

ISTITUZIONI DI MATEMATICA
CORSO DI INGEGNERIA – ARCHITETTURA
PARTE 2
— 2024 —

GABRIELE SICURO



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Indice

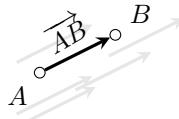
Capitolo 1. Elementi di algebra lineare e geometria	1
1. Spazi vettoriali	1
2. Lo spazio delle matrici	7
3. Sistemi di equazioni lineari	12
4. Endomorfismi e proprietà spettrali	23
5. Spazi prehilbertiani	33
6. Spazi affini e spazi euclidei	40
7. Geometria analitica nel piano	45
8. Geometria analitica nello spazio	51
Capitolo 2. Elementi di calcolo in più variabili	59
1. Intorni in \mathbb{R}^d : topologia, limiti e continuità	59
2. Derivate di una funzione a più variabili	62
3. Misura e integrazione su \mathbb{R}^d	67
4. Curve e superfici	77
5. Forme differenziali	87

CAPITOLO 1

Elementi di algebra lineare e geometria

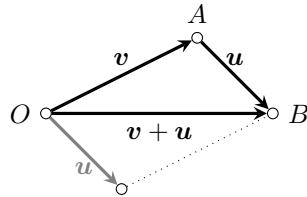
1. Spazi vettoriali

1.1. Vettori applicati e vettori geometrici. Consideriamo lo spazio ordinario tridimensionale \mathbb{E}_3 , e due punti A e B in esso. Immaginiamo di spostarci, in linea retta, da A a B e di associare a questo spostamento una freccia, che parte da A e punta in B : chiamiamo questa freccia *vettore applicato* \overrightarrow{AB} con *punto di applicazione* in A e *punto finale* in B ,



Un vettore applicato è caratterizzato dal suo punto di applicazione, A , da una direzione, un verso e una lunghezza o *modulo*. Se $A = B$, il vettore è detto *nullo* e viene indicato con $\mathbf{0}$: la sua lunghezza è zero e la sua direzione e il suo verso non sono definiti. Due vettori \overrightarrow{AB} e \overrightarrow{CD} sono *equipollenti* se hanno stessa direzione (ovvero giacciono su rette parallele), stessa lunghezza e stesso verso. L'equipollenza è una relazione di equivalenza e un *vettore geometrico* (o vettore libero) corrisponde ad una classe di equivalenza secondo equipollenza. Uno specifico vettore applicato è perciò un *rappresentante* di questa classe di equivalenza, mentre tutti i vettori con stessa direzione, stessa lunghezza e stesso verso corrispondono allo stesso “vettore geometrico”. Indicheremo una tale classe di equivalenza con una lettera minuscola in grassetto, per esempio \mathbf{v} , e con \mathbb{V} l'insieme di queste classi di equivalenza, ovvero l'insieme dei vettori liberi. Possiamo immaginare \mathbb{V} come l'insieme di tutti i vettori di qualsivoglia direzione, lunghezza e verso ma “privi di punto di applicazione”, ovvero liberi di essere applicati in un punto a piacimento.

1.1.1. *Somme tra vettori liberi e prodotti per scalari.* Dati due vettori geometrici \mathbf{v} e \mathbf{u} , è intuitivo definire la loro *somma* $\mathbf{v} + \mathbf{u}$ concatenando \mathbf{v} con \mathbf{u} : se immaginiamo questi vettori come spostamenti, ha senso che lo spostamento $\mathbf{v} + \mathbf{u}$ a partire da un certo punto O corrisponda al punto in cui si arriva spostandosi prima di \mathbf{u} e poi di \mathbf{v} , o equivalentemente prima di \mathbf{v} e poi di \mathbf{u} . Più precisamente, si applica prima \mathbf{u} ad un punto generico, sia O , e si ottiene il vettore \overrightarrow{OA} con testa nella nostra tappa intermedia A , e, consecutivamente, si applica un rappresentante di \mathbf{v} in A che punta verso la destinazione finale B : il vettore \overrightarrow{OB} che descrive lo spostamento da O a B è, per definizione, un rappresentante di $\mathbf{v} + \mathbf{u}$. La costruzione ora descritta corrisponde alla cosiddetta *regola del parallelogramma*, ed è equivalente a immaginare \mathbf{u} e \mathbf{v} applicati allo stesso punto O e costruire su di essi un parallelogramma, la cui diagonale con estremo in O corrisponde al vettore somma avente coda in O e testa in B :



È evidente dalla costruzione stessa che l'operazione è commutativa, ovvero $\mathbf{v} + \mathbf{u} = \mathbf{u} + \mathbf{v}$, e tale per cui $\mathbf{v} + \mathbf{0} = \mathbf{v}$. L'operazione di somma che abbiamo definito è anche *associativa*. Se abbiamo tre vettori \mathbf{v} , \mathbf{u} e \mathbf{w} , allora

$$\mathbf{v} + (\mathbf{u} + \mathbf{w}) = \mathbf{v} + \mathbf{u} + \mathbf{w} = (\mathbf{v} + \mathbf{u}) + \mathbf{w}.$$

Dato un vettore \mathbf{v} , denotiamo con $-\mathbf{v}$ il vettore geometrico tale per cui $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$ e diciamo che $-\mathbf{v}$ è l'*opposto* di \mathbf{v} : questo vettore corrisponde naturalmente ad una freccia con stessa direzione e verso di \mathbf{v} , ma verso opposto.

È possibile anche introdurre l'operazione *moltiplicazione per uno scalare* $c \in \mathbb{R}$: il vettore $c\mathbf{v}$ ha stessa direzione di \mathbf{v} , stesso verso se $c > 0$ e verso opposto se $c < 0$, e lunghezza uguale a $|c|$ volte quella di \mathbf{v} . Se $c = 0$, allora $c\mathbf{v} = \mathbf{0}$. Un vettore $-\mathbf{v}$ può essere interpretato come il risultato del prodotto di \mathbf{v} per -1 , sicché la differenza tra due vettori è semplicemente la somma del primo con l'opposto del secondo, $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v})$. È facile verificare che, se $a, b \in \mathbb{R}$, allora $(a+b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$, $(ab)\mathbf{v} = a(b\mathbf{v})$ e, dati due vettori \mathbf{v} e \mathbf{u} , $a(\mathbf{v} + \mathbf{u}) = a\mathbf{v} + a\mathbf{u}$ ¹. Quello che abbiamo costruito è uno *spazio vettoriale di vettori geometrici*, i cui elementi possono essere sommati tra loro o moltiplicati per uno scalare in modo da ottenere un nuovo vettore geometrico. Grazie a queste regole possiamo combinare i vettori geometrici per ottenerne di nuovi, o cambiarne il verso. Questa struttura algebrica è molto semplice, ma anche ubiqua.

1.2. Spazi vettoriali. L'idea ora è tentare di caratterizzare cosa sia uno spazio vettoriale senza necessariamente avere in mente una rappresentazione pittorica. Vorremmo cioè astrarre le regole introdotte sopra per definire un ente matematico che non abbia necessariamente a che fare con uno spazio di frecce nello spazio, ma che si comporti come i vettori liberi introdotti sopra. Tale oggetto è lo *spazio vettoriale*. La costruzione di uno spazio vettoriale richiede l'esistenza di un insieme \mathbb{V} di "vettori" e di un altro insieme di "scalari", \mathbb{K} . Ci limiteremo ad assumere che \mathbb{K} sia, a seconda dei casi \mathbb{Q} , \mathbb{R} o \mathbb{C} : ciò che importa, in realtà, è che abbia la struttura di , in modo da dare un senso a espressioni come $a+b$ o ab quando $a, b \in \mathbb{K}$. La definizione di spazio vettoriale è data come segue.

DEFINIZIONE 1.1 (Spazio vettoriale). *Uno spazio vettoriale su un campo \mathbb{K} è un insieme non vuoto \mathbb{V} tale per cui esistono due operazioni binarie, ovvero $+$: $\mathbb{V} \times \mathbb{V} \rightarrow \mathbb{V}$, detta somma, e \cdot : $\mathbb{K} \times \mathbb{V} \rightarrow \mathbb{V}$, detta prodotto per uno scalare, con le seguenti proprietà:*

- la somma $+$ è commutativa e associativa; esiste un elemento neutro $\mathbf{0} \in \mathbb{V}$, ovvero se $\mathbf{v} \in \mathbb{V}$, allora $\mathbf{v} + \mathbf{0} = \mathbf{0} + \mathbf{v} = \mathbf{v}$; esiste l'elemento opposto, ovvero se $\mathbf{v} \in \mathbb{V}$ esiste $-\mathbf{v} \in \mathbb{V}$ tale per cui $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$;
- il prodotto \cdot è distributivo sulla somma di vettori (ovvero $a \cdot (\mathbf{v} + \mathbf{u}) = a \cdot \mathbf{v} + a \cdot \mathbf{u}$ per $a \in \mathbb{K}$ e $\mathbf{v}, \mathbf{u} \in \mathbb{V}$) e sulla somma di scalari (ovvero $(a+b) \cdot \mathbf{v} = a \cdot \mathbf{v} + b \cdot \mathbf{v}$ per $a, b \in \mathbb{K}$ e $\mathbf{v} \in \mathbb{V}$); infine l'elemento neutro 1 di \mathbb{K} è tale per cui $1 \cdot \mathbf{v} = \mathbf{v}$.

¹Si noti che le regole introdotte sono analoghe alle regole di manipolazione delle frecce che rappresentavano i numeri complessi nel piano di Argand.

In seguito ometteremo spesso il simbolo \cdot , scrivendo $a\mathbf{v}$ per intendere $a \cdot \mathbf{v}$. La definizione precedente si applica ai vettori geometrici, ed è anzi ispirata da essi. Essa si applica anche ai campi numerici \mathbb{R} e \mathbb{C} che sono spazi vettoriali su \mathbb{R} (e \mathbb{C} è spazio vettoriale su \mathbb{C}).

Un esempio di spazio vettoriale è dato da $V = \mathbb{R}^2$, ovvero lo spazio delle coppie di numeri nella forma

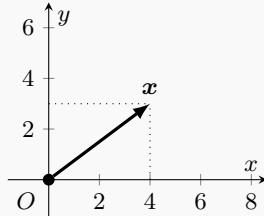
$$(1.1) \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad x_1, x_2 \in \mathbb{R}.$$

Dati due vettori \mathbf{x}_1 e \mathbf{x}_2 , una loro combinazione lineare con coefficienti reali c_1 e c_2 si ottiene come

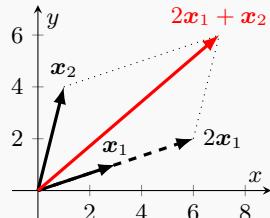
$$(1.2) \quad \mathbf{x}_1 = \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} x_{21} \\ x_{22} \end{pmatrix} \Rightarrow c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2 = \begin{pmatrix} c_1 x_{11} + c_2 x_{21} \\ c_1 x_{12} + c_2 x_{22} \end{pmatrix}.$$

Le operazioni “somma tra vettori” e “prodotto per scalare” soddisfano i requisiti della definizione di spazio vettoriale.

Esiste una corrispondenza tra \mathbb{R}^2 e il cosiddetto *piano cartesiano*. Nel piano cartesiano, si hanno due assi, quello delle *ascisse* (spesso detto *asse x*), orizzontale, e quello delle *ordinate* (spesso detto *asse y*), verticale e perpendicolare all’asse delle ascisse. I due assi si incrociano in un punto O detto *origine*. Dato questo sistema di riferimento, un punto X nel piano ha coordinate $(\frac{x_1}{x_2})$ se dista x_1 dall’asse delle ordinate e x_2 dall’asse delle ascisse. Esiste quindi una corrispondenza biunivoca tra i punti del piano cartesiano e \mathbb{R}^2 : ogni vettore $\mathbf{x} \in \mathbb{R}^2$ può essere associato ad un punto con coordinate date proprio dalle componenti di \mathbf{x} , e possiamo in effetti scrivere \mathbf{x} per significare il punto X nel piano:



Questa rappresentazione grafica aiuta a visualizzare le operazioni tra vettori nello spazio \mathbb{R}^2 . Per esempio, nella figura sotto $\mathbf{x}_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ e $\mathbf{x}_2 = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$, ed in rosso è rappresentata $2\mathbf{x}_1 + \mathbf{x}_2 = \begin{pmatrix} 7 \\ 6 \end{pmatrix}$.



Questo tipo di rappresentazione grafica tornerà utile in seguito.

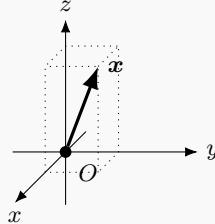
In analogia con il caso precedente, anche $V = \mathbb{R}^3$, insieme delle triplete di numeri reali, può essere dotato della struttura di spazio vettoriale. Un generico elemento di questo spazio ha la forma

$$(1.3) \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad x_1, x_2, x_3 \in \mathbb{R}.$$

Dati due vettori \mathbf{v} e \mathbf{u} in \mathbb{R}^3 , una loro combinazione lineare con coefficienti reali a e b si ottiene come

$$(1.4) \quad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \Rightarrow a\mathbf{v} + b\mathbf{u} = \begin{pmatrix} av_1 + bu_1 \\ av_2 + bu_2 \\ av_3 + bu_3 \end{pmatrix} \in \mathbb{R}^3.$$

Si può verificare che tutte le proprietà necessarie per poter considerare \mathbb{R}^3 uno spazio vettoriale sono verificate. Inoltre, come per \mathbb{R}^2 , è possibile costruire una corrispondenza biunivoca tra \mathbb{R}^3 e uno spazio cartesiano.



Per introdurre tale spazio, si parte da un piano cartesiano, con un suo asse x e un suo asse y , e si aggiunge un terzo asse, detto *asse z* , perpendicolare all'asse delle ascisse e a quello delle coordinate e passante per la stessa origine O . Questo asse viene orientato in modo che, osservando il piano cartesiano, la direzione positiva punti verso chi guarda. Dato questo sistema di riferimento, un punto X nello spazio ha coordinate come in Eq. (1.3) se dista x_1 dal piano in cui giacciono gli assi y e z , dista x_2 dal piano in cui giacciono gli assi x e z , e dista x_3 dal piano in cui giacciono gli assi x e y . Esiste così una corrispondenza biunivoca tra i punti dello spazio cartesiano e \mathbb{R}^3 : ogni vettore $\mathbf{x} \in \mathbb{R}^3$ può essere associato ad un punto X con coordinate date proprio dagli elementi ordinati di \mathbf{x} , e viceversa. Un vettore $\mathbf{x} \in \mathbb{R}^3$ può quindi essere rappresentato come una freccia che parte dall'origine O e arriva nel punto X associato a \mathbf{x} . Questa associazione tra \mathbb{R}^d e uno spazio d -dimensionale con un opportuno riferimento cartesiano vale per ogni d (anche se naturalmente non esiste una rappresentazione grafica per $d > 3$). Torneremo su questo punto in futuro, perché di grande importanza.

Se $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{V}$ e $c_1, \dots, c_k \in \mathbb{K}$, il vettore $\mathbf{v} = c_1\mathbf{v}_1 + \dots + c_k\mathbf{v}_k \in \mathbb{V}$ si dice essere una *combinazione lineare* di $\mathbf{v}_1, \dots, \mathbf{v}_k$ a coefficienti c_1, \dots, c_k . Il concetto di combinazione lineare è cruciale e verrà usato per caratterizzare molte proprietà di uno spazio vettoriale. Con un leggero abuso, utilizzeremo la notazione compatta

$$(1.5) \quad c_1\mathbf{v}_1 + \dots + c_k\mathbf{v}_k =: \sum_{i=1}^k c_i\mathbf{v}_i.$$

Chiamiamo l'insieme di tutte le combinazioni lineari dei vettori $\mathcal{V} := \{\mathbf{v}_i\}_{i=1}^k$ lo *span* di $\{\mathbf{v}_i\}_{i=1}^k$ e scriviamo $\text{Span}[\mathcal{V}]$: esso è cioè l'insieme di tutti i vettori che possono scriversi come in Eq. (1.5) al variare dei coefficienti c_1, \dots, c_k .

DEFINIZIONE 1.2 (Lineare dipendenza). *I vettori $\{\mathbf{v}_i\}_{i=1}^k$ sono linearmente dipendenti se esistono degli scalari $\{c_i\}_{i=1}^k$ non tutti nulli tali per cui*

$$(1.6) \quad \sum_{i=1}^k c_i\mathbf{v}_i = \mathbf{0};$$

diversamente si dicono linearmente indipendenti.

Per definizione, il singolo vettore \mathbf{v} è linearmente dipendente se e solo se uguale al vettore nullo. Viceversa, se $\mathbf{v}_2 = c\mathbf{v}_1$, \mathbf{v}_1 e \mathbf{v}_2 sono linearmente dipendenti: due vettori linearmente dipendenti sono anche detti *paralleli*. Tre vettori linearmente dipendenti si dicono, invece, *complanari*. La definizione implica che in un insieme di vettori linearmente dipendenti, almeno uno di essi può sempre esprimersi come combinazione lineare degli altri. Infine, vale la seguente proposizione.

PROPOSIZIONE 1.1. *Se $\{\mathbf{v}_i\}_{i=1}^k$ sono linearmente indipendenti, allora, dati due set di scalari in $\mathbb{K} \{a_i\}_{i=1}^k$ e $\{b_i\}_{i=1}^k$*

$$(1.7) \quad \sum_{i=1}^k a_i \mathbf{v}_i = \sum_{i=1}^k b_i \mathbf{v}_i \Leftrightarrow a_i = b_i \ \forall i = 1, \dots, k.$$

Infine, possiamo introdurre il concetto di *base*.

DEFINIZIONE 1.3 (Base di uno spazio vettoriale). *Il set $\mathcal{B} := \{\mathbf{b}_i\}_{i=1}^d$ di d vettori di \mathbb{V} è una base se essi sono linearmente indipendenti e se $\mathbb{V} = \text{Span}[\mathcal{B}]$.*

Questo significa che ogni vettore $\mathbf{v} \in \mathbb{V}$ può scriversi *in maniera unica* (per via della Proposizione 1.1) come $\mathbf{v} = \sum_{i=1}^d v_i \mathbf{b}_i$, con coefficienti v_i che sono detti *coordinate* di \mathbf{v} secondo la base \mathcal{B} , e si può rappresentare \mathbf{v} come una collezione di questi coefficienti

$$\mathbf{v} \mapsto \mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_d \end{pmatrix}.$$

È importante ricordare che, se $\mathbf{v} \neq \mathbf{0}$, le coordinate di \mathbf{v} dipendono dalla base \mathcal{B} . Tuttavia, a base fissata, c'è una corrispondenza biunivoca tra \mathbb{V} e \mathbb{K}^d , e un vettore $\mathbf{v} \in \mathbb{V}$ può essere identificato con le sue coordinate $\mathbf{v} \in \mathbb{K}^{d2}$.

La definizione di base comporta anche che non è possibile avere un set di vettori indipendenti con più di d elementi. Vale infatti il seguente

LEMMA 1.2 (Steinitz). *Sia $\mathcal{B} := \{\mathbf{b}_i\}_{i=1}^d$ una base di \mathbb{V} . Allora ogni set $\{\mathbf{u}_i\}_{i=1}^m$ di $m > d$ vettori di \mathbb{V} è costituito da elementi linearmente dipendenti. Di conseguenza, ogni base di \mathbb{V} ha d elementi: il numero d prende il nome di dimensione di \mathbb{V} , e si scrive $d = \dim(\mathbb{V})$*

Data una base \mathcal{B} , è possibile quindi costruire un *isomorfismo* tra un \mathbb{K} -spazio vettoriale d -dimensionale \mathbb{V} e \mathbb{K}^d , semplicemente associando ad ogni vettore \mathbf{v} le sue coordinate \mathbf{v} : la combinazione lineare $c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2$ di due vettori $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{V}$, con $c_1, c_2 \in \mathbb{K}$, corrisponde alla combinazione lineare $c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2$ dei vettori delle loro coordinate in \mathbb{K}^n , e viceversa. Questo ha una utilità rappresentativa importante. Se $d = 2$ e $\mathbb{K} = \mathbb{R}$, per esempio, possiamo rappresentare nel piano i vettori di \mathbf{v} utilizzando la rappresentazione cartesiana; possiamo in maniera simile rappresentare nello spazio i vettori di un \mathbb{R} -spazio vettoriale con $n = 3$.

Lo spazio vettoriale \mathbb{R}^d su \mathbb{R} è costituito da vettori d -dimensionali nella forma

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}.$$

In questo caso è facilmente identificata una *base canonica* $\mathcal{B} = \{\mathbf{e}_i\}_{i=1}^d$ di d vettori

$$(1.8) \quad \mathbf{e}_1 := \begin{pmatrix} 1 \\ 0 \\ \ddots \\ 0 \end{pmatrix} \quad \mathbf{e}_2 := \begin{pmatrix} 0 \\ 1 \\ \ddots \\ 0 \end{pmatrix} \quad \dots \quad \mathbf{e}_d := \begin{pmatrix} 0 \\ 0 \\ \ddots \\ 1 \end{pmatrix}$$

che permette di scrivere in maniera univoca qualunque vettore $\mathbf{x} \in \mathbb{R}^d$ come

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} = x_1 \mathbf{e}_1 + \dots + x_d \mathbf{e}_d \mapsto \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}.$$

La dimensione di \mathbb{R}^d è pertanto d . Si noti che è possibile scegliere un'altra base e scrivere una diversa rappresentazione di \mathbf{x} secondo questa base. Per fissare le idee, consideriamo per esempio $d = 2$, e si

²In queste note, tenterò di essere consistente ed adottare un grassetto corsivo \mathbf{v} per i vettori e un grassetto normale \mathbf{v} per la *rappresentazione* di tali vettori in termini delle loro coordinate via una certa base.

abbia un vettore

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2.$$

L'insieme $\mathcal{B}_b = \{\mathbf{b}_1, \mathbf{b}_2\}$ con

$$\mathbf{b}_1 = \begin{pmatrix} 1 & 1 \end{pmatrix}, \quad \mathbf{b}_2 = \begin{pmatrix} 1 & -1 \end{pmatrix}$$

è anch'esso una base, e possiamo sempre scrivere

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{x_1 + x_2}{2} \mathbf{b}_1 + \frac{x_1 - x_2}{2} \mathbf{b}_2 \mapsto \mathbf{x} = \begin{pmatrix} \frac{x_1 + x_2}{2} \\ \frac{x_1 - x_2}{2} \end{pmatrix}.$$

Come si vede, secondo questa nuova base le coordinate sono cambiate.

Se si ha un insieme di $k < \dim V$ vettori linearmente indipendenti, si può sempre usarli come elementi di una base, che però va naturalmente completata.

TEOREMA 1.3 (Teorema della base incompleta). *Sia V un \mathbb{K} -spazio vettoriale di dimensione d , e sia $\{\mathbf{v}_i\}_{i=1}^k$ un insieme di $k < d$ vettori linearmente indipendenti. È sempre possibile completare questo insieme con $n - k$ vettori, siano $\mathbf{v}_{k+1}, \dots, \mathbf{v}_n$, in modo che $\{\mathbf{v}_i\}_{i=1}^n$ sia una base.*

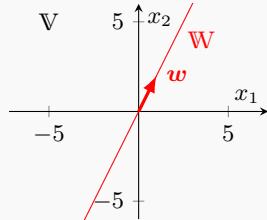
DIMOSTRAZIONE. Sia $\mathcal{B}_b = \{\mathbf{b}\}_{i=1}^d$ una base di V e si consideri l'insieme $\mathcal{S} := \{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{b}_1, \dots, \mathbf{b}_d\}$. Essendo tale insieme composto da $d + k$ vettori, questi saranno linearmente dipendenti, e d'altra parte $V = \text{Span}[\mathcal{S}]$ (dato che \mathcal{S} contiene una base). Possiamo quindi estrarre da \mathcal{S} una base procedendo algoritmicamente in questo modo: allo step n -esimo, si valuta se l' n -esimo vettore della lista è linearmente dipendente dai primi $n - 1$: in caso affermativo, si elimina dalla lista, diversamente si mantiene. Questo algoritmo manterrà i primi k elementi (per ipotesi) e scarterà $d - k$ elementi di \mathcal{B}_b ottenendo infine una base completa. \square

1.2.1. Sottospazi vettoriali. Dato uno spazio vettoriale V , è possibile identificare in alcuni casi un *sottospazio* vettoriale W , ovvero un sottoinsieme di V che è chiuso sotto le operazioni di somma e prodotto per uno scalare e si comporta a tutti gli effetti come uno spazio vettoriale a sé.

DEFINIZIONE 1.4 (Sottospazio vettoriale). *Dato un \mathbb{K} -spazio vettoriale V , un sottoinsieme non vuoto $W \subset V$ è un sottospazio vettoriale di V se, per ogni $\mathbf{w}, \mathbf{w}' \in W$, $\mathbf{w} + \mathbf{w}' \in W$, e per ogni $c \in \mathbb{K}$, se $\mathbf{w} \in W$ allora $c\mathbf{w} \in W$.*

Per esempio, se $\mathcal{V} := \{\mathbf{v}_i\}_{i=1}^k \subseteq V$, $\text{Span}[\mathcal{V}]$ è un sottospazio di V (in particolare, lo è anche per $k = 1$). La dimensione di un sottospazio è sempre limitata superiormente dalla dimensione dello spazio in cui vive: un insieme di vettori di V non può generare uno spazio di dimensione più grande di quella di V (per definizione di dimensione). È interessante notare che se U e W sono sottospazi vettoriali di V , allora $U \cap W$ è anche sottospazio vettoriale di V (mentre la loro unione, in generale, non lo è).

Se $V = \mathbb{R}^2$, per esempio, dato un vettore non nullo $\mathbf{w} \in \mathbb{R}^2$ un sottospazio vettoriale può costruirsi semplicemente considerando $W = \{\mathbf{v} \in V : \mathbf{v} = c\mathbf{w}, c \in \mathbb{R}\}$. Rappresentando i punti di W sul piano cartesiano, esso corrisponde ad una retta lungo la direzione di \mathbf{w} :



Dati due sottospazi vettoriali \mathbb{U} e \mathbb{W} di \mathbb{V} , è possibile costruire lo spazio *somma* come

$$\mathbb{U} + \mathbb{W} := \{\mathbf{u} + \mathbf{w} : \mathbf{u} \in \mathbb{U}, \mathbf{w} \in \mathbb{W}\} \subseteq \mathbb{V}.$$

Vale il seguente teorema.

TEOREMA 1.4 (Grassmann). *Siano \mathbb{U} e \mathbb{W} sottospazi vettoriali di \mathbb{V} , spazio vettoriale di dimensione finita. Allora*

$$(1.9) \quad \dim(\mathbb{U} + \mathbb{W}) = \dim(\mathbb{U}) + \dim(\mathbb{W}) - \dim(\mathbb{U} \cap \mathbb{W}).$$

DIMOSTRAZIONE. Indichiamo con n_u la dimensione di \mathbb{U} , con n_w la dimensione di \mathbb{W} e con d la dimensione di $\mathbb{U} \cap \mathbb{W}$. Sia $\mathcal{B} = \{\mathbf{v}_i\}_{i=1}^d$ la base dell'intersezione $\mathbb{U} \cap \mathbb{W}$: completiamola in \mathbb{U} aggiungendo l'insieme di vettori $\mathcal{B}_u = \{\mathbf{u}_i\}_{i=1}^{n_u-d}$ (in modo che $\mathcal{B}_u \cup \mathcal{B}$ sia una base di \mathbb{U}), e in \mathbb{W} aggiungendo l'insieme di vettori $\mathcal{B}_w = \{\mathbf{w}_i\}_{i=1}^{n_w-d}$ (in modo che $\mathcal{B}_w \cup \mathcal{B}$ sia una base di \mathbb{W}). Mostriamo che $\mathcal{B} \cup \mathcal{B}_u \cup \mathcal{B}_w$ è base di $\mathbb{U} + \mathbb{W}$. Per fare ciò, consideriamo una combinazione lineare tipo $\sum_{i=1}^{n_u-d} \mu_i \mathbf{u}_i + \sum_{i=1}^{n_w-d} \nu_i \mathbf{w}_i + \sum_{i=1}^d \lambda_i \mathbf{v}_i = \mathbf{0}$ cercando di mostrare che tutti i coefficienti λ_i , μ_i e ν_i devono essere nulli. Chiamando $\mathbf{v} = \sum_{i=1}^d \lambda_i \mathbf{v}_i$, $\mathbf{u} = \sum_{i=1}^{n_u-d} \mu_i \mathbf{u}_i$ e $\mathbf{w} = \sum_{i=1}^{n_w-d} \nu_i \mathbf{w}_i$, l'equazione stabilisce che $\mathbf{v} + \mathbf{u} + \mathbf{w} = \mathbf{0}$, ovvero $\mathbf{u} = -\mathbf{v} - \mathbf{w} \in \mathbb{W}$ che è possibile solo se $\mathbf{u} = \mathbf{0}$. Similmente, $\mathbf{w} = -\mathbf{v} - \mathbf{u} \in \mathbb{U}$, che è possibile solo se $\mathbf{w} = \mathbf{0}$. Infine, questo significa che $\mathbf{v} = \mathbf{0}$. Essendo i tre vettori \mathbf{u} , \mathbf{v} e \mathbf{w} ottenuti come combinazione lineare di vettori di base, essi sono nulli se i coefficienti della combinazione lineare sono nulli, il che prova che $\mathcal{B} \cup \mathcal{B}_u \cup \mathcal{B}_w$ è un insieme di vettori linearmente indipendenti e, dato che genera lo spazio somma, una base per esso. Essendo la dimensione di $\mathbb{U} + \mathbb{W}$ data dal numero di elementi di una sua base, essa è uguale a $n_u - d + n_w - d + d = n_u + n_w - d$. \square

Se $\mathbb{U} \cap \mathbb{W} = \{\mathbf{0}\}$, e quindi $\dim(\mathbb{U} \cap \mathbb{W}) = 0$, allora si dice che $\mathbb{U} + \mathbb{W}$ è una *somma diretta* di \mathbb{U} e \mathbb{W} e si indica con $\mathbb{U} \oplus \mathbb{W}$. Come in una somma usuale, spazio consiste di tutti i vettori \mathbf{v} nella forma $\mathbf{v} = \mathbf{u} + \mathbf{w}$ con $\mathbf{u} \in \mathbb{U}$, $\mathbf{w} \in \mathbb{W}$. Tuttavia, nel caso di una somma diretta, questa decomposizione è unica: se infatti assumessimo che esiste un'altra coppia tale per cui $\mathbf{v} = \mathbf{u}' + \mathbf{w}'$, con $\mathbf{u}' \in \mathbb{U}$, $\mathbf{w}' \in \mathbb{W}$, allora dovremmo avere $\mathbf{u} + \mathbf{w} = \mathbf{u}' + \mathbf{w}'$ e quindi $\mathbf{u} - \mathbf{u}' = \mathbf{w} - \mathbf{w}'$. Ma questa quantità può essere solo $\mathbf{0}$, unico elemento nell'intersezione dei due spazi.

2. Lo spazio delle matrici

Un esempio particolarmente importante e prototipico di spazio vettoriale su \mathbb{K} è costituito dallo spazio delle matrici di elementi in \mathbb{K} stesso. Consideriamo due interi positivi m e n . Una *matrice* di $m \times n$ elementi in \mathbb{K} è una tabella del tipo

$$(1.10) \quad \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \quad a_{\mu\nu} \in \mathbb{K} \quad \forall \mu, \nu.$$

In $a_{\mu\nu}$ il primo indice denota la riga dell'elemento nella tabella, il secondo la sua colonna. Lo spazio delle matrici di $m \times n$ elementi in \mathbb{K} si denota $\mathbb{M}_{m,n}(\mathbb{K})$ e scriveremo $\mathbf{A} = (a_{\mu\nu})_{\mu\nu} \in \mathbb{M}_{m,n}(\mathbb{K})$ per indicare la matrice $m \times n$ che ha $a_{\mu\nu}$ come elemento nella μ -esima riga e ν -esima colonna. Lo spazio $\mathbb{M}_{m,n}(\mathbb{K})$ è dotato della struttura di spazio vettoriale se introduciamo le seguenti operazioni. Date due matrici $\mathbf{A} = (a_{\mu\nu})_{\mu\nu}$, $\mathbf{B} = (b_{\mu\nu})_{\mu\nu} \in \mathbb{M}_{m,n}(\mathbb{K})$, e uno scalare $c \in \mathbb{K}$, allora definiamo

$$(1.11) \quad \mathbf{A} + \mathbf{B} := (a_{\mu\nu} + b_{\mu\nu})_{\mu\nu} \in \mathbb{M}_{m,n}(\mathbb{K}), \quad c\mathbf{A} := (ca_{\mu\nu})_{\mu\nu} \in \mathbb{M}_{m,n}(\mathbb{K}).$$

Queste operazioni soddisfano le proprietà richieste per rendere $\mathbb{M}_{m,n}(\mathbb{K})$ uno spazio vettoriale secondo la Definizione 1.1.

Lo spazio $\mathbb{M}_{m,n}(\mathbb{K})$ ha dimensione mn . Per vedere ciò basta osservare che una base è costituita dalle matrici $\mathbf{E}_{ij} = (\delta_{i\mu}\delta_{j\nu})_{\mu\nu}$, tali per cui tutti i loro elementi sono nulli eccezion fatta

per l'elemento nella i -esima riga e j -esima colonna, dove compare 1. Usando questa base, ogni matrice $\mathbf{A} = (a_{\mu\nu})_{\mu\nu}$ si può scrivere, in maniera univoca, come

$$(1.12) \quad \mathbf{A} = \sum_{\mu=1}^n \sum_{\nu=1}^n a_{\mu\nu} \mathbf{E}_{\mu\nu}.$$

Ogni matrice $\mathbf{A} \in \mathbb{M}_{m,n}(\mathbb{K})$ può essere associata ad una matrice $\mathbf{A}^\top \in \mathbb{M}_{n,m}(\mathbb{K})$ detta *trasposta* ottenuta invertendo righe con colonne:

$$(1.13) \quad \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \Rightarrow \mathbf{A}^\top = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix}$$

Si noti che l'operazione di trasposizione è quindi tale da far passare, in generale, da uno spazio vettoriale ad un altro, ovvero $\bullet^\top: \mathbb{M}_{n,m}(\mathbb{K}) \rightarrow \mathbb{M}_{m,n}(\mathbb{K})$. Matrici con $n = 1$ sono dette *vettori colonna*, e per esse viene tipicamente usata la notazione più compatta $\mathbb{M}_{m,1}(\mathbb{K}) \equiv \mathbb{K}^m$ (abbiamo già incontrato, nella precedente sezione, lo spazio vettoriale \mathbb{R}^n , e \mathbb{R}^2 in special modo, che è un caso particolare della discussione generale presentata qui).

2.1. Prodotto tra matrici. Oltre alle operazioni di somma tra matrici e prodotto tra matrice e scalare, è possibile definire una nuova operazione, quella di *prodotto tra matrici*, che coinvolge, in generale, matrici in diversi spazi vettoriali. Consideriamo una matrice $\mathbf{A} = (a_{\mu\nu})_{\mu\nu} \in \mathbb{M}_{n,k}(\mathbb{K})$ e una matrice $\mathbf{B} = (b_{\mu\nu})_{\mu\nu} \in \mathbb{M}_{k,m}(\mathbb{K})$. Allora è possibile introdurre una matrice $\mathbf{C} = (c_{\mu\nu})_{\mu\nu} := \mathbf{AB} \in \mathbb{M}_{n,m}(\mathbb{K})$ tale che

$$(1.14) \quad c_{\mu\nu} = \sum_{\rho=1}^k a_{\mu\rho} b_{\rho\nu}.$$

In particolare, dati due n -vettori colonna $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$

$$(1.15) \quad \mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n,$$

il prodotto

$$(1.16) \quad \mathbf{u}^\top \mathbf{v} = \mathbf{v}^\top \mathbf{u} = \sum_{\nu=1}^n u_\nu v_\nu \in \mathbb{R}.$$

è detto, per ragioni che vedremo in seguito, *prodotto scalare*. Chiameremo inoltre la quantità non negativa

$$(1.17) \quad \|\mathbf{v}\| := \sqrt{\sum_{\nu=1}^n v_\nu^2}$$

norma di \mathbf{v} . Torneremo in seguito sul concetto di prodotto scalare e norma.

L'operazione di prodotto tra matrici *non è abeliana* e anzi se $n \neq m$ il prodotto \mathbf{BA} è non definito: invertire l'ordine dei fattori quindi non è in generale ammesso. Anche qualora l'operazione fosse possibile, ovvero se $n = m$, $\mathbf{AB} \neq \mathbf{BA}$ in generale. Una speciale matrice in $\mathbb{M}_{n,n}(\mathbb{K})$, che ha il ruolo di “identità” rispetto al prodotto tra matrici, è la *matrice unità*

$$(1.18) \quad \mathbf{I}_n := \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \end{pmatrix}.$$

Le proprietà del prodotto tra matrici sono date dalla seguente Proposizione, la cui dimostrazione ometteremo (può essere ottenuta verificando esplicitamente la validità di ogni affermazione).

PROPOSIZIONE 2.1 (Proprietà del prodotto tra matrici). *Siano $\mathbf{A}, \mathbf{B} \in \mathbb{M}_{m,n}(\mathbb{K})$ e $\mathbf{C}, \mathbf{D} \in \mathbb{M}_{n,p}(\mathbb{K})$, e sia $c \in \mathbb{K}$. Allora*

- $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$;
- $\mathbf{A}(\mathbf{C} + \mathbf{D}) = \mathbf{AC} + \mathbf{AD}$;
- $\mathbf{A}(c\mathbf{C}) = c(\mathbf{AC}) = (c\mathbf{A})\mathbf{C}$;
- $\mathbf{A}\mathbf{I}_n = \mathbf{A} = \mathbf{I}_m\mathbf{A}$;
- $(\mathbf{AC})^\top = \mathbf{C}^\top\mathbf{A}^\top$;
- $(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$;
- $(\mathbf{AC})\mathbf{Q} = \mathbf{A}(\mathbf{CQ})$ se $\mathbf{Q} \in \mathbb{M}_{p,q}(\mathbb{K})$.

2.2. Matrici quadrate. Il campo $\mathbb{M}_{n,n}(\mathbb{K})$ corrisponde al caso di matrici quadrate ed ha una importanza speciale, dato che è legato, come vedremo dopo, ad un particolare tipo di applicazioni: lo denoteremo $\mathbb{M}_n(\mathbb{K})$, evitando il ridondante doppio pedice. Una matrice $\mathbf{A} \in \mathbb{M}_n(\mathbb{K})$ è detta *quadrata* di dimensione n . Una matrice quadrata $\mathbf{A} \in \mathbb{K}$ è *simmetrica* se $\mathbf{A} = \mathbf{A}^\top$ e *antisimmetrica* se $\mathbf{A} = -\mathbf{A}^\top$. Una matrice quadrata è *diagonale* se ha la forma

$$(1.19) \quad \mathbf{A} = (a_\mu \delta_{\mu\nu})_{\mu\nu} = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$$

ovvero se solo gli elementi fuori dalla sua diagonale sono nulli: la matrice \mathbf{I}_n è un caso particolare di matrice identità.

Data una matrice quadrata \mathbf{A} , si può ad essa associare un numero, detto *traccia* $\text{tr}(\mathbf{A})$, uguale alla somma dei suoi elementi sulla diagonale,

$$(1.20) \quad \text{tr}(\mathbf{A}) = \sum_{\mu=1}^n a_{\mu\mu}, \quad \mathbf{A} \in \mathbb{M}_n(\mathbb{K}).$$

Questo numero rimane inalterato se si prende la trasposta di \mathbf{A} , ovvero $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^\top)$, ed gode della cosiddetta *proprietà ciclica*, ovvero, date tre matrici $\mathbf{A} \in \mathbb{M}_{n,p}(\mathbb{K})$, $\mathbf{B} \in \mathbb{M}_{p,q}(\mathbb{K})$, $\mathbf{C} \in \mathbb{M}_{q,n}(\mathbb{K})$, si ha che

$$(1.21) \quad \text{tr}(\mathbf{ABC}) = \sum_{i=1}^n \sum_{j=1}^p \sum_{k=1}^q a_{ij} b_{jk} c_{ki} = \sum_{k=1}^q \sum_{i=1}^n \sum_{j=1}^p c_{ki} a_{ij} b_{jk} = \text{tr}(\mathbf{CBA}).$$

2.2.1. Matrice inversa. Le matrici quadrate hanno la caratteristica di avere il prodotto tra matrici come *operazione interna* del loro spazio $\mathbb{M}_n(\mathbb{K})$, ovvero, date $\mathbf{A}, \mathbf{B} \in \mathbb{M}_n(\mathbb{K})$, allora $\mathbf{AB} \in \mathbb{M}_n(\mathbb{K})$. Inoltre, il fatto che esista una sorta di identità rispetto alla moltiplicazione tra matrici in $\mathbb{M}_n(\mathbb{K})$ (ovvero \mathbf{I}_n) suggerisce di introdurre la seguente

DEFINIZIONE 2.1 (Matrice inversa). *Una matrice quadrata $\mathbf{A} \in \mathbb{M}_n(\mathbb{K})$ si dice invertibile se esiste una matrice quadrata in $\mathbb{M}_n(\mathbb{K})$, che denotiamo con \mathbf{A}^{-1} , tale che $\mathbf{A}^{-1}\mathbf{A} = \mathbf{AA}^{-1} = \mathbf{I}_n$.*

LEMMA 2.2 (Unicità). *Sia $\mathbf{A} \in \mathbb{M}_n(\mathbb{K})$. La sua corrispondente matrice inversa \mathbf{A}^{-1} , se esiste, è unica.*

DIMOSTRAZIONE. Supponiamo che la tesi non sia vera e procediamo per assurdo, ovvero ammettiamo che esista una matrice $\hat{\mathbf{A}}^{-1} \neq \mathbf{A}^{-1}$ tale che $\hat{\mathbf{A}}^{-1}\mathbf{A} = \mathbf{I}_n$. Allora $\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{AA}^{-1} = \hat{\mathbf{A}}^{-1}$ che è l'assurdo cercato. \square

Dalla definizione, $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$ e inoltre $\mathbf{I}_n^{-1} = \mathbf{I}_n$. Infine, vale la seguente

PROPOSIZIONE 2.3 (Inversa del prodotto). *Siano $\mathbf{A}, \mathbf{B} \in \mathbb{M}_n(\mathbb{K})$ entrambe dotate di inversa. Allora $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \in \mathbb{M}_n(\mathbb{K})$.*

DIMOSTRAZIONE. $(\mathbf{B}^{-1}\mathbf{A}^{-1})\mathbf{AB} = \mathbf{I}_n \Rightarrow \mathbf{B}^{-1}\mathbf{A}^{-1} = (\mathbf{AB})^{-1}$ da definizione. \square

Non tutte le matrici quadrate sono invertibili. Per esempio, consideriamo la matrice

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

Supponiamo di voler trovare la matrice

$$\mathbf{A}^{-1} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

tale che $\mathbf{AA}^{-1} = \mathbf{I}_2$. Imporre questa condizione fornisce

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ 0 & 0 \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

che è un assurdo: la matrice \mathbf{A} non è quindi invertibile.

Sia dato $\mathbb{M}_n(\mathbb{K})$ e identifichiamo due sottospazi, ovvero

$$\mathbb{S}_n(\mathbb{K}) = \{\mathbf{A} \in \mathbb{M}_n(\mathbb{K}) : \mathbf{M} = \mathbf{M}^\top\}, \quad \mathbb{A}_n(\mathbb{K}) = \{\mathbf{M} \in \mathbb{M}_n(\mathbb{K}) : \mathbf{M} = -\mathbf{M}^\top\},$$

che corrispondono rispettivamente all'insieme delle matrici simmetriche e all'insieme delle matrici antisimmetriche. Questi due insiemi sono entrambi *spazi vettoriali*, come si può facilmente verificare. Inoltre,

$$\dim(\mathbb{S}_n) = \frac{n(n+1)}{2}$$

dato che per costruire una matrice simmetrica abbiamo bisogno di specificare $\frac{n(n+1)}{2}$ valori, mentre viceversa

$$\dim(\mathbb{A}_n) = \frac{n(n-1)}{2}$$

dato che per costruire una matrice antisimmetrica abbiamo bisogno di specificare $\frac{n(n-1)}{2}$ valori (i valori sulla diagonale sono tutti nulli). Infine, $\mathbb{S}_n \cap \mathbb{A}_n = \{\mathbf{0}\}$. Osservando che

$$\dim(\mathbb{S}_n) + \dim(\mathbb{A}_n) + \dim(\mathbb{S}_n \cap \mathbb{A}_n) = \dim(\mathbb{S}_n) + \dim(\mathbb{A}_n) = \frac{n(n+1)}{2} + \frac{n(n-1)}{2} = n^2 = \dim(\mathbb{M}_n),$$

possiamo concludere che $\mathbb{M}_n(\mathbb{K}) = \mathbb{S}_n(\mathbb{K}) \oplus \mathbb{A}_n(\mathbb{K})$. In effetti ogni matrice $\mathbf{M} \in \mathbb{M}_n(\mathbb{K})$ si può scrivere in maniera univoca come

$$\mathbf{M} = \mathbf{S} + \mathbf{A}, \quad \mathbf{S} = \frac{\mathbf{M} + \mathbf{M}^\top}{2} \in \mathbb{S}_n(\mathbb{K}), \quad \mathbf{A} = \frac{\mathbf{M} - \mathbf{M}^\top}{2} \in \mathbb{A}_n(\mathbb{K}).$$

Il sottoinsieme di $\mathbb{M}_n(\mathbb{K})$ dato dalle matrici invertibili si denota $\mathrm{GL}_n(\mathbb{K})$ ed ha la struttura di *gruppo*. Se $\mathbb{K} = \mathbb{R}$, all'interno di $\mathrm{GL}_n(\mathbb{R})$ si può identificare $\mathrm{O}_n(\mathbb{R})$ delle matrici *ortogonalì*, ovvero delle matrici \mathbf{A} tali per cui $\mathbf{A}^{-1} = \mathbf{A}^\top$: anche questo sottoinsieme ha la struttura di gruppo ed è detto *gruppo ortogonale*, talvolta indicato come $\mathrm{O}(n)$.

2.2.2. Determinante. Ad ogni matrice quadrata può essere associato, oltre alla traccia, uno scalare in \mathbb{K} che, come vedremo, è estremamente ricco di informazioni sulle proprietà della matrice stessa. Questo scalare è il *determinante*. La sua definizione è relativamente complessa.

DEFINIZIONE 2.2 (Determinante). *Data una matrice quadrata $\mathbf{A} \in \mathbb{M}_n(\mathbb{K})$, il suo determinante è definito come*

$$(1.22) \quad \det(\mathbf{A}) := \sum_{\sigma \in \mathcal{S}_n} (-1)^{\mathbf{n}(\sigma)} \prod_{i=1}^n a_{i\sigma(i)},$$

dove la somma corre su tutte le permutazioni di n elementi e $\mathbf{n}(\sigma)$ è la parità della permutazione σ . Esso si indica talvolta come $|\mathbf{A}|$.

La complicazione e stranezza della definizione sta nella comparsa della somma su *tutte le permutazioni di n elementi* \mathcal{S}_n e in un prefattore che fa riferimento al numero di trasposizioni $\mathbf{n}(\sigma)$ che

producono σ , ovvero alla *parità* della permutazione. La definizione sopra è piuttosto convoluta e difficile da applicare. Estremamente utile è il seguente teorema, che non dimostriamo, dovuto a Pierre-Simon Laplace.

TEOREMA 2.4 (Teorema di Laplace). *Data una matrice quadrata $A \in \mathbb{M}_n(\mathbb{K})$, il determinante di A è uno scalare in \mathbb{K} definito ricorsivamente come*

$$(1.23) \quad \det(A) = \sum_{\nu=1}^n a_{\mu\nu} c_{\mu\nu} = \sum_{\nu=1}^n a_{\nu\mu} c_{\nu\mu} = (AC^\top)_{\mu\mu}, \quad c_{\mu\nu} := (-1)^{\mu+\nu} \det(A_{\mu\nu})$$

indipendentemente da μ , dove $A_{\mu\nu} \in \mathbb{K}^{(n-1) \times (n-1)}$ è il minore complementare di $a_{\mu\nu}$, ovvero la matrice $(n-1) \times (n-1)$ ottenuta da A rimuovendo la μ -esima riga e la ν -esima colonna. Inoltre, per definizione dato uno scalare $x \in \mathbb{K}$, $\det(x) = x$.

