$

Procediamo per induzione su n . Per $n = 2$ si ha

$$V(x_1, x_2) = \begin{vmatrix} 1 & 1 \\ x_1 & x_2 \end{vmatrix} = x_2 - x_1.$$

Supponiamo dunque che sia $n > 2$ e che la formula (3.2.3) valga per $n - 1$. Per calcolare $V(x_1, x_2, \dots, x_n)$ effettuiamo le seguenti operazioni elementari sulle righe della matrice di Vandermonde: alla n -esima riga sottraiamo la $(n - 1)$ -esima moltiplicata per x_1 , alla $(n - 1)$ -esima riga sottraiamo la $(n - 2)$ -esima moltiplicata per x_1, \dots , alla seconda riga sottraiamo la prima moltiplicata per x_1 . Si ottiene dunque

$$\begin{aligned} V(x_1, \dots, x_n) &= \begin{vmatrix} 1 & 1 & 1 & \dots & 1 \\ 0 & x_2 - x_1 & x_3 - x_1 & \dots & x_n - x_1 \\ 0 & x_2(x_2 - x_1) & x_3(x_3 - x_1) & \dots & x_n(x_n - x_1) \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & x_2^{n-2}(x_2 - x_1) & x_3^{n-2}(x_3 - x_1) & \dots & x_n^{n-2}(x_n - x_1) \end{vmatrix} \\ &= \begin{vmatrix} x_2 - x_1 & x_3 - x_1 & \dots & x_n - x_1 \\ x_2(x_2 - x_1) & x_3(x_3 - x_1) & \dots & x_n(x_n - x_1) \\ \vdots & \vdots & & \vdots \\ x_2^{n-2}(x_2 - x_1) & x_3^{n-2}(x_3 - x_1) & \dots & x_n^{n-2}(x_n - x_1) \end{vmatrix} \end{aligned}$$

Ora dalla prima colonna possiamo raccogliere il termine $x_2 - x_1$, dalla seconda il termine $x_3 - x_1, \dots$, dall'ultima colonna possiamo raccogliere $x_n - x_1$, ottenendo così

$$V(x_1, \dots, x_n) = (x_2 - x_1)(x_3 - x_1) \cdots (x_n - x_1) \begin{vmatrix} 1 & 1 & \dots & 1 \\ x_2 & x_3 & \dots & x_n \\ \vdots & \vdots & & \vdots \\ x_2^{n-2} & x_3^{n-2} & \dots & x_n^{n-2} \end{vmatrix}$$

Quest'ultimo determinante non è altro che il determinante di Vandermonde $V(x_2, \dots, x_n)$ che, per ipotesi induttiva, è uguale a $\prod_{2 \leq i < j \leq n} (x_j - x_i)$. Sostituendo nella formula precedente si ottiene

$$V(x_1, x_2, \dots, x_n) = \prod_{1 \leq i < j \leq n} (x_j - x_i).$$

3.2.2 Il determinante di un endomorfismo

Sia f un endomorfismo di uno spazio vettoriale V di dimensione finita su K . Fissando una base di V , è possibile associare a f una matrice $A \in M_n(K)$, ove $n = \dim V$. Naturalmente, a basi diverse di V corrispondono matrici diverse di f , tuttavia tutte queste matrici sono simili tra loro.

Proposizione 3.2.27. *Matrici simili hanno lo stesso determinante.*

Dimostrazione. Due matrici $A, A' \in M_n(K)$ sono simili se e solo se esiste una matrice invertibile P tale che

$$A' = PAP^{-1}.$$

Dal Teorema di Binet si deduce che

$$\det A' = (\det P)(\det A)(\det P^{-1}) = (\det P)(\det A)(\det P)^{-1} = \det A. \quad \square$$

Questo risultato ci permette di definire il determinante di un endomorfismo f come il determinante di una matrice A associata a f tramite la scelta di una base di V , in quanto tale determinante non dipende dalla base scelta.

Definizione 3.2.28. Sia f un endomorfismo di V . Il *determinante* di f è il determinante di una matrice di f rispetto a una qualche base di V .

Una conseguenza immediata del Teorema di Binet è la seguente:

Proposizione 3.2.29. *Siano f e g due endomorfismi di V . Allora*

$$\det(g \circ f) = (\det g)(\det f).$$

Se f è invertibile, si ha

$$\det(f^{-1}) = (\det f)^{-1}.$$

Vale inoltre il seguente risultato:

Proposizione 3.2.30. *Un endomorfismo f di uno spazio vettoriale V di dimensione finita su K è un isomorfismo se e solo se $\det f \neq 0$.*

Dimostrazione. Sia A la matrice di f rispetto a una qualche base di V . Allora f è un isomorfismo se e solo se A è invertibile ma, per il Corollario 3.2.22, A è invertibile se e solo se $\det A = \det f \neq 0$. \square

3.3 Determinanti e rango

Nel Capitolo 2, il rango (per colonne) di una matrice è stato definito come il massimo numero di colonne linearmente indipendenti. Naturalmente si può anche definire un analogo rango per righe, come il massimo numero di righe linearmente indipendenti. In questa sezione dimostreremo che le due definizioni di rango coincidono e analizzeremo le relazioni esistenti tra il rango di una matrice A e i determinanti delle sottomatrici quadrate che si possono estrarre da A .

Teorema 3.3.1. *Il rango per righe e il rango per colonne di una matrice $A \in M_{m,n}(K)$ coincidono.*

Dimostrazione. Indichiamo con r il rango per righe e con c il rango per colonne di A . Se $r = 0$ allora tutti gli elementi di A sono nulli e quindi anche $c = 0$. Supponiamo quindi che sia $r > 0$. Una relazione di dipendenza lineare tra le colonne di A

$$A_{(1)}x_1 + A_{(2)}x_2 + \cdots + A_{(n)}x_n = \mathbf{0}$$

è una soluzione non nulla del sistema omogeneo

$$AX = \mathbf{0}, \quad (3.3.1)$$

ove $X = {}^t(x_1, \dots, x_n)$. Pertanto il rango per colonne di A è determinato¹ dall'insieme delle soluzioni del sistema (3.3.1).

Poiché una permutazione delle righe di A non modifica l'insieme delle soluzioni del sistema (3.3.1), né tantomeno modifica il rango per righe di A , non è restrittivo supporre che le r righe linearmente indipendenti di A siano proprio le prime r . Di conseguenza, le righe $A^{(r+1)}, A^{(r+2)}, \dots, A^{(m)}$ sono delle combinazioni lineari delle prime r righe di A . Ciò significa che le ultime $m - r$ equazioni del sistema $AX = \mathbf{0}$ sono combinazioni lineari delle prime r e, pertanto, il sistema $AX = \mathbf{0}$ è equivalente al sistema $A'X = \mathbf{0}$, ove A' è la matrice costituita dalle prime r righe di A :

$$A' = \begin{pmatrix} A^{(1)} \\ \vdots \\ A^{(r)} \end{pmatrix}$$

Poiché A e A' individuano sistemi lineari aventi lo stesso insieme di soluzioni, e poiché l'insieme delle soluzioni determina il rango per colonne della matrice, le due matrici A e A' devono avere lo stesso rango per colonne, quindi il rango per colonne di A' è uguale a c . Dato che le colonne di A' sono vettori di K^r , si ha necessariamente $c \leq r$.

Ragionando allo stesso modo sulla matrice tA , si conclude che è anche $r \leq c$, e quindi deve essere $r = c$. \square

Da questo risultato discende che la distinzione tra rango per righe e rango per colonne di una matrice è del tutto inutile. Si potrà dunque parlare semplicemente del *rango* di una matrice per indicare il valore comune dei due tipi di rango, per righe e per colonne.

¹Ricordiamo che si ha $\text{rk}(A) = n - \text{null}(A)$, ove $\text{null}(A)$ è la dimensione dello spazio delle soluzioni del sistema omogeneo $AX = \mathbf{0}$ (vedi Capitolo 2, Definizione 2.2.15 e Osservazione 2.1.13).

Corollario 3.3.2. *Le operazioni elementari sulle righe o sulle colonne di una matrice non ne modificano il rango.*

Dimostrazione. È del tutto evidente che le operazioni elementari sulle righe di una matrice non ne modificano il rango per righe, mentre le operazioni elementari sulle colonne non ne modificano il rango per colonne. Poiché i due tipi di rango coincidono, si conclude. \square

Osservazione 3.3.3. Da quest'ultimo risultato discende che nel metodo dell'eliminazione di Gauss descritto nel Cap. 2, Sezione 2.3.2, per calcolare il rango di una matrice si possono usare sia operazioni elementari sulle righe che sulle colonne.

Proposizione 3.3.4. *Siano $A \in M_{m,n}(K)$ e $B \in M_{n,r}(K)$ due matrici.*

(i) *Si ha:*

$$\operatorname{rk}(AB) \leq \min(\operatorname{rk}(A), \operatorname{rk}(B)). \quad (3.3.2)$$

(ii) *Se $m = n$ e A è invertibile, si ha*

$$\operatorname{rk}(AB) = \operatorname{rk}(B).$$

(iii) *Se $n = r$ e B è invertibile, si ha*

$$\operatorname{rk}(AB) = \operatorname{rk}(A).$$

Dimostrazione. (i) Siano $F_A : K^n \rightarrow K^m$ e $F_B : K^r \rightarrow K^n$ le applicazioni lineari definite dalle matrici A e B rispettivamente. Allora AB è la matrice dell'applicazione composta $F_A \circ F_B : K^r \rightarrow K^m$. Ricordando che il rango di una matrice è uguale alla dimensione dell'immagine dell'applicazione lineare corrispondente, la diseguaglianza (3.3.2) è equivalente alla seguente:

$$\dim \operatorname{Im}(F_A \circ F_B) \leq \min(\dim \operatorname{Im}(F_A), \dim \operatorname{Im}(F_B)),$$

la dimostrazione della quale è un facile esercizio.

(ii) Se $m = n$, dire che A è invertibile equivale a dire che $F_A : K^n \rightarrow K^n$ è un isomorfismo. Da ciò discende che $\operatorname{Im}(F_B) \cong F_A(\operatorname{Im} F_B)$, quindi $\dim \operatorname{Im}(F_A \circ F_B) = \dim \operatorname{Im}(F_B)$, cioè $\operatorname{rk}(AB) = \operatorname{rk}(B)$.

(iii) Infine, se $n = r$, l'invertibilità di B equivale all'invertibilità di F_B e pertanto si ha $\operatorname{Im}(F_A \circ F_B) \cong \operatorname{Im}(F_A)$, e quindi $\operatorname{rk}(AB) = \operatorname{rk}(A)$. \square

Nel caso delle matrici quadrate, si ha:

Teorema 3.3.5. *Una matrice quadrata di ordine n è invertibile se e solo se ha rango n .*

Dimostrazione. Se A è invertibile, in base alla proposizione precedente si ha

$$\operatorname{rk}(A) = \operatorname{rk}(A^{-1}A) = \operatorname{rk}(\mathbf{1}_n) = n.$$

Viceversa, supponiamo che A abbia rango n . Allora le sue colonne $A_{(1)}, \dots, A_{(n)}$ costituiscono una base di K^n , quindi, per ogni $j = 1, \dots, n$, esistono $b_{1j}, b_{2j}, \dots, b_{nj} \in K$ tali che

$$A_{(1)}b_{1j} + A_{(2)}b_{2j} + \dots + A_{(n)}b_{nj} = e_j, \quad (3.3.3)$$

ove e_j è il j -esimo vettore della base canonica di K^n . Se poniamo $B = (b_{ij})$, le equazioni (3.3.3) possono essere riscritte come segue:

$$AB = \mathbf{1}_n,$$

da cui segue che A è invertibile. \square

Sia $A \in M_{m,n}(K)$ una matrice con m righe e n colonne. Fissiamo degli indici di riga $1 \leq i_1 < i_2 < \dots < i_p \leq m$ e degli indici di colonna $1 \leq j_1 < j_2 < \dots < j_q \leq n$ e poniamo $I = \{i_1, i_2, \dots, i_p\}$ e $J = \{j_1, j_2, \dots, j_q\}$. Indicheremo con A_{IJ} la sottomatrice $p \times q$ di A costituita dagli elementi comuni alle p righe e alle q colonne determinate dagli indici degli insiemi I e J rispettivamente.

Proposizione 3.3.6. *Se B è una sottomatrice della matrice A , allora $\text{rk}(B) \leq \text{rk}(A)$.*

Dimostrazione. Siano $A \in M_{m,n}(K)$, $I = \{i_1, i_2, \dots, i_p\}$ e $J = \{j_1, j_2, \dots, j_q\}$ due insiemi di indici fissati e poniamo $B = A_{IJ}$. Consideriamo la sottomatrice $C = A_{IN}$, ove $N = \{1, 2, \dots, n\}$. Se interpretiamo il rango come rango per righe (cioè come massimo numero di righe linearmente indipendenti), allora la relazione

$$\text{rk}(C) \leq \text{rk}(A)$$

è ovvia. D'altra parte, B è anche una sottomatrice di C e se, questa volta, interpretiamo il rango come rango per colonne, allora la relazione

$$\text{rk}(B) \leq \text{rk}(C)$$

è ovvia. Da queste due diseguaglianze segue che $\text{rk}(B) \leq \text{rk}(A)$. \square

Siamo ora in grado di dimostrare il seguente teorema, il quale mette in relazione la nozione di rango con quella di determinante.

Teorema 3.3.7. *Il rango di una matrice A a coefficienti in un campo è uguale al massimo degli ordini delle sue sottomatrici quadrate invertibili, cioè al massimo degli ordini dei minori non nulli di A .*

Dimostrazione. Sia ρ il massimo degli ordini delle sottomatrici quadrate invertibili di A . Dal Teorema 3.3.5 e dalla Proposizione 3.3.6 segue che $\rho \leq \text{rk}(A)$. D'altra parte, posto $r = \text{rk}(A)$, siano $A^{(i_1)}, A^{(i_2)}, \dots, A^{(i_r)}$ r righe linearmente indipendenti di A . Allora la sottomatrice A_{IN} di A (ove $I = \{i_1, i_2, \dots, i_r\}$ e $N = \{1, 2, \dots, n\}$) ha rango r , quindi possiede r colonne, di indici, j_1, j_2, \dots, j_r , che sono linearmente indipendenti. Posto $J = \{j_1, j_2, \dots, j_r\}$, ciò significa che la sottomatrice quadrata di ordine r A_{IJ} ha rango r , pertanto è invertibile. Da ciò segue che $\rho \geq r = \text{rk}(A)$. Si conclude dunque che deve essere $\rho = \text{rk}(A)$. \square

Osservazione 3.3.8. In base a questo risultato, il rango di una matrice A può anche essere definito come il massimo ordine dei minori non nulli di A . Utilizzando quest'ultima come definizione del rango di una matrice, e ricordando che il determinante di una matrice coincide con quello della sua trasposta, il fatto che il rango di una matrice coincida con il rango della sua trasposta risulta del tutto ovvio.

Da quest'ultimo teorema si deduce immediatamente il seguente criterio per stabilire se n vettori di K^n formano una base:

Corollario 3.3.9. *I vettori $v_1, v_2, \dots, v_n \in K^n$ formano una base se e solo se il determinante della matrice che ha come righe (o come colonne) i vettori dati è diverso da zero.*

Il principio dei minori orlati

Sia data una matrice $A \in M_{m,n}(K)$ e supponiamo di volerne calcolare il rango cercando di determinare il massimo ordine dei minori non nulli di A .

Supponiamo che A non sia la matrice nulla e quindi che il suo rango sia ≥ 1 . Dovremo quindi calcolare i minori di ordine via via crescente, a partire da quelli di ordine 2. Quando per un certo r si sarà trovato un minore di ordine r non nullo, mentre tutti i minori di ordine $r+1$ sono nulli (oppure non esistono, nel caso in cui $r = \min(m, n)$), si concluderà che $\text{rk}(A) = r$. Infatti dall'annullarsi di tutti i minori di ordine $r+1$ segue l'annullarsi di ogni minore di ordine superiore: ciò si dimostra facilmente per induzione su s , sviluppando un minore di ordine $s > r+1$ secondo una sua riga o una sua colonna.

In realtà, nella situazione appena descritta non è necessario verificare che tutti i minori di ordine $r+1$ siano nulli; basta limitarsi a quei minori di ordine $r+1$ che contengono la sottomatrice quadrata di ordine r con determinante diverso da zero che abbiamo considerato. Vale infatti il seguente risultato:

Teorema 3.3.10 (TEOREMA DEI MINORI ORLATI). *Sia $A \in M_{m,n}(K)$ e sia $B = A_{IJ}$ una sottomatrice quadrata di ordine r di A tale che $\det B \neq 0$. Supponiamo che ogni sottomatrice quadrata di ordine $r+1$ di A ottenuta aggiungendo a B una riga e una colonna di A (i cosiddetti minori orlati di B) abbia determinante nullo. Allora A ha rango r .*

Dimostrazione. Sia $B = A_{IJ}$ con $I = \{i_1, i_2, \dots, i_r\}$ e $J = \{j_1, j_2, \dots, j_r\}$. Dal'ipotesi $\det B \neq 0$ discende che le colonne di indici j_1, \dots, j_r di A sono linearmente indipendenti. La condizione sull'annullamento dei determinanti di tutti i minori orlati di B implica allora che ogni altra colonna di A è combinazione lineare delle colonne di indici j_1, \dots, j_r . Quindi A ha rango r . \square

3.4 Orientamenti

In questa sezione supporremo che K sia un campo ordinato; ad esempio $K = \mathbb{R}$, con la relazione d'ordine usuale. Tutti gli spazi vettoriali che considereremo saranno sempre finitamente generati.

Sia dunque V uno spazio vettoriale su K e siano $\mathbf{v} = \{v_1, v_2, \dots, v_n\}$ e $\mathbf{w} = \{w_1, w_2, \dots, w_n\}$ due basi di V . Indichiamo con $P = (p_{ij})$ la matrice di cambiamento di base, dalla base \mathbf{v} alla base \mathbf{w} . Si ha dunque

$$w_j = \sum_{i=1}^n p_{ij} v_i,$$

per ogni $j = 1, \dots, n$.

Definizione 3.4.1. Le basi \mathbf{v} e \mathbf{w} di V si dicono *equiorientate* se $\det(P) > 0$.

È immediato verificare che l'equiorientazione è una relazione di equivalenza sull'insieme \mathcal{B} di tutte le basi di V . Diamo quindi la seguente definizione:

Definizione 3.4.2. Un *orientamento* di uno spazio vettoriale V (finitamente generato su un campo ordinato K) è una classe di equivalenza per la relazione di equiorientazione. Uno *spazio vettoriale orientato* è uno spazio vettoriale in cui è stato scelto un orientamento.

Dato che, per una matrice P di cambiamento di base, si può solo avere $\det(P) > 0$ oppure $\det(P) < 0$, la relazione di equiorientazione ha solo due classi di equivalenza. Ciò significa che uno spazio vettoriale ammette solo due orientamenti. Fissare un orientamento di uno spazio vettoriale V equivale dunque a fissare una base di V , con la convenzione che due basi equiorientate definiscono lo stesso orientamento. I due possibili orientamenti di uno spazio vettoriale sono detti l'opposto uno dell'altro.

Esempio 3.4.3. Sia $V = \mathbb{R}^2$. Nella figura seguente le basi $\{v_1, v_2\}$ e $\{w_1, w_2\}$ sono equiorientate, mentre le basi $\{v_1, v_2\}$ e $\{u_1, u_2\}$ non sono equiorientate.



Siano V e W due spazi vettoriali sul campo K . Se $f : V \rightarrow W$ è un isomorfismo, l'immagine $f(\mathbf{v}) = \{f(v_1), \dots, f(v_n)\}$ di una base $\mathbf{v} = \{v_1, \dots, v_n\}$ di V è una base di W . L'orientamento di W definito dalla base $f(\mathbf{v})$ risulta essere indipendente dalla scelta della base \mathbf{v} all'interno della sua classe di equivalenza. Si ha infatti:

Lemma 3.4.4. *Sia $f : V \rightarrow W$ un isomorfismo di spazi vettoriali e siano $\mathbf{v} = \{v_1, \dots, v_n\}$ e $\mathbf{v}' = \{v'_1, \dots, v'_n\}$ due basi di V . Allora le basi $f(\mathbf{v}) = \{f(v_1), \dots, f(v_n)\}$ e $f(\mathbf{v}') = \{f(v'_1), \dots, f(v'_n)\}$ di W sono equiorientate se e solo se lo sono le basi \mathbf{v} e \mathbf{v}' .*

Dimostrazione. Sia $P = (p_{ij})$ la matrice di cambiamento di base, dalla base \mathbf{v} alla base \mathbf{v}' . Si ha dunque

$$v'_j = \sum_{i=1}^n p_{ij} v_i,$$

per ogni $j = 1, \dots, n$. Dalla linearità di f segue che

$$f(v'_j) = f\left(\sum_{i=1}^n p_{ij} v_i\right) = \sum_{i=1}^n p_{ij} f(v_i),$$

il che dimostra che P è anche la matrice di cambiamento di base dalla base $f(\mathbf{v})$ alla base $f(\mathbf{v}')$. \square

Il risultato precedente ci permette di dare la seguente definizione:

Definizione 3.4.5. Siano V e W due spazi vettoriali orientati, di dimensione n , sul campo K e siano $\mathbf{v} = \{v_1, \dots, v_n\}$ e $\mathbf{w} = \{w_1, \dots, w_n\}$ due basi che rappresentano gli orientamenti fissati di V e W rispettivamente. Diremo che un isomorfismo $f : V \rightarrow W$ è *compatibile con gli orientamenti* di V e W , o che f è un *isomorfismo di spazi vettoriali orientati*, se le basi $f(\mathbf{v}) = \{f(v_1), \dots, f(v_n)\}$ e $\mathbf{w} = \{w_1, \dots, w_n\}$ di W sono equiorientate.

Nel caso particolare in cui $W = V$, si ha:

Definizione 3.4.6. Sia V uno spazio vettoriale orientato e sia $\mathbf{v} = \{v_1, \dots, v_n\}$ una base che rappresenta l'orientamento di V . Diremo che un automorfismo $f : V \rightarrow V$ preserva l'orientamento di V se le basi $f(\mathbf{v}) = \{f(v_1), \dots, f(v_n)\}$ e $\mathbf{v} = \{v_1, \dots, v_n\}$ sono equiorientate. In caso contrario si dice che f inverte l'orientamento di V .

Il seguente risultato fornisce una caratterizzazione degli automorfismi che preservano l'orientamento:

Proposizione 3.4.7. *Sia V uno spazio vettoriale orientato. Un automorfismo $f : V \rightarrow V$ preserva l'orientamento di V se e solo se $\det(f) > 0$.*

Dimostrazione. Sia $\mathbf{v} = \{v_1, \dots, v_n\}$ una base che rappresenta l'orientamento di V . Se indichiamo con $A = (a_{ij})$ la matrice di f rispetto alla base \mathbf{v} , si ha

$$f(v_j) = \sum_{i=1}^n a_{ij} v_i,$$

per ogni $j = 1, \dots, n$. Ciò significa che A è la matrice di cambiamento di base dalla base $\mathbf{v} = \{v_1, \dots, v_n\}$ alla base $f(\mathbf{v}) = \{f(v_1), \dots, f(v_n)\}$ e dunque le basi \mathbf{v} e $f(\mathbf{v})$ sono equiorientate se e solo se $\det(A) > 0$. Ricordando che $\det(f) = \det(A)$, si conclude. \square

Osservazione 3.4.8. È immediato verificare che la composizione di due automorfismi di V che preservano l'orientamento è ancora un automorfismo che preserva l'orientamento. Analogamente, l'inverso di un automorfismo che preserva l'orientamento è ancora un automorfismo che preserva l'orientamento. Pertanto il sottoinsieme $\text{Aut}_+(V)$ di $\text{Aut}(V)$ formato dagli automorfismi che preservano l'orientamento è un sottogruppo del gruppo degli automorfismi di V .

Esercizi

Esercizio 3.4.1. Si calcoli

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

Esercizio 3.4.2. Si calcoli

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

Esercizio 3.4.3. Si calcoli l'inversa della matrice

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

Esercizio 3.4.4. Si calcoli l'inversa della matrice

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

Esercizio 3.4.5. Si calcoli il rango della seguente matrice mediante il calcolo dei suoi minori:

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

Esercizio 3.4.6. Utilizzando il metodo di eliminazione di Gauss, si calcoli il determinante della seguente matrice $n \times n$:

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 1 & \cdots & 1 \\ 1 & 1 & 3 & \cdots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 1 & \cdots & n \end{pmatrix}$$

Esercizio 3.4.7. Si indichi con A_n la seguente matrice $n \times n$:

$$A_n = \begin{pmatrix} 2 & 1 & 0 & \cdots & 0 \\ 3 & 2 & 1 & \cdots & 0 \\ 0 & 3 & 2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 0 & 0 & \cdots & 3 & 2 \end{pmatrix}$$

Determinare una formula ricorsiva per calcolare $\det A_n$, per ogni intero $n \geq 1$.

Esercizio 3.4.8. Si indichi con D_n (per $n \geq 1$) il determinante della seguente matrice tridiagonale di ordine n :

$$D_n = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 1 & -1 & \cdots & 0 \\ 0 & 1 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & -1 \\ 0 & 0 & \cdots & 1 & 1 \end{pmatrix}$$

Si dimostri che $D_n = D_{n-1} + D_{n-2}$. La successione dei D_n coincide pertanto con la successione dei *numeri di Fibonacci*: 1, 2, 3, 5, 8, 13, ...

Esercizio 3.4.9. Si indichi con A_n la seguente matrice $n \times n$:

$$A_n = \begin{pmatrix} 1 + a_1 & 1 & 1 & \cdots & 1 \\ 1 & 1 + a_2 & 1 & \cdots & 1 \\ 1 & 1 & 1 + a_3 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 + a_n \end{pmatrix}$$

ove $a_1, \dots, a_n \in K$. Si dimostri che, per ogni intero $n \geq 2$, si ha

$$\det A_n = a_1 a_2 \cdots a_n + \sum_{i=1}^n a_1 a_2 \cdots \hat{a}_i \cdots a_n,$$

dove \hat{a}_i significa che l'elemento a_i non compare nel prodotto.

Esercizio 3.4.10. Dati $a, x \in K$, si calcoli il determinante della seguente matrice $n \times n$:

$$\begin{pmatrix} x & a & a & \cdots & a \\ a & x & a & \cdots & a \\ a & a & x & \cdots & a \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a & a & a & \cdots & x \end{pmatrix}$$

Esercizio 3.4.11. Sia A una matrice $n \times n$ e sia A^* la sua matrice aggiunta. Si dimostri la seguente formula dovuta a Cauchy:

$$|A^*| = |A|^{n-1}.$$

Esercizio 3.4.12. Sia K un campo di caratteristica diversa da 2 e sia $A \in M_n(K)$ una matrice antisimmetrica, cioè tale che ${}^t A = -A$. Si dimostri che se n è dispari allora $\det A = 0$.

Esercizio 3.4.13. Data una matrice $A \in M_{m,n}(K)$, si dimostri che $\det({}^t A A) \geq 0$. Si provi inoltre che $\det({}^t A A) > 0$ se e solo se $\text{rk}(A) = n$.

Capitolo 4

Diagonalizzazione degli Endomorfismi

In questo capitolo ci occuperemo del seguente problema: dato un endomorfismo f di uno spazio vettoriale di dimensione finita V , è possibile trovare una base di V rispetto alla quale la matrice di f assuma una qualche forma particolarmente semplice (ad esempio, sia una matrice diagonale)?

Per rispondere a tale domanda introdurremo le nozioni di autovalore e autovettore di un endomorfismo f di uno spazio vettoriale V e determineremo delle condizioni necessarie e sufficienti affinché f sia diagonalizzabile. Per terminare, descriveremo la forma canonica di Jordan di un endomorfismo di uno spazio vettoriale di dimensione finita.

4.1 Autovalori e autovettori

Sia V uno spazio vettoriale di dimensione n sul campo K e sia $f : V \rightarrow V$ una funzione lineare. Nel Capitolo 2 abbiamo visto come, fissando una base di V , sia possibile associare a f una matrice quadrata di ordine n , a coefficienti in K . Naturalmente, a basi diverse di V corrispondono matrici diverse di f , tutte simili tra loro. Ci si può dunque chiedere se sia possibile trovare una base di V in modo tale che la corrispondente matrice di f assuma una qualche forma canonica (particolarmente semplice) prefissata.

Supponiamo, ad esempio, che esista una base di V costituita da vettori v_1, v_2, \dots, v_n tali che $f(v_i) = \lambda_i v_i$, per $i = 1, \dots, n$, per opportuni scalari $\lambda_i \in K$. Rispetto a tale base la matrice di f assumerebbe la seguente forma diagonale

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

Diamo quindi la seguente definizione:

Definizione 4.1.1. Un endomorfismo f di V è *diagonalizzabile* se esiste una base di V tale che la matrice di f rispetto a tale base sia diagonale.

Ricordando che due matrici quadrate si dicono simili quando esse rappresentano lo stesso endomorfismo rispetto a basi diverse, possiamo anche dare la seguente definizione:

Definizione 4.1.2. Una matrice quadrata A è *diagonalizzabile* se essa è simile a una matrice diagonale, cioè se esiste una matrice invertibile S e una matrice diagonale D tale che $A = SDS^{-1}$.

Il nostro obiettivo sarà quindi quello di cercare di determinare sotto quali condizioni una matrice quadrata (o un endomorfismo di uno spazio vettoriale) è diagonalizzabile.

Iniziamo col dare la seguente definizione:

Definizione 4.1.3. Un *autovalore* di un endomorfismo f di V è un elemento $\lambda \in K$ per cui esiste almeno un vettore non nullo $v \in V$ tale che $f(v) = \lambda v$. Un tale vettore v è detto un *autovettore* di f relativo all'autovalore λ .

Dato $\lambda \in K$ indicheremo semplicemente con $\lambda : V \rightarrow V$ l'applicazione λid_V , cioè l'applicazione che manda un vettore v nel vettore λv . L'equazione $f(v) = \lambda v$ può quindi essere riscritta nella seguente forma: $(f - \lambda)(v) = 0$. Da ciò si deduce che l'insieme degli autovettori relativi all'autovalore λ (assieme al vettore nullo) non è altro che il nucleo dell'applicazione lineare

$$f - \lambda : V \rightarrow V, \quad v \mapsto f(v) - \lambda v.$$

Poniamo

$$V_\lambda = \text{Ker}(f - \lambda) = \{v \in V \mid f(v) = \lambda v\}.$$

Si ha dunque:

Proposizione 4.1.4. Per ogni autovalore λ di un endomorfismo f di V , l'insieme

$$V_\lambda = \{v \in V \mid f(v) = \lambda v\}$$

è un sottospazio vettoriale di V . Esso è detto l'autospazio di f relativo all'autovalore λ .

Prima di iniziare lo studio delle principali proprietà degli autovalori e degli autovettori, vediamo come sia possibile determinarli.

Sia dunque f un endomorfismo di uno spazio vettoriale V di dimensione n sul campo K . Dalla definizione segue subito che $\lambda \in K$ è un autovalore di f se e solo se $\text{Ker}(f - \lambda) \neq \{\mathbf{0}\}$, il che equivale a dire che l'applicazione lineare $f - \lambda : V \rightarrow V$ non è iniettiva. Ricordiamo ora che richiedere che $f - \lambda$ non sia iniettiva equivale a richiedere che $\det(f - \lambda) = 0$ (vedi Cap. 3, Proposizione 3.2.30). Possiamo così concludere che λ è un autovalore di f se e solo se $\det(f - \lambda) = 0$.

Per calcolare questo determinante possiamo fissare arbitrariamente una base di V e considerare la corrispondente matrice A associata a f . Se indichiamo con

$$\lambda \cdot \mathbf{1} = \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda \end{pmatrix}$$

la matrice associata all'applicazione lineare λid_V , si ha

$$\det(f - \lambda) = \det(A - \lambda \cdot \mathbf{1}).$$

Diamo ora la seguente definizione:

Definizione 4.1.5. Sia A una matrice quadrata di ordine n a coefficienti in K e x una indeterminata. Il *polinomio caratteristico* di A è

$$P_A(x) = \det(A - x \cdot \mathbf{1}).$$

Osserviamo che, se $A \in M_n(K)$, $P_A(x)$ è un polinomio di grado n a coefficienti in K , il cui monomio di grado più elevato è $(-1)^n x^n$. A tale riguardo, facciamo notare che alcuni autori definiscono il polinomio caratteristico di una matrice A ponendo

$$P_A(x) = \det(x \cdot \mathbf{1} - A) = (-1)^n \det(A - x \cdot \mathbf{1}),$$

per fare in modo che $P_A(x)$ sia un polinomio monico.

Il prossimo risultato mostra che il polinomio caratteristico di una matrice quadrata dipende, in effetti, solo dalla sua classe di simiglianza.

Proposizione 4.1.6. *Sia f un endomorfismo di V e siano A e A' due matrici di f , rispetto a due basi diverse di V . Allora si ha $P_A(x) = P_{A'}(x)$.*

Dimostrazione. Ricordiamo che due matrici A e A' sono associate allo stesso endomorfismo f di V se e solo se esse sono simili, cioè se e solo se esiste una matrice invertibile S tale che $A' = SAS^{-1}$. In tal caso si ha:

$$\begin{aligned} \det(A' - x \cdot \mathbf{1}) &= \det(SAS^{-1} - x \cdot \mathbf{1}) \\ &= \det(S(A - x \cdot \mathbf{1})S^{-1}) \\ &= \det(S) \det(A - x \cdot \mathbf{1}) \det(S^{-1}) \\ &= \det(A - x \cdot \mathbf{1}), \end{aligned}$$

dato che $\det(S^{-1}) = (\det S)^{-1}$. □

In base a questo risultato, possiamo dare la seguente definizione:

Definizione 4.1.7. Il *polinomio caratteristico* $P_f(x)$ di un endomorfismo f di uno spazio vettoriale V di dimensione finita è il polinomio $P_f(x) = \det(f - x)$. Esso coincide con il polinomio caratteristico $P_A(x)$ di una qualsiasi matrice associata a f .

Da quanto detto in precedenza, concludiamo che $\lambda \in K$ è un autovalore di f se e solo se λ è una radice del polinomio caratteristico di f , cioè se e solo se $P_f(\lambda) = 0$.

Osservazione 4.1.8. L'equazione $\det(A - x \cdot \mathbf{1}) = 0$ è detta l'*equazione caratteristica*, o *equazione secolare*, della matrice A .

Osservazione 4.1.9. Notiamo che, poiché gli autovalori di una matrice quadrata di ordine n sono le soluzioni della sua equazione caratteristica, la quale ha grado n , non è detto che una matrice quadrata a coefficienti in un campo K abbia necessariamente degli autovalori in K . Già nel caso in cui $K = \mathbb{R}$ e $n = 2$, è noto che ci sono equazioni di secondo grado che non hanno soluzioni reali. Se invece

K è un campo algebricamente chiuso, come ad esempio il campo \mathbb{C} dei numeri complessi, allora ogni polinomio di grado $n \geq 1$ a coefficienti in K possiede n zeri in K (contati con le appropriate molteplicità), quindi ogni matrice quadrata a coefficienti in un campo algebricamente chiuso possiede degli autovalori.

Osservazione 4.1.10. Abbiamo visto che, se due matrici quadrate sono simili, esse hanno lo stesso polinomio caratteristico. Non vale invece il viceversa. Ad esempio, le seguenti matrici di ordine n

$$\begin{pmatrix} \alpha & 0 & \cdots & 0 & 0 \\ 0 & \alpha & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \alpha & 0 \\ 0 & 0 & \cdots & 0 & \alpha \end{pmatrix} \quad \text{e} \quad \begin{pmatrix} \alpha & 1 & 0 & \cdots & 0 \\ 0 & \alpha & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha & 1 \\ 0 & 0 & \cdots & 0 & \alpha \end{pmatrix}$$

hanno lo stesso polinomio caratteristico $(\alpha - x)^n$, ma, ovviamente, non sono simili, dato che una matrice diagonale è simile solo a sé stessa.

Una volta noti gli autovalori, la determinazione degli autovettori non presenta alcuna difficoltà. Se λ è un autovalore di f (o di una matrice A) gli autovettori corrispondenti sono gli elementi non nulli del sottospazio vettoriale $\text{Ker}(f - \lambda)$. Si tratta dunque di determinare le soluzioni non nulle del seguente sistema di equazioni lineari:

$$(A - \lambda \cdot \mathbf{1}) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \mathbf{0}.$$

Vediamo ora alcuni esempi.

Esempio 4.1.11. Consideriamo la matrice a coefficienti reali

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

Essa corrisponde alla seguente applicazione lineare:

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \quad \begin{pmatrix} a \\ b \end{pmatrix} \mapsto \begin{pmatrix} -b \\ a \end{pmatrix}$$

Il polinomio caratteristico della matrice A è

$$\det(A - x \cdot \mathbf{1}) = \det \begin{pmatrix} -x & -1 \\ 1 & -x \end{pmatrix} = x^2 + 1$$

che non ha zeri reali. La matrice A non ha dunque autovalori reali (ha tuttavia due autovalori complessi, dati da $x_1 = \sqrt{-1}$ e $x_2 = -\sqrt{-1}$).

Esempio 4.1.12. Consideriamo la matrice a coefficienti reali

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

Il polinomio caratteristico di A è

$$\det(A - x \cdot \mathbf{1}) = \det \begin{pmatrix} -x & 4 \\ -1 & 4 - x \end{pmatrix} = x^2 - 4x + 4 = (x - 2)^2.$$

Tale polinomio possiede un unico¹ zero reale $x = 2$. Gli autovettori sono dunque le soluzioni non nulle del seguente sistema:

$$(A - 2 \cdot \mathbf{1}) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{0},$$

cioè

$$\begin{pmatrix} -2 & 4 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Questo sistema si riscrive come segue

$$\begin{cases} -2x_1 + 4x_2 = 0 \\ -x_1 + 2x_2 = 0 \end{cases}$$

e le sue soluzioni sono dunque le soluzioni della singola equazione $x_1 = 2x_2$. Lo spazio delle soluzioni è pertanto

$$V_2 = \left\{ \begin{pmatrix} 2\alpha \\ \alpha \end{pmatrix} \mid \alpha \in \mathbb{R} \right\}$$

il quale ha dimensione 1.

Abbiamo così concluso che l'autospazio relativo all'autovalore $\lambda = 2$ ha dimensione 1. Poiché non ci sono altri autovalori, ciò significa che non esiste una base di \mathbb{R}^2 formata da autovettori di A . Quindi A non è diagonalizzabile.

Esempio 4.1.13. Consideriamo la matrice a coefficienti reali

$$A = \begin{pmatrix} -5 & 8 \\ -4 & 7 \end{pmatrix}.$$

Il polinomio caratteristico di A è

$$\det(A - x \cdot \mathbf{1}) = \det \begin{pmatrix} -5 - x & 8 \\ -4 & 7 - x \end{pmatrix} = x^2 - 2x - 3.$$

Tale polinomio possiede due radici reali $x_1 = -1$ e $x_2 = 3$; questi sono i due autovalori di A .

Consideriamo l'autovalore $\lambda_1 = -1$; i corrispondenti autovettori sono le soluzioni non nulle del seguente sistema:

$$(A - (-1) \cdot \mathbf{1}) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{0},$$

cioè

$$\begin{pmatrix} -4 & 8 \\ -4 & 8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Questo sistema si riscrive come segue

$$\begin{cases} -4x_1 + 8x_2 = 0 \\ -4x_1 + 8x_2 = 0 \end{cases}$$

¹In questo caso, l'unica soluzione dell'equazione $(x - 2)^2 = 0$ deve essere contata con molteplicità 2, cioè considerata come due soluzioni coincidenti.

e le soluzioni sono quindi quelle della singola equazione $x_1 = 2x_2$. Lo spazio delle soluzioni è dunque

$$V_{-1} = \left\{ \begin{pmatrix} 2\alpha \\ \alpha \end{pmatrix} \mid \alpha \in \mathbb{R} \right\}$$

il quale ha dimensione 1. Una base di tale sottospazio è costituita, ad esempio, dal vettore $v_1 = (2, 1)$.

Consideriamo ora l'autovalore $\lambda_2 = 3$; i corrispondenti autovettori sono le soluzioni non nulle del seguente sistema:

$$(A - 3 \cdot \mathbf{1}) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{0},$$

cioè

$$\begin{pmatrix} -8 & 8 \\ -4 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Questo sistema si riscrive come segue

$$\begin{cases} -8x_1 + 8x_2 = 0 \\ -4x_1 + 4x_2 = 0 \end{cases}$$

e le sue soluzioni sono quindi le soluzioni della singola equazione $x_1 = x_2$. Lo spazio delle soluzioni è dunque

$$V_3 = \left\{ \begin{pmatrix} \beta \\ \beta \end{pmatrix} \mid \beta \in \mathbb{R} \right\}$$

il quale ha dimensione 1. Una base di tale sottospazio è costituita, ad esempio, dal vettore $v_2 = (1, 1)$.

Si può facilmente verificare che i vettori $v_1 = (2, 1)$ e $v_2 = (1, 1)$ sono linearmente indipendenti, quindi formano una base di \mathbb{R}^2 .

In conclusione, esiste una base di \mathbb{R}^2 formata da autovettori di A , quindi A è diagonalizzabile.

Vediamo ora di studiare più in dettaglio alcune proprietà degli autovalori e degli autovettori di un endomorfismo f di V .

Definizione 4.1.14. Sia f un endomorfismo di uno spazio vettoriale di dimensione finita V e sia $P_f(x)$ il suo polinomio caratteristico. Sia $\lambda \in K$ un autovalore di f . La *molteplicità (algebrica)* di λ è il più grande intero m tale che $(x - \lambda)^m$ divida $P_f(x)$. La dimensione dell'autospazio $V_\lambda = \text{Ker}(f - \lambda)$ è detta la *molteplicità geometrica* (o la *nullità*) di λ .

Proposizione 4.1.15. *Autovettori relativi ad autovalori distinti sono linearmente indipendenti, cioè: siano $\lambda_1, \dots, \lambda_r$ autovalori di un endomorfismo f , a due a due distinti, e sia v_i un autovettore relativo all'autovalore λ_i , per $i = 1, \dots, r$. Allora i vettori v_1, \dots, v_r sono linearmente indipendenti.*

Dimostrazione. Dimostriamo l'asserto per induzione su r . Se $r = 1$ si ha un solo autovettore v_1 il quale, essendo non nullo, è linearmente indipendente. Supponiamo quindi che l'asserto sia vero per $r - 1$ autovettori. Consideriamo gli r autovettori v_1, \dots, v_r e consideriamo una loro combinazione lineare

$$\alpha_1 v_1 + \dots + \alpha_{r-1} v_{r-1} + \alpha_r v_r = \mathbf{0}. \quad (4.1.1)$$

Applicando l'endomorfismo f , e ricordando che $f(v_i) = \lambda_i v_i$, si ottiene

$$\alpha_1 \lambda_1 v_1 + \cdots + \alpha_{r-1} \lambda_{r-1} v_{r-1} + \alpha_r \lambda_r v_r = \mathbf{0}. \quad (4.1.2)$$

Moltiplicando la (4.1.1) per λ_r e sottraendo la (4.1.2) si ottiene

$$\alpha_1 (\lambda_r - \lambda_1) v_1 + \cdots + \alpha_{r-1} (\lambda_r - \lambda_{r-1}) v_{r-1} = \mathbf{0}.$$

Per ipotesi induttiva i vettori v_1, \dots, v_{r-1} sono linearmente indipendenti, quindi si deve avere

$$\alpha_1 (\lambda_r - \lambda_1) = \alpha_2 (\lambda_r - \lambda_2) = \cdots = \alpha_{r-1} (\lambda_r - \lambda_{r-1}) = 0.$$

Poiché gli autovalori λ_i sono a due a due distinti, si deduce che

$$\alpha_1 = \alpha_2 = \cdots = \alpha_{r-1} = 0.$$

Sostituendo questi valori nell'equazione (4.1.1), essa si riduce a $\alpha_r v_r = \mathbf{0}$, da cui segue $\alpha_r = 0$. Abbiamo così dimostrato che i vettori v_1, \dots, v_r sono linearmente indipendenti. \square

Le due molteplicità, algebrica e geometrica, di un autovalore di un endomorfismo soddisfano la seguente proprietà:

Proposizione 4.1.16. *Sia V uno spazio vettoriale di dimensione finita su K e sia f un endomorfismo di V . Sia $\lambda \in K$ un autovalore di f di molteplicità algebrica m . Allora si ha*

$$\dim V_\lambda \leq m,$$

cioè la molteplicità geometrica di un autovalore è minore o uguale della sua molteplicità algebrica.

Dimostrazione. Sia $r = \dim V_\lambda$ e sia v_1, \dots, v_r una base di V_λ . Completiamo, in modo arbitrario, tale base ad una base $v_1, \dots, v_r, v_{r+1}, \dots, v_n$ di V . Rispetto a questa base, la matrice A di f assume la seguente forma a blocchi

$$A = \begin{pmatrix} \lambda \cdot \mathbf{1}_r & B \\ \mathbf{0} & C \end{pmatrix}$$

ove $\mathbf{1}_r$ è la matrice identica di ordine r , B è una matrice con r righe e $n - r$ colonne e C è una matrice quadrata di ordine $n - r$.

La matrice A può dunque essere usata per calcolare il polinomio caratteristico di f , ottenendo

$$P_f(x) = \det(A - x \cdot \mathbf{1}) = \det((\lambda - x) \cdot \mathbf{1}_r) \det(C - x \cdot \mathbf{1}) = (\lambda - x)^r \det(C - x \cdot \mathbf{1}).$$

Da ciò si deduce che λ è una radice di molteplicità $\geq r$ del polinomio $P_f(x)$, pertanto si ha $m \geq r = \dim V_\lambda$, che è ciò che si voleva dimostrare. \square

Siamo ora in grado di dimostrare il seguente risultato:

Teorema 4.1.17. *Sia $f : V \rightarrow V$ un endomorfismo di uno spazio vettoriale V di dimensione n sul campo K . Indichiamo con $\lambda_1, \lambda_2, \dots, \lambda_r$, gli autovalori di f in K e con m_1, m_2, \dots, m_r le rispettive molteplicità algebriche. Allora f è diagonalizzabile se e solo se $m_1 + m_2 + \cdots + m_r = n$ e, per ogni autovalore λ_i , la sua molteplicità geometrica coincide con la molteplicità algebrica.*

Dimostrazione. Se f è diagonalizzabile esiste una base di V rispetto alla quale la matrice di f è una matrice a blocchi del tipo

$$\begin{pmatrix} \lambda_1 \cdot \mathbf{1}_{m_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \lambda_2 \cdot \mathbf{1}_{m_2} & & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \lambda_r \cdot \mathbf{1}_{m_r} \end{pmatrix}$$

Utilizzando questa matrice per calcolare il polinomio caratteristico di f , si trova

$$P_f(x) = (\lambda_1 - x)^{m_1} (\lambda_2 - x)^{m_2} \cdots (\lambda_r - x)^{m_r},$$

da cui si deduce che $m_1 + m_2 + \cdots + m_r = \deg P_f(x) = n$. Inoltre, poiché f è diagonalizzabile, esiste una base di V costituita da autovettori di f , da cui segue che

$$\dim V_{\lambda_1} + \dim V_{\lambda_2} + \cdots + \dim V_{\lambda_r} = n = \dim V.$$

Viceversa, supponiamo che esistano r autovalori di f in K , $\lambda_1, \lambda_2, \dots, \lambda_r$, di molteplicità algebrica rispettivamente m_1, m_2, \dots, m_r , tali che $m_1 + m_2 + \cdots + m_r = n$ e che, per ogni autovalore λ_i di f , la sua molteplicità geometrica coincida con la molteplicità algebrica; dobbiamo dimostrare che esiste una base di V costituita da autovettori di f .

Dato che, per ipotesi, $\dim V_{\lambda_i} = m_i$, per ogni autovalore λ_i , esiste una base $v_1^{(i)}, v_2^{(i)}, \dots, v_{m_i}^{(i)}$ dell'autospazio V_{λ_i} , per ogni $i = 1, \dots, r$. Poiché, sempre per ipotesi, è $m_1 + \cdots + m_r = n$, l'insieme

$$\{v_1^{(1)}, \dots, v_{m_1}^{(1)}, v_1^{(2)}, \dots, v_{m_2}^{(2)}, \dots, v_1^{(r)}, \dots, v_{m_r}^{(r)}\}$$

contiene esattamente n vettori, i quali sono linearmente indipendenti, come si verifica facilmente ricordando che autovettori relativi ad autovalori distinti sono linearmente indipendenti (Proposizione 4.1.15). Questi n vettori costituiscono quindi una base di V . Abbiamo così costruito una base di V formata da autovettori di f ; f è pertanto diagonalizzabile. \square

Osservazione 4.1.18. In base a quanto visto, possiamo affermare che un endomorfismo f di uno spazio vettoriale V di dimensione finita su K è diagonalizzabile se e solo se f possiede autovalori $\lambda_1, \dots, \lambda_r \in K$ e si ha

$$V = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_r}.$$

4.2 La forma canonica di Jordan

Sia V uno spazio vettoriale di dimensione n sul campo \mathbb{C} dei numeri complessi e sia $f : V \rightarrow V$ una funzione lineare.

Per ogni $\lambda \in \mathbb{C}$ indicheremo semplicemente con $\lambda : V \rightarrow V$ l'applicazione lineare data da $v \mapsto \lambda v$ (sarebbe più corretto, ma più noioso, indicare tale funzione con λid_V , ove id_V è l'applicazione identica di V in sé).

Ricordiamo che $\lambda \in \mathbb{C}$ è un autovalore di f se $\text{Ker}(f - \lambda) \neq \{\mathbf{0}\}$ e che un vettore non nullo $v \in V$ è un autovettore associato all'autovalore λ se $v \in \text{Ker}(f - \lambda)$, cioè se $f(v) = \lambda v$. Se λ è un autovalore di f , il sottospazio vettoriale $\text{Ker}(f - \lambda)$ di V è l'autospazio relativo all'autovalore λ .

Ora generalizzeremo la nozione di autovettore. Per ogni intero $m > 0$ indicheremo con $(f - \lambda)^m$ l'applicazione lineare composta

$$\underbrace{(f - \lambda) \circ (f - \lambda) \circ \cdots \circ (f - \lambda)}_m$$

Definizione 4.2.1. Sia $\lambda \in \mathbb{C}$ un autovalore di $f : V \rightarrow V$. Un vettore non nullo $v \in V$ è detto un *autovettore generalizzato* di f , relativo all'autovalore λ , se $v \in \text{Ker}(f - \lambda)^m$, per qualche $m > 0$.

Il minimo m per cui $v \in \text{Ker}(f - \lambda)^m$ è detto il *periodo* di v (se $m = 1$, v è un autovettore ordinario).

Si noti che valgono le seguenti inclusioni

$$\text{Ker}(f - \lambda) \subseteq \text{Ker}(f - \lambda)^2 \subseteq \cdots \subseteq \text{Ker}(f - \lambda)^m \subseteq \cdots \subseteq V. \quad (4.2.1)$$

Indicheremo con

$$V_\lambda = \bigcup_{m>0} \text{Ker}(f - \lambda)^m$$

il sottospazio vettoriale di V costituito dagli autovettori generalizzati relativi all'autovalore λ . Naturalmente si ha $V_\lambda = \text{Ker}(f - \lambda)^r$, per r sufficientemente grande.