La matrice $C = (c_{\mu\nu})_{\mu\nu}$ è detta *matrice dei cofattori*, mentre il termine $\det(A_{\mu\nu})$ è detto *complemento algebrico* di $a_{\mu\nu}$.

La definizione sopra è come detto ricorsiva ma permette di ottenere l'espressione del determinante in tutti i casi partendo dal fatto che il determinante di uno scalare è semplicemente lo scalare stesso. Per esempio, per $n = 2$

$$(1.24) \quad A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \Rightarrow \det(A) = a_{11}a_{22} - a_{12}a_{21}.$$

Per $n = 3$ produce la cosiddetta *regola di Sarrus*

$$(1.25) \quad A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \Rightarrow \det(A) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12}.$$

La regola si può memorizzare immaginando di disporre gli elementi della matrice come in

$$\begin{array}{ccccc} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{array}$$

e notando che i termini con $+$ corrispondono alle diagonali “da alto a sinistra a basso a destra”, mentre i termini con $-$ corrispondono alle diagonali “da basso a sinistra ad alto a destra”.

Il determinante gode di una serie di proprietà interessanti, elencate nel seguente

TEOREMA 2.5 (Proprietà del determinante). *Sia $A \in \mathbb{M}_n(\mathbb{K})$. Allora*

- (1) se A è una matrice diagonale, $\det(A) = \prod_{i=1}^n a_{ii}$.
- (2) $\det(A) = \det(A^\top)$ (e di conseguenza proprietà riferite alle colonne si applicano anche alle righe).
- (3) Dato uno scalare $c \in \mathbb{K}$, $\det(cA) = c^n \det(A)$.
- (4) Sia \mathbf{a}^ν la ν -esima colonna di A e supponiamo di scrivere $\mathbf{a}^\nu = cv + \mathbf{u}$, con $v, \mathbf{u} \in \mathbb{K}^n$ e $c \in \mathbb{K}$, i.e., $a_{\mu\nu} = cv_\nu + u_\nu$. Allora

$$(1.26) \quad \det(A) = \det \begin{pmatrix} a_{11} & \dots & a_{1\nu} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{n\nu} & \dots & a_{nn} \end{pmatrix} = c \det \begin{pmatrix} a_{11} & \dots & v_1 & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & v_n & \dots & a_{nn} \end{pmatrix} + \det \begin{pmatrix} a_{11} & \dots & u_k & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & u_n & \dots & a_{nn} \end{pmatrix}.$$

Si dice che di conseguenza il determinante è multilineare.

- (5) Sia $B \in \mathbb{M}_n(\mathbb{K})$; vale la formula di Binet, $\det(AB) = \det(A)\det(B)$.

(6) Il determinante è alternante, ovvero guadagna un segno se due colonne vengono scambiate:

$$(1.27) \quad \det \begin{pmatrix} a_{11} & \dots & a_{1k} & \dots & a_{1k'} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nk} & \dots & a_{nk'} & \dots & a_{nn} \end{pmatrix} = - \det \begin{pmatrix} a_{11} & \dots & a_{1k'} & \dots & a_{1k} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nk'} & \dots & a_{nk} & \dots & a_{nn} \end{pmatrix}.$$

Di conseguenza, $\det(\mathbf{A}) = 0$ se \mathbf{A} ha due righe o colonne identiche.

Le proprietà sopra implicano una serie di fatti molto importanti, che spiegano perché il determinante è una quantità così centrale nello studio delle matrici. Anzitutto, consideriamo l'insieme $\{\mathbf{a}^\nu\}_{\nu=1}^n$ delle colonne. Se una colonna può essere espressa come combinazione lineare di due o più altre colonne, per esempio $\mathbf{a}^n = \sum_{i=1}^{n-1} c_i \mathbf{a}^i$ con c_1, \dots, c_{n-1} non tutti nulli, allora il determinante è nullo. Infatti, dalla proprietà (4) abbiamo che

$$\det(\mathbf{A}) = \det \begin{pmatrix} a_{11} & \dots & a_{1\nu} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{n\nu} & \dots & a_{nn} \end{pmatrix} = \sum_{i=1}^{n-1} c_i \det \begin{pmatrix} a_{11} & \dots & a_{1\nu} & \dots & a_{1i} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{n\nu} & \dots & a_{ni} \end{pmatrix} = 0,$$

dove l'ultima uguaglianza discende dalla proprietà (6). La stessa cosa, per via della proprietà (2), vale per le righe. In definitiva, $\det(\mathbf{A}) \neq 0$ se e solo se le sue righe sono linearmente indipendenti e le sue colonne sono linearmente indipendenti, ovvero se lo spazio generato dalle sue righe e dalle sue colonne ha dimensione n .

Dato che $\det(\mathbf{I}_n) = 1$, una conseguenza della formula di Binet è che, se \mathbf{A} è una matrice invertibile, $\det(\mathbf{A}^{-1}) = \frac{1}{\det(\mathbf{A})}$: questo significa che se $\det(\mathbf{A}) = 0$ l'inversa non può esistere (e in questo senso il determinante *determina* se una matrice è invertibile o meno). Il ruolo cruciale del determinante nel calcolo dell'inversa emerge esplicitamente nel seguente

TEOREMA 2.6. Sia $\mathbf{A} \in \mathbb{M}_n(\mathbb{K})$; sia \mathbf{C} la matrice dei suoi cofattori, ovvero la matrice $n \times n$ con elementi

$$c_{\mu\nu} = (-1)^{\mu+\nu} \det(\mathbf{A}_{\mu\nu}).$$

Allora

$$(1.28) \quad \mathbf{A}^\top \mathbf{C} = \det(\mathbf{A}) \mathbf{I}_n.$$

In particolare, se $\det(\mathbf{A}) \neq 0$, l'inversa di \mathbf{A} esiste ed è data da

$$(1.29) \quad \mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \mathbf{C}^\top.$$

DIMOSTRAZIONE. Il lemma è una conseguenza del teorema di Laplace. Infatti, sappiamo che $\det(\mathbf{A}) = \sum_\rho a_{\mu\rho} c_{\mu\rho}$. Occorre solo dimostrare che $\sum_\rho a_{\mu\rho} c_{\nu\rho} = 0$ per $\mu \neq \nu$. Questa espressione corrisponde al determinante di una matrice ottenuta da \mathbf{A} rimpiazzando la sua ν -esima riga con la sua μ -esima ed è nullo proprio perché la matrice risultante ha due righe uguali. \square

3. Sistemi di equazioni lineari

Avendo familiarizzato con le proprietà generali di $\mathbb{M}_{m,n}(\mathbb{K})$, possiamo ora considerare una prima importante applicazione, ovvero il problema della soluzione di *sistemi di equazioni lineari*, ovvero sistemi nella forma

$$(1.30) \quad \begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \end{cases}$$

in cui si assume che i coefficienti $a_{\mu\nu} \in \mathbb{K}$ e $b_\nu \in \mathbb{K}$ siano noti e l'obiettivo è ottenere $x_\nu \in \mathbb{K}$. Indicando con $\mathbf{A} = (a_{\mu\nu})_{\mu\nu} \in \mathbb{M}_{m,n}(\mathbb{K})$, $\mathbf{b} = (b_\nu)_\nu \in \mathbb{K}^m$, $\mathbf{x} = (x_\nu)_\nu \in \mathbb{K}^n$, possiamo scrivere il sistema sopra in una forma molto più compatta come

$$(1.31) \quad \mathbf{Ax} = \mathbf{b}.$$

Il sistema di equazioni è detto *omogeneo* se $\mathbf{b} = \mathbf{0}$, mentre diversamente è detto *non omogeneo*. Ogni sistema è detto *compatibile* se ha almeno una soluzione: notare che un sistema omogeneo ha sempre la soluzione *banale* $\mathbf{x} = \mathbf{0}$. Se è dato un sistema non omogeneo $\mathbf{Ax} = \mathbf{b}$, $\mathbf{b} \neq \mathbf{0}$, allora il sistema omogeneo ad esso associato è $\mathbf{Ax} = \mathbf{0}$. Le soluzioni di questa equazione corrispondono ad un *sottospazio vettoriale* detto *kernel* (o *nucleo*) della matrice \mathbf{A} , $\text{Ker}(\mathbf{A})$:

$$(1.32) \quad \text{Ker}(\mathbf{A}) := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} = \mathbf{0}\}.$$

Ecco un sottospazio vettoriale, dato che, se $\mathbf{x}_1, \mathbf{x}_2 \in \text{Ker}(\mathbf{A})$, allora $c_1\mathbf{x}_1 + c_2\mathbf{x}_2 \in \text{Ker}(\mathbf{A})$ per qualunque coppia di scalari $c_1, c_2 \in \mathbb{K}$. Il kernel di \mathbf{A} è sempre non vuoto dato che $\mathbf{0} \in \text{Ker}(\mathbf{A})$. La dimensione $\dim \text{Ker}(\mathbf{A})$ si dice *nullità* di \mathbf{A} . Se \mathbf{x}_b è soluzione del sistema non omogeneo e $\mathbf{x}_0 \in \text{Ker } \mathbf{A}$, una nuova soluzione del sistema non omogeneo può essere ottenuta considerando $\mathbf{x}_b + c\mathbf{x}_0$ per qualunque $c \in \mathbb{K}$. Questo significa che se $\mathbf{Ax} = \mathbf{b}$ è compatibile, ha soluzione unica solo se $\dim \text{Ker}(\mathbf{A}) = 0$, ovvero la soluzione del problema omogeneo è la sola $\mathbf{x}_0 = \mathbf{0}$

3.1. Metodo di eliminazione di Gauss. Risolvere un sistema tipo $\mathbf{Ax} = \mathbf{b}$ è un problema centrale dell'algebra lineare. L'idea generale è utilizzare delle semplici operazioni matriciali per poter infine ottenere una espressione per la soluzione (o l'insieme di soluzioni) \mathbf{x} , se essa esiste, o concludere che il sistema non ha soluzione se questo è il caso. Un approccio classico e istruttivo è il cosiddetto *metodo di eliminazione* elaborato da Carl Friedrich Gauss. Supponiamo di avere un sistema come in Eq. (1.30) e introduciamo la seguente *matrice orlata*, ottenuta "incollando", ovvero concatenando, la matrice \mathbf{A} con il vettore \mathbf{b} ,

$$(1.33) \quad \mathbf{H} = (h_{\mu\nu})_{\mu\nu} = (\mathbf{A} \mid \mathbf{b}) = \left(\begin{array}{ccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array} \right).$$

Ogni matrice siffatta corrisponde ad un sistema come in Eq. (1.30). Possiamo eseguire una serie di operazioni di riga su questa matrice che lasciano inalterato il problema. In particolare

- possiamo moltiplicare una riga per una costante $c \in \mathbb{K}$ che sia non nulla;
- possiamo scambiare due righe;
- possiamo aggiungere ad una riga un multiplo di un'altra riga.

Questo tipo di operazioni può essere utilizzato per manipolare \mathbf{H} , dato che ciascuna di esse non cambia, di fatto, il sistema di equazioni lineari associato. Il metodo consiste nell'utilizzare queste operazioni per semplificare quanto più possibile le equazioni, procedendo come segue.

Si considera h_{11} , detto *primo pivot*, assumendolo diverso da zero (se $h_{11} = 0$, si scambia la prima riga con un'altra che ha primo elemento non nullo, cosa sempre possibile se la variabile x_1 appare nel sistema). A questo punto, per ogni $\mu > 1$, si sottrare alla μ -esima riga \mathbf{H}_μ la prima riga con un adeguato moltiplicatore in modo da annullare tutti gli elementi $h_{\mu 1}$ con $\mu > 1$ (cioè sotto h_{11}) ovvero

$$\mathbf{H}_\mu \mapsto \mathbf{H}_\mu - \frac{h_{\mu 1}}{h_{11}} \mathbf{H}_1.$$

Si ordinano quindi le righe dalla seconda in giù in modo che la nuova seconda riga sia una delle righe con l'elemento non nullo più a sinistra. Questo elemento sarà il secondo pivot. Si ripete quindi l'operazione come sopra, in modo da ottenere zeri sotto al secondo pivot. Si passa così alla terza riga, e si ripete la procedura fino ad essere arrivati all'ultima riga. Per chiarire meglio il funzionamento dell'algoritmo, consideriamo un esempio.

Cerchiamo di risolvere il sistema

$$(1.34) \quad \begin{cases} 2x_1 + 4x_2 - 2x_3 = 2 \\ 4x_1 + 9x_2 - 3x_3 = 8 \\ -2x_1 - 3x_2 + 7x_3 = 10 \end{cases} \Leftrightarrow \begin{pmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \\ 10 \end{pmatrix}$$

così che la matrice orlata corrispondente è

$$(1.35) \quad \left(\begin{array}{ccc|c} 2 & 4 & -2 & 2 \\ 4 & 9 & -3 & 8 \\ -2 & -3 & 7 & 10 \end{array} \right).$$

Nella matrice considerata $h_{11} \neq 0$ e può essere il nostro primo pivot. Eseguiamo l'operazione richiesta dall'algoritmo e sostituiamo la μ -esima riga \mathbf{H}_μ con $\mathbf{H}_\mu - \frac{h_{\mu 1}}{h_{11}} \mathbf{H}_1$. Si ottiene quindi

$$(1.36) \quad \left(\begin{array}{ccc|c} 2 & 4 & -2 & 2 \\ 0 & 1 & 1 & 4 \\ 0 & 1 & 5 & 12 \end{array} \right)$$

Ripetiamo la procedura partendo dalla *seconda* riga verso il basso. Questa volta il pivot nel nostro esempio è $h_{22} = 1$. Si procede come sopra per le righe $\mu > 2$, ottenendo

$$(1.37) \quad \left(\begin{array}{ccc|c} 2 & 4 & -2 & 2 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 4 & 8 \end{array} \right)$$

A questo punto non possiamo procedere oltre. Possiamo però notare che la matrice ottenuta ha una struttura particolarmente semplice, che permette una soluzione rapida del sistema. Essa corrisponde all'insieme di equazioni

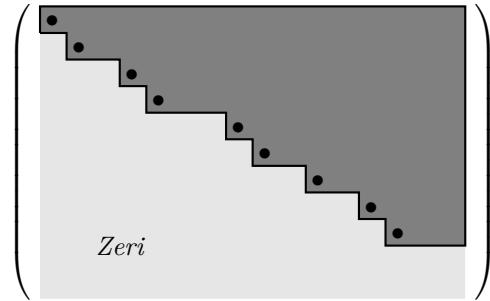
$$(1.38) \quad \begin{cases} 2x_1 + 4x_2 - 2x_3 = 2 \\ x_2 + x_3 = 4 \\ 4x_3 = 8 \end{cases}$$

che può essere risolto dal basso verso l'alto, ovvero risolvendo prima per x_3 , ottenendo $x_3 = 2$, che può essere sostituito nella seconda riga ottenendo $x_2 = 2$ e infine inserendo entrambe queste soluzioni nella prima equazione, che fornisce $x_1 = -1$, in modo che $(x_1, x_2, x_3) = (-1, 2, 2)$.

L'esempio sopra cattura l'idea del *metodo di eliminazione di Gauss*, che consiste nel trasformare la matrice orlata \mathbf{H} in una *matrice a gradini* \mathbf{H}^* che permette più facilmente di studiarne la soluzione.

DEFINIZIONE 3.1 (Matrice a gradini). *Una matrice $\mathbf{M} \in \mathbb{M}_{m,n}(\mathbb{K})$ è detta a gradini se in ogni riga il primo elemento non nullo è a destra del primo elemento non nullo della riga precedente, mentre le eventuali righe di soli zeri sono le ultime*

Una matrice a gradini ha quindi la forma



dove i punti scuri indicano elementi non nulli. A seguito della procedura di Gauss, la matrice orlata \mathbf{H} diventa a gradini, e i pivot corrispondono alle posizioni indicate dai punti scuri nella figura sopra, ovvero precisamente alla posizione dei “gradini”. Essendoci al più un solo pivot per riga e colonna, il numero totale di pivot di una matrice a gradini $m \times n$ è minore o uguale a $\min\{n, m\}$.

Possono naturalmente verificarsi diversi casi che permettono o meno di trovare una soluzione al sistema. Cominciamo considerando il caso $m \leq n$, e supponiamo che *m colonne abbiano il loro pivot*. Se $m = n$ e il nostro algoritmo produce una matrice orlata a gradini in questa forma, allora il sistema è risolubile, esattamente come nell'esempio dato sopra. Essendo la matrice tipo

$$\left(\begin{array}{c|c} \text{Zeri} & \mathbf{A}^* \\ \hline \mathbf{A}^* & \mathbf{b}^* \end{array} \right)$$

l'ultima riga fornisce x_n , la penultima per x_{n-1} e così via. In altre parole, in questo caso la soluzione del sistema è *unica*. Se $m < n$, la forma ottenuta sarà del tipo

$$\left(\begin{array}{c|c} \text{Zeri} & \mathbf{A}^* \\ \hline \mathbf{A}^* & \mathbf{b}^* \end{array} \right)$$

In questo caso non abbiamo sufficienti equazioni per fissare tutte le nostre variabili x_i : un colonna senza pivot è indice della presenza di un'equazione sottodeterminata che ci lascia un arbitrio nella determinazione di una variabile. Ciò significa che, alla fine della procedura, otterremo uno *spazio* di soluzioni in cui $n - m$ variabili (per esempio, quelle corrispondenti a colonne senza pivot) fungeranno da parametri t_1, \dots, t_{n-m} rispetto ai quali la nostra soluzione sarà data. Assumendo che le $n - m$ colonne senza pivot corrispondano in realtà a parametri (e quindi a “termini noti”), il sistema residuo ha m incognite che possono essere individuate in funzione dei parametri t_1, \dots, t_{n-m} . Il fatto che questi $n - m$ parametri possano essere fissati a piacere e indipendente, implica che lo spazio delle soluzioni è $(n - m)$ -dimensionale. Per capire meglio queste osservazioni, vediamo un esempio.

Consideriamo il sistema

$$\begin{cases} x_1 + x_2 = 4 \\ 2x_1 - x_2 + x_3 = 0 \end{cases} \Leftrightarrow \left(\begin{array}{ccc|c} 1 & 1 & 0 & 4 \\ 2 & -1 & 1 & 0 \end{array} \right) \left(\begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} \right) = \left(\begin{array}{c} 4 \\ 0 \end{array} \right)$$

La matrice orlata è

$$(1.39) \quad \mathbf{H} = \left(\begin{array}{ccc|c} 1 & 1 & 0 & 4 \\ 2 & -1 & 1 & 0 \end{array} \right).$$

Il metodo di Gauss ci fornisce la matrice a gradini

$$(1.40) \quad \mathbf{H}^* = \left(\begin{array}{ccc|c} 1 & 1 & 0 & 4 \\ 0 & -3 & 1 & -12 \end{array} \right).$$

Le prime due colonne hanno un pivot, ma, per ragioni di dimensione della matrice, non possiamo assegnarne alla terza, che quindi associamo ad un parametro t . In altre parole, assumiamo $x_3 = t$, parametro arbitrario, che equivale a lavorare sulla nuova matrice a gradini

$$(1.41) \quad \left(\begin{array}{cc|c} 1 & 1 & 4 \\ 0 & -3 & -12 - t \end{array} \right).$$

Di conseguenza, $x_2 = 4 + \frac{t}{3}$ e $x_1 = 4 - x_2 = -\frac{t}{3}$, per cui le soluzioni sono tutte le terne nella forma $(x_1, x_2, x_3) = (-t/3, 4 + t/3, t)$ con t arbitrario: esse costituiscono quindi uno spazio unidimensionale.

Consideriamo il sistema

$$(1.42) \quad \begin{cases} 2x_1 + 4x_2 - 2x_3 = 2 \\ 4x_1 + 8x_2 - 3x_3 = 8 \\ -8x_1 - 16x_2 + 6x_3 = 16 \end{cases} \Leftrightarrow \left(\begin{array}{ccc|c} 2 & 4 & -2 & 2 \\ 4 & 8 & -3 & 8 \\ -8 & -16 & 6 & 16 \end{array} \right) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \\ 16 \end{pmatrix}$$

così che la matrice orlata è

$$(1.43) \quad \mathbf{H} = \left(\begin{array}{ccc|c} 2 & 4 & -2 & 2 \\ 4 & 8 & -3 & 8 \\ -8 & -16 & 6 & 16 \end{array} \right)$$

Seguiamo la procedura di Gauss. Già al primo step otteniamo

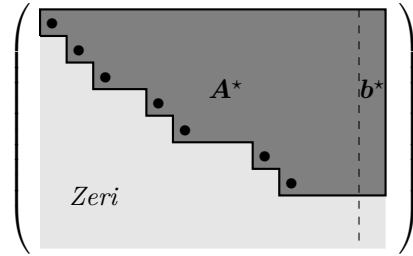
$$(1.44) \quad \left(\begin{array}{ccc|c} 2 & 4 & -2 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 1 & 4 \end{array} \right)$$

mentre al secondo

$$(1.45) \quad \mathbf{H}^* = \left(\begin{array}{ccc|c} 2 & 4 & -2 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

Come si vede, una colonna è priva di pivot. L'ultima riga inoltre corrisponde ad una equazione triviale, $0 = 0$. Il nostro sistema è *effettivamente* equivalente ad un sistema con $m = 2$ e $n = 3$, dato che l'ultima equazione non è informativa (in effetti, è uguale al doppio della seconda!). Dalla seconda equazione otteniamo $x_3 = 4$, ma questo non è sufficiente a trovare tutte le soluzioni dalla prima equazione, che è $2x_1 + 4x_2 - 2x_3 = 2 \Rightarrow x_1 = -2x_2 - 6$. Le soluzioni del sistema sono quindi tutte le terne $(x_1, x_2, x_3) = (-2t - 6, t, 4)$ per ogni $t \in \mathbb{R}$, ovvero esse costituiscono uno spazio 1-dimensionale.

L'esempio precedente corrisponde ad un caso tipo



in cui la procedura di Gauss rimuove completamente alcune equazioni (nell'esempio, il metodo rimuove l'ultima riga, lasciandoci due equazioni per tre incognite). Alcune colonne rimangono senza pivot, e le corrispondenti variabili possono essere fissate come parametri nello spazio delle soluzioni.

Finora abbiamo solo visto casi in cui i pivot compaiono in colonne di \mathbf{A} : ciò non è necessariamente sempre vero. Se il sistema è incompatibile, infatti, un pivot comparirà nell'ultima colonna, corrispondente a \mathbf{b} . Il seguente esempio corrisponde al caso in cui l'ultima riga non nulla fornisce appunto una inconsistenza.

Cerchiamo di risolvere il sistema

$$(1.46) \quad \begin{cases} 2x_1 + 4x_2 - 2x_3 = 2 \\ 4x_1 + 8x_2 - 3x_3 = 8 \\ -2x_1 - 4x_2 + 7x_3 = 10 \end{cases} \Leftrightarrow \begin{pmatrix} 2 & 4 & -2 \\ 4 & 8 & -3 \\ -2 & -4 & 7 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \\ 10 \end{pmatrix}$$

così che la matrice orlata è

$$(1.47) \quad \mathbf{H} = \left(\begin{array}{ccc|c} 2 & 4 & -2 & 2 \\ 4 & 8 & -3 & 8 \\ -2 & -4 & 7 & 10 \end{array} \right)$$

Seguiamo la procedura di Gauss. Al primo step otteniamo

$$(1.48) \quad \left(\begin{array}{ccc|c} 2 & 4 & -2 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 5 & 12 \end{array} \right)$$

mentre al secondo otteniamo la nostra matrice a gradini

$$(1.49) \quad \mathbf{H}^* = \left(\begin{array}{ccc|c} 2 & 4 & -2 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & -8 \end{array} \right)$$

Come si vede, un pivot è comparso nell'ultima colonna: la corrispondente riga corrisponde ad una equazione falsa, ovvero $0 = -8$. Questo significa che il sistema è *incompatibile*.

Se la matrice a gradini che compare alla fine della procedura è nella forma

$$\left(\begin{array}{ccc|c} \bullet & & & \\ & A^* & & b^* \\ \hline Zeri & & & \end{array} \right)$$

ovvero un pivot è nella colonna corrispondente a \mathbf{b} , questo significa che la riga dove questo pivot è collocato corrisponde all'equazione falsa $0 = b$ con $b \neq 0$, il che è un segnale del fatto che il sistema è *incompatibile*.

Il caso $m > n$ si studia in maniera simile: alla fine della procedura si ottiene una matrice orlata nella forma

$$(1.50) \quad \left(\begin{array}{ccc|c} \bullet & & & \\ & A^* & & b^* \\ \hline Zeri & & & \end{array} \right) \quad \text{oppure} \quad \left(\begin{array}{ccc|c} \bullet & & & \\ & A^* & & b^* \\ \hline Zeri & & & \end{array} \right)$$

Dato che possono esserci al più $n + 1$ pivot, appena terminata la procedura tutte le righe dalla $(n + 2)$ -esima in poi saranno messe a zero e quindi non saranno rilevanti. Se esiste un pivot nella colonna corrispondente a \mathbf{b} (Eq. (1.50), sinistra) allora la procedura ha prodotto una condizione falsa, che vuol dire che il nostro sistema è incompatibile e non esistono soluzioni. Se ciò non accade (Eq. (1.50), destra), vuol dire che la $(n + 1)$ -esima riga è vuota e possiamo trascurarla: siamo quindi tornati al caso di un sistema con un numero di righe effettive uguale o minore al numero di incognite, cioè il caso $m \leq n$.

3.2. Rango e compatibilità. Negli esempi sopra risulta evidente che, perché il sistema sia compatibile, la sottomatrice corrispondente a \mathbf{b} non deve avere pivot alla fine della procedura. Notare che questo non potrà mai accadere se il sistema in analisi è omogeneo (in questo caso infatti l'ultima colonna è data da soli zeri per tutta la durata della procedura di Gauss) ed infatti un sistema omogeneo è sempre compatibile. Se il sistema è compatibile, le colonne senza pivot ci segnalano quali variabili vanno fissate a parametri arbitrari. In ultima analisi, il numero di pivot sembra essere un elemento cruciale, e ciò suggerisce di introdurre una nuova quantità relativa alla matrice a gradini ottenuta alla fine della procedura corrispondente.

DEFINIZIONE 3.2 (Rango). Una matrice \mathbf{M} ha rango $\text{rank}(\mathbf{M}) = r$ se il numero di pivot nella matrice a gradini \mathbf{M}^* , ottenuta per mezzo della procedura di Gauss, è r .

Essendo uguale al numero dei pivot, naturalmente il rango di una matrice è minore o uguale alla più piccola delle sue dimensioni.

Nel primo esempio, partendo dalla matrice orlata in Eq. (1.35), siamo arrivati alla matrice a gradini in Eq. (1.37). Osservando quest'ultima, il rango della matrice \mathbf{H} è uguale a 3. Se ci focalizziamo sulla sottomatrice corrispondente ad \mathbf{A} , abbiamo trasformato

$$\mathbf{A} = \begin{pmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{pmatrix} \mapsto \mathbf{A}^* = \begin{pmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 4 \end{pmatrix}.$$

La matrice a gradini \mathbf{A}^* contiene anch'essa tre pivot, quindi il rango di \mathbf{A} qui è 3. Viceversa, nell'esempio in Eq. (1.47), la matrice orlata produce una matrice a gradini con tre pivot, e quindi \mathbf{H} ha rango 3, mentre abbiamo ottenuto

$$\mathbf{A} = \begin{pmatrix} 2 & 4 & -2 \\ 4 & 8 & -3 \\ -2 & -4 & 7 \end{pmatrix} \mapsto \mathbf{A}^* = \begin{pmatrix} 2 & 4 & -2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

La matrice a gradini contiene due pivot, quindi il rango di \mathbf{A} qui è 2.

L'insieme delle osservazioni raccolte sopra può essere riassunto nel seguente teorema, fondamentale per caratterizzare la risolubilità dei sistemi di equazioni lineari.

TEOREMA 3.1 (Kronecker–Rouché–Capelli). Un sistema lineare di m equazioni in n incognite $\mathbf{Ax} = \mathbf{b}$ è compatibile se e solo se la matrice orlata ha lo stesso rango della matrice \mathbf{A} . In tal caso, lo spazio delle soluzioni ha dimensione $n - \text{rank}(\mathbf{A})$.

3.2.1. Il teorema del rango. Il teorema di Kronecker–Rouché–Capelli ha in effetti una interpretazione profonda. Il rango di una matrice ha infatti anche una diversa, interessante e importante significato: esso è uguale alla dimensione dello spazio generato dalle sue righe o sue colonne, come stabilito dal seguente teorema, che diamo senza dimostrazione.

TEOREMA 3.2 (Teorema del rango). Sia $\mathbf{A} \in \mathbb{M}_{m,n}(\mathbb{K})$ una matrice di rango r . Allora, denotando $\mathcal{R} := \{\mathbf{a}_\mu\}_{\mu=1}^m$ l'insieme delle sue righe e con $\mathcal{C} := \{\mathbf{a}^\nu\}_{\nu=1}^n$ l'insieme delle sue colonne

$$(1.51) \quad \text{rank}(\mathbf{A}) = \dim \text{Span}[\mathcal{R}] = \dim \text{Span}[\mathcal{C}].$$

Se consideriamo quindi l'insieme dei vettori $\mathbf{y} = \mathbf{Ax} \in \mathbb{K}^m$, possiamo osservare che questo si scrive come

$$(1.52) \quad \mathbf{y} = \sum_{\nu=1}^n x_\nu \mathbf{a}^\nu = \mathbf{b},$$

Il rango non è altro che la dimensione dello spazio di questi vettori. Se cioè introduciamo l'immagine di \mathbb{K}^n via \mathbf{A} ,

$$(1.53) \quad \text{Im}(\mathbf{A}) = \{\mathbf{y} \in \mathbb{K}^n : \exists \mathbf{x} \in \mathbb{K}^n \text{ tale che } \mathbf{Ax} = \mathbf{y}\},$$

allora il teorema del rango dice che

$$(1.54) \quad \text{rank}(\mathbf{A}) = \dim \text{Im}(\mathbf{A}).$$

Questa ulteriore informazione chiarifica ancor meglio il fatto che il rango di una matrice sia legato alla risolubilità di un sistema lineare. L'equazione $\mathbf{Ax} = \mathbf{b}$ può interpretarsi come la domanda $\mathbf{b} \in \text{Span}[\mathcal{C}]$: in altre parole, perché il sistema abbia soluzione \mathbf{b} non deve vivere al di fuori dello spazio generato dai vettori colonna della matrice dei coefficienti.

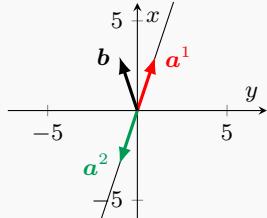
Consideriamo il sistema

$$(1.55) \quad \begin{cases} x_1 - x_2 = -1 \\ 3x_1 - 3x_2 = 3. \end{cases}$$

Usando il metodo di Gauss si trova

$$(1.56) \quad \left(\begin{array}{cc|c} 1 & -1 & -1 \\ 3 & -3 & 3 \end{array} \right) \rightarrow \left(\begin{array}{cc|c} 1 & -1 & -1 \\ 0 & 0 & 6 \end{array} \right)$$

che il teorema di Kronecker–Rouché–Capelli implica non avere soluzione (la matrice \mathbf{A} ha rango 1, mentre la matrice orlata ha rango 2). Il fatto che \mathbf{A} abbia rango 1 significa che i vettori colonna di \mathbf{A} generano uno spazio unidimensionale: in effetti, usando una rappresentazione cartesiana, si vede che i vettori colonna $\mathbf{a}^1 = (1, 3)^\top$ e $\mathbf{a}^2 = (-1, -3)^\top$ vivono sulla stessa retta



che corrisponde a $\text{Span}[\{\mathbf{a}^1, \mathbf{a}^2\}]$, mentre $\mathbf{b} = (-1, 3)^\top$ vive al di fuori di questa retta, come si vede dalla figura.

Consideriamo il seguente sistema:

$$(1.57) \quad \begin{cases} x_1 - 2x_3 = -2 \\ 2x_1 + 2x_2 - 3x_3 = -3 \\ 3x_1 + 4x_2 - 4x_3 = 4 \end{cases}$$

Applicando la procedura di Gauss otteniamo

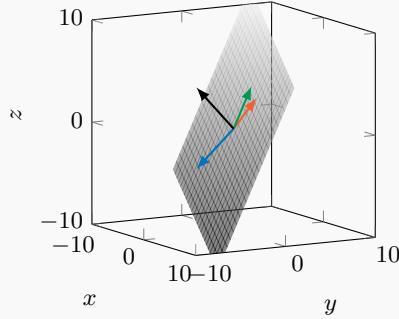
$$(1.58) \quad \left(\begin{array}{ccc|c} 1 & 0 & -2 & -2 \\ 2 & 2 & -3 & -3 \\ 3 & 4 & -4 & 4 \end{array} \right) \rightarrow \left(\begin{array}{ccc|c} 2 & 0 & -2 & -2 \\ 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & 8 \end{array} \right)$$

Il rango di \mathbf{A} e della matrice orlata sono diversi: la matrice orlata ha rango 3, mentre \mathbf{A} ha rango 2: il sistema è quindi incompatibile. Tentiamo di interpretare geometricamente questo risultato. Il

fatto che il rango di \mathbf{A} sia 2 vuol dire che i suoi vettori colonna vivono in uno spazio *bidimensionale*. In effetti, se rappresentiamo graficamente in \mathbb{R}^3 i vettori

$$(1.59) \quad \mathbf{a}^1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \mathbf{a}^2 = \begin{pmatrix} 0 \\ 2 \\ 4 \end{pmatrix}, \quad \mathbf{a}^3 = \begin{pmatrix} -2 \\ -3 \\ -4 \end{pmatrix},$$

si osserva che i vettori colonna della matrice \mathbf{A} vivono tutti sullo stesso *piano*, che rappresenta $\text{Span}\{\{\mathbf{a}^\nu\}_{\nu=1}^3\}$:



In effetti essi sono linearmente dipendenti (infatti, $4\mathbf{a}^1 - \mathbf{a}^2 + 2\mathbf{a}^3 = \mathbf{0}$). Viceversa, \mathbf{b} vive al di fuori di questo piano: di conseguenza non è possibile rappresentare \mathbf{b} usando i vettori colonna di \mathbf{A} e quindi trovare una soluzione al sistema $\mathbf{Ax} = \mathbf{b}$.

3.2.2. Il metodo dei minori. Il rango di una matrice è quindi, come il determinante, un numero ricco di informazioni ed è quindi importante saperlo calcolare. Una maniera per ottenerlo è applicare il metodo di Gauss: ridurre la matrice nella forma a gradini e contare i pivot. Un metodo differente utilizza l'informazione che ci ha fornito il teorema del rango: il rango di una matrice \mathbf{A} è la dimensione dello spazio generato da righe o colonne di \mathbf{A} . Questo vuol dire che se abbiamo una matrice $\mathbf{A} \in \mathbb{M}_{n,m}(\mathbb{K})$ e selezioniamo un suo minore \mathbf{M} , ovvero una sottomatrice quadrata di dimensione $k \times k$, $k \leq \min(n, m)$, se $\det(\mathbf{M}) \neq 0$ allora il rango di \mathbf{A} è almeno k : questo perché il fatto che $\det(\mathbf{M}) \neq 0$ significa che le righe o colonne di \mathbf{M} corrispondono a vettori linearmente indipendenti. Il rango di \mathbf{A} è quindi uguale alla dimensione del minore più grande in \mathbf{A} con determinante non nullo.

Supponiamo di voler calcolare il rango di

$$\mathbf{A} = \begin{pmatrix} 3 & 3 & 4 & 2 \\ 5 & 4 & 6 & 3 \\ 8 & 7 & 10 & 5 \end{pmatrix}$$

Ci sono quattro possibili minori di ordine tre, i cui rispettivi determinanti sono

$$\begin{vmatrix} 3 & 3 & 4 \\ 5 & 4 & 6 \\ 8 & 7 & 10 \end{vmatrix} = 0, \quad \begin{vmatrix} 3 & 3 & 2 \\ 5 & 4 & 3 \\ 8 & 7 & 5 \end{vmatrix} = 0, \quad \begin{vmatrix} 3 & 4 & 2 \\ 5 & 6 & 3 \\ 8 & 10 & 5 \end{vmatrix} = 0, \quad \begin{vmatrix} 3 & 4 & 2 \\ 4 & 6 & 3 \\ 7 & 10 & 5 \end{vmatrix} = 0.$$

Questo vuol dire che il rango di \mathbf{A} deve essere minore di 3. Consideriamo ora un minore di ordine 2, per esempio

$$\begin{vmatrix} 3 & 3 \\ 5 & 4 \end{vmatrix} = -3 \neq 0.$$

Avendo trovato un minore di ordine 2 con determinante non nullo, possiamo dire che $\text{rank}(\mathbf{A}) = 2$.

3.3. Regola di Cramer. Il caso di un sistema lineare $\mathbf{A}\mathbf{x} = \mathbf{b}$ con $m = n$, ovvero dove $\mathbf{A} \in \mathbb{M}_n(\mathbb{K})$, è particolarmente interessante dato che, se $\det(\mathbf{A}) \neq 0$, allora

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b},$$

che fornisce immediatamente la soluzione. Esplicitando, questo vuol dire che

$$x_\mu = \sum_{\nu=1}^n [\mathbf{A}^{-1}]_{\mu\nu} b_\nu.$$

Ma dalla forma per l'inversa data dall'Eq. (1.29) possiamo scrivere

$$x_\mu = \frac{1}{\det(\mathbf{A})} \sum_{\nu=1}^n c_{\nu\mu} b_\nu,$$

dove $\mathbf{C} = (c_{\mu\nu})_{\mu\nu}$ è la matrice dei cofattori di \mathbf{A} . Gabriel Cramer ha osservato l'interessante fatto che l'espressione nella somma ha la forma in Eq. (1.23) come se avessimo sostituito alla colonna μ -esima di \mathbf{A} il vettore \mathbf{b} : in altre parole, se indichiamo con \mathbf{A}_μ la matrice in cui la μ -esima colonna è sostituita da \mathbf{b} , allora

$$(1.60) \quad x_\mu = \frac{\det(\mathbf{A}_\mu)}{\det(\mathbf{A})}.$$

Vediamo un esempio. Supponiamo di voler studiare il sistema

$$\begin{cases} 2x_1 + x_2 = 3 \\ x_1 - x_2 + x_3 = 1 \\ 4x_1 - 2x_3 = 0. \end{cases} \Leftrightarrow \underbrace{\begin{pmatrix} 2 & 1 & 0 \\ 1 & -1 & 1 \\ 4 & 0 & -2 \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}_{\mathbf{x}} = \underbrace{\begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}}_{\mathbf{b}}$$

Dato che

$$\det(\mathbf{A}) = -2 \neq 0$$

possiamo risolvere il problema usando il metodo di Cramer. Occorre calcolare tre determinanti, ovvero

$$\det(\mathbf{A}_1) = \det \begin{pmatrix} 3 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 0 & -2 \end{pmatrix} = -8, \quad \det(\mathbf{A}_2) = \det \begin{pmatrix} 2 & 3 & 0 \\ 1 & 1 & 1 \\ 4 & 0 & -2 \end{pmatrix} = 10, \quad \det(\mathbf{A}_3) = \det \begin{pmatrix} 2 & 1 & 3 \\ 1 & -1 & 1 \\ 4 & 0 & 0 \end{pmatrix} = 16.$$

La soluzione è quindi

$$x_1 = \frac{\det(\mathbf{A}_1)}{\det(\mathbf{A})} = 4, \quad x_2 = \frac{\det(\mathbf{A}_2)}{\det(\mathbf{A})} = -5, \quad x_3 = \frac{\det(\mathbf{A}_3)}{\det(\mathbf{A})} = -8.$$

3.4. Metodo di Gauss–Jordan per l'inversa. Come abbiamo appena detto, dato il problema $\mathbf{A}\mathbf{x} = \mathbf{b}$, se $\mathbf{A} \in \mathbb{GL}_n(\mathbb{K})$ abbiamo che $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. D'altra parte il metodo di Gauss descritto sopra fornisce la stessa soluzione: in altre parole, il metodo “inverte”, quando possibile, la matrice \mathbf{A} , anche se in una maniera non esplicita. Se $\det(\mathbf{A}) \neq 0$ allora $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ deve essere la stessa soluzione trovata dall'algoritmo di Gauss. In qualche modo, quindi, il metodo produce esattamente il risultato di questa operazione applicando l'inversa di \mathbf{A} a \mathbf{b} , quando questa esiste. Forse non sorprende quindi che il metodo di Gauss possa essere utilizzato per ottenere l'inversa di una matrice. Come sappiamo, la matrice inversa \mathbf{A}^{-1} è tale per cui $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n$. Se denotiamo con \mathbf{x}^ν la ν -esima colonna di \mathbf{A}^{-1}

$$(1.61) \quad \mathbf{A}\mathbf{x}^\nu = \mathbf{e}_\nu, \quad \text{dove } \mathbf{e}_\nu = (\delta_{\mu\nu})_\mu \quad \text{per } i = 1, \dots, n.$$

In sostanza trovare l'inversa equivale a risolvere contemporaneamente n sistemi di equazioni lineari (uno per colonna), problema per il quale possiamo applicare il metodo di Gauss. Per meglio esemplificare questo fatto, ricorriamo ad un esempio.

Consideriamo la matrice quadrata

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

Per trovare l'inversa

$$\mathbf{A}^{-1} = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{22} & x_{33} \end{pmatrix}$$

occorre risolvere le equazioni

$$\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{21} \\ x_{31} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} x_{12} \\ x_{22} \\ x_{32} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} x_{13} \\ x_{23} \\ x_{33} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Possiamo immaginare di procedere parallelamente su tutti i sistemi usando una matrice orlata

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

La prima colonna a destra della linea verticale corrisponde al primo sistema, la seconda al secondo sistema, la terza al terzo sistema. Questa matrice orlata ha la forma

$$(\mathbf{A} \mid \mathbf{I}_3)$$

Usando la procedura di Gauss per mettere la matrice in forma a gradini si ottiene

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

Se, eseguendo altre operazioni di riga, riuscissimo a mettere il sistema nella forma

$$(\mathbf{I}_3 \mid \mathbf{B}^*)$$

avremmo che $\mathbf{B}^* = \mathbf{A}^{-1}$. La matrice orlata così ottenuta corrisponderebbe infatti all'insieme di equazioni del tipo $x_{\mu\nu} = b_{\mu\nu}^*$ fornendoci la soluzione che cerchiamo. Dobbiamo quindi lavorare per rimuovere gli zeri *sopra* la diagonale procedendo come nel metodo di Gauss ma provando ad azzerare gli elementi *sopra* i pivot. Nell'esempio,

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

Dividendo opportunamente ogni riga otteniamo

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

per cui

$$\mathbf{A}^{-1} = \begin{pmatrix} 3/4 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & 3/4 \end{pmatrix}.$$

Il metodo esemplificato è detto *di Gauss-Jordan*. La procedura fallisce se l'inversa cercata non esiste: questo è il caso se, *dopo* aver ottenuto la forma a gradini \mathbf{A}^* , qualche colonna difetta di pivot. Ma il fatto che nel caso di una matrice \mathbf{A} invertibile ogni colonna abbia il suo pivot equivale a dire che il rango della matrice è n , per cui vale la seguente

PROPOSIZIONE 3.3. *Una matrice quadrata $\mathbf{A} \in \mathbb{M}_n(\mathbb{K})$ è invertibile se e solo se $\text{rank}(\mathbf{A}) = n$.*

COROLLARIO 3.4. *Se $\mathbf{A} \in \mathbb{M}_n(\mathbb{K})$ allora*

$$\mathbf{A} \text{ invertibile} \Leftrightarrow \det(\mathbf{A}) \neq 0 \Leftrightarrow \text{rank}(\mathbf{A}) = n$$

4. Endomorfismi e proprietà spettrali

In questa sezione e nella successiva torniamo ad analizzare gli spazi vettoriali in generale. Le conoscenze acquisite sugli spazi di matrici e sui sistemi di equazioni lineari, tuttavia, saranno cruciali per lo studio delle proprietà che verranno discusse.

4.1. Applicazioni lineari. Introdurre funzioni tra spazi vettoriali è naturalmente possibile, e l'operazione di trasposizione che abbiamo già discusso ne è un esempio. Esiste però una famiglia di applicazioni tra spazi vettoriali che ha una particolare importanza.

DEFINIZIONE 4.1 (Applicazioni lineari). *Siano \mathbb{V} e \mathbb{W} due \mathbb{K} -spazi vettoriali. Una applicazione $\mathbf{F}: \mathbb{V} \rightarrow \mathbb{W}$ è lineare se, per ogni $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{V}$ e ogni $c_1, c_2 \in \mathbb{K}$, vale*

$$(1.62) \quad \mathbf{F}(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1\mathbf{F}(\mathbf{v}_1) + c_2\mathbf{F}(\mathbf{v}_2).$$

Si dice anche che \mathbf{F} è un omomorfismo tra spazi vettoriali e scrive $\mathbf{F} \in \text{Hom}(\mathbb{V}, \mathbb{W})$. Se $\mathbb{V} = \mathbb{W}$ l'applicazione è detto endomorfismo.

Consideriamo per esempio $\mathbb{V} = \mathbb{K}^n$ e $\mathbb{W} = \mathbb{K}^m$, entrambi su \mathbb{K} . Data una matrice $\mathbf{A} \in \mathbb{M}_{m,n}(\mathbb{K})$ e un vettore $\mathbf{v} \in \mathbb{K}^n$, l'applicazione $\mathbf{F}(\mathbf{v}) = \mathbf{A}\mathbf{v}$ di \mathbf{A} a \mathbf{v} restituisce un nuovo elemento di \mathbb{K}^m , ovvero $\mathbf{v} \in \mathbb{K}^n \mapsto \mathbf{A}\mathbf{v} \in \mathbb{K}^m$. Inoltre, l'applicazione di \mathbf{F} è una operazione lineare dato che la sua azione su una combinazione lineare di vettori equivale alla combinazione lineare della sua azione sui singoli vettori:

(1.63)

$$\mathbf{F}(c_1b\mathbf{v}_1 + c_2\mathbf{v}_2) = \mathbf{A}(c_1b\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1(\mathbf{A}\mathbf{v}_1) + c_2(\mathbf{A}\mathbf{v}_2) = c_1\mathbf{F}(\mathbf{v}_1) + c_2\mathbf{F}(\mathbf{v}_2), \quad c_1, c_2 \in \mathbb{K}, \quad \mathbf{v}_1, \mathbf{v}_2 \in \mathbb{K}^n.$$

Le matrici sono uno strumento importante per studiare le applicazioni lineari in generale e, di fatto, lo studio delle applicazioni lineari tra spazi finito dimensionali è lo studio delle matrici. Per capire la ragione di questa affermazione, supponiamo di avere due \mathbb{K} -spazi vettoriali di dimensione finita \mathbb{V} e \mathbb{W} con due basi rispettive $\mathcal{B}_v = \{\mathbf{v}_i\}_{i=1}^n$ e $\mathcal{B}_w = \{\mathbf{w}_i\}_{i=1}^m$. Possiamo completamente descrivere una applicazione lineare $\mathbf{F}: \mathbb{V} \rightarrow \mathbb{W}$ se conosciamo tutti gli elementi f_{ij} che compaiono espandendo nella base \mathcal{B}_w l'azione di \mathbf{F} sugli elementi della base \mathcal{B}_v , ovvero

$$(1.64) \quad \mathbf{F}(\mathbf{v}_i) = \sum_{j=1}^m \mathbf{w}_j f_{ji} \quad \forall \mathbf{v}_i \in \mathcal{B}_v.$$

In questo modo possiamo costruire la *matrice $m \times n$ associata a \mathbf{F}* rispetto alle due basi \mathcal{B}_v e \mathcal{B}_w

$$(1.65) \quad \mathbf{F}_{\mathbf{wv}} = \begin{pmatrix} f_{11} & f_{12} & \dots & f_{1n} \\ f_{21} & f_{22} & \dots & f_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{m1} & f_{m2} & \dots & f_{mn} \end{pmatrix}.$$

Notare che nella nostra convenzione calcolare l'azione di \mathbf{F} sul vettore \mathbf{v}_i fornisce i valori della matrice sulla i -esima colonna. Conoscere $\mathbf{F}_{\mathbf{wv}}$ è sufficientemente informativo per descrivere l'azione di \mathbf{F} . Supponiamo di partire da un generico vettore $\mathbf{x} = \sum_i x_i \mathbf{v}_i \in \mathbb{V}$ e chiamiamo $\mathbf{y} = \mathbf{F}(\mathbf{x}) \in \mathbb{W}$. Abbiamo che

$$(1.66) \quad \mathbf{F}(\mathbf{x}) = \mathbf{F} \left(\sum_{i=1}^n x_i \mathbf{v}_i \right) = \sum_{i=1}^n x_i \mathbf{F}(\mathbf{v}_i) = \sum_{j=1}^m \mathbf{w}_j \sum_{i=1}^n x_i f_{ji} \equiv \sum_{j=1}^m \mathbf{w}_j y_j,$$

dove y_j è il coefficiente di \mathbf{y} relativo all'elemento \mathbf{w}_j della base dello spazio di arrivo. L'espressione che abbiamo ottenuto si può scrivere in una maniera più compatta come

$$(1.67) \quad \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} f_{11} & f_{12} & \dots & f_{1n} \\ f_{21} & f_{22} & \dots & f_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{m1} & f_{m2} & \dots & f_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Notare che il vettore $\mathbf{x} = (x_1, \dots, x_n)^\top$ fa riferimento alle componenti di \mathbf{x} secondo la base \mathcal{B}_v , mentre il vettore $\mathbf{y} = (y_1, \dots, y_m)^\top$ fa riferimento alle componenti di $\mathbf{y} = \mathbf{F}(\mathbf{x})$ secondo la base \mathcal{B}_w : la relazione ottenuta afferma che $\mathbf{y} = \mathbf{F}_{vw}\mathbf{x}$. In altre parole, date due basi \mathcal{B}_v e \mathcal{B}_w di due spazi finiti dimensionali V e W , *ogni applicazione lineare $F: V \rightarrow W$ corrisponde univocamente ad una matrice $\mathbf{F}_{vw} \in \mathbb{M}_{m,n}(\mathbb{K})$* e l'azione di \mathbf{F} su un vettore \mathbf{x} può essere espressa come un *prodotto matriciale* di \mathbf{F}_{vw} per il vettore delle coordinate di \mathbf{x} secondo la base \mathcal{B}_v .

Questa corrispondenza tra applicazioni lineari e matrici vale anche nel caso di una *composizione* di applicazioni lineari. Per esempio, supponiamo di avere tre spazi vettoriale V , W e U di dimensione n , m e k rispettivamente, e due applicazioni lineari $\mathbf{F}: V \rightarrow W$ e $\mathbf{G}: W \rightarrow U$. Come rappresentare $\mathbf{H} := \mathbf{G} \circ \mathbf{F}: V \rightarrow U$? Se abbiamo tre basi \mathcal{B}_v , \mathcal{B}_w e \mathcal{B}_u per i tre spazi rispettivamente, è possibile mostrare che

$$\mathbf{H}_{uv} = \mathbf{G}_{uw}\mathbf{F}_{vv}.$$

Consideriamo un esempio di applicazione lineare sullo spazio vettoriale $\mathbb{R}_n[x]$ dei polinomi di grado n a coefficienti reali. Questo significa che un vettore $\mathbf{p} \in \mathbb{R}_n[x]$ è un polinomio, nella forma

$$\mathbf{p} \equiv p(x) = p_n x^n + p_{n-1} x^{n-1} + \dots + p_1 x + p_0.$$

Questo è effettivamente uno spazio vettoriale su \mathbb{R} . Infatti, dati $c_1, c_2 \in \mathbb{R}$ e due polinomi $\mathbf{p}, \mathbf{q} \in \mathbb{R}_n[x]$, la combinazione $c_1\mathbf{p} + c_2\mathbf{q}$ è ancora in $\mathbb{R}_n[x]$. Questo spazio ha dimensione $n+1$: possiamo individuare una base $\mathcal{B}_e = \{\mathbf{e}_0, \mathbf{e}_1, \dots, \mathbf{e}_n\}$ tale per cui $\mathbf{e}_i = x^i$ (dunque $\mathbf{e}_0 = 1$, $\mathbf{e}_1 = x$, $\mathbf{e}_2 = x^2$ e così via) e il generico polinomio \mathbf{p} si scrive $\mathbf{p} = \sum_{i=0}^n p_i \mathbf{e}_i$.

Consideriamo ora una applicazione $\mathbf{F}: \mathbb{R}_3[x] \rightarrow \mathbb{R}_2[x]$ tale per cui, dato un polinomio \mathbf{p} ,

$$\mathbf{p} = p_3 x^3 + p_2 x^2 + p_1 x^1 + p_0 \mapsto \mathbf{q} = \mathbf{F}(\mathbf{p}) = (p_3 + p_2)x^2 + (p_2 + p_1)x + p_1 + p_0 \in \mathbb{R}_2[x].$$

Questa applicazione, anzitutto, è *lineare*. Se infatti consideriamo due polinomi $\mathbf{p}, \mathbf{p}' \in \mathbb{R}_3[x]$, allora, per due scalari $c, c' \in \mathbb{R}$

$$\begin{aligned} F(c\mathbf{p} + c'\mathbf{p}') &= (cp_3 + c'p'_3 + cp_2 + c'p'_2)x^2 + (cp_2 + c'p'_2 + cp_1 + cp'_1)x + cp_1 + c'p'_1 + cp_0 + c'p'_0 \\ &= c((p_3 + p_2)x^2 + (p_2 + p_1)x + p_1 + p_0) + c'((p'_3 + p'_2)x^2 + (p'_2 + p'_1)x + p'_1 + p'_0) \\ &= cF(\mathbf{p}) + c'F(\mathbf{p}'). \end{aligned}$$

Usiamo la base $\mathcal{B}_3 = \{\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ per $\mathbb{R}_3[x]$. Secondo questa base, un vettore $\mathbf{p} \in \mathbb{R}_3[x]$ è rappresentato dalle coordinate

$$(1.68) \quad \mathbf{p} = p_3 \mathbf{e}_3 + p_2 \mathbf{e}_2 + p_1 \mathbf{e}_1 + p_0 \mathbf{e}_0 \mapsto \mathbf{p} = \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix}.$$

Usiamo la base $\mathcal{B}_2 = \{\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2\}$ per $\mathbb{R}_2[x]$, secondo cui, naturalmente, un vettore $\mathbf{q} \in \mathbb{R}_2[x]$ è rappresentato da tre coordinate,

$$\mathbf{q} = q_2 \mathbf{e}_2 + q_1 \mathbf{e}_1 + q_0 \mathbf{e}_0 \mapsto \mathbf{q} = \begin{pmatrix} q_0 \\ q_1 \\ q_2 \end{pmatrix}.$$

Usando queste basi, l'applicazione \mathbf{F} esegue la trasformazione

$$\mathbf{p} = \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} \xrightarrow{\mathbf{F}} \mathbf{q} = \begin{pmatrix} p_0 + p_1 \\ p_1 + p_2 \\ p_2 + p_3 \end{pmatrix}$$

Come rappresentare l'applicazione \mathbf{F} secondo la base introdotta? Occorre in effetti calcolare l'azione di \mathbf{F} secondo tutti i vettori della base di partenza, ovvero

$$\begin{aligned}\mathbf{F}(\mathbf{e}_0) &= \mathbf{e}_0 \\ \mathbf{F}(\mathbf{e}_1) &= \mathbf{e}_1 + \mathbf{e}_0 \\ \mathbf{F}(\mathbf{e}_2) &= \mathbf{e}_2 + \mathbf{e}_1 \\ \mathbf{F}(\mathbf{e}_3) &= \mathbf{e}_2\end{aligned}$$

quindi la sua rappresentazione matriciale è

$$\mathbf{F} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

Infatti, se \mathbf{p} ha la rappresentazione in coordinate in Eq. (1.68), allora

$$\mathbf{q} = \mathbf{F}\mathbf{p} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} p_0 + p_1 \\ p_1 + p_2 \\ p_2 + p_3 \end{pmatrix}.$$

4.2. Teorema del rango. In analogia con quanto fatto per le matrici, il *kernel* (o nucleo) di una mappa lineare $\mathbf{F}: \mathbb{V} \rightarrow \mathbb{W}$ è dato dall'insieme dei vettori $\mathbf{v} \in \mathbb{V}$ tale per cui $\mathbf{F}(\mathbf{v}) = \mathbf{0}$,

$$(1.69) \quad \text{Ker}(\mathbf{F}) := \{\mathbf{x} \in \mathbb{V} : \mathbf{F}(\mathbf{x}) = \mathbf{0}\} \subseteq \mathbb{W}.$$

Come nel caso delle matrici, questo è un *sottospazio vettoriale*: se $\mathbf{x}_1, \mathbf{x}_2 \in \text{Ker}(\mathbf{F})$, allora, per ogni coppia di scalari $c_1, c_2 \in \mathbb{K}$, anche $c_1\mathbf{x}_1 + c_2\mathbf{x}_2 \in \text{Ker}(\mathbf{F})$. L'*immagine* di \mathbf{F} , invece, è un sottoinsieme di \mathbb{W} dato da

$$(1.70) \quad \text{Im}(\mathbf{F}) := \{\mathbf{w} \in \mathbb{W} : \exists \mathbf{v} \in \mathbb{V} \text{ tale che } \mathbf{F}(\mathbf{v}) = \mathbf{w}\} \subseteq \mathbb{W}.$$

Anche questo è un sottospazio vettoriale: se $\mathbf{w}_1, \mathbf{w}_2 \in \text{Im}(\mathbf{F})$, vi appartiene anche una qualunque loro combinazione lineare. Vale il seguente teorema, che in un certo senso esprime il fatto che la dimensione dello spazio di partenza venga ripartita tra kernel e immagine.

TEOREMA 4.1 (Teorema del rango). *Sia $\mathbf{F}: \mathbb{V} \rightarrow \mathbb{W}$ applicazione lineare tra lo spazio \mathbb{V} , di dimensione n , e lo spazio \mathbb{W} , di dimensione m . Allora*

$$(1.71) \quad \dim \text{Ker}(\mathbf{F}) + \dim \text{Im}(\mathbf{F}) = \dim \mathbb{V}.$$

DIMOSTRAZIONE. Sia $r = \dim \text{Ker}(\mathbf{F}) \leq \dim \mathbb{V} = n$ la dimensione del kernel di \mathbf{F} , e sia $\mathcal{B}_k = \{\mathbf{k}_i\}_{i=1}^r$ una sua base. È possibile allora completare questa base aggiungendo $n - r$ vettori per ottenere una base per \mathbb{V} , $\mathcal{B} = \{\mathbf{k}_i\}_{i=1}^n$, in modo che i primi r vettori siano i vettori di \mathcal{B}_k . Chiamando ora $\mathbf{u}_i := \mathbf{F}(\mathbf{k}_{r+i})$, con $i = 1, \dots, n-r$, mostriamo che $\mathcal{B}_u = \{\mathbf{u}_i\}_{i=1}^{n-r}$ è una base per $\text{Im}(\mathbf{F})$ provando che i suoi elementi sono linearmente indipendenti. Supponiamo che ciò non sia vero: in questo caso, esisterebbe un set di scalari c_1, \dots, c_{n-r} tali per cui $\sum_{i=1}^{n-r} c_i \mathbf{u}_i = \mathbf{0}$. Ma ciò non può essere: se $\sum_{i=1}^{n-r} c_i \mathbf{u}_i = \mathbf{F}(\sum_{i=1}^{n-r} c_i \mathbf{k}_{r+i}) = \mathbf{0}$, allora $\sum_{i=1}^{n-r} c_i \mathbf{k}_{r+i} \in \text{Ker}(\mathbf{F})$, ovvero possiamo scrivere $\sum_{i=1}^{n-r} c_i \mathbf{k}_{r+i} = \sum_{i=1}^r \alpha_i \mathbf{k}_{r+i}$ per certi coefficienti $\alpha_1, \dots, \alpha_r$. Questa è una combinazione lineare di vettori di una base, quindi può essere nulla se e solo se tutti i coefficienti sono nulli. L'insieme \mathcal{B}_u è quindi una base per $\text{Im}(\mathbf{F})$ ed è costituita da $n - r$ elementi. \square

Questo teorema vale in particolare per una matrice $\mathbf{A} \in \mathbb{M}_{m,n}(\mathbb{K})$:

$$(1.72) \quad \dim \text{Ker}(\mathbf{A}) + \text{rank}(\mathbf{A}) = n.$$

Se $\dim \text{Ker}(\mathbf{A}) = 0$, quindi, $\text{rank}(\mathbf{A}) = n$: nel caso di una matrice quadrata, quindi, la condizione $\dim \text{Ker}(\mathbf{A}) = 0$ equivale all'invertibilità.

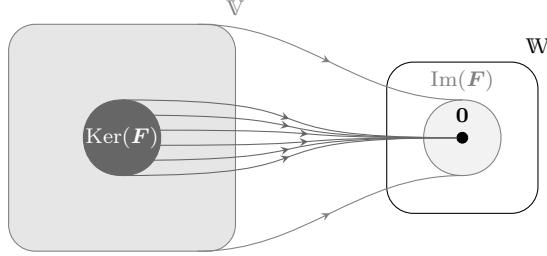


FIGURA 1. Rappresentazione dell’azione di \mathbf{F} su uno spazio V : il kernel $\text{Ker}(\mathbf{F}) \subseteq V$ viene mappato nel vettore nullo di W , mentre $\text{Im}(\mathbf{F})$ è il sottospazio di W in cui viene mappato l’intero spazio vettoriale di partenza.

4.3. Endomorfismi e cambi di base. Un discorso speciale meritano gli *endomorfismi*, ovvero le applicazioni lineari da V in se stesso. In questo caso la dimensione di partenza è uguale a quella di arrivo e pertanto una matrice che rappresenta un endomorfismo F è sempre *quadrata* e ha dimensione $n = \dim(V)$. Inoltre, possiamo scegliere anche di utilizzare una base di partenza *uguale* a quella di arrivo, sia essa per esempio $\mathcal{B}_b = \{\mathbf{b}_i\}_{i=1}^n$. Indichiamo con \mathbf{F}_b la matrice quadrata associata a F secondo questa base. Gli endomorfismi in V possono essere raggruppati in classi sulla base della relazione di *similitudine*, definita rispetto alle matrici che li rappresentano. Essa è una relazione di equivalenza nello spazio delle matrici.

DEFINIZIONE 4.2. Due matrici $\mathbf{A}, \mathbf{B} \in \mathbb{M}_n(\mathbb{K})$ sono dette simili se esiste una matrice $\mathbf{S} \in \text{GL}_n(\mathbb{K})$ tale che $\mathbf{B} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$.

Due matrici simili hanno lo stesso determinante. Inoltre, la rilevanza di questa definizione sta nella seguente Proposizione, che chiarisce l’effetto di un cambio di base sulla rappresentazione di un endomorfismo.

PROPOSIZIONE 4.2. Sia V un \mathbb{K} -spazio n -dimensionale. Due matrici $\mathbf{M}, \mathbf{M}' \in \mathbb{M}_n(\mathbb{K})$ sono simili se e solo se esiste un endomorfismo F su V e due basi $\mathcal{B}_{\hat{b}}$ e \mathcal{B}_b tali per cui $\mathbf{M} = \mathbf{F}_{\hat{b}}$ e $\mathbf{M}' = \mathbf{F}_b$.

In altre parole, *matrici simili rappresentano lo stesso endomorfismo in basi diverse*: $\mathbf{F}_{\hat{b}} = \mathbf{S}^{-1}\mathbf{F}_b\mathbf{S}$ e ciò che \mathbf{S} esegue è un semplice cambio di base per passare da \mathcal{B}_b a $\mathcal{B}_{\hat{b}}$. La matrice \mathbf{S} è la matrice di *cambio di base*: è costituita dalla *rappresentazione di una base* (quella di arrivo) in termini dell’altra (quella di partenza). Supponiamo infatti di avere due basi, $\mathcal{B}_{\hat{b}} = \{\hat{\mathbf{b}}_i\}_{i=1}^n$ e $\mathcal{B}_b = \{\mathbf{b}_i\}_{i=1}^n$, e che valgano le seguenti decomposizioni:

$$(1.73) \quad \mathbf{b}_\mu = \sum_{\nu=1}^n s_{\mu\nu} \hat{\mathbf{b}}_\nu \quad \hat{\mathbf{b}}_\mu = \sum_{\nu=1}^n \hat{s}_{\mu\nu} \mathbf{b}_\nu.$$

Le due matrici $\mathbf{S} = (s_{\mu\nu})_{\mu\nu}$ e $\hat{\mathbf{S}} = (\hat{s}_{\mu\nu})_{\mu\nu}$ non sono completamente indipendenti. Infatti

$$\mathbf{b}_\mu = \sum_{\nu=1}^n s_{\mu\nu} \hat{\mathbf{b}}_\nu = \sum_{\nu=1}^n \sum_{\rho=1}^n s_{\mu\nu} \hat{s}_{\nu\rho} \mathbf{b}_\rho = \sum_{\rho=1}^n \left[\sum_{\nu=1}^n s_{\mu\nu} \hat{s}_{\nu\rho} \right] \mathbf{b}_\rho = \sum_{\rho=1}^n [\mathbf{S}\hat{\mathbf{S}}]_{\nu\rho} \mathbf{b}_\rho.$$

Ora, \mathcal{B}_b è una base, e di conseguenza \mathbf{b}_μ è linearmente indipendente dai restanti componenti della base stessa: l’unico modo perché la relazione sopra sia consistente è che $[\mathbf{S}\hat{\mathbf{S}}]_{\nu\rho} = \delta_{\nu\rho}$, che vuol dire $\mathbf{S}\hat{\mathbf{S}} = \mathbf{I}_n$, ovvero $\hat{\mathbf{S}} = \mathbf{S}^{-1}$. La matrice di cambio di base \mathbf{S} ha, naturalmente, determinante non nullo (deve essere invertibile!). Si dice che due basi sono orientate in maniera *concorde* se, detta \mathbf{S} la corrispondente matrice di cambio di base, $\det(\mathbf{S}) > 0$; si dice invece che sono orientate

in maniera discorde se $\det(\mathbf{S}) < 0$. La relazione “essere concordi” è una relazione di equivalenza tra le basi.

Per vedere come opera \mathbf{S} , supponiamo di avere un vettore $\mathbf{x} \in \mathbb{V}$. Possiamo, come sappiamo, scriverlo univocamente in termini della base \mathcal{B}_b , come

$$\mathbf{x} = \sum_{\mu=1}^n x_\mu \mathbf{b}_\mu$$

che equivale a dire che, secondo la base \mathcal{B}_b , esso è rappresentato dal vettore di coordinate

$$\mathbf{x} \mapsto \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Proviamo ora a scrivere \mathbf{x} usando la base \mathcal{B}_e :

$$(1.74) \quad \mathbf{x} = \sum_{\mu=1}^n x_\mu \mathbf{b}_\mu = \sum_{\mu=1}^n \sum_{\nu=1}^n x_\mu s_{\mu\nu} \hat{\mathbf{b}}_\nu = \sum_{\nu=1}^n \left(\sum_{\mu=1}^n x_\mu s_{\mu\nu} \right) \hat{\mathbf{b}}_\nu$$

ovvero, secondo la base $\mathcal{B}_{\hat{b}}$, le coordinate di \mathbf{x} sono

$$(1.75) \quad \mathbf{x} \mapsto \hat{\mathbf{x}} = \begin{pmatrix} \sum_\mu x_\mu s_{\mu 1} \\ \vdots \\ \sum_\mu x_\mu s_{\mu n} \end{pmatrix} = \mathbf{S}^\top \mathbf{x}.$$

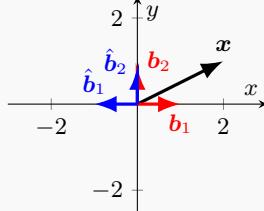
Consideriamo $\mathbb{V} = \mathbb{R}^2$ e le due basi

$$\mathcal{B}_b = \left\{ \mathbf{b}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{b}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

e

$$\mathcal{B}_{\hat{b}} = \left\{ \hat{\mathbf{b}}_1 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad \hat{\mathbf{b}}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}.$$

In un riferimento cartesiano queste possono essere rappresentate come in figura (\mathcal{B}_b in rosso, $\mathcal{B}_{\hat{b}}$ in blu)):



La matrice \mathbf{S} di cambiamento di base può essere ottenuta scrivendo la prima base utilizzando la seconda. È facile vedere che

$$\hat{\mathbf{b}}_1 = -\mathbf{b}_1, \quad \hat{\mathbf{b}}_2 = \mathbf{b}_2$$

quindi

$$\hat{\mathbf{S}} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \Rightarrow \mathbf{S} = \hat{\mathbf{S}}^{-1} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Se ora consideriamo un vettore \mathbf{x} , come in figura, tale che

$$\mathbf{x} = 2\mathbf{b}_1 + \mathbf{b}_2$$

esso ha quindi, secondo la prima base, una rappresentazione

$$\mathbf{x} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Per ottenere la sua rappresentazione secondo la *nuova* base possiamo scrivere

$$\hat{\mathbf{x}} = \mathbf{S}^T \mathbf{x} = \begin{pmatrix} -2 \\ 1 \end{pmatrix} \Leftrightarrow \mathbf{x} = -2\hat{\mathbf{b}}_1 + \hat{\mathbf{b}}_2.$$

Secondo la nostra definizione, le due basi sono discordi, dato che $\det(\mathbf{S}) = -1$.

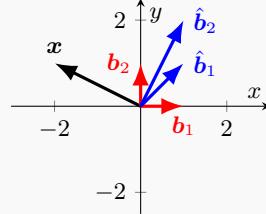
Consideriamo di nuovo $\mathbb{V} = \mathbb{R}^2$ e le due basi

$$\mathcal{B}_b = \left\{ \mathbf{b}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{b}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

e

$$\mathcal{B}_{\hat{b}} = \left\{ \hat{\mathbf{b}}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \hat{\mathbf{b}}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}.$$

In un riferimento cartesiano queste possono essere rappresentate come in figura (\mathcal{B}_b in rosso, $\mathcal{B}_{\hat{b}}$ in blu):



Come prima, matrice \mathbf{S} di cambiamento di base può essere ottenuta scrivendo la prima base utilizzando la seconda. Partendo dal fatto che

$$\hat{\mathbf{b}}_1 = \mathbf{b}_1 + \mathbf{b}_2, \quad \hat{\mathbf{b}}_2 = \mathbf{b}_1 + 2\mathbf{b}_2$$

abbiamo che

$$\hat{\mathbf{S}} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \Rightarrow \mathbf{S} = \hat{\mathbf{S}}^{-1} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}.$$

Se ora consideriamo un vettore \mathbf{x} , come in figura, tale che

$$\mathbf{x} = -2\mathbf{b}_1 + \mathbf{b}_2$$

esso ha quindi, secondo la prima base, una rappresentazione

$$\mathbf{x} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

Per ottenere la sua rappresentazione secondo la *nuova* base possiamo scrivere

$$\hat{\mathbf{x}} = \mathbf{S}^T \mathbf{x} = \begin{pmatrix} -5 \\ 3 \end{pmatrix} \Leftrightarrow \mathbf{x} = -5\hat{\mathbf{b}}_1 + 3\hat{\mathbf{b}}_2.$$

Secondo la nostra definizione, le due basi sono concordi, dato che $\det(\mathbf{S}) = 1 > 0$.

Vediamo ora come un cambiamento di base cambia la rappresentazione di una applicazione lineare \mathbf{F} . Per definizione, l'azione di \mathbf{F} è rappresentata, secondo la base \mathcal{B}_b , da una matrice \mathbf{F}_b , ovvero

$$\mathbf{F}(\mathbf{b}_\mu) = \sum_{\nu=1}^n [\mathbf{F}_b]_{\mu\nu} \mathbf{b}_\nu.$$

Indichiamo, per brevità, con $f_{\mu\nu}$ l'elemento $[\mathbf{F}_b]_{\mu\nu}$ della matrice \mathbf{F}_b . Per ottenere la rappresentazione di \mathbf{F} secondo la base $\mathcal{B}_{\hat{b}}$, dovremmo provare a scrivere una espressione del tipo

$$\mathbf{F}(\hat{\mathbf{b}}_\mu) = \sum_{\kappa=1}^n \hat{f}_{\mu\kappa} \hat{\mathbf{b}}_\kappa,$$

che coinvolge gli elementi $\hat{f}_{\mu\nu} := [\mathbf{F}_{\hat{\mathbf{b}}}]_{\mu\nu}$ della nuova rappresentazione. Tentiamo di calcolare quindi questa espressione:

$$\begin{aligned}
 F(\hat{\mathbf{b}}_\mu) &= \sum_{\nu=1}^n \hat{s}_{\mu\nu} \mathbf{F}(\mathbf{b}_\nu) = \sum_{\nu=1}^n \sum_{\rho=1}^n \hat{s}_{\mu\nu} f_{\nu\rho} \mathbf{b}_\rho \\
 &= \sum_{\nu=1}^n \sum_{\rho=1}^n \sum_{\kappa=1}^n \hat{s}_{\mu\nu} f_{\nu\rho} s_{\rho\kappa} \hat{\mathbf{b}}_\kappa \\
 (1.76) \quad &= \sum_{\kappa=1}^n \left[\sum_{\nu=1}^n \sum_{\kappa=1}^n \hat{s}_{\mu\nu} f_{\nu\rho} s_{\rho\kappa} \right] \hat{\mathbf{b}}_\kappa \\
 &\equiv \sum_{\kappa=1}^n [\mathbf{S}^{-1} \mathbf{F}_{\mathbf{b}} \mathbf{S}]_{\mu\kappa} \hat{\mathbf{b}}_\kappa.
 \end{aligned}$$

Dato che la rappresentazione di $\mathbf{F}(\hat{\mathbf{b}}_\mu)$ nella base $\mathcal{B}_{\hat{\mathbf{b}}}$ è unica, quel che abbiamo ottenuto implica che $\hat{f}_{\mu\kappa} = [\mathbf{S}^{-1} \mathbf{F}_{\mathbf{b}} \mathbf{S}]_{\mu\kappa}$, ovvero $\mathbf{F}_{\hat{\mathbf{b}}} = \mathbf{S}^{-1} \mathbf{F}_{\mathbf{b}} \mathbf{S}$. Si noti che, anche se gli elementi della rappresentazione di \mathbf{F} cambiano da una base all'altra, il *determinante* è una quantità invariante. Infatti

$$(1.77) \quad \det(\mathbf{F}_{\hat{\mathbf{b}}}) = \frac{1}{\det(\mathbf{S})} \det(\mathbf{F}_{\mathbf{b}}) \det(\mathbf{S}) = \det(\mathbf{F}_{\mathbf{b}}),$$

per cui scriveremo semplicemente $\det(\mathbf{F})$ per indicare questa quantità. Allo stesso modo, è invariante la *traccia*, per via della sua proprietà ciclica:

$$(1.78) \quad \text{tr}(\mathbf{F}_{\hat{\mathbf{b}}}) = \text{tr}(\mathbf{S}^{-1} \mathbf{F}_{\mathbf{b}} \mathbf{S}) = \text{tr}(\mathbf{S} \mathbf{S}^{-1} \mathbf{F}_{\mathbf{b}}) = \text{tr}(\mathbf{F}_{\mathbf{b}}).$$

per cui indicheremo la traccia di un endomorfismo \mathbf{F} semplicemente come $\text{tr}(\mathbf{F})$.

4.4. Diagonalizzabilità e polinomio caratteristico. Abbiamo visto che un cambio di base cambia la rappresentazione di un endomorfismo tramite una matrice di cambiamento di base \mathbf{S} . Ha senso quindi chiedersi se, tra tutte le basi, ne esista una che fornisce una rappresentazione particolarmente semplice dell'endomorfismo.

DEFINIZIONE 4.3. Un endomorfismo \mathbf{F} si dice *diagonalizzabile* se esiste una base $\mathcal{B}_e = \{\mathbf{e}_i\}_{i=1}^n$ di \mathbb{V} tale per cui

$$(1.79) \quad \mathbf{F}_e = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

I valori $\{\lambda_i\}_{i=1}^n$ costituiscono lo spettro di \mathbf{F} e sono detti *autovalori* di \mathbf{F} . La base \mathcal{B}_e si dice “diagonalizzante per \mathbf{F} ”.

Supponiamo di avere la rappresentazione matriciale \mathbf{F}_b di una applicazione F in una base $\mathcal{B}_b = \{\mathbf{b}_i\}_{i=1}^n$. Dire che \mathbf{F}_b è diagonalizzabile significa che esiste una diversa base $\mathcal{B}_e = \{\mathbf{e}_i\}_{i=1}^n$ in cui \mathbf{F}_e è diagonale. Il passaggio dalla base \mathcal{B}_b alla base \mathcal{B}_e sarà dato da una matrice \mathbf{S} che, come nell'Eq. (1.73), semplicemente esprime gli elementi di una base in termini dell'altra.

Avere una base diagonalizzante semplifica enormemente lo studio dell'azione di \mathbf{F} . Se la base \mathcal{B}_e diagonalizza \mathbf{F} , questo significa che

$$\mathbf{F}(\mathbf{e}_i) = \lambda_i \mathbf{e}_i, \quad i = 1, \dots, n.$$

Si dice che \mathbf{e}_i è un *autovettore* corrispondente all'autovalore λ_i : esso corrisponde ad una “direzione speciale” nello spazio vettoriale in cui l'azione di \mathbf{F} si riduce semplicemente ad un cambio di scala e possibilmente un'inversione (se $\lambda_i < 0$) mentre la direzione rimane inalterata. Essendo \mathcal{B}_e una

base, ciò significa che, dato un $\mathbf{x} \in \mathbb{V}$, possiamo scrivere $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i$ per qualche set di coefficienti opportuno. Scrivendo \mathbf{x} in questa base, l'azione di F su \mathbf{x} diventa semplice, dato che

$$\mathbf{F}(\mathbf{x}) = \sum_{i=1}^n x_i \mathbf{F}(\mathbf{e}_i) = \sum_{i=1}^n \lambda_i x_i \mathbf{e}_i,$$

e, per esempio, applicando \mathbf{F} k volte,

$$\mathbf{F}^{\otimes k}(\mathbf{x}) = \sum_{i=1}^n \lambda_i^k x_i \mathbf{e}_i.$$

Altre quantità legate all'endomorfismo possono calcolarsi altrettanto facilmente. Per esempio, dato che traccia e determinante rimangono invariate dopo un cambiamento di base, esse si ottengono prontamente utilizzando la forma diagonale,

$$(1.80) \quad \det(\mathbf{F}) = \prod_{i=1}^n \lambda_i, \quad \text{tr}(\mathbf{F}) = \sum_{i=1}^n \lambda_i.$$

La domanda a questo punto è se, dato un endomorfismo, esiste una *base di autovettori*, e se sì, come trovarla.

4.5. Calcolo di autovalori e autovettori. Consideriamo quindi una matrice quadrata $\mathbf{F}_b \in \mathbb{M}_n(\mathbb{K})$, che in generale rappresenta quindi un endomorfismo secondo la base \mathcal{B}_b . Per trovare un autovettore e il corrispondente autovalore dovremmo risolvere l'equazione $\mathbf{F}_b \mathbf{x} = \lambda \mathbf{x}$, dove sia λ che \mathbf{x} sono incogniti. Se esistono un $\mathbf{x} \neq \mathbf{0}$ e un λ tali per cui l'equazione è valida, essi sono rispettivamente autovettore e autovalore di \mathbf{F}_b . Si noti che \mathbf{x} non è univocamente determinato: se \mathbf{x} è autovettore di \mathbf{F}_b con autovalore λ , anche $c\mathbf{x}$ è autovettore di \mathbf{F}_b con lo stesso autovalore per ogni c non-nullo.

Cercare un autovettore equivale a cercare di risolvere il sistema omogeneo

$$(1.81) \quad (\mathbf{F}_b - \lambda \mathbf{I}) \mathbf{x} = \mathbf{0}.$$

Ora, tale sistema ha come unica soluzione $\mathbf{x} = \mathbf{0}$ se il determinante della matrice a moltiplicare è non nullo, e quindi la matrice è invertibile: questo tipo di soluzione però non è di nostro interesse, perché gli autovettori devono essere elementi di una base e quindi non nulli! Perché esistano soluzioni non triviali al problema deve essere

$$(1.82) \quad \det(\mathbf{F}_b - \lambda \mathbf{I}) = 0.$$

L'espressione sopra è un *polinomio*, detto *polinomio caratteristico*, ed è di grado n : come si vede non coinvolge \mathbf{x} ma fornisce una equazione per i soli autovalori, fornendo al più n radici. È importante osservare che l'equazione sopra è invariante per cambi di base e quindi gli autovalori non dipendono dalla base scelta per rappresentare l'endomorfismo. Se per esempio consideriamo una base $\mathcal{B}_{\hat{b}}$, tale per cui $\mathbf{F}_b = \mathbf{S}^{-1} \mathbf{F}_{\hat{b}} \mathbf{S}$, allora

$$(1.83) \quad \begin{aligned} 0 &= \det(\mathbf{F}_b - \lambda \mathbf{I}) \\ &= \det(\mathbf{S}^{-1} \mathbf{F}_{\hat{b}} \mathbf{S} - \lambda \mathbf{I}) \\ &= \det(\mathbf{S}^{-1} (\mathbf{F}_{\hat{b}} - \lambda \mathbf{I}) \mathbf{S}) \\ &= \det(\mathbf{F}_{\hat{b}} - \lambda \mathbf{I}). \end{aligned}$$

I singoli autovalori sono quindi proprietà intrinseche dell'endomorfismo \mathbf{F} , ovvero delle quantità che non cambiano cambiando base. La conseguenza delle considerazioni esposte finora è data dalla seguente

PROPOSIZIONE 4.3. *Sia \mathbb{V} uno spazio vettoriale di dimensione n e \mathbf{F} un suo endomorfismo. Allora $\lambda \in \mathbb{K}$ è autovalore di \mathbf{F} se e solo se λ è radice del polinomio caratteristico. Di conseguenza \mathbf{F} ha al più n autovalori distinti.*

Trovato un autovalore λ_i , possiamo cercare i corrispondenti autovettori risolvendo l'equazione $(\mathbf{F}_b - \lambda_i \mathbf{I})\mathbf{x} = 0$. Il nostro obiettivo ultimo sarà costruire una base usando questi autovettori. Un risultato utile, in questo senso, è il seguente

PROPOSIZIONE 4.4. *Autovettori corrispondenti ad autovalori distinti sono linearmente indipendenti.*

DIMOSTRAZIONE. Supponiamo che la tesi sia falsa. Sia \mathbf{v}_i autovettore con autovalore λ_i e \mathbf{v}_j autovettore con autovalore λ_j . Allora esistono due costanti c_1 e c_2 non nulli tali che $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}$. Applicando \mathbf{F} , otteniamo $c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 = \mathbf{0}$; ma d'altronde $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0} \Rightarrow c_1\lambda_1\mathbf{v}_1 + c_2\lambda_1\mathbf{v}_2 = \mathbf{0}$. Prendendo la differenza tra le due equazioni, troviamo $c_2(\lambda_2 - \lambda_1)\mathbf{v}_2 = \mathbf{0}$ che è assurdo. \square

Autovettori corrispondenti a diversi autovalori sono quindi linearmente indipendenti. A λ_i fissato, però, la condizione $\det(\mathbf{F}_b - \lambda_i \mathbf{I}) = 0$ fa sì che l'equazione $\mathbf{F}_b \mathbf{x} = \lambda_i \mathbf{x}$ abbia una soluzione non unica, come sappiamo: le soluzioni in effetti vivono in un certo spazio che, come specificato dal teorema di Kronecker–Rouché–Capelli, ha dimensione maggiore o uguale ad uno e che è $\text{Ker}(\mathbf{F}_b - \lambda_i \mathbf{I})$. Come sappiamo, questo è uno spazio vettoriale: infatti, se \mathbf{v}_1 e \mathbf{v}_2 sono elementi di $\text{Ker}(\mathbf{F}_b - \lambda_i \mathbf{I})$, allora $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \in \text{Ker}(\mathbf{F}_b - \lambda_i \mathbf{I})$. La dimensione $n_i := \dim \text{Ker}(\mathbf{F}_b - \lambda_i \mathbf{I})$ è detta *moltiplicità geometrica*: questo vuol dire che al suo interno possiamo trovare una base \mathcal{B}_i di n_i autovettori indipendenti corrispondenti all'autovalore λ_i che genera lo spazio degli autovettori di \mathbf{F} con autovalore λ_i . La molteplicità algebrica associata ad un autovalore è *almeno 1* (la condizione $\det(\mathbf{F}_b - \lambda_i \mathbf{I}) = 0$ implica che $n_i \geq 1$) e *al più n* (n_i non può essere più grande della dimensione dello spazio). Chiamiamo $\text{Ker}(\mathbf{F}_b - \lambda_i \mathbf{I})$ *autospazio* associato a λ_i e lo indichiamo con \mathbb{V}_i . D'altra parte, abbiamo visto che autovettori di autospazi diversi sono linearmente indipendenti. Se abbiamo quindi k autovalori distinti, in totale possiamo collezionare $\sum_{i=1}^k n_i$ autovettori linearmente indipendenti, considerando l'unione di tutte le basi dei vari autospazi, $\bigcup_{i=1}^k \mathcal{B}_i$. Il seguente teorema stabilisce che possiamo in effetti diagonalizzare la nostra matrice se e solo se questo numero è proprio uguale al numero di autovettori di cui abbiamo bisogno per costruire una base per \mathbb{V} , ovvero n .

TEOREMA 4.5 (Diagonalizzabilità). *Sia $\Lambda = \{\lambda_i\}_{i=1}^k$ lo spettro di \mathbf{F} , costituito da $k \leq n$ autovalori distinti. Allora*

$$(1.84) \quad \sum_{i=1}^k \dim \mathbb{V}_i \leq n.$$

\mathbf{F} è diagonalizzabile se e solo se vale l'uguaglianza. In particolare, ciò è verificato se $k = n$.

Per fissare le idee, consideriamo qualche esempio.

Consideriamo la rappresentazione

$$\mathbf{F}_b = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

di un endomorfismo su uno spazio bidimensionale rappresentato secondo una certa base $\mathcal{B}_b = \{\mathbf{b}_1, \mathbf{b}_2\}$. Vogliamo trovare autovettori e autovalori. Anzitutto calcoliamo il polinomio caratteristico come

$$\det \begin{pmatrix} 1 - \lambda & 2 \\ 2 & 1 - \lambda \end{pmatrix} = (1 - \lambda)^2 - 4.$$

Cerchiamone quindi gli zeri scrivendo $(1 - \lambda)^2 - 4 = 0 \Rightarrow \lambda_1 = -1 \wedge \lambda_2 = 3$. Per ottenere gli autovettori, dobbiamo risolvere *due* sistemi, uno per ogni autovalore trovato.

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = - \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Naturalmente questo sistema non ha un'unica soluzione. Otteniamo quindi che in generale la soluzione è nella forma $(x_1, x_2) = t(1, -1)$ per $t \in \mathbb{R}$: questo è uno spazio unidimensionale, cioè $\dim \mathbb{V}_1 = 1$. Usando l'autovalore λ_2 invece otteniamo, esattamente nello stesso modo, che gli autovettori corrispondenti hanno la forma $(x_1, x_2) = t(1, 1)$. Anche questo è uno spazio unidimensionale, $\dim \mathbb{V}_2 = 1$. Avendo ottenuto due autovalori distinti, il set

$$\mathcal{B}_e = \{\mathbf{e}_1, \mathbf{e}_2\}, \quad \mathbf{e}_1 = \mathbf{b}_1 - \mathbf{b}_2, \quad \mathbf{e}_2 = \mathbf{b}_1 + \mathbf{b}_2,$$

è una base diagonalizzante, ovvero

$$\hat{\mathbf{S}} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \Rightarrow \mathbf{S} = \hat{\mathbf{S}}^{-1} = \begin{pmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

Per vedere qual è la rappresentazione di \mathbf{F}_e nella nuova base, scriviamo

$$\mathbf{F}_e = \mathbf{S}^{-1} \mathbf{F}_b \mathbf{S} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}$$

come desiderato.

Consideriamo ora la matrice

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

e cerchiamone autovettori e autovalori esattamente come sopra. Come sopra, calcoliamo il polinomio caratteristico come

$$\det \begin{pmatrix} 1 - \lambda & 2 \\ 0 & 1 - \lambda \end{pmatrix} = (1 - \lambda)^2.$$

Si ha una sola soluzione, $\lambda = 1$. Cerchiamo ora gli autovettori risolvendo

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

La seconda equazione fornisce l'identità (non informativa) $x_2 = x_2$ mentre la prima ci dà $x_1 + 2x_2 = x_1 \Rightarrow x_2 = 0$. Non abbiamo quindi alcuna informazione su x_1 e la nostra soluzione è nella forma $(x_1, x_2) = t(1, 0)$ con $t \in \mathbb{R}$ qualsivoglia. La dimensione di questo spazio è 1 e corrisponde all'unico autovalore che abbiamo trovato, quindi la matrice *non* è diagonalizzabile alla luce del Teorema 4.5.

Consideriamo la matrice

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

rappresentazione di un endomorfismo su \mathbb{R}^3 secondo la base canonica

$$\mathcal{B}_b = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

e cerchiamo autovalori e autovettori come prima. Otteniamo che

$$\det(\mathbf{A} - \lambda \mathbf{I}_3) = -(\lambda + 1)(\lambda - 1)^2.$$

Abbiamo così due autovalori, $\lambda_1 = 1$ e $\lambda_2 = -1$.

Studiamo ora l'autospazio corrispondente a λ_1 . Se cerchiamo ora una soluzione al sistema $(\mathbf{A} - \mathbf{I})\mathbf{x} = \mathbf{0}$, vediamo che le soluzioni sono tutte le combinazioni lineari nella forma

$$\mathbf{x} = t_1 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + t_2 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad t_1, t_2 \in \mathbb{R}.$$

Questo è uno spazio *bidimensionale* \mathbb{V}_1 generato dai vettori

$$\mathcal{B}_1 = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}.$$

L'autovalore $\lambda_2 = -1$, invece, è associato a soluzioni nella forma

$$\mathbf{x} = t \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad t \in \mathbb{R},$$

ovvero ad uno spazio unidimensionale \mathbb{V}_2 generato dalla base

$$\mathcal{B}_2 = \left\{ \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}.$$

Essendo $\dim \mathbb{V}_1 + \dim \mathbb{V}_2 = 3$, la matrice è diagonalizzabile. Vediamo come effettuare il cambio di variabile dalla base canonica alla nuova base appena trovata

$$\mathcal{B}_e = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}, \quad \mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Scriviamo \mathbf{S} esprimendo la base canonica in termini della nuova base. Essendo

$$\mathbf{b}_1 = \frac{1}{2}\mathbf{e}_1 + \frac{1}{2}\mathbf{e}_2, \quad \mathbf{b}_2 = \frac{1}{2}\mathbf{e}_1 - \frac{1}{2}\mathbf{e}_2, \quad \mathbf{b}_3 = \mathbf{e}_3$$

per cui

$$\mathbf{S} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & -1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \Leftrightarrow \mathbf{S}^{-1} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

e infatti

$$\mathbf{S}^{-1} \mathbf{A} \mathbf{S} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

in cui come si vede λ_1 compare un numero di volte uguale alla sua molteplicità geometrica.

5. Spazi prehilbertiani

5.1. Forme bilineari. Il nostro punto di partenza sono stati vettori nello spazio usuale, in cui ha senso parlare di lunghezza, angolo, perpendicolarità e così via. Per introdurre concetti analoghi in uno spazio vettoriale \mathbb{V} , in questa sezione ci focalizzeremo su spazi *reali*, ovvero tali per cui $\mathbb{K} = \mathbb{R}$, e daremo, per cominciare, una definizione di grande importanza.

DEFINIZIONE 5.1 (Forma bilineare). *Dato uno spazio vettoriale reale, una applicazione $\varphi: \mathbb{V} \times \mathbb{V} \rightarrow \mathbb{R}$ si dice essere una forma bilineare se gode delle seguenti proprietà per ogni $\mathbf{v}, \mathbf{u}, \mathbf{w} \in \mathbb{V}$*

- (1) $\varphi(\mathbf{v} + \mathbf{u}, \mathbf{w}) = \varphi(\mathbf{v}, \mathbf{w}) + \varphi(\mathbf{u}, \mathbf{w})$;
- (2) $\varphi(\mathbf{v}, \mathbf{w} + \mathbf{u}) = \varphi(\mathbf{v}, \mathbf{w}) + \varphi(\mathbf{v}, \mathbf{u})$;
- (3) $\varphi(c\mathbf{v}, \mathbf{u}) = \varphi(\mathbf{v}, c\mathbf{u}) = c\varphi(\mathbf{v}, \mathbf{u}) \quad \forall c \in \mathbb{R}$.

La forma si dice simmetrica se $\varphi(\mathbf{v}, \mathbf{u}) = \varphi(\mathbf{u}, \mathbf{v})$ e antisimmetrica se $\varphi(\mathbf{v}, \mathbf{u}) = -\varphi(\mathbf{u}, \mathbf{v})$ (in questo secondo caso, $\varphi(\mathbf{v}, \mathbf{v}) = 0$).

Data una matrice quadrata $\mathbf{A} \in \text{M}_n(\mathbb{R})$, un esempio di forma bilineare indotta da \mathbf{A} su uno spazio n -dimensionale \mathbb{V} è $\varphi(\mathbf{v}, \mathbf{u}) = \mathbf{v}^\top \mathbf{A} \mathbf{u} = \sum_{\mu\nu} v_\mu A_{\mu\nu} u_\nu$, dove \mathbf{u}, \mathbf{v} sono le coordinate dei vettori \mathbf{v} e \mathbf{u} secondo una certa base prefissata \mathcal{B}_b . La forma è simmetrica se \mathbf{A} è simmetrica. Vedremo che in effetti *tutte* le forme bilineari possono essere scritte a questo modo. Un caso molto importante, come vedremo in seguito, è quello in cui $\mathbf{A} = \mathbf{I}_n$, per cui

$$\varphi(\mathbf{v}, \mathbf{u}) = \sum_{\mu=1}^n v_\mu u_\mu \equiv \mathbf{v}^\top \mathbf{u}.$$

Esattamente come avviene tra endomorfismo e sua matrice rappresentativa, esiste una corrispondenza, *a base fissata*, tra *forma bilineare* e *matrice quadrata*. Dato uno spazio vettoriale n -dimensionale \mathbb{V} , basta infatti fissare una base $\mathcal{B}_b = \{\mathbf{b}_i\}_{i=1}^n$ per notare che una forma bilineare corrisponde, nella base scelta, ad una matrice. Dati due vettori $\mathbf{v}, \mathbf{u} \in \mathbb{V}$, $\mathbf{v} = \sum_i v_i \mathbf{b}_i$ e

$\mathbf{u} = \sum_i u_i \mathbf{b}_i$, e dunque, se $\varphi_{ij} := \varphi(\mathbf{b}_i, \mathbf{b}_j)$,

$$(1.85) \quad \varphi(\mathbf{v}, \mathbf{u}) = \sum_{ij} v_i \varphi(\mathbf{b}_i, \mathbf{b}_j) u_j = \mathbf{v}^\top \Phi_b \mathbf{u}.$$

In altre parole, ogni forma bilineare φ corrisponde, secondo una data base \mathcal{B}_b , ad una matrice quadrata $\Phi_b = (\varphi_{ij})_{ij}$ di dimensione $n \times n$, mentre viceversa è possibile usare una matrice quadrata di dimensione $n \times n$ ed una base per definire una forma bilineare. La forma è simmetrica se la matrice associata è simmetrica: questa proprietà non dipende dalla base utilizzata. Come nel caso degli endomorfismi, la traccia e il determinante di Φ_b non dipendono dalla base utilizzata, e sono quindi due proprietà che si possono riferire alla forma stessa, ovvero si può dire *traccia* o *determinante* di φ .