Lemma 4.2.2. *Sia v un autovettore generalizzato per f , relativo all'autovalore λ , e sia m il periodo di v (ciò significa che $v \in \text{Ker}(f - \lambda)^m$ ma $v \notin \text{Ker}(f - \lambda)^{m-1}$). Allora gli m vettori*

$$v, (f - \lambda)(v), (f - \lambda)^2(v), \dots, (f - \lambda)^{m-1}(v)$$

sono linearmente indipendenti.

Dimostrazione. Consideriamo una combinazione lineare

$$\alpha_0 v + \alpha_1 (f - \lambda)(v) + \alpha_2 (f - \lambda)^2(v) + \cdots + \alpha_{m-1} (f - \lambda)^{m-1}(v) = \mathbf{0}.$$

Se applichiamo $(f - \lambda)^{m-1}$ ad ambo i membri (e ricordiamo che $(f - \lambda)^m(v) = \mathbf{0}$), otteniamo

$$\alpha_0 (f - \lambda)^{m-1}(v) + \mathbf{0} = \mathbf{0},$$

da cui segue $\alpha_0 = 0$. La precedente combinazione lineare si riduce pertanto a

$$\alpha_1 (f - \lambda)(v) + \alpha_2 (f - \lambda)^2(v) + \cdots + \alpha_{m-1} (f - \lambda)^{m-1}(v) = \mathbf{0}.$$

Applicando ora $(f - \lambda)^{m-2}$ ad ambo i membri, si ottiene

$$\alpha_1 (f - \lambda)^{m-1}(v) + \mathbf{0} = \mathbf{0},$$

da cui segue $\alpha_1 = 0$. Continuando in questo modo si dimostra che tutti i coefficienti α_i sono nulli. \square

Da questo risultato discende che ogni autovettore generalizzato per f ha periodo $\leq n$. Infatti, se un autovettore generalizzato v avesse periodo $m > n$, gli m vettori $v, (f - \lambda)(v), \dots, (f - \lambda)^{m-1}(v)$ sarebbero linearmente indipendenti, il che è assurdo dato che $m > n = \dim V$. Quindi si ha

$$V_\lambda = \text{Ker}(f - \lambda)^n.$$

Il risultato seguente è una generalizzazione, al caso degli autovettori generalizzati, del fatto che autovettori relativi ad autovalori distinti sono linearmente indipendenti.

Proposizione 4.2.3. *Siano V ed f come sopra e siano $\lambda_1, \dots, \lambda_r$ autovalori di f , a due a due distinti. Per ogni $i = 1, \dots, r$, sia $v_i \in V_{\lambda_i}$ un autovettore generalizzato relativo all'autovalore λ_i . Allora i vettori v_1, \dots, v_r sono linearmente indipendenti.*

Dimostrazione. Per ogni $i = 1, \dots, r$, indichiamo con m_i il periodo di v_i e consideriamo l'endomorfismo $(f - \lambda_i)^{m_i-1}$.

Come prima cosa osserviamo che, per ogni i e j , gli endomorfismi $(f - \lambda_i)^s$ e $(f - \lambda_j)^t$ commutano tra loro, per ogni $s, t \geq 1$. Ciò discende dal fatto che le potenze di f commutano tra loro e f commuta con la moltiplicazione per ogni scalare λ (perché f è lineare). Inoltre, per ogni $i = 1, \dots, r$, il vettore

$$w_i = (f - \lambda_i)^{m_i-1}(v_i)$$

è un autovettore di f relativo all'autovalore λ_i , infatti si ha

$$(f - \lambda_i)(w_i) = (f - \lambda_i)(f - \lambda_i)^{m_i-1}(v_i) = (f - \lambda_i)^{m_i}(v_i) = \mathbf{0}.$$

Da ciò discende che

$$(f - \lambda_j)^m(w_i) = (\lambda_i - \lambda_j)^m w_i,$$

per ogni $i, j = 1, \dots, r$ ed ogni $m \geq 1$.

Consideriamo ora una combinazione lineare

$$\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_r v_r = \mathbf{0}$$

e applichiamo ad ambo i membri l'endomorfismo

$$(f - \lambda_1)^{m_1-1}(f - \lambda_2)^{m_2}(f - \lambda_3)^{m_3} \dots (f - \lambda_r)^{m_r}.$$

Da quanto detto in precedenza si ottiene:

$$\alpha_1(\lambda_1 - \lambda_2)^{m_2}(\lambda_1 - \lambda_3)^{m_3} \dots (\lambda_1 - \lambda_r)^{m_r}(f - \lambda_1)^{m_1-1}(v_1) = \mathbf{0}$$

da cui, ricordando che gli autovalori $\lambda_1, \dots, \lambda_r$ sono a due a due distinti, discende che $\alpha_1 = 0$.

La combinazione lineare precedente si riduce quindi a

$$\alpha_2 v_2 + \dots + \alpha_r v_r = \mathbf{0}.$$

Applicando ora l'endomorfismo

$$(f - \lambda_2)^{m_2-1}(f - \lambda_3)^{m_3} \dots (f - \lambda_r)^{m_r}$$

si ottiene:

$$\alpha_2(\lambda_2 - \lambda_3)^{m_3} \dots (\lambda_2 - \lambda_r)^{m_r}(f - \lambda_2)^{m_2-1}(v_2) = \mathbf{0}$$

da cui segue $\alpha_2 = 0$.

Continuando in questo modo si dimostra che tutti gli α_i , $i = 1, \dots, r$, sono nulli. \square

Ricordiamo che, anche se una funzione lineare $f : V \rightarrow V$ possiede tutti i suoi autovalori nel campo di definizione K (il che accade sempre, se K è algebricamente chiuso), gli autovettori di f potrebbero non generare l'intero spazio vettoriale V . In tal caso non esiste una base di V costituita da autovettori di f , quindi f non è diagonalizzabile. Vedremo ora che con gli autovettori generalizzati un problema del genere non si presenta.

Proposizione 4.2.4. *Sia V uno spazio vettoriale di dimensione finita su \mathbb{C} . Per ogni endomorfismo $f : V \rightarrow V$, esiste una base di V costituita da autovettori generalizzati di f .*

Dimostrazione. Dimostreremo l'enunciato per induzione sulla dimensione di V . Se $\dim V = 1$ ogni funzione lineare $f : V \rightarrow V$ è data dalla moltiplicazione per uno scalare λ , quindi ogni vettore $v \in V$ è autovettore di f .

Supponiamo dunque che il risultato sia vero per tutti gli spazi vettoriali di dimensione strettamente minore di $n = \dim V$. Sia $\lambda \in \mathbb{C}$ un autovalore di f e sia $V_\lambda = \text{Ker}(f - \lambda)^n$ il sottospazio di V costituito dagli autovettori generalizzati di f relativi all'autovalore λ . Se $V = V_\lambda$ la dimostrazione è terminata (si veda il Lemma 4.2.2). In caso contrario poniamo $W_\lambda = \text{Im}(f - \lambda)^n$. Vogliamo dimostrare che $V = V_\lambda \oplus W_\lambda$.

Dato che $\dim V_\lambda + \dim W_\lambda = \dim V$, è sufficiente dimostrare che $V_\lambda \cap W_\lambda = \{\mathbf{0}\}$. Sia dunque $v \in V_\lambda \cap W_\lambda$. Dato che $v \in W_\lambda = \text{Im}(f - \lambda)^n$, si ha $v = (f - \lambda)^n(u)$, per qualche $u \in V$. Ma, dato che $v \in V_\lambda = \text{Ker}(f - \lambda)^n$, si ha $(f - \lambda)^n(v) = (f - \lambda)^{2n}(u) = \mathbf{0}$. Ciò significa che anche u è un autovettore generalizzato per f , relativo all'autovalore λ e, poiché il periodo di ogni autovettore generalizzato è $\leq n$, si deve avere $(f - \lambda)^n(u) = \mathbf{0}$, cioè $v = \mathbf{0}$.

Ora dimostriamo che W_λ è stabile per f , cioè che $f(W_\lambda) \subseteq W_\lambda$. Sia $v \in W_\lambda$, allora $v = (f - \lambda)^n(u)$, per qualche $u \in V$. Si ha

$$f(v) = f((f - \lambda)^n(u)) = (f - \lambda)^n(f(u)),$$

perché f commuta con $(f - \lambda)^n$. Ciò significa quindi che $f(v) \in W_\lambda$, che è quello che volevamo dimostrare.

Possiamo così considerare lo spazio vettoriale W_λ dotato della restrizione della funzione lineare f . Poiché $\dim W_\lambda < \dim V$, per l'ipotesi induttiva si ha che W_λ è generato da autovettori generalizzati di f . Questo conclude la dimostrazione. \square

Possiamo riassumere quanto visto finora nel seguente teorema:

Teorema 4.2.5. *Siano V ed f come sopra. Siano $\lambda_1, \dots, \lambda_r$ tutti gli autovalori di f , che supponiamo essere a due a due distinti, e sia*

$$P(x) = (x - \lambda_1)^{e_1}(x - \lambda_2)^{e_2} \cdots (x - \lambda_r)^{e_r}$$

il polinomio caratteristico di f . Allora si ha:

- (i) $V = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_r}$, ove V_{λ_i} è il sottospazio vettoriale costituito dagli autovettori generalizzati di f relativi all'autovalore λ_i .
- (ii) Ogni V_{λ_i} è stabile per f , cioè si ha

$$f(V_{\lambda_i}) \subseteq V_{\lambda_i},$$

per ogni $i = 1, \dots, r$.

(iii) $\dim V_{\lambda_i} = e_i$, quindi si ha

$$V_{\lambda_i} = \text{Ker}(f - \lambda_i)^n = \text{Ker}(f - \lambda_i)^{e_i},$$

per ogni $i = 1, \dots, r$.

Dimostrazione. (i) Per la proposizione precedente V è generato da autovettori generalizzati di f , cioè $V = V_{\lambda_1} + V_{\lambda_2} + \dots + V_{\lambda_r}$. Poiché abbiamo già dimostrato che autovettori generalizzati corrispondenti ad autovalori distinti sono linearmente indipendenti, tale somma è diretta.

(ii) Ricordiamo che f commuta con gli endomorfismi del tipo $(f - \lambda)^m$, per ogni λ e ogni $m \geq 1$. Se $v \in V_{\lambda_i}$ si ha $(f - \lambda_i)^n(v) = \mathbf{0}$, quindi

$$\mathbf{0} = f((f - \lambda_i)^n(v)) = (f - \lambda_i)^n(f(v)),$$

cioè $f(v) \in V_{\lambda_i}$.

(iii) Da quanto visto al punto (ii) segue che f induce un endomorfismo di V_{λ_i} , per $i = 1, \dots, r$. La restrizione di f a V_{λ_i} ha come unico autovalore λ_i , quindi il suo polinomio caratteristico è $(x - \lambda_i)^{d_i}$, ove $d_i = \dim V_{\lambda_i}$. Dato che il polinomio caratteristico di f è il prodotto dei polinomi caratteristici delle restrizioni di f ai vari sottospazi V_{λ_i} , si deve avere

$$(x - \lambda_1)^{e_1} \cdots (x - \lambda_r)^{e_r} = (x - \lambda_1)^{d_1} \cdots (x - \lambda_r)^{d_r},$$

da cui segue $d_i = e_i$, per ogni $i = 1, \dots, r$. Ricordando il Lemma 4.2.2, si ottiene

$$V_{\lambda_i} = \text{Ker}(f - \lambda_i)^n = \text{Ker}(f - \lambda_i)^{e_i}.$$

□

Osservazione 4.2.6. Il punto (iii) del teorema precedente afferma che il periodo di un autovettore generalizzato di f relativo all'autovalore λ_i è minore o uguale dell'esponente e_i con cui il fattore $(x - \lambda_i)$ compare nel polinomio caratteristico di f , cioè il periodo di ogni autovettore generalizzato è minore o uguale della molteplicità algebrica dell'autovalore corrispondente.

Dai punti (i) e (ii) del teorema precedente si deduce che, se scegliamo una base $\{w_1^{(i)}, w_2^{(i)}, \dots, w_{d_i}^{(i)}\}$ di V_{λ_i} , per $i = 1, \dots, r$, allora l'insieme

$$\{w_1^{(1)}, \dots, w_{d_1}^{(1)}, w_1^{(2)}, \dots, w_{d_2}^{(2)}, \dots, w_1^{(r)}, \dots, w_{d_r}^{(r)}\}$$

è una base di V rispetto alla quale la matrice di f è una matrice a blocchi del tipo

$$A = \begin{pmatrix} A_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \hline \mathbf{0} & A_2 & \cdots & \mathbf{0} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline \mathbf{0} & \mathbf{0} & \cdots & A_r \end{pmatrix}$$

ove ciascuna matrice A_i (per $i = 1, \dots, r$) è la matrice della restrizione di f al sottospazio V_{λ_i} rispetto alla base $\{w_1^{(i)}, w_2^{(i)}, \dots, w_{d_i}^{(i)}\}$.

Concentriamo ora la nostra attenzione sulla restrizione di f ad un singolo sottospazio V_{λ_i} . Sia m_i il massimo dei periodi degli elementi di V_{λ_i} ; è $m_i \leq d_i = \dim V_{\lambda_i}$ e si ha la seguente catena di inclusioni

$$\text{Ker}(f - \lambda_i) \subset \text{Ker}(f - \lambda_i)^2 \subset \cdots \subset \text{Ker}(f - \lambda_i)^{m_i} = V_{\lambda_i},$$

ove tutte le inclusioni sono proprie. Possiamo quindi scegliere la base di V_{λ_i} nel modo seguente: cominciamo scegliendo una base di $\text{Ker}(f - \lambda_i)$, completiamola poi ad una base di $\text{Ker}(f - \lambda_i)^2$, che completeremo a sua volta ad una base di $\text{Ker}(f - \lambda_i)^3$ e così via, sino ad ottenere una base di tutto V_{λ_i} .

Ora osserviamo che, se $v \in \text{Ker}(f - \lambda_i)^k$, si ha

$$f(v) = \lambda_i v + (f - \lambda_i)(v),$$

e $(f - \lambda_i)(v) \in \text{Ker}(f - \lambda_i)^{k-1}$. Da ciò segue che, se $\{w_1^{(i)}, w_2^{(i)}, \dots, w_{d_i}^{(i)}\}$ è la base di V_{λ_i} costruita nel modo appena descritto, si ha

$$f(w_j^{(i)}) = \lambda_i w_j^{(i)} + \sum_{h=1}^{j-1} \alpha_h w_h^{(i)},$$

per ogni $j = 1, \dots, d_i$. Questo significa che la matrice A_i di f rispetto ad una tale base è del tipo

$$A_i = \begin{pmatrix} \lambda_i & * & * & \cdots & * \\ 0 & \lambda_i & * & \cdots & * \\ 0 & 0 & \lambda_i & \cdots & * \\ \vdots & \vdots & \ddots & \ddots & * \\ 0 & 0 & \cdots & 0 & \lambda_i \end{pmatrix}$$

cioè è una matrice triangolare superiore con tutti gli elementi diagonali uguali all'autovalore λ_i .

Questo è un buon risultato: afferma che ogni matrice quadrata A , sul campo \mathbb{C} dei numeri complessi, è simile ad una matrice diagonale a blocchi, dove i blocchi diagonali A_i sono delle matrici triangolari superiori in cui gli elementi sulla diagonale principale sono gli autovalori λ_i della matrice A .

Se però ci impegnamo un po' di più, possiamo ottenere un risultato migliore. Prima di continuare, tuttavia, vogliamo introdurre la nozione di *polinomio minimo* di un endomorfismo di V (o, equivalentemente, di una matrice quadrata).

Siano V ed f come sopra e sia

$$P(x) = (x - \lambda_1)^{e_1} (x - \lambda_2)^{e_2} \cdots (x - \lambda_r)^{e_r}$$

il polinomio caratteristico di f , ove si suppone che gli autovalori $\lambda_1, \dots, \lambda_r$ siano a due a due distinti. Abbiamo già osservato che il periodo di ogni autovettore generalizzato di f , relativo all'autovalore λ_i , è minore o uguale a e_i . Sia dunque m_i il massimo dei periodi degli elementi di V_{λ_i} ; naturalmente è $m_i \leq e_i$, per ogni $i = 1, \dots, r$.

Diamo ora la seguente definizione:

Definizione 4.2.7. Con le notazioni precedentemente introdotte, il polinomio

$$Q(x) = (x - \lambda_1)^{m_1} (x - \lambda_2)^{m_2} \cdots (x - \lambda_r)^{m_r}$$

è detto il *polinomio minimo*² dell'endomorfismo f di V .

Notiamo subito che, poiché $m_i \leq e_i$, per ogni $i = 1, \dots, r$, il polinomio minimo di f divide il suo polinomio caratteristico. Inoltre il polinomio minimo e il polinomio caratteristico di un endomorfismo f hanno gli stessi zeri, che sono precisamente gli autovalori di f .

Possiamo ora fornire un criterio di diagonalizzabilità di un endomorfismo in termini del suo polinomio minimo.

Proposizione 4.2.8. *Siano V ed f come sopra. L'endomorfismo f è diagonalizzabile se e solo se il suo polinomio minimo è prodotto di fattori lineari distinti.*

Dimostrazione. Dire che il polinomio minimo di f è prodotto di fattori lineari distinti equivale a dire che i periodi m_1, \dots, m_r degli autovettori generalizzati sono tutti uguali a 1. Ma ciò equivale a dire che gli autovettori generalizzati sono, in realtà, dei veri e propri autovettori. Dal Teorema 4.2.5 (punto (i)) si deduce quindi l'esistenza di una base di V costituita da autovettori di f , il che equivale a dire che f è diagonalizzabile. \square

Possiamo anche dimostrare il seguente risultato:

Teorema 4.2.9 (HAMILTON–CAYLEY). *Siano V ed f come sopra e sia*

$$Q(x) = (x - \lambda_1)^{m_1} (x - \lambda_2)^{m_2} \cdots (x - \lambda_r)^{m_r}$$

il polinomio minimo di f . Allora si ha $Q(f) = 0$.

Dimostrazione. $Q(f)$ è l'endomorfismo di V dato da

$$(f - \lambda_1)^{m_1} (f - \lambda_2)^{m_2} \cdots (f - \lambda_r)^{m_r}.$$

Se $v \in V_{\lambda_i}$, per qualche $i = 1, \dots, r$, si ha $(f - \lambda_i)^{m_i}(v) = \mathbf{0}$, quindi $Q(f)(v) = \mathbf{0}$. Dato che ogni vettore di V si può scrivere come somma di vettori appartenenti ai vari sottospazi V_{λ_i} (punto (i) del Teorema 4.2.5), si ha $Q(f)(v) = \mathbf{0}$ per ogni $v \in V$, il che significa che $Q(f)$ è l'endomorfismo nullo. \square

Corollario 4.2.10. *Sia A una matrice quadrata di ordine n e sia $Q(x)$ il suo polinomio minimo. Allora $Q(A) = 0$. In particolare, se $P(x)$ è il polinomio caratteristico di A , si ha anche $P(A) = 0$ (basta ricordare che $Q(x)$ divide $P(x)$).*

Ritorniamo ora al problema di migliorare la scelta della base di ciascun sottospazio V_{λ_i} , al fine di semplificare ulteriormente la forma delle matrici A_i che compaiono come blocchi diagonali della matrice A di f .

Concentriamo la nostra attenzione su un singolo autovalore, che indicheremo con λ , e sul relativo sottospazio di autovettori generalizzati V_λ . Indichiamo con m il massimo dei periodi degli elementi di V_λ e consideriamo la catena di inclusioni proprie

$$\text{Ker}(f - \lambda) \subset \text{Ker}(f - \lambda)^2 \subset \cdots \subset \text{Ker}(f - \lambda)^m = V_\lambda.$$

²Il nome *polinomio minimo* deriva dal fatto che esso è il polinomio monico di grado minimo che si annulla su f , cioè tale che $Q(f)$ è l'endomorfismo nullo.

Per ogni $j = 1, \dots, m$, poniamo $d_j = \dim \text{Ker}(f - \lambda)^j$.

Preso un vettore non nullo $v \in \text{Ker}(f - \lambda)^m \setminus \text{Ker}(f - \lambda)^{m-1}$, gli m vettori $w_m = v, w_{m-1} = (f - \lambda)(v), w_{m-2} = (f - \lambda)^2(v), \dots, w_1 = (f - \lambda)^{m-1}(v)$, sono linearmente indipendenti (vedi Lemma 4.2.2). Notiamo che $(f - \lambda)(w_1) = (f - \lambda)^m(v) = \mathbf{0}$, cioè

$$f(w_1) = \lambda w_1,$$

mentre

$$(f - \lambda)(w_j) = (f - \lambda)(f - \lambda)^{m-j}(v) = (f - \lambda)^{m-j+1}(v) = w_{j-1},$$

cioè

$$f(w_j) = \lambda w_j + w_{j-1},$$

per $j = 2, \dots, m$.

Ciò significa che f induce un endomorfismo del sottospazio di V_λ generato dai vettori w_1, w_2, \dots, w_m , la cui matrice, rispetto alla base $\{w_1, w_2, \dots, w_m\}$ di tale sottospazio è la seguente matrice quadrata di ordine m :

$$J_\lambda = \begin{pmatrix} \lambda & 1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & 0 & \cdots & 0 \\ 0 & 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda & 1 \\ 0 & 0 & \cdots & 0 & 0 & \lambda \end{pmatrix}$$

Una matrice di questo tipo è detta un *blocco di Jordan* di ordine m relativo all'autovalore λ .

Sia $s = d_m - d_{m-1} = \dim \text{Ker}(f - \lambda)^m - \dim \text{Ker}(f - \lambda)^{m-1}$. Ricordiamo che $s \geq 1$. Vogliamo dimostrare che la matrice della restrizione di f al sottospazio V_λ contiene, lungo la diagonale principale, s blocchi di Jordan J_λ di ordine m . A tal fine consideriamo s vettori

$$v_1, v_2, \dots, v_s \in \text{Ker}(f - \lambda)^m \setminus \text{Ker}(f - \lambda)^{m-1}$$

tali che si abbia

$$\text{Ker}(f - \lambda)^m = \text{Ker}(f - \lambda)^{m-1} \oplus \langle v_1, \dots, v_s \rangle. \quad (4.2.2)$$

Applicando a ciascuno dei vettori v_1, \dots, v_s il procedimento descritto in precedenza, otteniamo il seguente insieme di vettori:

$$\begin{aligned} v_1, (f - \lambda)(v_1), (f - \lambda)^2(v_1), \dots, (f - \lambda)^{m-1}(v_1), \\ v_2, (f - \lambda)(v_2), (f - \lambda)^2(v_2), \dots, (f - \lambda)^{m-1}(v_2), \\ \vdots \quad \vdots \\ v_s, (f - \lambda)(v_s), (f - \lambda)^2(v_s), \dots, (f - \lambda)^{m-1}(v_s) \end{aligned} \quad (4.2.3)$$

(ciascuna di queste righe corrisponde a un blocco di Jordan di ordine m).

Si tratta ora di dimostrare che tutti questi vettori sono linearmente indipendenti. Consideriamo una loro combinazione lineare

$$\begin{aligned} \alpha_1^{(0)} v_1 + \cdots + \alpha_s^{(0)} v_s + \alpha_1^{(1)} (f - \lambda)(v_1) + \cdots + \alpha_s^{(1)} (f - \lambda)(v_s) + \\ + \alpha_1^{(2)} (f - \lambda)^2(v_1) + \cdots + \alpha_s^{(2)} (f - \lambda)^2(v_s) + \cdots \\ \cdots + \alpha_1^{(m-1)} (f - \lambda)^{m-1}(v_1) + \cdots + \alpha_s^{(m-1)} (f - \lambda)^{m-1}(v_s) = \mathbf{0}. \end{aligned}$$

Lasciando solo $\alpha_1^{(0)}v_1 + \cdots + \alpha_s^{(0)}v_s$ a primo membro e portando tutti gli altri addendi al secondo membro si deduce che

$$\alpha_1^{(0)}v_1 + \cdots + \alpha_s^{(0)}v_s \in \text{Ker}(f - \lambda)^{m-1},$$

da cui segue $\alpha_1^{(0)} = \alpha_2^{(0)} = \cdots = \alpha_s^{(0)} = 0$ (si ricordi che, in base a (4.2.2), si ha $\text{Ker}(f - \lambda)^{m-1} \cap \langle v_1, \dots, v_s \rangle = \{\mathbf{0}\}$).

La combinazione lineare precedente si riduce quindi a

$$\begin{aligned} & \alpha_1^{(1)}(f - \lambda)(v_1) + \cdots + \alpha_s^{(1)}(f - \lambda)(v_s) + \\ & + \alpha_1^{(2)}(f - \lambda)^2(v_1) + \cdots + \alpha_s^{(2)}(f - \lambda)^2(v_s) + \cdots \\ & \cdots + \alpha_1^{(m-1)}(f - \lambda)^{m-1}(v_1) + \cdots + \alpha_s^{(m-1)}(f - \lambda)^{m-1}(v_s) = \mathbf{0}. \end{aligned}$$

Procedendo come prima si ha

$$\alpha_1^{(1)}(f - \lambda)(v_1) + \cdots + \alpha_s^{(1)}(f - \lambda)(v_s) \in \text{Ker}(f - \lambda)^{m-2},$$

quindi

$$\alpha_1^{(1)}v_1 + \cdots + \alpha_s^{(1)}v_s \in \text{Ker}(f - \lambda)^{m-1},$$

da cui, esattamente come prima, segue che $\alpha_1^{(1)} = \alpha_2^{(1)} = \cdots = \alpha_s^{(1)} = 0$.

Continuando in questo modo si dimostra che tutti i coefficienti $\alpha_j^{(i)}$ sono nulli, il che conclude la dimostrazione dell'indipendenza lineare.

Ora osserviamo che se $d_m - d_{m-1} = s$ allora $d_i - d_{i-1} \geq s$ per ogni $i = 1, \dots, m$ (ove si intende che $d_0 = \dim \text{Ker}(f - \lambda)^0 = 0$). Infatti, dato che i vettori v_1, \dots, v_s hanno periodo m , i vettori

$$(f - \lambda)^{m-i}(v_1), (f - \lambda)^{m-i}(v_2), \dots, (f - \lambda)^{m-i}(v_s)$$

hanno periodo esattamente i e, come abbiamo appena visto, sono linearmente indipendenti.

Se accade che $d_i - d_{i-1} = s$, per ogni $i = 1, \dots, m$, si ha

$$\begin{aligned} \dim \text{Ker}(f - \lambda) &= s \\ \dim \text{Ker}(f - \lambda)^2 &= 2s \\ &\vdots \\ \dim \text{Ker}(f - \lambda)^m &= ms \end{aligned}$$

e, dato che $V_\lambda = \text{Ker}(f - \lambda)^m$, gli ms vettori di cui in (4.2.3) sono una base di V_λ . Ricordando quanto visto in precedenza è ora immediato verificare che la matrice della restrizione di f a V_λ consiste di s blocchi di Jordan di ordine m :

$$\left(\begin{array}{c|c|c|c} J_\lambda & \mathbf{0} & \cdots & \mathbf{0} \\ \hline \mathbf{0} & J_\lambda & \cdots & \mathbf{0} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline \mathbf{0} & \mathbf{0} & \cdots & J_\lambda \end{array} \right)$$

Se invece si ha $d_i - d_{i-1} > s$, per qualche i , indichiamo con j il massimo indice $< m$ tale che $d_j - d_{j-1} = t > s$ e prendiamo dei vettori w_1, w_2, \dots, w_{t-s} tali che si abbia

$$\begin{aligned} \text{Ker}(f - \lambda)^j &= \text{Ker}(f - \lambda)^{j-1} \oplus \langle (f - \lambda)^{m-j}(v_1), \dots, (f - \lambda)^{m-j}(v_s) \rangle \\ &\quad \oplus \langle w_1, \dots, w_{t-s} \rangle. \end{aligned}$$

Ragionando in modo analogo a quanto fatto in precedenza, si dimostra che i vettori

$$\begin{aligned} w_1, (f - \lambda)(w_1), (f - \lambda)^2(w_1), \dots, (f - \lambda)^{j-1}(w_1), \\ w_2, (f - \lambda)(w_2), (f - \lambda)^2(w_2), \dots, (f - \lambda)^{j-1}(w_2), \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ w_{t-s}, (f - \lambda)(w_{t-s}), (f - \lambda)^2(w_{t-s}), \dots, (f - \lambda)^{j-1}(w_{t-s}) \end{aligned}$$

sono linearmente indipendenti. Da ciò segue che, in questo caso, la matrice della restrizione di f a V_λ contiene anche, lungo la diagonale principale, $t - s$ blocchi di Jordan relativi all'autovalore λ , di ordine j (oltre agli s blocchi J_λ di ordine m già menzionati).

Procedendo in modo analogo con gli autovettori generalizzati di periodo via via minore, si arriva a concludere che, in generale, la matrice della restrizione di f a V_λ è una matrice diagonale a blocchi, in cui i blocchi diagonali sono dei blocchi di Jordan relativi all'autovalore λ di ordine $\leq m$, ove m è il massimo dei periodi degli elementi di V_λ (inoltre esiste almeno un blocco di Jordan di ordine esattamente m). Tale intero m non è altro che l'esponente con cui il fattore $(x - \lambda)$ compare nel polinomio minimo di f .

Abbiamo così dimostrato il seguente teorema:

Teorema 4.2.11 (JORDAN). *Sia V uno spazio vettoriale di dimensione n sul campo³ \mathbb{C} e sia $f : V \rightarrow V$ una funzione lineare. Sia*

$$P(x) = (x - \lambda_1)^{e_1}(x - \lambda_2)^{e_2} \cdots (x - \lambda_r)^{e_r}$$

il polinomio caratteristico di f (in cui si suppone che gli autovalori $\lambda_1, \dots, \lambda_r$ siano a due a due distinti). Allora esiste una base di V rispetto alla quale la matrice di f è una matrice a blocchi del tipo

$$J = \begin{pmatrix} J_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \hline \mathbf{0} & J_2 & \cdots & \mathbf{0} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline \mathbf{0} & \mathbf{0} & \cdots & J_r \end{pmatrix}$$

³Un risultato analogo vale, in realtà, su ogni campo algebricamente chiuso.

ove ogni J_i è, a sua volta, una matrice a blocchi del tipo

$$J_i = \left(\begin{array}{c|c|c|c} J_{\lambda_i}^1 & \mathbf{0} & \cdots & \mathbf{0} \\ \hline \mathbf{0} & J_{\lambda_i}^2 & \cdots & \mathbf{0} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline \mathbf{0} & \mathbf{0} & \cdots & J_{\lambda_i}^{s_i} \end{array} \right)$$

ove $J_{\lambda_i}^1, \dots, J_{\lambda_i}^{s_i}$ sono opportuni blocchi di Jordan relativi all'autovalore λ_i .

Una matrice J di questo tipo è detta *forma canonica di Jordan* e una base di V rispetto a cui f ha questa matrice è detta *base di Jordan*.

Cerchiamo ora di chiarire quanto sopra esposto applicando i risultati finora ottenuti in un esempio concreto.

Esercizio. Sia V uno spazio vettoriale di dimensione 5 sul campo complesso⁴ \mathbb{C} e sia $\{v_1, \dots, v_5\}$ una sua base. Sia $f : V \rightarrow V$ l'applicazione lineare di matrice

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

rispetto alla base data. Si determinino il polinomio caratteristico e il polinomio minimo di f . Si determini inoltre una matrice di Jordan di f e una base di Jordan di V .

Soluzione. Il polinomio caratteristico di f è

$$\begin{aligned} P(x) &= \det(x \cdot \mathbf{1} - A) = \begin{vmatrix} x-1 & 0 & 2 & 0 & 0 \\ 2 & x-3 & 2 & 0 & 0 \\ -2 & 0 & x-5 & 0 & 0 \\ 0 & -2 & -2 & x-2 & 1 \\ 0 & 0 & 0 & -1 & x-4 \end{vmatrix} \\ &= \begin{vmatrix} x-1 & 0 & 2 \\ 2 & x-3 & 2 \\ -2 & 0 & x-5 \end{vmatrix} \cdot \begin{vmatrix} x-2 & 1 \\ -1 & x-4 \end{vmatrix} \\ &= (x-3)[(x-1)(x-5) + 4][(x-2)(x-4) + 1] \\ &= (x-3)(x-3)^2(x-3)^2 = (x-3)^5, \end{aligned}$$

quindi f ha un unico autovalore $\lambda = 3$, con molteplicità 5. Poiché il polinomio minimo $Q(x)$ deve dividere il polinomio caratteristico, deve essere

$$Q(x) = (x-3)^m, \quad \text{con } 1 \leq m \leq 5.$$

⁴Come vedremo, in questo esempio non è necessario supporre che il campo sia algebricamente chiuso. Tutto quello che diremo vale anche per il campo \mathbb{Q} dei numeri razionali.

Osserviamo che

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

$$(A - 3)^2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & -2 & 0 & 0 \\ 0 & 2 & 2 & 0 & 0 \end{pmatrix}$$

e infine $(A - 3)^3 = \mathbf{0}$, quindi il polinomio minimo è $Q(x) = (x - 3)^3$. Questo significa che il massimo periodo degli autovettori generalizzati di f (relativi all'unico autovalore 3) è $m = 3$, quindi si ha la catena di inclusioni proprie

$$\text{Ker}(f - 3) \subset \text{Ker}(f - 3)^2 \subset \text{Ker}(f - 3)^3 = V.$$

Si ha:

$$\dim \text{Ker}(f - 3) = 5 - \text{rk}(A - 3 \cdot \mathbf{1}) = 5 - 3 = 2,$$

$$\dim \text{Ker}(f - 3)^2 = 5 - \text{rk}(A - 3 \cdot \mathbf{1})^2 = 5 - 1 = 4,$$

$$\dim \text{Ker}(f - 3)^3 = \dim V = 5.$$

Ponendo $d_i = \dim \text{Ker}(f - 3)^i$, l'intero $s = d_m - d_{m-1} = d_3 - d_2$ della dimostrazione del teorema di Jordan è, in questo caso, $s = 5 - 4 = 1$, quindi la matrice di Jordan J contiene un blocco di Jordan di ordine $m = 3$, relativo all'autovalore $\lambda = 3$,

$$\begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}.$$

Dato che si ha poi $\dim \text{Ker}(f - 3)^2 - \dim \text{Ker}(f - 3) = 4 - 2 = 2$, usando le stesse notazioni impiegate nella dimostrazione del teorema di Jordan, si ha $j = 2$ e $d_2 - d_1 = t = 2 > s = 1$, quindi $t - s = 1$ e la matrice di Jordan J contiene anche un blocco di Jordan di ordine $j = 2$

$$\begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}.$$

In conclusione, la matrice di Jordan di f è la seguente:

$$J = \begin{array}{cc|cc} 3 & 1 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ \hline 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 3 \end{array}$$

Rimane ora solo da determinare una base $\{w_1, \dots, w_5\}$ di V rispetto alla quale la matrice di f sia J . A tal fine scegliamo un vettore $v \in \text{Ker}(f - 3)^3$ tale che $v \notin \text{Ker}(f - 3)^2$: ad esempio, il vettore v_2 soddisfa a tali richieste. Poniamo allora

$$w_3 = v_2, \quad w_2 = (f - 3)(v_2) = 2v_4, \quad w_1 = (f - 3)^2(v_2) = -2v_4 + 2v_5.$$

I vettori $\{w_1, w_2, w_3\}$ sono linearmente indipendenti e formano quella parte della base di V che è responsabile della presenza del blocco di Jordan di ordine 3.

Per continuare dobbiamo ora scegliere un vettore u tale che si abbia

$$\text{Ker}(f - 3)^2 = \text{Ker}(f - 3) \oplus \langle (f - 3)^{3-2}(v_2) \rangle \oplus \langle u \rangle,$$

cioè u deve essere un vettore in $\text{Ker}(f - 3)^2$ che non appartenga al sottospazio generato da $\text{Ker}(f - 3)$ e dal vettore $(f - 3)(v_2) = w_2 = 2v_4$.

A tal fine determiniamo, innanzitutto, il nucleo di $f - 3$: esso è dato dai vettori $x_1v_1 + x_2v_2 + \dots + x_5v_5$ tali che le x_i siano soluzioni del seguente sistema di equazioni lineari:

$$\begin{cases} -2x_1 - 2x_3 = 0 \\ 2x_2 + 2x_3 - x_4 - x_5 = 0 \\ x_4 + x_5 = 0. \end{cases}$$

Risolvendo tale sistema si trova

$$\begin{cases} x_1 = -x_3 \\ x_2 = -x_3 \\ x_4 = -x_5 \end{cases}$$

quindi una base di $\text{Ker}(f - 3)$ è formata dai vettori $v_1 + v_2 - v_3$ e $v_4 - v_5$. Si ha pertanto

$$\begin{aligned} \text{Ker}(f - 3) \oplus \langle w_2 \rangle &= \langle v_1 + v_2 - v_3, v_4 - v_5, 2v_4 \rangle \\ &= \langle v_1 + v_2 - v_3, v_4, v_5 \rangle. \end{aligned}$$

Si può ora vedere facilmente che il vettore v_1 non appartiene a questo sottospazio, ma appartiene al nucleo di $(f - 3)^2$: questo è il vettore u che cercavamo. I due vettori mancanti per completare la base di Jordan sono quindi

$$w_5 = v_1, \quad w_4 = (f - 3)(v_1) = -2v_1 - 2v_2 + 2v_3.$$

In conclusione, i vettori

$$\begin{aligned} w_1 &= -2v_4 + 2v_5 \\ w_2 &= 2v_4 \\ w_3 &= v_2 \\ w_4 &= -2v_1 - 2v_2 + 2v_3 \\ w_5 &= v_1 \end{aligned}$$

formano una base di V rispetto alla quale la funzione lineare f ha matrice

$$J = \begin{pmatrix} 3 & 1 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

Per terminare, osserviamo che la matrice di cambiamento di base (cioè la matrice le cui colonne sono costituite dalle coordinate dei vettori della nuova base

$\{w_1, \dots, w_5\}$ rispetto alla vecchia base $\{v_1, \dots, v_5\}$) è la matrice

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

Si ha pertanto

$$J = P^{-1}AP$$

o, equivalentemente,

$$PJ = AP,$$

come si può facilmente verificare con un calcolo diretto.

Esercizi

Esercizio 4.1. Si determini la forma canonica di Jordan dell'endomorfismo $\phi : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ la cui matrice, rispetto alla base canonica, è

$$A = \begin{pmatrix} 3 & 4 & 4 \\ -3 & -6 & -8 \\ 3 & 7 & 9 \end{pmatrix}.$$

Si determini inoltre una base di \mathbb{R}^3 rispetto a cui la matrice di ϕ sia la forma canonica trovata.

Esercizio 4.2. Si determini la forma canonica di Jordan dell'endomorfismo $\phi : \mathbb{Q}^4 \rightarrow \mathbb{Q}^4$ la cui matrice, rispetto alla base canonica, è

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

Si determini inoltre una base di \mathbb{Q}^4 rispetto a cui la matrice di ϕ sia la forma canonica trovata.

Esercizio 4.3. Si determini la forma canonica di Jordan dell'endomorfismo $\phi : \mathbb{R}^4 \rightarrow \mathbb{R}^4$ la cui matrice, rispetto alla base canonica, è

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

Si determini inoltre una base di \mathbb{R}^4 rispetto a cui la matrice di ϕ sia la forma canonica trovata.

Esercizio 4.4. Si determini la forma canonica di Jordan dell'endomorfismo $\phi : \mathbb{R}^4 \rightarrow \mathbb{R}^4$ la cui matrice, rispetto alla base canonica, è

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

Si determini inoltre una base di \mathbb{R}^4 rispetto a cui la matrice di ϕ sia la forma canonica trovata.

Esercizio 4.5. Sia A una matrice $n \times n$ a coefficienti reali tale che $A^2 = A$. Si dimostri che A è diagonalizzabile e che i suoi autovalori sono solo 0 oppure 1. Si dimostri inoltre che, se anche B è una matrice tale che $B^2 = B$, allora A e B sono simili se e solo se hanno lo stesso rango.

Esercizio 4.6. Si fornisca un esempio di due matrici quadrate A e B dello stesso ordine, aventi lo stesso polinomio caratteristico e lo stesso polinomio minimo, ma tali che A non sia simile a B .

Esercizio 4.7. Sia V uno spazio vettoriale complesso di dimensione 9. Si determinino tutti gli endomorfismi ϕ di V che soddisfano le seguenti condizioni:

$$\begin{aligned}\dim \text{Ker}(\phi - 2) &= 1, & \dim \text{Ker}(\phi - 2)^3 &= 3, \\ \dim \text{Ker}(\phi - 3) &= 2, & \dim \text{Ker}(\phi - 3)^2 &= 4, \\ \dim \text{Im}(\phi^2) &= 7.\end{aligned}$$

Esercizio 4.8. Sia V uno spazio vettoriale di dimensione 10 su \mathbb{Q} . Si determini il polinomio caratteristico, il polinomio minimo e la matrice di Jordan di tutti gli endomorfismi ϕ di V che soddisfano le seguenti condizioni:

$$\begin{aligned}\dim \text{Ker}(\phi - 5) &= 2, & \dim \text{Ker}(\phi - 5)^2 &= 3, & \dim \text{Ker}(\phi - 5)^3 &= 4, \\ \dim \text{Ker}(\phi + 2) &= 2, & \dim \text{Ker}(\phi + 2)^2 &= 4, \\ \dim \text{Im}(\phi) &= 8.\end{aligned}$$

Esercizio 4.9. Sia V uno spazio vettoriale complesso di dimensione 6. Si determinino tutti gli endomorfismi ϕ di V che hanno rango ≥ 4 ed il cui polinomio minimo è $x^4 - 6x^3 + 9x^2$.

Esercizio 4.10. Sia A una matrice quadrata di ordine n a coefficienti in \mathbb{C} . Si dimostri che A è nilpotente⁵ se e solo se $\text{tr}(A) = \text{tr}(A^2) = \dots = \text{tr}(A^n) = 0$.

⁵Una matrice A si dice *nilpotente* se $A^r = \mathbf{0}$, per qualche $r \geq 1$.

Capitolo 5

Spazi Vettoriali Euclidei

In questo capitolo introdurremo il concetto di *prodotto scalare* di due vettori (si tratta di un particolare prodotto di due vettori il cui risultato è uno scalare). Vedremo poi come questo prodotto permetta di definire la lunghezza di un vettore e l'angolo compreso tra due vettori. Queste nozioni ci permetteranno poi, a loro volta, di calcolare aree, volumi, ecc.

5.1 Lunghezze e angoli

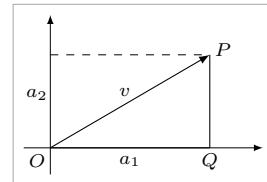
Nella geometria euclidea vengono definite, in modo del tutto naturale, le nozioni di lunghezza di un segmento (o distanza tra due punti) e di angolo tra due rette. Uno dei risultati più importanti riguardanti le lunghezze è il Teorema di Pitagora, che permette di ricavare la lunghezza dell'ipotenusa di un triangolo rettangolo quando sono note le lunghezze dei due cateti. Noi prenderemo spunto dalla validità del Teorema di Pitagora in \mathbb{R}^2 e \mathbb{R}^3 per definire la lunghezza di un vettore in \mathbb{R}^n .

5.1.1 Lunghezza di un vettore

Iniziamo considerando un vettore $v = (a_1, a_2) \in \mathbb{R}^2$ nel piano euclideo, che possiamo rappresentare graficamente come nella figura a lato.

Il vettore v è rappresentato dal segmento orientato \overrightarrow{OP} , mentre le lunghezze dei segmenti OQ e PQ sono rispettivamente $|a_1|$ e $|a_2|$, ossia le due componenti del vettore v prese in valore assoluto per evitare problemi nel caso in cui esse fossero negative. Poiché il triangolo $\triangle OPQ$ è un triangolo rettangolo, dal teorema di Pitagora segue che

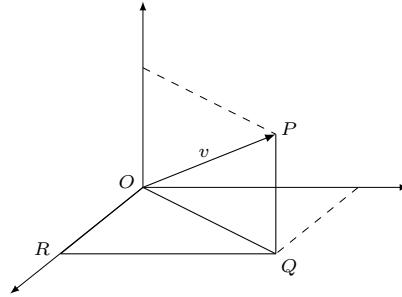
$$\|OP\|^2 = \|OQ\|^2 + \|PQ\|^2,$$



ove con il simbolo $\|AB\|$ indichiamo la lunghezza di un segmento AB . Ricordando che $\|OQ\| = |a_1|$ e $\|PQ\| = |a_2|$, si ottiene la seguente formula che esprime la lunghezza del segmento che rappresenta il vettore v :

$$\|OP\| = \sqrt{a_1^2 + a_2^2}.$$

Consideriamo ora un vettore $v = (a_1, a_2, a_3) \in \mathbb{R}^3$, rappresentato dal segmento orientato \overrightarrow{OP} come nella figura seguente:



Dato che $\triangle OPQ$ è un triangolo rettangolo, si ha

$$\|OP\|^2 = \|OQ\|^2 + \|PQ\|^2$$

e, osservando che anche il triangolo $\triangle OQR$ è rettangolo, si ha

$$\|OQ\|^2 = \|OR\|^2 + \|QR\|^2,$$

da cui segue che

$$\|OP\|^2 = \|OR\|^2 + \|QR\|^2 + \|PQ\|^2.$$

Notando che $\|OR\| = |a_1|$, $\|QR\| = |a_2|$ e $\|PQ\| = |a_3|$, si conclude che la lunghezza del (segmento che rappresenta il) vettore $v = (a_1, a_2, a_3)$ è data da

$$\|OP\| = \sqrt{a_1^2 + a_2^2 + a_3^2}.$$

Questi due risultati motivano la seguente definizione:

Definizione 5.1.1. Sia $v = (a_1, a_2, \dots, a_n) \in \mathbb{R}^n$. La *lunghezza* di v (detta anche la *norma* o il *modulo* di v) è data da

$$\|v\| = \sqrt{a_1^2 + a_2^2 + \dots + a_n^2}.$$

Osservazione 5.1.2. La lunghezza di un vettore v viene spesso chiamata il *modulo* di v e indicata con il simbolo $|v|$. Per evitare possibili confusioni con la nozione di modulo (cioè valore assoluto) di un numero, noi preferiamo usare il simbolo $\|v\|$, che chiameremo *norma* di v .

La *norma* definisce quindi una funzione

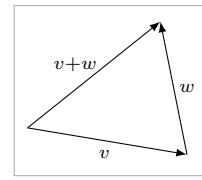
$$\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}, \quad v \mapsto \|v\|$$

che soddisfa le seguenti proprietà:

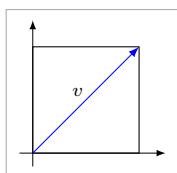
- (i) $\|v\| \geq 0$, per ogni $v \in \mathbb{R}^n$, e $\|v\| = 0$ se e solo se $v = \mathbf{0}$;
- (ii) $\|\lambda v\| = |\lambda| \|v\|$, per ogni $\lambda \in \mathbb{R}$ e ogni $v \in \mathbb{R}^n$;
- (iii) $\|v + w\| \leq \|v\| + \|w\|$, per ogni $v, w \in \mathbb{R}^n$.

La proprietà (i) è del tutto ovvia (una somma di quadrati di numeri reali è sempre ≥ 0 ed è nulla se e solo se tutti i numeri in questione sono zero), mentre la proprietà (ii) può essere verificata con un facile calcolo.

La proprietà (iii), detta *diseguaglianza triangolare*, ha un evidente significato geometrico: essa equivale al noto risultato di geometria euclidea che afferma che, in ogni triangolo, la lunghezza di un lato è minore della somma delle lunghezze degli altri due (vedi figura a lato). La dimostrazione algebrica di questa proprietà verrà data in seguito (vedi Proposizione 5.1.8).

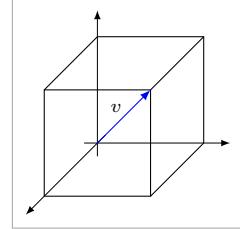


Esempio. A titolo di esempio, vedremo ora come si possa calcolare la lunghezza della diagonale di un (iper)cubo in \mathbb{R}^n , per ogni $n \geq 2$.



Nel caso del piano ($n = 2$), consideriamo un quadrato di lato unitario, che possiamo sempre supporre avente un vertice nell'origine e i lati paralleli agli assi coordinati. I vertici di tale quadrato sono dunque i punti di coordinate $(0, 0)$, $(0, 1)$, $(1, 0)$ e $(1, 1)$ e come diagonale possiamo considerare il segmento che ha come estremi i vertici $(0, 0)$ e $(1, 1)$. Tale segmento rappresenta il vettore $v = (1, 1)$ la cui lunghezza è data da $\|v\| = \sqrt{2}$.

Nel caso dello spazio tridimensionale ($n = 3$), consideriamo un cubo di lato unitario avente un vertice nell'origine e i lati paralleli agli assi coordinati. I vertici del cubo sono dunque i punti di coordinate $(0, 0, 0)$, $(0, 0, 1)$, $(0, 1, 0)$, $(0, 1, 1)$, $(1, 0, 0)$, $(1, 0, 1)$, $(1, 1, 0)$ e $(1, 1, 1)$, e la diagonale è il segmento che ha come estremi i vertici $(0, 0, 0)$ e $(1, 1, 1)$. Questo segmento rappresenta il vettore $v = (1, 1, 1)$ la cui lunghezza è $\|v\| = \sqrt{3}$.

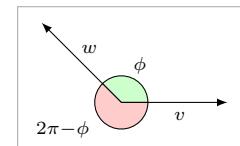


In modo del tutto analogo possiamo definire l'*iper cubo unitario* in \mathbb{R}^n : esso è la “figura” avente come vertici i 2^n punti di coordinate $(0, 0, \dots, 0)$, $(0, 0, \dots, 1)$, $(0, \dots, 1, 0)$, \dots , $(1, 1, \dots, 1)$. La diagonale è allora rappresentata dal vettore $v = (1, 1, \dots, 1)$, la cui lunghezza è $\|v\| = \sqrt{n}$. Si può osservare che, al crescere di n , la diagonale dell'iper cubo unitario aumenta tendendo a diventare infinitamente lunga, mentre la lunghezza del lato dell'iper cubo stesso rimane, naturalmente, sempre uguale a 1.

5.1.2 Angoli

Consideriamo ora il problema di determinare l'angolo compreso tra due vettori.

Due vettori non nulli v e w in \mathbb{R}^2 (oppure in \mathbb{R}^3) individuano due angoli la cui somma è un angolo giro. Con l'espressione “angolo compreso tra due vettori” noi intendiamo sempre l'angolo *convesso* (non orientato), cioè quello il cui valore è nell'intervallo $[0, \pi]$.