Supponiamo infine che \mathbf{e} sia un autovettore di Φ_b con autovalore λ . Questo significa che il corrispondente vettore \mathbf{e} (ovvero il vettore avente come coordinate \mathbf{e}) è tale che

$$\varphi(\mathbf{e}, \mathbf{e}) = \mathbf{e}^\top \Phi_b \mathbf{e} = \lambda \mathbf{e}^\top \mathbf{e} = \lambda \|\mathbf{e}\|^2.$$

Si dice che \mathbf{e} è un autovettore della forma, esattamente come \mathbf{e} lo è per la rappresentazione matriciale Φ_b .

5.1.1. *Forme simmetriche.* Un fatto molto interessante riguardo le forme bilineari simmetriche è il seguente teorema, che diamo senza dimostrazione.

TEOREMA 5.1. *Sia \mathbb{V} uno spazio vettoriale su \mathbb{R} di dimensione finita. Se φ è una forma bilineare simmetrica, allora esiste una base che diagonalizza φ , ovvero tale per cui la matrice che rappresenta φ è diagonale.*

Questo teorema può essere inteso nel seguente modo: ogni matrice simmetrica reale, essa è simile ad una matrice diagonale. Infatti c'è una corrispondenza biunivoca tra matrici simmetriche reali e forme bilineari simmetriche, e dire che le prime sono diagonalizzabili è equivalente a dire che lo sono le seconde.

Principalmente nel contesto delle forme bilineari simmetriche si dà la seguente definizione

DEFINIZIONE 5.2 (Forma definita positiva). *Una forma bilineare simmetrica φ si dice*

- *definita positiva se e solo se $\varphi(\mathbf{v}, \mathbf{v}) > 0 \forall \mathbf{v} \in \mathbb{V}$;*
- *semidefinita positiva se e solo se $\varphi(\mathbf{v}, \mathbf{v}) \geq 0 \forall \mathbf{v} \in \mathbb{V}$;*
- *definita negativa se e solo se $\varphi(\mathbf{v}, \mathbf{v}) < 0 \forall \mathbf{v} \in \mathbb{V}$;*
- *semidefinita negativa se e solo se $\varphi(\mathbf{v}, \mathbf{v}) \leq 0 \forall \mathbf{v} \in \mathbb{V}$.*

Gli autovalori della forma possono darci una informazione sulla natura della forma stessa, come stabilito dalla seguente proposizione che non dimostriamo.

PROPOSIZIONE 5.2. *Sia φ una forma bilineare simmetrica con spettro $\Lambda = \{\lambda_i\}_{i=1}^n$. Allora essa è*

- *definita positiva se e solo se $\lambda_i > 0 \forall i = 1, \dots, n$;*
- *semidefinita positiva se e solo se $\lambda_i \geq 0 \forall i = 1, \dots, n$;*
- *definita negativa se e solo se $\lambda_i < 0 \forall i = 1, \dots, n$;*
- *semidefinita negativa se e solo se $\lambda_i \leq 0 \forall i = 1, \dots, n$.*

Le forme bilineari simmetriche permettono di definire dei concetti che generalizzano quelli noti dalla geometria elementare, per esempio quello di *ortogonalità*.

DEFINIZIONE 5.3 (Ortogonalità). *Sia φ una forma bilineare simmetrica. Due vettori $\mathbf{v}, \mathbf{u} \in \mathbb{V}$ sono ortogonali, o perpendicolari, secondo φ se $\varphi(\mathbf{u}, \mathbf{v}) = 0$.*

Sia $\$$ un sottoinsieme di \mathbb{V} : possiamo definire l'insieme *ortogonale* a $\$$ come

$$\$^\perp := \{\mathbf{w} \in \mathbb{V}: \varphi(\mathbf{w}, \mathbf{v}) = 0 \ \forall \mathbf{v} \in \$\}.$$

Lo spazio $\$^\perp$ è un sottospazio vettoriale di \mathbb{V} .

5.2. Prodotti scalari. Data una forma bilineare simmetrica, è possibile associarvi una funzione detta *forma quadratica* $q_\varphi: \mathbb{V} \rightarrow \mathbb{R}$, definita come

$$q_\varphi(\mathbf{v}) := \varphi(\mathbf{v}, \mathbf{v}) = \sum_{ij} v_i v_j \varphi_{ij}.$$

Se la forma φ è anche definita positiva, allora essa si dice essere un *prodotto scalare*. Un prodotto scalare può essere utilizzato per dare una struttura *euclidea* ad uno spazio vettoriale: aiuta infatti a introdurre quantità equivalenti a lunghezza, distanza, angoli e così via. Uno spazio vettoriale dotato di prodotto scalare si dice *spazio prehilbertiano*. Un prodotto scalare è in genere indicato con, una notazione speciale, ovvero semplicemente

$$(1.86) \quad \varphi(\mathbf{v}, \mathbf{u}) = \langle \mathbf{v}, \mathbf{u} \rangle.$$

Si usa anche la notazione

$$(1.87) \quad q_\varphi(\mathbf{v}) = \|\mathbf{v}\|^2,$$

per la corrispondente forma quadratica: la quantità $\|\mathbf{v}\|$ è detta *norma* o *lunghezza* del vettore \mathbf{v} : se $\|\mathbf{v}\| = 1$ si dice che \mathbf{v} è un *versore*.

PROPOSIZIONE 5.3 (Disuguaglianza di Schwarz). *Se $\mathbf{v}, \mathbf{u} \in \mathbb{V}$, allora*

$$\langle \mathbf{v}, \mathbf{u} \rangle^2 \leq \|\mathbf{v}\|^2 \|\mathbf{u}\|^2,$$

mentre l'uguaglianza vale se e solo se \mathbf{v} e \mathbf{u} sono linearmente dipendenti, ovvero paralleli.

DIMOSTRAZIONE. Consideriamo due vettori $\mathbf{v}, \mathbf{u} \in \mathbb{V}$. Abbiamo che, per un $c \in \mathbb{R} \setminus \{0\}$

$$0 \leq \langle c\mathbf{v} + \mathbf{u}, c\mathbf{v} + \mathbf{u} \rangle = c^2 \|\mathbf{v}\|^2 + 2c\langle \mathbf{v}, \mathbf{u} \rangle + \|\mathbf{u}\|^2.$$

Questo polinomio quadratico in c deve essere sempre positivo, quindi il suo discriminante deve essere negativo o nullo, ovvero $4\langle \mathbf{v}, \mathbf{u} \rangle^2 - 4\|\mathbf{v}\|^2 \|\mathbf{u}\|^2 \leq 0 \Rightarrow \langle \mathbf{v}, \mathbf{u} \rangle^2 \leq \|\mathbf{v}\|^2 \|\mathbf{u}\|^2$. L'uguaglianza vale se e solo se $c\mathbf{v} + \mathbf{u} = \mathbf{0}$, ovvero i due vettori sono linearmente dipendenti. \square

La disuguaglianza di Schwarz si può scrivere anche come

$$(1.88) \quad -1 \leq \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{v}\| \|\mathbf{u}\|} \leq 1,$$

il che suggerisce di definire un “angolo” θ tra due vettori come

$$(1.89) \quad \theta := \arccos \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{v}\| \|\mathbf{u}\|} \in [0, \pi].$$

Si noti che questa definizione è generalissima e non è necessario che \mathbf{v} e \mathbf{u} siano associati ad un significato geometrico (abbiamo visto, per esempio, che potrebbero essere polinomi!), e di conseguenza non è sempre “visivamente intuitivo” a cosa questo angolo corrisponda: esso misura un certo grado di lineare indipendenza tra i vettori, ed è zero quando i vettori sono linearmente dipendenti. Una volta definito tale “angolo”, possiamo riscrivere l'espressione di prodotto scalare come

$$\langle \mathbf{v}, \mathbf{u} \rangle = \|\mathbf{v}\| \|\mathbf{u}\| \cos \theta.$$

La disuguaglianza di Schwarz implica infine che

$$\|\mathbf{v} + \mathbf{u}\|^2 = \|\mathbf{v}\|^2 + 2\langle \mathbf{v}, \mathbf{u} \rangle + \|\mathbf{u}\|^2 \leq \|\mathbf{v}\|^2 + 2\|\mathbf{u}\| \|\mathbf{v}\| + \|\mathbf{u}\|^2 = (\|\mathbf{u}\| + \|\mathbf{v}\|)^2,$$

ovvero, implica la cosiddetta *disuguaglianza triangolare*

$$(1.90) \quad \|\mathbf{v} + \mathbf{u}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$

L'ortogonalità dei vettori in uno spazio vettoriale secondo il prodotto scalare è anch'esso un fatto con importanti conseguenze. Supponiamo di avere un *insieme ortonormale di vettori*, ovvero un insieme di versori $\{\mathbf{v}_1, \dots, \mathbf{v}_t\}$ tale per cui $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \delta_{ij}$ per $i, j \in \{1, \dots, t\}$. Allora vale la seguente Proposizione.

PROPOSIZIONE 5.4 (Base ortonormale). *Sia $\{\mathbf{v}_1, \dots, \mathbf{v}_t\}$ un insieme di t vettori ortonormali in \mathbb{V} . Allora, essi sono linearmente indipendenti. In particolare, se $t = n = \dim(\mathbb{V})$, l'insieme ortonormale dato costituisce una base per \mathbb{V} .*

DIMOSTRAZIONE. Supponiamo per assurdo che esistano c_i , $i = 1, \dots, t$, non tutti nulli tali per cui $\sum_i c_i \mathbf{v}_i = \mathbf{0}$. Allora per ogni $k = 1, \dots, t$, $0 = \langle \mathbf{v}_k, \sum_i c_i \mathbf{v}_i \rangle = \sum_i c_i \langle \mathbf{v}_k, \mathbf{v}_i \rangle = c_k$, che è l'assurdo cercato. \square

Data una base ortonormale in uno spazio vettoriale d -dimensionale, sia essa per esempio $\mathcal{B}_e = \{\mathbf{e}_i\}_{i=1}^d$, il prodotto scalare tra due vettori può essere espresso in maniera molto semplice in termini dei loro coefficienti rispetto alla base. Se per esempio

$$\mathbf{v} = \sum_{i=1}^d v_i \mathbf{e}_i, \quad \mathbf{u} = \sum_{i=1}^d u_i \mathbf{e}_i,$$

allora

$$(1.91) \quad \langle \mathbf{v}, \mathbf{u} \rangle = \sum_{i,j} v_i u_j \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \sum_{i=1}^d v_i u_i = \mathbf{v}^\top \mathbf{u},$$

e in particolare $\|\mathbf{v}\|^2 = \sum_{i=1}^d v_i^2$. Come si vede, questa definizione è giustifica la nomenclatura data in Eq. (1.16): ciò che lì era definito tra vettori numerici (ovvero righe e colonne di numeri) qui viene applicato alla *rappresentazione in coordinate* di vettori generici.

5.3. Ortonormalizzazione di una base: la procedura di Gram–Schmidt. Ha senso chiedersi come costruire una base ortonormale, per esempio avendo a disposizione una base $\mathcal{B}_b = \{\mathbf{b}_i\}_{i=1}^d$ di un \mathbb{R} -spazio vettoriale che non è ortonormale. Esiste una procedura algoritmica, detta di Gram–Schmidt, che permette di raggiungere questo obiettivo “ortonormalizzando” \mathcal{B}_b . La procedura si basa sull'uso della seguente operazione di *proiezione di un vettore \mathbf{v} lungo \mathbf{u}* ,

$$(1.92) \quad \text{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|^2} \mathbf{u}.$$

Il primo step della procedura consiste nel normalizzare un vettore della base \mathcal{B}_b , per esempio \mathbf{b}_1 , scrivendo

$$\mathbf{u}_1 := \mathbf{b}_1, \quad \mathbf{e}_1 := \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|}.$$

A questo punto costruiamo, al secondo step,

$$\mathbf{u}_2 = \mathbf{b}_2 - \text{proj}_{\mathbf{u}_1}(\mathbf{b}_2), \quad \mathbf{e}_2 := \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|},$$

e in generale allo step k si rimuove la proiezione di \mathbf{b}_k lungo tutte le direzioni torvate fino a quel punto

$$\mathbf{u}_k = \mathbf{b}_k - \sum_{i=1}^{k-1} \text{proj}_{\mathbf{u}_i}(\mathbf{b}_k), \quad \mathbf{e}_k := \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|}.$$

I vettori $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$ ottenuti al d -esimo passo sono normalizzati per costruzione. Sono anche ortogonali (questo può essere dimostrato per induzione partendo dal fatto che $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle = 0$ e poi

procedendo per tutte le coppie $\langle \mathbf{e}_1, \mathbf{e}_k \rangle$ per $k = 3, \dots, d$). L'insieme $\mathcal{B}_{\mathbf{e}} = \{\mathbf{e}_i\}_{i=1}^d$ è quindi una base ortonormale.

5.4. Cambio di base ortonormale. Supponiamo di studiare uno spazio reale prehilbertiano \mathbb{V} e di avere due basi ortonormali $\mathcal{B}_{\mathbf{e}} = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_d\}$ e $\mathcal{B}_{\hat{\mathbf{e}}} = \{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \dots, \hat{\mathbf{e}}_d\}$. Abbiamo visto, nella Sezione 4.3, che, date le coordinate \mathbf{v} di un vettore \mathbf{v} espresse nella base $\mathcal{B}_{\mathbf{e}}$, possiamo ottenere le sue coordinate nella base $\mathcal{B}_{\hat{\mathbf{e}}}$ usando la matrice di cambiamento di base $\mathbf{S} = (s_{ij})_{ij} \in \mathbb{M}_d(\mathbb{R})$ che permette di scrivere gli elementi della nuova base in termini di quelli della vecchia,

$$(1.93) \quad \hat{\mathbf{e}}_i = \sum_{j=1}^d \hat{s}_{ij} \mathbf{e}_j, \quad \mathbf{e}_i = \sum_{j=1}^d s_{ij} \hat{\mathbf{e}}_j,$$

dove $\hat{\mathbf{S}} = (\hat{s}_{ij})_{ij} = \mathbf{S}^{-1}$. La *nuova* rappresentazione di \mathbf{v} nella nuova base, sia $\hat{\mathbf{v}}$, si otterrà dalla *vecchia* rappresentazione \mathbf{v} calcolando $\hat{\mathbf{v}} = \mathbf{S}^T \mathbf{v}$. Nel caso di basi ortonormali, però, la matrice \mathbf{S} ha alcune proprietà interessanti. Partiamo dall'osservazione:

$$\delta_{ij} = \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \sum_{k=1}^d \sum_{r=1}^d s_{ik} s_{jr} \langle \hat{\mathbf{e}}_k, \hat{\mathbf{e}}_r \rangle = \sum_{k=1}^d s_{ik} s_{jk} \Leftrightarrow \mathbf{S} \mathbf{S}^T = \mathbf{I}_d \Rightarrow \mathbf{S}^T = \mathbf{S}^{-1}.$$

Si dice quindi che \mathbf{S} è una matrice *ortogonale*, ovvero tale per cui $\mathbf{S}^T = \mathbf{S}^{-1}$, e si scrive $\mathbf{S} \in \mathrm{O}(d)$, come abbiamo già visto. Questo significa anche che $1 = \det(\mathbf{S} \mathbf{S}^T) = \det^2(\mathbf{S})$, ovvero $\det(\mathbf{S}) = \pm 1$: se vogliamo che le basi di partenza e di arrivo siano concordi, occorrerà che $\det(\mathbf{S}) = 1$. Ora, il *prodotto scalare è invariante per cambi di base ortogonali*. Se \mathbf{v} ha coordinate \mathbf{v} e \mathbf{u} ha coordinate \mathbf{u} nella base $\mathcal{B}_{\mathbf{e}}$, allora essi avranno coordinate rispettivamente $\hat{\mathbf{v}} = \mathbf{S}^T \mathbf{v}$ e $\hat{\mathbf{u}} = \mathbf{S}^T \mathbf{u}$ nella base $\mathcal{B}_{\hat{\mathbf{e}}}$: tuttavia,

$$\langle \mathbf{v}, \mathbf{u} \rangle = \mathbf{v}^T \mathbf{u} = \mathbf{v} \mathbf{S} \mathbf{S}^T \mathbf{u} = (\mathbf{S}^T \mathbf{v})^T \mathbf{S}^T \mathbf{u} = \hat{\mathbf{v}}^T \hat{\mathbf{u}}.$$

Questo significa che anche le *norme* sono invarianti. Ogni relazione scritta in termini di prodotti scalari e norme, quindi, non cambia forma se si cambia base ortonormale: questo è uno dei vantaggi di aver introdotto il formalismo di spazio prehilbertiano.

5.4.1. Il caso bidimensionale. Il caso bidimensionale è uno speciale caso particolare molto istruttivo. Una matrice ortogonale positiva avrà, per $d = 2$, la semplice forma

$$(1.94) \quad \mathbf{S} = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix}.$$

Imponiamo che la vecchia e la nuova base siano concordi, ovvero $\det(\mathbf{S}) = s_{11}s_{22} + s_{12}s_{21} = 1$ e, usando questo fatto, la condizione $\mathbf{S}^T = \mathbf{S}^{-1}$, che si scrive

$$(1.95) \quad \begin{pmatrix} s_{11} & s_{21} \\ s_{12} & s_{22} \end{pmatrix} = \begin{pmatrix} s_{22} & -s_{12} \\ -s_{21} & s_{11} \end{pmatrix}.$$

Questo significa che la matrice è in realtà parametrizzata da due quantità, siano a e b , tali per cui

$$(1.96) \quad \mathbf{S} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

col vincolo aggiuntivo che $a^2 + b^2 = 1$, che le rende dipendenti. Possiamo cioè parametrizzare a e b in termini di *una variabile angolare*, scrivendo $a = \cos \theta$ e $b = \sin \theta$, in modo che

$$(1.97) \quad \mathbf{S} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$$

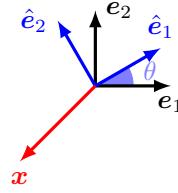


FIGURA 2. Due riferimenti cartesiani nel piano che differiscono di una rotazione di $\theta = \pi/6$: il vettore x , che nel primo è rappresentato da $\mathbf{x} = (-1, -1)^\top$, nel secondo è rappresentato da $\hat{\mathbf{x}}_1 = -\frac{1}{2}(\sqrt{3} + 1, \sqrt{3} - 1)^\top$: naturalmente, il vettore x non è cambiato, ma la nostra rappresentazione di x sì.

Se usiamo il fatto che uno spazio bidimensionale è isomorfo a \mathbb{R}^2 , e rappresentiamo i vettori $\mathbf{v} \in \mathbb{V}$ come frecce nel piano con coordinate date da \mathbf{v} in accordo con la vecchia base, la ragione di questa forma può risultare più chiara. L'unica maniera per cambiare un sistema ortonormale nel piano preservandone la positività (e l'origine) è, semplicemente, tramite una *rotazione*, come in Fig. 2. La matrice \mathbf{S} ci dice che, dato un vettore \mathbf{x} con coordinate $\mathbf{x} = (x_1, x_2)^\top$, questo diventerà secondo la nuova base

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \hat{\mathbf{x}} = \mathbf{S}^\top \mathbf{x} = \begin{pmatrix} x_1 \cos \theta + x_2 \sin \theta \\ x_2 \cos \theta - x_1 \sin \theta \end{pmatrix}$$

Le matrici ortogonali positive 2×2 descrivono quindi *rotazioni in senso antiorario* di un angolo θ : $\cos \theta$ è infatti proprio il coseno dell'angolo tra e_1 e \hat{e}_1 .

5.5. Prodotti vettoriali. Se \mathbb{V} è uno spazio prehilbertiano su \mathbb{R} e $\dim \mathbb{V} = 3$, è possibile definire una ulteriore operazione tra due vettori $\mathbf{v}, \mathbf{u} \in \mathbb{V}$. Supponiamo di avere una base ortonormale $\mathcal{B}_e = \{e_1, e_2, e_3\}$ e che

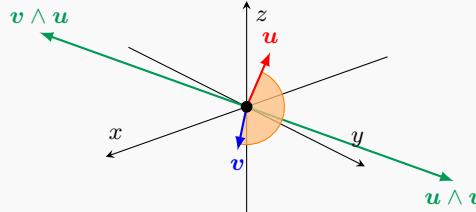
$$\mathbf{v} = v_1 e_1 + v_2 e_2 + v_3 e_3, \quad \mathbf{u} = u_1 e_1 + u_2 e_2 + u_3 e_3.$$

Considerando la sequenza (e_1, e_2, e_3) come *ordinata* (ovvero fissando l'ordine dei tre vettori della base), diamo la seguente definizione.

DEFINIZIONE 5.4 (Prodotto vettoriale). *Il prodotto vettoriale $\mathbf{v} \wedge \mathbf{u}$ è definito come*

$$(1.98) \quad \mathbf{v} \wedge \mathbf{u} = (v_2 u_3 - v_3 u_2) e_1 + (v_3 u_1 - v_1 u_3) e_2 + (v_1 u_2 - v_2 u_1) e_3.$$

Per esempio, sia $\mathbb{V} = \mathbb{R}^3$ e scegliendo come base la base canonica $\mathcal{B}_e = \{e_i\}_{i=1}^3$, $e_i = (\delta_{ij})_j \in \mathbb{R}^3$. Dati \mathbf{v} e \mathbf{u} due vettori in \mathbb{R}^3 , possiamo costruire due prodotti vettoriali, $\mathbf{v} \wedge \mathbf{u}$ e $\mathbf{u} \wedge \mathbf{v}$, entrambi applicati nell'origine, e rappresentarli come segue



La formula data nella definizione può essere ricordata nel seguente modo. Consideriamo la matrice

$$(1.99) \quad \begin{pmatrix} e_1 & e_2 & e_3 \\ v_1 & v_2 & v_3 \\ u_1 & u_2 & u_3 \end{pmatrix}.$$

Questa è, naturalmente, una matrice molto bizzarra i cui elementi sono in parte scalari e in parte vettori. Tuttavia quella che stiamo per dare è una regola *mnemonica*: possiamo *formalmente* scrivere che

$$(1.100) \quad \mathbf{v} \wedge \mathbf{u} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ v_1 & v_2 & v_3 \\ u_1 & u_2 & u_3 \end{vmatrix} = \begin{vmatrix} v_2 & v_3 \\ u_2 & u_3 \end{vmatrix} \mathbf{e}_1 - \begin{vmatrix} v_1 & v_3 \\ u_1 & u_3 \end{vmatrix} \mathbf{e}_2 + \begin{vmatrix} v_1 & v_2 \\ u_1 & u_2 \end{vmatrix} \mathbf{e}_3.$$

Il prodotto vettoriale ha una serie di interessanti proprietà, date in questo Teorema che non dimostriamo.

TEOREMA 5.5 (Proprietà del prodotto vettoriale). *Dati tre vettori $\mathbf{v}, \mathbf{u}, \mathbf{w} \in \mathbb{V}$ e uno scalare $c \in \mathbb{R}$, valgono le seguenti proprietà:*

- (1) $\mathbf{v} \wedge \mathbf{u} = -\mathbf{u} \wedge \mathbf{v}$;
- (2) $\mathbf{v} \wedge (\mathbf{u} + \mathbf{w}) = \mathbf{v} \wedge \mathbf{u} + \mathbf{v} \wedge \mathbf{w}$;
- (3) $c(\mathbf{v} \wedge \mathbf{u}) = (cv) \wedge \mathbf{u} = \mathbf{v} \wedge (cu)$;
- (4) $\langle \mathbf{v}, \mathbf{v} \wedge \mathbf{u} \rangle = 0$ e $\langle \mathbf{u}, \mathbf{v} \wedge \mathbf{u} \rangle = 0$;
- (5) $\|\mathbf{v} \wedge \mathbf{u}\|^2 = \|\mathbf{v}\|^2 \|\mathbf{u}\|^2 - \langle \mathbf{v}, \mathbf{u} \rangle^2$;
- (6) $\mathbf{v} \wedge \mathbf{u} = \mathbf{0}$ se e solo se \mathbf{v} e \mathbf{u} sono paralleli;
- (7) Se \mathbf{v} e \mathbf{u} sono linearmente indipendenti, ogni vettore ortogonale sia a \mathbf{v} che a \mathbf{u} è parallelo a $\mathbf{v} \wedge \mathbf{u}$. Inoltre, $\{\mathbf{v}, \mathbf{u}, \mathbf{v} \wedge \mathbf{u}\}$ è una base di \mathbb{V} concorde con la base ortonormale scelta \mathcal{B}_e .

La proprietà (5) in particolare può essere semplificata utilizzando l'angolo θ (come definito sopra) tra \mathbf{v} e \mathbf{u} . Infatti

$$\|\mathbf{v} \wedge \mathbf{u}\|^2 = \|\mathbf{v}\|^2 \|\mathbf{u}\|^2 - \langle \mathbf{v}, \mathbf{u} \rangle^2 = \|\mathbf{v}\|^2 \|\mathbf{u}\|^2 - \|\mathbf{v}\|^2 \|\mathbf{u}\|^2 \cos^2 \theta = \|\mathbf{v}\|^2 \|\mathbf{u}\|^2 \sin^2 \theta$$

ovvero la norma del prodotto vettoriale è

$$\|\mathbf{v} \wedge \mathbf{u}\| = \|\mathbf{v}\| \|\mathbf{u}\| |\sin \theta|.$$

Inoltre sappiamo, dalla proprietà (4), che $\mathbf{v} \wedge \mathbf{u}$ è ortogonale sia a \mathbf{v} che a \mathbf{u} . È facile vedere (per esempio, usando la regola mnemonica sopra) che se consideriamo precisamente la base ortonormale \mathcal{B}_e , allora

$$\mathbf{e}_1 \wedge \mathbf{e}_2 = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = \mathbf{e}_3$$

Possiamo scrivere una espressione simile per ogni altra coppia di vettori di base e in particolare vale la seguente relazione:

$$\mathbf{e}_i \wedge \mathbf{e}_j = \varepsilon_{ijk} \mathbf{e}_k$$

dove il simbolo ε_{ijk} è detto *di Levi–Civita* e vale

$$(1.101) \quad \varepsilon_{ijk} = \begin{cases} 1 & \text{se } (i, j, k) \in \{(1, 2, 3), (2, 3, 1), (3, 1, 2)\} \\ -1 & \text{se } (i, j, k) \in \{(2, 1, 3), (1, 3, 2), (3, 2, 1)\} \\ 0 & \text{diversamente.} \end{cases}$$

Il simbolo di Levi–Civita permette di scrivere in forma compatta i prodotti vettoriali tra due generici vettori, scrivendo

$$\mathbf{v} \wedge \mathbf{u} = \sum_{ijk} \varepsilon_{ijk} \mathbf{e}_i v_j u_k.$$

5.6. Prodotti misti. Avendo introdotto sia il prodotto scalare che il prodotto vettoriale, possiamo considerare espressioni in cui *entrambi* questi prodotti appaiono, ovvero espressioni del tipo

$$\langle \mathbf{v} \wedge \mathbf{u}, \mathbf{w} \rangle$$

per $\mathbf{v}, \mathbf{u}, \mathbf{w} \in \mathbb{V}$. Questa quantità può essere calcolata direttamente, per esempio usando l'espressione appena ottenuta con il simbolo di Levi-Civita. Se $\mathbf{w} = w_1 \mathbf{e}_1 + w_2 \mathbf{e}_2 + w_3 \mathbf{e}_3$, allora, essendo $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij}$,

$$\langle \mathbf{v} \wedge \mathbf{u}, \mathbf{w} \rangle = \left\langle \sum_{ijk} \varepsilon_{ijk} \mathbf{e}_i v_j u_k, \sum_r w_r \mathbf{e}_r \right\rangle = \sum_{ijk} \varepsilon_{ijk} w_r v_j u_k \langle \mathbf{e}_i, \mathbf{e}_r \rangle = \sum_{ijk} \varepsilon_{ijk} w_i v_j u_k = \begin{vmatrix} w_1 & w_2 & w_3 \\ v_1 & v_2 & v_3 \\ u_1 & u_2 & u_3 \end{vmatrix}.$$

L'espressione in termini di un determinante implica una utile identità, ovvero

$$(1.102) \quad \langle \mathbf{v} \wedge \mathbf{u}, \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{u} \wedge \mathbf{w} \rangle.$$

Sempre l'espressione in termini del determinante ci indica che se $\langle \mathbf{v} \wedge \mathbf{u}, \mathbf{w} \rangle = 0$, allora \mathbf{v} , \mathbf{u} e \mathbf{w} devono essere linearmente dipendenti, ovvero complanari.

6. Spazi affini e spazi euclidei

6.1. Spazi affini. L'analisi di questo capitolo è partita dai vettori liberi nello spazio euclideo. Abbiamo quindi visto che tali vettori liberi hanno una loro algebra che può essere astratta fino al concetto, molto generale, di spazio vettoriale, e abbiamo studiato le sue proprietà. Viene quindi da chiedersi come in generale il concetto di spazio vettoriale possa essere "trasferito" su un insieme in cui tale struttura va applicata (per esempio, lo spazio euclideo). Ovvero, ci si chiede come *partire* da un \mathbb{K} -spazio vettoriale \mathbb{V} per trasferire la sua struttura ad un generico *insieme di punti* \mathbb{A} . Lo spazio \mathbb{A} non è necessariamente l'usuale spazio tridimensionale (ma può esserlo): ciò che importa è che si possa legare a \mathbb{V} tramite una opportuna mappa, che permetterà di studiare le relazioni tra i punti di \mathbb{A} usando operazioni tra vettori di \mathbb{V} . Quando questo è possibile, si dice che \mathbb{A} è uno *spazio affine*.

DEFINIZIONE 6.1 (Spazio affine). *Una legge $\phi: \mathbb{A} \times \mathbb{A} \rightarrow \mathbb{V}$ definisce uno spazio affine su \mathbb{A} se per ogni punto $P \in \mathbb{A}$ ed ogni $\mathbf{v} \in \mathbb{V}$ esiste un unico punto $Q \in \mathbb{A}$ tale per cui $\phi(P, Q) = \mathbf{v}$ e inoltre dati tre punti $P, Q, R \in \mathbb{A}$*

$$(1.103) \quad \phi(P, Q) + \phi(Q, R) = \phi(P, R).$$

Uno spazio affine è quindi la coppia (\mathbb{A}, ϕ) . Si dice allora che \mathbb{V} è lo spazio vettoriale dei vettori liberi di \mathbb{A} . Se \mathbb{V} ha dimensione finita, allora si dice che la dimensione di \mathbb{A} su \mathbb{K} è $\dim \mathbb{V}$.

Come detto, non è importante ciò che \mathbb{A} sia *esattamente*: semplicemente, chiediamo che ci sia una corrispondenza tra coppie di elementi di \mathbb{A} e vettori di \mathbb{V} e, in più, questa corrispondenza sia tale che valga l'Eq. (1.103), detta *relazione di Chasles*³. La relazione di Chasles implica anche, scegliendo $P = Q = R$

$$\phi(P, P) = \phi(P, P) + \phi(P, P) \Rightarrow \phi(P, P) = \mathbf{0}$$

e similmente, se $P = R \neq Q$

$$\mathbf{0} = \phi(P, P) = \phi(P, Q) + \phi(Q, P) \Rightarrow \phi(Q, P) = -\phi(P, Q).$$

³L'applicazione ϕ viene a volte scritta come $\phi(P, Q) \equiv \overrightarrow{PQ} \in \mathbb{V}$, proprio in analogia con la notazione che abbiamo usato per introdurre i vettori applicati nello spazio usuale: tuttavia qui userò ϕ per l'applicazione, e \overrightarrow{PQ} per indicare precisamente il vettore applicato P con destinazione Q quando questi sono punti in uno spazio geometrico usuale

Tutto questo è consistente con quel che abbiamo visto riguardo i vettori geometrici nello spazio usuale, in cui il vettore era una freccia tra due punti.

Se scegliamo un'origine, ovvero un punto speciale $O \in \mathbb{A}$, possiamo associare *biunivocamente* ad ogni punto $P \in \mathbb{A}$ uno e un solo vettore $\mathbf{p} \in \mathbb{V}$: il vettore \mathbf{p} è il vettore da applicare a O per "spostarlo" in P , ovvero

$$\mathbf{p} = \phi(O, P)$$

creando una *corrispondenza esatta* tra lo spazio \mathbb{A} ("insieme di punti") e lo spazio vettoriale \mathbb{V} ("insieme degli spostamenti"). Se indichiamo $\phi(O, P) = \mathbf{p}$ e $\phi(O, Q) = \mathbf{q}$, possiamo inoltre scrivere

$$\phi(P, Q) = \phi(P, O) + \phi(O, Q) = \phi(O, Q) - \phi(O, P) = \mathbf{q} - \mathbf{p}.$$

Sia ora $\dim \mathbb{V} = d$ e scegliamo una base $\mathcal{B}_b := \{\mathbf{b}_i\}_{i=1}^d$ di \mathbb{V} : possiamo dire che $\{O, \mathcal{B}_b\}$ è un *riferimento affine* di \mathbb{A} . Grazie a questo riferimento, possiamo associare un punto $P \in \mathbb{A}$ ad un vettore $\mathbf{p} \in \mathbb{V}$ e quindi ad un insieme di coordinate $\mathbf{p} = (p_1, \dots, p_d)^\top$ secondo questa base, creando la corrispondenza

$$(1.104) \quad P \in \mathbb{A} \xleftarrow{\text{fissando } \phi \text{ e un'origine } O} \mathbf{p} \in \mathbb{V} \xleftarrow{\text{fissando una base } \mathcal{B}_b} \mathbf{p} = \begin{pmatrix} p_1 \\ \vdots \\ p_d \end{pmatrix} \in \mathbb{K}^d.$$

In questo modo abbiamo costruito un "ponte" tra \mathbb{A} e \mathbb{K}^d . È molto importante ricordare, ancora una volta, che i valori delle coordinate \mathbf{p} del punto P dipendono dalla base che abbiamo scelto. Riassumendo, data un'origine $O \in \mathbb{A}$, ogni punto P in \mathbb{A} può essere associato ad un vettore $\mathbf{p} \in \mathbb{V}$ che è rappresentabile a sua volta con un vettore di coordinate $\mathbf{p} \in \mathbb{K}^d$ secondo una data base \mathcal{B}_b . A riferimento fissato, possiamo identificare i punti di \mathbb{A} con le loro coordinate in \mathbb{K}^n .

Dato un \mathbb{K} -spazio vettoriale \mathbb{V} , si può sempre costruire uno spazio affine usando \mathbb{V} stesso come insieme, ovvero $\mathbb{A} \equiv \mathbb{V}$. In questo caso, una applicazione $\phi: \mathbb{V} \times \mathbb{V} \rightarrow \mathbb{V}$ con le proprietà giuste è semplicemente

$$\phi(\mathbf{v}, \mathbf{u}) = \mathbf{u} - \mathbf{v}.$$

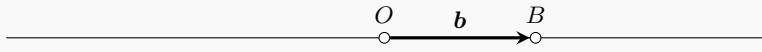
L'*n-spazio affine numerico* è un caso particolare: è costruito su $\mathbb{V} = \mathbb{K}^n$ si denota come $\mathbb{A}_{\mathbb{K}}^n$, ed è uno spazio affine di dimensione n in cui ciascun punto è associato ad un vettore colonna di n elementi di \mathbb{K} . Questo spazio ha un riferimento affine standard che abbiamo già incontrato quando abbiamo discusso \mathbb{K}^n nella veste di spazio vettoriale, ovvero una coppia origine-base $\{O, \mathcal{B}_e\}$ per cui

$$(1.105) \quad O \mapsto \mathbf{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

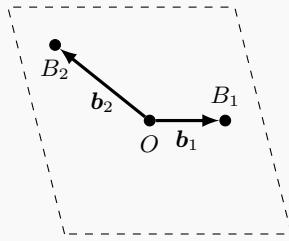
e $\mathcal{B}_e = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ tali per cui

$$(1.106) \quad \mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots \quad \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

La base dello spazio vettoriale permette di definire una (convenzionale) *orientazione* di uno spazio affine. Per esempio, nel caso della *retta affine orientata*, abbiamo uno spazio affine \mathbb{E}_1 che è associato ad uno spazio vettoriale \mathbb{V} unidimensionale. Dato un punto origine $O \in \mathbb{E}_1$ e una base $\mathcal{B}_b = \{\mathbf{b}\}$, il *verso* della retta corrisponde al verso indotto dal vettore di base \mathbf{b} . Ovvero, occorre trovare il punto B sulla retta tale che $\phi(O, B) = \mathbf{b}$: la direzione positiva sarà quella del segmento orientato \overrightarrow{OB} :



Il piano affine \mathbb{E}_2 è dato, per definizione, su uno spazio vettoriale che è bidimensionale reale \mathbb{V} . Fissato un punto origine $O \in \mathbb{E}_2$ e una base $\mathcal{B}_b = \{\mathbf{b}_1, \mathbf{b}_2\}$, questa identifierà due punti nel piano, B_1 e B_2 , tali per cui $\phi(O, B_1) = \mathbf{b}_1$ e $\phi(O, B_2) = \mathbf{b}_2$.



Anche in questo caso, la base induce una *orientazione* convenzionale che può essere definita una volta scelto un ordine all'interno della coppia $(\mathbf{b}_1, \mathbf{b}_2)$: il verso di rotazione del piano è quello che permette di sovrapporre il segmento orientato $\overrightarrow{OB_1}$ al segmento orientato $\overrightarrow{OB_2}$ tramite una rotazione attorno ad O che descriva un angolo convesso: se tale rotazione è in senso antiorario, si dice che l'orientazione è *positiva*. Nell'esempio nella figura sopra, per esempio, l'orientazione è positiva.

Dato che esiste una corrispondenza tra spazio \mathbb{V} e spazio \mathbb{A} , forse non sorprende che si può trovare una corrispondenza tra un sottospazio vettoriale \mathbb{W} di \mathbb{V} e un *sottospazio affine* di \mathbb{A} .

DEFINIZIONE 6.2 (Sottospazio affine). *Sia \mathbb{V} un \mathbb{K} -spazio vettoriale e sia \mathbb{A} spazio affine su di esso. Detto $Q \in \mathbb{A}$, e dato un sottospazio vettoriale \mathbb{W} di \mathbb{V} , il sottospazio affine \mathbb{S} passante per Q e parallelo a \mathbb{W} è dato da tutti i punti P di \mathbb{A} tali per cui $\phi(Q, P) \in \mathbb{W}$. Si dice allora che \mathbb{W} è la giacitura di \mathbb{S} .*

Se $\dim \mathbb{W} = 1$ si dice che il sottospazio \mathbb{S} è una retta, se $\dim(\mathbb{W}) = 2$ si dice che il sottospazio \mathbb{S} è un piano. Intuitivamente, un sottospazio affine corrisponde a tutti i punti dello spazio \mathbb{A} che si possono ottenere partendo da Q e “traslandarlo” secondo tutti i vettori di \mathbb{W} . Si può in effetti dimostrare che quale punto Q di \mathbb{S} si sia scelto non è importante: un sottospazio affine \mathbb{S} è definito dalla sua giacitura a da un qualsiasi suo punto. Una maniera differente, ma equivalente, di dare un sottospazio affine è quello di dare un set di $m + 1$ punti $Q, P_1, \dots, P_m \in \mathbb{A}$, e considerare i vettori $\phi(Q, P_i)$ con $i = 1, \dots, m$, e usare $\mathbb{W} = \text{Span}[\{\phi(Q, P_i)\}_{i=1}^m]$ per definire il sottospazio affine passante per Q con giacitura \mathbb{W} .

6.2. Spazi euclidei. A questo punto è chiaro che la funzione ϕ ideale per dare struttura affine allo spazio euclideo è proprio quella che abbiamo descritto a inizio capitolo, ovvero la mappa che associa due punti ad un vettore che va da un punto all'altro. Supponiamo quindi di avere lo spazio usuale in d dimensioni \mathbb{E}_d : associamo ogni coppia di punti P, Q ad un segmento orientato che li congiunge, con coda in P e testa in Q , e sia \mathbb{V} l'insieme di tali segmenti orientati. Se poi scegliamo un'origine O , allora possiamo associare univocamente ogni punto $X \in \mathbb{E}_d$ ad un segmento orientato $\mathbf{x} \in \mathbb{V}$, come abbiamo fatto in generalità sopra: in altre parole, con riferimento alla notazione di inizio capitolo, $\mathbf{x} = \overrightarrow{OX}$. Abbiamo visto, nella prima sezione, che \mathbb{V} può essere dotato della struttura di *spazio vettoriale reale*: possiamo definire somme tra questi vettori, e prodotto per uno scalare. In effetti, tutta la teoria che abbiamo sviluppato è *ispirata* da questo caso puramente geometrico: di conseguenza, lo spazio euclideo è uno spazio affine con $\phi(P, Q) = \overrightarrow{PQ}$.

Possiamo però aggiungere una struttura ulteriore. Se $\mathbf{x}, \mathbf{y} \in \mathbb{V}$ sono rappresentanti dei punti $X, Y \in \mathbb{E}_d$, indichiamo con $\|\mathbf{x}\|$ e $\|\mathbf{y}\|$ la distanza di X e Y dall'origine rispettivamente, e con θ l'angolo convesso \widehat{XOY} . Possiamo *definire* un “prodotto” tra i due vettori \mathbf{x} e \mathbf{y} in termini di queste quantità misurate in \mathbb{E}_d , ovvero

$$(1.107) \quad \mathbf{x} \cdot \mathbf{y} := \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta.$$

Questo prodotto moltiplica la lunghezza di un vettore per la proiezione dell'altro lungo la stessa direzione: $\|\mathbf{x}\| \cos \theta$ può essere infatti interpretato come la proiezione di \overrightarrow{OX} lungo \overrightarrow{OY} , o

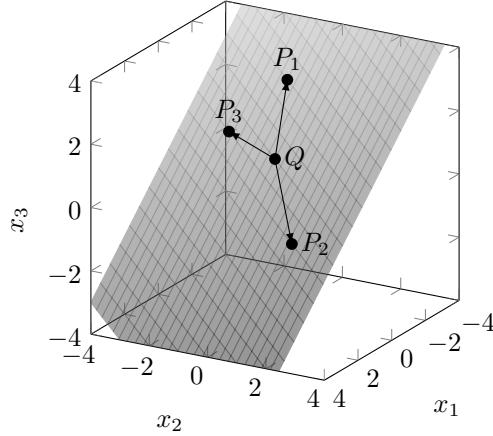
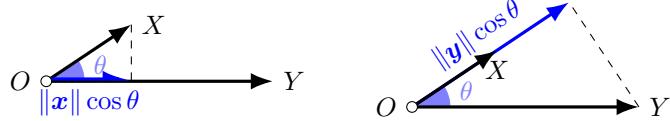
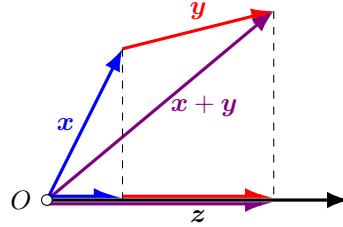


FIGURA 3. Spazio affine tridimensionale $\mathbb{A}_{\mathbb{R}}^n \equiv \mathbb{R}^3$ con sottospazio bidimensionale (piano) la cui giacitura $\mathbb{W} = \text{Span}[\{\phi(Q, P_1), \phi(Q, P_2), \phi(Q, P_3)\}]$ è generata da quattro punti $\{Q, P_1, P_2, P_3\}$. Ciascun vettore $\phi(Q, P_i)$ è associato ad un segmento orientato $\overrightarrow{QP_i}$: come si vede, i quattro punti sono complanari (ma non allineati), ovvero i tre segmenti giacciono nello stesso piano, e quindi \mathbb{W} è bidimensionale.

similmente $\|\mathbf{y}\| \cos \theta$ può essere interpretato come la proiezione di \overrightarrow{OY} lungo \overrightarrow{OX}



L'operazione introdotta è una operazione *bilineare simmetrica positiva*. Se $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{V}$, allora $(\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z}$, come esemplificato dalla seguente costruzione geometrica:



Usando la direzione di \mathbf{z} come riferimento, proiettare lungo \mathbf{z} la somma $\mathbf{x} + \mathbf{y}$ fornisce la stessa quantità ottenuta proiettando separatamente \mathbf{x} e \mathbf{y} e quindi sommando il risultato. Così si può vedere anche che $(c\mathbf{x}) \cdot \mathbf{y} = c(\mathbf{x} \cdot \mathbf{y})$ per $c \in \mathbb{R}$. Inoltre $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$, e infine

$$\mathbf{x} \cdot \mathbf{x} = \|\mathbf{x}\|^2 \geq 0$$

e in particolare, l'uguaglianza vale se e solo se $X = O$. In altre parole, il prodotto che abbiamo definito ha tutte le proprietà di un prodotto scalare: per esso viene perciò spesso usata la notazione che abbiamo già introdotto, $\mathbf{x} \cdot \mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle$. Esso è il prodotto scalare standard di vettori su \mathbb{E}_d che rende lo spazio \mathbb{V} prehilbertiano.

Data l'origine O , in uno spazio d dimensionale \mathbb{E}_d possiamo individuare d direzioni *ortogonal**i*. Possiamo quindi trovare al più d punti $E_1, E_2, \dots, E_d \in \mathbb{E}_d$ tali che $e_i = \overrightarrow{OE_i}$, $i = 1, \dots, d$ siano *ortonormali* tra loro. L'insieme $\mathcal{B}_e = \{e_1, \dots, e_d\}$ è quindi una base ortonormale secondo il prodotto scalare che abbiamo introdotto in Eq. (1.107), e questa fornisce, insieme ad un punto di origine O di riferimento, ad un *riferimento cartesiano* di versori lungo d direzioni perpendicolari (nel senso classico) nello spazio, che si indica talvolta con $Oe_1 \dots e_d$. Naturalmente, possono esistere al più d vettori ortonormali in \mathbb{E}_d , che perciò ha dimensione d . Secondo questo riferimento, ogni punto X è associato ad un vettore di coordinate in \mathbb{R}^d ,

$$X \in \mathbb{E}_d \leftrightarrow \mathbf{x} = x_1 e_1 + x_2 e_2 + \dots + x_d e_d \leftrightarrow \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix} \in \mathbb{R}^d.$$

La costruzione è tale che le lunghezze e gli angoli tra vettori in \mathbb{V} siano *esattamente* le lunghezze e gli angoli che conosciamo nello spazio usuale \mathbb{E}_d . Per costruzione, in questa base il prodotto scalare si può riscrivere come

$$(1.108) \quad \langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^d x_i y_i.$$

Questo significa anche che

$$(1.109) \quad \cos \theta = \frac{\sum_{i=1}^d x_i y_i}{\sqrt{\sum_{i=1}^d x_i^2} \sqrt{\sum_{i=1}^d y_i^2}}.$$

Inoltre, consistentemente con il teorema di Pitagora,

$$\|\mathbf{x}\|^2 = \sum_{i=1}^d x_i^2.$$

Quanto abbiamo detto significa che la linea, il piano e lo spazio possono essere visti come spazi affini su spazi prehilbertiani di dimensione $d = 1$, $d = 2$ e $d = 3$. In generale, uno spazio euclideo \mathbb{E}_d è uno spazio affine su uno spazio prehilbertiano di dimensione d : la struttura euclidea discende dalla definizione di prodotto scalare in Eq. (1.107), dato in termini puramente geometrici, che si può riscrivere come in Eq. (1.108) una volta ottenuta una rappresentazione in coordinate rispetto ad una data base ortonormale.

6.2.1. *Distanza e struttura metrica.* Avendo introdotto il concetto di prodotto scalare, che eredita ovviamente le proprietà generali discusse nel capitolo precedente, possiamo anche definire una quantità essenziale per introdurre una metrica nel nostro spazio.

DEFINIZIONE 6.3 (Distanza). *Siano $X, Y \in \mathbb{E}_d$. La distanza tra X e Y è definita come*

$$(1.110) \quad d(X, Y) := \|\phi(X, Y)\| = \|\mathbf{x} - \mathbf{y}\|.$$

Se abbiamo un sistema cartesiano $Oe_1 \dots e_d$, e $X \mapsto \mathbf{x} = (x_i)_{i=1}^d$ e $Y \mapsto \mathbf{y} = (y_i)_{i=1}^d$, vettori di coordinate dei vettori associati $\mathbf{x}, \mathbf{y} \in \mathbb{V}$ rispettivamente, allora

$$d(X, Y) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{i=1}^d (x_i - y_i)^2}.$$

Grazie a questa definizione, possiamo introdurre per esempio il *punto medio* M tra X e Y , come il punto di coordinate $\mathbf{m} = \frac{\mathbf{x} + \mathbf{y}}{2}$, dato che in questo modo $d(X, M) = d(M, Y)$.

La distanza gode di alcune proprietà che eredita da quelle di prodotto scalare: dati due punti $P, Q \in \mathbb{E}_d$, $d(P, Q)$ è sempre non-negativa (e pari a zero se e solo se $P = Q$), è simmetrica ($d(P, Q) = d(Q, P)$) e vale la diseguaglianza triangolare $d(P, Q) \leq d(P, R) + d(R, Q)$ per tre

punti $P, Q, R \in \mathbb{E}_d$. Queste tre proprietà, note dalla geometria euclidea, vengono utilizzate come assiomi fondamentali per definire il concetto di *spazio metrico*.

7. Geometria analitica nel piano

Consideriamo, per cominciare, la geometria del *piano*, ovvero dello spazio affine \mathbb{E}_2 su uno spazio vettoriale prehilbertiano bidimensionale \mathbb{V} . Assumeremo di aver introdotto un sistema ortonormale *positivamente orientato* $O\mathbf{e}_1\mathbf{e}_2$. Si noti che, in molti casi, si usa una notazione speciale per tale base nel piano, ovvero $\mathbf{e}_1 \equiv \mathbf{i}$ e $\mathbf{e}_2 \equiv \mathbf{j}$. Ogni punto $X \in \mathbb{E}_2$ è associato ad un vettore $\mathbf{x} = \phi(O, X)$ di uno spazio bidimensionale le cui coordinate, rispetto al sistema ortonormale menzionato, sono espresse da una coppia $\mathbf{x} \in \mathbb{R}^2$. Indicheremo questa associazione tra punto X e vettore \mathbf{x} con $X(\mathbf{x})$.

7.1. Vettori ortogonali. Cominciamo con una breve osservazione. Supponiamo di avere il punto $P(\mathbf{p})$ e di volere trovare $Q(\mathbf{q})$ tale per cui \mathbf{q} sia ortogonale a \mathbf{p} . Questo significa che $\langle \mathbf{q}, \mathbf{p} \rangle = 0$, ovvero, in coordinate, $q_1 p_1 + q_2 p_2 = 0$, il che implica che

$$(1.111) \quad \frac{q_1}{q_2} = -\frac{p_2}{p_1}.$$

Come si vede abbiamo ottenuto una informazione sul rapporto fra le due componenti (imporre l'ortogonalità non è infatti sufficiente a trovare ogni informazione su un vettore, ma solo la sua direzione) e abbiamo una infinità di soluzioni tra cui scegliere. Una di queste può essere per esempio il vettore \mathbf{q} con coordinate $\mathbf{q} = (-p_2, p_1)^\top$.

7.2. La retta. Dati due punti $P(\mathbf{p})$ e $Q(\mathbf{q})$ del piano, un punto $X(\mathbf{x})$ appartiene alla retta \mathcal{R} passante per essi se $\mathbf{x} - \mathbf{p}$ e $\mathbf{q} - \mathbf{p}$ sono paralleli, ovvero *linearmente dipendenti*. Ciò significa che, in coordinate,

$$(1.112) \quad \begin{vmatrix} x_1 - p_1 & x_2 - p_2 \\ q_1 - p_1 & q_2 - p_2 \end{vmatrix} = 0.$$

Sviluppando il determinante otteniamo l'equazione cartesiana della retta, ovvero

$$(1.113) \quad v_1 x_1 + v_2 x_2 + \gamma = 0, \quad v_1 = q_2 - p_2, \quad v_2 = p_1 - q_1, \quad \gamma = p_2 q_1 - p_1 q_2.$$

L'equazione esprime in coordinate la condizione

$$(1.114) \quad \langle \mathbf{v}, \mathbf{x} - \mathbf{p} \rangle = 0$$

dove \mathbf{v} è il vettore con coordinate $\mathbf{v} = (v_1, v_2)^\top$. L'equazione dice quindi che \mathbf{v} è un vettore ortogonale alla retta, ed infatti \mathbf{v} è un vettore ortogonale a $\mathbf{u} := \mathbf{q} - \mathbf{p}$, che esprime invece la direzione della retta essendo orientato come il segmento \overline{PQ} .

L'equazione della retta può essere data anche in forma *parametrica*, ovvero richiedendo che un vettore $\mathbf{x} - \mathbf{p}$ sia *proporzionale* a $\mathbf{q} - \mathbf{p}$,

$$\mathbf{x} - \mathbf{p} = t(\mathbf{q} - \mathbf{p}).$$

Ciò è equivalente a scrivere

$$(1.115) \quad \begin{cases} x_1 - p_1 = t(q_1 - p_1) \\ x_2 - p_2 = t(q_2 - p_2) \end{cases} \Rightarrow \begin{cases} x_1 = p_1 + t(q_1 - p_1) \\ x_2 = p_2 + t(q_2 - p_2) \end{cases} \Leftrightarrow \mathbf{x} = \mathbf{p} + t(\mathbf{q} - \mathbf{p}).$$

Una maniera ancora diversa di scrivere queste equazioni è porre

$$\frac{x_1 - p_1}{q_1 - p_1} = \frac{x_2 - p_2}{q_2 - p_2},$$

ovvero scrivere la retta in termini di rapporti uguali.

Infine, una retta può essere data anche per via non di due punti bensì di un punto, per esempio $P(\mathbf{p})$, e di una direzione, identificata da un vettore \mathbf{u} con rappresentazione in coordinate $\mathbf{u} = (u_1, u_2)^\top$. Assegnare la direzione \mathbf{u} è come assegnare un altro punto $Q(\mathbf{q})$ con $\mathbf{q} = \mathbf{p} + \mathbf{u}$ e l'equazione della retta è allora

$$\frac{x_1 - p_1}{u_1} = \frac{x_2 - p_2}{u_2}.$$

7.3. Intersezione tra rette. Date due rette,

$$\mathcal{R}: v_1x_1 + v_2x_2 + \gamma = 0 \quad \mathcal{R}': v'_1x_1 + v'_2x_2 + \gamma' = 0$$

si può studiare il luogo dei punti geometrici in comune alle due rette, $\mathcal{R} \cap \mathcal{R}'$, scrivendo il sistema di equazioni lineari

$$\begin{cases} v_1x_1 + v_2x_2 + \gamma = 0 \\ v'_1x_1 + v'_2x_2 + \gamma' = 0 \end{cases}$$

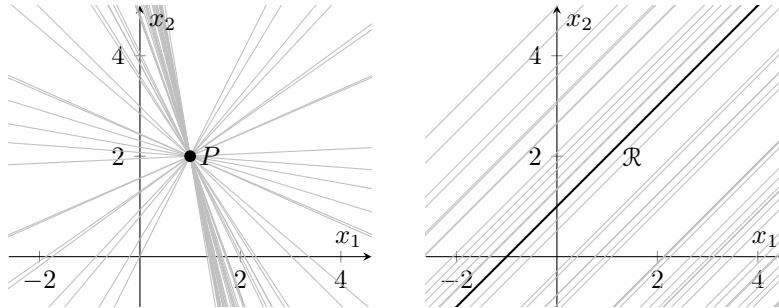
Sappiamo già che possono verificarsi vari casi a seconda del rango delle matrici

$$\mathbf{A} = \begin{pmatrix} v_1 & v_2 \\ v'_1 & v'_2 \end{pmatrix} \quad \mathbf{H} = \left(\begin{array}{cc|c} v_1 & v_2 & -\gamma \\ v'_1 & v'_2 & -\gamma' \end{array} \right)$$

Se $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{H}) = 2$, allora esiste un solo punto $X(\mathbf{x})$ comune alle due rette, che sono perciò *incidenti*, $\mathcal{R} \cap \mathcal{R}' = \{X\}$; se $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{H}) = 1$, invece, esistono infiniti punti in uno spazio unidimensionale e in effetti ciò significa che le due equazioni descrivono la stessa retta, $\mathcal{R} = \mathcal{R}'$; se $1 = \text{rank}(\mathbf{A}) \neq \text{rank}(\mathbf{H}) = 2$ il sistema non ha soluzione e le due rette non si incontrano mai, ovvero sono *parallele*, $\mathcal{R} \parallel \mathcal{R}'$. Questi sono i soli casi ammessi, per cui due rette nel piano sono coincidenti, incidenti o parallele.

7.3.1. Angolo tra due rette. Se due rette sono incidenti, può aver senso chiedersi quale sia il loro *angolo relativo*. Siano \mathbf{v} e \mathbf{v}' i corrispondenti vettori ortogonali. Essendo perpendicolari alle rispettive rette, il loro angolo relativo è uguale all'angolo tra le direzioni delle rette e possiamo scrivere $\theta = \arccos \frac{\langle \mathbf{v}, \mathbf{v}' \rangle}{\|\mathbf{v}\| \|\mathbf{v}'\|}$. In particolare, due rette sono perpendicolari se $\langle \mathbf{v}, \mathbf{v}' \rangle = 0$.

7.3.2. Fasci di rette. Dato un punto $P(\mathbf{p})$, l'insieme di tutte le rette del piano passanti per esso è detto *fascio proprio di rette* con *centro* P . Se invece consideriamo una retta \mathcal{R} , l'insieme delle rette ad essa parallela è detto *fascio improprio di rette* con *direzione* di \mathcal{R} .



Un fascio di rette proprio è in genere dato come combinazione lineare di due rette che vi appartengono. Se per esempio

$$\mathcal{R}: v_1x_1 + v_2x_2 + \gamma = 0 \quad \mathcal{R}': v'_1x_1 + v'_2x_2 + \gamma' = 0$$

sono due distinte rette del fascio, tutte le rette appartenenti ad esso possono essere ottenute utilizzando i parametri reali (λ, μ) nella combinazione lineare

$$(1.116) \quad \lambda(v_1x_1 + v_2x_2 + \gamma) + \mu(v'_1x_1 + v'_2x_2 + \gamma') = 0.$$

Il fascio è proprio o improprio se le rette sono incidenti (e quindi $\mathcal{R} \cap \mathcal{R}' = \{P\}$) o parallele rispettivamente. Assumendo $\lambda \neq 0$, possiamo introdurre $\rho = \mu/\lambda$ e riscrivere il fascio di rette utilizzando la cosiddetta *rappresentazione non omogenea*,

$$v_1x_1 + v_2x_2 + \gamma + \rho(v'_1x_1 + v'_2x_2 + \gamma') = 0.$$

Questo fascio contiene tutte le rette già contenute nell'espressione in Eq. (1.116), eccezion fatta per la retta \mathcal{R}' (dato che abbiamo assunto $\lambda \neq 0$). Questa rappresentazione rende però chiaro il fatto che un singolo parametro individua ogni membro del fascio. In particolare, se le due rette sono parallele, ovvero le equazioni sono nella forma $\mathcal{R}: \langle \mathbf{v}, \mathbf{x} - \mathbf{p} \rangle$ e $\mathcal{R}' \langle \mathbf{v}, \mathbf{x} - \mathbf{q} \rangle = 0$, il fascio può riscriversi come $\langle \mathbf{v}, \mathbf{x} \rangle + t = 0$, dove t è il parametro del fascio.

7.3.3. Distanze. Supponiamo di avere un punto $Q(\mathbf{q})$ e volerne calcolare la distanza dalla retta \mathcal{R} , individuata dalla direzione ortogonale \mathbf{v} e passante per il punto $P(\mathbf{p})$: la retta è quindi data da tutti gli \mathbf{x} tali per cui $\langle \mathbf{v}, \mathbf{x} - \mathbf{p} \rangle = 0$. Tale distanza è definita come la lunghezza del vettore $\mathbf{q}' - \mathbf{q}$, dove $Q'(\mathbf{q}') \in \mathcal{R}$ tale per cui $\mathbf{q}' - \mathbf{q}$ sia parallelo a \mathbf{v} , ovvero $\mathbf{q}' - \mathbf{q} = t\mathbf{v}$ per un qualche $t \in \mathbb{R}$, ma anche $\langle \mathbf{v}, \mathbf{q}' - \mathbf{p} \rangle = 0$. L'oggetto a cui siamo interessati è $d(Q, Q') = \|\mathbf{q}' - \mathbf{q}\|$. Per ottenerlo scriviamo

$$0 = t\langle \mathbf{v}, \mathbf{q}' - \mathbf{p} \rangle = \langle \mathbf{q}' - \mathbf{q}, \mathbf{q}' - \mathbf{q} + \mathbf{q} - \mathbf{p} \rangle = \|\mathbf{q}' - \mathbf{q}\|^2 + t\langle \mathbf{v}, \mathbf{q} - \mathbf{p} \rangle,$$

ovvero $\|\mathbf{q}' - \mathbf{q}\|^2 = -t\langle \mathbf{v}, \mathbf{q} - \mathbf{p} \rangle$. Possiamo ora rimuovere t osservando che, se $\mathbf{q}' - \mathbf{q} = t\mathbf{v}$, allora $\|\mathbf{q}' - \mathbf{q}\| = |t|\|\mathbf{v}\|$, e quindi $|t| = \frac{\|\mathbf{q}' - \mathbf{q}\|}{\|\mathbf{v}\|}$, quindi sostituendo questa espressione

$$\|\mathbf{q}' - \mathbf{q}\| = \text{sign}(t) \frac{\langle \mathbf{v}, \mathbf{q} - \mathbf{p} \rangle}{\|\mathbf{v}\|}$$

e, dato che tutte le quantità coinvolte sono positive eccetto il prodotto scalare, il segno di t deve essere tale che $t\langle \mathbf{v}, \mathbf{q} - \mathbf{p} \rangle > 0$, per cui possiamo scrivere in forma compatta

$$(1.117) \quad d(Q, \mathcal{R}) = d(Q, Q') = \|\mathbf{q}' - \mathbf{q}\| = \frac{|\langle \mathbf{v}, \mathbf{q} - \mathbf{p} \rangle|}{\|\mathbf{v}\|}.$$

Se passiamo ora in coordinate, ciò che appare al numeratore è esattamente l'equazione della retta calcolata in \mathbf{q} . Ovvero, se $\mathcal{R}: v_1x_1 + v_2x_2 + \gamma = 0$,

$$d(Q, \mathcal{R}) = \frac{|v_1q_1 + v_2q_2 + \gamma|}{\sqrt{v_1^2 + v_2^2}}.$$

Date due rette parallele, la loro distanza è data dalla distanza di un punto qualsiasi di una dall'altra.

7.4. Coniche nel piano. Se la retta è una funzione *lineare* della posizione \mathbf{x} dei suoi punti, del tipo $\langle \mathbf{v}, \mathbf{x} \rangle + \gamma = 0$, possiamo considerare il caso di domini determinati da equazioni quadratiche che, in un sistema cartesiano, assumono la forma

$$(1.118) \quad ax_1^2 + bx_1x_2 + cx_2^2 + 2w_1x_1 + 2w_2x_2 + \gamma = 0,$$

per una collezione di coefficienti reali (il fattore 2 compare per convenienza futura).

Questo tipo di equazioni si può scrivere sempre in termini di una forma bilineare simmetrica e un prodotto scalare, ovvero come

$$(1.119) \quad \varphi(\mathbf{x}, \mathbf{x}) + 2\langle \mathbf{w}, \mathbf{x} \rangle + \gamma = 0,$$

dove $\varphi: \mathbb{V} \times \mathbb{V} \rightarrow \mathbb{R}$ è una forma bilineare e $\mathbf{w} \in \mathbb{V}$ che, secondo il riferimento cartesiano Oe_1e_2 , ha matrice associata simmetrica

$$\Phi = \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}.$$

In termini di coordinate possiamo quindi riscrivere

$$(1.120) \quad \mathbf{x}^\top \Phi \mathbf{x} + 2\mathbf{w}^\top \mathbf{x} + \gamma = 0.$$

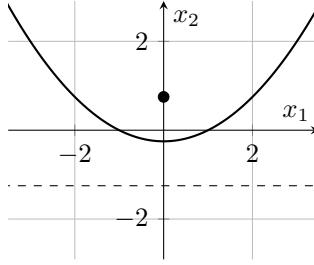


FIGURA 4. Parabola di equazione $x_2 = \frac{x_1^2 - 1}{4}$. Il punto indica la posizione del fuoco, mentre la direttrice è tratteggiata.

Le curve nel piano determinate da questa famiglia di equazioni sono dette *coniche*. Per studiarle, è utile eseguire un cambio di riferimento cartesiano. Essendo Φ simmetrica, esiste sempre una base ortonormale $\mathcal{B}_{\hat{\mathbf{e}}} = \{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2\}$ in cui questa forma può essere diagonalizzata: in altre parole, passando dal riferimento cartesiano $O\mathbf{e}_1\mathbf{e}_2$ al riferimento cartesiano $O\hat{\mathbf{e}}_1\hat{\mathbf{e}}_2$ con una opportuna matrice ortonormale \mathbf{S} , abbiamo che, chiamando $\hat{\mathbf{x}} = \mathbf{S}^\top \mathbf{x}$

$$\mathbf{x}^\top \Phi \mathbf{x} + 2\mathbf{w}^\top \mathbf{x} + \gamma = \mathbf{x}^\top \mathbf{S} \mathbf{S}^\top \Phi \mathbf{S} \mathbf{S}^\top \mathbf{x} + 2\mathbf{w}^\top \mathbf{S} \hat{\mathbf{x}} + \gamma = \hat{\mathbf{x}}^\top \hat{\Phi} \hat{\mathbf{x}} + 2\hat{\mathbf{w}}^\top \hat{\mathbf{x}} + \gamma = 0,$$

dove abbiamo introdotto $\hat{\mathbf{w}} = \mathbf{S}^\top \mathbf{w}$ e la rappresentazione di \mathbf{F} secondo la base diagonalizzante,

$$\hat{\Phi} := \mathbf{S}^\top \Phi \mathbf{S} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix},$$

con λ_1 e λ_2 sono gli autovalori della forma: questo vuol dire che l'equazione, in coordinate, ha ora la struttura

$$\lambda_1 \hat{x}_1^2 + \lambda_2 \hat{x}_2^2 + 2\hat{w}_1 \hat{x}_1 + 2\hat{w}_2 \hat{x}_2 + \gamma = 0.$$

A questo punto, possono verificarsi tre possibili casi a seconda del valore di $\det \Phi$.

7.4.1. *Il caso $\det \Phi = 0$ e la parabola.* Supponiamo che $\det \Phi = 0$, ovvero per esempio $\lambda_2 = 0$. Questo vuol dire che possiamo scrivere l'equazione per la conica in coordinate come

$$\lambda_1 \hat{x}_1^2 + 2\hat{w}_1 \hat{x}_1 + 2\hat{w}_2 \hat{x}_2 + \gamma = 0.$$

Se $\hat{w}_2 = 0$, si tratta di una equazione per x_1 , che quindi genera, al più, due rette (possibilmente coincidenti) nella forma $x_1 = x_0^\pm$ se $\hat{w}_1^2 - 4\lambda_1\gamma \geq 0$ (altrimenti, non corrisponde a nessun luogo del piano). Se $\hat{w}_2 \neq 0$, questa equazione può scriversi

$$\hat{x}_2 = -\frac{1}{2\hat{w}_2}(\lambda_1 \hat{x}_1^2 + 2\hat{w}_1 \hat{x}_1 + \gamma) = \alpha(\hat{x}_1 - p_1)^2 + p_2, \quad \alpha := -\frac{\lambda_1}{2\hat{w}_2}, \quad p_1 := \frac{\hat{w}_1}{\lambda_1}, \quad p_2 := \frac{\hat{w}_1^2 - \lambda_1\gamma}{2\lambda_1^2 \hat{w}_2}.$$

La curva descritta da una equazione siffatta è detta *parabola*.

Una parabola è infatti ottenuta come il luogo dei punti \mathbf{x} equidistanti da un *fuoco* prefissato \mathbf{f} e da una *direttrice* orizzontale descritta da un'equazione del tipo $\langle \hat{\mathbf{v}}, \mathbf{x} - \mathbf{q} \rangle = 0$, dove $\hat{\mathbf{v}}$ si assume essere un *versore*. In questo caso, sappiamo che deve valere l'equazione

$$\|\mathbf{x} - \mathbf{f}\|^2 = |\langle \hat{\mathbf{v}}, \mathbf{x} - \mathbf{q} \rangle|^2,$$

Un po' di algebra fa vedere che in effetti questa condizione può essere scritta come

$$\varphi(\mathbf{x}, \mathbf{x}) + 2\langle \mathbf{w}, \mathbf{x} \rangle + \gamma = 0$$

dove φ è una forma bilineare simmetrica e \mathbf{w} un vettore tali che

$$\varphi(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle - \langle \mathbf{x}, \hat{\mathbf{v}} \rangle \langle \hat{\mathbf{v}}, \mathbf{y} \rangle, \quad \mathbf{w} = \hat{\mathbf{v}} - \mathbf{f}, \quad \gamma = \|\mathbf{f}\|^2 - \langle \hat{\mathbf{v}}, \mathbf{p} \rangle^2.$$

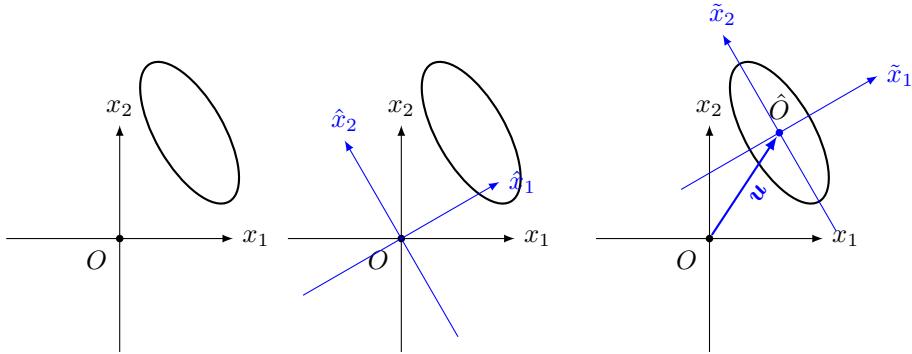


FIGURA 5. Esempio di cambio coordinate nel piano del tipo in Eq. (2.13) nel piano, consistente in una rotazione seguita da una traslazione. Ricordiamo che una matrice \mathbf{S} di cambio base in uno spazio bidimensionale può essere associata ad un angolo di rotazione θ : il cambio di coordinate permette di allineare gli assi della conica in analisi (nell'illustrazione, un'ellisse) con quelli di riferimento.

La forma bilinare, in particolare, ha determinante nullo: infatti, data una base ortonormale \mathcal{B}_e , la corrispondente matrice è

$$\Phi = \begin{pmatrix} 1 - \hat{v}_1^2 & \hat{v}_1 \hat{v}_2 \\ \hat{v}_1 \hat{v}_2 & 1 - \hat{v}_2^2 \end{pmatrix}$$

che, essendo $\|\hat{v}\|^2 = \hat{v}_1^2 + \hat{v}_2^2 = 1$, ha $\det \Phi = 0$. La parabola è una curva simmetrica rispetto al suo *asse*, costituito dalla retta passante per il fuoco e perpendicolare alla direttrice: l'intersezione tra tale retta e la curva stessa è detto *vertice* della parabola. Nel riferimento $O\hat{e}_1\hat{e}_2$ questo asse è semplicemente la retta $\hat{x}_1 = 0$.

7.4.2. Il caso $\det \Phi > 0$ e l'ellisse. Se $\det \Phi \neq 0$, possiamo semplificare ulteriormente l'equazione eseguendo uno shift, ovvero definendo $\hat{\mathbf{u}} = \hat{\Phi}^{-1}\hat{\mathbf{v}}$, corrispondente, secondo il riferimento $O\hat{e}_1\hat{e}_2$, ad un certo vettore \mathbf{u} nello spazio, e spostare l'origine del nostro riferimento cartesiano da O a $\hat{O}(\mathbf{u})$. Notare che, cambiando l'origine, i punti dello spazio saranno rappresentati da vettori *diversi*. Dovendo valere l'uguaglianza di Chasles, dato un punto X nello spazio, $\phi(O, X) = \phi(O, \hat{O}) + \phi(\hat{O}, X)$, ovvero, chiamando $\tilde{\mathbf{x}} = \phi(\hat{O}, X)$, $\mathbf{x} = \tilde{\mathbf{x}} + \mathbf{u}$. Grazie a questo shift, possiamo eliminare il termine lineare, con il cambio di variabili $\hat{\mathbf{x}} = \tilde{\mathbf{x}} + \hat{\mathbf{u}}$, ottenendo

$$\hat{\mathbf{x}}^\top \hat{\Phi} \hat{\mathbf{x}} + \tilde{\gamma} = 0,$$

dove $\tilde{\gamma} = \gamma - \hat{\mathbf{v}}^\top \Phi^{-1} \hat{\mathbf{v}} = \gamma - \mathbf{v}^\top \Phi^{-1} \mathbf{v}$. Avendo eseguito questo cambio di variabili, l'equazione assume una forma molto semplice, detta *forma standard*

$$(1.121) \quad \lambda_1 \tilde{x}_1^2 + \lambda_2 \tilde{x}_2^2 + \tilde{\gamma} = 0$$

che abbiamo ottenuto eseguendo una trasformazione di riferimento cartesiano da Oe_1e_2 a $\hat{O}\hat{e}_1\hat{e}_2$ costituita da *una traslazione dell'origine e una rotazione*,

$$(1.122) \quad \tilde{\mathbf{x}} = \mathbf{S}^\top (\mathbf{x} - \mathbf{v}).$$

Se ora $\det \Phi > 0$, questo significa che i due autovalori di Φ sono concordi, per esempio entrambi positivi (possiamo, in caso, moltiplicare per -1 l'equazione per rendere i due coefficienti positivi). Se $\tilde{\gamma} < 0$, allora l'equazione ammette delle soluzioni e possiamo scriverla come

$$(1.123) \quad \frac{\tilde{x}_1^2}{a^2} + \frac{\tilde{x}_2^2}{b^2} = 1, \quad a^2 := \frac{\tilde{\gamma}}{\lambda_1}, \quad b^2 := \frac{\tilde{\gamma}}{\lambda_2}.$$

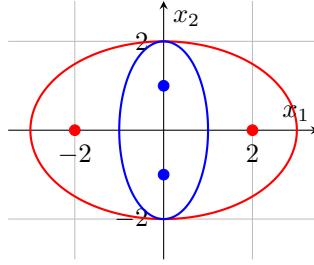


FIGURA 6. Ellissi con assi paralleli agli assi cartesiani e centrate nell'origine. In rosso, ellisse descritta dall'equazione $\frac{x_1^2}{3} + \frac{x_2^2}{2} = 1$; in blu, ellisse descritta dall'equazione $x_1^2 + \frac{x_2^2}{2} = 1$. I punti indicati sugli assi di corrispondente colore sono i rispettivi fuochi.

Questa è una equazione di un *ellisse*, ovvero il luogo dei punti geometrici tali per cui

$$\|\mathbf{x} - \mathbf{f}_1\| + \|\mathbf{x} - \mathbf{f}_2\| = 2 \max\{a, b\}$$

Qui \mathbf{f}_1 ha coordinate $\tilde{\mathbf{f}}_1 = (c, 0)$ e \mathbf{f}_2 ha coordinate $\tilde{\mathbf{f}}_2 = (-c, 0)$, con $c = \sqrt{a^2 - b^2}$, se $a > b$. Viceversa, \mathbf{f}_1 ha coordinate $\tilde{\mathbf{f}}_1 = (0, c)$ e \mathbf{f}_2 ha coordinate $\tilde{\mathbf{f}}_2 = (0, -c)$, con $c = \sqrt{b^2 - a^2}$, se $a < b$. L'*eccentricità* di una ellisse è data da

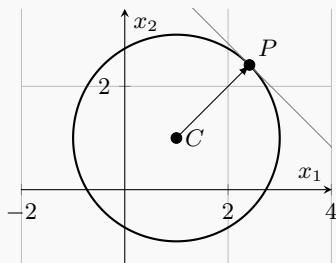
$$e = \frac{c}{\max\{a, b\}}.$$

La quantità $\max\{a, b\}$ è detta *semiasse maggiore*, mentre $\min\{a, b\}$ è il *semiasse minore*. Se $a = b$ abbiamo una circonferenza, con centro l'origine \hat{O} e raggio a . In questo riferimento, l'*asse principale dell'ellisse*, ovvero la retta su cui giacciono i suoi fuochi, coincide sempre con uno dei due assi cartesiani: l'altro asse è detto *asse secondario*. Se $\tilde{\gamma} > 0$, invece, si dice che l'equazione descrive una *ellisse immaginaria*, e non corrisponde ad alcun punto. Nel caso $\tilde{\gamma} = 0$, l'unica soluzione corrisponde al punto nell'origine \hat{O} .

Nel sistema di riferimento $O\hat{\mathbf{e}}_1\hat{\mathbf{e}}_2$, l'equazione ha la forma

$$\frac{(\hat{x}_1 - \hat{u}_1)^2}{a^2} + \frac{(\hat{x}_2 - \hat{u}_2)^2}{b^2} = 1.$$

che è una ellisse centrata in $\hat{O}(\mathbf{u})$ punto nel riferimento $O\hat{\mathbf{e}}_1\hat{\mathbf{e}}_2$ e con asse principale parallelo a uno dei due assi cartesiani. Se l'ellisse ha assi ruotati rispetto al riferimento, invece, compaiono termini misti nell'espressione cartesiana come in Eq. (1.118).



Sia \mathcal{C} una circonferenza di centro $C(\mathbf{c})$. Consideriamo un punto $P(\mathbf{p}) \in \mathcal{C}$ e tracciamone la tangente alla circonferenza: come scriverne l'equazione? Occorre tracciare la retta passante per P orientata lungo la direzione ortogonale a $\mathbf{p} - \mathbf{c}$. Questa direzione, sia \mathbf{u} , è espressa in coordinate come $\mathbf{u} =$

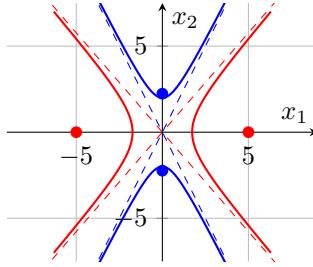


FIGURA 7. Iperboli di equazione $\frac{x_1^2}{3} - \frac{x_2^2}{4} = 1$ (rosso) e $-x_1^2 + \frac{x_2^2}{4} = 1$ (blu). I punti indicati sugli assi sono i fuochi, mentre le linee tratteggiate sono gli asintoti delle iperboli.

$(c_2 - p_2, p_1 - c_1)$, per cui l'equazione è

$$(1.124) \quad \frac{x_1 - p_1}{c_2 - p_2} = \frac{x_2 - c_2}{p_1 - c_1} \Leftrightarrow (x_1 - p_1)(p_1 - c_1) + (x_2 - c_2)(p_2 - c_2) = 0.$$

7.4.3. *Il caso $\det \Phi < 0$ e l'iperbole.* Se $\det \Phi < 0$ possiamo ugualmente ottenere la forma standard in Eq. (1.121), ma questa volta $\lambda_1 \lambda_2 < 0$. Come nel caso precedente, se $\tilde{\gamma} = 0$ l'equazione descrive il solo punto \hat{O} . Se invece $\tilde{\gamma} \neq 0$, supponiamo per esempio che λ_2 e $\tilde{\gamma}$ abbiano lo stesso segno, dividiamo per $\tilde{\gamma}$ e scriviamo

$$\frac{\tilde{x}_1^2}{a^2} - \frac{\tilde{x}_2^2}{b^2} = 1, \quad a^2 := -\frac{\tilde{\gamma}}{\lambda_1}, \quad b^2 := \frac{\tilde{\gamma}}{\lambda_2}.$$

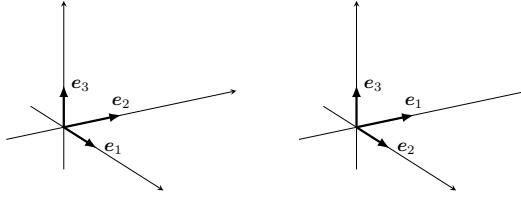
Questa equazione descrive una *iperbole*. Infatti, essa fornisce tutti i punti \mathbf{x} le cui distanze da due fuochi, **1** e **2**, hanno una differenza in modulo fissa, ovvero $2a$, con $a > 0$

$$\|\mathbf{x} - \mathbf{f}_1\| - \|\mathbf{x} - \mathbf{f}_2\| = 2a,$$

con fuochi aventi coordinate $\tilde{\mathbf{f}}_1 = (c, 0)$ e $\tilde{\mathbf{f}}_2 = (-c, 0)$, dove $c^2 = a^2 + b^2$. Nel caso in cui invece $\tilde{\gamma}$ e λ_2 abbiano lo segno opposto, e quindi l'equazione sia nella forma $\frac{\tilde{x}_1^2}{a^2} - \frac{\tilde{x}_2^2}{b^2} = 1$, i fuochi sono in posizione $\tilde{\mathbf{f}}_1 = (0, c)$ e $\tilde{\mathbf{f}}_2 = (0, -c)$, con $c^2 = a^2 + b^2$. Infine, una iperbole è dotata di due *asintoti*, ovvero rette che l'iperbole avvicina (senza mai toccare) quando \mathbf{x} si allontana dall'origine. Queste rette hanno equazione $b\tilde{x}_1 \pm a\tilde{x}_2 = 0$: questo si può vedere osservando che l'equazione della parabola implica che, per $\tilde{x}_1^2, \tilde{x}_2^2 \gg 0$, $\frac{\tilde{x}_1^2}{\tilde{x}_2^2} \simeq \frac{a^2}{b^2}$.

8. Geometria analitica nello spazio

Consideriamo ora il caso $d = 3$ e vediamo come studiare i piani nello spazio e la loro distanza. Anche in questa sezione useremo la notazione $P(\mathbf{p})$ per indicare un punto P associato al vettore $\mathbf{p} \in \mathbb{V}$ le cui coordinate sono $\mathbf{p} = (p_1, p_2, p_3)^\top$ secondo un riferimento ortogonale cartesiano fissato $Oe_1e_2e_3$. Come nel caso del piano, una base ortonormale nello spazio è a volte denotata con una notazione speciale, $e_1 = \mathbf{i}$, $e_2 = \mathbf{j}$, $e_3 = \mathbf{k}$. Anche questa base viene indicizzata in accordo con una convenzione che ne definisce una *orientazione positiva* e che sfrutta la definizione di prodotto vettoriale. Ricordiamo che, in uno spazio euclideo tridimensionale, data la base ortonormale \mathcal{B}_e , il prodotto vettoriale è tale che $e_3 = e_1 \wedge e_2$. La convenzione *positiva* si ottiene seguendo la cosiddetta “regola della mano destra”: puntando il pollice nella direzione di e_1 , l'indice in quella di e_2 , e tenendo il medio in direzione perpendicolare al palmo della mano, quest'ultimo punterà nella direzione di e_3 . Per esempio, nella figura sotto a sinistra la convenzione è positiva, mentre a destra è negativa.



Come nota finale, osserviamo che il modulo del prodotto vettoriale ha anche una interessante interpretazione geometrica: $\|\mathbf{x} \wedge \mathbf{y}\|$ è infatti il doppio dell'area del triangolo con vertici O , X e Y . Basta infatti osservare che $\|\mathbf{y}\| |\sin \theta|$ è proprio l'altezza del triangolo siffatto rispetto alla base individuata da \overrightarrow{OX} . Torneremo su questo punto dato che è legato ad una interpretazione geometrica generale del determinante.

8.1. Il piano. Come nel caso delle rette nel piano, anche un piano Π nello spazio può essere individuato in diversi modi. Una maniera può essere dare un punto $P(\mathbf{p})$ e un vettore \mathbf{v} non nullo: il piano viene quindi individuato da tutti i punti $X(\mathbf{x})$ tali per cui $\mathbf{x} - \mathbf{p}$ è ortogonale alla direzione individuata da \mathbf{v} . Sia quindi $\mathbf{v} = (v_1, v_2, v_3)^\top$ una rappresentazione di \mathbf{v} nel nostro riferimento cartesiano: la condizione di ortogonalità si scrive

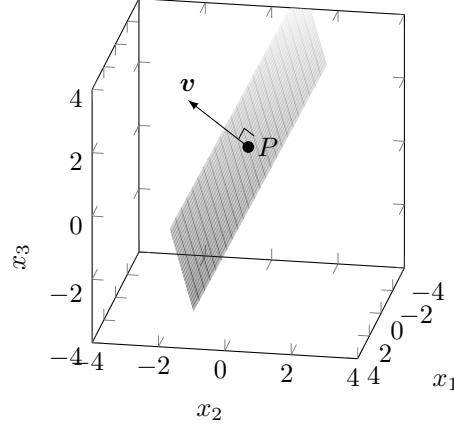
$$(1.125) \quad \Pi: \langle \mathbf{v}, \mathbf{x} - \mathbf{p} \rangle = 0 \Leftrightarrow v_1(x_1 - p_1) + v_2(x_2 - p_2) + v_3(x_3 - p_3) = 0,$$

che possiamo riarrangiare nella forma

$$\Pi: v_1x_1 + v_2x_2 + v_3x_3 + \gamma = 0,$$

dove $\gamma = -\langle \mathbf{v}, \mathbf{p} \rangle$. Si noti che ciò che è importante è la direzione di \mathbf{v} , non il suo modulo o verso (quindi $c\mathbf{v}$, con $c \in \mathbb{R}$, per esempio, individua lo stesso piano). Le coordinate \mathbf{v} di \mathbf{v} sono dette *parametri di giacitura* e, per quanto appena detto, sono individuate a meno di un fattore non nullo di proporzionalità. L'insieme dei piani individuato da $\langle \mathbf{v}, \mathbf{x} - \mathbf{p} \rangle = 0$ al variare di \mathbf{v} è detta *stella dei piani con centro in P* . Dato che \mathbf{v} individua la direzione ortogonale al piano rispetto al riferimento cartesiano, a seconda del suo valore possono succedere diversi fatti.

- Se $\langle \mathbf{v}, \mathbf{p} \rangle = 0$, il piano passa per l'origine O .
- Se $v_i = 0$, allora il piano è parallelo all'asse e_i ; se due elementi di \mathbf{v} sono nulli, siano per esempio v_i e v_j con $i \neq j$, allora il piano è parallelo al piano generato da $\{e_i, e_j\}$.



Una maniera diversa di individuare il piano è dare *tre punti*, $P(\mathbf{p})$, $Q(\mathbf{q})$ e $R(\mathbf{r})$ non allineati, e richiedere che il piano Π passi per questi tre punti. Ciò vuol dire che, se $X(\mathbf{x}) \in \Pi$, $\mathbf{x} - \mathbf{p}$,

$\mathbf{q} - \mathbf{p}$ e $\mathbf{r} - \mathbf{p}$ devono essere complanari, ovvero linearmente dipendenti, e quindi in coordinate

$$(1.126) \quad \Pi: \begin{vmatrix} x_1 - p_1 & x_2 - p_2 & x_3 - p_3 \\ q_1 - p_1 & q_2 - p_2 & q_3 - p_3 \\ r_1 - p_1 & r_2 - p_2 & r_3 - p_3 \end{vmatrix} = 0$$

che fornisce l'equazione del piano passante per i tre punti dati. È anche possibile individuare il piano usando un punto $P(\mathbf{p})$ e due vettori non paralleli $\mathbf{u}, \mathbf{w} \in \mathbb{V}$ che individuano due direzioni nello spazio in modo che il loro span sia il piano che ci interessa. Possiamo usare questi due vettori per costruire $\mathbf{v} = \mathbf{u} \wedge \mathbf{w}$, ortogonale a entrambi e quindi anche al piano, e usare \mathbf{v} esattamente come sopra: un punto $X(\mathbf{x}) \in \Pi$ se $\mathbf{x} - \mathbf{p}$ è ortogonale a \mathbf{v} , ovvero a $\mathbf{u} \wedge \mathbf{w}$, o in altre parole $\langle \mathbf{x} - \mathbf{p}, \mathbf{u} \wedge \mathbf{w} \rangle = 0$. In coordinate questo significa

$$(1.127) \quad \Pi: \begin{vmatrix} x_1 - p_1 & x_2 - p_2 & x_3 - p_3 \\ u_1 & u_2 & u_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = 0.$$

O ancora diversamente, si può chiedere che il vettore $\mathbf{x} - \mathbf{p}$ sia nello span di \mathbf{u} e \mathbf{v} , ovvero in coordinate

$$(1.128) \quad \mathbf{x} - \mathbf{p} = t_1 \mathbf{u} + t_2 \mathbf{v} \Rightarrow \mathbf{x} = \mathbf{p} + t_1 \mathbf{u} + t_2 \mathbf{v}, \quad t_1, t_2 \in \mathbb{R}.$$

Quella ottenuta è la *rappresentazione parametrica del piano*. Come nel caso della retta nel piano, tutte le rappresentazioni date sono equivalenti.

8.1.1. *Intersezioni di piani*. Esattamente come nel caso delle rette nel piano, dati due piani

$$\Pi: v_1 x_1 + v_2 x_2 + v_3 x_3 + \gamma = 0, \quad \Pi': v'_1 x_1 + v'_2 x_2 + v'_3 x_3 + \gamma' = 0$$

associati rispettivamente ai vettori normali \mathbf{v} e \mathbf{v}' , possiamo studiare la loro intersezione risolvendo il sistema

$$\begin{cases} v_1 x_1 + v_2 x_2 + v_3 x_3 + \gamma = 0, \\ v'_1 x_1 + v'_2 x_2 + v'_3 x_3 + \gamma' = 0 \end{cases}$$

la cui matrice dei coefficienti e matrice orlata sono rispettivamente

$$\mathbf{A} = \begin{pmatrix} v_1 & v_2 & v_3 \\ v'_1 & v'_2 & v'_3 \end{pmatrix}, \quad \mathbf{H} = \left(\begin{array}{ccc|c} v_1 & v_2 & v_3 & -\gamma \\ v'_1 & v'_2 & v'_3 & -\gamma' \end{array} \right).$$

Possono verificarsi diverse eventualità:

$\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{H}) = 2$: l'intersezione è non vuota e ha dimensione 1, ovvero è una retta;

$\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{H}) = 1$: l'intersezione è non vuota e ha dimensione 2, ovvero è un piano: le due equazioni descrivono lo stesso piano nello spazio;

$1 = \text{rank}(\mathbf{A}) \neq \text{rank}(\mathbf{H}) = 2$: l'intersezione tra i due piani è vuota: essi sono paralleli.

Due piani sono paralleli se i rispettivi vettori normali \mathbf{v} e \mathbf{v}' sono paralleli; viceversa sono ortogonali se lo sono i rispettivi vettori normali, $\langle \mathbf{v}, \mathbf{v}' \rangle = 0$.

8.1.2. *Fasci di piani*. Esattamente come nel caso delle rette, è possibile definire dei *fasci di piani* come una famiglia di piani con una certa proprietà. Un fascio di piani *improprio* è una famiglia di piani paralleli: ciò vuol dire che tutti i piani hanno lo stesso vettore ortogonale \mathbf{v} , ma diverso \mathbf{p} , per cui una fascio di piani improprio si scrive

$$(1.129) \quad \langle \mathbf{v}, \mathbf{x} \rangle + t = v_1 x_1 + v_2 x_2 + v_3 x_3 + t = 0, \quad t \in \mathbb{R}.$$

Un fascio *proprio* di piani, invece, è una famiglia di piani con una *retta* in comune. Se

$$\Pi: v_1 x_1 + v_2 x_2 + v_3 x_3 + \gamma = 0, \quad \Pi': v'_1 x_1 + v'_2 x_2 + v'_3 x_3 + \gamma' = 0$$

hanno $\Pi \cap \Pi' = \mathcal{R}$, con \mathcal{R} retta, il fascio può essere scritto come

$$(1.130) \quad \lambda(v_1 x_1 + v_2 x_2 + v_3 x_3 + \gamma) + \mu(v'_1 x_1 + v'_2 x_2 + v'_3 x_3 + \gamma') = 0, \quad \lambda, \mu \in \mathbb{R},$$

oppure, in forma non omogenea,

$$(1.131) \quad v_1x_1 + v_2x_2 + v_3x_3 + \gamma + \rho(v'_1x_1 + v'_2x_2 + v'_3x_3 + \gamma') = 0, \quad \rho \in \mathbb{R},$$

dove al variare di ρ tutti gli elementi del fascio sono riprodotti eccezion fatta per Π .

8.2. La retta nello spazio. Consideriamo ora il problema della rappresentazione di una retta \mathcal{R} nello spazio. Dati tre punti distinti $X, P, Q \in \mathcal{R}$, essi devono essere tali che $\mathbf{q} - \mathbf{p}$ e $\mathbf{x} - \mathbf{p}$ sono linearmente dipendenti, che significa che

$$(1.132) \quad \mathbf{x} - \mathbf{p} = t(\mathbf{q} - \mathbf{p}) \Rightarrow \mathbf{x} = \mathbf{p} + t(\mathbf{q} - \mathbf{p}), \quad t \in \mathbb{R}.$$

Questa è la stessa equazione parametrica già ottenuta per il caso del piano in Eq. (1.115), ma questa volta i nostri vettori sono tridimensionali, quindi implica che

$$(1.133) \quad \frac{x_1 - p_1}{q_1 - p_1} = \frac{x_2 - p_2}{q_2 - p_2} = \frac{x_3 - p_3}{q_3 - p_3}.$$

Come nel caso bidimensionale, il vettore $\mathbf{u} = \mathbf{q} - \mathbf{p}$ individua la direzione della retta. Se è dato direttamente \mathbf{u} , le cui coordinate siano $\mathbf{u} = (u_1, u_2, u_3)^T$, allora

$$(1.134) \quad \frac{x_1 - p_1}{u_1} = \frac{x_2 - p_2}{u_2} = \frac{x_3 - p_3}{u_3}.$$

individua la retta passante per $P(\mathbf{p})$ nella direzione individuata da \mathbf{u} . Tale condizione si può scrivere come un insieme di *due* equazioni,

$$(1.135) \quad \begin{cases} v_1x_1 + v_2x_2 + v_3x_3 + \gamma = 0 \\ v'_1x_1 + v'_2x_2 + v'_3x_3 + \gamma' = 0 \end{cases},$$

con opportuni coefficienti, cioè, come forse atteso, come un problema di intersezione tra due piani: è questa la rappresentazione *cartesiana* di \mathcal{R} . Si noti che i due piani non sono identificati univocamente: ci sono, come abbiamo visto nello studio dei fasci propri, infiniti piani passanti per una retta, quindi quelli ottenuti sono solo due diversi piani tra infiniti possibili. Inoltre, perché il sistema sopra abbia come spazio delle soluzioni uno spazio unidimensionale, il rango della matrice dei coefficienti e della matrice orlata deve essere massimo, ovvero 2.

8.3. Posizioni reciproche di rette e piani. Avendo introdotto due oggetti nello spazio, la retta e il piano, ha senso chiedersi come misurare la loro posizione reciproca, ovvero come valutare se essi si intersecano o no, o qual è il loro angolo relativo. In questo compito, la nostra conoscenza dell'algebra lineare sarà di grande aiuto.