Consideriamo dunque lo spazio vettoriale $V = \mathbb{R}^2$, oppure $V = \mathbb{R}^3$. In fisica viene definito il *prodotto scalare* di due vettori v e $w \in V$ ponendo

$$v \cdot w = \|v\| \|w\| \cos \phi,$$

ove ϕ è l'angolo compreso tra i due vettori (a tal proposito si noti che $\cos \phi = \cos(2\pi - \phi)$ quindi, nella definizione di prodotto scalare, non è importante quale dei due angoli determinati da v e w si considera). Il prodotto scalare definisce quindi una funzione

$$\cdot : V \times V \rightarrow \mathbb{R}, \quad (v, w) \mapsto v \cdot w.$$

Vogliamo dimostrare che questa funzione è *bilineare* (cioè lineare rispetto a ciascuno dei suoi due argomenti) e *simmetrica*, ossia che soddisfa le seguenti proprietà:

- (i) $(\lambda v) \cdot w = \lambda(v \cdot w)$,
- (ii) $(v_1 + v_2) \cdot w = v_1 \cdot w + v_2 \cdot w$,
- (iii) $v \cdot (\mu w) = \mu(v \cdot w)$,
- (iv) $v \cdot (w_1 + w_2) = v \cdot w_1 + v \cdot w_2$,
- (v) $v \cdot w = w \cdot v$,

per ogni $\lambda, \mu \in \mathbb{R}$ e ogni $v, v_1, v_2, w, w_1, w_2 \in V$.

La proprietà di simmetria (v) è ovvia. Grazie a questa è quindi sufficiente dimostrare la linearità del prodotto scalare rispetto a uno solo dei suoi due argomenti, ad esempio rispetto al secondo (proprietà (iii) e (iv)).

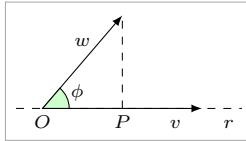
Per dimostrare la proprietà (iii) consideriamo separatamente i casi $\mu = 0$, $\mu > 0$ e $\mu < 0$. Se $\mu = 0$, si ha $v \cdot (0w) = v \cdot \mathbf{0} = 0 = 0(v \cdot w)$. Nel caso in cui $\mu > 0$ si ha $\|\mu w\| = |\mu| \|w\| = \mu \|w\|$ e il vettore μw è parallelo e ha lo stesso verso del vettore w . Di conseguenza l'angolo compreso tra i vettori v e μw coincide con l'angolo ϕ compreso tra v e w . Si ha pertanto:

$$v \cdot (\mu w) = \|v\| \|\mu w\| \cos \phi = \mu \|v\| \|w\| \cos \phi = \mu(v \cdot w).$$

Se invece $\mu < 0$ si ha $\|\mu w\| = |\mu| \|w\| = -\mu \|w\|$ e in questo caso il vettore μw è parallelo ma ha verso opposto al vettore w . Di conseguenza l'angolo compreso tra i vettori v e μw è $\pi - \phi$, se ϕ denota l'angolo compreso tra v e w . Si ha pertanto:

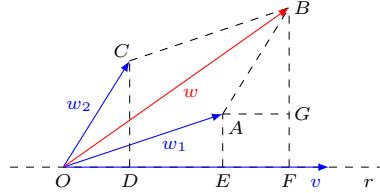
$$v \cdot (\mu w) = \|v\| \|\mu w\| \cos(\pi - \phi) = -\mu \|v\| \|w\| (-\cos \phi) = \mu(v \cdot w).$$

Dimostriamo ora l'additività del prodotto scalare rispetto al secondo argomento (proprietà (iv)). Premettiamo la seguente osservazione. Consideriamo due vettori $v, w \in V$ (con $v \neq \mathbf{0}$) e indichiamo con \overrightarrow{OP} la proiezione ortogonale del vettore w sulla retta r generata dal vettore v , come indicato nella figura a lato. Il prodotto $\|w\| \cos \phi$ non è altro che la lunghezza del segmento OP (considerata negativa se il vettore \overrightarrow{OP} ha verso opposto a quello del vettore v , cioè se $\frac{\pi}{2} < \phi \leq \pi$). Possiamo quindi affermare che il prodotto scalare di v per w è il prodotto della norma di v per la norma



della proiezione ortogonale di w sulla retta generata dal vettore v (sempre con la convenzione che tale norma va considerata con il segno negativo se $\frac{\pi}{2} < \phi \leq \pi$).

Siano w_1 e w_2 due vettori di V e poniamo $w = w_1 + w_2$. Dobbiamo dimostrare che $v \cdot w = v \cdot w_1 + v \cdot w_2$. Consideriamo dunque le proiezioni ortogonali dei vettori w_1 , w_2 e w sulla retta r generata dal vettore v , come illustrato nella figura seguente



ove, per semplicità, abbiamo considerato solo il caso in cui gli angoli compresi tra v e w_1 , tra v e w_2 e tra v e w sono tutti minori di un angolo retto. Indicando rispettivamente con \overrightarrow{OE} la proiezione ortogonale del vettore w_1 sulla retta r , con \overrightarrow{OD} quella di w_2 e con \overrightarrow{OF} quella del vettore w , si ha:

$$v \cdot w_1 = \|v\| \|OE\|, \quad v \cdot w_2 = \|v\| \|OD\|, \quad v \cdot w = \|v\| \|OF\|.$$

Dall'uguaglianza dei triangoli $\triangle OCD$ e $\triangle ABG$ si deduce l'uguaglianza $\|OD\| = \|AG\| = \|EF\|$, quindi si ha $\|OF\| = \|OE\| + \|EF\| = \|OE\| + \|OD\|$. Si ha pertanto:

$$v \cdot w = \|v\| \|OF\| = \|v\| \|OE\| + \|v\| \|OD\| = v \cdot w_1 + v \cdot w_2,$$

come volevasi dimostrare.

Possiamo ora dimostrare il seguente risultato:

Proposizione 5.1.3. *Siano $v = (a_1, a_2) \in \mathbb{R}^2$ (oppure, $v = (a_1, a_2, a_3) \in \mathbb{R}^3$), $w = (b_1, b_2) \in \mathbb{R}^2$ (oppure, $w = (b_1, b_2, b_3) \in \mathbb{R}^3$). Allora si ha:*

$$v \cdot w = a_1 b_1 + a_2 b_2 \quad (\text{risp., } v \cdot w = a_1 b_1 + a_2 b_2 + a_3 b_3).$$

Dimostrazione. Effettuiamo la dimostrazione per \mathbb{R}^3 (il caso in cui $v, w \in \mathbb{R}^2$ è del tutto analogo). Siano dunque $v = (a_1, a_2, a_3)$, $w = (b_1, b_2, b_3) \in \mathbb{R}^3$ e indichiamo con e_1, e_2, e_3 i tre vettori della base canonica di \mathbb{R}^3 . Possiamo quindi scrivere $v = a_1 e_1 + a_2 e_2 + a_3 e_3$ e $w = b_1 e_1 + b_2 e_2 + b_3 e_3$. Utilizzando la bilinearità del prodotto scalare si ottiene:

$$v \cdot w = (a_1 e_1 + a_2 e_2 + a_3 e_3) \cdot (b_1 e_1 + b_2 e_2 + b_3 e_3) = \sum_{i,j=1}^3 a_i b_j (e_i \cdot e_j).$$

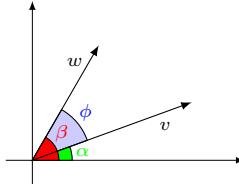
Notiamo che $e_i \cdot e_i = \|e_i\|^2 = 1$ per $i = 1, 2, 3$, mentre $e_i \cdot e_j = 0$ per ogni $i \neq j$, perché i vettori della base canonica sono a due a due ortogonali. Sostituendo nell'espressione precedente si ottiene quindi

$$v \cdot w = \sum_{i=1}^3 a_i b_i.$$

□

Osservazione 5.1.4. Nel caso di due vettori di \mathbb{R}^2 si può fornire una dimostrazione diretta del risultato precedente, senza ricorrere alla bilinearità del prodotto scalare (ma utilizzando qualche risultato di trigonometria).

Siano infatti $v = (a_1, a_2)$ e $w = (b_1, b_2)$ due vettori di \mathbb{R}^2 , come rappresentato nella figura seguente:



L'angolo ϕ compreso tra v e w è dato dalla differenza $\phi = \beta - \alpha$ tra l'angolo β che il vettore w forma con l'asse X e l'angolo α compreso tra l'asse X e il vettore v . Il prodotto scalare di v per w è pertanto:

$$\begin{aligned} v \cdot w &= \|v\| \|w\| \cos \phi \\ &= \|v\| \|w\| \cos(\beta - \alpha) \\ &= \|v\| \|w\| \cos \beta \cos \alpha + \|v\| \|w\| \sin \beta \sin \alpha \\ &= a_1 b_1 + a_2 b_2, \end{aligned}$$

dato che $a_1 = \|v\| \cos \alpha$, $a_2 = \|v\| \sin \alpha$, $b_1 = \|w\| \cos \beta$ e $b_2 = \|w\| \sin \beta$.

I risultati precedenti permettono di ricavare l'angolo ϕ compreso tra due vettori non nulli v e w di \mathbb{R}^2 o \mathbb{R}^3 . Si ha infatti:

$$\cos \phi = \frac{v \cdot w}{\|v\| \|w\|}. \quad (5.1.1)$$

Possiamo quindi concludere che, nel caso di vettori di \mathbb{R}^2 o \mathbb{R}^3 , il prodotto scalare permette di calcolare sia la norma di un vettore che l'angolo (non orientato) formato da due vettori non nulli. Nella prossima sezione vedremo come questi risultati si possano estendere a vettori di \mathbb{R}^n , per ogni $n \geq 2$.

5.1.3 Il prodotto scalare in \mathbb{R}^n

I risultati ottenuti nei paragrafi precedenti motivano la seguente definizione:

Definizione 5.1.5. Dati due vettori $v = (a_1, a_2, \dots, a_n)$, $w = (b_1, b_2, \dots, b_n) \in \mathbb{R}^n$, il loro *prodotto scalare* è definito ponendo

$$v \cdot w = a_1 b_1 + a_2 b_2 + \dots + a_n b_n = \sum_{i=1}^n a_i b_i.$$

In base a questa definizione, per ogni $v \in \mathbb{R}^n$, si ha

$$v \cdot v = \sum_{i=1}^n a_i^2 = \|v\|^2,$$

quindi

$$\|v\| = \sqrt{v \cdot v}.$$

Poiché sappiamo che in \mathbb{R}^2 e in \mathbb{R}^3 vale la formula (5.1.1), vorremmo usare una formula analoga per definire l'angolo ϕ compreso tra due vettori non nulli di

\mathbb{R}^n , per ogni $n \geq 2$. Tuttavia, dato che il coseno di un angolo assume solo valori compresi nell'intervallo $[-1, 1]$, dobbiamo prima verificare che, per ogni coppia di vettori non nulli $v, w \in \mathbb{R}^n$, si abbia

$$-1 \leq \frac{v \cdot w}{\|v\| \|w\|} \leq 1.$$

Proposizione 5.1.6 (DISUGUAGLIANZA DI CAUCHY–SCHWARZ). *Per ogni coppia di vettori $v, w \in \mathbb{R}^n$, si ha*

$$|v \cdot w| \leq \|v\| \|w\|. \quad (5.1.2)$$

Inoltre vale il segno di uguaglianza se e solo se i due vettori sono linearmente dipendenti.

Dimostrazione. Iniziamo col dimostrare la prima affermazione. Osserviamo che se uno dei due vettori è nullo, il risultato è banalmente verificato. Supponiamo quindi che v e w siano vettori non nulli. Ponendo $u_\alpha = v + \alpha w$, con $\alpha \in \mathbb{R}$, si ha:

$$\|u_\alpha\|^2 = u_\alpha \cdot u_\alpha = (v + \alpha w) \cdot (v + \alpha w) = \|v\|^2 + 2\alpha(v \cdot w) + \alpha^2\|w\|^2 \geq 0,$$

per ogni $\alpha \in \mathbb{R}$. Il trinomio di secondo grado in α

$$\alpha^2\|w\|^2 + 2\alpha(v \cdot w) + \|v\|^2$$

assume dunque sempre valori ≥ 0 , pertanto il suo discriminante deve essere ≤ 0 . Si ha quindi

$$\Delta = 4(v \cdot w)^2 - 4\|v\|^2\|w\|^2 \leq 0,$$

da cui si deduce che $|v \cdot w| \leq \|v\| \|w\|$.

Dimostriamo ora la seconda affermazione. Se i due vettori sono linearmente dipendenti uno dei due deve essere multiplo dell'altro. Possiamo supporre, ad esempio, che sia $w = \lambda v$, per qualche $\lambda \in \mathbb{R}$. In tal caso $\|w\| = |\lambda| \|v\|$ e si ottiene

$$|v \cdot w| = |v \cdot \lambda v| = |\lambda| \|v\|^2 = \|v\| \|w\|.$$

Viceversa, supponiamo che nell'espressione (5.1.2) valga il segno di uguaglianza. Si ha quindi

$$v \cdot w = \pm \|v\| \|w\|.$$

Se $w = 0$ i due vettori sono (banalmente) linearmente dipendenti; in caso contrario poniamo

$$\bar{\alpha} = \begin{cases} -\frac{\|v\|}{\|w\|} & \text{se } v \cdot w = \|v\| \|w\| \\ \frac{\|v\|}{\|w\|} & \text{se } v \cdot w = -\|v\| \|w\| \end{cases}$$

e consideriamo il vettore $u = v + \bar{\alpha}w$. Si ha:

$$\begin{aligned} \|u\|^2 &= (v + \bar{\alpha}w) \cdot (v + \bar{\alpha}w) \\ &= \|v\|^2 + 2\bar{\alpha}(v \cdot w) + \bar{\alpha}^2\|w\|^2 \\ &= \|v\|^2 \pm 2\bar{\alpha}\|v\|\|w\| + \bar{\alpha}^2\|w\|^2 \\ &= (\|v\| \pm \bar{\alpha}\|w\|)^2 \\ &= \left(\|v\| - \frac{\|v\|}{\|w\|}\|w\|\right)^2 = 0. \end{aligned}$$

Poiché l'unico vettore che ha norma nulla è il vettore nullo, si conclude che $u = v + \bar{\alpha}w = \mathbf{0}$, il che dimostra che v e w sono linearmente dipendenti. \square

Dalla diseguaglianza di Cauchy–Schwarz segue che, per ogni coppia di vettori non nulli $v, w \in \mathbb{R}^n$, è

$$-1 \leq \frac{v \cdot w}{\|v\| \|w\|} \leq 1.$$

Inoltre, dalla dimostrazione della Proposizione 5.1.6, si deduce facilmente che è $\frac{v \cdot w}{\|v\| \|w\|} = 1$ se e solo se i vettori v e w sono paralleli e hanno lo stesso verso, mentre $\frac{v \cdot w}{\|v\| \|w\|} = -1$ se e solo se v e w sono paralleli ma hanno versi opposti. Osservando che, per ogni numero reale $t \in [-1, 1]$ esiste un unico angolo $\phi \in [0, \pi]$ tale che $t = \cos \phi$, possiamo dare la seguente definizione:

Definizione 5.1.7. Dati due vettori non nulli $v, w \in \mathbb{R}^n$, l'*angolo* (non orientato) tra essi compreso è l'unico $\phi \in [0, \pi]$ tale che

$$\cos \phi = \frac{v \cdot w}{\|v\| \|w\|}. \quad (5.1.3)$$

Si noti che, in base alle osservazioni precedenti, l'angolo compreso tra due vettori è nullo se e solo se i due vettori sono paralleli e hanno lo stesso verso, mentre esso è pari a π se e solo se i due vettori sono paralleli ma hanno versi opposti.

Un'altra conseguenza della diseguaglianza di Cauchy–Schwarz è la cosiddetta *diseguaglianza triangolare*, che ora dimostreremo.

Proposizione 5.1.8 (DISEGUAGLIAZIONE TRIANGOLARE). *Per ogni $v, w \in \mathbb{R}^n$, si ha*

$$\|v + w\| \leq \|v\| + \|w\|.$$

Dimostrazione. Si ha:

$$\begin{aligned} \|v + w\|^2 &= (v + w) \cdot (v + w) \\ &= \|v\|^2 + 2(v \cdot w) + \|w\|^2 \\ &\leq \|v\|^2 + 2|v \cdot w| + \|w\|^2 \\ &\leq \|v\|^2 + 2\|v\| \|w\| + \|w\|^2 \\ &= (\|v\| + \|w\|)^2. \end{aligned}$$

Estraendo la radice quadrata si ottiene $\|v + w\| \leq \|v\| + \|w\|$, come volevansi dimostrare. \square

Osservazione 5.1.9. Dalla definizione di angolo tra due vettori segue la seguente condizione di perpendicolarità: due vettori non nulli sono ortogonali se e solo se il loro prodotto scalare è zero. Infatti, si ha $v \cdot w = 0$ se e solo se $\|v\| = 0$ (e quindi $v = \mathbf{0}$) oppure $\|w\| = 0$ (cioè $w = \mathbf{0}$) oppure ancora $\cos \phi = 0$ (cioè $\phi = \frac{\pi}{2}$).

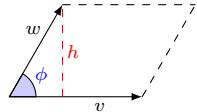
5.2 Aree e volumi

In questa sezione vedremo come, utilizzando il prodotto scalare, sia possibile calcolare aree e volumi in \mathbb{R}^n .

Consideriamo due vettori $v, w \in \mathbb{R}^n$. Poiché due vettori qualunque sono sempre contenuti in un piano (cioè in un sottospazio di dimensione 2 di \mathbb{R}^n) possiamo concentrare la nostra attenzione su tale piano. Indichiamo con

$$\mathcal{P}(v, w) = \{\lambda v + \mu w \mid 0 \leq \lambda, \mu \leq 1\}$$

il parallelogramma determinato dai due vettori v e w , come nella figura seguente:



L'area di tale parallelogramma è data dal prodotto della lunghezza del vettore v (la base) per l'altezza h relativa a tale base. Indicando con ϕ l'angolo compreso tra i vettori v e w , si ha $h = \|w\| \sin \phi$. Otteniamo quindi:

$$\text{Area } \mathcal{P}(v, w) = \|v\| \|w\| \sin \phi.$$

Ricordando che $\sin^2 \phi = 1 - \cos^2 \phi$ e utilizzando la formula (5.1.3), si trova

$$\sin^2 \phi = 1 - \frac{(v \cdot w)^2}{\|v\|^2 \|w\|^2} = \frac{\|v\|^2 \|w\|^2 - (v \cdot w)^2}{\|v\|^2 \|w\|^2}.$$

Si ha dunque

$$(\text{Area } \mathcal{P}(v, w))^2 = \|v\|^2 \|w\|^2 - (v \cdot w)^2 = (v \cdot v)(w \cdot w) - (v \cdot w)^2.$$

Questa formula può essere riscritta come segue:

$$\text{Area } \mathcal{P}(v, w) = \sqrt{\det \begin{pmatrix} v \cdot v & v \cdot w \\ w \cdot v & w \cdot w \end{pmatrix}}.$$

Vedremo ora che se $n = 2$, cioè nel caso di due vettori $v, w \in \mathbb{R}^2$, è possibile calcolare l'area del parallelogramma $\mathcal{P}(v, w)$ senza ricorrere al prodotto scalare.

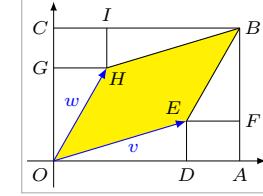
Siano dunque $v = (a_1, a_2)$, $w = (b_1, b_2) \in \mathbb{R}^2$ e consideriamo la situazione descritta nella figura a lato (ove, per semplicità, abbiamo supposto $a_1, a_2, b_1, b_2 \geq 0$). L'area del parallelogramma $\mathcal{P}(v, w)$ può essere ottenuta, per differenza di aree, dall'area del rettangolo $OABC$ sottraendo le aree dei due rettangoli $ADEF$ e $CGHI$ e dei quattro triangoli ODE , BEF , BHI e OGH . Osservando che $\|OD\| = \|BI\| = a_1$, $\|DE\| = \|AF\| = \|CG\| = \|HI\| = a_2$, $\|CI\| = \|GH\| = \|DA\| = \|EF\| = b_1$, $\|OG\| = \|BF\| = b_2$, si ottiene:

$$\text{Area}(OABC) = (a_1 + b_1)(a_2 + b_2),$$

$$\text{Area}(ODE) = \text{Area}(BHI) = \frac{1}{2} a_1 a_2,$$

$$\text{Area}(OGH) = \text{Area}(BEF) = \frac{1}{2} b_1 b_2,$$

$$\text{Area}(ADEF) = \text{Area}(CGHI) = b_1 a_2,$$



da cui si ricava:

$$\begin{aligned} \text{Area } \mathcal{P}(v, w) &= (a_1 + b_1)(a_2 + b_2) - a_1 a_2 - b_1 b_2 - 2b_1 a_2 \\ &= a_1 b_2 - b_1 a_2 = \det \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}. \end{aligned}$$

Il determinante della matrice quadrata costituita dalle componenti dei due vettori v e w (scritti indifferentemente in colonna oppure in riga) fornisce quindi il valore dell'area del parallelogramma $\mathcal{P}(v, w)$. Tuttavia, ricordando che

$$\det \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} = - \det \begin{pmatrix} b_1 & a_1 \\ b_2 & a_2 \end{pmatrix}$$

conviene prendere tale determinante in valore assoluto, per evitare il rischio di trovare valori negativi per l'area di un parallelogramma. Sarà dunque più corretto scrivere la seguente formula:

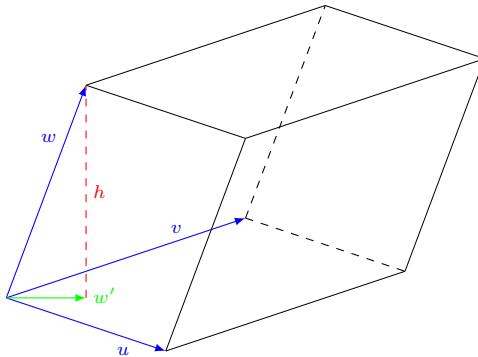
$$\text{Area } \mathcal{P}(v, w) = \left| \det \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \right|, \quad (5.2.1)$$

per ogni $v = (a_1, a_2)$, $w = (b_1, b_2) \in \mathbb{R}^2$.

Passiamo ora al caso dei volumi. A tal fine consideriamo tre vettori $u, v, w \in \mathbb{R}^n$ e indichiamo con

$$\mathcal{P}(u, v, w) = \{ \lambda u + \mu v + \nu w \mid 0 \leq \lambda, \mu, \nu \leq 1 \}$$

il parallelepipedo da essi determinato. Poiché tre vettori qualunque di \mathbb{R}^n sono sempre contenuti in un sottospazio di dimensione 3, possiamo concentrare la nostra attenzione su tale sottospazio tridimensionale. Consideriamo dunque la situazione schematizzata nella figura seguente:



Il volume di tale parallelepipedo è dato dal prodotto dell'area del parallelogramma $\mathcal{P}(u, v)$ (l'area di base) per l'altezza h relativa a tale base. La determinazione dell'altezza h può essere effettuata come segue: consideriamo un generico vettore $\alpha u + \beta v$ appartenente al piano generato da u e v . Dobbiamo determinare i valori di α e β che rendono minima¹ la norma del vettore $w - \alpha u - \beta v$. Il valore minimo della norma di tale vettore è precisamente l'altezza h che stiamo

¹In alternativa si potrebbe richiedere che il vettore $w - \alpha u - \beta v$ sia ortogonale al piano generato da u e v (l'altezza di un parallelepipedo deve essere perpendicolare alla sua base). Questo equivale a richiedere che

$$\begin{cases} (w - \alpha u - \beta v) \cdot u = 0 \\ (w - \alpha u - \beta v) \cdot v = 0. \end{cases}$$

Si trova così un sistema equivalente a (5.2.2).

cercando. Sviluppando i calcoli, si trova:

$$\begin{aligned}\|w - \alpha u - \beta v\|^2 &= (w - \alpha u - \beta v) \cdot (w - \alpha u - \beta v) \\ &= w \cdot w - 2\alpha(u \cdot w) - 2\beta(v \cdot w) + 2\alpha\beta(u \cdot v) \\ &\quad + \alpha^2(u \cdot u) + \beta^2(v \cdot v).\end{aligned}$$

Per trovare il valore minimo di tale espressione basta calcolarne le due derivate parziali, rispetto a α e β , e imporre che queste siano nulle. Si ottiene così il seguente sistema:

$$\begin{cases} 2\alpha(u \cdot u) + 2\beta(u \cdot v) - 2(u \cdot w) = 0 \\ 2\alpha(u \cdot v) + 2\beta(v \cdot v) - 2(v \cdot w) = 0 \end{cases} \quad (5.2.2)$$

la cui soluzione è data da

$$\begin{cases} \bar{\alpha} = \frac{(u \cdot w)(v \cdot v) - (u \cdot v)(v \cdot w)}{(u \cdot u)(v \cdot v) - (u \cdot v)^2} \\ \bar{\beta} = \frac{(u \cdot u)(v \cdot w) - (u \cdot v)(u \cdot w)}{(u \cdot u)(v \cdot v) - (u \cdot v)^2}. \end{cases}$$

Ponendo $w' = \bar{\alpha}u + \bar{\beta}v$, il vettore $w - w'$ rappresenta l'altezza del parallelepipedo $\mathcal{P}(u, v, w)$, relativa alla base $\mathcal{P}(u, v)$. Ricordando che l'area del parallelogramma $\mathcal{P}(u, v)$ è data da

$$\text{Area } \mathcal{P}(u, v) = \sqrt{\det \begin{pmatrix} u \cdot u & u \cdot v \\ v \cdot u & v \cdot v \end{pmatrix}},$$

siamo ora in grado di calcolare il volume del parallelepipedo $\mathcal{P}(u, v, w)$ moltiplicando tale espressione per $h = \|w - w'\|$. Sviluppando i calcoli si trova la seguente formula per il volume:

$$\text{Vol } \mathcal{P}(u, v, w) = \sqrt{\det \begin{pmatrix} u \cdot u & u \cdot v & u \cdot w \\ v \cdot u & v \cdot v & v \cdot w \\ w \cdot u & w \cdot v & w \cdot w \end{pmatrix}}.$$

Se $n = 3$, cioè nel caso di tre vettori $u, v, w \in \mathbb{R}^3$, è possibile calcolare il volume del parallelepipedo $\mathcal{P}(u, v, w)$ anche come il valore assoluto del determinante della matrice le cui colonne (o righe) sono costituite dalle componenti dei vettori dati. Si può cioè dimostrare che, dati $u = (a_1, a_2, a_3)$, $v = (b_1, b_2, b_3)$, $w = (c_1, c_2, c_3) \in \mathbb{R}^3$, si ha

$$\text{Vol } \mathcal{P}(u, v, w) = \left| \det \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} \right|. \quad (5.2.3)$$

I risultati che abbiamo ottenuto per le aree dei parallelogrammi e i volumi dei parallelepipedi motivano la seguente definizione:

Definizione 5.2.1. Dati r vettori $v_1, v_2, \dots, v_r \in \mathbb{R}^n$ (con $r \leq n$), il *parallelotopo* da essi generato è il sottoinsieme di \mathbb{R}^n dato da

$$\mathcal{P}(v_1, v_2, \dots, v_r) = \left\{ \sum_{i=1}^r \lambda_i v_i \mid 0 \leq \lambda_i \leq 1, i = 1, \dots, r \right\}.$$

La sua *misura* (o volume) r -dimensionale è definita ponendo

$$\text{Mis } \mathcal{P}(v_1, \dots, v_r) = \sqrt{\det(v_i \cdot v_j)_{i,j=1,\dots,r}}.$$

Si noti che, nei casi $r = 2$ o $r = 3$, questa definizione si riduce rispettivamente alla definizione di area di un parallelogramma o volume di un parallelepipedo, mentre, per $r = 1$, si ritrova la definizione della norma di un vettore.

Osservazione 5.2.2. Se gli r vettori v_1, v_2, \dots, v_r sono linearmente dipendenti è facile dimostrare che la misura r -dimensionale di $\mathcal{P}(v_1, v_2, \dots, v_r)$ è nulla. Infatti in tal caso uno degli r vettori si può esprimere come combinazione lineare degli altri. A meno di un riordinamento, possiamo quindi supporre che sia

$$v_r = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_{r-1} v_{r-1}.$$

È ora immediato verificare che, nella matrice $(v_i \cdot v_j)_{i,j=1,\dots,r}$, l'ultima colonna è una combinazione lineare delle colonne precedenti, quindi il determinante di tale matrice è nullo. Questa è la ragione per cui, nella Definizione 5.2.1, abbiamo supposto $r \leq n$; infatti nel caso $r > n$, gli r vettori v_1, \dots, v_r sarebbero sempre linearmente dipendenti.

Scrivendo in colonna le componenti degli r vettori $v_1 = {}^t(a_{11}, a_{21}, \dots, a_{n1})$, $v_2 = {}^t(a_{12}, a_{22}, \dots, a_{n2})$, ..., $v_r = {}^t(a_{1r}, a_{2r}, \dots, a_{nr})$, otteniamo la seguente matrice con n righe e r colonne (ove abbiamo supposto $r \leq n$)

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1r} \\ a_{21} & a_{22} & \cdots & a_{2r} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nr} \end{pmatrix}.$$

La matrice dei prodotti scalari $v_i \cdot v_j$ può allora essere scritta come segue:

$$(v_i \cdot v_j)_{i,j=1,\dots,r} = {}^t A A.$$

Nel caso particolare in cui $r = n$, la matrice A è quadrata e dal Teorema di Binet segue che

$$\det(v_i \cdot v_j)_{i,j=1,\dots,n} = (\det A)^2,$$

da cui si ottiene

$$\text{Mis } \mathcal{P}(v_1, \dots, v_n) = |\det A|, \tag{5.2.4}$$

ove A è la matrice le cui colonne sono costituite dalle componenti degli n vettori $v_1, \dots, v_n \in \mathbb{R}^n$. Per $n = 2$ e $n = 3$ si ritrovano così le due formule (5.2.1) e (5.2.3) citate in precedenza. Notiamo che, da quest'ultima formula, risulta evidente che $\text{Mis } \mathcal{P}(v_1, \dots, v_n) = 0$ se e solo se $\det A = 0$, cioè se e solo se i vettori v_1, v_2, \dots, v_n sono linearmente dipendenti.

5.3 Forme bilineari

Sia V uno spazio vettoriale definito sul campo K .

Definizione 5.3.1. Una *forma bilineare* su V è una funzione

$$g : V \times V \rightarrow K, \quad (v, w) \mapsto g(v, w),$$

lineare rispetto a ciascuno dei suoi due argomenti, cioè tale che

- (i) $g(\lambda_1 v_1 + \lambda_2 v_2, w) = \lambda_1 g(v_1, w) + \lambda_2 g(v_2, w),$
- (ii) $g(v, \mu_1 w_1 + \mu_2 w_2) = \mu_1 g(v, w_1) + \mu_2 g(v, w_2),$

per ogni $\lambda_1, \lambda_2, \mu_1, \mu_2 \in K$ e ogni $v, v_1, v_2, w, w_1, w_2 \in V$.

Definizione 5.3.2. Una forma bilineare $g : V \times V \rightarrow K$ è detta *simmetrica* se $g(v, w) = g(w, v)$, per ogni $v, w \in V$. Se la caratteristica del campo K è diversa da 2, g è detta *antisimmetrica* (o *alternante*) se, per ogni $v, w \in V$, si ha $g(v, w) = -g(w, v)$.

Osservazione 5.3.3. Notiamo che, se K è un campo di caratteristica diversa da 2, ogni forma bilineare g si può decomporre come segue

$$g(v, w) = g_s(v, w) + g_a(v, w),$$

ove g_s e g_a sono le due forme bilineari definite ponendo

$$g_s(v, w) = \frac{g(v, w) + g(w, v)}{2}, \quad g_a(v, w) = \frac{g(v, w) - g(w, v)}{2}.$$

Poiché g_s è una forma bilineare simmetrica mentre g_a è una forma bilineare alternante, ciò significa che ogni forma bilineare su V può essere espressa come somma di una forma bilineare simmetrica e di una forma bilineare alternante. Lo studio delle forme bilineari è quindi riconducibile allo studio delle forme bilineari simmetriche e di quelle alternanti. Nel seguito ci occuperemo esclusivamente dello studio delle forme bilineari simmetriche.

Osservazione 5.3.4. Se g_1 e g_2 sono due forme bilineari definite su uno spazio vettoriale V , è immediato verificare che la funzione $g_1 + g_2 : V \times V \rightarrow K$ definita ponendo $(g_1 + g_2)(v, w) = g_1(v, w) + g_2(v, w)$, per ogni $v, w \in V$, è una forma bilineare su V . Analogamente, per ogni $\lambda \in K$ e ogni forma bilineare g , anche la funzione λg è una forma bilineare su V . L'insieme $\text{Bil}(V)$ delle forme bilineari su V è quindi dotato, in modo naturale, di una struttura di spazio vettoriale su K . Poiché una combinazione lineare di forme bilineari simmetriche è ancora una forma bilineare simmetrica, l'insieme delle forme bilineari simmetriche è un sottospazio vettoriale di $\text{Bil}(V)$. Un risultato analogo vale naturalmente anche per le forme bilineari alternanti.

Sia V uno spazio vettoriale e siano U, W due sottospazi vettoriali di V tali che $V = U \oplus W$. Siano inoltre

$$g_U : U \times U \rightarrow K, \quad g_W : W \times W \rightarrow K,$$

due forme bilineari definite, rispettivamente, su U e su W . Ricordando che ogni vettore $v \in V$ si può scrivere in modo unico come $v = u + w$, con $u \in U$ e $w \in W$, definiamo una funzione

$$g_V : V \times V \rightarrow K$$

ponendo

$$g_V(u_1 + w_1, u_2 + w_2) = g_U(u_1, u_2) + g_W(w_1, w_2),$$

per ogni $u_1, u_2 \in U$, $w_1, w_2 \in W$. Si verifica facilmente che g_V è una forma bilineare su V e che essa è simmetrica se e solo se lo sono sia g_U che g_W .

Tale forma bilineare verrà indicata con $g_V = g_U \oplus g_W$ e detta la *somma diretta*² di g_U e g_W .

Consideriamo ora una forma bilineare simmetrica g definita su uno spazio vettoriale V . Diamo la seguente definizione:

Definizione 5.3.5. Il *nucleo* di g è il seguente sottoinsieme di V :

$$\text{Ker}(g) = \{v \in V \mid g(v, w) = 0, \text{ per ogni } w \in V\}.$$

Osservazione 5.3.6. Si dimostra facilmente, usando la bilinearità di g , che $\text{Ker}(g)$ è un sottospazio vettoriale di V . Inoltre, dalla simmetria di g , segue che è anche

$$\text{Ker}(g) = \{w \in V \mid g(v, w) = 0, \text{ per ogni } v \in V\}.$$

Definizione 5.3.7. Una forma bilineare simmetrica $g : V \times V \rightarrow K$ è detta *non degenera* se $\text{Ker}(g) = \{\mathbf{0}\}$. In caso contrario essa è detta *degenera*.

Per ogni spazio vettoriale V indicheremo con g_0 la *forma bilineare nulla* su V , cioè la forma bilineare definita ponendo $g_0(v, w) = 0$, per ogni $v, w \in V$. Si noti che, per ogni forma bilineare simmetrica g su V , la restrizione di g a $\text{Ker}(g) \times \text{Ker}(g)$ è la forma bilineare nulla sul sottospazio $\text{Ker}(g)$ di V .

Siamo ora in grado di enunciare e dimostrare il seguente risultato:

Proposizione 5.3.8. *Sia g una forma bilineare simmetrica definita su uno spazio vettoriale V . Esiste un sottospazio vettoriale U di V tale che:*

- (i) $V = U \oplus \text{Ker}(g)$;
- (ii) la restrizione di g a $U \times U$, che indicheremo con g_U , è una forma bilineare simmetrica non degenera su U ;
- (iii) $g = g_U \oplus g_0$, ove g_0 è la forma bilineare nulla su $\text{Ker}(g)$.

Dimostrazione. Dati V e g , esiste certamente un sottospazio vettoriale U di V tale che $V = U \oplus \text{Ker}(g)$. Indicando con g_U la restrizione di g a $U \times U$ e con g_0 la forma bilineare nulla definita su $\text{Ker}(g)$, consideriamo la forma bilineare $g_U \oplus g_0$ definita su $U \oplus \text{Ker}(g) = V$. Dobbiamo dimostrare che $g = g_U \oplus g_0$.

Consideriamo dunque due vettori $v_1, v_2 \in V$ e scriviamo $v_1 = u_1 + w_1$ e $v_2 = u_2 + w_2$, ove $u_1, u_2 \in U$ e $w_1, w_2 \in \text{Ker}(g)$. Si ha:

$$\begin{aligned} (g_U \oplus g_0)(v_1, v_2) &= (g_U \oplus g_0)(u_1 + w_1, u_2 + w_2) \\ &= g_U(u_1, u_2) + g_0(w_1, w_2) \\ &= g_U(u_1, u_2) = g(u_1, u_2), \end{aligned}$$

e

$$\begin{aligned} g(v_1, v_2) &= g(u_1 + w_1, u_2 + w_2) \\ &= g(u_1, u_2) + g(u_1, w_2) + g(w_1, u_2) + g(w_1, w_2). \end{aligned}$$

²La forma bilineare $g_V = g_U \oplus g_W$ ha la seguente proprietà: $g_V(v_1, v_2) = 0$ se $v_1 \in U$ e $v_2 \in W$, oppure se $v_1 \in W$ e $v_2 \in U$. Per tale motivo è anche detta la *somma ortogonale* di g_U e g_W ed è spesso indicata con la notazione $g_V = g_U \boxplus g_W$.

Dato che $w_1, w_2 \in \text{Ker}(g)$, si ha $g(u_1, w_2) = g(w_1, u_2) = g(w_1, w_2) = 0$ e pertanto $g(v_1, v_2) = (g_U \oplus g_0)(v_1, v_2)$.

Rimane solo da dimostrare che la forma bilineare $g_U : U \times U \rightarrow K$ è non degenere. Consideriamo un vettore $u_0 \in \text{Ker}(g_U)$ e un generico vettore $v \in V$. Scrivendo v nella forma $v = u + w$, con $u \in U$ e $w \in \text{Ker}(g)$, si ha

$$\begin{aligned} g(u_0, v) &= g(u_0, u + w) \\ &= g(u_0, u) + g(u_0, w) \\ &= g_U(u_0, u) + 0 = 0 \end{aligned}$$

perché $w \in \text{Ker}(g)$ e $u_0 \in \text{Ker}(g_U)$. Da ciò segue che $u_0 \in \text{Ker}(g)$, quindi deve essere $u_0 = \mathbf{0}$, dato che $U \cap \text{Ker}(g) = \{\mathbf{0}\}$. Abbiamo così dimostrato che $\text{Ker}(g_U) = \{\mathbf{0}\}$, quindi g_U è non degenere. \square

Osservazione 5.3.9. La proposizione precedente afferma che ogni forma bilineare simmetrica è somma diretta di una forma bilineare simmetrica non degenere e di una forma bilineare nulla. Possiamo quindi limitarci a studiare le forme bilineari simmetriche non degeneri.

Ricordando che il prodotto scalare usuale in \mathbb{R}^n è una forma bilineare simmetrica non degenere, e ricordando inoltre la condizione di ortogonalità tra due vettori di \mathbb{R}^n , possiamo dare la seguente definizione:

Definizione 5.3.10. Sia V uno spazio vettoriale dotato di una forma bilineare simmetrica non degenere g . Due vettori $v, w \in V$ si dicono *ortogonali* se $g(v, w) = 0$. Due sottospazi vettoriali $U_1, U_2 \subseteq V$ si dicono *ortogonali* se $g(u_1, u_2) = 0$, per ogni $u_1 \in U_1$ e ogni $u_2 \in U_2$.

Esempio 5.3.11. Nello spazio vettoriale $V = \mathbb{R}^2$ consideriamo la forma bilineare simmetrica g definita ponendo

$$g((x_1, x_2), (y_1, y_2)) = x_1 y_2 + x_2 y_1.$$

È immediato verificare che $\text{Ker}(g) = \{(0, 0)\}$, quindi g è non degenere. Si noti che $g((1, 1), (1, -1)) = 0$, quindi i vettori $v_1 = (1, 1)$ e $v_2 = (1, -1)$ sono ortogonali. Tuttavia, si ha anche $g((1, 0), (1, 0)) = 0$, pertanto il vettore $e_1 = (1, 0)$ risulta essere ortogonale a sé stesso!

Definizione 5.3.12. Sia V uno spazio vettoriale dotato di una forma bilineare simmetrica g . Un vettore $v \in V$ tale che $g(v, v) = 0$ è detto *isotropo*. Un sottospazio $U \subseteq V$ è detto *isotropo* se la restrizione di g a $U \times U$ è la forma bilineare nulla, cioè se $g(u_1, u_2) = 0$, per ogni $u_1, u_2 \in U$.

Osservazione 5.3.13. Si noti che se $U \subseteq V$ è un sottospazio isotropo, allora ogni vettore $u \in U$ è un vettore isotropo. Il viceversa, in generale, è falso. Se $u \in V$ è un vettore isotropo, certamente il sottospazio $\langle u \rangle$ da esso generato è un sottospazio isotropo, tuttavia se $u_1, u_2 \in V$ sono due vettori isotropi, non è detto che il sottospazio $U = \langle u_1, u_2 \rangle$ da essi generato sia un sottospazio isotropo. Infatti è ovviamente $g(u_1, u_1) = 0$ e $g(u_2, u_2) = 0$, ma non è detto che sia anche $g(u_1, u_2) = 0$. La forma bilineare considerata nell'Esempio 5.3.11 illustra precisamente questo fatto: i due vettori della base canonica $e_1 = (1, 0)$ ed $e_2 = (0, 1)$ sono due vettori isotropi, ma si ha $g(e_1, e_2) = 1$.

Definizione 5.3.14. Sia V uno spazio vettoriale dotato di una forma bilineare simmetrica g . Dato un sottoinsieme S di V definiamo il suo *ortogonale* ponendo

$$S^\perp = \{v \in V \mid g(v, w) = 0, \forall w \in S\}.$$

Proposizione 5.3.15. Siano V uno spazio vettoriale dotato di una forma bilineare simmetrica g e S un sottoinsieme di V . Allora:

- (i) S^\perp è un sottospazio vettoriale di V ;
- (ii) Si ha $S \subseteq \langle S \rangle \subseteq (S^\perp)^\perp$, ove $\langle S \rangle$ denota il sottospazio vettoriale di V generato da S .

Dimostrazione. Per dimostrare la prima affermazione consideriamo due vettori $v_1, v_2 \in S^\perp$ e due scalari $\lambda_1, \lambda_2 \in K$. Per ogni $w \in S$, si ha:

$$g(\lambda_1 v_1 + \lambda_2 v_2, w) = \lambda_1 g(v_1, w) + \lambda_2 g(v_2, w) = \lambda_1 0 + \lambda_2 0 = 0,$$

quindi $\lambda_1 v_1 + \lambda_2 v_2 \in S^\perp$.

Per quanto riguarda la seconda affermazione, l'inclusione $S \subseteq \langle S \rangle$ è ovvia. Per dimostrare che $\langle S \rangle \subseteq (S^\perp)^\perp$ consideriamo un generico vettore $w \in \langle S \rangle$; esso si potrà scrivere nella forma

$$w = \lambda_1 w_1 + \lambda_2 w_2 + \cdots + \lambda_r w_r,$$

per qualche $\lambda_1, \dots, \lambda_r \in K$ e qualche $w_1, \dots, w_r \in S$. Per ogni $v \in S^\perp$, si ha

$$g(w, v) = g\left(\sum_{i=1}^r \lambda_i w_i, v\right) = \sum_{i=1}^r \lambda_i g(w_i, v) = 0,$$

dato che $w_i \in S$ e $v \in S^\perp$. Ciò significa che w è ortogonale a tutti i vettori di S^\perp , quindi $w \in (S^\perp)^\perp$. \square

Proposizione 5.3.16. Siano V uno spazio vettoriale dotato di una forma bilineare simmetrica g e S, T due sottoinsiemi di V . Valgono le seguenti proprietà:

- (i) $S \subseteq T \Rightarrow T^\perp \subseteq S^\perp$;
- (ii) $S^\perp = \langle S \rangle^\perp$.

Dimostrazione. (i) Sia $v \in T^\perp$; si ha $g(v, w) = 0$, per ogni $w \in T$. Dato che $S \subseteq T$, si ha quindi anche $g(v, w) = 0$, per ogni $w \in S$, il che significa che $v \in S^\perp$.

(ii) Dall'inclusione $S \subseteq \langle S \rangle$ e dalla proprietà (i) segue che $\langle S \rangle^\perp \subseteq S^\perp$. Per dimostrare l'inclusione opposta, consideriamo un vettore $v \in S^\perp$. Ogni vettore $w \in \langle S \rangle$ si può esprimere nella forma

$$w = \lambda_1 w_1 + \lambda_2 w_2 + \cdots + \lambda_r w_r,$$

per qualche $\lambda_1, \dots, \lambda_r \in K$ e qualche $w_1, \dots, w_r \in S$. Si ha quindi

$$g(v, w) = g\left(v, \sum_{i=1}^r \lambda_i w_i\right) = \sum_{i=1}^r \lambda_i g(v, w_i) = 0.$$

Ciò significa che v è ortogonale a tutti i vettori di $\langle S \rangle$, quindi $v \in \langle S \rangle^\perp$. \square

Definizione 5.3.17. Sia V uno spazio vettoriale su K e $g : V \times V \rightarrow K$ una forma bilineare simmetrica. Una base $\{v_1, v_2, \dots, v_n\}$ di V costituita da vettori a due a due ortogonali è detta una *base ortogonale* di V .

In uno spazio vettoriale V dotato di una forma bilineare simmetrica g , potremmo definire il quadrato della norma di un vettore v ponendo $\|v\|^2 = g(v, v)$. Tale definizione è motivata dal fatto che, nello spazio vettoriale \mathbb{R}^n dotato del prodotto scalare usuale, si ha $\|v\|^2 = v \cdot v$, per ogni $v \in \mathbb{R}^n$. In generale, non è però possibile definire la *norma* di v ; infatti non è detto che nel campo K esista una radice quadrata di $g(v, v)$.

A titolo di esempio, consideriamo lo spazio vettoriale $V = \mathbb{Q}^2$, sul campo \mathbb{Q} dei numeri razionali, dotato del prodotto scalare usuale. Il vettore $v = (1, 1)$ è tale che $\|v\|^2 = v \cdot v = 2$, quindi nel campo \mathbb{Q} non è possibile definire la norma di v , dato che dovrebbe essere $\|v\| = \sqrt{2}$, ma $\sqrt{2} \notin \mathbb{Q}$. Un esempio analogo si ottiene considerando lo spazio vettoriale reale $V = \mathbb{R}^2$, dotato della forma bilineare simmetrica definita nell'Esempio 5.3.11. In questo caso, per il vettore $v = (1, -1)$ si ha $\|v\|^2 = g(v, v) = -2$, pertanto dovrebbe essere $\|v\| = \sqrt{-2}$, ma $\sqrt{-2} \notin \mathbb{R}$.

Naturalmente, nel caso particolare in cui $g(v, v) = 1$, si ha anche $\|v\| = 1$. Possiamo quindi dare la seguente definizione:

Definizione 5.3.18. Sia V uno spazio vettoriale dotato di una forma bilineare simmetrica g . Un vettore $v \in V$ si dice *normalizzato* se $g(v, v) = 1$.

Osservazione 5.3.19. Se v non è un vettore isotropo, cioè se $g(v, v) \neq 0$, e se nel campo K esiste una radice quadrata di $g(v, v)$, è possibile *normalizzare* il vettore v dividendolo per $\sqrt{g(v, v)}$. Infatti, ponendo $v' = v / \sqrt{g(v, v)}$, si ha:

$$\|v'\|^2 = g(v', v') = \frac{1}{(\sqrt{g(v, v)})^2} g(v, v) = 1.$$

Definizione 5.3.20. Sia V uno spazio vettoriale dotato di una forma bilineare simmetrica g . Una base $\{v_1, v_2, \dots, v_n\}$ di V si dice *ortonormale* se essa è una base ortogonale e se tutti i vettori v_1, \dots, v_n sono normalizzati, cioè se $g(v_i, v_j) = 0$ per $i \neq j$, mentre $g(v_i, v_i) = 1$, per ogni $i = 1, \dots, n$.

Consideriamo ora più in dettaglio il problema di definire la norma di un vettore in uno spazio vettoriale V dotato di una forma bilineare simmetrica g . Come già accennato in precedenza, ciò è possibile se, per ogni $v \in V$, esiste nel campo K una radice quadrata di $g(v, v)$. È quindi naturale imporre delle condizioni su g che garantiscano l'esistenza di $\sqrt{g(v, v)}$, per ogni $v \in V$.

Notiamo, ad esempio, che se K è il campo \mathbb{C} dei numeri complessi tale richiesta è soddisfatta per ogni forma bilineare simmetrica g . Ciò non vale invece se K è il campo \mathbb{R} dei numeri reali: in tal caso è necessario richiedere che $g(v, v) \geq 0$, per ogni $v \in V$.

Diamo quindi la seguente definizione:

Definizione 5.3.21. Sia V uno spazio vettoriale reale. Una forma bilineare simmetrica g su V è detta:

- (i) *definita positiva* se $g(v, v) > 0$, per ogni $v \in V$, $v \neq \mathbf{0}$;
- (ii) *definita negativa* se $g(v, v) < 0$, per ogni $v \in V$, $v \neq \mathbf{0}$;

- (iii) *semidefinita positiva* se $g(v, v) \geq 0$, per ogni $v \in V$;
- (iv) *semidefinita negativa* se $g(v, v) \leq 0$, per ogni $v \in V$;
- (v) *indefinita* se esistono $v, w \in V$ tali che $g(v, v) > 0$ e $g(w, w) < 0$.

Nel caso in cui la forma bilineare simmetrica g sia semidefinita positiva è possibile porre $\|v\| = \sqrt{g(v, v)}$. Notiamo tuttavia che se si desidera che l'unico vettore avente norma nulla sia il vettore nullo, è necessario richiedere che g sia definita positiva. Diamo quindi la seguente definizione:

Definizione 5.3.22. Sia V uno spazio vettoriale reale dotato di una forma bilineare simmetrica definita positiva g . La *norma* di un vettore $v \in V$ è definita ponendo

$$\|v\| = \sqrt{g(v, v)}.$$

La norma così definita soddisfa le seguenti proprietà, la cui verifica è immediata:

- (i) $\|v\| \geq 0$, per ogni $v \in V$, e $\|v\| = 0$ se e solo se $v = \mathbf{0}$;
- (ii) $\|\lambda v\| = |\lambda| \|v\|$, per ogni $\lambda \in \mathbb{R}$ e ogni $v \in V$.

Come per l'usuale norma dei vettori in \mathbb{R}^n , vale poi il seguente risultato:

Proposizione 5.3.23 (DISUGUAGLIANZA DI CAUCHY–SCHWARZ). *Sia g una forma bilineare simmetrica definita positiva su uno spazio vettoriale reale V . Per ogni coppia di vettori $v, w \in V$, si ha*

$$|g(v, w)| \leq \|v\| \|w\|. \quad (5.3.1)$$

Inoltre vale il segno di uguaglianza se e solo se i due vettori sono linearmente dipendenti.