8.3.1. *Posizione reciproca tra retta e piano.* La posizione reciproca tra una retta \mathcal{R} e un piano Π può essere facilmente studiata utilizzando, ancora una volta, il teorema di Krockecker–Rouché–Capelli. Sia la retta

$$\mathcal{R}: \begin{cases} u_1x_1 + u_2x_2 + u_3x_3 + \gamma_u = 0 \\ w_1x_1 + w_2x_2 + w_3x_3 + \gamma_w = 0 \end{cases}$$

e il piano

$$\Pi: v_1x_1 + v_2x_2 + v_3x_3 + \gamma_v = 0.$$

Per trovare la loro intersezione dobbiamo risolvere il sistema

$$\mathcal{R} \cap \Pi: \begin{cases} v_1x_1 + v_2x_2 + v_3x_3 + \gamma_v = 0 \\ u_1x_1 + u_2x_2 + u_3x_3 + \gamma_u = 0 \\ w_1x_1 + w_2x_2 + w_3x_3 + \gamma_w = 0 \end{cases}$$

corrispondente alla matrice dei coefficienti \mathbf{A} e alla matrice orlata \mathbf{H}

$$\mathbf{A} = \begin{pmatrix} v_1 & v_2 & v_3 \\ u_1 & u_2 & u_3 \\ w_1 & w_2 & w_3 \end{pmatrix}, \quad \mathbf{H} = \left(\begin{array}{ccc|c} v_1 & v_2 & v_3 & -\gamma_v \\ u_1 & u_2 & u_3 & -\gamma_u \\ w_1 & w_2 & w_3 & -\gamma_w \end{array} \right).$$

Osservando che il rango della matrice \mathbf{A} e quindi della matrice \mathbf{H} deve essere *almeno* pari a 2 (per via del fatto che il sottosistema corrispondente ad \mathcal{R} deve appunto rappresentare una retta), possono verificarsi diversi casi:

$\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{H}) = 3$: esiste un'unica soluzione del sistema, e quindi un unico punto P , tale che $\mathcal{R} \cap \Pi = \{P\}$;

$\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{H}) = 2$: esiste un sottospazio di dimensione 1, ovvero una retta, soluzione del sistema: questa retta non può che essere \mathcal{R} stessa, che quindi giace nel piano, per cui $\mathcal{R} \cap \Pi = \mathcal{R}$;

$2 = \text{rank}(\mathbf{A}) \neq \text{rank}(\mathbf{H}) = 3$: non esistono soluzioni, ovvero la retta è *parallela* al piano.

8.3.2. *Angoli*. Date due rette \mathcal{R} e \mathcal{R}' , supposte incidenti, siano \mathbf{u} e \mathbf{u}' i vettori che ne individuano la direzione: il loro angolo è $\theta = \arccos \frac{\langle \mathbf{u}, \mathbf{u}' \rangle}{\|\mathbf{u}\| \|\mathbf{u}'\|}$, esattamente come nel caso delle rette nel piano. È possibile anche definire i cosiddetti *coseni direttori* rispetto agli assi di una retta \mathcal{R} , ovvero i coseni degli angoli tra la retta e le tre direzioni individuate da \mathbf{e}_1 , \mathbf{e}_2 e \mathbf{e}_3 , dati da

$$\cos \theta_i = \frac{\langle \mathbf{u}, \mathbf{e}_i \rangle}{\|\mathbf{u}\|} = \frac{u_i}{\sqrt{u_1^2 + u_2^2 + u_3^2}}.$$

Questi coseni non sono altro che i coefficienti che appaiono scrivendo il vettore \mathbf{u} nella base ortogonale,

$$\mathbf{u} = \|\mathbf{u}\| (\cos \theta_1 \mathbf{e}_1 + \cos \theta_2 \mathbf{e}_2 + \cos \theta_3 \mathbf{e}_3).$$

Se consideriamo invece due piani Π e Π' con vettori ortogonali \mathbf{v} e \mathbf{v}' , si può analogamente scrivere che il loro angolo relativo è $\theta = \arccos \frac{\langle \mathbf{v}, \mathbf{v}' \rangle}{\|\mathbf{v}\| \|\mathbf{v}'\|}$ (e quindi i piani sono ortogonali se $\langle \mathbf{v}, \mathbf{v}' \rangle = 0$).

Infine, siano dati una retta nella direzione \mathbf{u} e un piano con direzione ortogonale \mathbf{v} . Questo vuol dire che se ϑ è l'angolo tra \mathbf{v} e \mathbf{u} , l'angolo tra la retta e la direzione *ortogonale a* \mathbf{v} (ovvero, parallela al piano) è $\theta = \pi - \vartheta$, per cui

$$(1.136) \quad \sin \theta = \cos \vartheta = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{u}\| \|\mathbf{v}\|}.$$

8.3.3. *Distanze*. Gli argomenti che ci hanno permesso di ottenere l'Eq. (1.117) possono essere riprodotti esattamente qui per il piano: se Π è individuato dalla condizione $\langle \mathbf{v}, \mathbf{x} - \mathbf{p} \rangle = 0$, e $Q(\mathbf{q})$ è un punto nello spazio, allora la distanza tra Q e il piano è

$$(1.137) \quad d(Q, \Pi) = \frac{|\langle \mathbf{v}, \mathbf{q} - \mathbf{p} \rangle|}{\|\mathbf{v}\|}.$$

Ovvero, se il piano è individuato dall'equazione cartesiana Π : $v_1 x_1 + v_2 x_2 + v_3 x_3 + \gamma$,

$$(1.138) \quad d(Q, \Pi) = \frac{|v_1 q_1 + v_2 q_2 + v_3 q_3 + \gamma|}{\sqrt{v_1^2 + v_2^2 + v_3^2}}.$$

Volendo invece calcolare la distanza di un punto $P(\mathbf{p})$ da una retta \mathcal{R} nello spazio, si può osservare che se la retta ha direzione \mathbf{u} , allora il piano individuato da Π : $\langle \mathbf{u}, \mathbf{x} - \mathbf{p} \rangle = 0$ passa per P ed è ortogonale alla retta. Se $H = \mathcal{R} \cap \Pi$, allora $d(P, \mathcal{R}) = d(P, H)$.

Infine, consideriamo il problema del calcolo della distanza di due rette \mathcal{R} e \mathcal{R}' nello spazio. Esse possono essere incidenti (e allora la loro distanza è zero), parallele (e allora la loro distanza è uguale alla distanza di un qualunque punto di una dall'altra) o *sghembe*. Due rette sono

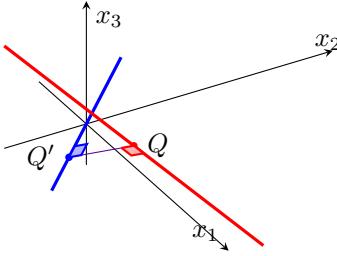


FIGURA 8. Due rette sghembe e la distanza tra esse.

sghembe se non si intersecano ma i due vettori \mathbf{u} e \mathbf{u}' che ne indicano la direzione non sono paralleli. Supponiamo quindi che le due rette abbiano la forma $\mathbf{x} = \mathbf{p} + t\mathbf{u}$, $\mathbf{x} = \mathbf{p}' + t\mathbf{u}'$, $t, t' \in \mathbb{R}$. In questo caso, è possibile individuare un vettore \mathbf{w} tale che $\langle \mathbf{w}, \mathbf{u} \rangle = 0$ e $\langle \mathbf{w}, \mathbf{u}' \rangle = 0$. Volendone individuare la direzione, questo vuol dire cercare la soluzione del sistema

$$(1.139) \quad \langle \mathbf{u}, \mathbf{w} \rangle = \langle \mathbf{u}', \mathbf{w} \rangle = 0 \Leftrightarrow \begin{cases} u_1 w_1 + u_2 w_2 + u_3 w_3 = 0 \\ u'_1 w_1 + u'_2 w_2 + u'_3 w_3 = 0 \end{cases}$$

la cui matrice dei coefficienti ha rango massimo (ovvero 2) dato che i vettori \mathbf{u} e \mathbf{u}' sono linearmente indipendenti. Una volta trovato \mathbf{w} , occorre trovare $Q(\mathbf{q}) \in \mathcal{R}$ e $Q'(\mathbf{q}') \in \mathcal{R}'$ tali per cui $\mathbf{q} - \mathbf{q}' = c\mathbf{w}$ per un qualche $c \in \mathbb{R}$, ovvero tali per cui $\mathbf{q} - \mathbf{q}'$ è ortogonale a entrambe le rette. Dato che $Q(\mathbf{q}) \in \mathcal{R}$ e $Q'(\mathbf{q}') \in \mathcal{R}'$ questo significa che esistono τ e τ' per cui $\mathbf{q} = \mathbf{p} + \tau\mathbf{u}$ e $\mathbf{q} = \mathbf{p}' + \tau'\mathbf{u}'$,

$$(1.140) \quad \|\mathbf{q} - \mathbf{q}'\|^2 = \langle c\mathbf{w}, \mathbf{p} - \mathbf{p}' + \tau\mathbf{u} - \tau'\mathbf{u}' \rangle = c\langle \mathbf{w}, \mathbf{p} - \mathbf{p}' \rangle.$$

Ma $\|\mathbf{q} - \mathbf{q}'\| = |c| \|\mathbf{w}\|$ ovvero

$$(1.141) \quad d(Q, Q') = \frac{|\langle \mathbf{w}, \mathbf{p} - \mathbf{p}' \rangle|}{\|\mathbf{w}\|}.$$

Per definizione $d(\mathcal{R}, \mathcal{R}') := d(Q, Q')$.

8.4. Sfere e circonferenze nello spazio. La sfera è l'analogo della circonferenza in tre dimensioni ed è descritta dalla stessa equazione: dato un punto $C(\mathbf{c})$ e un reale positivo r , la sfera \mathcal{S} di centro C e raggio r è data dai punti $X(\mathbf{x})$ tali che

$$(1.142) \quad \mathcal{S}: d^2(X, C) = \|\mathbf{x} - \mathbf{c}\|^2 = \|\mathbf{x}\|^2 + 2\langle \mathbf{x}, \mathbf{c} \rangle + \|\mathbf{c}\|^2 = r^2.$$

In coordinate questo significa

$$(1.143) \quad (x_1 - c_1)^2 + (x_2 - c_2)^2 + (x_3 - c_3)^2 = r^2,$$

che può mettersi nella forma

$$(1.144) \quad x_1^2 + x_2^2 + x_3^2 + a_1 x_1 + a_2 x_2 + a_3 x_3 + \gamma = 0.$$

Esattamente come abbiamo visto nel caso della circonferenza, possiamo tentare di mettere una generica equazione tipo Eq. (1.144) nella forma in Eq. (1.143) identificando

$$c_i = -\frac{a_i}{2} \text{ per } i = 1, 2, 3, \quad r^2 = \frac{a_1^2 + a_2^2 + a_3^2}{4} - \gamma.$$

Se $r^2 > 0$ abbiamo una sfera ordinaria, mentre se $r^2 = 0$ la sfera ha raggio nullo. Se $r^2 < 0$, invece, abbiamo una inconsistenza, ovvero la sfera è “immaginaria” ed in effetti ciò significa che $\mathcal{S} = \emptyset$.

Dato un piano Π , esso può avere diverse posizioni rispetto alla sfera:

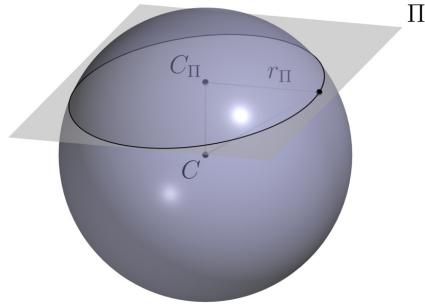


FIGURA 9. Sfera di centro C intersecata da piano Π . L'intersezione è una circonferenza di centro C_Π , punto del piano di minima distanza da C .

- se $d(C, \Pi) > r$, il piano è esterno e non tocca la sfera;
- se $d(C, \Pi) = r$, il piano è tangente e tocca la sfera in un punto;
- se $d(C, \Pi) < r$, il piano è secante come in figura, e $S \cap \Pi$ sia l'insieme dei punti dell'intersezione. Questi punti distano tutti r dal centro della sfera, mentre distano $r_\Pi^2 = r^2 - d^2(C, \Pi)$ dall'intersezione C_Π tra il piano Π e la retta R ad esso ortogonale passante per C : l'insieme è quindi una circonferenza di centro C_Π e raggio r_Π .

L'osservazione sopra implica che una circonferenza C può in effetti essere rappresentata nello spazio proprio come una intersezione tra una sfera e un piano: si noti che esiste un unico piano che contiene una certa circonferenza (quello in cui essa giace) ma infinite sfere che intersecano quel piano per produrre la stessa circonferenza.

CAPITOLO 2

Elementi di calcolo in più variabili

Premessa. Nel precedente capitolo abbiamo già visto come sia possibile associare *in maniera biunivoca* ad ogni punto $X \in \mathbb{E}_d$ un vettore \mathbf{x} di uno spazio bidimensionale, le cui coordinate, una volta specificato un riferimento cartesiano $Oe_1 \dots e_d$, sono espresse da un vettore di numeri reali $\mathbf{x} \in \mathbb{R}^d$. Alla luce di questa corrispondenza, immagineremo in generale *fissato* il riferimento cartesiano: ciò permette di associare *univocamente* ogni punto $X(\mathbf{x})$ al corrispondente set di coordinate \mathbf{x} e quindi di identificare i tre oggetti (punto, vettore nello spazio vettoriale, coordinate del vettore), $X \simeq \mathbf{x} \simeq \mathbf{x}$. In definitiva, cioè, possiamo identificare $\mathbb{E}_d \simeq \mathbb{R}^d$. Scriveremo quindi per brevità $\mathbf{x} \in \mathbb{R}^d$ significando il punto X nello spazio associato ad un vettore \mathbf{x} con coordinate $\mathbf{x} = (x_1, \dots, x_d)^\top \in \mathbb{R}^d$.

1. Intorni in \mathbb{R}^d : topologia, limiti e continuità

1.1. Topologia di \mathbb{R}^d . In questa prima sezione introduciamo un “dizionario” per l’analisi in molte dimensioni. Molte definizioni che verranno date generalizzano il caso unidimensionale. Abbiamo già visto che, dati due punti $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$, possiamo associare una *distanza* tra essi,

$$d(\mathbf{x}, \mathbf{x}') = \|\mathbf{x} - \mathbf{x}'\| = \sqrt{\sum_{i=1}^d (x_i - x'_i)^2}.$$

Esattamente come nel caso delle funzioni reali su \mathbb{R} , è utile partire dal concetto di *intorno*, che possiamo introdurre avendo già dato una distanza.

DEFINIZIONE 1.1 (Intorno sferico). *Sia $\mathbf{x} \in \mathbb{R}^d$. Data la quantità reale $r > 0$, si dice intorno sferico di \mathbf{x} di raggio r l’insieme dei punti \mathbf{p} tali che*

$$(2.1) \quad \mathcal{I}_r(\mathbf{x}) := \{\mathbf{p} \in \mathbb{R}^d : 0 < \|\mathbf{p} - \mathbf{x}\| \leq r^2\}.$$

Un insieme di punti $\mathcal{A} \subseteq \mathbb{R}^d$ si dice *limitato* se esiste un intorno che lo contiene: diversamente si dice che \mathcal{A} è *illimitato*. Come nel caso unidimensionale, possiamo dare una definizione di *punto di accumulazione* rispetto ad un insieme di punti $\mathcal{A} \subseteq \mathbb{R}^d$.

DEFINIZIONE 1.2. *Un punto $\mathbf{x} \in \mathbb{R}^d$ si dice di accumulazione di \mathcal{A} se per ogni intorno circolare $\mathcal{I}_r(\mathbf{x})$, $\mathcal{I}_r(\mathbf{x}) \cap \mathcal{A} \neq \emptyset$.*

In particolare questo significa che se \mathbf{x} è un punto di accumulazione di \mathcal{A} , in ogni suo intorno cadono *infiniti* punti di \mathcal{A} : d’altra parte questo implica anche che se \mathcal{A} ha un numero finito di punti, esso non può avere punti di accumulazione. Vale il seguente teorema:

TEOREMA 1.1 (di Bolzano–Weierstrass). *Ogni insieme limitato contenente infiniti punti ammette almeno un punto di accumulazione.*

Dato un insieme $\mathcal{A} \subseteq \mathbb{R}^d$, definiamo il complemento di \mathcal{A} in \mathbb{R}^d come $\bar{\mathcal{A}} := \mathbb{R}^d \setminus \mathcal{A}$. Possiamo allora distinguere punti *interni*, *esterni* e *di frontiera*.

DEFINIZIONE 1.3. Un punto $\mathbf{x} \in \mathcal{A} \subseteq \mathbb{R}^d$ si dice interno se esiste $\mathcal{I}_r(\mathbf{x}) \subseteq \mathcal{A}$. Si dice viceversa esterno se esiste $\mathcal{I}_r(\mathbf{x}) \subset \bar{\mathcal{A}}$. Si dice infine di frontiera se per ogni intorno $\mathcal{I}_r(\mathbf{x})$, $\mathcal{I}_r(\mathbf{x}) \cap \mathcal{A} \neq \emptyset$ e $\mathcal{I}_r(\mathbf{x}) \cap \bar{\mathcal{A}} \neq \emptyset$. L'insieme di tutti i punti di frontiera di \mathcal{A} si dice frontiera o contorno di \mathcal{A} e si indica con $\partial\mathcal{A}$. L'insieme di punti interni di \mathcal{A} si dice interno di \mathcal{A} e si indica con \mathcal{A}° .

Infine, possiamo dare la definizione di insieme aperto e insieme chiuso in questo contesto.

DEFINIZIONE 1.4 (Insieme aperto e chiuso). Un insieme \mathcal{A} è aperto se ogni suo punto è interno, ovvero non contiene punti di frontiera. Viceversa, un insieme è chiuso se contiene tutti i suoi punti di frontiera.

Da questa definizione discende che possono avversi insiemi che non sono né aperti né chiusi.

DEFINIZIONE 1.5 (Dominio). Un dominio di \mathbb{R}^d è un insieme di punti che è chiuso e tale per cui ogni suo punto è di accumulazione di punti interni.

Concludiamo con la definizione di insieme connesso per archi e insieme semplicemente connesso.

DEFINIZIONE 1.6 (Insieme connesso per archi). Un insieme $\mathcal{A} \subseteq \mathbb{R}^d$ si dice connesso per archi se, dati due punti $\mathbf{x}, \mathbf{y} \in \mathcal{A}$, esiste una funzione continua $\gamma: [0, 1] \in \mathcal{A}$, tale che $\gamma(0) = \mathbf{x}$ e $\gamma(1) = \mathbf{y}$.

Una funzione continua $\gamma: [0, 1] \in \mathcal{A}$, tale che $\gamma(0) = \mathbf{x}$ e $\gamma(1) = \mathbf{y}$ è detta *arco* tra \mathbf{x} e \mathbf{y} . Se $\mathbf{x} = \mathbf{y}$ essa è detta *arco centrato in \mathbf{x}* . Un arco centrato in \mathbf{x} è *contraibile* se esiste una funzione continua $\mathbf{h}: [0, 1] \times [0, 1] \rightarrow \mathbb{R}^d$ tale che $\mathbf{h}(t, 0) = \gamma(t)$ e $\mathbf{h}(t, 1) = \mathbf{p} \forall t \in [0, 1]$.

DEFINIZIONE 1.7 (Insieme semplicemente connesso). Sia $\mathcal{A} \subseteq \mathbb{R}^d$ connesso per archi. Esso si dice semplicemente connesso se, per ogni $\mathbf{p} \in \mathcal{A}$, un arco centrato in \mathbf{p} è contraibile.

1.2. Funzioni. Una certa quantità reale z si dice funzione di $\mathbf{x} \in \mathcal{A} \subseteq \mathbb{R}^d$ quando ad ogni $\mathbf{x} \in \mathcal{A}$ corrisponde uno e un solo valore $z \in \mathbb{R}$. L'insieme \mathcal{A} si dice *insieme di esistenza o definizione* e si scrive che $z = f(\mathbf{x})$ per indicare che esiste una legge, indicata con f , che associa ad ogni $\mathbf{x} \in \mathcal{A}$ uno e un solo $z \in \mathbb{R}$, ovvero in simboli

$$f: \mathcal{A} \rightarrow \mathbb{R}.$$

L'insieme

$$f(\mathcal{A}) := \{z \in \mathbb{R}: \exists \mathbf{x} \in \mathcal{A} \text{ tale che } z = f(\mathbf{x})\} \subseteq \mathbb{R}.$$

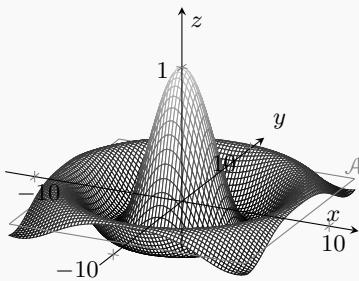
è l'immagine di \mathcal{A} per mezzo di f : f si dice limitata se $f(\mathcal{A})$ è limitato. Il luogo geometrico dei punti $(\mathbf{x}, f(\mathbf{x}))^\top \in \mathbb{R}^{d+1}$ ottenuti facendo variare \mathbf{x} in \mathcal{A} è detto *grafico* della funzione f . Infine, possiamo definire il *massimo assoluto* $\max_{\mathbf{x} \in \mathcal{A}} f(\mathbf{x}) := \max f(\mathcal{A})$ e il *minimo assoluto* di f su \mathcal{A} , $\min_{\mathbf{x} \in \mathcal{A}} f(\mathbf{x}) := \min f(\mathcal{A})$, se questi due valori esistono.

Tutte le definizioni sopra sono una generalizzazione di quanto visto per $d = 1$. Un caso speciale è il caso $d = 2$ semplicemente per via del fatto che è possibile *rappresentare* graficamente una funzione. In questo caso si usa una notazione speciale anche per le componenti di \mathbf{x} , ovvero si denota con x la componente x_1 e con y la componente x_2 , scrivendo $\mathbf{x} = (x, y)^\top$. La rappresentazione grafica si ottiene per mezzo di un diagramma cartesiano in cui ogni punto individuato dalla funzione f è identificato dalla terna $(x, y, f(\mathbf{x})) \in \mathbb{R}^3$.

Per esempio, supponiamo di avere la funzione

$$z = f(\mathbf{x}) = \frac{\sin \sqrt{x^2 + y^2}}{\sqrt{x^2 + y^2}}, \quad \mathcal{A} := \{\mathbf{x} = (x, y)^\top \in \mathbb{R}^2: |x| < 8, |y| < 8\}.$$

Otteniamo la seguente rappresentazione:



1.3. Limiti. Come nel caso delle funzioni in una variabile, è possibile definire il limite di una funzione $f: \mathcal{A} \rightarrow \mathbb{R}$. Diamo di seguito le seguenti definizioni di limite di f verso un punto di accumulazione \mathbf{p} per \mathcal{A} .

DEFINIZIONE 1.8. Sia $f: \mathcal{A} \rightarrow \mathbb{R}$ e sia \mathbf{p} punto di accumulazione di \mathcal{A} . Allora si dice che $\ell \in \mathbb{R}$ è il limite di f per $\mathbf{x} \rightarrow \mathbf{p}$, e si scrive

$$\lim_{\mathbf{x} \rightarrow \mathbf{p}} f(\mathbf{x}) = \ell,$$

se per ogni $\varepsilon > 0$ esiste $\delta > 0$ tale che

$$|\ell - f(\mathbf{x})| < \varepsilon \quad \forall \mathbf{x} \in \mathcal{J}_\delta(\mathbf{p}).$$

DEFINIZIONE 1.9. Sia $f: \mathcal{A} \rightarrow \mathbb{R}$ e sia \mathbf{p} punto di accumulazione di \mathcal{A} . Si dice che la funzione ha limite infinito positivo, e si scrive

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

se per ogni $M > 0$ esiste $\delta > 0$ tale che

$$f(\mathbf{x}) > M \quad \forall \mathbf{x} \in \mathcal{J}_\delta(\mathbf{p}).$$

DEFINIZIONE 1.10. Sia $f: \mathcal{A} \rightarrow \mathbb{R}$ e sia \mathbf{p} punto di accumulazione di \mathcal{A} . Si dice che la funzione ha limite infinito negativo, e si scrive

$$\lim_{\mathbf{x} \rightarrow \mathbf{p}} f(\mathbf{x}) = -\infty,$$

se per ogni $M > 0$ esiste $\delta > 0$ tale che

$$f(\mathbf{x}) < -M \quad \forall \mathbf{x} \in \mathcal{J}_\delta(\mathbf{p}).$$

Le definizioni sopra possono ripetersi per dare il limite di una funzione f all'infinito, scrivendo $\lim_{\mathbf{x} \rightarrow \infty} f(\mathbf{x})$. In questo caso, occorre dare un senso al concetto di "intorno dell'infinito" $\mathcal{J}_\delta(\infty)$ per un dato $\delta > 0$: si può utilizzare la definizione

$$\mathcal{J}_\delta(\infty) := \mathbb{R}^d \setminus \mathcal{J}_{1/\delta}(\mathbf{0}).$$

I teoremi fondamentali riguardo la composizione di limiti di due funzioni diverse visti per il caso in una variabile continuano a valere nella forma seguente.

TEOREMA 1.2. Siano f e g due funzioni definite sullo stesso insieme $\mathcal{A} \subseteq \mathbb{R}^d$, avente punto di accumulazione \mathbf{p} . Se esistono $\ell, \ell' \in \mathbb{R}$ tali per cui

$$\lim_{\mathbf{x} \rightarrow \mathbf{p}} f(\mathbf{x}) = \ell, \quad \lim_{\mathbf{x} \rightarrow \mathbf{p}} g(\mathbf{x}) = \ell',$$

allora

$$\lim_{\mathbf{x} \rightarrow \mathbf{p}} [f(\mathbf{x}) \pm g(\mathbf{x})] = \ell \pm \ell', \quad \lim_{\mathbf{x} \rightarrow \mathbf{p}} f(\mathbf{x})g(\mathbf{x}) = \ell\ell', \quad \text{se } \ell' \neq 0 \quad \lim_{\mathbf{x} \rightarrow \mathbf{p}} \frac{f(\mathbf{x})}{g(\mathbf{x})} = \frac{\ell}{\ell'}.$$

In fine, se $\ell \neq 0$, allora esiste $\delta > 0$ tale che $\forall \mathbf{x} \in \mathcal{J}_\delta(\mathbf{p})$, $f(\mathbf{x})$ ha lo stesso segno di ℓ .

1.4. Continuità. Ancora una volta, la definizione di continuità può essere data in analogia col caso unidimensionale.

DEFINIZIONE 1.11. Una funzione f con insieme di definizione \mathcal{A} è continua in $\mathbf{p} \in \mathcal{A}$, punto di accumulazione per \mathcal{A} , se

$$\lim_{\mathbf{x} \rightarrow \mathbf{p}} f(\mathbf{x}) = f(\mathbf{p}).$$

Se f è continua in ogni punto $\mathbf{x} \in \mathcal{A}$, essa si dice continua in \mathcal{A} .

Viceversa, se f non è continua nel punto di accumulazione $\mathbf{p} \in \mathcal{A}$, essa si dice discontinua in \mathbf{p} . Possiamo riassumere i principali fatti riguardo le funzioni continue nei seguenti teoremi, che sono gli equivalenti di analoghi risultati per il caso di funzioni in una variabile.

TEOREMA 1.3. Somme, differenze e prodotti di funzioni continue in un dato punto risultano in funzioni continue nello stesso punto. Quozienti di funzioni continue in un punto sono continue nel dato punto purché il denominatore del rapporto sia non nullo in esso.

TEOREMA 1.4. Se f è continua su un dominio limitato \mathcal{D} , essa ha sempre massimo e minimo assoluti su tale dominio.

2. Derivate di una funzione a più variabili

2.1. Derivate direzionali e derivate parziali. Supponiamo di avere una funzione f definita su un dominio \mathcal{D} , e sia \mathbf{p} un punto interno del dominio. Nel caso unidimensionale una *derivata* in un punto quantifica la variazione di una funzione, in un intorno del punto considerato, nella direzione dell'asse reale. Per una funzione su \mathbb{R}^d ci sono, naturalmente, infinite direzioni in cui si può studiare la variazione di una funzione f . Supponiamo quindi di scegliere una direzione specifica, indicata da un versore $\mathbf{v} \in \mathbb{R}^d$ e diamo la seguente definizione:

DEFINIZIONE 2.1 (Derivata direzionale). Sia $\mathcal{A} \subseteq \mathbb{R}^d$ un aperto e sia f definita su \mathcal{A} . Si dice che f è derivabile in $\mathbf{p} \in \mathcal{A}$ nella direzione \mathbf{v} se il limite

$$\lim_{t \rightarrow 0} \frac{f(\mathbf{p} + t\mathbf{v}) - f(\mathbf{p})}{t}$$

esiste finito. Questo limite si indica con

$$\frac{\partial f(\mathbf{p})}{\partial \mathbf{v}}.$$

La derivata introdotta è una *derivata direzionale*. Un caso particolarmente importante è quello in cui la direzione scelta è proprio quella fornita da un vettore della base ortonormale del riferimento cartesiano, $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$ di \mathbb{R}^d . In questo caso la derivata si chiama *derivata parziale* e si utilizza una convenzione speciale, ovvero

$$\frac{\partial f(\mathbf{p})}{\partial \mathbf{e}_i} \equiv \frac{\partial f(\mathbf{p})}{\partial x_i} \equiv \partial_i f(\mathbf{p}),$$

che equivale a dire che

$$\frac{\partial f(\mathbf{p})}{\partial x_i} := \lim_{t \rightarrow 0} \frac{f(p_1, p_2, \dots, p_{i-1}, p_i + t, p_{i+1}, \dots, p_d) - f(\mathbf{p})}{t}.$$

In altre parole, questa è l'usuale derivata di una funzione in una variabile dove le variabili x_j con $j \neq i$ vengono assunte fisse. Qui e nel seguito sarà molto utile introdurre il *gradiente* di una funzione f calcolato in un punto \mathbf{p} come il seguente vettore colonna

$$(2.2) \quad \nabla f(\mathbf{p}) = \left(\frac{\partial f(\mathbf{p})}{\partial x_1} \quad \frac{\partial f(\mathbf{p})}{\partial x_2} \quad \dots \quad \frac{\partial f(\mathbf{p})}{\partial x_d} \right)^T.$$

Una funzione dotata di tutte le derivate parziali in un punto \mathbf{p} si dice *derivabile* in \mathbf{p} . Esiste però una proprietà più stringente, quella di *differenziabilità*.

DEFINIZIONE 2.2 (Differenziabilità). *Una funzione $f: \mathbb{R}^d \rightarrow \mathbb{R}$ si dice differenziabile in \mathbf{p} se essa è derivabile e se*

$$\lim_{\mathbf{h} \rightarrow 0} \frac{f(\mathbf{p} + \mathbf{h}) - f(\mathbf{p}) - \langle \mathbf{h}, \nabla f(\mathbf{p}) \rangle}{\|\mathbf{h}\|} = 0.$$

Questa definizione comporta alcuni fatti. Per esempio, se f è differenziabile allora essa ha derivata in ogni direzione: basta scegliere $\mathbf{h} = t\mathbf{v}$, $t > 0$ e \mathbf{v} versore, per vedere che

$$\langle \mathbf{v}, \nabla f(\mathbf{p}) \rangle = \lim_{t \rightarrow 0} \frac{f(\mathbf{p} + t\mathbf{v}) - f(\mathbf{p})}{t} \equiv \frac{\partial f(\mathbf{p})}{\partial \mathbf{v}},$$

per cui, esplicitando,

$$\frac{\partial f(\mathbf{p})}{\partial \mathbf{v}} = \langle \mathbf{v}, \nabla f(\mathbf{p}) \rangle = \sum_{i=1}^d v_i \frac{\partial f(\mathbf{p})}{\partial x_i}.$$

Se una funzione f è differenziabile in un punto, si può usare il formalismo

$$d f(\mathbf{p}) = \sum_{i=1}^d \frac{\partial f(\mathbf{p})}{\partial x_i} d x_i,$$

per esprimere la variazione infinitesima di f quando ciascuna componente p_i dell'argomento viene modificata di un infinitesimo $d x_i$. Questa quantità si dice *differenziale*, o *differenziale totale*. Torneremo sui differenziali, e sul loro uso, in seguito.

La differenziabilità di una funzione ha alcune conseguenze molto interessanti. Per esempio, vale il seguente teorema che generalizza, ancora una volta, il caso unidimensionale.

TEOREMA 2.1 (Continuità di una funzione differenziabile). *Se una funzione f è differenziabile in \mathbf{p} allora è continua in \mathbf{p} .*

DIMOSTRAZIONE. Sia

$$\sigma(\mathbf{p}, \mathbf{x}) := f(\mathbf{x}) - f(\mathbf{p}) - \langle \mathbf{x} - \mathbf{p}, \nabla f(\mathbf{p}) \rangle \Rightarrow f(\mathbf{x}) - f(\mathbf{p}) = \sigma(\mathbf{p}, \mathbf{x}) + \langle \mathbf{x} - \mathbf{p}, \nabla f(\mathbf{p}) \rangle$$

che implica, usando la diseguaglianza di Cauchy–Schwarz,

$$|f(\mathbf{x}) - f(\mathbf{p})| \leq |\sigma(\mathbf{p}, \mathbf{x})| + \|\mathbf{x} - \mathbf{p}\| \|\nabla f(\mathbf{p})\|$$

per cui, usando la definizione di differenziabilità (e cioè che $\sigma(\mathbf{p}, \mathbf{x}) = o(\|\mathbf{x} - \mathbf{p}\|)$),

$$\lim_{\mathbf{x} \rightarrow \mathbf{p}} f(\mathbf{x}) = f(\mathbf{p}).$$

□

Verificare la differenziabilità in un punto per mezzo della sua definizione può essere a volte convoluto. Questo teorema ci viene in aiuto per risolvere molti casi di uso pratico, fornendo una condizione sufficiente per la differenziabilità.

TEOREMA 2.2 (Teorema del differenziale totale). *Se $f: \mathbb{R}^d \rightarrow \mathbb{R}$ ha derivate parziali in un intorno di \mathbf{p} ed esse sono continue in \mathbf{p} , allora è differenziabile in \mathbf{p} .*

L'esistenza di tutte le derivate direzionali non è purtroppo sufficiente a far dire che una funzione è differenziabile. Per esempio, consideriamo la funzione $f: \mathbb{R}^2 \rightarrow \mathbb{R}$,

$$(2.3) \quad f(\mathbf{x}) \equiv f(x, y) = \begin{cases} \frac{xy^2}{x^2+y^2} & \mathbf{x} \neq \mathbf{0}, \\ 0 & \mathbf{x} = \mathbf{0}, \end{cases}$$

rappresentata in Fig. 1 (sinistra). Calcoliamo la sua derivata lungo la generica direzione $\mathbf{v} = (v_1, v_2)^\top$ in $\mathbf{p} = \mathbf{0}$,

$$\frac{\partial f(\mathbf{0})}{\partial \mathbf{v}} = \lim_{t \rightarrow 0} \frac{f(t\mathbf{v}) - f(\mathbf{0})}{t} = \frac{v_1 v_2^2}{v_1^2 + v_2^2}.$$

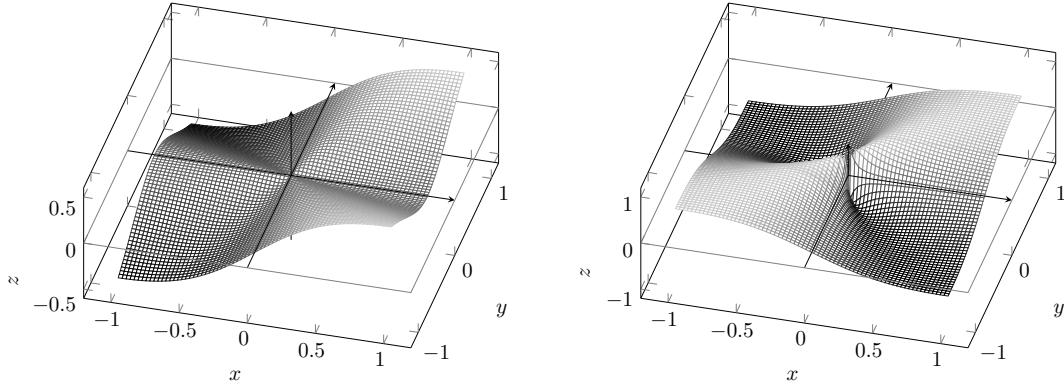


FIGURA 1. Esempi di funzione non differenziabile nell'origine: a sinistra, funzione continua ma non differenziabile, data in Eq. (2.3); a destra, funzione non continua e non differenziabile, data in Eq. (2.4).

In particolare, abbiamo che $\partial_x f(\mathbf{0}) = \partial_y f(\mathbf{0}) = 0$, quindi $\nabla f(\mathbf{0}) = \mathbf{0}$, ma presa una direzione generica la derivata non è nulla: questo significa che la funzione non è differenziabile, dato che $\frac{\partial f(\mathbf{0})}{\partial \mathbf{v}} \neq \sum_i v_i \partial_i f(\mathbf{0})$. Si noti che però la funzione è continua nell'origine: $\lim_{\mathbf{x} \rightarrow \mathbf{0}} f(\mathbf{x}) = f(\mathbf{0}) = 0$. La differenziabilità implica la continuità ma non vale il viceversa.

Consideriamo ora la funzione

$$(2.4) \quad f(\mathbf{x}) \equiv f(x, y) = \begin{cases} \frac{xy}{x^2+y^2} & \mathbf{x} \neq \mathbf{0}, \\ 0 & \mathbf{x} = \mathbf{0}, \end{cases}$$

rappresentata in Fig. 1 (destra). Calcolando la derivata direzionale come nel caso precedente, si trova

$$\frac{\partial f(\mathbf{0})}{\partial \mathbf{v}} = \lim_{t \rightarrow 0} \frac{1}{t} \frac{v_1 v_2}{v_1^2 + v_2^2}.$$

Questo limite è identicamente nullo se v_1 o v_2 sono identicamente nulli, ovvero se stiamo calcolando $\partial_x f(\mathbf{0})$ o $\partial_y f(\mathbf{0})$. Viceversa è *infinito* in tutti gli altri casi. La funzione non è quindi differenziabile. Graficando questa funzione, si può osservare che essa ha effettivamente un comportamento piuttosto complesso vicino all'asse z : La funzione è identicamente nulla lungo l'asse x e lungo l'asse y , ma se ci si avvicina all'asse z secondo una generica direzione $\mathbf{v} = (\cos \theta, \sin \theta)^\top$ si ottiene limite: $\lim_{t \rightarrow 0} f(t\mathbf{v}) = 1/2 \sin(2\theta)$ che dipende da θ , variando tra un massimo valore pari a $1/2$ (lungo la direzione indicata dalla linea rossa) e un minimo pari a $-1/2$ (lungo la direzione della linea bianca): ciò mostra che la funzione è discontinua in $\mathbf{0}$, e quindi non può essere differenziabile.

2.1.1. Derivate di ordine superiore. Possiamo naturalmente iterare la procedura e definire nella stessa maniera derivate parziali di ordine superiore rispetto alla stessa componente, ma soprattutto possiamo anche definire *derivate parziali miste*, che scriviamo

$$\frac{\partial^2 f(\mathbf{p})}{\partial x_j \partial x_i}.$$

In seguito useremo anche, per questa quantità, la notazione compatta $\partial_{ij}^2 f(\mathbf{p})$. Questa scrittura intende che eseguiremo una derivata di f *prima* rispetto a x_i e *poi* rispetto a x_j . Quando in generale si parla di derivate di funzioni a più variabili, quindi, l'ordine è, in linea di principio, importante. Il seguente teorema, che non dimostriamo, mostra che, in alcuni casi, è però possibile ignorare l'ordine di derivazione.

TEOREMA 2.3 (Schwartz). *Sia f una funzione definita nell'aperto $\mathcal{A} \subset \mathbb{R}^d$ e sia essa dotata di derivate miste*

$$\frac{\partial^2 f(\mathbf{p})}{\partial x_j \partial x_i} = \frac{\partial^2 f(\mathbf{p})}{\partial x_i \partial x_j}$$

in un intorno di $\mathbf{p} \in \mathcal{A}$. Se tali derivate sono continue, allora

$$\frac{\partial^2 f(\mathbf{p})}{\partial x_j \partial x_i} = \frac{\partial^2 f(\mathbf{p})}{\partial x_i \partial x_j}.$$

Il teorema può essere applicato a derivate di ordine superiore e permette così di ignorare il problema dell'ordine di derivazione purché sussista la continuità. Se per esempio calcoliamo la derivata n -esima volendo derivare n_1 volte rispetto a x_1 , n_2 volte rispetto a $x_2 \dots$, n_d volte rispetto a x_d , l'ordine delle singole derivate non è importante e possiamo semplicemente scrivere

$$\frac{\partial^{n_1 + \dots + n_d} f(\mathbf{p})}{\partial x_1^{n_1} \cdots \partial x_d^{n_d}}$$

purché sussista la continuità.

2.2. Matrice Jacobiana. La definizione di differenziabilità data sopra si può generalizzare al caso generico di una funzione $\mathbf{f}: \mathbb{R}^d \rightarrow \mathbb{R}^k$: in fondo, una siffatta funzione è semplicemente una collezione di k funzioni, ciascuna del tipo della funzione f studiata sopra. Possiamo quindi scrivere la seguente definizione:

DEFINIZIONE 2.3. *La funzione $\mathbf{f}: \mathcal{A} \rightarrow \mathbb{R}^k$ definita su un aperto $\mathcal{A} \subset \mathbb{R}^d$ è differenziabile in $\mathbf{p} \in \mathcal{A}$ se ciascuna sua componente è differenziabile, ovvero ha gradiente in \mathbf{p} e*

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{f_i(\mathbf{p} + \mathbf{h}) - f_i(\mathbf{p}) - \langle \mathbf{h}, \nabla f_i(\mathbf{p}) \rangle}{\|\mathbf{h}\|} = 0 \quad i = 1, \dots, k.$$

La formula sopra si può scrivere in una maniera più compatta raccogliendo tutti i gradienti in una matrice, detta *matrice jacobiana*,

$$(2.5) \quad \mathbf{J}_{\mathbf{f}}(\mathbf{p}) = \begin{pmatrix} \partial_1 f_1(\mathbf{p}) & \partial_1 f_2(\mathbf{p}) & \dots & \partial_1 f_k(\mathbf{p}) \\ \partial_2 f_1(\mathbf{p}) & \partial_2 f_2(\mathbf{p}) & \dots & \partial_2 f_k(\mathbf{p}) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_n f_1(\mathbf{p}) & \partial_n f_2(\mathbf{p}) & \dots & \partial_n f_k(\mathbf{p}) \end{pmatrix} \in \mathbb{M}_{n,k}(\mathbb{R}).$$

Usando questa matrice, la condizione di differenziabilità si può scrivere in forma matriciale come

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\mathbf{f}(\mathbf{p} + \mathbf{h}) - \mathbf{f}(\mathbf{p}) - \mathbf{h}^\top \mathbf{J}_{\mathbf{f}}(\mathbf{p})}{\|\mathbf{h}\|} = \mathbf{0}.$$

2.3. Derivate di funzioni composte. Esattamente come nel caso di variabili reali, può verificarsi che si abbia a che fare con una funzione composta, ovvero che, detto $\mathcal{A} \subseteq \mathbb{R}^d$ un aperto, $f: \mathcal{A} \rightarrow \mathbb{R}$ venga composta con una funzione $\phi: \mathbb{R} \rightarrow \mathcal{A}$, in modo che $g := f \circ \phi: \mathbb{R} \rightarrow \mathbb{R}$. Questa situazione si presenta in numerose applicazioni: la funzione $\phi(t)$ di argomento reale t può descrivere per esempio l'evoluzione di un punto nello spazio secondo un parametro "temporale" t , in modo che f sia poi qualche funzione della posizione di detto punto. Per calcolare la derivata di $g(t) := f(\phi(t))$ rispetto a t , possiamo utilizzare il seguente teorema.

TEOREMA 2.4. *Sia $\phi: \mathbb{R} \rightarrow \mathcal{A}$ differenziabile in t e f differenziabile in $\mathbf{p} = \phi(t)$. Allora $g = f \circ \phi$ è differenziabile in t e*

$$\frac{d g(t)}{d t} = \sum_{i=1}^d \frac{d \phi_i(t)}{d t} \frac{\partial f(\phi(t))}{\partial x_i} = \langle \dot{\phi}(t), \nabla f(\phi(t)) \rangle,$$

dove $\dot{\phi}(t) = (\frac{d \phi_i(t)}{d t})_i$.

Un caso particolare è dato da dall'eventualità in cui $f: \mathbb{R}^{1+D} \rightarrow \mathbb{R}$ e $\phi(t) = (t, \mathbf{x}(t)) \equiv (t, x_1(t), \dots, x_D(t))$: qui ϕ descrive una dipendenza di tutte le componenti x_i dalla prima coordinata, che svolge il ruolo di “tempo”. Si definisce quindi *derivata totale* di f rispetto a t la quantità

$$\frac{d f(t, \mathbf{x})}{dt} = \frac{\partial f(t, \mathbf{x})}{\partial t} + \sum_{i=1}^d \frac{\partial f(t, \mathbf{x})}{\partial x_i} \frac{d x_i}{dt}.$$

Concludiamo dando la forma più generale di derivata di una funzione composta.

TEOREMA 2.5. *Sia $\mathbf{h}: \mathcal{A} \rightarrow \mathbb{R}^d$ differenziabile in $\mathbf{p} \in \mathcal{A} \subseteq \mathbb{R}^p$ e $\mathbf{f}: \mathbb{R}^d \rightarrow \mathbb{R}^q$ differenziabile in $\mathbf{q} = \mathbf{f}(\mathbf{p})$. Allora $\mathbf{g} = \mathbf{f} \circ \mathbf{h}$ è differenziabile in \mathbf{p} con matrice jacobiana*

$$\mathbf{J}_{\mathbf{g}}(\mathbf{p}) = \mathbf{J}_{\mathbf{f}}(\mathbf{h}(\mathbf{p})) \mathbf{J}_{\mathbf{h}}(\mathbf{p}).$$

2.4. La matrice hessiana: massimi e minimi. Si dice che \mathbf{p} è un massimo relativo per una funzione f se esiste un intorno $\mathcal{I}_r(\mathbf{p})$ tale per cui $f(\mathbf{x}) \leq f(\mathbf{p})$ per ogni punto \mathbf{x} in tale intorno. Viceversa si dice che \mathbf{p} è un minimo relativo per una funzione f se esiste un intorno $\mathcal{I}_r(\mathbf{p})$ tale per cui $f(\mathbf{x}) \geq f(\mathbf{p})$ per ogni punto \mathbf{x} in tale intorno. Un punto di massimo o minimo relativo di una funzione f differenziabile in \mathbf{p} può essere identificato grazie al fatto che il gradiente in esso è uguale al vettore nullo.

TEOREMA 2.6. *Sia $f: \mathcal{A} \rightarrow \mathbb{R}$, con \mathcal{A} aperto, e $\mathbf{p} \in \mathcal{A}$ punto di massimo o minimo relativo. Se f è differenziabile in \mathbf{p} , allora*

$$\nabla f(\mathbf{p}) = \mathbf{0}.$$

DIMOSTRAZIONE. Supponiamo che \mathbf{p} sia un massimo. Dato un generico versore \mathbf{v} , allora $g(t) := f(\mathbf{p} + t\mathbf{v})$ è definita in un intorno di \mathbf{p} per t sufficientemente piccolo. Essendo \mathbf{p} un massimo, questo significa che $\dot{g}(0) = 0$, ovvero $\langle \mathbf{v}, \nabla f(\mathbf{p}) \rangle = 0$. Essendo \mathbf{v} generico, segue che $\nabla f(\mathbf{p}) = \mathbf{0}$. \square

Notare che possono esistere punti \mathbf{p} in cui $\nabla f(\mathbf{p}) = \mathbf{0}$ e che tuttavia *non* sono di massimo o minimo. Si tratta dei cosiddetti *punti di sella*: in caso di differenzibilità quindi, il teorema fornisce una condizione necessaria ma non sufficiente per identificare punti estremali. La condizione $\nabla f(\mathbf{p}) = \mathbf{0}$ si dice *di stazionarietà*, e \mathbf{p} si dice punto stazionario.

Consideriamo la funzione $f: \mathbb{R}^2 \rightarrow \mathbb{R}$,

$$(2.6) \quad f(\mathbf{x}) \equiv f(x, y) = \sin(x) \sin(y).$$

Se ne calcoliamo il gradiente

$$\nabla f(\mathbf{x}) = (\cos(x) \sin(y), \sin(x) \cos(y))$$

e cherchiamo i punti \mathbf{p} tali per cui $\nabla f(\mathbf{p}) = \mathbf{0}$, possiamo vedere che essi sono nelle posizioni

$$(x, y) = \left(n \frac{\pi}{2}, m \frac{\pi}{2} \right), \quad n, m \in \mathbb{Z},$$

vedasi Fig. 2. Alcuni di questi sono di massimo relativo (corrispondono ai casi in cui m ed n sono entrambi dispari e concordi), altri sono di minimo relativo (corrispondono ai casi in cui m ed n sono entrambi dispari e discordi), altri ancora sono *di sella* (tutti gli altri casi), ovvero né di massimo né di minimo. Uno di questi è proprio l'origine.

Per *distinguere* massimi e minimi occorre ricorrere a derivate di ordine superiore, esattamente come accadeva nel caso unidimensionale, e introdurre un oggetto nuovo detta *matrice hessiana*, che è la matrice $d \times d$ di tutte le derivate seconde di f in un punto,

$$\text{Hess}[f](\mathbf{x}) := \left(\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} \right)_{ij} = \begin{pmatrix} \partial_{11}^2 f(\mathbf{x}) & \partial_{12}^2 f(\mathbf{x}) & \dots & \partial_{1n}^2 f(\mathbf{x}) \\ \partial_{21}^2 f(\mathbf{x}) & \partial_{22}^2 f(\mathbf{x}) & \dots & \partial_{2n}^2 f(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_{n1}^2 f(\mathbf{x}) & \partial_{n2}^2 f(\mathbf{x}) & \dots & \partial_{nn}^2 f(\mathbf{x}) \end{pmatrix}.$$

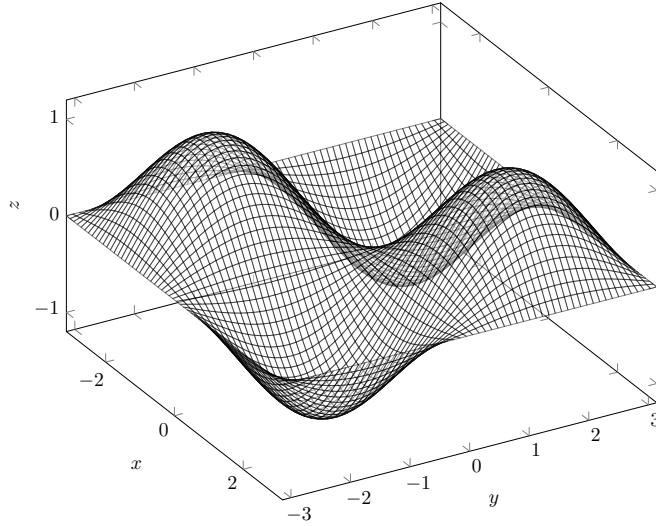


FIGURA 2. Funzione in Eq. (2.6).

Notare che questa è un matrice simmetrica reale, quindi diagonalizzabile. Vale il seguente teorema.

TEOREMA 2.7. *Sia $f: \mathcal{A} \rightarrow \mathbb{R}$ continua con derivate fino al secondo ordine continue sull'aperto \mathcal{A} . Se \mathbf{p} è un punto stazionario, esso è un minimo relativo se la matrice hessiana $\text{Hess}[f](\mathbf{p})$ ha solo autovalori positivi, ovvero è definita positiva; viceversa è di massimo relativo se ha solo autovalori negativi.*

Nei casi in cui esistano sia autovalori positivi che negativi, o autovalori nulli, il punto è detto *di sella*, e non è né di massimo né di minimo. Se siamo interessati ai punti di massimo e minimo *assoluto* di una funzione definita su un dominio \mathcal{D} , questi andranno cercati tra i punti stazionari, i punti in cui il gradiente non esiste e i punti di frontiera (dato che per questi ultimi due gruppi i criteri sopra non si applicano).

Concludiamo osservando che la matrice hessiana permette di scrivere una espansione di Taylor per una funzione f attorno ad un punto, in analogia col caso unidimensionale.

PROPOSIZIONE 2.8 (Formula di Taylor). *Sia f una funzione con derivate continue fino al secondo ordine in un aperto $\mathcal{A} \in \mathbb{R}^d$. In un intorno di $\mathbf{p} \in \mathcal{A}$ possiamo scrivere*

$$(2.7) \quad f(\mathbf{x}) = f(\mathbf{p}) + \langle \mathbf{x} - \mathbf{p}, \nabla f(\mathbf{p}) \rangle + \frac{1}{2} + \langle \mathbf{x} - \mathbf{p}, \text{Hess}[f](\mathbf{p})(\mathbf{x} - \mathbf{p}) \rangle + R(\mathbf{x}, \mathbf{p}),$$

dove $R(\mathbf{x}, \mathbf{p}) = o(\|\mathbf{x} - \mathbf{p}\|^2)$, ovvero $\lim_{\mathbf{x} \rightarrow \mathbf{p}} \frac{R(\mathbf{x}, \mathbf{p})}{\|\mathbf{x} - \mathbf{p}\|^2} = 0$.

3. Misura e integrazione su \mathbb{R}^d

3.1. Plurintervalli e misura secondo Peano–Jordan. Per introdurre l'*integrazione* su un dominio d -dimensionale, rivedremo i concetti di intervallo e misura, secondo una elaborazione data nella seconda metà dell'Ottocento da Giuseppe Peano e Camille Jordan. Il punto di partenza per studiare la teoria dell'integrazione in \mathbb{R}^d è, esattamente come nel caso $d = 1$, l'introduzione di una unità semplice la cui "misura" è di facile definizione. Nel caso unidimensionale, tale unità era l'*intervallo*, ovvero un insieme del tipo $[a, b] := \{x \in \mathbb{R}: a \leq x < b\} \subset \mathbb{R}$, la cui misura era

data da $b - a$. Nel caso di dimensione più alta possiamo pensare a “rettangoli d -dimensionali”, e dare la seguente definizione.

DEFINIZIONE 3.1 (Intervallo). *Un intervallo d -dimensionale $\mathcal{R} \subset \mathbb{R}^d$ ha la forma*

$$\mathcal{R} := \{\mathbf{x} \in \mathbb{R}^d : a_i \leq x_i < b_i \text{ } i = 1, \dots, d\},$$

dove, per ogni $i = 1, \dots, d$, $a_i, b_i \in \mathbb{R}$ con $a_i \leq b_i$. La sua misura $\mathfrak{m}(\mathcal{R})$ è definita come

$$\mathfrak{m}(\mathcal{R}) = (b_1 - a_1)(b_2 - a_2) \dots (b_d - a_d) \geq 0.$$

Un intervallo in \mathbb{R}^d si può anche scrivere come

$$\mathcal{R} = [a_1, b_1] \times [a_2, b_2] \times \dots [a_d, b_d],$$

interpretandolo cioè come *prodotto cartesiano* di intervalli in \mathbb{R} . In questo modo, per esempio, se \mathcal{R}_1 è un intervallo in \mathbb{R}^p e \mathcal{R}_2 è un intervallo in \mathbb{R}^q , allora $\mathcal{R} := \mathcal{R}_1 \times \mathcal{R}_2$ è un intervallo in \mathbb{R}^{p+q} e la sua misura è

$$\mathfrak{m}(\mathcal{R}) = \mathfrak{m}(\mathcal{R}_1)\mathfrak{m}(\mathcal{R}_2).$$

Immaginiamo ora che, in \mathbb{R}^d , per ogni componente $i = 1, \dots, d$, si scelgano k_i numeri reali distinti $\{a_{ij}\}_{j=1}^{k_i}$: con questi numeri, costruiamo un insieme di $k := \sum_{i=1}^d k_i$ iperpiani dati dalle equazioni

$$x_i = a_{ij}, \quad j = 1, \dots, k_i, \quad i = 1, \dots, d.$$

Questi iperpiani formano un *reticolato* \mathcal{P} di \mathbb{R}^d , dividendolo in porzioni, alcune delle quali limitate: tali cellette sono degli intervalli definiti come sopra, e che saranno detti gli intervalli associati al reticolato \mathcal{P} . A questo punto, possiamo dare la seguente definizione.

DEFINIZIONE 3.2 (Plurintervallo). *Un insieme $\mathcal{Y} \subset \mathbb{R}^d$ è detto plurintervallo se esiste un reticolato \mathcal{P} tale per cui \mathcal{Y} è l'unione di n intervalli di \mathcal{P} ,*

$$\mathcal{Y} = \bigcup_{i=1}^n \mathcal{R}_i,$$

per un qualche n finito. La sua misura è definita come $\mathfrak{m}(\mathcal{P}) = \sum_{i=1}^n \mathfrak{m}(\mathcal{R}_i)$. Diremo che \mathcal{P} induce la partizione di \mathcal{Y} negli intervalli $\{\mathcal{R}_i\}_{i=1}^k$.

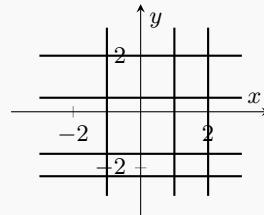
Consideriamo \mathbb{R}^2 e costruiamo un reticolato scegliendo, per esempio,

$$x = -1, \quad x = 1, \quad x = 2,$$

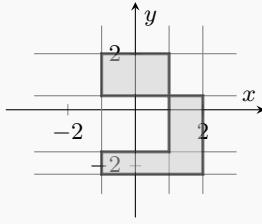
e

$$y = -7/3, \quad y = -3/2, \quad y = 1/3, \quad y = 2.$$

Gli “iperpiani” sono delle rette come in figura



Possiamo individuare sei intervalli associati a questo reticolato. La regione colorata in figura è un plurintervallo:



La definizione sopra generalizza quanto fatto in $d = 1$. Un plurintervallo può essere rappresentato a partire da reticolati diversi: per esempio, se è dato in termini degli intervalli di un reticolato \mathcal{P} , lo si potrà dare anche in termini degli intervalli di un *raffinamento* \mathcal{P}' di \mathcal{P} , ovvero un reticolato \mathcal{P}' che contiene tutti gli iperpiani di \mathcal{P} più altri aggiuntivi. La misura di \mathcal{Y} però non dipenderà dal reticolato scelto. Dati due plurintervalli \mathcal{Y}_1 e \mathcal{Y}_2 , abbiamo che $\mathfrak{m}(\mathcal{Y}_1 \cup \mathcal{Y}_2) \leq \mathfrak{m}(\mathcal{Y}_1) + \mathfrak{m}(\mathcal{Y}_2)$ (se $\mathcal{Y}_1 \cap \mathcal{Y}_2 = \emptyset$ vale l'uguaglianza).

Avendo definito la misura di intervalli e plurintervalli, possiamo utilizzare questi oggetti come strumenti per definire la misura di insiemi qualsiasi. Per esempio possiamo dare la seguente definizione.

DEFINIZIONE 3.3. *Sia $\mathcal{A} \subset \mathbb{R}^d$. Si dice che esso ha misura nulla secondo Peano–Jordan se, per ogni $\varepsilon > 0$, esiste un plurintervallo \mathcal{Y}_ε tale che*

$$\mathcal{A} \subseteq \mathcal{Y}_\varepsilon, \quad \mathfrak{m}(\mathcal{Y}_\varepsilon) \leq \varepsilon.$$

Questo vuol dire per esempio che un singolo punto ha misura zero. Inoltre, se $\mathcal{B} \subseteq \mathcal{A}$ e \mathcal{A} ha misura nulla, allora \mathcal{B} ha misura nulla, e, infine, se $\{\mathcal{A}_k\}_{k=1}^K$ è una lista finita di K insiemi di misura nulla, $\mathfrak{m}(\bigcup_{k=1}^K \mathcal{A}_k) = 0$.

La definizione di plurintervallo ci permette di definire la misura di un *generico insieme* $\mathcal{A} \subseteq \mathbb{R}^d$.

DEFINIZIONE 3.4 (Misura di Jordan). *Sia $\mathcal{A} \subseteq \mathbb{R}^d$. Definiamo la sua misura interna di Jordan*

$$(2.8) \quad \underline{\mathfrak{m}}(\mathcal{A}) = \sup_{\mathcal{Y}} \{\mathfrak{m}(\mathcal{Y}) : \mathcal{Y} \subset \mathcal{A} \text{ plurintervallo}\},$$

e la sua misura esterna di Jordan

$$(2.9) \quad \overline{\mathfrak{m}}(\mathcal{A}) = \inf_{\mathcal{Y}} \{\mathfrak{m}(\mathcal{Y}) : \mathcal{A} \subset \mathcal{Y} \text{ plurintervallo}\}.$$

Diciamo che \mathcal{A} è misurabile secondo Peano–Jordan se

$$(2.10) \quad \underline{\mathfrak{m}}(\mathcal{A}) = \overline{\mathfrak{m}}(\mathcal{A}) =: \mathfrak{m}(\mathcal{A}).$$

La definizione data implica che se \mathcal{A} è un insieme limitato, esso è misurabile secondo Peano–Jordan se e solo se la sua frontiera ha misura nulla. La definizione sopra infatti è equivalente a dire che, $\forall \varepsilon > 0$, esistono due plurintervalli $\mathcal{Y}_i, \mathcal{Y}_e$ tali che $\mathcal{Y}_e \subset \mathcal{A} \subset \mathcal{Y}_i$, così che $\mathfrak{m}(\mathcal{Y}_e) - \mathfrak{m}(\mathcal{Y}_i) < \varepsilon$: ma questa è proprio la misura di $\mathcal{Y}_e \setminus \mathcal{Y}_i$, che è l'insieme in cui vive la frontiera. La definizione sopra quindi ha come conseguenza la proposizione

PROPOSIZIONE 3.1 (Misurabilità secondo Peano–Jordan). *Un insieme $\mathcal{A} \subseteq \mathbb{R}^d$ è misurabile secondo Peano–Jordan se e solo se, per ogni $\varepsilon > 0$, esiste un plurintervallo \mathcal{Z} con $\mathfrak{m}(\mathcal{Z}) = \varepsilon$ tale che $\partial\mathcal{A} \subseteq \mathcal{Z}$, ovvero $\mathfrak{m}(\partial\mathcal{A}) = 0$.*

DEFINIZIONE 3.5 (Unione quasi disgiunta). *Si dice che un insieme \mathcal{A} è una unione quasi disgiunta di K insiemi misurabili $\{\mathcal{A}_k\}_{k=1}^K$ se $\mathcal{A} = \bigcup_{k=1}^K \mathcal{A}_k$ e dati, due insiemi \mathcal{A}_k e $\mathcal{A}_{k'}$ della lista, $\mathfrak{m}(\mathcal{A}_k \cap \mathcal{A}_{k'}) = 0$.*

Di conseguenza $\mathfrak{m}(\mathcal{A}) = \sum_{k=1}^K \mathfrak{m}(\mathcal{A}_k)$ se si tratta di un'unione quasi disgiunta.

Chiudiamo questa rassegna con il seguente teorema nel caso di prodotti cartesiani di insiemi.

TEOREMA 3.2 (Misura di prodotti cartesiani). *Siano $\mathcal{A} \subset \mathbb{R}^p$ e $\mathcal{B} \subset \mathbb{R}^q$ due insiemi misurabili secondo Peano–Jordan. Allora $\mathcal{A} \times \mathcal{B} \subset \mathbb{R}^{p+q}$ è misurabile secondo Peano–Jordan e*

$$\mathfrak{m}(\mathcal{A} \times \mathcal{B}) = \mathfrak{m}(\mathcal{A})\mathfrak{m}(\mathcal{B}).$$

3.2. Integrazione secondo Riemann. Avendo definito la misura di insiemi in \mathbb{R}^d , possiamo introdurre l'integrazione di una funzione f su tali insiemi secondo la teoria dovuta a Bernhard Riemann.

3.2.1. *Integrazione su plurintervalli.* Consideriamo per cominciare una funzione $f: \mathcal{Y} \rightarrow \mathbb{R}$, limitata su un plurintervallo $\mathcal{Y} \subset \mathbb{R}^d$. Sia \mathcal{P} un reticolo che induce una partizione di \mathcal{Y} in intervalli quasi disgiunti, e scriviamo $\mathcal{Y} = \bigcup_{i=1}^k \mathcal{R}_i$. Possiamo definire due quantità, ovvero

$$L_{\mathcal{Y}}[f; \mathcal{P}] := \sum_{i=1}^k \mathfrak{m}(\mathcal{R}_i) \min_{\mathbf{x} \in \mathcal{R}_i} f(\mathbf{x}), \quad U_{\mathcal{Y}}[f; \mathcal{P}] := \sum_{i=1}^k \mathfrak{m}(\mathcal{R}_i) \max_{\mathbf{x} \in \mathcal{R}_i} f(\mathbf{x}).$$

Queste due quantità dipendono dal reticolato che abbiamo usato per partizionare \mathcal{Y} , ma possiamo tentare di essere quanto più raffinati possibile e definire

$$(2.11) \quad L_{\mathcal{Y}}[f] := \sup_{\mathcal{P}} L_{\mathcal{Y}}[f; \mathcal{P}], \quad U_{\mathcal{Y}}[f] := \inf_{\mathcal{P}} U_{\mathcal{Y}}[f; \mathcal{P}].$$

Se $L_{\mathcal{Y}}[f] = U_{\mathcal{Y}}[f] =: I_{\mathcal{Y}}[f]$, allora si dice che f è integrabile su \mathcal{Y} e scriviamo

$$(2.12) \quad I_{\mathcal{Y}}[f] = \int_{\mathcal{Y}} f(\mathbf{x}) d^d x.$$

Le definizioni date per i plurintervalli possono facilmente estendersi a insiemi qualsiasi \mathcal{A} su cui f sia definita e che siano limitati. Basta infatti considerare un plurintervallo \mathcal{Y} tale per cui $\mathcal{A} \subseteq \mathcal{Y}$ e quindi introdurre la funzione

$$\hat{f}(\mathbf{x}) := f(\mathbf{x})1_{\mathcal{A}}(\mathbf{x}), \quad 1_{\mathcal{A}}(\mathbf{x}) = \begin{cases} 1 & \text{se } \mathbf{x} \in \mathcal{A} \\ 0 & \text{se } \mathbf{x} \in \mathcal{Y} \setminus \mathcal{A}. \end{cases}$$

La funzione $1_{\mathcal{A}}$ si chiama *funzione indicatrice di \mathcal{A}* . Allora per definizione f è integrabile su \mathcal{A} se esiste \mathcal{Y} plurintervallo contenente \mathcal{A} su cui lo è \hat{f}

$$\int_{\mathcal{A}} f(\mathbf{x}) d^d x := \int_{\mathcal{Y}} \hat{f}(\mathbf{x}) d^d x.$$

La definizione di integrale definito che abbiamo dato ha come conseguenze una serie di proprietà. Date due funzioni f e g integrabili su un insieme $\mathcal{A} \subset \mathbb{R}^d$ misurabile, si ha che

Linerarità: $\int_{\mathcal{A}} [c_1 f(\mathbf{x}) + c_2 g(\mathbf{x})] d^d x = c_1 \int_{\mathcal{A}} f(\mathbf{x}) d^d x + c_2 \int_{\mathcal{A}} g(\mathbf{x}) d^d x$ per $c_1, c_2 \in \mathbb{R}$;

Additività: se $\mathcal{A} = \bigcup_{k=1}^K \mathcal{A}_k$, unione quasi disgiunta di insiemi misurabili, allora $\int_{\mathcal{A}} f(\mathbf{x}) d^d x = \sum_{k=1}^K \int_{\mathcal{A}_k} f(\mathbf{x}) d^d x$;

Monotonicità: se $f(\mathbf{x}) \leq g(\mathbf{x}) \forall \mathbf{x} \in \mathcal{A}$, allora $\int_{\mathcal{A}} f(\mathbf{x}) d^d x \leq \int_{\mathcal{A}} g(\mathbf{x}) d^d x$;

Valore assoluto: $|\int_{\mathcal{A}} f(\mathbf{x}) d^d x| \leq \int_{\mathcal{A}} |f(\mathbf{x})| d^d x$.

Infine, la misura stessa di un insieme misurabile \mathcal{A} si può scrivere in forma di integrale: per definizione infatti, possiamo scrivere

$$\mathfrak{m}(\mathcal{A}) = \int_{\mathcal{A}} d^d x.$$

Dato un sottoinsieme $\mathcal{B} \subseteq \mathcal{A}$, possiamo calcolarne la misura usando la funzione indicatrice $1_{\mathcal{B}}$ e scrivendo $\mathfrak{m}(\mathcal{B}) = \int_{\mathcal{A}} 1_{\mathcal{B}}(\mathbf{x}) d^d x$.

Notare che, invece di $\int_{\mathcal{A}} f(\mathbf{x}) d^d x$, si usa a volte la notazione $\int_{\mathcal{A}} f(x_1, \dots, x_d) d x_1 \dots d x_d$ per sottolineare la dipendenza dalle diverse componenti. Se $d = 2$, si scrive a volte $\iint_{\mathcal{A}} f(x, y) d x d y$; se $d = 3$, in maniera simile, si scrive talvolta $\iiint_{\mathcal{A}} f(x, y, z) d x d y d z$.

3.2.2. Il teorema di Fubini. La definizione di integrale data sopra non aiuta particolarmente, eccetto che in rari e semplici casi, a calcolare effettivamente il valore dell'integrale stesso. Il teorema di Guido Fubini aiuta a trasformare, in molti casi, il problema dell'integrazione di una funzione in più variabili in un problema di integrazione successiva di singole variabili. Enunceremo questo teorema, detto anche *di riduzione*, per una funzione definita su un aperto $\mathcal{A} \subseteq \mathbb{R}^2$.

TEOREMA 3.3 (Fubini). *Sia $f: \mathcal{A} \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ integrabile sul rettangolo $\mathcal{R} = [a_1, b_1] \times [a_2, b_2]$. Supponiamo che, ad $x \in [a_1, b_1]$ fissata, la funzione $f(x, y)$ sia integrabile in y su $[a_2, b_2]$. Allora*

$$F(x) := \int_{a_2}^{b_2} f(x, y) d y$$

è integrabile su $[a_1, b_1]$ e in particolare possiamo scrivere

$$\int_{\mathcal{R}} f(\mathbf{x}) d^2 x = \int_{a_1}^{b_1} F(x) d x = \int_{a_1}^{b_1} d x \int_{a_2}^{b_2} d y f(x, y).$$

In altre parole, il teorema ci permette di integrare “una variabile per volta” eseguendo quindi integrali semplici unidimensionali. Il teorema si può estendere a insiemi *normali* rispetto all'asse x , ovvero nella forma

$$\mathcal{E} = \{(x, y) \in \mathbb{R}^2 : a \leq x \leq b, g(x) \leq y \leq h(x)\}$$

con g e h funzioni continue su $[a, b]$. In tal caso

$$\int_{\mathcal{E}} f(\mathbf{x}) d^2 x = \int_a^b d x \int_{g(x)}^{h(x)} d y f(x, y).$$

Il teorema di Fubini si può estendere al caso di tre o più variabili. Sia per esempio $f: A \subseteq \mathbb{R}^3 \rightarrow \mathbb{R}$ integrabile su $\mathcal{R} = [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3] \subseteq \mathcal{A}$ e sia $\mathcal{R}' = [a_1, b_1] \times [a_2, b_2] \subset \mathbb{R}^2$. Allora

$$\int_{\mathcal{R}} f(\mathbf{x}) d^3 x = \int_{\mathcal{R}'} d y d z \int_{a_1}^{b_1} f(x, y, z) d x,$$

e similmente per insiemi normali rispetto all'asse z

$$\mathcal{E} = \{(x, y, z) \in \mathbb{R}^3 : (x, y) \in \mathcal{R}', g(x, y) \leq z \leq h(x, y)\}$$

con g e h funzioni continue su $\mathcal{R}' \subset \mathbb{R}^2$ intervallo, allora

$$\int_{\mathcal{E}} f(\mathbf{x}) d^d x = \int_{\mathcal{R}'} d x d y \int_{g(x,y)}^{h(x,y)} d z f(x, y, z).$$

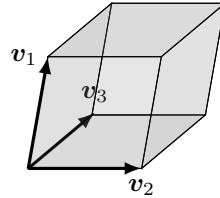
Supponiamo di avere la seguente funzione in $d = 2$,

$$f(\mathbf{x}) = f(x, y) = \frac{1}{(x+y)^2}$$

e supponiamo di volerne calcolare l'integrale su $\mathcal{R} = [3, 4] \times [1, 2]$. Il teorema di Fubini semplifica molto il calcolo, dato che

$$\int_{\mathcal{R}} \frac{1}{(x+y)^2} dx dy = \int_3^4 dx \int_1^2 dy \frac{1}{(x+y)^2} = \int_3^4 dx \left(\frac{1}{x+1} - \frac{1}{x+2} \right) = \ln \frac{x+1}{x+2} \Big|_{x=3}^{x=4} = \ln \frac{25}{24}.$$

3.2.3. Volume di un parallelepipedo d -dimensionale. Abbiamo visto che è molto intuitivo assegnare un volume ad un intervallo. La situazione è un po' più complessa quando si considera un parallelepipedo in d dimensioni, ovvero un volume generato da d vettori $\mathbf{v}_1, \dots, \mathbf{v}_d \in \mathbb{R}^d$, ciascuno rappresentante un lato del parallelepipedo, da immaginarsi con un vertice nell'origine.

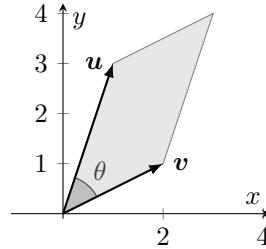


Ciò che si può dimostrare è che il volume V di tale parallelepipedo è uguale al modulo del *determinante* della matrice avente come colonne (o righe) i vettori che lo generano,

$$V = |\det(\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_d)|.$$

Come atteso, se due o più vettori sono linearmente dipendenti il volume è nullo, in quanto questo vorrebbe dire che il parallelepipedo difetta di estensione in una o più dimensioni.

Possiamo verificare facilmente questo risultato per $d = 2$ e $d = 3$. In $d = 2$, supponiamo di avere due vettori $\mathbf{v} = (v_1, v_2)^\top$ e $\mathbf{u} = (u_1, u_2)^\top$.



In questo caso, il “parallelepipedo” è un parallelogramma. Con riferimento alla figura sopra, la sua area A si può scrivere come $\|\mathbf{u}\| \|\mathbf{v}\| |\sin \theta|$. Questa quantità può essere interpretata come il modulo di un prodotto vettoriale tra due vettori in \mathbb{R}^3 nella forma

$$\hat{\mathbf{v}} = \begin{pmatrix} v_1 \\ v_2 \\ 0 \end{pmatrix}, \quad \hat{\mathbf{u}} = \begin{pmatrix} u_1 \\ u_2 \\ 0 \end{pmatrix} \Rightarrow \hat{\mathbf{v}} \wedge \hat{\mathbf{u}} = \det \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ v_1 & v_2 & 0 \\ u_1 & u_2 & 0 \end{pmatrix} = \mathbf{e}_3 \det \begin{pmatrix} v_1 & u_1 \\ v_2 & u_2 \end{pmatrix} \Rightarrow A = |\det \begin{pmatrix} v_1 & u_1 \\ v_2 & u_2 \end{pmatrix}|,$$

che mostra quanto intendevamo.

In $d = 3$, siano $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ tre vettori non complanari. Scelti, per esempio, \mathbf{v} e \mathbf{u} come vettori “di base”, l’area di base del parallelepipedo è quindi $\|\mathbf{u} \wedge \mathbf{v}\|$. D’altra parte $\mathbf{u} \wedge \mathbf{v}$ è un vettore normale alla base, e quindi $\langle \mathbf{w}, \mathbf{u} \wedge \mathbf{v} \rangle = \|\mathbf{w}\| \|\mathbf{u} \wedge \mathbf{v}\| \cos \theta$, dove θ è l’angolo tra \mathbf{w} e la superficie di base, ovvero $\|\mathbf{w}\| \cos \theta$ è l’altezza del parallelepipedo. In definitiva cioè il volume V del parallelepipedo è

$$V = |\langle \mathbf{w}, \mathbf{u} \wedge \mathbf{v} \rangle| = \left| \det \begin{pmatrix} v_1 & u_1 & w_1 \\ v_2 & u_2 & w_2 \\ v_3 & u_3 & w_3 \end{pmatrix} \right|.$$

3.3. Cambio di variabili. Le formule date sopra sono fornite rispetto ad un set di coordinate \mathbf{x} date rispetto ad un riferimento ortogonale fissato. In questa sezione, cercheremo di capire come passare da un vecchio sistema di riferimento associato a coordinate $\mathbf{x} = (x_i)_{i=1}^d$ ad un nuovo insieme di variabili $\mathbf{u} = (u_i)_{i=1}^d$, utilizzando una funzione $\mathbf{s}: \mathbb{R}^d \rightarrow \mathbb{R}^d$ tale che¹

$$(2.13) \quad \mathbf{x} = \mathbf{s}(\mathbf{u}).$$

Il nuovo insieme di variabili *non corrisponde necessariamente ad una base ortonormale*, ed in generale non corrisponde ad una base dello spazio: può essere intesa come una rappresentazione diversa dei punti di \mathbb{R}^d , tale per cui ogni \mathbf{u} sia associato ad uno e un solo punto di \mathbb{R}^d . Si richiede perciò $\mathbf{s}: \mathbb{R}^d \rightarrow \mathbb{R}^d$ sia un *diffeomorfismo*, ovvero deve essere invertibile (e quindi ogni \mathbf{u} corrisponde ad uno e un solo \mathbf{x} e viceversa) e differenziabile con inversa differenziabile. Questo significa che

$$\mathbf{u} = \mathbf{s}^{-1}(\mathbf{s}(\mathbf{u})) \Rightarrow \mathbf{I}_d = \mathbf{J}_{\mathbf{s}^{-1}}(\mathbf{s}(\mathbf{u}))\mathbf{J}_{\mathbf{s}}(\mathbf{u}) \Rightarrow \mathbf{J}_{\mathbf{s}^{-1}}(\mathbf{x}) = \mathbf{J}_{\mathbf{s}}^{-1}(\mathbf{u}),$$

ovvero la matrice jacobiana di \mathbf{s} deve essere invertibile in ogni \mathbf{u} . Il seguente teorema ci dice come eseguire il cambio di variabili nell'esecuzione di una integrazione.

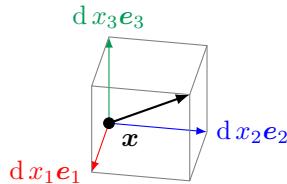
TEOREMA 3.4 (Cambio di variabili). *Sia $\mathbf{s}: \mathbb{R}^d \rightarrow \mathbb{R}^d$ un diffeomorfismo, e $\mathcal{A} \subseteq \mathbb{R}^d$. Chiamiamo*

$$\mathbf{s}(\mathcal{A}) := \{\mathbf{x} \in \mathbb{R}^d : \exists \mathbf{u} \in \mathcal{A} \text{ tale che } \mathbf{x} = \mathbf{s}(\mathbf{u})\}.$$

Allora vale la seguente formula di cambio di variabili

$$\int_{\mathbf{s}(\mathcal{A})} f(\mathbf{x}) d^d x = \int_{\mathcal{A}} f(\mathbf{s}(\mathbf{u})) |\det \mathbf{J}_{\mathbf{s}}(\mathbf{u})| d^d u.$$

Come si vede, oltre all'attesa sostituzione nell'argomento della funzione integranda e nel dominio di integrazione, nella formula comprare uno *jacobiano*. Possiamo dare una ragione intuitiva di questa formula, senza dimostrarla. Immaginiamo di fissare un punto \mathbf{x} nel nostro dominio di integrazione e considerare il rettangolo infinitesimo $\mathcal{R} = [x_1, x_1 + dx_1] \times \cdots \times [x_d, x_d + dx_d]$:



Questo volumetto ha misura $\prod_{i=1}^d dx_i$. Per ogni direzione i abbiamo

$$dx_i = ds_i(\mathbf{u}) = \sum_{j=1}^d du_j \frac{\partial s_i(\mathbf{u})}{\partial u_j}.$$

In generale, quindi, stiamo eseguendo *localmente a livello infinitesimo* una trasformazione lineare: un vero e proprio cambio di base. Osservando che du_j sono infinitesimi corrispondenti alla nuova base (non necessariamente ortogonale), l'espressione fornisce una rappresentazione dei vettori che generano un parallelepipedo nel nuovo riferimento, per cui il volume dell'infinitesimo può calcolarsi prendendo il determinante della matrice avente come colonne tali vettori $\mathbf{v}_i = (\partial_j s_i(\mathbf{u}) du_j)_j$, che fornisce esattamente l'espressione $|\det \mathbf{J}_{\mathbf{s}}(\mathbf{u})| d^d u$.

¹Si noti che questo tipo di cambiamento di variabile è più generale di quello che abbiamo visto nel capitolo precedente, quando abbiamo studiato cambi di *base ortogonale* per l'intero spazio, che corrispondevano a trasformazioni $\mathbf{x} = \mathbf{S}\mathbf{u}$ con $\mathbf{S} \in \mathrm{O}(d)$: una trasformazione tipo quella in Eq. (2.13) *non* corrisponde in generale ad un cambio di base dello spazio.

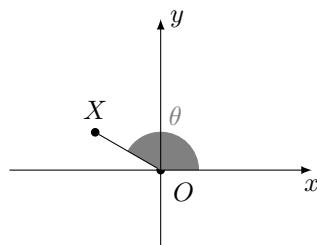
3.3.1. Coordinate polari. Se i riferimenti cartesiani, ovvero fondati sull'uso una base ortonormale, sono senz'altro i più "standard" in letteratura, esistono però altri modi di rappresentare i punti nel piano e nello spazio che possono presentare importanti vantaggi e sono per questo molto utilizzati. Un esempio importante sono le coordinate polari nel piano. Supponiamo quindi di essere nel piano, e di avere su esso il nostro usuale riferimento cartesiano, in modo che ogni punto abbia coordinate $\mathbf{x} = (x, y)^\top$. Introduciamo ora due nuove quantità, ovvero

$$\rho(x, y) = \sqrt{x^2 + y^2} = \|\mathbf{x}\| > 0,$$

che è detto *modulo* di \mathbf{x} e misura la distanza di \mathbf{x} dall'origine, e l'*anomalia*

$$\theta(x, y) = \text{sign}(y) \arccos \frac{x}{\sqrt{x^2 + y^2}} \in (-\pi, \pi],$$

angolo tra \mathbf{x} e la semiretta orientata con origine O e direzione e verso determinati dall'asse delle ascisse. Le coordinate polari di \mathbf{x} sono costituite esattamente dalla coppia (ρ, θ) .



Viceversa, data la coppia (ρ, θ) di coordinate polari di un punto nel piano, le sue coordinate cartesiane possono scriversi come

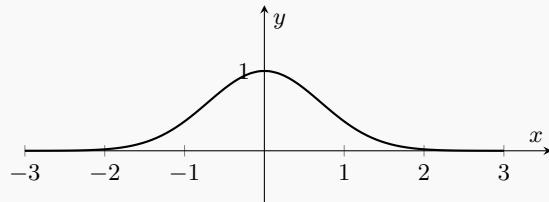
$$\begin{cases} x = \rho \cos \theta, \\ y = \rho \sin \theta. \end{cases}$$

Lo jacobiano della trasformazione si ottiene subito come

$$|\mathbf{J}_s(\rho, \theta)| = \begin{vmatrix} \cos \theta & -\rho \sin \theta \\ \sin \theta & \rho \cos \theta \end{vmatrix} = \rho \Rightarrow dxdy = \rho d\rho d\theta.$$

Il punto O ha modulo 0 e anomalia indeterminata: questo punto fa sì che il cambio di variabili da cartesiane a polari *non* goda della proprietà di biettività richiesta in un diffeomorfismo. Tuttavia, dato che questa condizione viene meno per il solo punto all'origine, dal punto di vista pratico non costituisce un problema.

Sia data la funzione $f(x) = e^{-x^2}$ su \mathbb{R} .



Calcoliamone l'integrale sui reali,

$$I = \int_{-\infty}^{+\infty} e^{-x^2} dx.$$

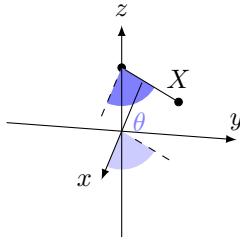
Il calcolo di questa quantità è di molto semplificato se calcoliamo piuttosto

$$I^2 = \left(\int_{-\infty}^{+\infty} e^{-x^2} dx \right)^2 = \int_{-\infty}^{+\infty} e^{-x^2-y^2} dx dy = \int_0^{\infty} \rho d\rho \int_0^{2\pi} d\theta e^{-\rho^2} = \pi,$$

dove abbiamo eseguito un cambio di variabili passando ad un riferimento polare. Abbiamo così che

$$I = \int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}.$$

3.3.2. Coordinate cilindriche e volumi di rotazione. Nello spazio euclideo tridimensionale con un riferimento cartesiano $Oe_1e_2e_3$ è possibile individuare un punto $X(x)$ utilizzando la sua distanza ρ dall'asse z , la sua quota z e l'angolo θ che esso forma con la direzione positiva dell'asse delle ascisse x .



Questo sistema di coordinate è detto *cilindrico*, e può essere visto come una estensione semplice delle coordinate polari nello spazio: abbiamo che

$$\begin{cases} x = \rho \cos \theta, \\ y = \rho \sin \theta, \\ z = z. \end{cases}$$

Questa trasformazione corrisponde allo jacobiano

$$|\mathbf{J}_s(\rho, \theta, z)| = \begin{vmatrix} \cos \theta & -\rho \sin \theta & 0 \\ \sin \theta & \rho \cos \theta & 0 \\ 0 & 0 & 1 \end{vmatrix} = \rho \Rightarrow dx dy dz = \rho d\rho d\theta dz.$$

Un cambio di coordinate siffatto è particolarmente conveniente se si vuole calcolare il *volume di un solido di rotazione*, ovvero un solido costruito ruotando un certo dominio bidimensionale attorno all'asse z . Supponiamo per esempio di considerare il seguente dominio:

$$\mathcal{E} = \{(x, y, z) \in \mathbb{R}^3 : z \in [a, b], \quad 0 \leq x^2 + y^2 \leq r^2(z)\}$$

per una certa funzione $r: [a, b] \rightarrow \mathbb{R}^+$. Questo dominio può essere interpretato come il risultato di una rotazione del dominio

$$\mathcal{A} = \{(x, z) \in \mathbb{R}^2 : z \in [a, b], \quad 0 \leq x \leq r(z)\}$$

del piano xz attorno all'asse z . Per quanto detto sopra, il suo volume può scriversi come

$$(2.14) \quad V(\mathcal{E}) = \int_{\mathcal{E}} dx dy dz = 2\pi \int_a^b dz \int_0^{r(z)} \rho d\rho = \pi \int_a^b r^2(z) dz.$$

Per esempio una sfera S_R di raggio $R \geq 0$ si può scrivere come

$$S_R := \{(x, y, z) \in \mathbb{R}^3 : z \in [-R, R], \quad 0 \leq x^2 + y^2 \leq R^2 - z^2\}$$

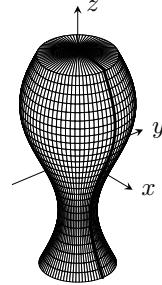
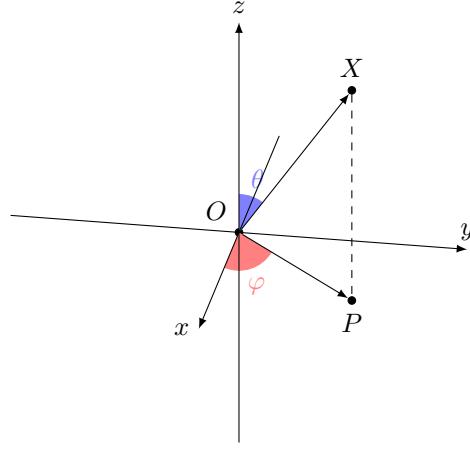


FIGURA 3. Esempio di superficie di rotazione \mathcal{E} ottenuta ruotando attorno all'asse z la curva evidenziata, che esprime la distanza $r(z)$ della superficie dall'asse z ad una quota data. Il volume di $V(\mathcal{E})$ può essere calcolato in questo caso usando la formula in Eq. (2.14).

per cui il suo volume è

$$V(S_R) = \pi \int_{-1}^1 (R^2 - z^2) dz = \frac{4}{3}\pi R^3.$$

3.3.3. Coordinate sferiche. Una maniera alternativa di individuare un punto $X(x)$ nello spazio \mathbb{R}^3 consiste nell'utilizzare la sua distanza ρ dall'origine O , l'angolo formato da \overrightarrow{OX} con il semiasse z positivo e l'angolo φ formato dalla proiezione \overrightarrow{OP} del segmento \overrightarrow{OX} sul piano $z = 0$ con il verso positivo dell'asse x .



In termini di queste tre quantità possiamo scrivere

$$(2.15) \quad \begin{cases} x = \rho \sin \theta \cos \varphi, \\ y = \rho \sin \theta \sin \varphi, \\ z = \rho \cos \theta. \end{cases}$$

La terna (ρ, θ, φ) costituisce le *coordinate sferiche* del punto. Lo jacobiano di questo cambio di coordinate si ottiene tramite calcolo diretto come

$$|\mathbf{J}_s(\rho, \theta, \varphi)| = \begin{vmatrix} \sin \theta \cos \varphi & \rho \cos \theta \cos \varphi & -\rho \sin \theta \sin \varphi \\ \sin \theta \sin \varphi & \rho \cos \theta \sin \varphi & \rho \sin \theta \cos \varphi \\ \cos \theta & -\rho \sin \theta & 0 \end{vmatrix} = \rho^2 \sin \theta \Rightarrow dx dy dz = \rho^2 \sin \theta d\rho d\theta d\varphi.$$

Superfici a simmetria sferica si scrivono in maniera molto semplice in termini di queste coordinate.

Per esempio, il dominio racchiuso da una sfera di raggio R si scrive

$$\mathcal{S}_R = \{(\rho, \theta, \varphi) : \rho \in [0, R], \quad \theta \in [0, \pi], \quad \varphi \in [0, 2\pi]\}.$$

In questo modo

$$V(\mathcal{S}_R) = \int_0^R \rho^2 d\rho \int_0^\pi \sin \theta d\theta \int_0^\pi d\varphi = \frac{4\pi}{3} R^3,$$

come sopra.

3.4. Baricentro. Supponiamo di avere un insieme misurabile $\mathcal{E} \subset \mathbb{R}^d$ e una funzione $\varrho : \mathcal{E} \rightarrow \mathbb{R}^+$ integrabile su \mathcal{E} , detta *densità volumetrica*, o semplicemente densità. Chiamiamo

$$m = \int_{\mathcal{E}} \varrho(\mathbf{x}) d^d x$$

la *massa* di \mathcal{E} : se $\varrho = 1$, la massa di \mathcal{E} coincide numericamente col suo volume. Se ϱ è una funzione costante allora si dice che \mathcal{E} è *omogeneo*. Il *baricentro* di \mathcal{E} è definito per mezzo di un integrale di volume come quelli visti finora. Sia $\mathcal{E} \subset \mathbb{R}^d$. Allora formalmente possiamo scrivere

$$\bar{\mathbf{x}} := \int_{\mathcal{E}} \mathbf{x} \varrho(\mathbf{x}) d^d x$$

ovvero è un punto $\bar{\mathbf{x}} \in \mathbb{R}^d$ tale che

$$\bar{x}_i := \frac{1}{m} \int_{\mathcal{E}} x_i \varrho(\mathbf{x}) d^d x.$$

Esso ha importanti applicazioni in *meccanica*, che qui non approfondiremo.

Consideriamo lo spicchio di semisfera in Fig. 4 che può essere facilmente parametrizzato usando coordinate sferiche,

$$\mathcal{E} = \{(\rho, \theta, \varphi) : 0 \leq \rho \leq 1, \quad 0 \leq \theta \leq \pi/2, \quad 0 \leq \varphi \leq \pi/4\}.$$

Assumendo una densità unitaria $\varrho = 1$, la massa di \mathcal{E} corrisponde al suo volume e può essere trovata calcolando

$$m = V(\mathcal{E}) = \int_0^1 d\rho \int_0^{\pi/2} d\theta \int_0^{\pi/4} d\varphi \rho^2 \sin \theta = \frac{\pi}{12}.$$

Assumendo una densità pari ad 1, possiamo calcolare le coordinate cartesiane del suo baricentro usando la definizione e ricordando l'espressione delle coordinate cartesiane in termini delle coordinate sferiche in Eq. (2.15), ovvero

$$\begin{aligned} \bar{x} &= \frac{1}{m} \int_0^1 d\rho \int_0^{\pi/2} d\theta \int_0^{\pi/4} d\varphi \rho^3 \cos \varphi \sin^2 \theta = \frac{\sqrt{2}}{8}, \\ \bar{y} &= \frac{1}{m} \int_0^1 d\rho \int_0^{\pi/2} d\theta \int_0^{\pi/4} d\varphi \rho^3 \sin \varphi \sin^2 \theta = \frac{2-\sqrt{2}}{8}, \\ \bar{z} &= \frac{1}{m} \int_0^1 d\rho \int_0^{\pi/2} d\theta \int_0^{\pi/4} d\varphi \rho^3 \cos \theta \sin \theta = \frac{1}{8}. \end{aligned}$$

4. Curve e superfici

4.1. Curve. Una *curva* in \mathbb{R}^d è una applicazione $\gamma : [a, b] \rightarrow \mathbb{R}^d$, dove $[a, b] \subset \mathbb{R}$. L'equazione $\mathbf{x} = \gamma(t)$ è detta *equazione parametrica della curva*: in essa, t è una sorta di "variabile temporale" in modo tale che $\gamma(t)$ possa essere interpretata come una "posizione nello spazio al

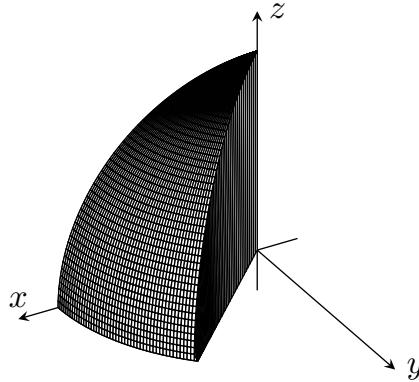


FIGURA 4. Spicchio di semisfera di raggio unitario nello spazio.

tempo t ". Se le componenti di γ sono derivabili in un certo $t_0 \in (a, b)$, possiamo associare un vettore velocità al punto $\gamma(t_0)$,

$$(2.16) \quad \dot{\gamma}(t_0) := \lim_{t \rightarrow t_0} \frac{\gamma(t) - \gamma(t_0)}{t - t_0}.$$

Se $\dot{\gamma}(t_0) \neq \mathbf{0}$, questo significa che la retta $\mathbf{x} = \gamma(t_0) + (t - t_0)\dot{\gamma}(t_0)$ è tangente alla curva. Se invece $\dot{\gamma}(t_0) = \mathbf{0}$, allora non esiste una retta tangente alla curva in $\gamma(t_0)$. Diamo le seguenti definizioni:

DEFINIZIONE 4.1 (Curva semplice, curva chiusa e curva di Jordan). Una curva $\gamma: [a, b] \rightarrow \mathbb{R}^d$ è semplice se, per ogni coppia di valori distinti $t, t' \in (a, b)$, $\gamma(t) \neq \gamma(t')$. È detta chiusa se $\gamma(a) = \gamma(b)$. Una curva semplice e chiusa è detta di Jordan.

DEFINIZIONE 4.2 (Curva regolare). Una curva $\gamma: [a, b] \rightarrow \mathbb{R}^d$ è regolare se ha vettore di derivate continue che non si annulla in nessun punto di (a, b) . La curva invece si dice regolare a tratti se è continua e si può partizionare in un numero finito di intervalli in cui è regolare.

Denotiamo Γ l'immagine di $[a, b]$ attraverso γ e definiamo versore tangente in $\mathbf{x} = \gamma(t)$ il vettore

$$(2.17) \quad \mathbf{u}_\gamma(t) := \frac{\dot{\gamma}(t)}{\|\dot{\gamma}(t)\|}, \quad \mathbf{x} = \gamma(t).$$

Si noti che la regolarità di una curva dipende dalla funzione γ e non dal supporto Γ : dipende cioè dalla rappresentazione di Γ .

Consideriamo la seguente curva in \mathbb{R}^2 :

$$\gamma(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}, \quad t \in \mathbb{R}.$$

Essa descrive una circonferenza Γ ed è una curva regolare e chiusa, ma non semplice. La curva

$$\psi(t) = \begin{pmatrix} \cos t^2 \\ \sin t^2 \end{pmatrix}, \quad t \in \mathbb{R}$$

rappresenta esattamente la stessa circonferenza Γ , e tuttavia non è regolare, perché $\psi'(0) = \mathbf{0}$.