Dimostrazione. La dimostrazione è esattamente la stessa di quella della Proposizione 5.1.6. Per adattarla al presente contesto è sufficiente definire il “prodotto scalare” di due vettori $v, w \in V$ ponendo $v \cdot w = g(v, w)$. \square

Una conseguenza immediata della disuguaglianza di Cauchy–Schwarz è la disuguaglianza triangolare:

Proposizione 5.3.24 (DISUGUAGLIANZA TRIANGOLARE). *Sia V uno spazio vettoriale reale dotato di una forma bilineare simmetrica definita positiva g . Per ogni $v, w \in V$, si ha*

$$\|v + w\| \leq \|v\| + \|w\|.$$

Dimostrazione. La dimostrazione è del tutto analoga a quella della Proposizione 5.1.8. \square

La validità della disuguaglianza di Cauchy–Schwarz ci permette, a sua volta, di definire l'angolo compreso tra due vettori in modo del tutto analogo a quanto abbiamo fatto per i vettori di \mathbb{R}^n :

Definizione 5.3.25. Sia V uno spazio vettoriale reale e sia g una forma bilineare simmetrica definita positiva su V . Dati due vettori non nulli $v, w \in V$, l'*angolo* (non orientato) tra essi compreso è l'unico $\phi \in [0, \pi]$ tale che

$$\cos \phi = \frac{g(v, w)}{\|v\| \|w\|}. \quad (5.3.2)$$

I risultati descritti finora mostrano come una forma bilineare simmetrica definita positiva su uno spazio vettoriale reale V permetta di definire le lunghezze dei vettori e gli angoli tra vettori in modo del tutto analogo a quanto accadeva per lo spazio vettoriale \mathbb{R}^n dotato del prodotto scalare usuale. Per tale ragione, uno spazio vettoriale reale dotato di una forma bilineare simmetrica definita positiva viene detto uno *spazio vettoriale euclideo*.

In uno spazio vettoriale euclideo (V, g) si definisce il *prodotto scalare* di due vettori $v, w \in V$ ponendo $v \cdot w = g(v, w)$. Con tale definizione nello spazio vettoriale V valgono, senza alcuna modifica, (quasi)³ tutti i risultati sulle aree e i volumi descritti nella Sezione 5.2.

Prima di affrontare lo studio dettagliato degli spazi vettoriali euclidei ci occuperemo delle matrici associate alle forme bilineari.

5.4 Forme bilineari e matrici

Sia V uno spazio vettoriale di dimensione finita su un campo K e sia g una forma bilineare definita su V . Consideriamo una base $\mathbf{v} = \{v_1, \dots, v_n\}$ di V e poniamo $g_{ij} = g(v_i, v_j)$, per ogni $i, j = 1, \dots, n$. Gli scalari g_{ij} formano una matrice quadrata, di ordine n , a coefficienti in K .

Definizione 5.4.1. Con le notazioni precedenti, la matrice $G = (g_{ij})$ è detta la *matrice della forma bilineare g* , rispetto alla base \mathbf{v} di V .

La conoscenza della matrice di g permette di calcolare $g(v, w)$ per ogni coppia di vettori $v, w \in V$. Infatti, esprimendo v e w come combinazioni lineari dei vettori della base di V ,

$$v = \sum_{i=1}^n \lambda_i v_i, \quad w = \sum_{j=1}^n \mu_j v_j,$$

grazie alla bilinearità di g , si ha:

$$\begin{aligned} g(v, w) &= g\left(\sum_{i=1}^n \lambda_i v_i, \sum_{j=1}^n \mu_j v_j\right) \\ &= \sum_{i,j=1}^n \lambda_i \mu_j g(v_i, v_j) \\ &= \sum_{i,j=1}^n \lambda_i \mu_j g_{ij}. \end{aligned}$$

Considerando i vettori $(\lambda_1, \lambda_2, \dots, \lambda_n)$ e $(\mu_1, \mu_2, \dots, \mu_n)$, costituiti dalle componenti di v e w rispetto alla base fissata di V , è possibile esprimere $g(v, w)$ in termini di prodotti tra matrici e vettori come segue:

$$g(v, w) = (\lambda_1, \dots, \lambda_n) G \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}.$$

³Più precisamente, tutti tranne quelli che usano il fatto che la base canonica è una base ortonormale.

Ad ogni forma bilineare g su V possiamo quindi associare una matrice quadrata G a coefficienti in K la quale, ovviamente, dipende dalla scelta di una base di V . È immediato verificare che la forma bilineare g è simmetrica se e solo se $g_{ij} = g_{ji}$, per ogni $i, j = 1, \dots, n$, cioè se e solo se G è una matrice simmetrica. Analogamente, g è antisimmetrica se e solo se $g_{ij} = -g_{ji}$, per ogni $i, j = 1, \dots, n$, cioè se e solo se G è una matrice antisimmetrica.

Osservazione 5.4.2. Siano U e W due sottospazi vettoriali di uno spazio vettoriale V e sia $g_V = g_U \oplus g_W$, ove g_U e g_W sono delle forme bilineari definite su U e W rispettivamente. Sia $\mathbf{u} = \{u_1, \dots, u_r\}$ una base di U e indichiamo con G_U la matrice di g_U rispetto alla base \mathbf{u} . Analogamente, consideriamo una base $\mathbf{w} = \{w_1, \dots, w_s\}$ di W e indichiamo con G_W la matrice di g_W rispetto alla base \mathbf{w} . Se indichiamo con G la matrice di g_V rispetto alla base \mathbf{v} di V costituita dai vettori $u_1, \dots, u_r, w_1, \dots, w_s$, è immediato verificare che la matrice G ha la seguente forma a blocchi

$$G = \left(\begin{array}{c|c} G_U & \mathbf{0}_{r,s} \\ \hline \mathbf{0}_{s,r} & G_W \end{array} \right)$$

ove $\mathbf{0}_{r,s}$ indica la matrice nulla con r righe e s colonne.

Descriviamo ora il nucleo di una forma bilineare simmetrica g in termini della matrice ad essa associata:

Proposizione 5.4.3. *Siano V uno spazio vettoriale su K , g una forma bilineare simmetrica su V e $\mathbf{v} = \{v_1, \dots, v_n\}$ una base di V . Indichiamo con $G \in M_n(K)$ la matrice di g rispetto alla base \mathbf{v} di V . Il nucleo di g è l'insieme dei vettori v del tipo $v = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n$, ove la n -upla $(\lambda_1, \dots, \lambda_n)$ è una soluzione del sistema lineare omogeneo $G\mathbf{x} = \mathbf{0}$.*

Dimostrazione. Un vettore $v = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n$ appartiene al nucleo di g se e solo se $g(w, v) = 0$ per ogni $w \in V$. Poiché i vettori v_1, v_2, \dots, v_n formano una base di V , ciò equivale a richiedere che $g(v_i, v) = 0$, per $i = 1, \dots, n$. Si ha dunque, per ogni vettore v_i della base di V ,

$$\begin{aligned} g(v_i, v) &= g(v_i, \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n) \\ &= \lambda_1 g(v_i, v_1) + \lambda_2 g(v_i, v_2) + \dots + \lambda_n g(v_i, v_n) \\ &= g_{i1}\lambda_1 + g_{i2}\lambda_2 + \dots + g_{in}\lambda_n = 0. \end{aligned}$$

Queste sono precisamente le n equazioni del sistema lineare

$$G \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \mathbf{0}.$$

□

Corollario 5.4.4. *Sia V uno spazio vettoriale di dimensione finita su K . Una forma bilineare simmetrica g definita su V è non degenere se e solo se la sua matrice G , rispetto a una qualunque base di V , è non singolare, cioè se e solo se $\det G \neq 0$.*

Dimostrazione. Ricordiamo che g è non degenere se e solo se $\text{Ker}(g) = \{\mathbf{0}\}$. In base alla proposizione precedente, ciò equivale a richiedere che il sistema lineare $G\mathbf{X} = \mathbf{0}$ abbia come unica soluzione il vettore nullo. Per il Teorema di Cramer questo avviene se e solo se la matrice G è non singolare. \square

Proposizione 5.4.5. *Siano V uno spazio vettoriale di dimensione n su K e g una forma bilineare simmetrica non degenere su V . Per ogni sottospazio U di V , si ha*

$$\dim U^\perp = n - \dim U.$$

Dimostrazione. Sia U un sottospazio di dimensione r di V e consideriamo una base v_1, v_2, \dots, v_r di U . Completiamo questa base in modo da ottenere una base $v_1, \dots, v_r, v_{r+1}, \dots, v_n$ di V . Un vettore $v = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n$ appartiene a U^\perp se e solo se $g(u, v) = 0$, per ogni $u \in U$. Poiché i vettori v_1, v_2, \dots, v_r sono una base di U , ciò equivale a richiedere che $g(v_i, v) = 0$ per $i = 1, \dots, r$. Si hanno dunque le seguenti equazioni

$$\begin{aligned} g(v_i, v) &= g(v_i, \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n) \\ &= \lambda_1 g(v_i, v_1) + \lambda_2 g(v_i, v_2) + \dots + \lambda_n g(v_i, v_n) \quad (5.4.1) \\ &= g_{i1} \lambda_1 + g_{i2} \lambda_2 + \dots + g_{in} \lambda_n = 0, \end{aligned}$$

per $i = 1, \dots, r$. In questo modo si ottiene un sistema di equazioni lineari omogenee la cui matrice G' è precisamente la sottomatrice costituita dalle prime r righe della matrice G . Poiché, per ipotesi, g è non degenere, la matrice G è non singolare, quindi tutte le sue righe sono linearmente indipendenti. Da ciò deriva che la sottomatrice G' ha rango massimo, pari a r . Dal Teorema di Rouché–Capelli segue quindi che lo spazio delle soluzioni del sistema lineare (5.4.1) ha dimensione $n - r$. Dato che le soluzioni di tale sistema forniscono le componenti, rispetto alla base v_1, \dots, v_n , dei vettori appartenenti al sottospazio U^\perp di V , si conclude che $\dim U^\perp = n - r$. \square

Corollario 5.4.6. *Sia V uno spazio vettoriale di dimensione finita su K e g una forma bilineare simmetrica non degenere su V . Per ogni sottospazio U di V , si ha*

$$(U^\perp)^\perp = U.$$

Dimostrazione. Se $n = \dim V$, per la proposizione precedente si ha:

$$\dim(U^\perp)^\perp = n - \dim U^\perp = n - (n - \dim U) = \dim U.$$

Pertanto U e $(U^\perp)^\perp$ sono due sottospazi vettoriali della stessa dimensione. Poiché sappiamo che $U \subseteq (U^\perp)^\perp$ (vedi Proposizione 5.3.15), deve necessariamente essere $U = (U^\perp)^\perp$. \square

Osservazione 5.4.7. Si noti che la dimostrazione del risultato precedente usa in modo essenziale il fatto che V abbia dimensione finita. Si potrebbe infatti dimostrare che, nel caso di uno spazio V di dimensione infinita, per un sottospazio U di V si ha, in generale, solo un'inclusione propria $U \subset (U^\perp)^\perp$.

Terminiamo questa sezione dimostrando il seguente risultato:

Proposizione 5.4.8. *Siano V uno spazio vettoriale sul campo K e g una forma bilineare simmetrica non degenere su V . Per ogni W_1, W_2 sottospazi di V , si ha:*

- (i) $(W_1 + W_2)^\perp = W_1^\perp \cap W_2^\perp$;
- (ii) $(W_1 \cap W_2)^\perp \supseteq W_1^\perp + W_2^\perp$.

Inoltre, se V ha dimensione finita, nella (ii) vale l'uguaglianza.

Dimostrazione. (i) Sia $v \in (W_1 + W_2)^\perp$. Si ha $g(v, w) = 0$ per ogni $w \in W_1 + W_2$ e dunque anche $g(v, w_1) = 0$ e $g(v, w_2) = 0$, per ogni $w_1 \in W_1$ e ogni $w_2 \in W_2$. Ciò dimostra che $v \in W_1^\perp \cap W_2^\perp$.

Viceversa, supponiamo che v appartenga a $W_1^\perp \cap W_2^\perp$, cioè che sia $g(v, w_1) = 0$ e $g(v, w_2) = 0$, per ogni $w_1 \in W_1$ e ogni $w_2 \in W_2$. Ricordando che ogni vettore $w \in W_1 + W_2$ si scrive nella forma $w = w_1 + w_2$, con $w_1 \in W_1$ e $w_2 \in W_2$, si ha $g(v, w) = g(v, w_1) + g(v, w_2) = 0$, il che dimostra che $v \in (W_1 + W_2)^\perp$.

(ii) Sia $v \in W_1^\perp + W_2^\perp$. Allora si ha $v = v_1 + v_2$, con $v_1 \in W_1^\perp$ e $v_2 \in W_2^\perp$. Per ogni $w \in W_1 \cap W_2$ si ha dunque $g(v, w) = g(v_1, w) + g(v_2, w) = 0$, quindi $v \in (W_1 \cap W_2)^\perp$, come volevasi dimostrare.

Infine, se V ha dimensione finita, poniamo $n = \dim V$, $m_1 = \dim W_1$ e $m_2 = \dim W_2$. Dalla Proposizione 5.4.5 segue che

$$\begin{aligned}\dim W_1^\perp &= n - m_1 \\ \dim W_2^\perp &= n - m_2 \\ \dim(W_1 \cap W_2)^\perp &= n - \dim(W_1 \cap W_2).\end{aligned}$$

Per la formula di Grassmann, si ha

$$\begin{aligned}\dim(W_1^\perp + W_2^\perp) &= \dim W_1^\perp + \dim W_2^\perp - \dim(W_1^\perp \cap W_2^\perp) \\ &= 2n - m_1 - m_2 - \dim(W_1^\perp \cap W_2^\perp).\end{aligned}\tag{5.4.2}$$

Per il punto (i) si ha

$$\begin{aligned}\dim(W_1^\perp \cap W_2^\perp) &= \dim(W_1 + W_2)^\perp \\ &= n - \dim(W_1 + W_2)\end{aligned}$$

e dunque, usando ancora la formula di Grassmann, si trova

$$\dim(W_1^\perp \cap W_2^\perp) = n - m_1 - m_2 + \dim(W_1 \cap W_2).$$

Sostituendo nella (5.4.2) si trova infine

$$\dim(W_1^\perp + W_2^\perp) = n - \dim(W_1 \cap W_2) = \dim(W_1 \cap W_2)^\perp.$$

Dall'inclusione (ii) e dall'uguaglianza delle dimensioni dei due sottospazi si deduce che vale l'uguaglianza $(W_1 \cap W_2)^\perp = W_1^\perp + W_2^\perp$. \square

5.4.1 Cambiamenti di base

La matrice associata a una forma bilineare g , definita su uno spazio vettoriale di dimensione finita V , dipende dalla scelta di una base di V . In questa sezione ci proponiamo di descrivere come cambia la matrice di g in corrispondenza di un cambiamento della base di V .

Consideriamo quindi uno spazio vettoriale V di dimensione n su K e una forma bilineare g su V . Siano $\mathbf{v} = \{v_1, \dots, v_n\}$ e $\mathbf{v}' = \{v'_1, \dots, v'_n\}$ due basi di V

e indichiamo con $G = (g_{ij})$ la matrice di g rispetto alla base \mathbf{v} e con $G' = (g'_{ij})$ la matrice di g rispetto alla base \mathbf{v}' . Ricordiamo che ciò significa che

$$g_{ij} = g(v_i, v_j), \quad \text{e} \quad g'_{ij} = g(v'_i, v'_j),$$

per ogni $j = 1, \dots, n$.

Indichiamo con $\alpha_{\mathbf{v}} : V \xrightarrow{\sim} K^n$ l'isomorfismo che associa ad ogni vettore $v \in V$ la n -upla (x_1, \dots, x_n) delle sue componenti rispetto alla base \mathbf{v} e con $\alpha_{\mathbf{v}'} : V \xrightarrow{\sim} K^n$ l'isomorfismo che associa ad ogni $v \in V$ la n -upla (x'_1, \dots, x'_n) delle sue componenti rispetto alla base \mathbf{v}' .

Componendo $\alpha_{\mathbf{v}'}$ con l'inverso dell'isomorfismo $\alpha_{\mathbf{v}}$ otteniamo un isomorfismo di K^n in sé, il quale corrisponde alla moltiplicazione per una qualche matrice $P \in M_n(K)$. Indicheremo questo isomorfismo con $F_P : K^n \xrightarrow{\sim} K^n$. Si ottiene così il seguente diagramma commutativo:

$$\begin{array}{ccc} & V & \\ \alpha_{\mathbf{v}} \swarrow & & \searrow \alpha_{\mathbf{v}'} \\ K^n & \xrightarrow{F_P} & K^n \end{array}$$

Facciamo notare che la matrice P è invertibile, dato che la corrispondente applicazione lineare F_P è un isomorfismo.

Vediamo ora di ottenere una descrizione più esplicita della matrice P . Ricordiamo che le colonne di P sono date dalle immagini, tramite l'isomorfismo F_P , dei vettori della base canonica di K^n . Sia $e_j = {}^t(0, \dots, 0, 1, 0, \dots, 0)$ il j -esimo vettore della base canonica di K^n (tutte le coordinate sono nulle tranne la j -esima che è uguale a 1). Tramite l'isomorfismo $\alpha_{\mathbf{v}}$, il vettore $e_j \in K^n$ corrisponde al j -esimo vettore v_j della base \mathbf{v} di V . Si ha quindi

$$F_P(e_j) = \alpha_{\mathbf{v}'}(\alpha_{\mathbf{v}}^{-1}(e_j)) = \alpha_{\mathbf{v}'}(v_j),$$

dove ricordiamo che $\alpha_{\mathbf{v}'}(v_j) \in K^n$ è il vettore costituito dalle componenti del vettore v_j calcolate rispetto alla base \mathbf{v}' ; questo vettore è la j -esima colonna di P .

In conclusione, possiamo affermare che le *colonne* della matrice P non sono altro che le componenti dei vettori v_1, \dots, v_n della base \mathbf{v} di V calcolate rispetto alla seconda base \mathbf{v}' . Con un analogo ragionamento, scambiando i ruoli delle due basi, si potrebbe dimostrare che le colonne della matrice inversa P^{-1} sono precisamente le componenti dei vettori v'_1, \dots, v'_n della base \mathbf{v}' di V calcolate rispetto alla prima base \mathbf{v} .

In dettaglio, se $X = {}^t(x_1, x_2, \dots, x_n)$ è il vettore costituito dalle componenti di un vettore v rispetto alla base \mathbf{v} e se $X' = {}^t(x'_1, x'_2, \dots, x'_n)$ è il vettore costituito dalle componenti dello stesso vettore v rispetto alla base \mathbf{v}' , si ha

$$v = \sum_{j=1}^n x_j v_j = \sum_{i=1}^n x'_i v'_i, \quad X' = P X.$$

Si ha pertanto $x'_i = \sum_{j=1}^n p_{ij}x_j$, da cui si ricava

$$\begin{aligned} \sum_{j=1}^n x_j v_j &= \sum_{i=1}^n x'_i v'_i \\ &= \sum_{i=1}^n \left(\sum_{j=1}^n p_{ij}x_j \right) v'_i \\ &= \sum_{i,j=1}^n p_{ij}x_j v'_i \\ &= \sum_{j=1}^n x_j \left(\sum_{i=1}^n p_{ij}v'_i \right) \end{aligned}$$

da cui si deduce che

$$v_j = \sum_{i=1}^n p_{ij}v'_i.$$

Siamo ora in grado di determinare la relazione esistente tra le due matrici G e G' di g . Si ha infatti:

$$\begin{aligned} g_{ij} &= g(v_i, v_j) \\ &= g\left(\sum_{h=1}^n p_{hi}v'_h, \sum_{k=1}^n p_{kj}v'_k\right) \\ &= \sum_{h,k=1}^n p_{hi}p_{kj}g(v'_h, v'_k) \\ &= \sum_{h,k=1}^n p_{hi}g'_{hk}p_{kj}. \end{aligned}$$

Questa uguaglianza può essere riscritta in termini di matrici come segue:

$$G = {}^t P G' P. \quad (5.4.3)$$

Un metodo alternativo per ricavare quest'ultima formula è il seguente. Indichiamo con $X = {}^t(x_1, x_2, \dots, x_n)$ e $Y = {}^t(y_1, y_2, \dots, y_n)$ i vettori costituiti dalle componenti di v e w rispetto alla base \mathbf{v} e con $X' = {}^t(x'_1, x'_2, \dots, x'_n)$ e $Y' = {}^t(y'_1, y'_2, \dots, y'_n)$ i vettori costituiti dalle componenti di v e w rispetto alla base \mathbf{v}' . Si ha pertanto:

$$g(v, w) = {}^t X G Y = {}^t X' G' Y', \quad X' = P X, \quad Y' = P Y.$$

Sostituendo, si ottiene:

$${}^t X G Y = {}^t X' G' Y' = {}^t (P X) G' (P Y) = {}^t X {}^t P G' P Y.$$

Poiché questa uguaglianza vale per ogni $v, w \in V$, cioè per ogni $X, Y \in K^n$, si deduce che deve sussistere l'uguaglianza (5.4.3).

Diamo ora la seguente definizione:

Definizione 5.4.9. Due matrici quadrate G e G' di ordine n a coefficienti in K si dicono *congruenti* se esiste una matrice invertibile $P \in M_n(K)$ (cioè $P \in \text{GL}_n(K)$) tale che

$$G = {}^t P G' P.$$

Da quanto sopra detto si deduce il seguente risultato:

Corollario 5.4.10. Due matrici $G, G' \in M_n(K)$ rappresentano la stessa forma bilineare g definita su uno spazio vettoriale V di dimensione n su K , rispetto a basi diverse, se e solo se esse sono congruenti.

Osservazione 5.4.11. Si verifica facilmente che la relazione di congruenza è una relazione di equivalenza sull'insieme $M_n(K)$ delle matrici quadrate di ordine n a coefficienti in K .

Terminiamo questa sezione estendendo alle matrici la Definizione 5.3.21 data per le forme bilineari simmetriche:

Definizione 5.4.12. Sia $G \in M_n(\mathbb{R})$ una matrice simmetrica e sia $g : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ la corrispondente forma bilineare simmetrica (cioè la forma bilineare la cui matrice, rispetto alla base canonica di \mathbb{R}^n , è G). La matrice G è detta:

- (i) *definita positiva* se g è definita positiva;
- (ii) *definita negativa* se g è definita negativa;
- (iii) *semidefinita positiva* se g è semidefinita positiva;
- (iv) *semidefinita negativa* se g è semidefinita negativa;
- (v) *indefinita* se g è indefinita.

5.4.2 Basi ortogonali e ortonormali

In questa sezione dimostreremo che ogni spazio vettoriale reale V di dimensione finita, dotato di una forma bilineare simmetrica definita positiva, ammette una base ortonormale. Descriveremo inoltre un procedimento che permette di costruire una base ortonormale partendo da una base qualunque di V .

Iniziamo col dimostrare il seguente risultato:

Proposizione 5.4.13. Sia V uno spazio vettoriale reale e sia g una forma bilineare simmetrica definita positiva su V . Se i vettori v_1, v_2, \dots, v_r sono a due a due ortogonali, essi sono anche linearmente indipendenti.

Dimostrazione. Consideriamo una combinazione lineare

$$\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_r v_r = \mathbf{0}.$$

Per ogni $i = 1, \dots, r$, si ha

$$g\left(v_i, \sum_{j=1}^r \lambda_j v_j\right) = \sum_{j=1}^r \lambda_j g(v_i, v_j) = \lambda_i g(v_i, v_i) = 0,$$

dato che $g(v_i, v_j) = 0$ per ogni $i \neq j$. Essendo g definita positiva, è $g(v_i, v_i) > 0$, quindi deve essere $\lambda_i = 0$. Questo dimostra che i vettori v_1, \dots, v_r sono linearmente indipendenti. \square

Possiamo ora dimostrare il seguente risultato:

Teorema 5.4.14. *Ogni spazio vettoriale V di dimensione finita sul campo dei numeri reali, dotato di una forma bilineare simmetrica definita positiva g , possiede una base ortonormale.*

Dimostrazione. Procediamo per induzione su $n = \dim V$. Se $n = 1$ consideriamo un vettore non nullo $v \in V$. Poiché g è definita positiva, si ha $g(v, v) > 0$, quindi possiamo porre $w = v/\sqrt{g(v, v)}$. Il vettore w è normalizzato, cioè si ha $g(w, w) = 1$, e constituisce pertanto una base ortonormale di V .

Supponiamo quindi che il teorema valga per ogni spazio vettoriale reale di dimensione $n - 1$ e proviamo che allora esso vale anche per spazi di dimensione n . Sia V uno spazio vettoriale di dimensione n , come nell'enunciato, e sia $v \in V$ un vettore non nullo. Consideriamo il sottospazio $W = \langle v \rangle^\perp$ dotato della forma bilineare simmetrica definita positiva indotta da g . Per la Proposizione 5.4.5, si ha $\dim W = n - 1$ pertanto, per l'ipotesi induttiva, W possiede una base ortonormale w_1, \dots, w_{n-1} . Poiché, per ipotesi, g è definita positiva, si ha $g(v, v) > 0$; possiamo quindi porre $w_n = v/\sqrt{g(v, v)}$. È ora immediato verificare che $\{w_1, \dots, w_{n-1}, w_n\}$ è una base ortonormale di V . \square

Descriviamo ora in dettaglio un metodo, noto come *procedimento di ortonormalizzazione di Gram–Schmidt*, che permette di costruire una base ortonormale partendo da una base qualsiasi di V .

Consideriamo quindi uno spazio vettoriale V di dimensione n sul campo dei numeri reali, dotato di una forma bilineare simmetrica definita positiva g . Sia $\mathbf{v} = \{v_1, \dots, v_n\}$ una base qualunque di V . Poniamo $w_1 = v_1$ e cerchiamo un vettore w_2 , ortogonale a w_1 , della forma $w_2 = \alpha_1 w_1 + v_2$ (notiamo che, in questo modo, il sottospazio vettoriale generato da w_1 e w_2 coincide con quello generato da v_1 e v_2). La condizione di ortogonalità tra w_1 e w_2 si esprime ponendo $g(w_1, w_2) = 0$. Si ha pertanto

$$g(w_1, w_2) = g(w_1, \alpha_1 w_1 + v_2) = \alpha_1 g(w_1, w_1) + g(w_1, v_2) = 0,$$

da cui si ottiene

$$\alpha_1 = -\frac{g(w_1, v_2)}{g(w_1, w_1)}.$$

Il vettore cercato è quindi

$$w_2 = v_2 - \frac{g(w_1, v_2)}{g(w_1, w_1)} w_1. \quad (5.4.4)$$

Cerchiamo ora un vettore w_3 , ortogonale al sottospazio generato da w_1 e w_2 , della forma $w_3 = \alpha_1 w_1 + \alpha_2 w_2 + v_3$. Imponendo che w_3 sia ortogonale a w_1 e w_2 , si ottengono le equazioni $g(w_1, w_3) = 0$ e $g(w_2, w_3) = 0$. Sviluppando i calcoli, si trova:

$$\begin{aligned} g(w_1, w_3) &= g(w_1, \alpha_1 w_1 + \alpha_2 w_2 + v_3) \\ &= \alpha_1 g(w_1, w_1) + \alpha_2 g(w_1, w_2) + g(w_1, v_3) \\ &= \alpha_1 g(w_1, w_1) + g(w_1, v_3) = 0, \end{aligned}$$

$$\begin{aligned} g(w_2, w_3) &= g(w_2, \alpha_1 w_1 + \alpha_2 w_2 + v_3) \\ &= \alpha_1 g(w_2, w_1) + \alpha_2 g(w_2, w_2) + g(w_2, v_3) \\ &= \alpha_2 g(w_2, w_2) + g(w_2, v_3) = 0, \end{aligned}$$

da cui si ottiene

$$\alpha_1 = -\frac{g(w_1, v_3)}{g(w_1, w_1)}, \quad \alpha_2 = -\frac{g(w_2, v_3)}{g(w_2, w_2)}.$$

Il vettore cercato è quindi

$$w_3 = v_3 - \frac{g(w_1, v_3)}{g(w_1, w_1)} w_1 - \frac{g(w_2, v_3)}{g(w_2, w_2)} w_2. \quad (5.4.5)$$

Continuando in questo modo, si ottiene una base ortogonale $\{w_1, w_2, \dots, w_n\}$ di V . Più precisamente, supponiamo (per ipotesi induttiva) di aver già costruito i vettori w_1, w_2, \dots, w_{r-1} , che sono tra essi a due a due ortogonali. Cerchiamo allora un vettore w_r , della forma

$$w_r = \alpha_1 w_1 + \alpha_2 w_2 + \dots + \alpha_{r-1} w_{r-1} + v_r,$$

ortogonale al sottospazio generato da w_1, \dots, w_{r-1} (notiamo che, in questo modo, il sottospazio generato dai vettori w_1, \dots, w_r coincide con il sottospazio generato dai vettori v_1, \dots, v_r). Per ogni $i = 1, \dots, r-1$, la condizione di ortogonalità tra w_r e w_i fornisce la seguente equazione

$$\begin{aligned} g(w_i, w_r) &= g\left(w_i, \sum_{j=1}^{r-1} \alpha_j w_j + v_r\right) \\ &= \sum_{j=1}^{r-1} \alpha_j g(w_i, w_j) + g(w_i, v_r) \\ &= \alpha_i g(w_i, w_i) + g(w_i, v_r) = 0, \end{aligned}$$

da cui si ricava

$$\alpha_i = -\frac{g(w_i, v_r)}{g(w_i, w_i)}.$$

Il vettore w_r cercato è quindi dato da:

$$w_r = v_r - \frac{g(w_1, v_r)}{g(w_1, w_1)} w_1 - \frac{g(w_2, v_r)}{g(w_2, w_2)} w_2 - \dots - \frac{g(w_{r-1}, v_r)}{g(w_{r-1}, w_{r-1})} w_{r-1}.$$

Gli n vettori w_1, \dots, w_n così costruiti sono a due a due ortogonali, quindi sono linearmente indipendenti (vedi Proposizione 5.4.13); essi sono pertanto una base ortogonale di V . Per ottenere una base ortonormale non rimane altro che normalizzare i vettori trovati. A tal fine è sufficiente porre

$$w'_i = \frac{w_i}{\sqrt{g(w_i, w_i)}},$$

per ogni $i = 1, \dots, n$ (ancora una volta, ciò è possibile perché si è supposto che g sia definita positiva). I vettori w'_1, \dots, w'_n così costruiti sono una base ortonormale di V .

Se indichiamo con G la matrice di g rispetto alla base $\mathbf{v} = \{v_1, \dots, v_n\}$ di V e con G' la matrice di g rispetto alla base $\mathbf{w}' = \{w'_1, \dots, w'_n\}$, si ha

$$G' = {}^t P G P,$$

ove P è la matrice di cambiamento di base, cioè la matrice le cui colonne sono le componenti dei vettori w'_1, \dots, w'_n della nuova base, rispetto ai vettori v_1, \dots, v_n della base originale di V . Dato che $\mathbf{w}' = \{w'_1, \dots, w'_n\}$ è una base ortonormale, si ha

$$g'_{ij} = g(w'_i, w'_j) = \begin{cases} 1 & \text{se } i = j, \\ 0 & \text{se } i \neq j, \end{cases}$$

quindi G' è la matrice identica.

Osservazione 5.4.15. Si noti che ogni vettore w'_j della base ortonormale \mathbf{w}' si scrive come combinazione lineare dei vettori v_1, v_2, \dots, v_j della base \mathbf{v} . Ciò significa che, nella matrice di cambiamento di base P , tutti gli elementi sotto la diagonale principale sono nulli; P è quindi una matrice triangolare superiore.

Possiamo riassumere quanto detto finora nella seguente proposizione:

Proposizione 5.4.16. *Sia $G \in M_n(\mathbb{R})$ una matrice simmetrica definita positiva. Esiste una matrice invertibile $P \in M_n(\mathbb{R})$ tale che ${}^t P G P = \mathbf{1}$. Inoltre, tale matrice P può essere scelta triangolare superiore.*

Esempio 5.4.17. Applichiamo ora su un esempio concreto il metodo di ortogonalizzazione di Gram–Schmidt descritto in precedenza. Sia V uno spazio vettoriale reale di dimensione 4 e sia g la forma bilineare simmetrica su V di matrice

$$G = \begin{pmatrix} 4 & 2 & -2 & 2 \\ 2 & 10 & -7 & -2 \\ -2 & -7 & 6 & 3 \\ 2 & -2 & 3 & 10 \end{pmatrix}$$

rispetto alla base $\mathbf{v} = \{v_1, v_2, v_3, v_4\}$ di V . Ci proponiamo di costruire una base ortonormale di V .

Iniziamo ponendo $w_1 = v_1$. Si ha

$$g(w_1, w_1) = g(v_1, v_1) = 4, \quad g(w_1, v_2) = g(v_1, v_2) = 2,$$

quindi, dalla formula (5.4.4), otteniamo

$$w_2 = v_2 - \frac{g(w_1, v_2)}{g(w_1, w_1)} w_1 = v_2 - \frac{1}{2} v_1.$$

Ora si ha:

$$g(w_1, v_3) = g(v_1, v_3) = -2,$$

$$g(w_2, v_3) = g\left(v_2 - \frac{1}{2} v_1, v_3\right) = g(v_2, v_3) - \frac{1}{2} g(v_1, v_3) = -6,$$

$$g(w_2, w_2) = g\left(v_2 - \frac{1}{2} v_1, v_2 - \frac{1}{2} v_1\right) = g(v_2, v_2) - g(v_2, v_1) + \frac{1}{4} g(v_1, v_1) = 9.$$

Dalla formula (5.4.5) si ricava

$$w_3 = v_3 - \frac{g(w_1, v_3)}{g(w_1, w_1)} w_1 - \frac{g(w_2, v_3)}{g(w_2, w_2)} w_2 = v_3 + \frac{1}{6} v_1 + \frac{2}{3} v_2.$$

Infine, in modo del tutto analogo, si ha

$$w_4 = v_4 - \frac{g(w_1, v_4)}{g(w_1, w_1)} w_1 - \frac{g(w_2, v_4)}{g(w_2, w_2)} w_2 - \frac{g(w_3, v_4)}{g(w_3, w_3)} w_3.$$

Sviluppando i calcoli, si trova

$$\begin{aligned} g(w_1, v_4) &= 2 & g(w_2, v_4) &= -3 \\ g(w_3, v_4) &= 2 & g(w_3, w_3) &= 1, \end{aligned}$$

da cui si ricava

$$w_4 = -v_1 - v_2 - 2v_3 + v_4.$$

Calcoliamo infine $g(w_4, w_4)$:

$$g(w_4, w_4) = g(-v_1 - v_2 - 2v_3 + v_4, -v_1 - v_2 - 2v_3 + v_4) = 4.$$

Ora non rimane altro che normalizzare i vettori trovati:

$$\begin{aligned} w'_1 &= \frac{w_1}{\sqrt{g(w_1, w_1)}} = \frac{w_1}{\sqrt{4}} = \frac{1}{2} v_1, \\ w'_2 &= \frac{w_2}{\sqrt{g(w_2, w_2)}} = \frac{w_2}{\sqrt{9}} = -\frac{1}{6} v_1 + \frac{1}{3} v_2, \\ w'_3 &= \frac{w_3}{\sqrt{g(w_3, w_3)}} = \frac{w_3}{\sqrt{1}} = \frac{1}{6} v_1 + \frac{2}{3} v_2 + v_3, \\ w'_4 &= \frac{w_4}{\sqrt{g(w_4, w_4)}} = \frac{w_4}{\sqrt{4}} = -\frac{1}{2} v_1 - \frac{1}{2} v_2 - v_3 + \frac{1}{2} v_4. \end{aligned}$$

La matrice di cambiamento di base è quindi

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

che, come si vede, è triangolare superiore. È ora immediato verificare che ${}^t P G P = \mathbf{1}$.

5.5 Classificazione delle forme bilineari simmetriche reali

In questa sezione ci occuperemo dello studio e della classificazione delle forme bilineari simmetriche definite su uno spazio vettoriale reale di dimensione finita.

Sia dunque V uno spazio vettoriale di dimensione n sul campo \mathbb{R} dei numeri reali e sia g una forma bilineare simmetrica definita su V . Per la Proposizione 5.3.8, esiste un sottospazio vettoriale U di V tale che $V = \text{Ker}(g) \oplus U$ e tale che la restrizione di g a $U \times U$ sia una forma bilineare non degenere. Ciò significa che è sempre possibile scegliere una base $\{v_1, \dots, v_k, v_{k+1}, \dots, v_n\}$ di V (ove $\{v_1, \dots, v_k\}$ è una base di $\text{Ker}(g)$) rispetto alla quale la matrice G di g sia una matrice a blocchi del tipo

$$G = \left(\begin{array}{c|c} \mathbf{0}_{k,k} & \mathbf{0}_{k,n-k} \\ \hline \mathbf{0}_{n-k,k} & G_U \end{array} \right)$$

ove $\mathbf{0}_{r,s}$ indica la matrice nulla con r righe e s colonne mentre G_U è la matrice della restrizione di g al sottospazio U generato dai vettori $\{v_{k+1}, \dots, v_n\}$. Per classificare le forme bilineari simmetriche è quindi sufficiente classificare quelle non degeneri. Descriveremo ora la classificazione reale di tali forme.

Sia dunque g una forma bilineare simmetrica non degenera su uno spazio vettoriale reale V . Se g è definita positiva il procedimento di Gram–Schmidt permette di costruire una base ortonormale di V . Rispetto a tale base la matrice di g è la matrice identica. Se g è definita negativa allora $-g$ è definita positiva. Da ciò si deduce che esiste una base di V rispetto a cui la matrice di g è l'opposto della matrice identica. Poiché g è supposta non degenera, rimane dunque solo da studiare il caso in cui g è indefinita, cioè il caso in cui esistono dei vettori v tali che $g(v, v) > 0$ e dei vettori w tali che $g(w, w) < 0$.

L'idea per trattare tale caso è la seguente: si parte da una base qualsiasi di V e si usa un metodo simile al procedimento di Gram–Schmidt per costruire una base ortogonale di V . La matrice di g rispetto ad una tale base sarà dunque diagonale. Ogni elemento diagonale positivo $\alpha > 0$ può essere reso uguale a 1 dividendo il corrispondente vettore di base per $\sqrt{\alpha}$, mentre ogni elemento diagonale negativo $\beta < 0$ può essere reso uguale a -1 dividendo il corrispondente vettore di base per $\sqrt{-\beta}$; non ci possono essere elementi diagonali nulli perché la forma g è non degenera. In questo modo si ottiene una base di V rispetto alla quale la matrice di g è una matrice diagonale, in cui gli elementi sulla diagonale sono 1 o -1 .

Teorema 5.5.1 (TEOREMA DI SYLVESTER). *Siano V uno spazio vettoriale reale di dimensione n e $g : V \times V \rightarrow \mathbb{R}$ una forma bilineare simmetrica non degenera. Esiste una base di V rispetto alla quale la matrice di g è una matrice a blocchi del tipo*

$$\left(\begin{array}{c|c} \mathbf{1}_{r,r} & \mathbf{0}_{r,s} \\ \hline \mathbf{0}_{s,r} & -\mathbf{1}_{s,s} \end{array} \right)$$

ove $r+s = n$ e ove $\mathbf{1}_{r,r}$ indica la matrice identica di ordine r e $-\mathbf{1}_{s,s}$ è l'opposto della matrice identica di ordine s . Inoltre i numeri r ed s (i numeri di 1 e -1 sulla diagonale) sono univocamente determinati (dipendono solo da g e non dalla base scelta). La coppia (r, s) è detta la *segnatura della forma g* , mentre la differenza $r - s$ è anche detta l'*indice d'inerzia di g* .

Dimostrazione. Per quanto detto in precedenza l'enunciato del teorema è banalmente verificato se g è definita positiva o negativa. Supponiamo quindi che g sia indefinita. In tal caso esiste un vettore $v \in V$ tale che $g(v, v) = \alpha > 0$. Se poniamo $v_1 = v/\sqrt{\alpha}$, si ha $g(v_1, v_1) = 1$. Sia $W = \langle v_1 \rangle^\perp$; si ha $\dim W = n - 1$. Se $\{w_2, \dots, w_n\}$ è una base di W , i vettori $\{v_1, w_2, \dots, w_n\}$ sono una base di V , rispetto alla quale la matrice G di g è una matrice a blocchi del tipo

$$G = \left(\begin{array}{c|c} 1 & \mathbf{0}_{1, n-1} \\ \hline \mathbf{0}_{n-1, 1} & G_W \end{array} \right)$$

ove G_W è la matrice della restrizione di g a W (rispetto alla base di W indicata). Ora basta ripetere il ragionamento appena descritto applicandolo alla restrizione di g a W (in alternativa, si ragioni per induzione sulla dimensione di V).

Questo dimostra l'esistenza di una base di V rispetto alla quale la matrice di g ha la forma voluta. Rimane ora da dimostrare l'unicità dei numeri r e s , ovvero la loro indipendenza dalla particolare base scelta.

Supponiamo che esistano due basi $\{v_1, \dots, v_n\}$ e $\{v'_1, \dots, v'_n\}$ di V rispetto alle quali le matrici di g siano

$$G = \begin{pmatrix} \mathbf{1}_{r,r} & \mathbf{0}_{r,s} \\ \mathbf{0}_{s,r} & -\mathbf{1}_{s,s} \end{pmatrix}, \quad G' = \begin{pmatrix} \mathbf{1}_{r',r'} & \mathbf{0}_{r',s'} \\ \mathbf{0}_{s',r'} & -\mathbf{1}_{s',s'} \end{pmatrix}$$

con $r+s = r'+s' = n$.

Poniamo $W^+ = \langle v_1, \dots, v_r \rangle$, $W^- = \langle v_{r+1}, \dots, v_n \rangle$, $Z^+ = \langle v'_1, \dots, v'_{r'} \rangle$, $Z^- = \langle v'_{r'+1}, \dots, v'_n \rangle$. Si ha $\dim W^+ = r$, $\dim W^- = s$, $\dim Z^+ = r'$, $\dim Z^- = s'$,

$$V = W^+ \oplus W^- = Z^+ \oplus Z^-,$$

inoltre la restrizione di g a W^+ e Z^+ è definita positiva, mentre g ristretta a W^- e Z^- è definita negativa.

Supponiamo, per assurdo, che sia $r < r'$. Ogni vettore v'_i ($i = 1, \dots, r'$) della base di Z^+ si può scrivere (in modo unico) nella forma

$$v'_i = u_i + w_i,$$

con $u_i \in W^+$ e $w_i \in W^-$.

Dato che $r' > r = \dim W^+$, i vettori $u_1, \dots, u_{r'}$ devono essere linearmente dipendenti. Si ha quindi

$$\lambda_1 u_1 + \dots + \lambda_{r'} u_{r'} = \mathbf{0},$$

con $\lambda_1, \dots, \lambda_{r'}$ non tutti nulli.

Sia $z = \lambda_1 v'_1 + \dots + \lambda_{r'} v'_{r'}$. Allora $z \neq \mathbf{0}$, perché i λ_j non sono tutti nulli e i vettori $\{v'_1, \dots, v'_{r'}\}$ sono una base di Z^+ , quindi $g(z, z) > 0$ (perché la restrizione di g a Z^+ è definita positiva). Tuttavia si ha

$$\begin{aligned} z &= \lambda_1 v'_1 + \dots + \lambda_{r'} v'_{r'} \\ &= \lambda_1 (u_1 + w_1) + \dots + \lambda_{r'} (u_{r'} + w_{r'}) \\ &= \lambda_1 w_1 + \dots + \lambda_{r'} w_{r'} \in W^- \end{aligned}$$

quindi $g(z, z) < 0$, perché la restrizione di g a W^- è definita negativa. Siamo così arrivati ad un assurdo, che nasce dall'aver supposto $r' > r$. Deve quindi essere $r' \leq r$.

Ripetendo il ragionamento dopo aver scambiato i ruoli di r e r' , si conclude che deve anche essere $r \leq r'$, il che dimostra l'uguaglianza $r = r'$ (e quindi anche $s = s'$). \square

Possiamo riassumere quanto visto finora nel seguente risultato:

Teorema 5.5.2. *Sia V uno spazio vettoriale di dimensione n sul campo \mathbb{R} dei numeri reali e sia $g : V \times V \rightarrow \mathbb{R}$ una forma bilineare simmetrica. Esiste una base di V rispetto alla quale la matrice di g è una matrice a blocchi del tipo*

$$\begin{pmatrix} \mathbf{0}_{k,k} & \mathbf{0}_{k,r} & \mathbf{0}_{k,s} \\ \hline \mathbf{0}_{r,k} & \mathbf{1}_{r,r} & \mathbf{0}_{r,s} \\ \hline \mathbf{0}_{s,k} & \mathbf{0}_{s,r} & -\mathbf{1}_{s,s} \end{pmatrix}$$

ove $k = \dim \text{Ker}(g)$ e $k + r + s = n$. Gli interi k , r e s sono unicamente determinati e $r + s$ coincide con il rango della matrice di g rispetto ad una base qualunque di V .

Dimostrazione. L'unica affermazione che non è stata ancora dimostrata è quella che identifica $r + s$ con il rango di una matrice di g rispetto a una base qualunque di V . Ciò discende dal fatto che il rango della matrice indicata nell'enunciato è precisamente $r + s$ e che due diverse matrici G e G' di g sono collegate dall'uguaglianza $G' = {}^t PGP$, per una qualche matrice invertibile P . Basta ora osservare che moltiplicare una matrice G a destra oppure a sinistra per una matrice invertibile non ne modifica il rango. \square

Vediamo ora alcuni esempi concreti di applicazione di quanto precedentemente descritto.

Esercizio 1. Sia V uno spazio vettoriale reale e sia g la forma bilineare simmetrica di matrice

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

rispetto alla base $\{v_1, v_2, v_3\}$ di V . Si determini una base di V rispetto alla quale la matrice di g sia diagonale, con soli elementi 1 e -1 sulla diagonale.

Soluzione. Utilizziamo un analogo del procedimento di Gram–Schmidt. Osserviamo tuttavia che il vettore v_1 è isotropo, pertanto non possiamo prendere come primo vettore v'_1 della nuova base un suo multiplo. Per ovviare a questo inconveniente è sufficiente scambiare tra loro i vettori v_1 e v_2 della base data ed applicare il procedimento di ortogonalizzazione partendo dunque da v_2 . Dato che $g(v_2, v_2) = 3$, bisognerà porre $v'_1 = v_2/\sqrt{3}$. Si ha così $g(v'_1, v'_1) = 1$. Contnuando, poniamo $v''_2 = \alpha v'_1 + v_1$ ed imponiamo che sia $g(v'_1, v''_2) = 0$. Si trova allora $\alpha = -g(v'_1, v_1) = -2/\sqrt{3}$, e quindi $v''_2 = -2/\sqrt{3} v'_1 + v_1 = -\frac{2}{3} v_2 + v_1$. Dato che si ha $g(v''_2, v''_2) = -\frac{4}{3}$, e dato che $\sqrt{-\frac{4}{3}}$ non esiste in \mathbb{R} , dovremo porre $v'_2 = v''_2/\sqrt{\frac{4}{3}} = \frac{\sqrt{3}}{2} v''_2 = -\frac{\sqrt{3}}{3} v_2 + \frac{\sqrt{3}}{2} v_1$, ottenendo però $g(v'_2, v'_2) = -1$.

Per trovare il terzo e ultimo vettore della base, poniamo $v''_3 = \alpha v'_1 + \beta v'_2 + v_3$ ed imponiamo che $g(v'_1, v''_3) = 0$ e $g(v'_2, v''_3) = 0$. Si ottiene così $\alpha = -g(v'_1, v_3) = 1/\sqrt{3}$ e $\beta = g(v'_2, v_3) = 5\sqrt{3}/6$ (attenzione che in questo caso si ha $g(v'_2, v'_2) = -1$). Si ha dunque $v''_3 = \frac{1}{\sqrt{3}} v'_1 + \frac{5\sqrt{3}}{6} v'_2 + v_3 = -\frac{1}{2} v_2 + \frac{5}{4} v_1 + v_3$. Si ha allora $g(v''_3, v''_3) = 15/4$ e quindi si deve porre $v'_3 = \frac{2}{\sqrt{15}} v''_3 = \frac{5}{2\sqrt{15}} v_1 - \frac{1}{\sqrt{15}} v_2 + \frac{2}{\sqrt{15}} v_3$, ottenendo così $g(v'_3, v'_3) = 1$.

La matrice di g nella base $\{v'_1, v'_2, v'_3\}$ è dunque

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

Per terminare possiamo osservare che si ha

$$G' = {}^t PGP,$$

ove la matrice di cambiamento di base P è la matrice le cui colonne sono le componenti dei vettori v'_1 , v'_2 e v'_3 rispetto alla base $\{v_1, v_2, v_3\}$. Si ha dunque

$$P = \begin{pmatrix} 0 & \frac{\sqrt{3}}{2} & \frac{5}{2\sqrt{15}} \\ \frac{1}{\sqrt{3}} & -\frac{\sqrt{3}}{3} & -\frac{1}{\sqrt{15}} \\ 0 & 0 & \frac{2}{\sqrt{15}} \end{pmatrix} = \begin{pmatrix} 0 & \frac{\sqrt{3}}{2} & \frac{\sqrt{15}}{6} \\ \frac{\sqrt{3}}{3} & -\frac{\sqrt{3}}{3} & -\frac{\sqrt{15}}{15} \\ 0 & 0 & \frac{2\sqrt{15}}{15} \end{pmatrix}.$$

Esercizio 2. Sia V uno spazio vettoriale reale di dimensione 3, e indichiamo con $\{v_1, v_2, v_3\}$ una sua base. Si consideri la forma bilineare simmetrica g di matrice

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

rispetto alla base data.

- (i) Si verifichi che g è non degenere e si determini una base ortogonale di V relativamente a g .
- (ii) Si calcoli l'indice d'inerzia $i(g)$.
- (iii) Si dica se esistono vettori isotropi non nulli relativamente a g e, in caso affermativo, si determini un sottospazio isotropo di dimensione massima.

Soluzione. (i) Si ha

$$\det G = -3 \begin{vmatrix} 2 & -1 \\ -1 & -1 \end{vmatrix} - \begin{vmatrix} 1 & -1 \\ 0 & -1 \end{vmatrix} = 10,$$

quindi g è non degenere.

Procediamo quindi alla determinazione di una base ortogonale. Poniamo $w_1 = v_1$; si ha $g(w_1, w_1) = -3$. Si noti che non è necessario normalizzare i vettori della nuova base, perché è richiesta solo una base ortogonale.

Osservando la matrice G si nota che il vettore v_3 è ortogonale a w_1 , quindi come secondo vettore della base ortogonale possiamo prendere $w_2 = v_3$; si ha quindi $g(w_2, w_2) = -1$.

Rimane ora solo da determinare un terzo vettore w_3 ortogonale ai due precedenti. Poniamo $w_3 = \lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3$ ed imponiamo a w_3 di essere ortogonale a w_1 e w_2 . Si ottiene il seguente sistema:

$$\begin{cases} g(w_1, w_3) = -3\lambda_1 + \lambda_2 = 0 \\ g(w_2, w_3) = -\lambda_2 - \lambda_3 = 0, \end{cases}$$

da cui si ricava

$$\begin{cases} \lambda_1 = -\frac{1}{3}\lambda_3 \\ \lambda_2 = -\lambda_3. \end{cases}$$

Ponendo, ad esempio, $\lambda_3 = 3$, si ottiene il vettore $w_3 = -v_1 - 3v_2 + 3v_3$. Si trova ora $g(w_3, w_3) = 30$, da cui si deduce che la matrice di g rispetto alla base ortogonale $\{w_1, w_2, w_3\}$ è

$$\begin{pmatrix} -3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 30 \end{pmatrix}.$$

(ii) Osservando la matrice appena trovata si scopre che la restrizione di g al sottospazio $\langle w_1, w_2 \rangle$, generato da w_1 e w_2 , è definita negativa, mentre la restrizione di g al sottospazio generato da w_3 è definita positiva. Si deduce quindi che l'indice d'inerzia di g è dato da $i(g) = 1 - 2 = -1$.

(iii) Dato che g è indefinita, esistono sicuramente dei vettori isotropi non nulli. Ad esempio, se consideriamo un vettore $w = \lambda_1 w_1 + \lambda_2 w_2 + \lambda_3 w_3$, la condizione che w sia isotropo è

$$g(w, w) = -3\lambda_1^2 - \lambda_2^2 + 30\lambda_3^2 = 0,$$

che ha ovviamente soluzioni reali non nulle.

Sia dunque U un sottospazio isotropo; si ha quindi $U \subset U^\perp$, e $\dim U^\perp = 3 - \dim U$. Da ciò si deduce che U può avere al più dimensione 1. Quindi per determinare un sottospazio isotropo di dimensione massima è sufficiente trovare un vettore isotropo non nullo; ad esempio il vettore $w = \sqrt{30} w_2 + w_3$. Il sottospazio $U = \langle w \rangle$ è pertanto un sottospazio isotropo di dimensione massima (ovviamente tale sottospazio non è unico).

Esercizio 3. Sullo spazio vettoriale \mathbb{R}^4 si consideri la forma bilineare simmetrica g di matrice

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

rispetto alla base canonica.

(i) Si determini l'indice d'inerzia di g .

(ii) Si determini, se esiste, una base di \mathbb{R}^4 rispetto a cui g ha matrice

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

(iii) Si determini, se esiste, una base di \mathbb{R}^4 rispetto a cui g ha matrice

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

(iv) Si determini, se esiste, una base di \mathbb{R}^4 rispetto a cui g ha matrice

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

Soluzione. (i) Per determinare l'indice d'inerzia cerchiamo innanzitutto una base ortogonale (dato che sarà utile anche nel seguito). Indichiamo con $\{e_1, \dots, e_4\}$ la base canonica di \mathbb{R}^4 . In questo caso tutti e quattro i vettori della base canonica sono isotropi, quindi non possiamo prendere nessuno di questi come primo vettore di una base ortogonale. Prendiamo allora $v_1 = e_1 + e_2$; questo vettore non è isotropo, dato che si ha $g(v_1, v_1) = 2g(e_1, e_2) = 4$.

Come secondo vettore si vede immediatamente che si può prendere $v_2 = e_1 - e_2$, dato che esso è ortogonale a v_1 : $g(v_1, v_2) = 0$. Si ha poi $g(v_2, v_2) = -4$.

A questo punto possiamo prendere

$$v_3 = \lambda_1 v_1 + \lambda_2 v_2 + e_3,$$

ed imporre che $g(v_1, v_3) = g(v_2, v_3) = 0$. Si trova allora $\lambda_1 = \frac{1}{4}$ e $\lambda_2 = -\frac{3}{4}$. Si ha quindi

$$v_3 = \frac{1}{4}v_1 - \frac{3}{4}v_2 + e_3 = -\frac{1}{2}e_1 + e_2 + e_3,$$

da cui si ottiene $g(v_3, v_3) = 2$. Analogamente, ponendo

$$v_4 = \lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 + e_4,$$

ed imponendo che $g(v_1, v_4) = g(v_2, v_4) = g(v_3, v_4) = 0$, si trova $\lambda_1 = \frac{3}{4}$, $\lambda_2 = \frac{3}{4}$ e $\lambda_3 = 2$, da cui si ottiene

$$v_4 = \frac{3}{4}v_1 + \frac{3}{4}v_2 + 2v_3 + e_4 = \frac{1}{2}e_1 + 2e_2 + 2e_3 + e_4,$$

e $g(v_4, v_4) = -8$. La matrice di g rispetto alla base $\{v_1, \dots, v_4\}$ è quindi

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

da cui si deduce che l'indice d'inerzia di g è $i(g) = 2 - 2 = 0$.

(ii) Dato che, come si verifica immediatamente, la matrice $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ ha indice d'inerzia 0, si deduce che la matrice

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

ha indice d'inerzia $2+0 = 2$, che è diverso dall'indice d'inerzia di g . Dal teorema di Sylvester segue allora che non esiste una base di \mathbb{R}^4 rispetto a cui g abbia la matrice indicata.

(iii) Da quanto detto sopra segue che la matrice

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

ha indice d'inerzia $0+0 = 0$, che coincide con l'indice d'inerzia di g . Il teorema di Sylvester afferma allora che esiste una base di \mathbb{R}^4 rispetto a cui g abbia la

matrice indicata. Ricordando la matrice di g rispetto alla base $\{v_1, \dots, v_4\}$ trovata in precedenza, è facile vedere che ponendo, ad esempio, $w_1 = v_1 + v_2$, $w_2 = \frac{1}{4}(v_1 - v_2)$, $w_3 = v_3 + \frac{1}{2}v_4$ e $w_4 = \frac{1}{4}(v_3 - \frac{1}{2}v_4)$, si ottiene una base rispetto alla quale la matrice di g coincide con la matrice in questione.

(iv) Da quanto visto in precedenza segue che la matrice

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

ha indice d'inerzia $0 - 2 = -2$, che è diverso dall'indice d'inerzia di g . Dal teorema di Sylvester segue allora che non esiste una base di \mathbb{R}^4 rispetto a cui g abbia la matrice indicata.

5.6 Isometrie

Siano (V, g) e (W, h) due spazi vettoriali sul campo K dotati di due forme bilineari simmetriche

$$g : V \times V \rightarrow K, \quad h : W \times W \rightarrow K.$$

Definizione 5.6.1. Una funzione lineare $f : V \rightarrow W$ è *compatibile* con le forme bilineari simmetriche g e h se, per ogni $v_1, v_2 \in V$, si ha

$$h(f(v_1), f(v_2)) = g(v_1, v_2). \quad (5.6.1)$$

In tal caso diremo anche che $f : (V, g) \rightarrow (W, h)$ è un *omomorfismo di spazi vettoriali dotati di forme bilineari simmetriche*.

La condizione di compatibilità (5.6.1) determina l'esistenza di alcune relazioni tra i nuclei delle funzioni f , g e h . Più precisamente, si ha:

Lemma 5.6.2. *Sia $f : (V, g) \rightarrow (W, h)$ un omomorfismo di spazi vettoriali dotati di forme bilineari simmetriche. Valgono le seguenti inclusioni:*

$$\text{Ker } f \subseteq f^{-1}(\text{Ker } h) \subseteq \text{Ker } g. \quad (5.6.2)$$

Dimostrazione. La prima inclusione è ovvia: se $v \in \text{Ker } f$ è $f(v) = \mathbf{0}$, pertanto $h(f(v), w) = 0$ per ogni $w \in W$, quindi $f(v) \in \text{Ker } h$.

Dimostriamo ora la seconda inclusione. Per ogni $v \in f^{-1}(\text{Ker } h)$ e ogni $v' \in V$, si ha $g(v, v') = h(f(v), f(v')) = 0$, dato che $f(v) \in \text{Ker } h$. Si conclude pertanto che $v \in \text{Ker } g$. \square

Da questo risultato discendono alcune interessanti conseguenze:

Corollario 5.6.3. *Sia $f : (V, g) \rightarrow (W, h)$ un omomorfismo di spazi vettoriali dotati di forme bilineari simmetriche. Allora:*

- (i) se h è non degenere, $\text{Ker } f = f^{-1}(\text{Ker } h)$;
- (ii) se g è non degenere, si ha $\text{Ker } f = f^{-1}(\text{Ker } h) = \{\mathbf{0}\}$. In particolare, f è iniettiva e la restrizione di h a $\text{Im}(f) \times \text{Im}(f)$ è non degenere;

- (iii) se g è non degenere e f è suriettiva, allora anche h è non degenere. Inoltre, in tal caso, f è un isomorfismo di spazi vettoriali e $f^{-1} : (W, h) \rightarrow (V, g)$ è compatibile con le forme bilineari h e g .

Dimostrazione. Per dimostrare la proprietà (i) basta ricordare che h è non degenere se e solo se $\text{Ker } h = \{\mathbf{0}\}$.

(ii) Se g è non degenere, cioè se $\text{Ker } g = \{\mathbf{0}\}$, la (5.6.2) si riduce a $\text{Ker } f = f^{-1}(\text{Ker } h) = \{\mathbf{0}\}$, quindi f è iniettiva. Indicando con \tilde{h} la restrizione di h a $\text{Im}(f) \times \text{Im}(f)$, per ogni $w_1, w_2 \in \text{Im}(f)$ si ha $\tilde{h}(w_1, w_2) = g(v_1, v_2)$, ove $v_1, v_2 \in V$ sono tali che $w_1 = f(v_1)$ e $w_2 = f(v_2)$. Dato che g è non degenere si conclude quindi che anche \tilde{h} è non degenere.

Dimostriamo infine la proprietà (iii). Se f è suriettiva, si ha $\text{Im } f = W$ quindi, poiché g è non degenere, dalla (ii) segue che h è non degenere e che f è iniettiva. La funzione $f : V \rightarrow W$ è quindi biiettiva e pertanto è un isomorfismo di spazi vettoriali. Rimane solo da dimostrare che la sua inversa $f^{-1} : W \rightarrow V$ è compatibile con le forme bilineari h e g . Siano dunque $v_1, v_2 \in V$ e poniamo $w_1 = f(v_1)$ e $w_2 = f(v_2)$. Si ha $v_1 = f^{-1}(w_1)$, $v_2 = f^{-1}(w_2)$ e quindi

$$g(f^{-1}(w_1), f^{-1}(w_2)) = g(v_1, v_2) = h(f(v_1), f(v_2)) = h(w_1, w_2).$$

Ciò dimostra che anche f^{-1} è un omomorfismo di spazi vettoriali dotati di forme bilineari simmetriche. \square

Notiamo anche che, sotto opportune ipotesi, la condizione di compatibilità (5.6.1) implica la linearità di f .

Proposizione 5.6.4. *Siano (V, g) e (W, h) due spazi vettoriali dotati di forme bilineari simmetriche e sia $f : V \rightarrow W$ una funzione qualunque (in particolare, non stiamo supponendo che f sia lineare) che soddisfa l'uguaglianza (5.6.1), per ogni $v_1, v_2 \in V$. Se h è non degenere e f è suriettiva, allora f è lineare.*

Dimostrazione. Siano $v_1, v_2 \in V$ e $\lambda_1, \lambda_2 \in K$. Vogliamo dimostrare che

$$f(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 f(v_1) + \lambda_2 f(v_2).$$

Per ogni $v \in V$, si ha:

$$\begin{aligned} h(f(\lambda_1 v_1 + \lambda_2 v_2) - \lambda_1 f(v_1) - \lambda_2 f(v_2), f(v)) &= \\ &= h(f(\lambda_1 v_1 + \lambda_2 v_2), f(v)) - \lambda_1 h(f(v_1), f(v)) - \lambda_2 h(f(v_2), f(v)) \\ &= g(\lambda_1 v_1 + \lambda_2 v_2, v) - \lambda_1 g(v_1, v) - \lambda_2 g(v_2, v) = 0, \end{aligned}$$

per la bilinearità di h e g . Poiché f è suriettiva e h è non degenere, da ciò segue che

$$f(\lambda_1 v_1 + \lambda_2 v_2) - \lambda_1 f(v_1) - \lambda_2 f(v_2) = \mathbf{0},$$

il che dimostra la linearità di f . \square

Vediamo ora come si esprime in termini di matrici la condizione di compatibilità (5.6.1).

Proposizione 5.6.5. *Siano (V, g) e (W, h) due spazi vettoriali dotati di forme bilineari simmetriche e sia $f : V \rightarrow W$ una funzione lineare. Siano $\mathbf{v} = \{v_1, \dots, v_n\}$ e $\mathbf{w} = \{w_1, \dots, w_m\}$ basi di V e W rispettivamente. Indichiamo*

con A la matrice di f rispetto alle basi \mathbf{v} e \mathbf{w} , con G la matrice di g rispetto alla base \mathbf{v} e con H la matrice di h rispetto alla base \mathbf{w} . Allora f è compatibile con le forme bilineari g e h se e solo se

$$G = {}^t A H A. \quad (5.6.3)$$

Dimostrazione. Dalla definizione di matrice associata a un'applicazione lineare segue che, per ogni vettore v_i della base \mathbf{v} , si ha

$$f(v_i) = \sum_{r=1}^m a_{ri} w_r.$$

Per ogni $i, j = 1, \dots, n$, la condizione di compatibilità (5.6.1) equivale a:

$$\begin{aligned} g_{ij} &= g(v_i, v_j) \\ &= h(f(v_i), f(v_j)) \\ &= h\left(\sum_{r=1}^m a_{ri} w_r, \sum_{s=1}^m a_{sj} w_s\right) \\ &= \sum_{r,s=1}^m a_{ri} a_{sj} h(w_r, w_s) \\ &= \sum_{r,s=1}^m a_{ri} a_{sj} h_{rs}. \end{aligned}$$

Questa uguaglianza, riscritta in termini di matrici, non è altro che la (5.6.3). \square

Specializzando questo risultato al caso in cui $(W, h) = (V, g)$, si ottiene:

Corollario 5.6.6. *Sia V uno spazio vettoriale sul campo K e sia g una forma bilineare simmetrica definita su V . Indichiamo con G la matrice di g rispetto a una base $\mathbf{v} = \{v_1, \dots, v_n\}$ di V . Un endomorfismo f di V , di matrice A rispetto alla base \mathbf{v} , è compatibile con la forma bilineare g se e solo se A soddisfa la seguente uguaglianza:*

$$G = {}^t A G A.$$

Restringiamo ora la nostra attenzione al caso di spazi vettoriali dotati di forme bilineari simmetriche non degeneri.

Definizione 5.6.7. Siano (V, g) e (W, h) due spazi vettoriali dotati di forme bilineari simmetriche non degeneri. Una funzione lineare $f : V \rightarrow W$ compatibile con le forme bilineari g e h è detta una *isometria* di V in W .

Osserviamo che, per il Corollario 5.6.3, f è necessariamente iniettiva, pertanto determina un isomorfismo di V con il sottospazio $\text{Im}(f)$ di W .

Definizione 5.6.8. Due spazi vettoriali dotati di forme bilineari simmetriche non degeneri (V, g) e (W, h) si dicono *isometrici* se esiste un'isometria di V su tutto W , cioè se esiste un isomorfismo $f : V \xrightarrow{\sim} W$ compatibile con le forme bilineari g e h .

Osservazione 5.6.9. Se $f : (V, g) \rightarrow (W, h)$ è un'isometria di V in W , lo spazio vettoriale V è isometrico al sottospazio vettoriale $\text{Im}(f)$ di W dotato della forma bilineare simmetrica non degenera indotta, per restrizione, da h .

Osservazione 5.6.10. Nel caso degli spazi vettoriali euclidei (cioè degli spazi vettoriali reali dotati di forme bilineari simmetriche definite positive) le isometrie preservano le lunghezze dei vettori e gli angoli tra vettori. Consideriamo infatti due spazi vettoriali euclidei (V, g) e (W, h) e un'isometria $f : V \rightarrow W$. Per ogni $v \in V$, si ha:

$$\|f(v)\| = \sqrt{h(f(v), f(v))} = \sqrt{g(v, v)} = \|v\|.$$

Inoltre, se indichiamo con ϕ l'angolo compreso tra due vettori $v_1, v_2 \in V$ e con ϕ' l'angolo compreso tra le loro immagini $f(v_1), f(v_2) \in W$, si ha

$$\cos \phi' = \frac{h(f(v_1), f(v_2))}{\|f(v_1)\| \|f(v_2)\|} = \frac{g(v_1, v_2)}{\|v_1\| \|v_2\|} = \cos \phi.$$

È facile dimostrare che una qualunque funzione lineare $f : V \rightarrow W$ che preservi la norma dei vettori è un'isometria. Infatti, dati $v_1, v_2 \in V$, si ha

$$\begin{aligned} \|v_1 + v_2\|^2 &= g(v_1 + v_2, v_1 + v_2) \\ &= g(v_1, v_1) + 2g(v_1, v_2) + g(v_2, v_2) \\ &= \|v_1\|^2 + 2g(v_1, v_2) + \|v_2\|^2. \end{aligned}$$

Analogamente,

$$\begin{aligned} \|f(v_1 + v_2)\|^2 &= \|f(v_1) + f(v_2)\|^2 \\ &= h(f(v_1) + f(v_2), f(v_1) + f(v_2)) \\ &= h(f(v_1), f(v_1)) + 2h(f(v_1), f(v_2)) + h(f(v_2), f(v_2)) \\ &= \|f(v_1)\|^2 + 2h(f(v_1), f(v_2)) + \|f(v_2)\|^2. \end{aligned}$$

Poiché f , per ipotesi, preserva la norma dei vettori, si ha

$$\|f(v_1)\| = \|v_1\|, \quad \|f(v_2)\| = \|v_2\|, \quad \|f(v_1 + v_2)\| = \|v_1 + v_2\|.$$

Da ciò segue che si ha anche $h(f(v_1), f(v_2)) = g(v_1, v_2)$, quindi f è un'isometria.

Abbiamo già osservato, nel Corollario 5.6.3, che se un'isometria $f : V \rightarrow W$ è biettiva, allora anche la sua inversa $f^{-1} : W \rightarrow V$ è un'isometria. Inoltre, è immediato verificare che la composizione di due isometrie è anch'essa un'isometria. Da queste osservazioni segue che l'insieme $O(V, g)$ delle isometrie $f : V \rightarrow V$ di uno spazio vettoriale V dotato di una forma bilineare simmetrica non degenere g , rispetto alla legge di composizione, è un gruppo. Tale gruppo, detto il *gruppo delle isometrie* o *gruppo ortogonale* di (V, g) , è un sottogruppo del gruppo $GL(V)$ di tutti gli automorfismi di V .

Consideriamo ora uno spazio vettoriale V dotato di una forma bilineare simmetrica non degenere g , fissiamo una base $\mathbf{v} = \{v_1, \dots, v_n\}$ di V e indichiamo con G la matrice di g rispetto alla base \mathbf{v} . Come abbiamo visto, un endomorfismo $f : V \rightarrow V$ è un'isometria se e solo se la sua matrice A , rispetto alla base \mathbf{v} , soddisfa l'uguaglianza ${}^t A G A = G$. In termini di matrici, il gruppo delle isometrie di (V, g) ha quindi la seguente espressione:

$$O(G) = \{A \in GL(n) \mid {}^t A G A = G\}.$$

Dal Teorema di Binet segue che, per ogni $A \in O(G)$, si ha

$$\det(G) = \det({}^t A G A) = \det(G) \det(A)^2.$$

Poiché g è non degenere si ha $\det(G) \neq 0$, quindi deve essere $\det(A)^2 = 1$, cioè $\det(A) = \pm 1$. Si verifica facilmente che le isometrie con determinante 1 formano un sottogruppo (di indice 2) del gruppo di tutte le isometrie di (V, g) . Tale gruppo, indicato con $\mathrm{SO}(V, g)$, è detto il *gruppo delle isometrie dirette*, o anche il *gruppo speciale ortogonale*, dello spazio vettoriale V dotato della forma bilineare simmetrica g . In termini di matrici tale gruppo ha la seguente descrizione:

$$\mathrm{SO}(G) = \{A \in \mathrm{GL}(n) \mid {}^t A G A = G, \det(A) = 1\}.$$

Se (V, g) è uno spazio vettoriale euclideo, il Teorema 5.4.14 garantisce l'esistenza di una base ortonormale $\mathbf{v} = \{v_1, \dots, v_n\}$ di V . Rispetto a tale base la matrice di g è la matrice identica, pertanto le matrici delle isometrie di (V, g) sono caratterizzate dalla seguente equazione:

$${}^t A A = \mathbf{1}.$$

Il gruppo costituito da tali matrici è il *gruppo ortogonale di ordine n*

$$\mathrm{O}(n) = \{A \in \mathrm{GL}(n) \mid {}^t A A = \mathbf{1}\}.$$

Una matrice $A \in \mathrm{O}(n)$, cioè una matrice quadrata A , di ordine n , tale che ${}^t A A = \mathbf{1}$ o, equivalentemente, tale che ${}^t A = A^{-1}$, è detta una *matrice ortogonale*. Indicando con $A_{(1)}, \dots, A_{(n)}$ le colonne di A , l'equazione ${}^t A A = \mathbf{1}$ equivale a

$$A_{(i)} \cdot A_{(j)} = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j, \end{cases}$$

il che significa che le colonne di A formano una base ortonormale di \mathbb{R}^n , rispetto al prodotto scalare usuale.

5.6.1 Isometrie di \mathbb{R}^2

Studiamo ora in dettaglio la struttura delle isometrie dello spazio vettoriale euclideo \mathbb{R}^2 , dotato del prodotto scalare usuale. Il gruppo di tali isometrie è il gruppo ortogonale di ordine 2

$$\mathrm{O}(2) = \{A \in \mathrm{GL}(2) \mid {}^t A A = \mathbf{1}\}.$$

Data una matrice

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

la condizione ${}^t A A = \mathbf{1}$ equivale al seguente sistema:

$$\begin{cases} a^2 + c^2 = 1 \\ b^2 + d^2 = 1 \\ ab + cd = 0. \end{cases}$$

Dall'equazione $a^2 + c^2 = 1$ si deduce che esiste un angolo α tale che

$$a = \cos \alpha, \quad c = \sin \alpha.$$

Analogamente, da $b^2 + d^2 = 1$ si deduce che esiste un angolo β tale che

$$b = \sin \beta, \quad d = \cos \beta.$$

La terza equazione, $ab + cd = 0$, diventa allora

$$\cos \alpha \sin \beta + \sin \alpha \cos \beta = \sin(\alpha + \beta) = 0,$$

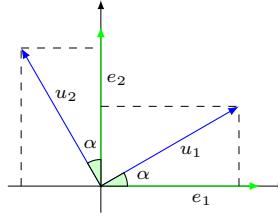
da cui si ottiene $\alpha + \beta = k\pi$, cioè $\beta = k\pi - \alpha$, con $k \in \mathbb{Z}$. Si ottengono così due diversi tipi di isometrie, a seconda che k sia pari o dispari. Più precisamente, se k è pari la matrice A ha la forma seguente:

$$\begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}. \quad (5.6.4)$$

Il determinante di questa matrice è pari a 1; si tratta quindi di una isometria diretta. In termini geometrici, questa è la matrice di una rotazione del piano \mathbb{R}^2 , attorno all'origine, di un angolo α in senso antiorario. Le due colonne

$$u_1 = \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix} \quad u_2 = \begin{pmatrix} -\sin \alpha \\ \cos \alpha \end{pmatrix}$$

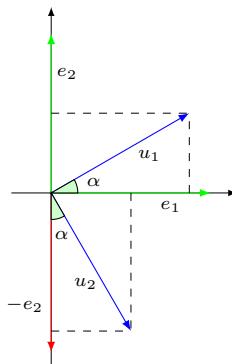
della matrice (5.6.4) sono infatti i trasformati dei vettori e_1 ed e_2 della base canonica di \mathbb{R}^2 mediante la suddetta rotazione, come indicato nella figura seguente:



Se invece k è dispari, si ottengono matrici del tipo seguente:

$$\begin{pmatrix} \cos \alpha & \sin \alpha \\ \sin \alpha & -\cos \alpha \end{pmatrix} \quad (5.6.5)$$

Queste matrici hanno determinante pari a -1 e rappresentano pertanto delle isometrie inverse. Geometricamente esse corrispondono a una riflessione del piano rispetto all'asse delle ascisse, seguita da una rotazione, attorno all'origine, di un angolo α in senso antiorario. La situazione è illustrata nella figura seguente:



Come si vede facilmente da questa figura, i vettori u_1 e u_2 , immagini dei vettori e_1 ed e_2 della base canonica, hanno le seguenti coordinate:

$$u_1 = \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix} \quad u_2 = \begin{pmatrix} \sin \alpha \\ -\cos \alpha \end{pmatrix}$$

Essi coincidono quindi con le colonne della matrice (5.6.5).

Esercizi

Esercizio 5.1. Sia V uno spazio vettoriale sul campo complesso \mathbb{C} con base $\{v_1, v_2, v_3\}$ e sia g la forma bilineare simmetrica di matrice, rispetto alla base data,

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

Si determini una matrice invertibile P tale che ${}^tPGP = \mathbf{1}$.

Esercizio 5.2. Sia V uno spazio vettoriale sul campo dei numeri reali, con base $\{v_1, v_2, v_3\}$, e sia g la forma bilineare simmetrica di matrice, rispetto alla base data,

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

Si determini una base di V rispetto alla quale la matrice di g sia diagonale, con soli elementi 1 e -1 sulla diagonale.

Esercizio 5.3. Sia V uno spazio vettoriale sul campo dei numeri reali, con base $\{v_1, v_2, v_3\}$, e sia g la forma bilineare simmetrica di matrice, rispetto alla base data,

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

Si determini una base di V rispetto alla quale la matrice di g sia diagonale, con soli elementi 1 e -1 sulla diagonale.

Esercizio 5.4. Sia V uno spazio vettoriale reale di dimensione 3, e sia $\{v_1, v_2, v_3\}$ una sua base. Si consideri la forma bilineare simmetrica g di matrice

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

rispetto alla base data.

- (1) Si verifichi che g è non-degenera e si determini una base ortogonale di V relativamente a g .
- (2) Si calcoli l'indice d'inerzia $i(g)$.
- (3) Si dica se esistono vettori isotropi non nulli relativamente a g e, in caso affermativo, si determini un sottospazio isotropo di dimensione massima.

Esercizio 5.5. Sia $V = \mathbb{R}^4$ e sia g la forma bilineare simmetrica di matrice

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

rispetto alla base canonica.

- (1) Si verifichi che g è non-degenere, se ne calcoli l'indice d'inerzia e si determini una base ortogonale di V relativamente a g .
- (2) (\mathbb{R}^4, g) è isometrico allo spazio \mathbb{R}^4 dotato del prodotto scalare usuale?

Esercizio 5.6. Sia V uno spazio vettoriale reale di dimensione finita e sia $g : V \times V \rightarrow \mathbb{R}$ una forma bilineare simmetrica definita positiva. Se $\phi : V \rightarrow V$ è una isometria, si dimostri che si ha

$$\text{Im}(\phi - \text{id}) = \text{Ker}(\phi - \text{id})^\perp.$$

Esercizio 5.7. Si consideri \mathbb{C} come spazio vettoriale su \mathbb{R} e si ponga su di esso l'applicazione bilineare $g : \mathbb{C} \times \mathbb{C} \rightarrow \mathbb{R}$ definita da

$$g(z_1, z_2) = 2\Re(z_1 z_2),$$

ove $\Re(z)$ indica la parte reale del numero complesso z .

- (1) Si dimostri che g è bilineare e non-degenere.
- (2) Si determini la matrice di g rispetto alla base $\{1, i\}$ di \mathbb{C} .
- (3) Siano $z_1, z_2 \in \mathbb{C}$, linearmente indipendenti su \mathbb{R} (cioè z_1 e z_2 sono una base di \mathbb{C} , visto come \mathbb{R} -spazio vettoriale). Si determini la matrice G di g rispetto alla base $\{z_1, z_2\}$ e si dimostri che

$$\det G = \left[\det \begin{pmatrix} z_1 & z_2 \\ \bar{z}_1 & \bar{z}_2 \end{pmatrix} \right]^2,$$

ove \bar{z} indica il numero complesso coniugato di $z \in \mathbb{C}$.

Esercizio 5.8. Sia A una matrice simmetrica ad elementi in \mathbb{R} . È vero o falso che la matrice $\mathbf{1} + A^2$ è invertibile?

Capitolo 6

Geometria Affine

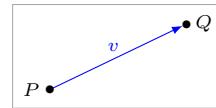
In questo capitolo vedremo come l'algebra lineare possa essere utilizzata per lo studio di problemi geometrici. Partendo dalla nozione di vettore e di spazio vettoriale, introdurremo il concetto di spazio affine come ambiente naturale per lo studio della Geometria Euclidea. Studieremo poi le principali proprietà geometriche di tali spazi e dei loro sottospazi.

6.1 Spazi affini

Il concetto di spazio vettoriale, che è alla base degli sviluppi dell'algebra lineare, si rivela inadeguato a fornire un modello di "spazio" per lo studio della geometria euclidea.¹ Come abbiamo visto nel Capitolo 1, i vettori sono stati originariamente introdotti per rappresentare degli spostamenti; è quindi necessario disporre di uno "spazio" (cioè di un insieme di punti) del quale i vettori possano rappresentare le traslazioni.

Queste semplici considerazioni ci portano a concludere che un modello di spazio adatto allo studio della geometria dovrà consistere di un insieme di punti e di un insieme di vettori, cioè di uno spazio vettoriale, il quale dovrà rappresentare l'insieme delle traslazioni di tali punti.

Due modi equivalenti di esprimere questo fatto sono i seguenti: (1) stabilire che una coppia ordinata di punti P e Q determina un vettore v (rappresentato graficamente come un segmento orientato che ha l'origine nel punto P e l'estremità della freccia in Q), oppure (2) stabilire che un punto P e un vettore v determinano un altro punto Q (che rappresenta l'effetto della traslazione indicata dal vettore v , applicata al punto P).



¹Un esempio di questa inadeguatezza è dato dal fatto che in ogni spazio vettoriale esiste un elemento privilegiato, il vettore nullo $\mathbf{0}$. Al contrario, nello spazio della Geometria Euclidea, non esiste alcun punto privilegiato O che si possa far corrispondere al vettore nullo.

Da questa seconda interpretazione segue che il nostro spazio dovrà essere dotato di una legge di composizione, che indicheremo con il simbolo $+$, la quale associa alla coppia punto-vettore (P, v) il punto Q ; scriveremo dunque $Q = P + v$.

Possiamo ora formalizzare il concetto di spazio affine seguendo l'idea appena descritta.

Definizione 6.1.1. Uno *spazio affine* \mathbb{A} sul campo K è il dato di un insieme non vuoto \mathcal{A} , detto l'*insieme dei punti* di \mathbb{A} , di uno spazio vettoriale V su K e di un'operazione

$$+_{\mathbb{A}} : \mathcal{A} \times V \rightarrow \mathcal{A}, \quad (P, v) \mapsto Q = P +_{\mathbb{A}} v,$$

che soddisfa le seguenti proprietà:

- (i) $(P +_{\mathbb{A}} v_1) +_{\mathbb{A}} v_2 = P +_{\mathbb{A}} (v_1 + v_2)$, per ogni $P \in \mathcal{A}$ e ogni $v_1, v_2 \in V$;
- (ii) $P +_{\mathbb{A}} \mathbf{0} = P$, per ogni $P \in \mathcal{A}$;
- (iii) per ogni $P, Q \in \mathcal{A}$ esiste un unico vettore $v \in V$ tale che $Q = P +_{\mathbb{A}} v$.

Lo spazio vettoriale V è detto lo *spazio vettoriale soggiacente* allo spazio affine \mathbb{A} (o lo spazio *direttore* di \mathbb{A}) e la dimensione di V è detta la *dimensione* dello spazio affine \mathbb{A} . Uno spazio affine è detto di *dimensione finita* se lo spazio vettoriale soggiacente è finitamente generato.

Nel seguito uno spazio affine $\mathbb{A} = (\mathcal{A}, V, +_{\mathbb{A}})$ verrà spesso identificato con il suo insieme di punti \mathcal{A} . Per indicare che P è un punto di \mathbb{A} scriveremo dunque $P \in \mathbb{A}$ al posto di $P \in \mathcal{A}$. Analogamente, per indicare che v è un vettore appartenente allo spazio vettoriale V soggiacente allo spazio affine \mathbb{A} , potremo scrivere $v \in \mathbb{A}$. Il significato di tali espressioni sarà sempre chiaro dal contesto.

Osservazione 6.1.2. L'operazione di “somma tra punti e vettori” in uno spazio affine $\mathbb{A} = (\mathcal{A}, V, +_{\mathbb{A}})$ è stata indicata con il simbolo $+_{\mathbb{A}}$ al fine di distinguerla dall'operazione di somma tra vettori dello spazio vettoriale V , indicata semplicemente con $+$. Tuttavia, poiché sarà sempre chiaro dal contesto di quale delle due operazioni di somma si tratti, nel seguito indicheremo con il simbolo $+$ sia l'operazione di somma tra vettori che l'operazione di somma tra un punto e un vettore.

Osservazione 6.1.3. Per ogni coppia di punti P e Q di uno spazio affine \mathbb{A} , la proprietà (iii) della Definizione 6.1.1 garantisce l'esistenza e l'unicità di un vettore v tale che $Q = P + v$; tale vettore verrà indicato con la notazione \overrightarrow{PQ} . Ricavando formalmente v dall'espressione $Q = P + v$, scriveremo anche

$$v = \overrightarrow{PQ} = Q - P,$$

in modo da avere l'identità $P + (Q - P) = Q$. Come già menzionato, il vettore $\overrightarrow{PQ} = Q - P$ verrà rappresentato graficamente come un segmento orientato avente l'origine in P e l'estremità della freccia nel punto Q .

Osservazione 6.1.4. In uno spazio affine $\mathbb{A} = (\mathcal{A}, V, +_{\mathbb{A}})$ il gruppo additivo dello spazio vettoriale V agisce sull'insieme \mathcal{A} dei punti di \mathbb{A} ; tale azione è libera e transitiva. È facile verificare che uno spazio affine può essere definito, in modo del tutto equivalente, come un insieme non vuoto dotato di un'azione libera e transitiva del gruppo additivo di uno spazio vettoriale.

Per ogni punto P di uno spazio affine $\mathbb{A} = (\mathcal{A}, V, +)$ possiamo definire la funzione

$$f_P : V \rightarrow \mathcal{A}, \quad v \mapsto f_P(v) = P + v.$$

Una conseguenza immediata della Definizione 6.1.1 è il seguente risultato:

Proposizione 6.1.5. *Per ogni punto P di uno spazio affine \mathbb{A} , la funzione $f_P : V \rightarrow \mathcal{A}$ è biiettiva.*

Dimostrazione. Siano $v_1, v_2 \in V$ due vettori tali che $f_P(v_1) = f_P(v_2)$. Si ha dunque $P + v_1 = P + v_2$ da cui, sommando ad ambo i membri il vettore $-v_2$, si ottiene

$$(P + v_1) + (-v_2) = (P + v_2) + (-v_2) = P + (v_2 - v_2) = P + \mathbf{0} = P.$$

Poiché $(P + v_1) + (-v_2) = P + (v_1 - v_2)$, dall'uguaglianza $P + (v_1 - v_2) = P$ e dalle proprietà (ii) e (iii) della definizione di spazio affine, segue che $v_1 - v_2 = \mathbf{0}$ e quindi $v_1 = v_2$; f_P è dunque una funzione iniettiva. La suriettività di f_P è una conseguenza immediata della proprietà (iii) della Definizione 6.1.1. \square

Osservazione 6.1.6. L'esistenza di una biiezione tra l'insieme dei punti \mathcal{A} e lo spazio vettoriale V di uno spazio affine $\mathbb{A} = (\mathcal{A}, V, +)$ permette di concludere che \mathcal{A} può essere identificato con V . Si noti tuttavia che non esiste alcuna biiezione canonica tra V e \mathcal{A} ; una tale biiezione dipende infatti dalla scelta di un punto P di \mathcal{A} . Osserviamo inoltre che nella biiezione $f_P : V \rightarrow \mathcal{A}$ il punto P corrisponde al vettore nullo di V . Possiamo quindi concludere che, in ogni spazio affine \mathbb{A} , l'insieme dei punti può essere identificato con lo spazio vettoriale soggiacente solo dopo aver fissato (arbitrariamente) un punto di \mathbb{A} .

Osservazione 6.1.7. Dato uno spazio affine $\mathbb{A} = (\mathcal{A}, V, +)$, per ogni vettore $v \in V$ la funzione

$$\tau_v : \mathcal{A} \rightarrow \mathcal{A}, \quad P \mapsto P + v,$$

è detta la *traslazione* parallela al vettore v . Se v è il vettore nullo, τ_0 è l'identità, mentre per ogni $v \neq \mathbf{0}$, la traslazione τ_v è una biiezione priva di punti fissi, cioè $\tau_v(P) \neq P$, per ogni $P \in \mathcal{A}$ e ogni $v \neq \mathbf{0}$.

Esempio 6.1.8. Mostreremo ora come ogni spazio vettoriale possieda una struttura canonica di spazio affine.

Dato uno spazio vettoriale V sul campo K , definiamo lo spazio affine $\mathbb{V} = (\mathcal{V}, V, +_{\mathbb{V}})$ ad esso associato ponendo $\mathcal{V} = V$. In questo modo l'operazione $+_{\mathbb{V}}$ coincide con l'operazione di somma tra elementi di V :

$$+_{\mathbb{V}} : \mathcal{V} \times V \rightarrow \mathcal{V}, \quad (P, v) \mapsto P +_{\mathbb{V}} v = P +_V v.$$

È facile verificare che, con queste definizioni, \mathbb{V} risulta essere uno spazio affine sul campo K , avente V come spazio vettoriale soggiacente. Notiamo che questa struttura di spazio affine dipende esclusivamente dalla struttura di spazio vettoriale di V : essa è detta pertanto la *struttura affine canonica* dello spazio vettoriale V .

Esempio 6.1.9. L'esempio fondamentale di spazio affine è fornito dallo *spazio affine n -dimensionale standard sul campo K* , che indicheremo con \mathbb{A}_K^n . Si tratta dello spazio affine associato, in modo canonico, allo spazio vettoriale K^n , come descritto nell'esempio precedente. \mathbb{A}_K^n è dunque lo spazio affine il cui

insieme di punti è l'*insieme* K^n e il cui spazio direttore è lo *spazio vettoriale* K^n . L'operazione di somma tra punti e vettori è definita come l'usuale somma componente per componente di due n -uple di elementi di K ; più precisamente, se $P = (p_1, p_2, \dots, p_n)$ è un punto e $v = (a_1, a_2, \dots, a_n)$ è un vettore di \mathbb{A}_K^n , si ha

$$Q = P + v = (p_1 + a_1, p_2 + a_2, \dots, p_n + a_n).$$

È immediato verificare che le proprietà (i), (ii) e (iii) della Definizione 6.1.1 sono soddisfatte, quindi \mathbb{A}_K^n è uno spazio affine. Si noti inoltre che, dalla definizione data, segue subito che se $P = (p_1, p_2, \dots, p_n)$ e $Q = (q_1, q_2, \dots, q_n)$ sono due punti di \mathbb{A}_K^n , il vettore $v = Q - P$ è dato da

$$v = (q_1 - p_1, q_2 - p_2, \dots, q_n - p_n).$$

Come vedremo in seguito, questo esempio ha un'importanza particolare: infatti ogni spazio affine di dimensione n sul campo K risulta essere *isomorfo*² (anche se non in modo canonico) allo spazio affine standard \mathbb{A}_K^n .

Osservazione 6.1.10. Nello spazio affine \mathbb{A}_K^n sia i punti che i vettori sono semplicemente delle n -uple di elementi di K . Osserviamo però che mentre la somma di un punto e un vettore, oppure la differenza di due punti, sono operazioni lecite, la somma di due punti, benché algebricamente possibile, non è un'operazione lecita. A tale riguardo facciamo notare che per distinguere le n -uple di elementi di K che rappresentano dei punti da quelle che rappresentano dei vettori è possibile adottare la seguente convenzione: ad ogni n -upla $(a_1, a_2, \dots, a_n) \in K^n$ viene aggiunto un elemento $a_0 \in \{0, 1\}$, con la convenzione che se $a_0 = 0$ la n -upla (a_1, a_2, \dots, a_n) rappresenta un vettore, mentre se $a_0 = 1$ tale n -upla rappresenta un punto dello spazio affine. In altre parole, i vettori di \mathbb{A}_K^n si scrivono nella forma

$$v = (0, a_1, a_2, \dots, a_n),$$

mentre i punti si scrivono nella forma

$$A = (1, a_1, a_2, \dots, a_n).$$

Si noti che tale convenzione è compatibile con la definizione delle operazioni tra punti e vettori di uno spazio affine: una combinazione lineare di vettori è ancora un vettore (la prima componente è 0), la somma di un punto e di un vettore dà come risultato un punto (infatti, se osserviamo la prima componente, si ha $1+0=1$) e la differenza tra due punti è un vettore (nella prima componente si ha $1-1=0$). La somma di due punti, al contrario, non è un'operazione lecita; ciò è evidenziato dal fatto che, sommando due punti, si otterrebbe una $(n+1)$ -upla avente come prima componente $a_0 = 2$, che non è un valore permesso.

6.2 Sottospazi affini

Sia $\mathbb{A} = (\mathcal{A}, V, +_{\mathbb{A}})$ uno spazio affine sul campo K .

Definizione 6.2.1. Un *sottospazio affine* (o *sottovarietà lineare*) di \mathbb{A} è uno spazio affine $\mathbb{B} = (\mathcal{B}, W, +_{\mathbb{B}})$, ove \mathcal{B} è un sottoinsieme di \mathcal{A} , W è un sottospazio vettoriale di V e l'operazione $+_{\mathbb{B}} : \mathcal{B} \times W \rightarrow \mathcal{B}$ è indotta, per restrizione, dall'operazione $+_{\mathbb{A}} : \mathcal{A} \times V \rightarrow \mathcal{A}$ definita nello spazio affine \mathbb{A} .

²La definizione di isomorfismo tra spazi affini verrà data nel Paragrafo 6.6.

La struttura dei sottospazi affini di uno spazio affine \mathbb{A} è ulteriormente precisata dal seguente risultato:

Proposizione 6.2.2. *Sia $\mathbb{B} = (\mathcal{B}, W, +_{\mathbb{B}})$ un sottospazio affine dello spazio affine $\mathbb{A} = (\mathcal{A}, V, +_{\mathbb{A}})$. Per ogni punto $P \in \mathcal{B}$, si ha*

$$\mathcal{B} = P +_{\mathbb{B}} W = \{P +_{\mathbb{B}} w \mid w \in W\}.$$

Dimostrazione. Fissato $P \in \mathcal{B}$, dalla definizione di spazio affine segue che, per ogni vettore $w \in W$, il punto $Q = P +_{\mathbb{B}} w$ appartiene all'insieme \mathcal{B} . Viceversa, dalla proprietà (iii) della Definizione 6.1.1, segue che, per ogni $Q \in \mathcal{B}$, esiste un unico vettore $w \in W$ tale che $Q = P +_{\mathbb{B}} w$. \square

Osservazione 6.2.3. Se $\mathbb{A} = (\mathcal{A}, V, +_{\mathbb{A}})$ è uno spazio affine e W è un sottospazio vettoriale di V , per ogni punto $P \in \mathcal{A}$ si può costruire un sottospazio affine $\mathbb{B} = (\mathcal{B}, W, +_{\mathbb{B}})$ di \mathbb{A} ponendo

$$\mathcal{B} = P +_{\mathbb{A}} W = \{P +_{\mathbb{A}} w \mid w \in W\}$$

e definendo l'operazione $+_{\mathbb{B}}$ come la restrizione dell'operazione di somma tra punti e vettori definita in \mathbb{A} . La verifica che \mathbb{B} è uno spazio affine è immediata. A titolo di esempio verifichiamo che la funzione

$$+_{\mathbb{B}} : \mathcal{B} \times W \rightarrow \mathcal{B}$$

è ben definita. Più precisamente, verifichiamo che, per ogni $A \in \mathcal{B}$ e ogni $w \in W$, si ha $A +_{\mathbb{B}} w \in \mathcal{B}$.

Sia dunque $A \in \mathcal{B}$ e sia $u \in W$ tale che $A = P +_{\mathbb{A}} u$. Allora, per ogni $w \in W$, si ha:

$$A +_{\mathbb{B}} w = A +_{\mathbb{A}} w = (P +_{\mathbb{A}} u) +_{\mathbb{A}} w = P +_{\mathbb{A}} (u + w) \in \mathcal{B},$$

dato che $u + w \in W$, essendo W un sottospazio vettoriale di V .

Il sottospazio affine \mathbb{B} così definito verrà spesso indicato semplicemente con $\mathbb{B} = P + W$: esso è detto il *sottospazio affine di \mathbb{A} passante per il punto P e parallelo al sottospazio W* . Il sottospazio vettoriale W è anche detto la *giacitura* di \mathbb{B} .

Osservazione 6.2.4. Un sottospazio affine di \mathbb{A} di dimensione 0 è semplicemente un punto di \mathbb{A} . Un sottospazio affine di dimensione 1 è detto una *retta*, un sottospazio affine di dimensione 2 è detto un *piano* e, se $\dim \mathbb{A} = n$, un sottospazio affine di \mathbb{A} di dimensione $n - 1$ è detto un *iperpiano*.

Definizione 6.2.5. Sia \mathcal{S} un sottoinsieme non vuoto (dell'insieme dei punti) di uno spazio affine \mathbb{A} . Il più piccolo sottospazio affine di \mathbb{A} contenente \mathcal{S} è detto il *sottospazio affine generato da \mathcal{S}* ed è indicato con $\langle \mathcal{S} \rangle$.

Proposizione 6.2.6. *Sia $\mathbb{A} = (\mathcal{A}, V, +)$ uno spazio affine e \mathcal{S} un sottoinsieme non vuoto di \mathcal{A} . Fissato arbitrariamente un punto $P \in \mathcal{S}$ indichiamo con W il sottospazio vettoriale di V generato dall'insieme dei vettori $v_Q = Q - P$, al variare del punto Q in \mathcal{S} . Allora si ha $\langle \mathcal{S} \rangle = P + W$.*

Dimostrazione. Indichiamo con \mathbb{B} il sottospazio affine $\mathbb{B} = P + W$ di \mathbb{A} . Per ogni $Q \in \mathcal{S}$, si ha $Q = P + (Q - P) = P + v_Q \in \mathbb{B}$, il che dimostra che $\mathcal{S} \subseteq \mathbb{B}$. Per dimostrare che $\langle \mathcal{S} \rangle = \mathbb{B}$ bisogna quindi dimostrare che \mathbb{B} è il più piccolo sottospazio affine di \mathbb{A} contenente \mathcal{S} , cioè che ogni sottospazio affine di \mathbb{A} contenente \mathcal{S} contiene anche \mathbb{B} . Sia dunque $\mathbb{A}' = (\mathcal{A}', V', +)$ un sottospazio affine di \mathbb{A} contenente \mathcal{S} . Per ogni $P, Q \in \mathcal{S}$ si ha anche $P, Q \in \mathbb{A}'$ e quindi $v_Q = Q - P \in V'$. Il sottospazio vettoriale V' contiene dunque l'insieme dei vettori $v_Q = Q - P$, al variare di Q in \mathcal{S} , e pertanto contiene anche il sottospazio vettoriale W da essi generato. Per ogni vettore $w \in W$ si ha quindi $P + w \in \mathbb{A}'$, il che dimostra che $\mathbb{B} \subseteq \mathbb{A}'$. \square

Come già accadeva nel caso degli spazi vettoriali, l'unione di due sottospazi affini di uno spazio affine \mathbb{A} non è, in generale, un sottospazio affine di \mathbb{A} . Definiamo dunque la *somma* di due sottospazi affini \mathbb{L} e \mathbb{M} , che indicheremo³ con $\mathbb{L} \vee \mathbb{M}$, come il sottospazio affine generato dall'unione di \mathbb{L} e \mathbb{M} , cioè come il più piccolo sottospazio affine di \mathbb{A} contenente \mathbb{L} e \mathbb{M} . Nel caso particolare in cui i sottospazi affini in questione sono due punti, P e Q , con la notazione $P \vee Q$ si indica dunque il più piccolo sottospazio affine contenente i punti P e Q , cioè la retta passante per P e Q (se i punti P e Q sono distinti). Analogamente, dati tre punti P, Q e R di uno spazio affine \mathbb{A} , la notazione $P \vee Q \vee R$ indica il più piccolo sottospazio affine di \mathbb{A} contenente i tre punti dati, cioè il piano passante per P, Q e R , se i tre punti in questione non sono allineati.

Per la somma di due sottospazi affini vale il seguente risultato:

Proposizione 6.2.7. *Siano $\mathbb{L} = (\mathcal{L}, L, +)$ e $\mathbb{M} = (\mathcal{M}, M, +)$ due sottospazi affini di uno spazio affine \mathbb{A} . Se $\mathcal{L} \cap \mathcal{M} \neq \emptyset$, allora*

$$\mathbb{L} \vee \mathbb{M} = P + (L + M) = \{P + v \mid v \in L + M\},$$

per ogni punto $P \in \mathcal{L} \cup \mathcal{M}$. Se invece $\mathcal{L} \cap \mathcal{M} = \emptyset$, allora

$$\mathbb{L} \vee \mathbb{M} = P + (L + M + \langle u \rangle) = \{P + v \mid v \in L + M + \langle u \rangle\},$$

ove $P \in \mathcal{L} \cup \mathcal{M}$ e $u = B - A$, con $A \in \mathcal{L}$ e $B \in \mathcal{M}$.

Dimostrazione. Consideriamo dapprima il caso in cui $\mathcal{L} \cap \mathcal{M} \neq \emptyset$. Sia $A \in \mathcal{L} \cap \mathcal{M}$ e fissiamo un punto $P \in \mathcal{L} \cup \mathcal{M}$. A meno di scambiare i ruoli di \mathcal{L} e \mathcal{M} , non è restrittivo supporre che $P \in \mathcal{L}$. Consideriamo dunque il sottospazio affine $\mathbb{B} = P + (L + M)$. Per dimostrare che $\mathbb{B} = \mathbb{L} \vee \mathbb{M}$ dobbiamo dimostrare che \mathbb{B} è il più piccolo sottospazio affine di \mathbb{A} che contiene \mathbb{L} e \mathbb{M} . Per ogni punto $Q \in \mathbb{L}$ il vettore $v = Q - P$ appartiene allo spazio vettoriale L , quindi $Q = P + v \in P + L \subseteq \mathbb{B}$; ciò dimostra che $\mathbb{L} \subseteq \mathbb{B}$. Per dimostrare che anche $\mathbb{M} \subseteq \mathbb{B}$, consideriamo un punto $R \in \mathcal{M}$ e poniamo $w = R - A$ (notiamo che $w \in M$, dato che anche il punto A appartiene a \mathcal{M}). Dato che i punti A e P appartengono a \mathbb{L} , il vettore $u = A - P$ appartiene al sottospazio vettoriale L . Si ha dunque

$$R = P + (A - P) + (R - A) = P + u + w \in P + (L + M),$$

pertanto $R \in \mathbb{B}$. Abbiamo così dimostrato che \mathbb{B} contiene \mathbb{L} e \mathbb{M} .

³Un'altra notazione comunemente usata per indicare la somma dei due sottospazi affini \mathbb{L} e \mathbb{M} è $\mathbb{L} + \mathbb{M}$.

Sia ora \mathbb{B}' un sottospazio affine di \mathbb{A} contenente \mathbb{L} e \mathbb{M} . Se indichiamo con W lo spazio vettoriale soggiacente a \mathbb{B}' , si ha $W \supseteq \mathbb{L}$ e $W \supseteq \mathbb{M}$, pertanto $W \supseteq \mathbb{L} + \mathbb{M}$. Poiché anche $P \in \mathbb{B}'$, si ha $\mathbb{B} = P + (\mathbb{L} + \mathbb{M}) \subseteq \mathbb{B}'$, il che dimostra che \mathbb{B} è il più piccolo sottospazio affine di \mathbb{A} contenente \mathbb{L} e \mathbb{M} , cioè $\mathbb{B} = \mathbb{L} \vee \mathbb{M}$.

Consideriamo ora il caso in cui $\mathcal{L} \cap \mathcal{M} = \emptyset$. Siano $A \in \mathcal{L}$, $B \in \mathcal{M}$ e poniamo $u = B - A$, fissiamo inoltre un punto $P \in \mathcal{L} \cup \mathcal{M}$. Anche in questo caso, a meno di scambiare i ruoli di \mathcal{L} e \mathcal{M} , non è restrittivo supporre che $P \in \mathcal{L}$. Consideriamo il sottospazio affine $\mathbb{B} = P + (L + M + \langle u \rangle)$. Per dimostrare che $\mathbb{B} = \mathbb{L} \vee \mathbb{M}$ bisogna dimostrare che \mathbb{B} è il più piccolo sottospazio affine di \mathbb{A} che contiene \mathbb{L} e \mathbb{M} . Come nel caso precedente, per ogni punto $Q \in \mathbb{L}$ il vettore $v = Q - P$ appartiene allo spazio vettoriale L , quindi $Q = P + v \in P + L \subseteq \mathbb{B}$; ciò dimostra che $\mathbb{L} \subseteq \mathbb{B}$. Per dimostrare che anche $\mathbb{M} \subseteq \mathbb{B}$, consideriamo un punto $R \in \mathbb{M}$ e notiamo che si ha

$$R = P + (A - P) + (B - A) + (R - B).$$

Ora basta osservare che il vettore $A - P$ appartiene a L , dato che $A, P \in \mathbb{L}$, che $R - B \in M$ dato che i punti B e R appartengono a \mathbb{M} e che $B - A = u$, per concludere che $R \in P + (L + M + \langle u \rangle) = \mathbb{B}$. Abbiamo così dimostrato che \mathbb{B} contiene \mathbb{L} e \mathbb{M} .

Consideriamo ora un sottospazio affine \mathbb{B}' contenente \mathbb{L} e \mathbb{M} . Se indichiamo con W lo spazio vettoriale soggiacente a \mathbb{B}' , si ha $W \supseteq \mathbb{L}$ e $W \supseteq \mathbb{M}$. Inoltre, poiché $A, B \in \mathbb{B}'$, si ha anche $u = B - A \in W$, quindi $W \supseteq L + M + \langle u \rangle$. Si ha dunque $\mathbb{B} = P + (L + M + \langle u \rangle) \subseteq \mathbb{B}'$. Pertanto \mathbb{B} è il più piccolo sottospazio affine di \mathbb{A} contenente \mathbb{L} e \mathbb{M} , cioè $\mathbb{B} = \mathbb{L} \vee \mathbb{M}$. \square

A differenza dell'unione, l'intersezione di due o più sottospazi affini è, solitamente, un sottospazio affine. Si ha infatti:

Proposizione 6.2.8. *Sia $\mathbb{B}_i = (\mathcal{B}_i, W_i, +)$, $i \in I$, una famiglia di sottospazi affini di uno spazio affine $\mathbb{A} = (\mathcal{A}, V, +)$. Se $\bigcap_{i \in I} \mathcal{B}_i \neq \emptyset$, l'intersezione dei sottospazi affini \mathbb{B}_i è un sottospazio affine di \mathbb{A} , il cui insieme dei punti è $\bigcap_{i \in I} \mathcal{B}_i$ e il cui spazio direttore è $\bigcap_{i \in I} W_i$. Si ha cioè*

$$\bigcap_{i \in I} \mathbb{B}_i = \left(\bigcap_{i \in I} \mathcal{B}_i, \bigcap_{i \in I} W_i, + \right).$$

Dimostrazione. La dimostrazione si riduce alla verifica che per la terna

$$\left(\bigcap_{i \in I} \mathcal{B}_i, \bigcap_{i \in I} W_i, + \right)$$

valgono le proprietà (i), (ii) e (iii) della definizione di spazio affine, il che è del tutto ovvio. \square

Osservazione 6.2.9. L'unico caso in cui l'intersezione di due o più sottospazi affini non è un sottospazio affine si ha quando tale intersezione è l'insieme vuoto. Risulta pertanto conveniente considerare anche l'insieme vuoto come uno spazio affine. Naturalmente, la dimensione di un tale spazio affine non è definita. È comunque possibile attribuire una dimensione allo spazio affine vuoto ponendo $\dim \emptyset = -1$ (a tal proposito, si ricordi che uno spazio affine di dimensione zero non è vuoto, ma è costituito da un solo punto).

Osservazione 6.2.10. Dato un sottoinsieme \mathcal{S} di uno spazio affine \mathbb{A} , si può ora facilmente verificare che il sottospazio affine di \mathbb{A} generato da \mathcal{S} coincide con l'intersezione di tutti i sottospazi affini di \mathbb{A} contenenti \mathcal{S} .

Per quanto riguarda le posizioni reciproche di due sottospazi affini, possiamo dare la seguente definizione:

Definizione 6.2.11. Siano \mathbb{L} e \mathbb{M} due sottospazi affini di uno spazio affine \mathbb{A} , di sottospazi direttori L e M , rispettivamente.

- (i) \mathbb{L} e \mathbb{M} sono *incidenti* se $\mathbb{L} \cap \mathbb{M} \neq \emptyset$;
- (ii) \mathbb{L} e \mathbb{M} sono *paralleli* se $L \subseteq M$ oppure $M \subseteq L$;
- (iii) \mathbb{L} e \mathbb{M} sono *sghembì* se $\mathbb{L} \cap \mathbb{M} = \emptyset$ e $L \cap M = \{\mathbf{0}\}$.

Dalle Proposizioni 6.2.7 e 6.2.8 discende subito il seguente risultato:

Corollario 6.2.12. Siano \mathbb{L} e \mathbb{M} due sottospazi affini di \mathbb{A} . Se \mathbb{L} e \mathbb{M} sono incidenti oppure sghembì, si ha

$$\dim(\mathbb{L} \vee \mathbb{M}) = \dim \mathbb{L} + \dim \mathbb{M} - \dim(\mathbb{L} \cap \mathbb{M}),$$

altrimenti

$$\dim(\mathbb{L} \vee \mathbb{M}) < \dim \mathbb{L} + \dim \mathbb{M} - \dim(\mathbb{L} \cap \mathbb{M}).$$

Dimostrazione. Indichiamo con L e M i sottospazi direttori di \mathbb{L} e \mathbb{M} , rispettivamente. Se \mathbb{L} e \mathbb{M} sono incidenti, $\mathbb{L} \cap \mathbb{M}$ è un sottospazio affine il cui spazio direttore è $L \cap M$, mentre lo spazio vettoriale soggiacente a $\mathbb{L} \vee \mathbb{M}$ è la somma dei sottospazi L e M . Dalla formula di Grassmann (Cap. 1, Proposizione 1.3.49) segue allora che

$$\begin{aligned} \dim(\mathbb{L} \vee \mathbb{M}) &= \dim(L + M) \\ &= \dim L + \dim M - \dim(L \cap M) \\ &= \dim \mathbb{L} + \dim \mathbb{M} - \dim(\mathbb{L} \cap \mathbb{M}). \end{aligned}$$

Se \mathbb{L} e \mathbb{M} sono sghembì, si ha $\mathbb{L} \cap \mathbb{M} = \emptyset$ e $L \cap M = \{\mathbf{0}\}$. Avendo posto, per convenzione, $\dim \emptyset = -1$ (vedi Osservazione 6.2.9), dalla Proposizione 6.2.7 e dalla formula di Grassmann si deduce che

$$\begin{aligned} \dim(\mathbb{L} \vee \mathbb{M}) &= \dim(L + M + \langle u \rangle) \\ &= \dim L + \dim M + 1 \\ &= \dim \mathbb{L} + \dim \mathbb{M} - \dim(\mathbb{L} \cap \mathbb{M}). \end{aligned}$$

Se invece \mathbb{L} e \mathbb{M} non sono incidenti né sghembì, si ha $\mathbb{L} \cap \mathbb{M} = \emptyset$ e $L \cap M \neq \{\mathbf{0}\}$, e dunque $\dim(L \cap M) > 0$. In questo caso si deduce facilmente che

$$\begin{aligned} \dim(\mathbb{L} \vee \mathbb{M}) &= \dim(L + M + \langle u \rangle) \\ &= \dim L + \dim M - \dim(L \cap M) + 1 \\ &< \dim L + \dim M + 1 \\ &= \dim \mathbb{L} + \dim \mathbb{M} - \dim(\mathbb{L} \cap \mathbb{M}). \end{aligned}$$

□

6.3 Sistemi di riferimento

Un sistema di riferimento in uno spazio affine è l'analogo di una base per uno spazio vettoriale. Ricordiamo che la scelta di una base di uno spazio vettoriale n -dimensionale V permette di associare ad ogni vettore $v \in V$ una n -upla di elementi del campo K . In modo del tutto analogo, la scelta di un sistema di riferimento in uno spazio affine n -dimensionale \mathbb{A} permette di associare a ogni punto (e a ogni vettore) di \mathbb{A} una n -upla di elementi di K .

Definizione 6.3.1. Sia $\mathbb{A} = (\mathcal{A}, V, +)$ uno spazio affine n -dimensionale sul campo K . Un *sistema di riferimento* \mathcal{R} in \mathbb{A} è il dato di un punto $O \in \mathbb{A}$ (detto *origine*) e di una base v_1, v_2, \dots, v_n di V .

Equivalentemente, un sistema di riferimento in \mathbb{A} è il dato di $n + 1$ punti, P_0, P_1, \dots, P_n , tali che i vettori $v_1 = P_1 - P_0, v_2 = P_2 - P_0, \dots, v_n = P_n - P_0$, siano una base di V (in tal caso si può prendere come origine del sistema di riferimento il punto $O = P_0$).

Sia dunque $\mathcal{R} = \{O, v_1, v_2, \dots, v_n\}$ un sistema di riferimento in uno spazio affine \mathbb{A} sul campo K . Per ogni punto $P \in \mathbb{A}$ esiste un vettore $v \in V$ tale che $P = O + v$. Scrivendo v come combinazione lineare dei vettori di base v_1, v_2, \dots, v_n ,

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n,$$

si ottiene

$$P = O + \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n.$$

Gli scalari $\lambda_1, \dots, \lambda_n$ sono detti le *coordinate* del punto P nel sistema di riferimento dato. Ad ogni punto $P \in \mathbb{A}$ possiamo quindi associare una n -upla $(\lambda_1, \dots, \lambda_n) \in K^n$; si noti che al punto O (l'origine del sistema di riferimento) viene associata la n -upla nulla $(0, 0, \dots, 0)$. Analogamente, ad ogni vettore $v \in V$ risulta associata la n -upla delle sue componenti rispetto alla base v_1, v_2, \dots, v_n . La scelta di un sistema di riferimento determina quindi una biiezione tra lo spazio affine n -dimensionale \mathbb{A} sul campo K e lo spazio affine \mathbb{A}_K^n ; tale biiezione è ottenuta associando ad ogni punto $P \in \mathbb{A}$ le sue coordinate $(\lambda_1, \dots, \lambda_n) \in K^n$ e ad ogni vettore $v \in V$ le sue componenti rispetto alla base data.

Definizione 6.3.2. Sia $\mathcal{R} = \{O, v_1, v_2, \dots, v_n\}$ un sistema di riferimento in uno spazio affine \mathbb{A} . Le n rette passanti per O e parallele ai vettori v_1, v_2, \dots, v_n sono dette gli *assi* del sistema di riferimento (o *assi coordinati*). Gli n iperpiani π_i ($i = 1, \dots, n$) passanti per O e paralleli ai sottospazi generati dai vettori $v_1, \dots, \hat{v}_i, \dots, v_n$ (la notazione \hat{v}_i significa che tale vettore va escluso) sono detti gli *iperpiani coordinati*.

Definizione 6.3.3. Dati $n + 1$ punti di uno spazio affine \mathbb{A} , diremo che essi sono *affinamente indipendenti* (o *in posizione generica*) se il sottospazio affine da essi generato ha dimensione n .

Proposizione 6.3.4. *Un sistema di riferimento in uno spazio affine $\mathbb{A} = (\mathcal{A}, V, +)$ di dimensione n è costituito da $n + 1$ punti affinamente indipendenti.*

Dimostrazione. Dati $n + 1$ punti P_0, P_1, \dots, P_n , poniamo $v_1 = P_1 - P_0, v_2 = P_2 - P_0, \dots, v_n = P_n - P_0$ e indichiamo con W il sottospazio vettoriale di V

generato dai vettori v_i , per $i = 1, \dots, n$. Per la Proposizione 6.2.6, il sottospazio affine generato dai punti P_0, \dots, P_n è

$$\langle P_0, \dots, P_n \rangle = P_0 + W.$$

Da ciò segue che i punti P_0, \dots, P_n sono affinamente indipendenti se e solo se $\dim W = n$, cioè se e solo se i vettori v_1, \dots, v_n sono una base di V . Ciò equivale ad affermare che gli $n+1$ punti P_0, P_1, \dots, P_n costituiscono un sistema di riferimento in \mathbb{A} . \square

Osservazione 6.3.5. Sia P_0, P_1, \dots, P_n un sistema di riferimento in uno spazio affine \mathbb{A} . Si verifica facilmente che, per ogni permutazione $\sigma \in \mathfrak{S}_{n+1}$, i punti $P_{\sigma(0)}, P_{\sigma(1)}, \dots, P_{\sigma(n)}$ formano anch'essi un sistema di riferimento in \mathbb{A} .

6.4 Equazioni dei sottospazi affini

Come abbiamo visto nel paragrafo precedente, la scelta di un sistema di riferimento in uno spazio affine permette di introdurre un sistema di coordinate. In questo modo ogni spazio affine di dimensione n sul campo K può essere identificato con lo spazio affine standard \mathbb{A}_K^n .

In questa sezione mostreremo come i sottospazi affini di uno spazio affine dotato di un sistema di riferimento possano essere identificati con gli insiemi delle soluzioni dei sistemi di equazioni lineari.

Indichiamo dunque con $\mathbb{A} = \mathbb{A}_K^n$ lo spazio affine n -dimensionale sul campo K e consideriamo un sottospazio affine \mathbb{B} , di dimensione r , di \mathbb{A} . Sia $W \subseteq K^n$ il sottospazio direttore di \mathbb{B} e sia w_1, w_2, \dots, w_r una base di W . Dato un punto $P \in \mathbb{B}$, si ha:

$$\mathbb{B} = P + W = \{P + w \mid w \in W\}.$$

Poiché ogni vettore $w \in W$ può essere espresso come combinazione lineare dei vettori di base w_1, w_2, \dots, w_r , il generico punto $X = (x_1, x_2, \dots, x_n)$ di \mathbb{B} può essere dunque espresso nella forma

$$X = P + \lambda_1 w_1 + \lambda_2 w_2 + \dots + \lambda_r w_r, \quad (6.4.1)$$

al variare di $\lambda_1, \lambda_2, \dots, \lambda_r \in K$. In coordinate, l'equazione (6.4.1) si traduce nel seguente sistema di equazioni:

$$\begin{cases} x_1 = p_1 + \lambda_1 a_{11} + \lambda_2 a_{12} + \dots + \lambda_r a_{1r} \\ x_2 = p_2 + \lambda_1 a_{21} + \lambda_2 a_{22} + \dots + \lambda_r a_{2r} \\ \vdots \\ x_n = p_n + \lambda_1 a_{n1} + \lambda_2 a_{n2} + \dots + \lambda_r a_{nr} \end{cases} \quad (6.4.2)$$

ove $P = (p_1, p_2, \dots, p_n)$ e $w_i = (a_{1i}, a_{2i}, \dots, a_{ni}) \in K^n$, per ogni $i = 1, \dots, r$. Gli scalari $\lambda_1, \lambda_2, \dots, \lambda_r$ sono anche chiamati *parametri* e le equazioni precedenti sono dette le *equazioni parametriche* del sottospazio affine \mathbb{B} .

Nel sistema di equazioni (6.4.2) è possibile eliminare i parametri $\lambda_1, \dots, \lambda_r$ ottenendo così un sistema di equazioni lineari (non omogenee) nelle indeterminate

nate x_1, x_2, \dots, x_n , del tipo

$$S : \begin{cases} a'_{11}x_1 + a'_{12}x_2 + \cdots + a'_{1n}x_n = b_1 \\ a'_{21}x_1 + a'_{22}x_2 + \cdots + a'_{2n}x_n = b_2 \\ \vdots \quad \vdots \quad \vdots \\ a'_{s1}x_1 + a'_{s2}x_2 + \cdots + a'_{sn}x_n = b_s \end{cases} \quad (6.4.3)$$

per opportuni coefficienti a'_{ij} e b_i in K . Quest'ultime sono dette le *equazioni cartesiane* di \mathbb{B} .

Viceversa, dato un sistema S di equazioni lineari come in (6.4.3), indichiamo con W il sottospazio vettoriale di K^n costituito dalle soluzioni del sistema omogeneo S_0 ad esso associato (vedi Cap. 2, Proposizione 2.3.2):

Se il sistema S è incompatibile l'insieme Σ delle soluzioni di S è vuoto: esso può quindi essere considerato un sottospazio affine di \mathbb{A} (vedi Osservazione 6.2.9). Se invece S ammette soluzioni ogni tale soluzione può essere espressa come somma di una soluzione particolare $P = (p_1, \dots, p_n)$ di S con una soluzione $w \in W$ del sistema omogeneo associato (vedi Cap. 2, Proposizione 2.3.3). In tal caso l'insieme Σ delle soluzioni di S è dato da $\Sigma = P + W$ e dunque Σ è un sottospazio affine di \mathbb{A} .

Possiamo quindi concludere che i sottospazi affini di \mathbb{A}_K^n possono essere descritti come insiemi delle soluzioni di sistemi di equazioni lineari in n incognite. Si noti tuttavia che tale corrispondenza tra sistemi di equazioni lineari e sottospazi affini non è biunivoca; infatti esistono sistemi diversi aventi lo stesso insieme di soluzioni.

Vediamo ora alcuni esempi particolarmente significativi.

Esempio 6.4.1 (EQUAZIONE DI UNA RETTA NEL PIANO AFFINE). Sia r una retta nel piano affine \mathbb{A}_K^2 . Il sottospazio direttrice di r ha dimensione 1 ed è quindi generato da un vettore (non nullo) $v = (a_1, a_2) \in K^2$. Se $P = (p_1, p_2)$ è un punto di r e se indichiamo con $X = (x_1, x_2)$ un punto generico del piano, si ha

$$X \in r \iff X = P + \lambda v, \lambda \in K.$$

L'equazione vettoriale $X = P + \lambda v$ si traduce nel seguente sistema di equazioni:

$$\begin{cases} x_1 = p_1 + \lambda a_1 \\ x_2 = p_2 + \lambda a_2. \end{cases}$$

Queste sono dunque le equazioni parametriche di una retta nel piano affine.

Ricavando λ da una delle due equazioni precedenti e sostituendo l'espressione trovata nell'altra, si ottiene un'equazione di primo grado nelle incognite x_1 e x_2 , del tipo

$$ax_1 + bx_2 + c = 0,$$

ove $a, b, c \in K$. Possiamo quindi concludere che una retta nel piano affine è il luogo degli zeri di un polinomio di primo grado nelle indeterminate x_1 e x_2 .

Esempio 6.4.2 (EQUAZIONE DI UN PIANO NELLO SPAZIO AFFINE TRIDIMENSIONALE). Sia π un piano nello spazio affine \mathbb{A}_K^3 . Il sottospazio diretore di π ha dimensione 2 ed è quindi generato da due vettori (linearmente indipendenti) $v_1 = (a_{11}, a_{21}, a_{31})$ e $v_2 = (a_{12}, a_{22}, a_{32})$. Se $P = (p_1, p_2, p_3)$ è un punto del piano π e se indichiamo con $X = (x_1, x_2, x_3)$ un punto generico dello spazio affine, si ha

$$X \in \pi \iff X = P + \lambda_1 v_1 + \lambda_2 v_2, \quad \lambda_1, \lambda_2 \in K.$$

L'equazione vettoriale $X = P + \lambda_1 v_1 + \lambda_2 v_2$ corrisponde al seguente sistema di equazioni:

$$\begin{cases} x_1 = p_1 + \lambda_1 a_{11} + \lambda_2 a_{12} \\ x_2 = p_2 + \lambda_1 a_{21} + \lambda_2 a_{22} \\ x_3 = p_3 + \lambda_1 a_{31} + \lambda_2 a_{32}. \end{cases}$$

Queste sono pertanto le equazioni parametriche di un piano nello spazio affine tridimensionale.

Eliminando i parametri λ_1 e λ_2 dalle equazioni precedenti si ottiene un'equazione di primo grado nelle incognite x_1 , x_2 e x_3 , del tipo

$$ax_1 + bx_2 + cx_3 + d = 0,$$

ove $a, b, c, d \in K$. Possiamo quindi concludere che un piano nello spazio affine tridimensionale è il luogo degli zeri di un polinomio di primo grado nelle indeterminate x_1 , x_2 , x_3 .

Esempio 6.4.3 (EQUAZIONE DI UN IPERPIANO NELLO SPAZIO AFFINE n -DIMENSIONALE). Generalizzando i due esempi precedenti è possibile ricavare l'equazione di un sottospazio affine di dimensione $n - 1$ dello spazio affine \mathbb{A}_K^n .

Sia dunque π un iperpiano in \mathbb{A}_K^n . Il sottospazio diretore di π ha dimensione $n - 1$ ed è quindi generato da $n - 1$ vettori (linearmente indipendenti) $v_1 = (a_{11}, a_{21}, \dots, a_{n1})$, \dots , $v_{n-1} = (a_{1,n-1}, a_{2,n-1}, \dots, a_{n,n-1}) \in K^n$. Se $P = (p_1, p_2, \dots, p_n)$ è un punto dell'iperpiano π e se indichiamo con $X = (x_1, x_2, \dots, x_n)$ un punto generico di \mathbb{A}_K^n , si ha

$$X \in \pi \iff X = P + \lambda_1 v_1 + \dots + \lambda_{n-1} v_{n-1}, \quad \lambda_1, \dots, \lambda_{n-1} \in K.$$

L'equazione vettoriale $X = P + \lambda_1 v_1 + \dots + \lambda_{n-1} v_{n-1}$ si traduce nel seguente sistema di equazioni:

$$\begin{cases} x_1 = p_1 + \lambda_1 a_{11} + \lambda_2 a_{12} + \dots + \lambda_{n-1} a_{1,n-1} \\ x_2 = p_2 + \lambda_1 a_{21} + \lambda_2 a_{22} + \dots + \lambda_{n-1} a_{2,n-1} \\ \quad \ddots \\ x_n = p_n + \lambda_1 a_{n1} + \lambda_2 a_{n2} + \dots + \lambda_{n-1} a_{n,n-1}. \end{cases}$$

Queste sono dunque le equazioni parametriche di un iperpiano nello spazio affine n -dimensionale.

Eliminando i parametri $\lambda_1, \lambda_2, \dots, \lambda_{n-1}$ dalle equazioni precedenti si ottiene un'equazione di primo grado nelle incognite x_1, x_2, \dots, x_n , del tipo

$$a_1 x_1 + a_2 x_2 + \dots + a_n x_n + a_{n+1} = 0,$$

con $a_1, \dots, a_{n+1} \in K$. Possiamo quindi concludere che un iperpiano nello spazio affine n -dimensionale è il luogo degli zeri di un polinomio di primo grado nelle indeterminate x_1, x_2, \dots, x_n .

Esempio 6.4.4 (EQUAZIONE DI UNA RETTA NELLO SPAZIO AFFINE TRIDIMENSIONALE). Sia r una retta nello spazio affine \mathbb{A}_K^3 . Il sottospazio direttore di r ha dimensione 1 ed è quindi generato da un vettore (non nullo) $v = (a_1, a_2, a_3) \in K^3$. Se $P = (p_1, p_2, p_3)$ è un punto di r e se indichiamo con $X = (x_1, x_2, x_3)$ un punto generico dello spazio affine, si ha

$$X \in r \iff X = P + \lambda v, \quad \lambda \in K.$$

L'equazione vettoriale $X = P + \lambda v$ fornisce il seguente sistema di equazioni:

$$\begin{cases} x_1 = p_1 + \lambda a_1 \\ x_2 = p_2 + \lambda a_2 \\ x_3 = p_3 + \lambda a_3. \end{cases}$$

Queste sono dunque le equazioni parametriche di una retta nello spazio affine tridimensionale.

Ricavando λ da una delle tre equazioni precedenti e sostituendo l'espressione trovata nelle altre due, si ottiene un sistema di due equazioni di primo grado nelle incognite x_1, x_2 e x_3 , del tipo

$$\begin{cases} ax_1 + bx_2 + cx_3 + d = 0 \\ ex_1 + fx_2 + gx_3 + h = 0. \end{cases}$$

Possiamo quindi concludere che una retta nello spazio affine tridimensionale può essere identificata con l'insieme delle soluzioni di un sistema di due equazioni di primo grado nelle incognite x_1, x_2, x_3 . Non vale invece il viceversa, cioè non è vero che ogni sistema di due equazioni lineari in tre incognite determina una retta: ad esempio, l'insieme delle soluzioni di un tale sistema potrebbe essere vuoto.

Osserviamo che questo risultato ha un'ovvia interpretazione geometrica. Infatti, come abbiamo visto nell'Esempio 6.4.2, l'insieme delle soluzioni di un'equazione lineare nelle incognite x_1, x_2, x_3 è un piano nello spazio affine tridimensionale. Pertanto l'insieme delle soluzioni di un sistema di due equazioni lineari corrisponde all'insieme dei punti comuni a due piani. Se tali piani sono paralleli e distinti, la loro intersezione è l'insieme vuoto; in questo caso il corrispondente sistema di equazioni lineari è incompatibile. Se invece i due piani non sono paralleli, la loro intersezione è una retta.

Osservazione 6.4.5. Dalle considerazioni precedenti si deduce anche che, se il sottospazio affine (non vuoto) $\mathbb{B} = (\mathcal{B}, W, +)$ dello spazio affine \mathbb{A} è dato dalle soluzioni del sistema di equazioni lineari

$$S : AX = B,$$

allora il sottospazio direttore W di \mathbb{B} è l'insieme dalle soluzioni del sistema omogeneo associato

$$S_0 : AX = \mathbf{0}.$$

Inoltre, se \mathbb{B} non è vuoto, la sua dimensione è data da

$$\dim \mathbb{B} = n - \text{rk}(A),$$

ove n è la dimensione dello spazio affine \mathbb{A} . In particolare, \mathbb{B} è un iperpiano di \mathbb{A} se e solo se $\text{rk}(A) = 1$, cioè se e solo se \mathbb{B} può essere descritto da una sola equazione non banale.

Consideriamo ora il problema di determinare l'intersezione di due sottospazi affini in termini delle loro equazioni cartesiane. Siano dunque $\mathbb{L} = (\mathcal{L}, L, +)$ e $\mathbb{M} = (\mathcal{M}, M, +)$ due sottospazi affini dello spazio affine $\mathbb{A} = \mathbb{A}_K^n$ e siano

$$AX = B \quad \text{e} \quad CX = D$$

i corrispondenti sistemi di equazioni lineari. Le coordinate dei punti appartenenti a $\mathbb{L} \cap \mathbb{M}$ soddisfano sia le equazioni del sistema $AX = B$ che quelle del sistema $CX = D$; esse sono pertanto le soluzioni del sistema lineare ottenuto considerando tutte le equazioni dei due sistemi precedenti. Se denotiamo con $EX = F$ tale sistema, la matrice E e il vettore colonna F sono dati da

$$E = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right), \quad F = \left(\begin{array}{c} B \\ D \end{array} \right).$$

Dal Teorema di Rouché-Capelli si deduce ora il seguente risultato, che fornisce un criterio per determinare la posizione reciproca di due sottovarietà lineari di uno spazio affine.

Proposizione 6.4.6. *Siano $\mathbb{L} = (\mathcal{L}, L, +)$ e $\mathbb{M} = (\mathcal{M}, M, +)$ due sottovarietà lineari dello spazio affine $\mathbb{A} = \mathbb{A}_K^n$, rappresentate dai sistemi lineari $AX = B$ e $CX = D$, rispettivamente. Allora si ha:*

(i) \mathbb{L} e \mathbb{M} sono incidenti se e solo se

$$\text{rk} \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right) = \text{rk} \left(\begin{array}{c} A \\ C \end{array} \right).$$

In tal caso $\mathbb{L} \cap \mathbb{M}$ è una sottovarietà lineare di dimensione $n - t$, ove t è il valore comune dei due ranghi precedenti.

(ii) \mathbb{L} e \mathbb{M} sono sghembe se e solo se

$$\text{rk} \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right) > \text{rk} \left(\begin{array}{c} A \\ C \end{array} \right) \quad \text{e} \quad \text{rk} \left(\begin{array}{c} A \\ C \end{array} \right) = n.$$

(iii) Se si ha

$$\text{rk} \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right) > \text{rk} \left(\begin{array}{c} A \\ C \end{array} \right) \quad \text{e} \quad \text{rk} \left(\begin{array}{c} A \\ C \end{array} \right) = t < n,$$

allora \mathbb{L} contiene una sottovarietà lineare di dimensione $n - t$ parallela a \mathbb{M} .

Dimostrazione. (i) \mathbb{L} e \mathbb{M} sono incidenti se e solo se il sistema

$$\left(\begin{array}{c} A \\ C \end{array} \right) X = \left(\begin{array}{c} B \\ D \end{array} \right) \tag{6.4.4}$$

ammette soluzioni. Per il Teorema di Rouché-Capelli questo avviene se e solo se le matrici completa e incompleta di tale sistema hanno lo stesso rango, cioè se e solo se

$$\text{rk} \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right) = \text{rk} \left(\begin{array}{c} A \\ C \end{array} \right).$$

(ii) Ricordiamo che le due sottovarietà lineari \mathbb{L} e \mathbb{M} sono sghembe se e solo se $\mathcal{L} \cap \mathcal{M} = \emptyset$ e $L \cap M = \{\mathbf{0}\}$. Per il Teorema di Rouché-Capelli, la condizione $\mathcal{L} \cap \mathcal{M} = \emptyset$ equivale a

$$\text{rk} \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right) > \text{rk} \left(\begin{array}{c} A \\ \hline C \end{array} \right).$$

La seconda condizione,

$$\text{rk} \left(\begin{array}{c} A \\ \hline C \end{array} \right) = n,$$

equivale a richiedere che il sistema omogeneo associato

$$\left(\begin{array}{c} A \\ \hline C \end{array} \right) X = \mathbf{0}$$

abbia come unica soluzione il vettore nullo, cioè che sia $L \cap M = \{\mathbf{0}\}$.

(iii) La diseguaglianza

$$\text{rk} \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right) > \text{rk} \left(\begin{array}{c} A \\ \hline C \end{array} \right)$$

implica che il sistema (6.4.4) non ammette soluzioni, cioè che $\mathcal{L} \cap \mathcal{M} = \emptyset$. La condizione

$$\text{rk} \left(\begin{array}{c} A \\ \hline C \end{array} \right) = t < n$$

implica che l'insieme delle soluzioni del sistema omogeneo associato

$$\left(\begin{array}{c} A \\ \hline C \end{array} \right) X = \mathbf{0}$$

forma un sottospazio vettoriale Z di dimensione $n - t$; si ha dunque $Z = L \cap M$. Scelto arbitrariamente un punto $P \in \mathbb{L}$, la sottovarietà lineare $\mathbb{L}' = P + Z$ è contenuta in \mathbb{L} , ha dimensione $n - t$ ed è parallela a \mathbb{M} . \square

Corollario 6.4.7. *Sia \mathbb{L} un sottospazio affine di dimensione r di uno spazio affine n -dimensionale \mathbb{A} . Allora \mathbb{L} è intersezione di $n - r$ iperpiani di \mathbb{A} .*

Dimostrazione. Abbiamo già visto che \mathbb{L} può essere identificato con l'insieme delle soluzioni di un sistema formato da $n - r$ equazioni lineari. Basta ora ricordare che ogni tale equazione rappresenta un iperpiano di \mathbb{A} . \square

Definizione 6.4.8. Sia \mathbb{L} una sottovarietà lineare di dimensione r di uno spazio affine n -dimensionale \mathbb{A} . L'insieme degli iperpiani di \mathbb{A} contenenti \mathbb{L} è detto la *stella di iperpiani di centro* \mathbb{L} (nel caso in cui sia $r = n - 2$, si parla di *fascio di iperpiani*). Più in generale, dato m con $r \leq m \leq n$, la *stella di sottovarietà lineari m -dimensionali di centro* \mathbb{L} è l'insieme di tutti i sottospazi affini di dimensione m di \mathbb{A} che contengono \mathbb{L} .

Definizione 6.4.9. Sia W un sottospazio vettoriale di dimensione r dello spazio direttore di uno spazio affine n -dimensionale \mathbb{A} . L'insieme degli iperpiani di \mathbb{A} contenenti W nel loro spazio direttore è detto la *stella impropria di iperpiani paralleli a* W (se $r = n - 2$ si parla di *fascio improprio di iperpiani*). Più in generale, dato m con $r \leq m \leq n$, la *stella impropria di sottovarietà lineari di dimensione m parallele a* W è l'insieme di tutti i sottospazi affini di dimensione m di \mathbb{A} che contengono W nel loro spazio direttore.

Esempio 6.4.10. Nello spazio affine \mathbb{A}_K^n sia \mathbb{L} la sottovarietà lineare data dalle soluzioni del seguente sistema:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + b_1 = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + b_2 = 0 \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n + b_m = 0 \end{cases}$$

Allora le equazioni cartesiane degli iperpiani che formano la stella di centro \mathbb{L} sono date da

$$\begin{aligned} & \lambda_1(a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + b_1) \\ & + \lambda_2(a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + b_2) \\ & + \cdots + \lambda_m(a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n + b_m) = 0, \end{aligned}$$

per ogni $\lambda_1, \lambda_2, \dots, \lambda_m \in K$, non tutti nulli.

6.5 Alcuni risultati di geometria affine

Data una sottovarietà lineare \mathbb{L} in uno spazio affine \mathbb{A} , per ogni punto $P \in \mathbb{A}$ esiste un'unica sottovarietà lineare \mathbb{L}' , con $\dim \mathbb{L}' = \dim \mathbb{L}$, passante per P e parallela a \mathbb{L} ; tale sottovarietà lineare è infatti data da

$$\mathbb{L}' = P + \mathbb{L},$$

ove \mathbb{L} è il sottospazio direttore di \mathbb{L} . Questo risultato, nel caso in cui \mathbb{L} sia una retta, non è altro che il famoso “Quinto Postulato” di Euclide. È poi altrettanto immediato verificare che uno spazio affine soddisfa tutti i rimanenti assiomi che sono alla base della classica Geometria Euclidea. Pertanto, come già osservato all'inizio di questo capitolo, uno spazio affine fornisce un modello di spazio adeguato per lo studio di tale geometria.

In questa sezione presenteremo dunque alcuni classici risultati di geometria euclidea, che dimostreremo utilizzando il linguaggio e gli strumenti dell'algebra lineare sviluppati finora. Vedremo come tali tecniche permettano non solo di ottenere delle dimostrazioni particolarmente semplici, ma anche di estendere tali risultati al caso di spazi affini di dimensione qualunque.

Teorema 6.5.1 (TEOREMA DI TALETE). *Siano π_1, π_2 e π_3 tre iperpiani paralleli e distinti di uno spazio affine n -dimensionale \mathbb{A} sul campo K e siano r , s due rette di \mathbb{A} , non parallele a tali iperpiani. Per ogni $i = 1, 2, 3$, poniamo*

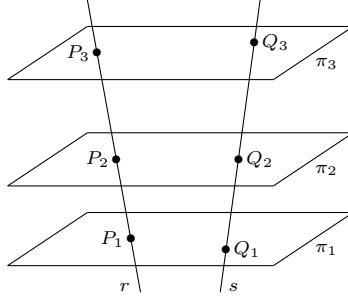
$$P_i = r \cap \pi_i, \quad Q_i = s \cap \pi_i.$$

Siano inoltre $\alpha, \beta \in K$ tali che

$$\overrightarrow{P_1 P_3} = \alpha \overrightarrow{P_1 P_2}, \quad \overrightarrow{Q_1 Q_3} = \beta \overrightarrow{Q_1 Q_2}.$$

Allora si ha $\alpha = \beta$.

Dimostrazione. Innanzitutto osserviamo che il fatto che ciascuna delle rette r e s intersechi l'iperpiano π_i in un punto è una conseguenza del Corollario 6.2.12. La situazione, nel caso in cui $\dim \mathbb{A} = 3$, è rappresentata nella seguente figura:



Indichiamo con W il sottospazio direttore degli iperpiani π_1, π_2, π_3 . Dall'uguaglianza

$$\overrightarrow{P_1Q_1} + \overrightarrow{Q_1Q_2} = \overrightarrow{P_1P_2} + \overrightarrow{P_2Q_2}$$

si ottiene

$$\overrightarrow{Q_1Q_2} - \overrightarrow{P_1P_2} = \overrightarrow{P_2Q_2} - \overrightarrow{P_1Q_1} \in W,$$

quindi

$$\overrightarrow{Q_1Q_2} = \overrightarrow{P_1P_2} + w, \quad (6.5.1)$$

per qualche $w \in W$. Analogamente, da

$$\overrightarrow{P_1Q_1} + \overrightarrow{Q_1Q_3} = \overrightarrow{P_1P_3} + \overrightarrow{P_3Q_3}$$

si deduce che

$$\overrightarrow{Q_1Q_3} - \overrightarrow{P_1P_3} = \overrightarrow{P_3Q_3} - \overrightarrow{P_1Q_1} \in W.$$

Ricordando che $\overrightarrow{P_1P_3} = \alpha \overrightarrow{P_1P_2}$ e $\overrightarrow{Q_1Q_3} = \beta \overrightarrow{Q_1Q_2}$, si ha dunque

$$\beta \overrightarrow{Q_1Q_2} - \alpha \overrightarrow{P_1P_2} \in W$$

e quindi, dato che è $\beta \neq 0$,

$$\overrightarrow{Q_1Q_2} - \frac{\alpha}{\beta} \overrightarrow{P_1P_2} \in W.$$

Da quest'ultima espressione e dalla (6.5.1) si deduce infine che

$$\left(1 - \frac{\alpha}{\beta}\right) \overrightarrow{P_1P_2} \in W.$$

Se fosse $\alpha/\beta \neq 1$, si avrebbe $\overrightarrow{P_1P_2} \in W$ e dunque la retta r sarebbe parallela agli iperpiani dati, contro l'ipotesi. Si conclude pertanto che deve essere $\alpha = \beta$. \square

Notiamo che, nel caso di uno spazio affine di dimensione due, il risultato appena dimostrato non è altro che il classico Teorema di Talete della geometria piana. Vale la pena osservare, inoltre, che il Teorema di Talete si può ulteriormente generalizzare come segue:

Teorema 6.5.2. *Siano \mathbb{L}_1 , \mathbb{L}_2 e \mathbb{L}_3 tre sottovarietà lineari di dimensione r , parallele e distinte, di uno spazio affine n -dimensionale \mathbb{A} sul campo K e siano \mathbb{H} , \mathbb{K} due sottovarietà lineari di dimensione $n-r$ di \mathbb{A} che intersecano ciascuna delle sottovarietà \mathbb{L}_i in un punto. Per ogni $i = 1, 2, 3$, poniamo*

$$P_i = \mathbb{H} \cap \mathbb{L}_i, \quad Q_i = \mathbb{K} \cap \mathbb{L}_i.$$

Siano inoltre $\alpha, \beta \in K$ tali che

$$\overrightarrow{P_1P_3} = \alpha \overrightarrow{P_1P_2}, \quad \overrightarrow{Q_1Q_3} = \beta \overrightarrow{Q_1Q_2}.$$

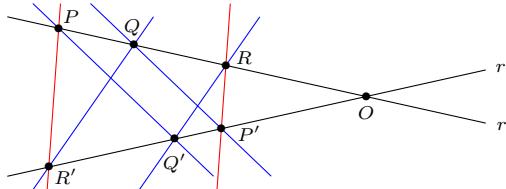
Allora si ha $\alpha = \beta$.

Dimostrazione. La dimostrazione di questo risultato è del tutto analoga a quella del Teorema 6.5.1. \square

Utilizzando il Teorema di Talete possiamo ora dimostrare il seguente risultato di geometria piana:

Teorema 6.5.3 (TEOREMA DI PAPPO). *In un piano affine \mathbb{A} siano date due rette distinte r e r' . Siano poi $P, Q, R \in r$, $P', Q', R' \in r'$, punti distinti tra loro e distinti anche dall'eventuale punto di intersezione di r e r' . Se la retta passante per P e Q' è parallela alla retta passante per P' e Q e la retta passante per Q e R' è parallela alla retta passante per Q' e R , allora la retta passante per P e R' è parallela alla retta passante per P' e R .*

Dimostrazione. Supponiamo che le rette r e r' siano incidenti e indichiamo con O il loro punto di intersezione.



Per il Teorema di Talete, si ha:

$$\begin{aligned} \overrightarrow{OQ} &= \alpha \overrightarrow{OR} & \overrightarrow{OP} &= \beta \overrightarrow{OQ} \\ \overrightarrow{OR'} &= \alpha \overrightarrow{OQ'} & \text{e} & \overrightarrow{OQ'} &= \beta \overrightarrow{OP}, \end{aligned}$$

con $\alpha, \beta \neq 0$. Da ciò segue che

$$\begin{aligned} \overrightarrow{PR'} &= \overrightarrow{OR'} - \overrightarrow{OP} = \alpha \overrightarrow{OQ'} - \beta \overrightarrow{OQ} \\ \overrightarrow{RP'} &= \overrightarrow{OP'} - \overrightarrow{OR} = \frac{1}{\beta} \overrightarrow{OQ'} - \frac{1}{\alpha} \overrightarrow{OQ}. \end{aligned}$$

Si conclude pertanto che $\overrightarrow{PR'} = \alpha \beta \overrightarrow{RP'}$ e dunque la retta passante per i punti P e R' è parallela a quella passante per P' e R .

Se invece le rette r e r' sono parallele, i quadrilateri $PQP'Q'$ e $QRQ'R'$ sono dei parallelogrammi, quindi si ha $\overrightarrow{PQ} = \overrightarrow{Q'P'}$ e $\overrightarrow{QR} = \overrightarrow{R'Q'}$. Da ciò si deduce che

$$\overrightarrow{PR} = \overrightarrow{PQ} + \overrightarrow{QR} = \overrightarrow{Q'P'} + \overrightarrow{R'Q'} = \overrightarrow{R'P'}$$

e quindi

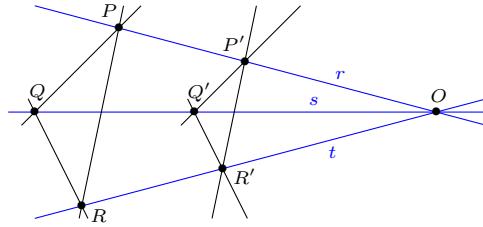
$$\overrightarrow{PR'} = \overrightarrow{PR} + \overrightarrow{RR'} = \overrightarrow{R'P'} + \overrightarrow{RR'} = \overrightarrow{RP'}.$$

Anche in questo caso si conclude dunque che la retta passante per i punti P e R' è parallela a quella passante per P' e R . \square

Un'altra conseguenza del Teorema di Talete è il seguente risultato:

Teorema 6.5.4 (TEOREMA DI DESARGUES). *In un piano affine \mathbb{A} siano dati due triangoli PQR e $P'Q'R'$, con vertici a due a due distinti. Se la retta passante per P e Q è parallela alla retta per P' e Q' , la retta per Q e R è parallela a quella passante per Q' e R' e la retta per P e R è parallela a quella per P' e R' , allora le rette passanti per P e P' , per Q e Q' e per R e R' rispettivamente, sono parallele oppure hanno un punto in comune.*

Dimostrazione. Indichiamo con r la retta passante per P e P' , con s quella passante per Q e Q' e con t quella per R e R' . Supponiamo che le rette r , s e t non siano parallele. Allora due di esse, non è restrittivo supporre che siano r e s , si intersecano in un punto che indichiamo con O .



Per il Teorema di Talete, applicato alle rette parallele passanti per P , Q e per P' , Q' , tagliate dalle trasversali r e s , si ha

$$\overrightarrow{OP} = \alpha \overrightarrow{OP'}, \quad \overrightarrow{OQ} = \alpha \overrightarrow{OQ'},$$

per qualche $\alpha \in K$. Consideriamo ora la retta passante per i punti O e R . Se tale retta è parallela alla retta per P' e R' , essa coincide necessariamente con la retta per P e R . Da ciò si deduce che i punti P , P' , R , R' e O sono allineati, quindi le rette r e t coincidono. In questo caso le rette r , s e t passano per il punto O , come volevasi dimostrare. Supponiamo quindi che le rette passanti per i punti O , R e per P' , R' non siano parallele; esse si intersecano dunque in un punto, che indicheremo con R'' . Se applichiamo il Teorema di Talete alle rette parallele passanti per P , R e per P' , R' , si conclude che

$$\overrightarrow{OR} = \beta \overrightarrow{OR''}, \quad \overrightarrow{OP} = \beta \overrightarrow{OP'},$$

per qualche $\beta \in K$. Dal confronto con le espressioni precedenti si deduce che $\beta = \alpha$, quindi si ha

$$\overrightarrow{OR} = \alpha \overrightarrow{OR''}.$$

Ora consideriamo le rette passanti per O , R e per Q' , R' . Se esse sono parallele, la retta passante per i punti O e R coincide necessariamente con quella passante per Q e R . Da ciò si deduce che i punti Q , Q' , R , R' e O sono allineati, quindi le rette s e t coincidono. Anche in questo caso le rette r , s e t passano per il

punto O . Supponiamo allora che le rette passanti per i punti O, R e per Q', R' non siano parallele; esse si intersecano dunque in un punto, che indicheremo con R''' . Applicando il Teorema di Talete alle rette parallele passanti per Q, R e per Q', R' , si ottiene

$$\overrightarrow{OR} = \gamma \overrightarrow{OR'''}, \quad \overrightarrow{OQ} = \gamma \overrightarrow{OQ'},$$

per qualche $\gamma \in K$. Dal confronto con le espressioni precedenti si deduce che $\gamma = \alpha$, quindi si ha

$$\overrightarrow{OR} = \alpha \overrightarrow{OR'''},$$

Da ciò segue che i punti R'' e R''' coincidono, ma ciò è possibile solo se entrambi coincidono con il punto R' . Si ha pertanto $R' = R'' = R'''$, il che dimostra che anche la retta t passa per il punto O . \square

Osservazione 6.5.5. I teoremi di Pappo e Desargues ammettono altre formulazioni, diverse da quelle che abbiamo presentato. In particolare, essi si possono riformulare nell'ambito della geometria proiettiva. L'importanza di tali risultati è dovuta, in modo particolare, alla relazione che essi hanno con la caratterizzazione degli spazi affini per mezzo di proprietà di natura grafica.

Terminiamo questa sezione con una discussione relativa alla costruzione del punto medio di un segmento e, più in generale, del baricentro di un insieme finito di punti.

Definizione 6.5.6. Sia K un campo ordinato⁴ e sia \mathbb{A} uno spazio affine su K . Dati due punti $A, B \in \mathbb{A}$, il *segmento* di estremi A e B è l'insieme dei punti

$$X = A + \lambda(B - A),$$

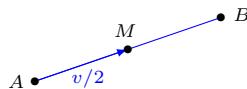
al variare di $\lambda \in K$, con $0 \leq \lambda \leq 1$.

Definizione 6.5.7. Sia K un campo ordinato e sia \mathbb{A} uno spazio affine sul campo K . Siano A e B due punti di \mathbb{A} e poniamo $v = B - A$. Il punto

$$M = A + \frac{1}{2}v$$

è detto il *punto medio* del segmento AB .

Si ha infatti $M - A = B - M = \frac{1}{2}v$.



Formalmente si può dunque scrivere

$$M = A + \frac{1}{2}(B - A) = \frac{A + B}{2}.$$

⁴Un *campo ordinato* è un campo K dotato di una relazione d'ordine totale \leq che soddisfa le seguenti proprietà:

- (i) $a \leq b \implies a + c \leq b + c$, per ogni $c \in K$;
- (ii) per ogni $a, b \in K$ con $a > 0$ e $b > 0$, si ha $ab > 0$.

Due esempi classici di campi ordinati sono dati dal campo \mathbb{Q} dei numeri razionali e dal campo \mathbb{R} dei numeri reali, dotati della relazione d'ordine usuale.

L'ultima espressione va intesa in senso puramente formale dato che, in uno spazio affine, la *somma di due punti* non è un'operazione definita. Naturalmente, in termini di coordinate, le due espressioni

$$A + \frac{1}{2}(B - A) \quad \text{e} \quad \frac{A + B}{2}$$

forniscono lo stesso risultato. Osserviamo infine che, in coordinate, la scrittura

$$M = \frac{A + B}{2}$$

è compatibile con la convenzione descritta nell'Osservazione 6.1.10. Infatti, se $A = (1, a_1, \dots, a_n)$ e $B = (1, b_1, \dots, b_n)$, si ha

$$\frac{A + B}{2} = \left(1, \frac{a_1 + b_1}{2}, \dots, \frac{a_n + b_n}{2}\right)$$

che rappresenta effettivamente un punto dello spazio affine, dato che la prima coordinata è 1.

Osservazione 6.5.8. Se il campo K non è ordinato non ha più senso parlare del “segmento” di estremi A e B . Tuttavia, se la caratteristica di K è diversa da 2, è comunque possibile costruire il punto

$$M = \frac{A + B}{2}.$$

In tal caso M è detto il *baricentro* dell'insieme di punti $\{A, B\}$.

Prendendo spunto da quanto detto sopra, possiamo dare la seguente definizione:

Definizione 6.5.9. Sia K un campo di caratteristica zero e sia $\mathbb{A} = \mathbb{A}_K^n$. Dati r punti $A_1, A_2, \dots, A_r \in \mathbb{A}$, il punto G le cui coordinate sono date da

$$G = \frac{A_1 + A_2 + \dots + A_r}{r}$$

è detto il *baricentro*⁵ dell'insieme di punti $\{A_1, A_2, \dots, A_r\}$.

Come applicazione delle tecniche finora introdotte, dimostriamo il seguente risultato di geometria elementare:

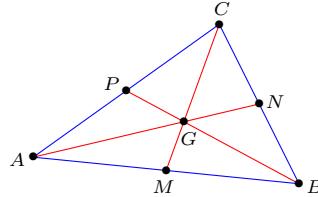
Proposizione 6.5.10. *Sia K un campo ordinato.⁶ Nello spazio affine \mathbb{A}_K^n , il baricentro di un triangolo è il punto d'incontro delle tre mediane. Inoltre, ogni mediana viene divisa da tale punto in due parti, delle quali una è doppia dell'altra.*

⁵La nozione di *baricentro* di un insieme di punti deriva dalla fisica. Più in generale, se A_1, A_2, \dots, A_r rappresentano dei punti materiali di masse rispettivamente m_1, m_2, \dots, m_r , il loro baricentro è il punto le cui coordinate sono date dalla seguente espressione:

$$G = \frac{m_1 A_1 + m_2 A_2 + \dots + m_r A_r}{m_1 + m_2 + \dots + m_r}.$$

⁶In effetti è sufficiente richiedere che K sia un campo di caratteristica $\neq 2, 3$.

Dimostrazione. Consideriamo un triangolo di vertici A, B e C nello spazio affine \mathbb{A}_K^n . Indichiamo con M, N, P i punti medi dei segmenti AB, BC e CA , rispettivamente, e indichiamo con G il baricentro dei tre punti A, B e C .



Si ha dunque

$$M = \frac{A+B}{2}, \quad N = \frac{B+C}{2}, \quad P = \frac{A+C}{2}, \quad G = \frac{A+B+C}{3}.$$

Le tre mediane del triangolo ABC sono le rette AN, BP e CM , le cui equazioni parametriche sono:

$$\begin{aligned} \text{retta } AN : & \quad X = A + \alpha(N - A) \\ \text{retta } BP : & \quad X = B + \beta(P - B) \\ \text{retta } CM : & \quad X = C + \gamma(M - C). \end{aligned}$$

È ora immediato verificare che, per $\alpha = \beta = \gamma = 2/3$, le tre equazioni precedenti forniscono precisamente le coordinate del baricentro G . Ad esempio, considerando l'equazione della retta AN , si ha infatti

$$A + \frac{2}{3}(N - A) = A + \frac{2}{3}\left(\frac{B+C}{2} - A\right) = \frac{A+B+C}{3}.$$

Ciò dimostra che il punto G giace sulle tre mediane. Per dimostrare l'ultima asserzione basta osservare che

$$\overrightarrow{AG} = G - A = \frac{A+B+C}{3} - A = \frac{B+C-2A}{3}$$

mentre

$$\overrightarrow{GN} = N - G = \frac{B+C}{2} - \frac{A+B+C}{3} = \frac{B+C-2A}{6},$$

quindi $\overrightarrow{AG} = 2\overrightarrow{GN}$ (per le altre due mediane il ragionamento è analogo). \square

6.6 Applicazioni affini

In questa sezione studieremo le funzioni tra due spazi affini che “rispettano” la struttura di spazio affine. Tali funzioni saranno dette *applicazioni affini*.

Definizione 6.6.1. Siano $\mathbb{A} = (\mathcal{A}, V, +_{\mathbb{A}})$ e $\mathbb{A}' = (\mathcal{A}', V', +_{\mathbb{A}'})$ due spazi affini sul campo K . Un'applicazione affine $F : \mathbb{A} \rightarrow \mathbb{A}'$ è il dato di una funzione tra gli insiemi di punti $f : \mathcal{A} \rightarrow \mathcal{A}'$ e di una funzione lineare tra gli spazi vettoriali $\phi : V \rightarrow V'$ che soddisfano la seguente proprietà:

$$f(P +_{\mathbb{A}} v) = f(P) +_{\mathbb{A}'} \phi(v),$$

per ogni $P \in \mathcal{A}$ e ogni $v \in V$. L'applicazione ϕ è detta l'*applicazione lineare soggiacente* all'applicazione affine F .

Per ogni spazio affine $\mathbb{A} = (\mathcal{A}, V, +_{\mathbb{A}})$ è immediato verificare che l'identità $I = (\text{id}_{\mathcal{A}}, \text{id}_V) : \mathbb{A} \rightarrow \mathbb{A}$ è un'applicazione affine.

Se $\mathbb{A} = (\mathcal{A}, V, +_{\mathbb{A}})$, $\mathbb{A}' = (\mathcal{A}', V', +_{\mathbb{A}'})$ e $\mathbb{A}'' = (\mathcal{A}'', V'', +_{\mathbb{A}''})$ sono tre spazi affini su K e se $F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{A}'$ e $G = (g, \psi) : \mathbb{A}' \rightarrow \mathbb{A}''$ sono due applicazioni affini, l'applicazione composta

$$G \circ F = (g \circ f, \psi \circ \phi) : \mathbb{A} \rightarrow \mathbb{A}''$$

è anch'essa un'applicazione affine. Si ha infatti

$$\begin{aligned} (g \circ f)(P + v) &= g(f(P + v)) \\ &= g(f(P) + \phi(v)) \\ &= g(f(P)) + \psi(\phi(v)) \\ &= (g \circ f)(P) + (\psi \circ \phi)(v). \end{aligned}$$

Infine, data un'applicazione affine $F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{A}'$, se le funzioni f e ϕ sono biettive è possibile definire l'applicazione $F^{-1} = (f^{-1}, \phi^{-1}) : \mathbb{A}' \rightarrow \mathbb{A}$, inversa di F . Questa è un'applicazione affine: infatti, se poniamo $Q = f(P)$ e $w = \phi(v)$, si ha

$$\begin{aligned} f^{-1}(Q + w) &= f^{-1}(f(P) + \phi(v)) \\ &= f^{-1}(f(P + v)) \\ &= P + v \\ &= f^{-1}(Q) + \phi^{-1}(w). \end{aligned}$$

Osservazione 6.6.2. Se $F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{A}'$ è un'applicazione affine, le due funzioni f e ϕ sono strettamente collegate tra loro. Infatti, per ogni coppia di punti $P, Q \in \mathcal{A}$, posto $v = Q - P$, si ha $Q = P + v$, quindi $f(Q) = f(P) + \phi(v)$, da cui segue che $\phi(v) = f(Q) - f(P)$. La funzione lineare $\phi : V \rightarrow V'$ è dunque completamente determinata dalla funzione insiemistica $f : \mathcal{A} \rightarrow \mathcal{A}'$.

Data una funzione insiemistica $f : \mathcal{A} \rightarrow \mathcal{A}'$ tale che, per ogni $P, Q, P', Q' \in \mathcal{A}$, si abbia

$$Q - P = Q' - P' \implies f(Q) - f(P) = f(Q') - f(P'), \quad (6.6.1)$$

è possibile definire un'applicazione $\phi : V \rightarrow V'$ ponendo, per ogni $v \in V$,

$$\phi(v) = f(P + v) - f(P),$$

per qualche $P \in \mathcal{A}$. La proprietà (6.6.1) garantisce che tale definizione non dipende dalla scelta del punto P . Se poi f è tale che la funzione ϕ così definita è lineare, la coppia (f, ϕ) definisce un'applicazione affine $F : \mathbb{A} \rightarrow \mathbb{A}'$.

Mostriamo ora che, dati due spazi affini $\mathbb{A} = (\mathcal{A}, V, +_{\mathbb{A}})$ e $\mathbb{A}' = (\mathcal{A}', V', +_{\mathbb{A}'})$ sul campo K , per ogni funzione lineare $\phi : V \rightarrow V'$ esiste un'applicazione affine $F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{A}'$ che invia un qualsiasi punto prefissato $P \in \mathcal{A}$ in un qualsiasi punto $P' \in \mathcal{A}'$.

Proposizione 6.6.3. *Siano $\mathbb{A} = (\mathcal{A}, V, +_{\mathbb{A}})$ e $\mathbb{A}' = (\mathcal{A}', V', +_{\mathbb{A}'})$ due spazi affini sul campo K e sia $\phi : V \rightarrow V'$ una funzione lineare. Per ogni $P \in \mathcal{A}$ e ogni $P' \in \mathcal{A}'$ esiste un'unica applicazione affine $F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{A}'$ tale che $f(P) = P'$.*

Dimostrazione. Fissato $P \in \mathcal{A}$, per ogni punto $Q \in \mathcal{A}$ sia $v_Q = Q - P$. Possiamo definire una funzione $f : \mathcal{A} \rightarrow \mathcal{A}'$ ponendo, per ogni $Q \in \mathcal{A}$,

$$f(Q) = P' + \phi(v_Q).$$

Dato che $v_P = P - P = \mathbf{0}$, si ha $f(P) = P' + \phi(v_P) = P'$. La funzione $F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{A}'$ così definita è un'applicazione affine: infatti, dati $Q \in \mathcal{A}$ e $v \in V$, ponendo $R = Q + v$, si ha

$$\begin{aligned} f(R) &= P' + \phi(v_R) \\ &= P' + \phi(Q + v - P) \\ &= P' + \phi(v_Q + v) \\ &= P' + \phi(v_Q) + \phi(v) \\ &= f(Q) + \phi(v). \end{aligned}$$

Per dimostrare l'unicità di F , supponiamo che $G = (g, \phi) : \mathbb{A} \rightarrow \mathbb{A}'$ sia anch'essa un'applicazione affine tale che $g(P) = P'$. Per ogni $Q \in \mathcal{A}$ si ha

$$g(Q) = g(P + v_Q) = g(P) + \phi(v_Q) = P' + \phi(v_Q) = f(Q),$$

pertanto $f = g$ e quindi $F = G$. \square

Le applicazioni affini trasformano sottospazi affini del loro dominio in sottospazi affini del codominio. Più precisamente, vale il seguente risultato:

Proposizione 6.6.4. *Siano $\mathbb{A} = (\mathcal{A}, V, +_{\mathbb{A}})$ e $\mathbb{A}' = (\mathcal{A}', V', +_{\mathbb{A}'})$ due spazi affini sul campo K e sia $F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{A}'$ un'applicazione affine.*

- (i) *Per ogni sottospazio affine $\mathbb{B} = (\mathcal{B}, W, +)$ di \mathbb{A} , ponendo $\mathcal{B}' = f(\mathcal{B})$ e $W' = \phi(W)$, si ottiene un sottospazio affine $\mathbb{B}' = (\mathcal{B}', W', +)$ di \mathbb{A}' . Tale sottospazio affine è detto l'immagine tramite F del sottospazio \mathbb{B} , e sarà indicato con $F(\mathbb{B})$.*
- (ii) *Per ogni sottospazio affine $\mathbb{B}' = (\mathcal{B}', W', +)$ di \mathbb{A}' , ponendo $\mathcal{B} = f^{-1}(\mathcal{B}')$ e $W = \phi^{-1}(W')$, si ottiene un sottospazio affine $\mathbb{B} = (\mathcal{B}, W, +)$ di \mathbb{A} . Tale sottospazio affine è detto l'immagine inversa tramite F del sottospazio \mathbb{B}' , e sarà indicato con $F^{-1}(\mathbb{B}')$.*

Dimostrazione. (i) Osserviamo innanzitutto che $W' = \phi(W)$ è un sottospazio vettoriale di V' . Dimostriamo ora che l'operazione $+_{\mathbb{A}'} : \mathcal{A}' \times V' \rightarrow \mathcal{A}'$ induce, per restrizione, un'operazione

$$+ : \mathcal{B}' \times W' \rightarrow \mathcal{B}'.$$

Siano dunque $P' \in \mathcal{B}'$, $w' \in W'$ e poniamo $Q' = P' + w' \in \mathcal{A}'$. Dobbiamo dimostrare che $Q' \in \mathcal{B}'$. Siano $P \in \mathcal{B}$ e $w \in W$ tali che $P' = f(P)$ e $w' = \phi(w)$. Il punto $Q = P + w$ appartiene a \mathcal{B} , dato che \mathbb{B} è uno spazio affine. Poiché F è un'applicazione affine, si ha $f(Q) = f(P) + \phi(w) = P' + w' = Q'$, quindi $Q' \in \mathcal{B}'$. È ora del tutto evidente che la terna $\mathbb{B}' = (\mathcal{B}', W', +)$ soddisfa le proprietà (i) e (ii) della Definizione 6.1.1. Dimostriamo quindi che vale anche la proprietà (iii). Dati due punti $P', Q' \in \mathcal{B}'$, esiste un unico vettore $w' \in V'$ tale che $Q' = P' + w'$: dobbiamo solo dimostrare che $w' \in W'$, cioè che esiste un vettore $w \in W$ tale che $\phi(w) = w'$. Siano $P, Q \in \mathcal{B}$ tali che $f(P) = P'$ e

$f(Q) = Q'$ e poniamo $w = Q - P$; notiamo che $w \in W$ perché \mathbb{B} è un sottospazio affine. Si ha dunque $Q = P + w$ e quindi $Q' = f(Q) = f(P) + \phi(w) = P' + \phi(w)$, da cui si deduce che $\phi(w) = Q' - P' = w'$, come volevasi dimostrare.

(ii) Iniziamo con l'osservare che $W = \phi^{-1}(W')$ è un sottospazio vettoriale di V . Dimostriamo ora che l'operazione $+_{\mathbb{A}} : \mathcal{A} \times V \rightarrow \mathcal{A}$ induce, per restrizione, un'operazione

$$+ : \mathcal{B} \times W \rightarrow \mathcal{B}.$$

Siano dunque $P \in \mathcal{B}$, $w \in W$ e poniamo $Q = P + w \in \mathcal{A}$. Dobbiamo dimostrare che $Q \in \mathcal{B}$, cioè che $Q' = f(Q) \in \mathcal{B}'$. Ponendo $P' = f(P)$ e $w' = \phi(w)$, si ha

$$Q' = f(Q) = f(P) + \phi(w) = P' + w' \in \mathcal{B}',$$

dato che \mathbb{B}' è un sottospazio affine. È ora del tutto evidente che la terna $\mathbb{B} = (\mathcal{B}, W, +)$ soddisfa le proprietà (i) e (ii) della Definizione 6.1.1. Dimostriamo quindi che vale anche la proprietà (iii). Dati due punti $P, Q \in \mathcal{B}$, esiste un unico vettore $w \in V$ tale che $Q = P + w$: dobbiamo solo dimostrare che $w \in W$, cioè che $w' = \phi(w) \in W'$. Ponendo $P' = f(P)$ e $w' = \phi(w)$, si ha

$$Q' = f(Q) = f(P) + \phi(w) = P' + w',$$

quindi $w' = Q' - P' \in W'$, dato che \mathbb{B}' è un sottospazio affine. \square

Definizione 6.6.5. Un'applicazione affine $F : \mathbb{A} \rightarrow \mathbb{A}'$ è un *isomorfismo di spazi affini* se esiste un'applicazione affine $G : \mathbb{A}' \rightarrow \mathbb{A}$ tale che le applicazioni composte $F \circ G$ e $G \circ F$ siano l'identità. Un'isomorfismo di spazi affini di \mathbb{A} in sé stesso è detto un'*affinità* di \mathbb{A} .

Osservazione 6.6.6. Dato uno spazio affine \mathbb{A} , indicheremo con $\text{Aff}(\mathbb{A})$ l'insieme delle affinità di \mathbb{A} . Ricordando che la composizione di due affinità è un'affinità e che l'inversa di un'affinità è un'affinità, è immediato verificare che $\text{Aff}(\mathbb{A})$ è un gruppo per la legge di composizione delle applicazioni.

Notiamo che, per quanto visto in precedenza, un'applicazione affine $F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{A}'$ è un'isomorfismo se e solo se le funzioni f e ϕ sono biettive. In effetti, come ora dimostreremo, per un'applicazione affine $F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{A}'$ l'iniettività (risp., la suriettività) dell'applicazione insiemistica f equivale all'iniettività (risp., alla suriettività) dell'applicazione lineare ϕ .

Proposizione 6.6.7. Sia $F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{A}'$ un'applicazione affine. Valgono le seguenti proprietà:

- (i) f è iniettiva se e solo se ϕ è iniettiva;
- (ii) f è suriettiva se e solo se ϕ è suriettiva.

Dimostrazione. (i) Supponiamo che $f : \mathcal{A} \rightarrow \mathcal{A}'$ sia iniettiva. Siano $v_1, v_2 \in V$ tali che $\phi(v_1) = \phi(v_2)$. Dato un punto $P \in \mathcal{A}$, si ha:

$$f(P + v_1) = f(P) + \phi(v_1) = f(P) + \phi(v_2) = f(P + v_2).$$

Dall'iniettività di f discende allora che $P + v_1 = P + v_2$, da cui segue $v_1 = v_2$.

Viceversa, supponiamo che $\phi : V \rightarrow V'$ sia iniettiva. Dati $P, Q \in \mathcal{A}$ con $f(P) = f(Q)$, poniamo $v = Q - P$. Si ha dunque

$$f(P) = f(Q) = f(P + v) = f(P) + \phi(v),$$

da cui si deduce che $\phi(v) = \mathbf{0}$. Dall'iniettività di ϕ segue che $v = \mathbf{0}$, quindi $P = Q$.

(ii) Supponiamo ora che $f : \mathcal{A} \rightarrow \mathcal{A}'$ sia suriettiva. Dato $v' \in V'$, scegliamo un punto $P' \in \mathcal{A}'$ e poniamo $Q' = P' + v'$. Poiché f è suriettiva, esistono due punti $P, Q \in \mathcal{A}$ tali che $f(P) = P'$ e $f(Q) = Q'$. Sia dunque $v = Q - P$: vogliamo dimostrare che $\phi(v) = v'$. Si ha infatti

$$Q' = f(Q) = f(P + v) = f(P) + \phi(v) = P' + \phi(v),$$

da cui segue che $\phi(v) = Q' - P' = v'$, come volevasi dimostrare.

Viceversa, supponiamo che $\phi : V \rightarrow V'$ sia suriettiva. Dato $P' \in \mathcal{A}'$ scegliamo un punto $A \in \mathcal{A}$, poniamo $A' = f(A)$ e $v' = P' - A'$. Dato che ϕ è suriettiva, esiste $v \in V$ tale che $\phi(v) = v'$. Ponendo $P = A + v$, si ha

$$f(P) = f(A + v) = f(A) + \phi(v) = A' + v' = P',$$

il che dimostra la suriettività di f . \square

Come applicazione di quanto abbiamo visto finora, possiamo dimostrare il seguente risultato:

Proposizione 6.6.8. *Siano $\mathbb{A} = (\mathcal{A}, V, +)$ e $\mathbb{A}' = (\mathcal{A}', V', +)$ due spazi affini di dimensione n sul campo K . Dati due sistemi di riferimento*

$$\mathcal{R} = \{P_0, P_1, \dots, P_n\}, \quad \mathcal{R}' = \{P'_0, P'_1, \dots, P'_n\},$$

in \mathbb{A} e \mathbb{A}' rispettivamente, esiste un unico isomorfismo di spazi affini

$$F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{A}'$$

tale che $f(P_i) = P'_i$, per ogni $i = 0, \dots, n$.

Dimostrazione. Poiché \mathcal{R} è un sistema di riferimento in \mathbb{A} , i vettori

$$v_1 = P_1 - P_0, v_2 = P_2 - P_0, \dots, v_n = P_n - P_0$$

costituiscono una base dello spazio vettoriale V . In modo del tutto analogo, i vettori

$$v'_1 = P'_1 - P'_0, v'_2 = P'_2 - P'_0, \dots, v'_n = P'_n - P'_0$$

sono una base dello spazio vettoriale V' , dato che \mathcal{R}' è un sistema di riferimento in \mathbb{A}' . Esiste pertanto un'unica applicazione lineare $\phi : V \rightarrow V'$ tale che $\phi(v_i) = v'_i$, per $i = 1, \dots, n$; inoltre tale applicazione lineare è un isomorfismo di spazi vettoriali. Dalla Proposizione 6.6.3 si deduce che esiste un'unica applicazione affine $F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{A}'$ tale che $f(P_0) = P'_0$; tale applicazione è un isomorfismo di spazi affini, visto che ϕ è un isomorfismo di spazi vettoriali. Notiamo infine che, per ogni $i = 1, \dots, n$, si ha $P_i = P_0 + v_i$ quindi, poiché F è un'applicazione affine, si ha

$$f(P_i) = f(P_0) + \phi(v_i) = P'_0 + v'_i = P'_i.$$

F soddisfa pertanto le proprietà richieste. \square

Terminiamo questa sezione dimostrando che ogni spazio affine di dimensione n sul campo K è isomorfo (non in modo canonico) allo spazio affine standard \mathbb{A}_K^n .

Sia dunque $\mathbb{A} = (\mathcal{A}, V, +)$ uno spazio affine di dimensione n sul campo K e sia $\mathcal{R} = \{O, v_1, \dots, v_n\}$ un sistema di riferimento in \mathbb{A} . Ad ogni punto $P \in \mathcal{A}$ è possibile associare la n -upla (p_1, \dots, p_n) delle sue coordinate nel sistema di riferimento \mathcal{R} ; ricordiamo che ciò significa che P si scrive nella forma

$$P = O + p_1 v_1 + \dots + p_n v_n.$$

Come già osservato nel Paragrafo 6.3, risulta così definita una biezione

$$f = f_{\mathcal{R}} : \mathcal{A} \rightarrow K^n, \quad P \mapsto f(P) = (p_1, \dots, p_n).$$

Analogamente, associando ad ogni vettore $v \in V$ la n -upla $(\alpha_1, \dots, \alpha_n)$ delle sue componenti rispetto alla base v_1, \dots, v_n , si ottiene un isomorfismo di spazi vettoriali

$$\phi = \phi_{\mathcal{R}} : V \rightarrow K^n.$$

Possiamo ora dimostrare il seguente risultato:

Proposizione 6.6.9. *Con le notazioni precedenti, la coppia $(f_{\mathcal{R}}, \phi_{\mathcal{R}})$ definisce un isomorfismo di spazi affini $\Phi_{\mathcal{R}} : \mathbb{A} \xrightarrow{\sim} \mathbb{A}_K^n$.*

Dimostrazione. Per dimostrare che $\Phi_{\mathcal{R}} = (f_{\mathcal{R}}, \phi_{\mathcal{R}})$ è un'applicazione affine consideriamo un punto $P \in \mathbb{A}$ e un vettore $v \in V$. Poniamo $(p_1, \dots, p_n) = f_{\mathcal{R}}(P)$ e $(\alpha_1, \dots, \alpha_n) = \phi_{\mathcal{R}}(v)$: si ha dunque

$$P = O + p_1 v_1 + \dots + p_n v_n, \quad v = \alpha_1 v_1 + \dots + \alpha_n v_n.$$

Da ciò segue che

$$P + v = O + (p_1 + \alpha_1) v_1 + \dots + (p_n + \alpha_n) v_n$$

e dunque le coordinate del punto $P + v$ nel sistema di riferimento \mathcal{R} sono date da $(p_1 + \alpha_1, \dots, p_n + \alpha_n)$. Si ha pertanto:

$$f_{\mathcal{R}}(P + v) = (p_1 + \alpha_1, \dots, p_n + \alpha_n) = f_{\mathcal{R}}(P) + \phi_{\mathcal{R}}(v),$$

il che dimostra che $\Phi_{\mathcal{R}} = (f_{\mathcal{R}}, \phi_{\mathcal{R}})$ è un'applicazione affine. Dalla biettività delle funzioni $f_{\mathcal{R}}$ e $\phi_{\mathcal{R}}$ si deduce che $\Phi_{\mathcal{R}}$ è un isomorfismo. \square

6.6.1 Matrici associate alle applicazioni affini

Siano $\mathbb{A} = (\mathcal{A}, V, +)$ e $\mathbb{B} = (\mathcal{B}, W, +)$ due spazi affini sul campo K , di dimensioni n e m rispettivamente, e sia $F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{B}$ un'applicazione affine. Indichiamo con $\mathcal{R} = \{O_{\mathbb{A}}, v_1, \dots, v_n\}$ un sistema di riferimento in \mathbb{A} e con $\mathcal{S} = \{O_{\mathbb{B}}, w_1, \dots, w_m\}$ un sistema di riferimento in \mathbb{B} .

Come abbiamo visto nel Paragrafo 6.3, ogni punto $P \in \mathbb{A}$ si può scrivere, in modo unico, nella forma

$$P = O_{\mathbb{A}} + \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n,$$

ove $(\lambda_1, \lambda_2, \dots, \lambda_n) \in K^n$ sono le coordinate di P nel sistema di riferimento \mathcal{R} . Dalla definizione di applicazione affine si deduce allora che

$$f(P) = f(O_{\mathbb{A}}) + \lambda_1 \phi(v_1) + \lambda_2 \phi(v_2) + \dots + \lambda_n \phi(v_n).$$

Se indichiamo con $A = (a_{ij})$ la matrice dell'applicazione lineare $\phi : V \rightarrow W$ rispetto alle basi $\{v_1, \dots, v_n\}$ di V e $\{w_1, \dots, w_m\}$ di W , si ha

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

per ogni $j = 1, \dots, n$. Indicando poi con $(t_1, \dots, t_m) \in K^m$ le coordinate del punto $f(O_{\mathbb{A}})$ nel sistema di riferimento \mathcal{S} , si ha

$$f(O_{\mathbb{A}}) = O_{\mathbb{B}} + t_1 w_1 + t_2 w_2 + \dots + t_m w_m$$

e quindi

$$\begin{aligned} f(P) &= O_{\mathbb{B}} + \sum_{i=1}^m t_i w_i + \sum_{j=1}^n \lambda_j \phi(v_j) \\ &= O_{\mathbb{B}} + \sum_{i=1}^m t_i w_i + \sum_{j=1}^n \lambda_j \sum_{i=1}^m a_{ij} w_i \\ &= O_{\mathbb{B}} + \sum_{i=1}^m t_i w_i + \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij} \lambda_j \right) w_i \\ &= O_{\mathbb{B}} + \sum_{i=1}^m \left(t_i + \sum_{j=1}^n a_{ij} \lambda_j \right) w_i. \end{aligned}$$

L'applicazione affine $F = (f, \phi)$ è dunque completamente determinata dalle coordinate (t_1, \dots, t_m) del punto $f(O_{\mathbb{A}})$ e dalla matrice $A = (a_{ij})$ dell'applicazione lineare ϕ . Indicando con (x_1, \dots, x_n) le coordinate di un generico punto $X \in \mathbb{A}$ e con (x'_1, \dots, x'_m) le coordinate del punto $f(X) \in \mathbb{B}$, la formula precedente si riscrive come segue:

$$\begin{pmatrix} x'_1 \\ \vdots \\ x'_m \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} t_1 \\ \vdots \\ t_m \end{pmatrix} \quad (6.6.2)$$

Quest'ultima è dunque l'espressione, in coordinate, di una generica applicazione affine F tra due spazi affini \mathbb{A} e \mathbb{B} , di dimensioni n e m rispettivamente.

Osservazione 6.6.10. Se rappresentiamo le coordinate dei punti X e $X' = f(X)$ nella forma $X = (1, x_1, \dots, x_n)$ e $X' = (1, x'_1, \dots, x'_m)$ (vedi la convenzione introdotta nell'Osservazione 6.1.10) e se associamo alla matrice $A = (a_{ij})$ e alla m -upla (t_1, \dots, t_m) la matrice

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ t_1 & a_{11} & a_{12} & \dots & a_{1n} \\ t_2 & a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ t_m & a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

possiamo riscrivere la formula (6.6.2) come segue:

$$\begin{pmatrix} 1 \\ x'_1 \\ \vdots \\ x'_m \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ t_1 & a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ t_m & a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{pmatrix} \quad (6.6.3)$$

È dunque possibile rappresentare un'applicazione affine $F = (f, \phi)$ tra due spazi affini \mathbb{A} e \mathbb{B} , di dimensioni n e m rispettivamente, mediante una matrice $(m+1) \times (n+1)$ del tipo

$$\left(\begin{array}{c|c} 1 & \mathbf{0} \\ \mathbf{t} & A \end{array} \right)$$

ove $\mathbf{t} = (t_1, \dots, t_m) \in K^m$ sono le coordinate dell'immagine tramite f dell'origine del sistema di riferimento fissato in \mathbb{A} , $A \in M_{m,n}(K)$ è la matrice dell'applicazione lineare ϕ rispetto alle basi $\{v_1, \dots, v_n\}$ di V e $\{w_1, \dots, w_m\}$ di W e $\mathbf{0} = (0, \dots, 0) \in K^n$ è il vettore nullo (scritto in riga).

Osservazione 6.6.11. Prendendo spunto dalla formula (6.6.2) possiamo definire, per ogni matrice $A \in M_{m,n}(K)$ e ogni m -upla $\mathbf{t} = (t_1, \dots, t_m) \in K^m$, un'applicazione insiemistica $f : K^n \rightarrow K^m$ ponendo $f(X) = AX + \mathbf{t}$ e un'applicazione lineare $\phi : K^n \rightarrow K^m$ ponendo $\phi(v) = Av$. Per ogni punto $P \in \mathbb{A}_K^n$ e ogni vettore $v \in K^n$ si ha

$$f(P+v) = A(P+v) + \mathbf{t} = AP + \mathbf{t} + Av = f(P) + \phi(v),$$

quindi la coppia (f, ϕ) definisce un'applicazione affine

$$F = F_{(\mathbf{t}, A)} : \mathbb{A}_K^n \rightarrow \mathbb{A}_K^m.$$

È immediato verificare che $F_{(\mathbf{t}, A)}$ è un isomorfismo se e solo se la matrice A è invertibile: infatti sappiamo che $F_{(\mathbf{t}, A)} = (f, \phi)$ è un isomorfismo di spazi affini se e solo se ϕ è un isomorfismo di spazi vettoriali, il che accade precisamente se e solo se A è invertibile.

Osservazione 6.6.12. Sia $F : \mathbb{A}_K^n \rightarrow \mathbb{A}_K^m$, $F(X) = AX + \mathbf{t}$, un'applicazione affine e sia \mathbb{L} la sottovarietà lineare di \mathbb{A}_K^m costituita dalle soluzioni di un sistema di equazioni lineari $CY = D$. Nella Proposizione 6.6.4 abbiamo dimostrato che l'immagine inversa di \mathbb{L} tramite F è una sottovarietà lineare di \mathbb{A}_K^n :

$$F^{-1}(\mathbb{L}) = \{X \in \mathbb{A}_K^n \mid F(X) \in \mathbb{L}\}.$$

Vogliamo determinare il sistema lineare corrispondente alla sottovarietà $F^{-1}(\mathbb{L})$. A tale scopo basta osservare che, dalle definizioni date, si ha

$$\begin{aligned} X \in F^{-1}(\mathbb{L}) &\iff F(X) \in \mathbb{L} \\ &\iff AX + \mathbf{t} \in \mathbb{L} \\ &\iff C(AX + \mathbf{t}) = D \\ &\iff (CA)X = D - C\mathbf{t}. \end{aligned}$$

Si conclude quindi che la sottovarietà lineare $F^{-1}(\mathbb{L})$ di \mathbb{A}_K^n è costituita dalle soluzioni del sistema di equazioni lineari $EX = G$, ove $E = CA$ e $G = D - C\mathbf{t}$.

Nel caso in cui $F : \mathbb{A}_K^n \rightarrow \mathbb{A}_K^m$ sia un isomorfismo di spazi affini (e quindi $m = n$), l'applicazione inversa $F^{-1} : \mathbb{A}_K^m \rightarrow \mathbb{A}_K^n$ è data da $F^{-1}(Y) = A^{-1}Y - A^{-1}\mathbf{t}$. Se \mathbb{L}' è una sottovarietà lineare di \mathbb{A}_K^n , costituita dalle soluzioni di un sistema di equazioni lineari $C'X = D'$, la sua immagine tramite F è una sottovarietà lineare dello spazio affine \mathbb{A}_K^m e si ha

$$F(\mathbb{L}') = (F^{-1})^{-1}(\mathbb{L}').$$

Applicando quanto detto sopra alla funzione inversa di F , si deduce che il sistema di equazioni lineari corrispondente alla sottovarietà $F(\mathbb{L}')$ è $E'Y = G'$, ove $E' = C'A^{-1}$ e $G' = D' + C'A^{-1}\mathbf{t}$.

Osservazione 6.6.13. Siano $\mathbb{A} = (\mathcal{A}, V, +)$ e $\mathbb{B} = (\mathcal{B}, W, +)$ due spazi affini sul campo K , di dimensioni n e m rispettivamente. Fissiamo un sistema di riferimento $\mathcal{R} = \{O_{\mathbb{A}}, v_1, \dots, v_n\}$ in \mathbb{A} e indichiamo con $\Phi_{\mathcal{R}} : \mathbb{A} \xrightarrow{\sim} \mathbb{A}_K^n$ l'isomorfismo di spazi affini definito nella Proposizione 6.6.9. Fissiamo poi un sistema di riferimento $\mathcal{S} = \{O_{\mathbb{B}}, w_1, \dots, w_m\}$ in \mathbb{B} e indichiamo con $\Psi_{\mathcal{S}} : \mathbb{B} \xrightarrow{\sim} \mathbb{A}_K^m$ l'isomorfismo di spazi affini corrispondente. Data un'applicazione affine $F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{B}$, indichiamo con $\mathbf{t} = (t_1, \dots, t_m)$ le coordinate del punto $f(O_{\mathbb{A}})$ nel sistema di riferimento \mathcal{S} e con $A \in M_{m,n}(K)$ la matrice dell'applicazione lineare ϕ rispetto alle basi $\{v_1, \dots, v_n\}$ di V e $\{w_1, \dots, w_m\}$ di W . Sia poi $F_{(\mathbf{t}, A)} : \mathbb{A}_K^n \rightarrow \mathbb{A}_K^m$ l'applicazione affine definita nell'Osservazione precedente. Ricordando la formula (6.6.2), è immediato verificare che il seguente diagramma è commutativo:

$$\begin{array}{ccc} \mathbb{A} & \xrightarrow{F} & \mathbb{B} \\ \Phi_{\mathcal{R}} \downarrow \wr & & \downarrow \wr \Psi_{\mathcal{S}} \\ \mathbb{A}_K^n & \xrightarrow{F_{(\mathbf{t}, A)}} & \mathbb{A}_K^m \end{array}$$

Osservazione 6.6.14. Quanto visto finora permette di affermare che esiste una corrispondenza biunivoca tra l'insieme delle applicazioni affini di \mathbb{A}_K^n in \mathbb{A}_K^m e l'insieme delle coppie (\mathbf{t}, A) , ove $\mathbf{t} \in K^m$ e $A \in M_{m,n}(K)$ è una matrice $m \times n$ a coefficienti in K . Vedremo ora come si esprime, in termini di coppie (\mathbf{t}, A) , l'operazione di composizione di due applicazioni affini.

Consideriamo dunque le applicazioni affini

$$F_{(\mathbf{t}, A)} : \mathbb{A}_K^n \rightarrow \mathbb{A}_K^m \quad \text{e} \quad F_{(\mathbf{u}, B)} : \mathbb{A}_K^m \rightarrow \mathbb{A}_K^r,$$

corrispondenti alle coppie (\mathbf{t}, A) e (\mathbf{u}, B) , ove $A \in M_{m,n}(K)$, $B \in M_{r,m}(K)$, $\mathbf{t} \in K^m$ e $\mathbf{u} \in K^r$. Dato che l'applicazione composta

$$F_{(\mathbf{u}, B)} \circ F_{(\mathbf{t}, A)} : \mathbb{A}_K^n \rightarrow \mathbb{A}_K^r$$

è un'applicazione affine, essa corrisponderà a una coppia (\mathbf{v}, C) , per qualche $C \in M_{r,n}(K)$ e $\mathbf{v} \in K^r$. Per ogni punto $X \in \mathbb{A}_K^n$, si ha dunque:

$$\begin{aligned} F_{(\mathbf{v}, C)}(X) &= (F_{(\mathbf{u}, B)} \circ F_{(\mathbf{t}, A)})(X) \\ &= F_{(\mathbf{u}, B)}(AX + \mathbf{t}) \\ &= B(AX + \mathbf{t}) + \mathbf{u} \\ &= (BA)X + (B\mathbf{t} + \mathbf{u}). \end{aligned}$$

D'altra parte, si ha anche

$$F_{(\mathbf{v}, C)}(X) = CX + \mathbf{v}$$

da cui si deduce che $C = BA$ e $\mathbf{v} = \mathbf{u} + B\mathbf{t}$. Concludiamo quindi che, in termini di coppie, l'operazione di composizione di due applicazioni affini si esprime come segue:

$$(\mathbf{u}, B) \circ (\mathbf{t}, A) = (\mathbf{u} + B\mathbf{t}, BA).$$

Osservazione 6.6.15. Per ogni spazio affine \mathbb{A} di dimensione n sul campo K , la scelta di un sistema di riferimento \mathcal{R} in \mathbb{A} permette di associare ad ogni affinità F di \mathbb{A} una coppia (\mathbf{t}, A) , ove $\mathbf{t} \in K^n$ e $A \in M_n(K)$ è una matrice invertibile. Si stabilisce così una biiezione (dipendente dalla scelta di \mathcal{R}) tra l'insieme $\text{Aff}(\mathbb{A})$ e il prodotto cartesiano $K^n \times \text{GL}_n(K)$. Come abbiamo visto nell'Osservazione precedente, la legge di gruppo di $\text{Aff}(\mathbb{A})$ si traduce nella seguente legge di composizione di coppie:

$$(\mathbf{u}, B) \cdot (\mathbf{t}, A) = (\mathbf{u} + B\mathbf{t}, BA).$$

L'insieme $K^n \times \text{GL}_n(K)$, con la legge di composizione sopra descritta, è un gruppo, detto il *prodotto semidiretto* del gruppo additivo $(K^n, +)$ per il gruppo moltiplicativo $(\text{GL}_n(K), \cdot)$. Tale gruppo viene indicato con la notazione seguente:⁷

$$K^n \rtimes \text{GL}_n(K).$$

L'elemento identico per tale legge di gruppo è la coppia $(\mathbf{0}, \mathbf{1})$ (la quale corrisponde all'applicazione affine identica di \mathbb{A} in sé) mentre l'inverso di una coppia (\mathbf{t}, A) risulta essere la coppia $(-A^{-1}\mathbf{t}, A^{-1})$. Concludiamo quindi che, per ogni spazio affine \mathbb{A} di dimensione n sul campo K , il gruppo $\text{Aff}(\mathbb{A})$ delle affinità di \mathbb{A} è isomorfo (non canonicamente) al prodotto semidiretto $K^n \rtimes \text{GL}_n(K)$.

Osservazione 6.6.16. Come abbiamo visto nell'Osservazione 6.6.10, ad ogni coppia $(\mathbf{t}, A) \in K^n \rtimes \text{GL}_n(K)$ è possibile associare la matrice

$$\left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{t} & A \end{array} \right) \in \text{GL}_{n+1}(K). \quad (6.6.4)$$

Vogliamo dimostrare che tale associazione definisce un omomorfismo iniettivo di gruppi

$$j : K^n \rtimes \text{GL}_n(K) \rightarrow \text{GL}_{n+1}(K),$$

il quale permette dunque di identificare il prodotto semidiretto $K^n \rtimes \text{GL}_n(K)$ con il sottogruppo di $\text{GL}_{n+1}(K)$ costituito dalle matrici della forma (6.6.4).

A tale scopo basta osservare che si ha

$$j((\mathbf{u}, B) \circ (\mathbf{t}, A)) = j(\mathbf{u} + B\mathbf{t}, BA) = \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{u} + B\mathbf{t} & BA \end{array} \right)$$

e che tale matrice coincide con il prodotto righe per colonne delle matrici $j(\mathbf{u}, B)$ e $j(\mathbf{t}, A)$.

⁷La notazione $G = N \rtimes H$, per indicare il prodotto semidiretto di due gruppi N e H , serve a ricordare che N è un sottogruppo *normale* di G , $N \triangleleft G$, mentre H è solo un sottogruppo di G , $H < G$.

6.7 Spazi affini euclidei

In questa sezione studieremo gli spazi affini i cui spazi direttori sono degli spazi vettoriali euclidei, cioè degli spazi vettoriali reali dotati di un prodotto scalare. Come vedremo, la presenza di una tale struttura nello spazio vettoriale V soggiacente a uno spazio affine $\mathbb{A} = (\mathcal{A}, V, +)$ permetterà di introdurre in \mathbb{A} nozioni di natura metrica come distanze, angoli, aree, volumi.

Definizione 6.7.1. Uno *spazio affine euclideo* è uno spazio affine \mathbb{A} , definito sul campo \mathbb{R} dei numeri reali, in cui lo spazio direttore V è dotato di una struttura di spazio vettoriale euclideo, cioè di una forma bilineare simmetrica definita positiva $g : V \times V \rightarrow \mathbb{R}$. Commettendo un abuso di notazione, dati due vettori $v, w \in V$, nel seguito scriveremo $v \cdot w$ al posto di $g(v, w)$.

Esempio 6.7.2. L'esempio fondamentale di spazio affine euclideo è costituito dallo spazio affine $\mathbb{A}_{\mathbb{R}}^n$, in cui lo spazio vettoriale soggiacente \mathbb{R}^n è dotato del prodotto scalare usuale. Tale spazio verrà chiamato semplicemente lo spazio affine euclideo $\mathbb{A}_{\mathbb{R}}^n$.

Definizione 6.7.3. In uno spazio affine euclideo \mathbb{A} (di dimensione finita), un sistema di riferimento $\mathcal{R} = \{O, v_1, v_2, \dots, v_n\}$ è detto *ortogonale* (risp., *ortonormale*) se i vettori v_1, v_2, \dots, v_n sono una base ortogonale (risp., ortonormale) dello spazio vettoriale euclideo V . Un sistema di riferimento ortonormale è anche detto *cartesiano*.

Osservazione 6.7.4. Si noti che l'esistenza di sistemi di riferimento ortonormali in uno spazio affine euclideo è garantita dall'esistenza di basi ortonormali negli spazi vettoriali euclidei (vedi Cap. 5, Paragrafo 5.4.2). Inoltre, la scelta di un sistema di riferimento ortonormale \mathcal{R} in uno spazio affine euclideo \mathbb{A} di dimensione n , determina un isomorfismo di spazi affini (vedi Proposizione 6.6.9)

$$\Phi_{\mathcal{R}} : \mathbb{A} \xrightarrow{\sim} \mathbb{A}_{\mathbb{R}}^n$$

Mediante tale isomorfismo la forma bilineare g definita sullo spazio vettoriale V soggiacente allo spazio affine \mathbb{A} viene identificata con il prodotto scalare usuale di \mathbb{R}^n .

Dalla Definizione 6.7.1 segue che tutti i risultati ottenuti nel Capitolo 5 per gli spazi vettoriali euclidei si possono riformulare, in modo del tutto ovvio, nel contesto degli spazi affini euclidei.

La presenza di un prodotto scalare, definito nello spazio vettoriale V soggiacente a uno spazio affine euclideo \mathbb{A} , permette di definire la distanza tra due punti di \mathbb{A} :

Definizione 6.7.5. Siano \mathbb{A} uno spazio affine euclideo e P, Q due punti di \mathbb{A} . La *distanza* $d(P, Q)$ tra P e Q è la norma del vettore $Q - P$:

$$d(P, Q) = \|Q - P\| = \sqrt{(Q - P) \cdot (Q - P)}.$$

Dalle proprietà della norma di un vettore studiate nel Capitolo 5, si deduce che, in uno spazio affine euclideo $\mathbb{A} = (\mathcal{A}, V, +)$, la funzione distanza

$$d : \mathcal{A} \times \mathcal{A} \rightarrow \mathbb{R}$$

gode delle seguenti proprietà:

- (i) $d(P, Q) \geq 0$, per ogni $P, Q \in \mathcal{A}$, e $d(P, Q) = 0$ se e solo se $P = Q$;
- (ii) $d(P, Q) = d(Q, P)$, per ogni $P, Q \in \mathcal{A}$;
- (iii) $d(P, R) \leq d(P, Q) + d(Q, R)$, per ogni $P, Q, R \in \mathcal{A}$.

La proprietà (iii) è detta *disuguaglianza triangolare*: essa afferma che, in un triangolo, ogni lato è minore o uguale della somma degli altri due (vedi Cap. 5, Proposizioni 5.1.8 e 5.3.24).

Se \mathcal{S} e \mathcal{T} sono due sottoinsiemi non vuoti (dell'insieme dei punti) di uno spazio affine euclideo \mathbb{A} , risulta naturale definire la distanza di \mathcal{S} da \mathcal{T} come l'estremo inferiore delle distanze dei punti di \mathcal{S} dai punti di \mathcal{T} :

$$d(\mathcal{S}, \mathcal{T}) = \inf\{d(P, Q) \mid P \in \mathcal{S}, Q \in \mathcal{T}\}.$$

Dalla definizione data segue subito che

$$\mathcal{S} \cap \mathcal{T} \neq \emptyset \implies d(\mathcal{S}, \mathcal{T}) = 0.$$

L'osservazione seguente mostra che non vale l'implicazione opposta.

Osservazione 6.7.6. Se \mathcal{S} e \mathcal{T} sono due sottoinsiemi di uno spazio affine euclideo \mathbb{A} , non è detto che esistano dei punti $P \in \mathcal{S}$ e $Q \in \mathcal{T}$ tali che $d(P, Q) = d(\mathcal{S}, \mathcal{T})$. Ad esempio, considerando la retta affine reale $\mathbb{A}_{\mathbb{R}}^1$ e ponendo

$$\mathcal{S} = \{x \in \mathbb{R} \mid x \leq 0\}, \quad \mathcal{T} = \{x \in \mathbb{R} \mid x > 0\},$$

si ha $d(\mathcal{S}, \mathcal{T}) = 0$, ma $d(P, Q) > 0$ per ogni $P \in \mathcal{S}$ e ogni $Q \in \mathcal{T}$.

Nel caso particolare in cui i sottoinsiemi \mathcal{S} e \mathcal{T} sono (gli insiemi dei punti di) due sottospazi affini \mathbb{L} e \mathbb{M} dello spazio affine euclideo \mathbb{A} , la funzione distanza

$$d : \mathbb{L} \times \mathbb{M} \rightarrow \mathbb{R}, \quad (P, Q) \mapsto d(P, Q)$$

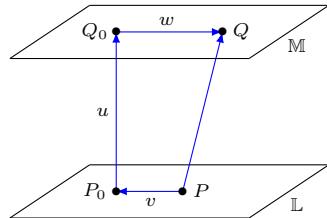
ammette minimo, cioè esistono dei punti $P_0 \in \mathbb{L}$ e $Q_0 \in \mathbb{M}$ tali che

$$d(\mathbb{L}, \mathbb{M}) = d(P_0, Q_0).$$

Prima di dimostrare questo risultato enunciamo e dimostriamo il seguente lemma:

Lemma 6.7.7. *Siano $\mathbb{L} = (\mathcal{L}, L, +)$ e $\mathbb{M} = (\mathcal{M}, M, +)$ due sottospazi affini di uno spazio affine euclideo \mathbb{A} e siano $P_0 \in \mathbb{L}$ e $Q_0 \in \mathbb{M}$ due punti tali che il vettore $u = Q_0 - P_0$ sia ortogonale a \mathbb{L} e \mathbb{M} . Allora, per ogni $P \in \mathbb{L}$ e ogni $Q \in \mathbb{M}$, si ha $d(P_0, Q_0) \leq d(P, Q)$. Inoltre, se $P_1 \in \mathbb{L}$ e $Q_1 \in \mathbb{M}$ sono due punti tali che $d(P_1, Q_1) = d(P_0, Q_0)$, deve necessariamente essere $Q_1 - P_1 = Q_0 - P_0$.*

Dimostrazione. Per ogni $P \in \mathbb{L}$ e ogni $Q \in \mathbb{M}$, si ha $Q - P = (Q - Q_0) + (Q_0 - P_0) + (P_0 - P)$, cioè $Q - P = u + v + w$, ove $u = Q_0 - P_0$, $v = P_0 - P \in L$ e $w = Q - Q_0 \in M$.



Dato che u è ortogonale a \mathbb{L} e \mathbb{M} , si ha $u \cdot v = u \cdot w = 0$. Da ciò segue che

$$\begin{aligned}\|Q - P\|^2 &= (u + v + w) \cdot (u + v + w) \\ &= \|u\|^2 + \|v\|^2 + \|w\|^2 + 2v \cdot w \\ &= \|u\|^2 + \|v + w\|^2 \\ &\geq \|u\|^2,\end{aligned}$$

e quindi $d(P, Q) \geq d(P_0, Q_0)$. Specializzando ora il ragionamento precedente al caso in cui $P = P_1$ e $Q = Q_1$, si ha

$$\|Q_1 - P_1\|^2 = \|u\|^2 + \|v + w\|^2 = \|Q_0 - P_0\|^2 + \|v + w\|^2.$$

Poiché, per ipotesi, si ha $d(P_1, Q_1) = d(P_0, Q_0)$, deve essere $\|Q_1 - P_1\|^2 = \|Q_0 - P_0\|^2$ e quindi $\|v + w\|^2 = 0$. Da ciò si deduce che $v + w = \mathbf{0}$ e dunque $w = -v$. Si ha pertanto $Q_1 - P_1 = u + v + w = u + v - v = u = Q_0 - P_0$, come volevasi dimostrare. \square

Proposizione 6.7.8. *Siano $\mathbb{L} = (\mathcal{L}, L, +)$ e $\mathbb{M} = (\mathcal{M}, M, +)$ due sottospazi affini (non vuoti) di uno spazio affine euclideo \mathbb{A} . Allora esistono dei punti $P_0 \in \mathbb{L}$ e $Q_0 \in \mathbb{M}$ tali che*

$$d(P_0, Q_0) \leq d(P, Q),$$

per ogni $P \in \mathbb{L}$ e ogni $Q \in \mathbb{M}$.

Dimostrazione. Se \mathbb{L} e \mathbb{M} sono incidenti, è sufficiente prendere $P_0 = Q_0 \in \mathbb{L} \cap \mathbb{M}$. Possiamo quindi supporre che i due sottospazi affini \mathbb{L} e \mathbb{M} non abbiano punti in comune.

Consideriamo due punti qualsiasi $P_1 \in \mathbb{L}$, $Q_1 \in \mathbb{M}$ e indichiamo con u_1 il vettore $Q_1 - P_1$. Osserviamo che u_1 non può appartenere alla somma dei due sottospazi L e M . Infatti, se fosse $u_1 = Q_1 - P_1 \in L + M$, si avrebbe $Q_1 - P_1 = v + w$, per qualche $v \in L$, $w \in M$. Da ciò seguirebbe che il punto $R = P_1 + v = Q_1 - w$ apparterrebbe sia a \mathbb{L} che a \mathbb{M} , contro l'ipotesi che \mathbb{L} e \mathbb{M} siano disgiunti.

Indichiamo con u'_1 la proiezione ortogonale di u_1 sul sottospazio $L + M$. Si ha dunque

$$u_1 = u'_1 + u''_1,$$

con $u'_1 \in L + M$ e $u''_1 \in (L + M)^\perp = L^\perp \cap M^\perp$ (si veda la Proposizione 5.4.8 del Cap. 5). Notiamo che, da quanto detto sopra, segue che $u''_1 \neq \mathbf{0}$. Dato che $u'_1 \in L + M$, possiamo scrivere $u'_1 = v + w$, per qualche $v \in L$ e $w \in M$ (si noti che tale decomposizione non è, in generale, unica; lo è solo nel caso in cui $L \cap M = \{\mathbf{0}\}$, cioè quando i due sottospazi affini \mathbb{L} e \mathbb{M} sono sgombri). Ora poniamo $P_0 = P_1 + v$ e $Q_0 = Q_1 - w$; si ha ovviamente $P_0 \in \mathbb{L}$ e $Q_0 \in \mathbb{M}$. Affermiamo che P_0 e Q_0 sono i punti cercati. Infatti, si ha

$$\begin{aligned}Q_0 - P_0 &= (Q_1 - w) - (P_1 + v) \\ &= (Q_1 - P_1) - (v + w) \\ &= u_1 - u'_1 \\ &= u''_1 \in L^\perp \cap M^\perp,\end{aligned}$$

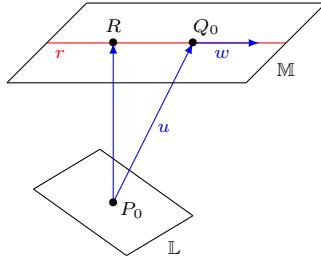
quindi il vettore $u = u''_1 = Q_0 - P_0$ è ortogonale ad entrambi i sottospazi affini \mathbb{L} e \mathbb{M} . Per il Lemma 6.7.7 si conclude. \square

Il risultato seguente afferma che la distanza tra due sottospazi affini non incidenti \mathbb{L} e \mathbb{M} di uno spazio affine euclideo \mathbb{A} si misura lungo una direzione perpendicolare ad entrambi i sottospazi. Inoltre, in base al Lemma 6.7.7, una tale direzione risulta essere unica.

Proposizione 6.7.9. *Siano $\mathbb{L} = (\mathcal{L}, L, +)$ e $\mathbb{M} = (\mathcal{M}, M, +)$ due sottospazi affini (non vuoti) di uno spazio affine euclideo \mathbb{A} e siano $P_0 \in \mathcal{L}$ e $Q_0 \in \mathcal{M}$ due punti tali che $d(P_0, Q_0) = d(\mathbb{L}, \mathbb{M})$. Allora, indicando con u il vettore $Q_0 - P_0$, si ha $u \in L^\perp \cap M^\perp$, cioè, per ogni vettore $v \in L$ e ogni $w \in M$, è $u \cdot v = u \cdot w = 0$.*

Dimostrazione. Se \mathbb{L} e \mathbb{M} sono incidenti si ha $P_0 = Q_0$, quindi $u = Q_0 - P_0 = \mathbf{0}$. In questo caso la tesi è banalmente verificata.

Supponiamo quindi che $\mathbb{L} \cap \mathbb{M} = \emptyset$. Siano dunque $P_0 \in \mathcal{L}$ e $Q_0 \in \mathcal{M}$ tali che $d(P_0, Q_0) = d(\mathbb{L}, \mathbb{M})$: si ha pertanto $d(P_0, Q_0) \leq d(A, B)$, per ogni $A \in \mathcal{L}$ e ogni $B \in \mathcal{M}$. Poniamo $u = Q_0 - P_0$ e supponiamo, per assurdo, che $u \notin M^\perp$. Ciò significa che esiste un vettore $w \in M$ tale che $u \cdot w \neq 0$. La situazione è schematizzata nella figura seguente:



Indichiamo con r la retta passante per Q_0 e parallela al vettore w : tale retta è contenuta in \mathbb{M} , dato che $Q_0 \in \mathcal{M}$ e $w \in M$. Sia R il punto della retta r per cui il vettore $\overrightarrow{P_0R}$ è ortogonale al vettore w ; R è dato da

$$R = Q_0 - \left(\frac{u \cdot w}{w \cdot w} \right) w.$$

Il quadrato della distanza di P_0 da R è:

$$\begin{aligned} d(P_0, R)^2 &= \|R - P_0\|^2 \\ &= \left\| u - \left(\frac{u \cdot w}{w \cdot w} \right) w \right\|^2 \\ &= \|u\|^2 - \left(\frac{u \cdot w}{\|w\|} \right)^2. \end{aligned}$$

Poiché abbiamo supposto che sia $u \cdot w \neq 0$, si ha

$$d(P_0, R) < \|u\| = d(P_0, Q_0),$$

il che contraddice l'ipotesi che P_0 e Q_0 siano i punti di minima distanza di \mathbb{L} e \mathbb{M} . L'assurdo deriva dall'aver supposto che il vettore u non sia perpendicolare al sottospazio M ; deve pertanto essere $u \in M^\perp$. Scambiando i ruoli di \mathbb{L} e \mathbb{M} si dimostra, in modo del tutto analogo, che vale anche $u \in L^\perp$. \square

Osservazione 6.7.10. Nel caso in cui il sottospazio affine \mathbb{L} si riduce a un punto P i risultati precedenti permettono di concludere che esiste un unico punto $Q \in \mathbb{M}$

ta che $d(P, Q) = d(P, \mathbb{M})$. In tal caso, inoltre, il vettore \overrightarrow{PQ} risulta essere ortogonale a \mathbb{M} .

Possiamo dunque affermare che, per ogni punto P e ogni sottovarietà lineare \mathbb{M} di uno spazio affine euclideo \mathbb{A} , esiste un unico punto $Q \in \mathbb{M}$ tale che il vettore \overrightarrow{PQ} sia ortogonale a \mathbb{M} ; tale punto Q è detto la *proiezione ortogonale* di P su \mathbb{M} . Risulta così definita una funzione

$$\pi_{\mathbb{M}} : \mathbb{A} \rightarrow \mathbb{M}$$

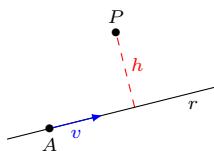
che associa ad ogni $P \in \mathbb{A}$ la sua proiezione ortogonale $\pi_{\mathbb{M}}(P) = Q$ sul sottospazio affine \mathbb{M} . Lasciamo come esercizio la verifica che $\pi_{\mathbb{M}}$ è un'applicazione affine, la cui applicazione lineare soggiacente è la funzione $\pi_M : V \rightarrow M$ che associa ad ogni vettore $v \in V$ la sua proiezione ortogonale sul sottospazio M (ove, come di consueto, abbiamo indicato con V e M gli spazi vettoriali soggiacenti agli spazi affini \mathbb{A} e \mathbb{M} , rispettivamente).

Si noti infine che, per ogni punto $P \in \mathbb{M}$, si ha $\pi_{\mathbb{M}}(P) = P$. Da ciò segue che $\pi_{\mathbb{M}}(\pi_{\mathbb{M}}(P)) = \pi_{\mathbb{M}}(P)$, per ogni $P \in \mathbb{A}$. Questa proprietà viene comunemente espressa con la seguente notazione: $\pi_{\mathbb{M}}^2 = \pi_{\mathbb{M}} \circ \pi_{\mathbb{M}} = \pi_{\mathbb{M}}$.

I risultati precedenti suggeriscono un metodo per determinare la distanza tra due sottovarietà lineari \mathbb{L} e \mathbb{M} , di dimensioni rispettivamente r e s , di uno spazio affine euclideo \mathbb{A} . Si considerari un punto generico $P \in \mathbb{L}$, le cui coordinate dipenderanno quindi da r parametri, e un punto generico $Q \in \mathbb{M}$, le cui coordinate dipenderanno da s parametri. Si calcoli il vettore $u = Q - P$, le cui componenti dipenderanno dunque da $r + s$ parametri, e si imponga che u sia ortogonale ai sottospazi \mathbb{L} e \mathbb{M} . La condizione di ortogonalità a \mathbb{L} si esprime imponendo che il prodotto scalare di u con gli r vettori di una base del sottospazio direttore di \mathbb{L} sia nullo (si ottengono così r equazioni lineari). Analogamente, la condizione di ortogonalità a \mathbb{M} si esprime imponendo che il prodotto scalare di u con gli s vettori di una base del sottospazio direttore di \mathbb{M} sia nullo (si ottengono così s equazioni lineari). Si ottiene pertanto un sistema di $r + s$ equazioni lineari in $r + s$ incognite, la cui soluzione permette di determinare due punti $P_0 \in \mathbb{L}$ e $Q_0 \in \mathbb{M}$ tali che il vettore $Q_0 - P_0$ sia ortogonale a \mathbb{L} e \mathbb{M} . Per quanto visto in precedenza, la distanza tra P_0 e Q_0 coincide con la distanza tra \mathbb{L} e \mathbb{M} .

Illustriamo ora quanto sopra esposto mediante alcuni esempi.

Esempio 6.7.11 (DISTANZA DI UN PUNTO DA UNA RETTA). Sia \mathbb{A} uno spazio affine euclideo e indichiamo con r la retta passante per un punto A e parallela a un vettore v . Dato un punto $P \in \mathbb{A}$, vogliamo determinare la distanza h di P dalla retta r .



Un generico punto X della retta r è dato da $X = A + \lambda v$, al variare del parametro $\lambda \in \mathbb{R}$. Il vettore $u = \overrightarrow{PX}$ è quindi dato da $u = X - P = (A - P) + \lambda v$. Imponendo che questo vettore sia ortogonale alla retta r (cioè al vettore v), si ottiene

$$0 = u \cdot v = (A - P) \cdot v + \lambda v \cdot v,$$

da cui si ricava

$$\lambda = -\frac{(A - P) \cdot v}{\|v\|^2}.$$

Se indichiamo con Q il punto di r corrispondente a tale valore del parametro λ , si ha

$$Q = A - \frac{(A - P) \cdot v}{\|v\|^2} v.$$

Il punto Q è dunque il punto della retta r per cui il vettore \overrightarrow{PQ} è ortogonale alla retta r stessa: Q è dunque il piede della perpendicolare tracciata per P alla retta r . Si ha pertanto

$$d(P, r) = d(P, Q) = \|Q - P\| = \left\| \left(A - \frac{(A - P) \cdot v}{\|v\|^2} v \right) - P \right\|.$$

Sviluppando i calcoli, si ottiene

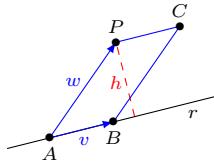
$$\begin{aligned} d(P, r)^2 &= \left(\left(A - \frac{(A - P) \cdot v}{\|v\|^2} v \right) - P \right) \cdot \left(\left(A - \frac{(A - P) \cdot v}{\|v\|^2} v \right) - P \right) \\ &= \|A - P\|^2 - 2 \frac{((A - P) \cdot v)^2}{\|v\|^2} + \frac{((A - P) \cdot v)^2}{\|v\|^2} \\ &= \|A - P\|^2 - \frac{((A - P) \cdot v)^2}{\|v\|^2}, \end{aligned}$$

da cui si ricava la seguente formula per la distanza di un punto da una retta:

$$d(P, r) = \sqrt{\|A - P\|^2 - \frac{((A - P) \cdot v)^2}{\|v\|^2}}.$$

Il problema della determinazione della distanza di un punto da una retta può essere affrontato anche in un altro modo, come ora spiegheremo.

Utilizzando le notazioni precedenti, poniamo $B = A + v$, $C = P + v$ e consideriamo il parallelogramma $ABCP$, avente come base il segmento AB e come altezza la distanza h del punto P dalla retta r .



Poiché l'area di un parallelogramma è il prodotto della sua base per la relativa altezza, si ha

$$h = \frac{\text{Area}(ABCP)}{AB}.$$

Indicando con w il vettore \overrightarrow{AP} e ricordando che $\overrightarrow{AB} = v$, l'area del parallelogramma $ABCP$ è data dalla seguente formula (vedi Cap. 5, Sezione 5.2)

$$\text{Area}(ABCP) = \sqrt{\det \begin{pmatrix} v \cdot v & v \cdot w \\ w \cdot v & w \cdot w \end{pmatrix}}.$$

Si ottiene pertanto:

$$h = \frac{\sqrt{\det \begin{pmatrix} v \cdot v & v \cdot w \\ w \cdot v & w \cdot w \end{pmatrix}}}{\|v\|} = \sqrt{\frac{1}{v \cdot v} \det \begin{pmatrix} v \cdot v & v \cdot w \\ w \cdot v & w \cdot w \end{pmatrix}}.$$

Nel caso particolare in cui \mathbb{A} è lo spazio affine euclideo $\mathbb{A}_{\mathbb{R}}^3$, l'area di un parallelogramma di lati v e w coincide con la norma del prodotto vettoriale $v \times w$:

$$\text{Area}(ABCP) = \|v \times w\|.$$

In questo caso la distanza h di P dalla retta r si può dunque calcolare come segue:

$$h = \frac{\|v \times w\|}{\|v\|}.$$

Nel prossimo esempio ricaveremo un'utile formula per determinare la distanza di un punto da un iperpiano dello spazio affine euclideo $\mathbb{A}_{\mathbb{R}}^n$. Prima però abbiamo bisogno del seguente risultato:

Proposizione 6.7.12. *Nello spazio affine euclideo $\mathbb{A}_{\mathbb{R}}^n$ consideriamo un iperpiano π di equazione*

$$\pi : a_1x_1 + a_2x_2 + \cdots + a_nx_n + b = 0$$

e indichiamo con u il vettore le cui componenti sono i coefficienti delle incognite x_1, x_2, \dots, x_n nell'equazione di π ,

$$u = (a_1, a_2, \dots, a_n).$$

Allora il vettore u è ortogonale all'iperpiano π .

Dimostrazione. Sia $P = (p_1, p_2, \dots, p_n)$ un punto di π ; si ha dunque

$$a_1p_1 + a_2p_2 + \cdots + a_np_n + b = 0.$$

Per ogni vettore $v = (\alpha_1, \alpha_2, \dots, \alpha_n)$ appartenente al sottospazio direttrice di π , il punto $Q = P + v = (p_1 + \alpha_1, p_2 + \alpha_2, \dots, p_n + \alpha_n)$ appartiene all'iperpiano π , quindi si ha

$$a_1(p_1 + \alpha_1) + a_2(p_2 + \alpha_2) + \cdots + a_n(p_n + \alpha_n) + b = 0.$$

Sottraendo le due uguaglianze precedenti, si ottiene

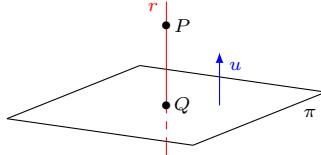
$$a_1\alpha_1 + a_2\alpha_2 + \cdots + a_n\alpha_n = 0,$$

cioè $u \cdot v = 0$. Poiché questo vale per ogni vettore v appartenente al sottospazio direttrice di π , si conclude che il vettore u è ortogonale all'iperpiano π . \square

Esempio 6.7.13 (DISTANZA DI UN PUNTO DA UN IPERPIANO). Nello spazio affine euclideo $\mathbb{A} = \mathbb{A}_{\mathbb{R}}^n$ indichiamo con π l'iperpiano di equazione

$$\pi : a_1x_1 + a_2x_2 + \cdots + a_nx_n + b = 0$$

e con $P = (p_1, p_2, \dots, p_n)$ un punto di \mathbb{A} . Vogliamo determinare la distanza di P dall'iperpiano π . Per quanto visto in precedenza, tale distanza coincide con la distanza di P dall'unico punto $Q \in \pi$ per cui il vettore \overrightarrow{PQ} è ortogonale a π .



Come abbiamo visto nella Proposizione 6.7.12, il vettore $u = (a_1, a_2, \dots, a_n)$ è ortogonale all'iperpiano π , quindi la retta r passante per il punto P e perpendicolare a π ha equazione

$$r : X = P + \lambda u.$$

Il punto Q cercato è dunque il punto di intersezione tra la retta r e l'iperpiano π :

$$Q = r \cap \pi.$$

In coordinate, le equazioni parametriche della retta r sono

$$\begin{cases} x_1 = p_1 + \lambda a_1 \\ x_2 = p_2 + \lambda a_2 \\ \dots \\ x_n = p_n + \lambda a_n. \end{cases}$$

Sostituendo queste espressioni nell'equazione di π si ottiene l'equazione

$$a_1(p_1 + \lambda a_1) + a_2(p_2 + \lambda a_2) + \dots + a_n(p_n + \lambda a_n) + b = 0,$$

la cui soluzione è

$$\bar{\lambda} = -\frac{a_1 p_1 + a_2 p_2 + \dots + a_n p_n + b}{a_1^2 + a_2^2 + \dots + a_n^2}.$$

Il punto della retta r corrispondente a tale valore di λ è il punto Q cercato:

$$Q = P + \bar{\lambda} u.$$

Si ha pertanto

$$d(P, \pi) = d(P, Q) = \|Q - P\| = \|\bar{\lambda} u\| = |\bar{\lambda}| \|u\|.$$

Sviluppando i calcoli, si ottiene così

$$d(P, \pi) = \frac{|a_1 p_1 + a_2 p_2 + \dots + a_n p_n + b|}{\sqrt{a_1^2 + a_2^2 + \dots + a_n^2}}.$$

Generalizzando gli esempi precedenti, vediamo ora come si possa determinare la distanza di un punto da una sottovarietà lineare qualunque di uno spazio affine euclideo.

Esempio 6.7.14 (DISTANZA DI UN PUNTO DA UNA SOTTOVARIETÀ LINEARE QUALUNQUE). Sia \mathbb{A} uno spazio affine euclideo e sia $\mathbb{L} = (\mathcal{L}, L, +)$ una sottovarietà lineare di dimensione r di \mathbb{A} . Dato un punto $P \in \mathbb{A}$, vogliamo determinare la distanza di P da \mathbb{L} ; tale distanza coinciderà con la distanza di P dall'unico punto Q di \mathbb{L} per cui il vettore \overrightarrow{PQ} è ortogonale a \mathbb{L} .

Sia A un punto qualunque di \mathbb{L} e indichiamo con $\{v_1, v_2, \dots, v_r\}$ una base ortonormale⁸ di L . Un generico punto X di \mathbb{L} è dato da

$$X = A + \lambda_1 v_1 + \dots + \lambda_r v_r,$$

al variare dei parametri $\lambda_1, \dots, \lambda_r \in \mathbb{R}$. Indicando con u il vettore $X - P$ e con w il vettore $A - P$, si ha dunque

$$u = w + \sum_{j=1}^r \lambda_j v_j.$$

La richiesta che questo vettore sia ortogonale a \mathbb{L} si esprime imponendo che il prodotto scalare di u per i vettori della base di L sia nullo. Ricordando che i vettori v_1, \dots, v_r formano una base ortonormale di L , per ogni $i = 1, \dots, r$, si ha

$$0 = u \cdot v_i = w \cdot v_i + \sum_{j=1}^r \lambda_j v_j \cdot v_i = w \cdot v_i + \lambda_i,$$

da cui si ricava

$$\lambda_i = -w \cdot v_i.$$

Il punto Q cercato è quindi dato da

$$Q = A - (w \cdot v_1)v_1 - \dots - (w \cdot v_r)v_r.$$

Si ha pertanto

$$d(P, \mathbb{L}) = d(P, Q) = \|Q - P\| = \|w - (w \cdot v_1)v_1 - \dots - (w \cdot v_r)v_r\|.$$

Sviluppando i calcoli si trova

$$d(P, \mathbb{L})^2 = \|w\|^2 - (w \cdot v_1)^2 - \dots - (w \cdot v_r)^2,$$

da cui si ottiene infine la seguente formula per la distanza di P da \mathbb{L} :

$$d(P, \mathbb{L}) = \sqrt{\|w\|^2 - (w \cdot v_1)^2 - \dots - (w \cdot v_r)^2}.$$

Esempio 6.7.15 (DISTANZA TRA DUE RETTE). Indichiamo con r e s due rette in uno spazio affine euclideo \mathbb{A} . Per determinare la distanza di r da s cerchiamo due punti $P \in r$ e $Q \in s$ tali che il vettore \overrightarrow{PQ} sia ortogonale alle rette r e s .

Sia v un vettore direttore di r e scegliamo arbitrariamente un punto $A \in r$: un generico punto X di r è quindi dato da

$$X = A + \lambda v.$$

⁸L'ipotesi che la base $\{v_1, v_2, \dots, v_r\}$ sia ortonormale non è indispensabile, ma permette di semplificare i calcoli.

Analogamente, indicando con w un vettore direttore di s e fissando un punto $B \in s$, un generico punto Y della retta s è dato da:

$$Y = B + \mu w.$$

Se indichiamo con u il vettore \overrightarrow{XY} , si ha

$$u = (B - A) - \lambda v + \mu w.$$

La richiesta che questo vettore sia ortogonale alle rette r e s si esprime imponendo che i prodotti scalari di u con v e w siano nulli:

$$\begin{cases} u \cdot v = (B - A) \cdot v - \lambda v \cdot v + \mu w \cdot v = 0 \\ u \cdot w = (B - A) \cdot w - \lambda v \cdot w + \mu w \cdot w = 0. \end{cases}$$

Se i vettori v e w sono linearmente indipendenti (cioè se le rette r e s non sono parallele) il determinante della matrice dei coefficienti di questo sistema è diverso da zero, quindi il sistema ammette un'unica soluzione $\lambda = \bar{\lambda}$ e $\mu = \bar{\mu}$. Sostituendo tali valori nelle espressioni di X e Y si ottengono i punti P e Q cercati. Se invece le rette r e s sono parallele, il determinante della matrice dei coefficienti del sistema precedente è nullo. In questo caso esso ammette infinite soluzioni (dipendenti da un parametro). Una qualunque di queste soluzioni fornisce una coppia di punti P e Q tali che $d(r, s) = d(P, Q)$.

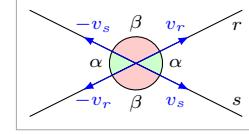
6.7.1 Angoli

Parliamo ora del concetto di angolo in uno spazio affine euclideo. Come già osservato nel Capitolo 5, la presenza di una forma bilineare simmetrica definita positiva su uno spazio vettoriale reale V permette di definire la nozione di angolo compreso tra due vettori non nulli di V . Le stesse considerazioni si applicano quindi al caso di uno spazio affine euclideo. Utilizzando la nozione di angolo compreso tra due vettori è poi possibile definire l'angolo compreso tra due rette incidenti.

Siano dunque r e s due rette incidenti in uno spazio affine euclideo A e indichiamo con v_r e v_s rispettivamente due generatori dei sottospazi direttori di r e s .

Si noti che i vettori v_r e v_s sono determinati a meno della moltiplicazione per uno scalare non nullo. Ricordiamo ora che il coseno dell'angolo α compreso tra v_r e v_s è dato dal rapporto

$$\frac{v_r \cdot v_s}{\|v_r\| \|v_s\|} \tag{6.7.1}$$



Sostituendo v_r e v_s con i vettori λv_r e μv_s , con $\lambda, \mu \in \mathbb{R}$, $\lambda, \mu \neq 0$, il rapporto precedente diventa

$$\frac{\lambda v_r \cdot \mu v_s}{|\lambda| \|v_r\| |\mu| \|v_s\|} = \pm \frac{v_r \cdot v_s}{\|v_r\| \|v_s\|}$$

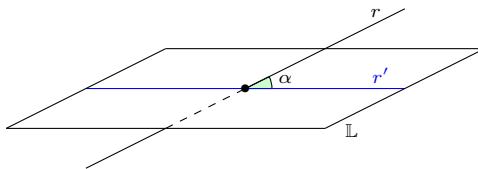
ove il segno dipende dal segno del prodotto $\lambda\mu$. In altri termini, ciò significa che, date due rette r e s , il rapporto (6.7.1) è ben definito solo a meno del segno. Ciò corrisponde al fatto che, se indichiamo con α l'angolo compreso tra i vettori v_r e v_s , allora l'angolo compreso tra v_r e $-v_s$, oppure tra $-v_r$ e v_s è $\beta = \pi - \alpha$, come illustrato nella figura precedente.

Fatte queste premesse, possiamo dare la seguente definizione:

Definizione 6.7.16. Con le notazioni precedenti, chiameremo *angolo compreso tra le due rette incidenti* r e s quello, tra i due angoli α e $\beta = \pi - \alpha$, che risulta compreso nell'intervallo $[0, \pi/2]$.

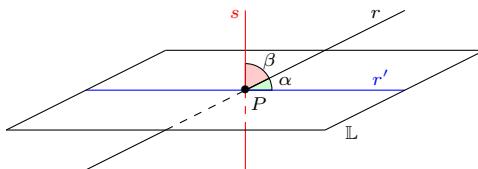
La definizione appena data può essere generalizzata al caso dell'angolo compreso tra una retta r e una sottovarietà lineare \mathbb{L} , di dimensione ≥ 1 , ad essa incidente.

Sia dunque \mathbb{L} una sottovarietà lineare di dimensione ≥ 1 di uno spazio affine euclideo \mathbb{A} e indichiamo con $\pi_{\mathbb{L}} : \mathbb{A} \rightarrow \mathbb{L}$ la proiezione ortogonale su \mathbb{L} . Data una retta r incidente a \mathbb{L} , indichiamo con $r' = \pi_{\mathbb{L}}(r)$ la sua proiezione ortogonale⁹ su \mathbb{L} . Se la retta r è ortogonale a \mathbb{L} , r' è un punto, il quale è precisamente il punto di intersezione tra r e \mathbb{L} ; in caso contrario r' è una retta.



Definizione 6.7.17. Con le notazioni precedenti, chiameremo *angolo compreso tra la retta r e la sottovarietà lineare \mathbb{L} ad essa incidente* l'angolo α compreso tra r e la sua proiezione ortogonale r' su \mathbb{L} , con la convenzione che, se r' è un punto, α è un angolo retto.

Nel caso in cui la sottovarietà lineare \mathbb{L} è un iperpiano di \mathbb{A} , il sottospazio \mathbb{L}^\perp , ortogonale del sottospazio direttore di \mathbb{L} , ha dimensione 1. Indicando con P il punto di intersezione tra la retta r e l'iperpiano \mathbb{L} , la retta $s = P + \mathbb{L}^\perp$ è la retta perpendicolare a \mathbb{L} passante per il punto P . Se r non è ortogonale a \mathbb{L} , indicata con r' la proiezione ortogonale di r sull'iperpiano \mathbb{L} , le tre rette s , r e r' sono complanari e, inoltre, s e r' sono tra loro ortogonali. La situazione è illustrata nella figura seguente:



Di conseguenza, se indichiamo con α l'angolo compreso tra la retta r e l'iperpiano \mathbb{L} (cioè l'angolo tra r e r') e con β l'angolo compreso tra le rette r e s , si ha

$$\alpha + \beta = \frac{\pi}{2}.$$

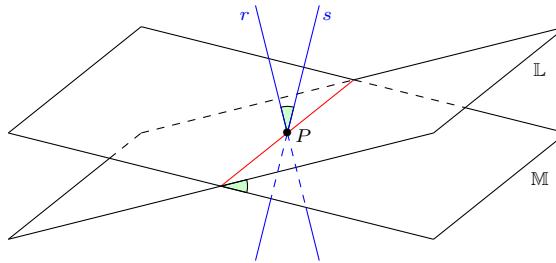
Questa osservazione fornisce un metodo alternativo per calcolare l'angolo α compreso tra la retta r e l'iperpiano \mathbb{L} : basta infatti determinare la retta s

⁹Per determinare la proiezione ortogonale della retta r sulla sottovarietà lineare \mathbb{L} è sufficiente considerare due punti $P, Q \in r$ e determinare le loro proiezioni ortogonali $P' = \pi_{\mathbb{L}}(P)$ e $Q' = \pi_{\mathbb{L}}(Q)$ su \mathbb{L} . Se $P' \neq Q'$, la retta passante per P' e Q' è la retta r' cercata, se invece $P' = Q'$ la retta r risulta essere ortogonale a \mathbb{L} e pertanto la sua proiezione ortogonale su \mathbb{L} si riduce a un punto.

perpendicolare a \mathbb{L} , calcolare l'angolo β tra r e s e ricordare che $\alpha + \beta$ è un angolo retto.

Per terminare, osserviamo che le considerazioni precedenti ci permettono di definire la nozione di angolo compreso tra due iperpiani incidenti o, più in generale, tra un iperpiano e una sottovarietà lineare di dimensione ≥ 1 ad esso incidente.

Definizione 6.7.18. Siano \mathbb{L} e \mathbb{M} due iperpiani incidenti in uno spazio affine euclideo \mathbb{A} . Sia $P \in \mathbb{L} \cap \mathbb{M}$ e indichiamo con r e s rispettivamente le rette passanti per il punto P e perpendicolari agli iperpiani \mathbb{L} e \mathbb{M} . Definiamo l'angolo α , compreso tra gli iperpiani \mathbb{L} e \mathbb{M} , come l'angolo compreso tra le rette r e s (nel caso in cui $\dim \mathbb{A} = 3$ la situazione è illustrata nella figura seguente).



Definizione 6.7.19. Sia \mathbb{L} un iperpiano in uno spazio affine euclideo \mathbb{A} e sia \mathbb{M} una sottovarietà lineare di \mathbb{A} , di dimensione ≥ 1 , incidente a \mathbb{L} . Sia $P \in \mathbb{L} \cap \mathbb{M}$ e indichiamo con r la retta perpendicolare a \mathbb{L} passante per il punto P . L'angolo α compreso tra \mathbb{L} e \mathbb{M} è definito ponendo $\alpha = \pi/2 - \beta$, ove β è l'angolo compreso tra la retta r e la sottovarietà lineare \mathbb{M} .

Osservazione 6.7.20. Come già accennato in precedenza, le nozioni di distanza e angolo permettono poi di introdurre anche i concetti di area e volume. La trattazione delle aree e dei volumi negli spazi affini euclidei è del tutto analoga è quella svolta nel caso degli spazi vettoriali euclidei, a cui rimandiamo (vedi Cap. 5, Sezione 5.2).

6.8 Isometrie degli spazi affini euclidei

In questo paragrafo studieremo le applicazioni tra due spazi affini euclidei \mathbb{A} e \mathbb{B} che rispettano la struttura di spazio affine euclideo, cioè le applicazioni affini che sono compatibili con i prodotti scalari definiti negli spazi vettoriali soggiacenti agli spazi affini \mathbb{A} e \mathbb{B} .

Siano dunque $\mathbb{A} = (\mathcal{A}, V, +)$, $\mathbb{B} = (\mathcal{B}, W, +)$ due spazi affini euclidei e indichiamo con g e h i prodotti scalari definiti sugli spazi vettoriali V e W , rispettivamente.

Definizione 6.8.1. Un'applicazione affine $F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{B}$ è detta un'*isometria* (di spazi affini euclidei) se la funzione lineare $\phi : V \rightarrow W$ è un'isometria di spazi vettoriali euclidei, cioè se, per ogni $v_1, v_2 \in V$, si ha

$$h(\phi(v_1), \phi(v_2)) = g(v_1, v_2).$$

Tutti i risultati riguardanti le isometrie degli spazi vettoriali euclidei, ottenuti nel Capitolo 5, si estendono in modo ovvio al contesto degli spazi affini euclidei. Il seguente risultato, ad esempio, è una conseguenza diretta del Corollario 5.6.3:

Proposizione 6.8.2. *Siano $\mathbb{A} = (\mathcal{A}, V, +)$ e $\mathbb{B} = (\mathcal{B}, W, +)$ due spazi affini euclidei e sia $F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{B}$ un'applicazione affine. Se F è un'isometria allora essa è iniettiva.*

Dimostrazione. Se F è un'isometria di spazi affini euclidei, l'applicazione lineare $\phi : V \rightarrow W$ è un'isometria di spazi vettoriali euclidei e pertanto è iniettiva. Come abbiamo visto nella Proposizione 6.6.7, l'iniettività di ϕ equivale all'iniettività di F . \square

Le isometrie degli spazi affini euclidei preservano le distanze tra i punti. Se \mathbb{A}, \mathbb{B} sono due spazi affini euclidei e $F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{B}$ è un'isometria, per ogni coppia di punti $P, Q \in \mathbb{A}$ si ha infatti

$$d(P, Q) = \|Q - P\| = \|\phi(Q - P)\| = \|f(Q) - f(P)\| = d(f(P), f(Q)).$$

È facile verificare che questa proprietà caratterizza le isometrie:

Proposizione 6.8.3. *Siano \mathbb{A} e \mathbb{B} due spazi affini euclidei e sia $F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{B}$ un'applicazione affine tale che $d(P, Q) = d(f(P), f(Q))$, per ogni $P, Q \in \mathbb{A}$. Allora F è un'isometria.*

Dimostrazione. Dati $P \in \mathbb{A}$ e $v \in V$, poniamo $Q = P + v$. Si ha dunque $\phi(v) = f(Q) - f(P)$ e quindi $\|v\| = d(P, Q) = d(f(P), f(Q)) = \|f(Q) - f(P)\| = \|\phi(v)\|$. Si deduce pertanto che $\phi : V \rightarrow W$ è un'applicazione lineare che preserva le norme dei vettori e quindi è un'isometria (vedi Cap. 5, Osservazione 5.6.10). \square

Come nel caso delle applicazioni affini, è del tutto evidente che l'applicazione identica di uno spazio affine euclideo in sé è un'isometria e che la composizione di due isometrie è ancora un'isometria. Inoltre se un'isometria è biettiva, anche la sua inversa è un'isometria. Pertanto le isometrie di uno spazio affine euclideo \mathbb{A} formano un sottogruppo del gruppo delle affinità di \mathbb{A} ; tale sottogruppo è indicato con $\text{Isom}(\mathbb{A})$.

Esempio 6.8.4. Dato uno spazio affine euclideo $\mathbb{A} = (\mathcal{A}, V, +)$, le traslazioni costituiscono degli esempi di isometrie di \mathbb{A} (vedi Osservazione 6.1.7). Per ogni $v \in V$, la funzione $\tau_v : \mathbb{A} \rightarrow \mathbb{A}$ definita ponendo $\tau_v(P) = P + v$, per ogni $P \in \mathbb{A}$, è un'isometria di \mathbb{A} ; infatti l'applicazione lineare $\phi : V \rightarrow V$ sottogiacente a τ_v è l'applicazione identica. Si noti che, se $v \neq \mathbf{0}$, la traslazione τ_v è un'isometria priva di punti fissi.

L'insieme delle traslazioni è un sottogruppo del gruppo delle isometrie di \mathbb{A} , isomorfo al gruppo additivo dello spazio vettoriale V .

Esempio 6.8.5. Sia $\mathbb{A} = (\mathcal{A}, V, +)$ uno spazio affine euclideo e sia $\phi : V \rightarrow V$ un'isometria. Per ogni punto $P_0 \in \mathcal{A}$, definiamo una funzione $f : \mathcal{A} \rightarrow \mathcal{A}$ ponendo $f(P) = P_0 + \phi(P - P_0)$. Si ottiene così un'isometria $F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{A}$ per la quale P_0 è un punto fisso.

Se \mathbb{A} e \mathbb{B} sono due spazi affini euclidei, la scelta di due sistemi di riferimento $\mathcal{R} = \{O_{\mathbb{A}}, v_1, \dots, v_n\}$ in \mathbb{A} e $\mathcal{S} = \{O_{\mathbb{B}}, w_1, \dots, w_m\}$ in \mathbb{B} permette di associare a ogni applicazione affine $F = (f, \phi) : \mathbb{A} \rightarrow \mathbb{B}$ una coppia formata da un elemento

$\mathbf{t} = (t_1, \dots, t_m) \in \mathbb{R}^m$ e una matrice $A \in M_{m,n}(\mathbb{R})$, la quale non è altro che la matrice dell'applicazione lineare $\phi : V \rightarrow W$, rispetto alle basi $\{v_1, \dots, v_n\}$ di V e $\{w_1, \dots, w_m\}$ di W . Da quanto detto si deduce che F è un'isometria di spazi affini euclidei se e solo se A è la matrice di un'isometria di spazi vettoriali euclidei (vedi Cap. 5, Paragrafo 5.6). Se inoltre i sistemi di riferimento \mathcal{R} e \mathcal{S} sono ortonormali, e quindi le matrici associate alle forme bilineari simmetriche g e h sono le matrici identiche, F è un'isometria se e solo se A è una matrice ortogonale, cioè se e solo se ${}^t A A = \mathbf{1}$.

Osservazione 6.8.6. Come abbiamo già notato in precedenza, dato uno spazio affine \mathbb{A} di dimensione n sul campo K , la scelta di un sistema di riferimento \mathcal{R} in \mathbb{A} permette di identificare il gruppo delle affinità di \mathbb{A} con il prodotto semidiretto $K^n \rtimes \mathrm{GL}_n(K)$. Nel caso in cui \mathbb{A} sia uno spazio affine euclideo, dalle considerazioni precedenti si deduce che la scelta di un sistema di riferimento ortonormale in \mathbb{A} permette di identificare il gruppo delle isometrie di \mathbb{A} con il prodotto semidiretto $\mathbb{R}^n \rtimes \mathrm{O}_n(\mathbb{R})$.

Esercizi

Esercizio 6.1. Nello spazio affine $\mathbb{A}_{\mathbb{R}}^3$ sia r la retta di equazioni

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

Si determini la retta s passante per l'origine e parallela ad r , ed il piano del fascio di asse r passante per l'origine.

Esercizio 6.2. Siano r_1 e r_2 rispettivamente le rette passanti per $P_1 = (0, 1, 2)$ e $P_2 = (2, -1, 0)$ e parallele ai vettori $v_1 = (2, 1, 1)$ e $v_2 = (-2, 3, 0)$. Siano \mathcal{F}_1 e \mathcal{F}_2 rispettivamente i fasci di piani di asse r_1 e r_2 . Si determinino i piani π_1 di \mathcal{F}_1 e π_2 di \mathcal{F}_2 passanti per il punto $P = (1, 1, 1)$, e le intersezioni della retta $r = \pi_1 \cap \pi_2$ con i piani coordinati. Si determini infine il triangolo individuato dalle intersezioni delle rette r , r_1 e r_2 con il piano di equazione $x = 0$.

Esercizio 6.3. Determinare l'equazione del piano contenente le rette r e s di equazioni

$$r : \begin{cases} y + 1 = 0 \\ 2x + 2z - 3 = 0 \end{cases} \quad s : \begin{cases} 2x - 3y + 2z - 1 = 0 \\ x - 2y + z = 0. \end{cases}$$

Esercizio 6.4. Sia r la retta passante per i punti $P = (1, 0, -2)$ e $Q = (0, -1, 3)$, e sia s la retta passante per il punto $R = (-1, 1, 0)$ e parallela al vettore $v = (1, -1, -1/2)$. Si determinino il piano π_1 , passante per i punti P , Q e R , il piano π_2 contenente la retta s e passante per il punto medio del segmento PQ , ed infine un vettore u parallelo alla retta $\pi_1 \cap \pi_2$.

Esercizio 6.5. Si determini la distanza del punto $P = (2, 0, -1)$ dalla retta r di equazioni parametriche

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

Esercizio 6.6. Si determinino le rette passanti per il punto $Q = (0, 1, -1)$, distanti 1 dal punto $P = (1, 1, 0)$ e contenute nel piano π di equazione $y + z = 0$.

Esercizio 6.7. Si determini l'equazione del luogo dei punti dello spazio affine euclideo $\mathbb{A}_{\mathbb{R}}^3$ appartenenti alle rette passanti per il punto $P = (2, 1, 1)$ che formano un angolo pari a $\frac{\pi}{6}$ con il piano π di equazione $2x + y - z = 0$.

Esercizio 6.8. Si determini l'equazione del luogo dei punti dello spazio affine $\mathbb{A}_{\mathbb{R}}^3$ appartenenti alle rette passanti per il punto $P = (3, 3, 3)$ che intersecano la curva C di equazione

$$\begin{cases} x^2 + 2y^2 + 3z^2 = 4 \\ x + z = 0. \end{cases}$$

Esercizio 6.9. Si determinino le rette del piano π di equazione $x - y + 2z = 0$ parallele al vettore $v = (1, 3, 1)$ e distanti $\sqrt{6}$ dal punto $P = (1, 0, 0)$.

Esercizio 6.10. Nello spazio affine euclideo reale tridimensionale, si determini l'equazione del luogo dei punti appartenenti alle rette distanti 1 dal punto $P = (1, 1, 2)$ e formanti un angolo pari a $\frac{\pi}{3}$ con i piani $\pi_1 : x + y = 0$ e $\pi_2 : x - z = 0$.

Esercizio 6.11. Nello spazio affine euclideo $\mathbb{A}_{\mathbb{R}}^3$ sia π il piano contenente la retta r di equazioni $\frac{1}{2}(x - 1) = y + 1 = \frac{1}{3}z$ ed il punto $P = (0, 1, 1)$. Si determinino le rette del piano π passanti per P e distanti $\sqrt{3}$ dal punto $Q = (1, 1, -1)$. Si determinino inoltre la proiezione ortogonale Q_0 di Q su π e le rette di π passanti per Q_0 e ortogonali alle rette trovate in precedenza.

Esercizio 6.12. Nello spazio affine euclideo $\mathbb{A}_{\mathbb{R}}^3$ si determinino le equazioni delle rette passanti per $P = (1, 0, 0)$, incidenti la retta r di equazioni

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

e formanti con quest'ultima un angolo pari a $\frac{\pi}{6}$.

Esercizio 6.13. Nel piano affine euclideo $\mathbb{A}_{\mathbb{R}}^2$ siano r_1 e r_2 le rette di equazioni $x - 2y + 1 = 0$ e $3x - y + 2 = 0$, rispettivamente. Si determinino le equazioni delle rette s_1 e s_2 bisettrici degli angoli formati da r_1 e r_2 .

Bibliografia

- [1] S. Abeasis, *Elementi di Algebra Lineare e Geometria*, Zanichelli, 2000.
- [2] M. Artin, *Algebra*, Bollati Boringhieri, Torino, 1997.
- [3] M. Audin, *Geometry*, Springer-Verlag, 2002.
- [4] I. Barsotti, *Appunti di Algebra*, Zanichelli, 1968.
- [5] I. Barsotti, *Appunti di Geometria II*, note ciclostilate, Università di Padova, 1972-73.
- [6] N. Bourbaki, *Algèbre*, Masson, Paris, 1981.
- [7] M. Candilera, *Algebra Lineare e Geometria*, Edizioni Libreria Progetto, Padova, 1992.
- [8] P.R. Halmos, *Finite-dimensional Vector Spaces*, Springer-Verlag, 1987.
- [9] R. Hartshorne, *Geometry: Euclid and Beyond*, Springer-Verlag, 2000.
- [10] J. Hefferon, *Linear Algebra*, <http://joshua.smcvt.edu/linearalgebra/>
- [11] D. Hilbert, S. Cohn-Vossen, *Geometria Intuitiva*, Bollati Boringhieri, Torino, 1972.
- [12] K. Hoffman, R. Kunze, *Linear Algebra*, Prentice-Hall, 1971.
- [13] L. Hogben, *Handbook of Linear Algebra*, Chapman & Hall/CRC, 2007.
- [14] R.A. Horn, C.R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
- [15] N. Jacobson, *Lectures in Abstract Algebra*, Springer-Verlag, 1953.
- [16] S. Lang, *Algebra*, Springer-Verlag, 2002.
- [17] H. Lütkepohl, *Handbook of Matrices*, John Wiley & Sons, 1996.
- [18] R.C. Lyndon, *Groups and Geometry*, Cambridge University Press, Cambridge, 1985.
- [19] C.D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, 2000.
- [20] E. Sernesi, *Geometria 1*, Bollati Boringhieri, Torino, 1989.
- [21] G.E. Shilov, *Linear Algebra*, Dover Publications, 1977.
- [22] G. Strang, *Linear Algebra and its Applications*, Thomson Learning, 1988.

Indice analitico

A

additiva (funzione), 27
affinamente indipendenti, 171
affinità, 187
aggiunta (matrice), 80
alternante (funzione), 76
angolo, 122, 127, 137, 204, 205
applicazione
 affine, 184
 lineare soggiacente, 184
area, 128
assi (coordinati), 171
 $\text{Aut}(V)$, 32
automorfismo, 32, 95
autospazio, 99
autovalore, 99
autovettore, 99
 generalizzato, 106

B

baricentro, 183
base
 cambiamento di, 46
 canonica, 20
 di Jordan, 115
 di uno spazio vettoriale, 17
 ortogonale, 136
 ortonormale, 136
basi equiorientate, 93
Binet (Teorema di), 79
blocco di Jordan, 112

C

campo, 9
 ordinato, 182
Cauchy (formula di), 97
cicli disgiunti, 68
ciclo, 68
cofattore, 80
colonne, 35
combinazione lineare, 11
complementare (sottospazio), 23
complemento algebrico, 80
congruenti (matrici), 144
coordinate, 35, 171

Cramer (regola di), 85

D

definita negativa, 136, 144
definita positiva, 136, 144
degenera, 133
determinante, 71
 di un endomorfismo, 89
 di Vandermonde, 88
diagonale principale, 43
diagonalizzabile, 99
diagramma commutativo, 32
dimensione, 20, 164
distanza, 194–196, 198, 200, 202
disuguaglianza
 di Cauchy–Schwarz, 126, 137
 triangolare, 122, 127, 137, 195
divisori di zero, 45

E

eliminazione, 53, 58, 60, 86
 $\text{End}(V)$, 32
endomorfismo, 32
 diagonalizzabile, 99
epimorfismo, 29
equazione
 caratteristica, 100
 di un iperpiano affine, 174
 di un piano affine, 174
 di una retta affine, 173, 175
 secolare, 100
equazioni
 cartesiane, 173
 parametriche, 172
equiorientate (basi), 93
equiorientazione, 93
equipollenti, 5
equipollenza, 5
estensione (di campi), 22
Euclide (Quinto Postulato), 178

F

fascio, 177
finitamente generato, 18
forma a scala, 54

- forma bilineare, 132
 forma canonica di Jordan, 115
Formula
 di Grassmann, 22
 di Laplace, 80
 funzione
 additiva, 27
 additiva non lineare, 34
 alternante, 76
 lineare, 27, 32, 155
 multilineare, 76
- G**
 Gauss (eliminazione), 53, 58, 60, 86
 generatori (insieme di), 15
 giacitura, 167
 $GL(n, K)$, 46
 Gram–Schmidt, 145
 Grassmann (formula di), 22
 gruppo
 delle affinità, 193
 delle isometrie, 158, 206, 207
 delle isometrie dirette, 159
 generale lineare, 46
 ortogonale, 158, 159
 speciale ortogonale, 159
- H**
 $\text{Hom}(V, W)$, 32
- I**
 immagine, 29
 di un sottospazio affine, 186
 incidenti, 170
 indefinita, 136, 144
 insieme
 di generatori, 15
 libero, 16
 intersezione
 di sottospazi affini, 169
 di sottospazi vettoriali, 12
 inversa (matrice), 45, 60
 inversione, 66
 ipercubo, 122
 iperpiani coordinati, 171
 iperpiano affine, 167
 isometria, 157
 di spazi affini euclidei, 205
 isometrici, 157
 isomorfi (spazi vettoriali), 28
 isomorfismo, 28
 compatibile con gli orientamenti, 94
 di spazi affini, 187
 di spazi vettoriali orientati, 94
- isotropo, 134
- J**
 Jordan
 base di, 115
 blocco di, 112
 forma canonica di, 115
- L**
 Laplace (formula di), 80
 libero (insieme), 16
 linearmente (in)dipendenti, 16
 lunghezza di un vettore, 121
- M**
 matrice, 35
 aggiunta, 80
 cofattore, 80
 completa, 52
 di un'applicazione affine, 191
 di una forma bilineare, 138
 diagonale, 44
 diagonalizzabile, 99
 identica, 43
 inversa, 45
 nulla, 43
 ortogonale, 159
 quadrata, 43
 scalare, 44
 trasposta, 39
 triangolare, 44
- matrici
 congruenti, 144
 simili, 48
 minore, 80, 92
 minori orlati, 93
 misura, 131
 modulo, 5, 121
 modulo (su un anello), 11
 molteplicità, 103
 monomorfismo, 29
 multilineare, 76
- N**
 nilpotenti (elementi), 45
 non degenera, 133
 norma, 5, 121, 137
 normalizzato (vettore), 136
 nucleo, 29
 di una forma bilineare, 133
- nullità
 di un autovalore, 103
 di una funzione lineare, 31
 di una matrice, 42

O

- omomorfismo, 27, 155
- operazioni elementari, 54, 62
- orientamento, 94
- origine, 171
- ortogonale
 - base, 136
 - di un sottospazio vettoriale, 135
- ortogonali, 134
- ortonormale (base), 136

P

- paralleli, 170
- parallelogramma (regola del), 6
- parallelotopo, 130
- parametri, 172
- periodo, 106
- permutazione, 65
 - dispari, 66
 - pari, 66
 - segno di una, 66
- piano affine, 167
- polinomi, 10
- polinomio
 - caratteristico, 100
 - minimo, 111
- posizione generica, 171
- Postulato di Euclide, 178
- procedimento di Gram–Schmidt, 145
- prodotto
 - di matrici, 37, 40
 - di una matrice per un vettore, 38
 - di una matrice per uno scalare, 36
 - righe per colonne, 38
 - scalare, 39, 123, 125, 138
 - semidiretto, 193
- proiezione ortogonale, 198
- punto medio, 182

Q

- Quinto Postulato di Euclide, 178

R

- rango, 90, 92
 - di una funzione lineare, 31
 - di una matrice, 42
 - per colonne, 43, 58
 - per righe, 43, 58
- regola
 - del parallelogramma, 6
 - di Cramer, 85
 - di Sarrus, 72
- retta affine, 167
- righe, 35

S

- Sarrus (regola di), 72
- scalare, 9
- scambio di elementi contigui, 69
- segmenti equipollenti, 5
- segmento, 182
 - orientato, 5
- segno (di una permutazione), 66
- semidefinita negativa, 136, 144
- semidefinita positiva, 136, 144
- sghembi, 170
- simili (matrici), 48
- similitudine, 48
- sistema
 - di equazioni lineari, 1
 - lineare, 50
 - non omogeneo, 52
 - omogeneo, 2, 51
- sistema di riferimento, 171
 - cartesiano, 194
 - ortogonale, 194
 - ortonormale, 194
- somma
 - di vettori, 6
 - di matrici, 36
 - di sottospazi affini, 168
 - di sottospazi vettoriali, 13
- somma diretta, 14
 - di forme bilineari, 133
 - esterna, 14
 - interna, 14
- somma ortogonale, 133
- sostituzione, 2
- sottospazio
 - affine, 166, 167
 - complementare, 23
 - nullo, 12
 - vettoriale, 11, 12
- sottovarietà lineare, 166
- spazio
 - affine, 164
 - affine euclideo, 194
 - direttore, 164
 - euclideo, 138
 - vettoriale, 7, 9
 - vettoriale orientato, 94
- stella, 177
- struttura canonica (spazio affine), 165

T

- Teorema
 - dei minori orlati, 93
 - di Binet, 79
 - di Desargues, 181

di Hamilton–Cayley, 111
di Jordan, 114
di Pappo, 180
di Rouché–Capelli, 53
di Sylvester, 149
di Talete, 178, 180
traslazione, 165
trasposizione, 39, 68
trasposta, 39

U

unione (di sottospazi vettoriali), 12

V

Vandermonde (determinante), 88
vettore, 9
 applicato, 6
 geometrico, 5
 nullo, 7
volume, 129, 131