L'interpretazione di t come di variabile "temporale" che parametrizza la curva porta anche alla seguente definizione, che esprime, in un certo senso, l'indipendenza dell'oggetto "curva" Γ dall'"orologio" con cui misuriamo la percorrenza su di essa.

DEFINIZIONE 4.3 (Equivalenza). *Due curve $\gamma: [a, b] \rightarrow \mathbb{R}^d$ e $\psi: [c, d] \rightarrow \mathbb{R}^d$ si dicono equivalenti se esiste una funzione $\tau: [a, b] \rightarrow [c, d]$ di classe C^1 , suriettiva e con derivata non nulla, tale che $\gamma(t) = \psi(\tau(t))$. Si dice che le due curve hanno lo stesso verso se $\tau' > 0$, diversamente si dice che hanno verso opposto.*

PROPOSIZIONE 4.1. *Se γ e ψ sono due curve equivalenti e con lo stesso verso hanno lo stesso versore tangente.*

DIMOSTRAZIONE. Basta osservare che, detto $\mathbf{x} = \gamma(t)$

$$\mathbf{u}_\gamma(t) = \frac{\dot{\gamma}(t)}{\|\dot{\gamma}(t)\|} = \frac{\dot{\psi}(\tau(t))\tau'(t)}{\|\dot{\psi}(\tau(t))\|\tau'(t)} = \frac{\dot{\psi}(\tau(t))}{\|\dot{\psi}(\tau(t))\|} = \mathbf{u}_\psi(\tau(t)).$$

□

In altre parole, il versore tangente nel punto $\gamma(t)$ è una funzione del punto e della direzione di percorrenza, e non della parametrizzazione.

4.1.1. Lunghezza di una curva e integrali curvilinei. Se $\gamma(t)$ fornisce la posizione di un punto lungo la curva al "tempo" t , e $\dot{\gamma}(t)$ la sua velocità, la distanza percorsa lungo la curva nell'intervallo $[t, t + dt]$ è $\|\dot{\gamma}(t)\| dt$: informalmente, infatti, $\|\dot{\gamma}(t)\| dt \simeq \sqrt{\sum_{i=1}^d d\gamma_i(t)^2}$, coerentemente con il teorema di Pitagora. Possiamo quindi dare la seguente definizione.

DEFINIZIONE 4.4 (Lunghezza di una curva). *Data una curva $\gamma: [a, b] \rightarrow \mathbb{R}^d$ continua con derivate continue, la sua lunghezza $\ell(\Gamma)$ è definita come*

$$\ell(\Gamma) := \int_a^b \|\dot{\gamma}(t)\| dt.$$

Vale la seguente diseguaglianza, che stabilisce che il percorso più breve fra due punti in geometria euclidea è dato dal segmento che li unisce.

TEOREMA 4.2. *La lunghezza di una curva $\gamma: [a, b] \rightarrow \mathbb{R}^d$ è sempre maggiore o uguale allo spostamento, ovvero alla distanza tra i suoi estremi,*

$$\ell(\Gamma) \geq \|\gamma(b) - \gamma(a)\|.$$

DIMOSTRAZIONE. Assumendo uno spostamento non nullo, abbiamo che, per un generico vettore \mathbf{v}

$$\langle \mathbf{u}, \gamma(b) - \gamma(a) \rangle = \int_a^b \langle \mathbf{v}, \dot{\gamma}(t) \rangle dt \leq \|\mathbf{u}\| \int_a^b \|\dot{\gamma}(t)\| dt = \|\mathbf{u}\| \ell(\Gamma).$$

Scegliendo $\mathbf{u} = \gamma(b) - \gamma(a)$ si prova la tesi. □

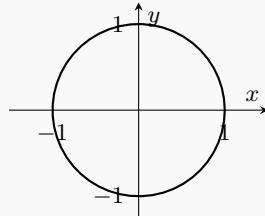
Avendo due curve equivalenti lo stesso supporto, forse non sorprende il seguente Teorema.

TEOREMA 4.3. *Due curve equivalenti $\gamma: [a, b] \rightarrow \mathbb{R}^d$ e $\psi: [c, d] \rightarrow \mathbb{R}^d$ hanno la stessa lunghezza.*

DIMOSTRAZIONE. La dimostrazione si basa sul fatto che, supponendo per esempio $\tau'(t) > 0$, $\|\dot{\gamma}(t)\| = \|\dot{\psi}(\tau(t))\|\tau'(t)$, per cui $\ell(\Gamma) = \int_a^b \|\dot{\gamma}(t)\| dt = \int_a^b \|\dot{\psi}(\tau(t))\|\tau'(t) dt = \int_c^d \|\dot{\psi}(\hat{t})\| d\hat{t}$. La dimostrazione è analoga se $\tau'(t) < 0$. □

La circonferenza nel piano con centro l'origine e raggio 1 può essere scritta in forma parametrica come

$$\gamma(t) = \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}, \quad t \in [0, 2\pi].$$



La curva è regolare: per ogni $t \in [0, 2\pi]$, abbiamo

$$\dot{\gamma}(t) = \begin{pmatrix} -\sin(t) \\ \cos(t) \end{pmatrix},$$

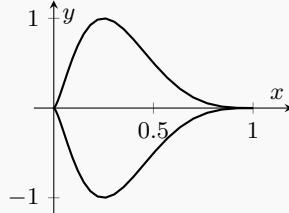
che non è mai uguale al vettore nullo. La sua lunghezza è

$$\ell(\Gamma) = \int_0^{2\pi} \sqrt{\sin^2(t) + \cos^2(t)} dt = 2\pi,$$

come ben sappiamo.

La curva

$$\gamma(t) = \begin{pmatrix} (t/\pi)^2 \\ \sin^3(t) \end{pmatrix}, \quad t \in [-\pi, \pi].$$



non è regolare perché ha, per esempio, $\dot{\gamma}(0) = \mathbf{0}$: in effetti presenta una cuspidé nell'origine (ma è regolare a tratti). Tuttavia il vettore delle derivate esiste continuo

$$\dot{\gamma}(t) = \begin{pmatrix} 2/\pi^2 t \\ 3 \cos(t) \sin^2(t) \end{pmatrix}$$

e possiamo calcolare la lunghezza

$$\ell(\Gamma) = \int_{-\pi}^{\pi} \sqrt{\frac{4t^2}{\pi^4} + 9 \cos^2(t) \sin^4(t)} dt = 4.78142\dots$$

dove l'integrale è stato eseguito numericamente.

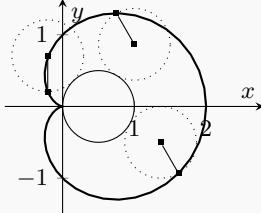
Talvolta le curve nel piano sono date non in coordinate cartesiane ma polari, ovvero scrivendo $(\rho(t), \theta(t))$ come funzioni di un certo parametro t . Per esempio, la cardioide può essere ottenuta facendo rotolare una circonferenza attorno ad un'altra fissa, con stesso raggio, nell'origine: la sua equazione può essere scritta come $\rho = 1 + \cos \theta$ per $\theta \in [0, 2\pi]$. In forma parametrica, possiamo riscrivere

$$\rho(t) = 1 + \cos t, \quad \theta = t, \quad t \in [0, 2\pi].$$

Questo significa che

$$(2.18) \quad \gamma(t) = \begin{pmatrix} (1 + \cos t) \cos t \\ (1 + \cos t) \sin t \end{pmatrix} \quad t \in [0, 2\pi],$$

e quindi, grazie ad un po' di algebra, $\|\dot{\gamma}(t)\| = \sqrt{2(1 + \cos t)}$.



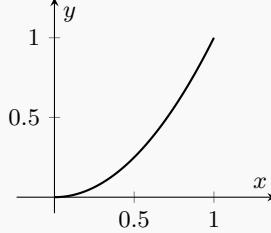
La sua lunghezza è quindi

$$\ell(\Gamma) = \sqrt{2} \int_0^{2\pi} \sqrt{1 + \cos t} dt = 2 \int_0^{2\pi} \left| \cos \frac{t}{2} \right| dt = 8.$$

Un importante esempio è il caso della lunghezza dei grafici di una funzione reale, ovvero delle curve descritte da una funzione $y = f(x)$ su un intervallo $[a, b] \in \mathbb{R}$. Queste curve possono essere interpretate come curve $\gamma(t) = (t, f(t))^\top \in \mathbb{R}^2$ con $t \in [a, b]$. Allora la lunghezza di un grafico è data da

$$\ell(\Gamma) = \int_a^b \|\dot{\gamma}(t)\| dt = \int_a^b \sqrt{1 + (\partial_x f(x))^2} dx.$$

Per esempio, consideriamo il ramo di parabola $y = \frac{1}{2}x^2$ con $x \in [0, 1]$.



Allora

$$\ell(\Gamma) = \int_0^1 \sqrt{1 + x^2} dx = \frac{\sqrt{2} + \ln(1 + \sqrt{2})}{2}.$$

Data una curva $\gamma: [a, b] \rightarrow \mathbb{R}^d$, possiamo introdurre la seguente quantità, che misura la lunghezza percorsa nell'intervallo $[a, t]$,

$$s(t) := \int_a^t \|\dot{\gamma}(\tilde{t})\| d\tilde{t}.$$

Se la curva è regolare, questa è una quantità strettamente crescente, essendo $\frac{ds}{dt} = \|\dot{\gamma}(t)\| > 0$, ed è quindi invertibile (“a ogni posizione corrisponde un tempo e viceversa”). La quantità s è a volte detta *ascissa curvilinea*. La lunghezza di una certa curva γ si indica quindi come $\ell(\Gamma) = \int_{\gamma} ds$, sottolineando che ds sia lo spostamento infinitesimo *lungo* la curva,

$$ds = \|\gamma(t + dt) - \gamma(t)\| = \|\dot{\gamma}(t)\| dt.$$

Si può in generale definire un *integrale curvilineo* di una funzione $f: \mathcal{A} \rightarrow \mathbb{R}$ lungo una curva $\Gamma \subset \mathcal{A}$ come

$$(2.19) \quad \int_{\gamma} f \, d s := \int_a^b f(\gamma(t)) \|\dot{\gamma}(t)\| \, d t.$$

Se γ e ψ sono curve equivalenti, l'integrale curvilineo di f su γ ha lo stesso valore dell'integrale curvilineo di f su ψ : in altre parole, esattamente come avviene per la lunghezza, il suo valore dipende dal *supporto* della curva, e non da come questa è parametrizzata. Questo tipo di integrale è detto *integrale di linea di prima specie*.

4.2. Superfici. Vediamo ora come trattare le superfici e gli integrali di superficie. Considereremo solo superfici in \mathbb{R}^3 , per semplicità.

DEFINIZIONE 4.5 (Superficie regolare). *Una superficie regolare σ è una applicazione $\sigma: \mathcal{K} \rightarrow \mathbb{R}^3$, con $\mathcal{K} \subset \mathbb{R}^2$ chiusura di un aperto limitato e connesso, tale che $\sigma \in C^1(\mathcal{K})$ (ovvero con derivate parziali continue), la sua restrizione all'interno \mathcal{K}° di \mathcal{K} è iniettiva e infine*

$$(2.20) \quad \frac{\partial \sigma(w)}{\partial u} \wedge \frac{\partial \sigma(w)}{\partial v} \neq \mathbf{0}, \quad \forall w = (u, v) \in \mathcal{K}^\circ.$$

L'ipotesi che \mathcal{K} sia la chiusura di un aperto connesso esclude il fatto che la superficie sia ottenuta da due pezzi uniti da una curva. La condizione con il prodotto vettoriale, invece, potrebbe apparire misteriosa ma è dovuta al fatto seguente. Una curva regolare era definita in maniera tale che esistesse un vettore tangente non nullo in ogni punto. Se ora consideriamo $w_0 \in \mathcal{K}^\circ$, possiamo costruire una curva regolare $w: [a, b] \rightarrow \mathcal{K}^\circ$ passante per w_0 a $t_0 \in [a, b]$. Possiamo anche costruire $\gamma(t) := \sigma(w(t))$ curva sulla superficie σ immagine della curva w in \mathcal{K}° . Indicando con $w(t) = (u(t), v(t))^\top$ la parametrizzazione della curva in \mathcal{K}° , il vettore tangente a γ sarà

$$\dot{\gamma}(t) = \frac{\partial \sigma(w_0)}{\partial u} \dot{u}(t_0) + \frac{\partial \sigma(w_0)}{\partial v} \dot{v}(t_0).$$

Perché questa curva sia regolare, occorre che questo vettore sia non nullo: essendo \dot{u} e \dot{v} non entrambi nulli per via della regolarità di w , ciò è garantito se i due vettori che appaiono nella somma non sono paralleli. Questa condizione si può esprimere dicendo che il loro prodotto vettoriale deve essere diverso dal vettore nullo, che è la condizione in Eq. (2.20).

Essendo $\partial_u \sigma$ e $\partial_v \sigma$ tangentini alla superficie, il vettore $\partial_u \sigma \wedge \partial_v \sigma$ è normale alla superficie e in particolare si può definire il versore normale a σ come

$$(2.21) \quad n_\sigma := \frac{\partial_u \sigma \wedge \partial_v \sigma}{\|\partial_u \sigma \wedge \partial_v \sigma\|}$$

4.2.1. Area di una superficie e integrali di superficie. L'area di una superficie σ è definita come segue

DEFINIZIONE 4.6 (Area di una superficie). *Sia $\sigma: \mathcal{K} \rightarrow \mathbb{R}^3$ una superficie regolare. La sua area è definita come*

$$A(\sigma) := \int_{\mathcal{K}} \|\partial_u \sigma \wedge \partial_v \sigma\| \, d u \, d v$$

La ragione *intuitiva* per questa definizione è la seguente: il quadratino di lati $d u$ e $d v$ in \mathcal{K} attorno ad un certo punto w viene mappato, tramite σ , in un parallelogramma nello spazio tridimensionale i cui lati sono $d u \partial_u \sigma$ e $d v \partial_v \sigma$: la quantità $\|\partial_u \sigma \wedge \partial_v \sigma\| d u d v$ corrisponde esattamente all'area di questo parallelogramma infinitesimo. L'integrale quindi somma questi

contributi infinitesimi che ricoprono la superficie che ci interessa. Informalmente, quindi, $d\sigma := \|\partial_u \boldsymbol{\sigma} \wedge \partial_v \boldsymbol{\sigma}\| dudv$ è un “elemento di superficie” e si scrive talvolta $A(\boldsymbol{\sigma}) = \int_{\boldsymbol{\sigma}} d\sigma$.

Un'applicazione analoga a quella che abbiamo visto per gli integrali curvilinei è il calcolo di aree di superfici definite da funzioni tipo $z = f(x, y)$, dove $f \in C^1(\mathcal{K})$ per un qualche insieme $\mathcal{K} \subset \mathbb{R}^2$. Queste funzioni possono essere scritte nel formalismo sopra come

$$\boldsymbol{\sigma}(u, v) = (u, v, f(u, v)), \quad (u, v) \in \mathcal{K}.$$

Di conseguenza

$$\|\partial_u \boldsymbol{\sigma} \wedge \partial_v \boldsymbol{\sigma}\| = \sqrt{1 + \|\nabla f\|^2},$$

ovvero l'area della superficie è data da

$$A = \int_{\mathcal{K}} \sqrt{1 + \|\nabla f(u, v)\|^2} du dv.$$

Per esempio, consideriamo il paraboloido in Fig. 5, individuato dall'equazione

$$f(x, y) = x^2 + y^2, \quad \mathcal{K} = \{(x, y) \in \mathbb{R}^2 : 0 \leq x^2 + y^2 \leq 1\}.$$

Allora la superficie del paraboloido è

$$A = \int_{\mathcal{K}} \sqrt{1 + 4x^2 + 4y^2} dx dy = 2\pi \int_0^1 \sqrt{1 + 4\rho^2} \rho d\rho = \frac{\pi}{6} (\sqrt[3]{25} - 1).$$

Usando quanto visto nella sezione precedente, possiamo calcolare anche il volume contenuto dal paraboloido, ovvero la misura di

$$\mathcal{E} = \{(x, y, z) \in \mathbb{R}^3 : z \in [0, 1], \quad 0 \leq x^2 + y^2 \leq z\}.$$

Osservando che si tratta di un solido di rotazione,

$$V = \iiint_{\mathcal{E}} dx dy dz = \pi \int_0^1 z dz = \frac{\pi}{2}.$$

Una seconda applicazione, estremamente utile, del formalismo sopra è il calcolo delle aree di *superficie di rotazione*, ovvero superfici ottenute ruotando attorno all'asse z funzioni definite sul piano $y = 0$ da un'equazione del tipo $x = f(z)$, con $z \in [a, b] \subset \mathbb{R}$ e $f(z) > 0$. Questo vuol dire che la superficie generata può essere scritta in forma parametrica come

$$\boldsymbol{\sigma}(u, v) = (f(u) \cos v, f(u) \sin v, u), \quad (u, v) = [a, b] \times [0, 2\pi].$$

In questo caso un calcolo diretto mostra che

$$\|\partial_u \boldsymbol{\sigma} \wedge \partial_v \boldsymbol{\sigma}\| = f(u) \sqrt{1 + (\partial_z f(u))^2}$$

Questo significa che l'area di tali superfici si può scrivere come

$$A = 2\pi \int_a^b f(u) \sqrt{1 + (\partial_z f(u))^2} du.$$

Per esempio, il paraboloido considerato nell'esempio precedente può essere visto come la rotazione attorno all'asse z della parabola $z = x^2$ nel piano $y = 0$, ovvero $x = f(z) = \sqrt{z}$ per $z \in [0, 1]$. La formula che abbiamo ottenuto dice quindi che possiamo calcolarne l'area come

$$A = 2\pi \int_0^1 \sqrt{u} \sqrt{1 + \frac{1}{4u}} du = \pi \int_0^1 \sqrt{1 + 4u} du = \frac{\pi}{6} (\sqrt[3]{25} - 1).$$

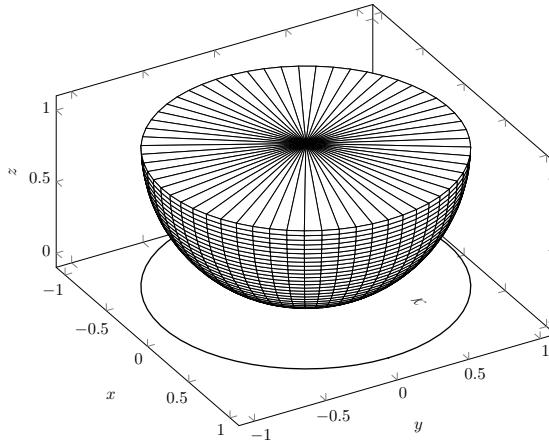


FIGURA 5. Paraboloide su disco.

L'integrale di superficie di una funzione $f: \mathcal{A} \rightarrow \mathbb{R}$ su $\Sigma \subseteq \mathcal{A}$ si definisce facilmente come

$$(2.22) \quad \int_{\sigma} f d\sigma := \int_{\mathcal{K}} f(\boldsymbol{\sigma}(u, v)) \|\partial_u \boldsymbol{\sigma} \wedge \partial_v \boldsymbol{\sigma}\| dudv.$$

Come vedremo in seguito, è possibile anche definire integrali di superficie riguardanti *campi vettoriali* $\mathbf{f}: \mathcal{A} \subseteq \mathbb{R}^3 \rightarrow \mathbb{R}^3$, ovvero funzioni vettoriali 3-dimensionali in 3 variabili. Si definisce *flusso* del campo vettoriale \mathbf{f} attraverso la superficie $\boldsymbol{\sigma}$ l'integrale

$$(2.23) \quad \int_{\sigma} \langle \mathbf{f}, \mathbf{n}_{\boldsymbol{\sigma}} \rangle d\sigma = \iint_{\mathcal{K}} \langle \mathbf{f}(\boldsymbol{\sigma}(u, v)), \partial_u \boldsymbol{\sigma}(u, v) \wedge \partial_v \boldsymbol{\sigma}(u, v) \rangle du dv.$$

4.2.2. Centroide. Analogamente al baricentro introdotto per i volumi, i *centroide* di una superficie è definito come una sorta di posizione media della superficie stessa. Sia A l'area di una superficie $\boldsymbol{\sigma}$: allora il suo centroide ha coordinate

$$\bar{x} = \frac{1}{A} \int x d\sigma, \quad \bar{y} = \frac{1}{A} \int y d\sigma, \quad \bar{z} = \frac{1}{A} \int z d\sigma.$$

Consideriamo nuovamente lo spicchio di semisfera in Fig. 4. La sua superficie curva è parametrizzabile come

$$\boldsymbol{\sigma}(u, v) = \begin{pmatrix} \cos v \sin u \\ \sin v \sin u \\ \cos u \end{pmatrix}, \quad u \in [0, \pi/2], \quad v \in [0, \pi/4].$$

Da questo possiamo derivare, con un po' di algebra,

$$\|\partial_u \boldsymbol{\sigma}(u, v) \wedge \partial_v \boldsymbol{\sigma}(u, v)\| = \sin u$$

(dove abbiamo osservato che nel nostro intervallo di interesse, $\sin u$ è sempre positivo) da cui possiamo calcolare l'area dello spicchio come

$$A = \int_0^{\frac{\pi}{2}} du \int_0^{\frac{\pi}{4}} dv \sin u = \frac{\pi}{4}.$$

Calcoliamo ora le coordinate del centroide:

$$\begin{aligned}\bar{x} &= \frac{1}{A} \int_0^{\frac{\pi}{2}} du \int_0^{\frac{\pi}{4}} dv \cos v \sin^2 u = \frac{1}{\sqrt{2}}, \\ \bar{y} &= \frac{1}{A} \int_0^{\frac{\pi}{2}} du \int_0^{\frac{\pi}{4}} dv \sin v \sin^2 u = 1 - \frac{1}{\sqrt{2}}, \\ \bar{z} &= \frac{1}{A} \int_0^{\frac{\pi}{2}} du \int_0^{\frac{\pi}{4}} dv \cos u \sin u = \frac{1}{2}.\end{aligned}$$

4.3. Massimi e minimi vincolati. Abbiamo finora considerato il problema di trovare massimi e minimi di una funzione $f: \mathcal{A} \rightarrow \mathbb{R}$, con $\mathcal{A} \subseteq \mathbb{R}^d$, e abbiamo trovato un criterio per individuare i punti interni di un insieme che siano *stazionari* e poi, tra essi, individuare massimi e minimi per mezzo dello studio della matrice hessiana.

In molte applicazioni, tuttavia, si è interessati a cercare un punto di massimo e minimo non genericamente su \mathcal{A}° ma, per esempio, lungo una curva $\Gamma \subset \mathcal{A}$: in altre parole, parametrizzando la curva come $\mathbf{x} = \gamma(t)$, i punti estremali saranno tali che

$$\frac{d f(\gamma(t))}{dt} = \langle \dot{\gamma}(t), \nabla f(\gamma(t)) \rangle = 0.$$

Essendo $\dot{\gamma}$ un vettore parallelo alla curva, ciò significa che perché f abbia un massimo in $\mathbf{x} = \gamma(t)$, in tale punto il suo gradiente deve essere ortogonale alla curva: è questa la condizione di punto stazionario *vincolato* a Γ .

Se $d = 3$ e invece vogliamo trovare un punto estremale vincolando la funzione su una superficie Σ descritta da una funzione $\sigma: \mathcal{K} \rightarrow \mathbb{R}^d$. Possiamo procedere in maniera simile cercando i punti in cui

$$\begin{pmatrix} \partial_u f(\sigma(u, v)) \\ \partial_v f(\sigma(u, v)) \end{pmatrix} = \begin{pmatrix} \langle \nabla f, \partial_u \sigma \rangle \\ \langle \nabla f, \partial_v \sigma \rangle \end{pmatrix} = \mathbf{0},$$

ovvero ancora una volta ∇f deve essere perpendicolare al vincolo (dato che $\partial_u \sigma$ e $\partial_v \sigma$ sono vettori tangenti alla superficie).

Ciò che abbiamo ottenuto può essere riassunto nel seguente Teorema.

TEOREMA 4.4. *Sia \mathcal{G} un vincolo (ovvero una superficie o una curva) ed $f: \mathcal{A} \rightarrow \mathbb{R}$ una funzione continua con derivate continue su un aperto \mathcal{A} contenente \mathcal{G} . Allora se f ha un punto di massimo o minimo relativo vincolato in $\mathbf{p} \in \mathcal{G}^\circ$, il gradiente di f in \mathbf{p} è ortogonale a \mathcal{G} .*

Se il vincolo è dato con equazioni parametriche del tipo che abbiamo visto sopra, ovvero tipo $\mathbf{x} = \gamma(t)$ o $\mathbf{x} = \sigma(u, v)$, possiamo semplicemente considerare la composizione di f con queste funzioni e calcolare le opportune derivate o gli opportuni gradienti per imporre poi la condizione di stazionarietà. Può accadere tuttavia che il vincolo \mathcal{G} sia dato con una condizione del tipo

$$\mathcal{G}: g(\mathbf{x}) = 0,$$

ovvero con una *equazione cartesiana*. In tal caso, la direzione *ortogonale* al vincolo nel punto \mathbf{x} è quella corrispondente al gradiente ∇g , esattamente per le stesse ragioni sopra. Per esempio, se il vincolo specificasse una curva $\mathbf{x} = \gamma(t)$, allora ciò vorrebbe dire che $g(\gamma(t)) = 0$, e di conseguenza $\langle \nabla g, \dot{\gamma} \rangle = 0$. Ciò che dobbiamo imporre è quindi una condizione di parallelismo, nella forma

$$\nabla f(\mathbf{x}) + \lambda \nabla g(\mathbf{x}) = \mathbf{0} \quad \text{per qualche } \lambda \in \mathbb{R}.$$

Questo fatto, insieme alle osservazioni precedenti, può essere riassunto nel seguente importante teorema, dalle molteplici applicazioni, dovuto a Giuseppe Lagrange.

TEOREMA 4.5 (Moltiplicatori di Lagrange). *Supponiamo di avere un vincolo \mathcal{G} identificato da una condizione del tipo $g(\mathbf{x}) = 0$, con $g \in \mathcal{C}^1(\mathcal{A})$, $\mathcal{A} \subset \mathbb{R}^d$, e sia f una funzione di classe \mathcal{C}^1*

in un aperto contenente \mathcal{G} . Allora i punti di massimo e minimo relativo vincolati di f interni al vincolo sono i punti stazionari liberi della funzione

$$H(\mathbf{x}, \lambda) := f(\mathbf{x}) + \lambda g(\mathbf{x})$$

rispetto a \mathbf{x} e λ .

Consideriamo il seguente problema: trovare la il parallelepipedo *senza coperchio* a base quadrata di massimo volume di area A data. Se x è il lato della base e y l'altezza del parallelepipedo, allora il suo volume è $V = x^2y$, mentre la sua area è vincolata come $A = 4xy + x^2$. Scriviamo la nostra funzione da estremizzare,

$$H(x, y, \lambda) = x^2y + \lambda(x^2 + 4xy - A).$$

Abbiamo le condizioni imposte dal gradiente

$$x^2 + 4xy = A, \quad 2y + 2\lambda x + 4\lambda y = 0, \quad x^2 + 4x\lambda = 0.$$

Dovendo essere $x \neq 0$, l'ultima equivale a $x = -4\lambda$ (per cui $\lambda < 0$) mentre la seconda si può semplificare come $y = -2\lambda$. Possiamo trovare il valore di λ imponendo il vincolo,

$$A = x^2 + 4xy = 48\lambda^2 \Rightarrow \lambda = -\frac{1}{4}\sqrt{\frac{A}{3}},$$

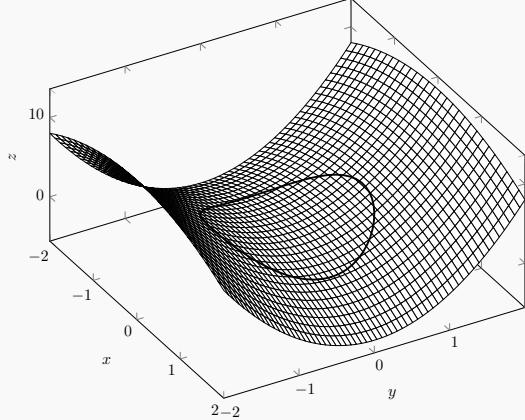
per cui

$$x = \sqrt{\frac{A}{3}}, \quad y = \frac{1}{2}\sqrt{\frac{A}{3}}.$$

Supponiamo di avere la funzione

$$f(\mathbf{x}) = f(x, y) = -x^2 + 3y^2$$

e di doverla estremizzare col vincolo che $x^2 + y^2 = 1$.



Procediamo utilizzando il metodo di Lagrange, scrivendo anzitutto

$$H(\mathbf{x}, \lambda) = -x^2 + 3y^2 + \lambda(x^2 + y^2 - 1).$$

Le diverse derivate forniscono

$$x^2 + y^2 = 1, \quad 2(\lambda - 1)x = 0, \quad (6 + \lambda)y = 0.$$

La prima equazione fornisce $\lambda = 1$. Dalla seconda, in questo caso, si ottiene $y = 0$ e quindi dall'equazione del vincolo $x = \pm 1$. Viceversa, scegliendo $\lambda = -6$, la seconda equazione è soddisfatta mentre la prima fornisce $x = 0$, che, insieme al vincolo, dà $y = \pm 1$. I punti estremali sono quindi $(0, \pm 1)$ e $(\pm 1, 0)$. In particolare, i massimi sono $(0, \pm 1)$, mentre i minimi sono su $(\pm 1, 0)$.

5. Forme differenziali

In questa sezione discuteremo le *forme differenziali*. Iniziamo dalla loro definizione.

DEFINIZIONE 5.1. Sia $\mathcal{A} \subseteq \mathbb{R}^d$ un aperto e sia $\mathbf{a}(\mathbf{x})$ un campo vettoriale d -dimensionale, ovvero una funzione di d componenti su \mathcal{A} , $\mathbf{a}: \mathcal{A} \rightarrow \mathbb{R}^d$. La forma differenziale associata ad \mathbf{a} è data dall'espressione

$$(2.24) \quad \omega := \langle d\mathbf{x}, \mathbf{a}(\mathbf{x}) \rangle = \sum_{i=1}^d a_i(\mathbf{x}) d x_i.$$

Data una curva regolare $\gamma: [a, b] \rightarrow \mathbb{R}^d$, si scrive

$$(2.25) \quad \int_{\gamma} \omega := \int_a^b \langle \dot{\gamma}(t), \mathbf{a}(\gamma(t)) \rangle dt.$$

Si dice anche che la quantità sopra è l'*integrale di linea di seconda specie*, o la *circuitazione*, del campo vettoriale \mathbf{a} lungo γ . Se γ è una curva chiusa, si usa a volte la notazione

$$\oint_{\gamma} \omega.$$

Vale il seguente teorema, che diamo senza dimostrazione (la prova è basata ancora su un cambio di variabile).

TEOREMA 5.1. Se γ e ψ sono equivalenti, allora $\int_{\gamma} \omega = \int_{\psi} \omega$ se hanno lo stesso verso, $\int_{\gamma} \omega = -\int_{\psi} \omega$ se hanno verso opposto.

5.1. Forme differenziali esatte. Abbiamo già visto come scrivere il differenziale di una funzione differenziabile, quando abbiamo esplicitato

$$d f(\mathbf{x}) = \sum_{i=1}^d \partial_i f(\mathbf{x}) d x_i,$$

Il differenziale di una funzione f è quindi associato al campo vettoriale ∇f e si dice essere una *forma esatta*.

DEFINIZIONE 5.2 (Forma esatta). Una forma differenziale ω definita su un aperto $\mathcal{A} \subset \mathbb{R}^d$ e detta esatta se esiste una funzione f , detta primitiva di ω , tale che $\omega = \langle d\mathbf{x}, \nabla f(\mathbf{x}) \rangle$.

Le forme esatte godono di una proprietà molto importante. Se infatti consideriamo un integrale curvilineo,

$$\int_{\gamma} \omega = \int_a^b \langle \dot{\gamma}(t), \nabla f(\gamma(t)) \rangle dt = \int_a^b \frac{d f(\gamma(t))}{dt} dt = f(\gamma(b)) - f(\gamma(a)),$$

ovvero l'integrale della forma $d f$ lungo una curva dipende solo dal punto iniziale e finale della curva. Il campo $\mathbf{a}(\mathbf{x}) = \nabla f(\mathbf{x})$ si dice *conservativo* e f si dice essere la *primitiva* di ω . Una forma differenziale esatta gode quindi della proprietà anzidetta di avere integrale su una curva dipendente solo dagli estremi della curva. Data quindi la classe $\Gamma(\mathbf{p}, \mathbf{q})$ di tutte le curve con estremi \mathbf{p} e \mathbf{q} , abbiamo che

$$(2.26) \quad \int_{\gamma} \omega = \int_{\psi} \omega, \quad \gamma, \psi \in \Gamma(\mathbf{p}, \mathbf{q}).$$

In particolare, se la curva è chiusa,

$$(2.27) \quad \oint_{\gamma} \omega = 0.$$

La condizione in Eq. (2.26) è *caratterizzante* per le forme esatte, ovvero tutte e sole le forme che soddisfano questa proprietà sono esatte, come stabilito dal seguente teorema che non dimostriamo.

TEOREMA 5.2. *Sia $\omega = \langle d\mathbf{x}, \mathbf{a}(\mathbf{x}) \rangle$ una forma differenziale continua in un aperto \mathcal{A} . Se, comunque presi $\mathbf{p}, \mathbf{q} \in \mathcal{A}$ e comunque prese due curve $\gamma, \psi \in \Gamma(\mathbf{p}, \mathbf{q})$ vale l'Eq. (2.26), allora ω è esatta. Equivalentemente, condizione necessaria e sufficiente perché la forma sia esatta è che l'Eq. (2.27) valga per ogni curva chiusa.*

Il teorema di Schwarz ci fornisce una condizione necessaria (ma *non* sufficiente) perché una forma sia esatta.

TEOREMA 5.3. *Sia $\omega = \langle d\mathbf{x}, \mathbf{a}(\mathbf{x}) \rangle$ una forma differenziale di classe $C^1(\mathcal{A})$. Condizione necessaria perché sia esatta è che*

$$(2.28) \quad \partial_i a_j(\mathbf{x}) = \partial_j a_i(\mathbf{x}), \quad \forall \mathbf{x} \in \mathcal{A} \quad \forall i, j \in \{1, \dots, d\}.$$

Se ω soddisfa l'Eq. (2.28) si dice che essa è *chiusa*: in fisica si dice che $\mathbf{a}(\mathbf{x})$ è *irrotazionale*. In termini fisici quindi ogni campo conservativo è irrotazionale ma il viceversa non è sempre vero. Una condizione perché una curva chiusa sia esatta è che il dominio su cui essa è data sia semplicemente connesso (ovvero se ogni curva chiusa nel dominio può essere deformata con continuità in un punto).

La forma definita su \mathbb{R}^2

$$\omega = xy \, dx + \frac{x^2}{2} \, dy$$

ha $\mathbf{a}(\mathbf{x}) = (-y, x^2)^\top$ ed è chiusa, dato che $\partial_y a_2 = \partial_y a_1 = x$. Essendo \mathbb{R}^2 semplicemente connesso, essa è anche esatta. Viceversa, la forma su \mathbb{R}^2

$$\omega = -xy \, dx + \frac{x^2}{2} \, dy$$

ha $x = \partial_y a_2 \neq \partial_y a_1 = -x$ ovunque eccetto in $x = 0$. La forma pertanto non è chiusa (e quindi neppure esatta) su \mathbb{R}^2 .

Consideriamo la forma su $\mathbb{R}^2 \setminus \{\mathbf{0}\}$

$$\omega = \left(\frac{1}{x^2} + \frac{1}{y^2} \right) (y \, dx - x \, dy).$$

Si verifica facilmente che questa forma è chiusa, dato che

$$\partial_x a_2 = \partial_y a_1 = \frac{1}{x^2} - \frac{1}{y^2}.$$

Per vedere se essa è esatta (ciò non è scontato visto che $\mathbb{R}^2 \setminus \{\mathbf{0}\}$ non è semplicemente connesso) cerchiamo una funzione f primitiva. Dobbiamo avere

$$\partial_x f = \left(\frac{1}{x^2} + \frac{1}{y^2} \right) y \Rightarrow f(x, y) = -\frac{y}{x} + \frac{x}{y} + c(y) \quad \partial_y f = -\left(\frac{1}{x^2} + \frac{1}{y^2} \right) x \Rightarrow f(x, y) = -\frac{y}{x} + \frac{x}{y} + \hat{c}(x).$$

Questo suggerisce

$$f(x, y) = -\frac{x}{y} + \frac{y}{x} + c$$

per una arbitraria costante $c \in \mathbb{R}$. Abbiamo così ottenuto una famiglia di primitive per ω , che quindi è esatta.

Diamo ora un esempio di forma chiusa ma non esatta. Sia data

$$(2.29) \quad \omega = \frac{y \, dx - x \, dy}{x^2 + y^2}$$

su $\mathbb{R}^2 \setminus \{\mathbf{0}\}$. Questa forma è chiusa, dato che

$$\partial_y a_1 = \partial_x a_2 = \frac{x^2 - y^2}{(x^2 + y^2)^2}.$$

Per mostrare che essa non è esatta, calcoliamo il suo integrale lungo la circonferenza di raggio uno centrata nell'origine, ovvero

$$(2.30) \quad \gamma = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}, \quad t \in [0, 2\pi].$$

Otteniamo

$$\oint_{\gamma} \omega = - \int_0^{2\pi} dt = -2\pi \neq 0.$$

5.2. La formula di Gauss–Green. Focalizziamoci ora sul caso bidimensionale. In $d = 2$ si usa una notazione speciale per una forma differenziale, e si scrive

$$\omega = M(x, y) \, dx + N(x, y) \, dy.$$

Supponiamo ora che $\gamma: [a, b] \rightarrow \mathbb{R}^2$ sia una curva semplice e chiusa, che assumiamo *orientata positivamente*, ovvero in maniera tale che percorrendola secondo t crescente la porzione di spazio racchiusa dalla curva venga tenuta a sinistra. Questa curva è quindi la frontiera $\partial\mathcal{G}$ di un insieme $\mathcal{G} \subset \mathbb{R}^2$. Vale il seguente teorema, che non dimostreremo.

TEOREMA 5.4 (Gauss–Green). *Sia $\omega = M(x, y) \, dx + N(x, y) \, dy$ una forma differenziale di classe C^1 su un aperto $\mathcal{A} \subseteq \mathbb{R}^2$ e sia $\mathcal{G} \subseteq \mathcal{A}$ una regione limitata avente come frontiera $\partial\mathcal{G}$ una curva regolare γ orientata positivamente, semplice e chiusa. Allora*

$$\oint_{\gamma} M \, dx = - \iint_{\mathcal{G}} \partial_y M \, dx \, dy, \quad \int_{\gamma} N \, dy = \iint_{\mathcal{G}} \partial_x N \, dx \, dy$$

per cui

$$(2.31) \quad \oint_{\gamma} \omega = \iint_{\mathcal{G}} (\partial_x N(x, y) - \partial_y M(x, y)) \, dx \, dy.$$

Come specificato nelle ipotesi, \oint_{γ} denota l'integrale lungo la frontiera di \mathcal{G} secondo una qualche parametrizzazione che ne renda l'orientazione positiva. Questo teorema, detto di Gauss–Green, è ricco di conseguenze. Può per esempio aiutare a calcolare l'area di una regione \mathcal{G} : basta infatti scegliere come forma

$$\omega = -\frac{1}{2}y \, dx + \frac{1}{2}x \, dy,$$

in modo che

$$\oint_{\gamma} \omega = \iint_{\mathcal{G}} dx \, dy.$$

Consideriamo la cardioide definita in Eq. (2.18) come la curva di equazione

$$\gamma(t) = \begin{pmatrix} (1+\cos(t)) \cos(t) \\ (1+\cos(t)) \sin(t) \end{pmatrix}, \quad t \in [0, 2\pi]$$

Abbiamo già calcolato la lunghezza della sua frontiera. Proviamo ora a calcolarne l'area. Indichiamo

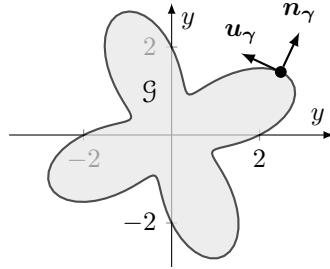


FIGURA 6. Insieme limitato \mathcal{G} avente come frontiera una curva regolare γ . Dato un punto sulla frontiera, sono evidenziati il vettore u_γ tangente alla frontiera e il vettore n_γ normale ad essa.

con \mathcal{C} il dominio delimitato dalla cardioide e usiamo, come detto sopra, la forma

$$\omega = -\frac{1}{2}y \, dx + \frac{1}{2}x \, dy.$$

Le formule di Gauss–Green ci dicono che l'integrale che vogliamo calcolare è uguale a

$$\iint_{\mathcal{C}} dx \, dy = \oint_{\gamma} \omega = \int_0^{2\pi} \frac{(1 + \cos(t))^2}{2} \, dt = \frac{3}{2}\pi.$$

Consideriamo la forma in Eq. (2.29),

$$\omega = \frac{y \, dx - x \, dy}{x^2 + y^2}$$

e il cerchio di raggio unitario centrato nell'origine \mathcal{C} con frontiera γ in Eq. (2.30). Abbiamo visto che

$$\oint_{\gamma} \omega = -2\pi.$$

D'altra parte

$$\iint_{\mathcal{C}} \left(-\frac{\partial}{\partial x} \left(\frac{x \, dy}{x^2 + y^2} \right) - \frac{\partial}{\partial y} \left(\frac{y \, dx}{x^2 + y^2} \right) \right) dx \, dy = 0$$

che mostra che in questo caso il teorema di Gauss–Green *non* si applica: la forma, infatti, ha una singolarità nell'origine, condizione non prevista dalle ipotesi del teorema.

5.2.1. Il teorema della divergenza. Il teorema ha come corollario il caso bidimensionale del cosiddetto *teorema della divergenza*. Per introdurlo, vediamo come possiamo riscrivere l'enunciazione del teorema di Gauss–Green. Ricordiamo che se scriviamo l'equazione della frontiera γ come $\mathbf{x} = \gamma(t)$, con $\gamma: [a, b] \rightarrow \mathbb{R}^2$, allora

$$\mathbf{u}_\gamma(\mathbf{x}) := \frac{\dot{\gamma}(t)}{\|\dot{\gamma}(t)\|} = \frac{(\dot{\gamma}_1(t), \dot{\gamma}_2(t))^\top}{\|\dot{\gamma}(t)\|}, \quad \mathbf{n}_\gamma(\mathbf{x}) := \frac{(\dot{\gamma}_2(t), -\dot{\gamma}_1(t))^\top}{\|\dot{\gamma}(t)\|}$$

sono rispettivamente il versore tangente alla curva e il versore ortogonale alla curva che, secondo l'orientazione scelta, punta all'esterno di \mathcal{G} ².

²Il versore ortogonale che punta verso l'esterno può essere ottenuto immaginando \mathbf{u}_γ in uno spazio tridimensionale, e costruendo $\mathbf{n}_\gamma = \mathbf{u}_\gamma \wedge \mathbf{e}_3$: facendo in questo modo, infatti, $\mathbf{v}_{\partial\mathcal{G}}$ è perpendicolare a $\mathbf{u}_{\partial\mathcal{G}}$ e giace nel piano, ed è orientato nella direzione desiderata.

Abbiamo allora che

$$\oint_{\gamma} M \, dx = \int_a^b M(\gamma(t)) \dot{\gamma}_1(t) \, dt = \int_a^b M(\gamma(t)) \frac{\dot{\gamma}_1(t)}{\|\dot{\gamma}(t)\|} \|\dot{\gamma}(t)\| \, dt$$

e allo stesso modo si vede che

$$\int_{\gamma} N \, dy = \int_a^b N(\gamma(t)) \frac{\dot{\gamma}_2(t)}{\|\dot{\gamma}(t)\|} \|\dot{\gamma}(t)\| \, dt,$$

per cui, se introduciamo la coordinata curvilinea $d s = \|\dot{\gamma}(t)\| \, dt$ e il campo vettoriale

$$\mathbf{f}(x, y) = \begin{pmatrix} N(x, y) \\ M(x, y) \end{pmatrix}$$

possiamo sommare le due equazioni e scrivere

$$(2.32) \quad \oint_{\gamma} (N \, dy - M \, dx) = \oint_{\gamma} \langle \mathbf{n}_{\gamma}, \mathbf{f} \rangle \, ds$$

per le formule di Gauss–Green = $\iint_{\mathcal{G}} (\partial_x N + \partial_y M) \, dx \, dy =: \iint_{\mathcal{G}} \operatorname{div} \mathbf{f} \, dx \, dy$

dove abbiamo definita la *divergenza* di \mathbf{f} . Per un campo vettoriale \mathbf{f} in dimensione d generica, essa è data da

$$\operatorname{div} \mathbf{f}(\mathbf{x}) = \sum_{i=1}^d \partial_i f_i(\mathbf{x}).$$

Questa quantità è a volte indicata con $\nabla \cdot \mathbf{f}$, interpretando la scrittura come un “prodotto scalare” formale tra il vettore $\nabla = (\partial_i)_{i=1}^d$ e \mathbf{f} . Il *teorema della divergenza* ottenuto dalla formula di Gauss–Green si generalizza in dimensione d generica e si può scrivere

$$(2.33) \quad \int_{\partial \mathcal{G}} \langle \mathbf{n}_{\partial \mathcal{G}}, \mathbf{f} \rangle \, ds = \int_{\mathcal{G}} \operatorname{div} \mathbf{f} \, d^d x.$$

Nella formula precedente, $\partial \mathcal{G}$ è la frontiera di \mathcal{G} e $\mathbf{n}_{\partial \mathcal{G}}$ è il versore normale ad essa in un dato punto, orientato verso l'esterno. Il teorema stabilisce che *l'integrale della divergenza di un campo vettoriale $\mathbf{f}(\mathbf{x})$ su \mathcal{G} è uguale al flusso del campo attraverso la frontiera di \mathcal{G} .*

5.3. La formula di Stokes. Una forma differenziale in $d = 3$ si scrive solitamente usando la notazione

$$\omega = M(x, y, z) \, dx + N(x, y, z) \, dy + Q(x, y, z) \, dz.$$

Presenteremo un risultato di importanza analoga al teorema della divergenza nel caso di forme differenziali in $d = 3$. Esso riguarda una quantità legata al campo vettoriale $\mathbf{a} = (M, N, Q)^T$ associato alla forma, ovvero il *rotore* di \mathbf{a} ,

$$(2.34) \quad \operatorname{rot} \mathbf{a} = \begin{pmatrix} \partial_y Q - \partial_z N \\ \partial_z M - \partial_x Q \\ \partial_x N - \partial_y M \end{pmatrix}$$

che talvolta si scrive

$$\operatorname{rot} \mathbf{a} = \nabla \wedge \mathbf{a},$$

intendendo un prodotto vettoriale formale tra \mathbf{a} e il vettore $\nabla = (\partial_x, \partial_y, \partial_z)$.

Sia data ora una superficie regolare $\mathcal{S} \subset \mathbb{R}^3$, parametrizzata come $\mathbf{x} = \boldsymbol{\sigma}(u, v)$ con $(u, v) \in \mathcal{K} \subset \mathbb{R}^2$. Supponiamo di considerare un aperto $\mathcal{A} \subseteq \mathcal{K}$ e che $\partial \mathcal{A}$ sia dato da una curva semplice

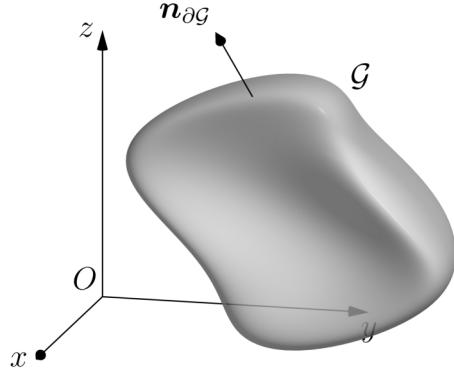


FIGURA 7. Il teorema della divergenza afferma che il flusso attraverso la frontiera ∂G di un campo \mathbf{f} può scriversi come l'integrale della divergenza di \mathbf{f} su G .

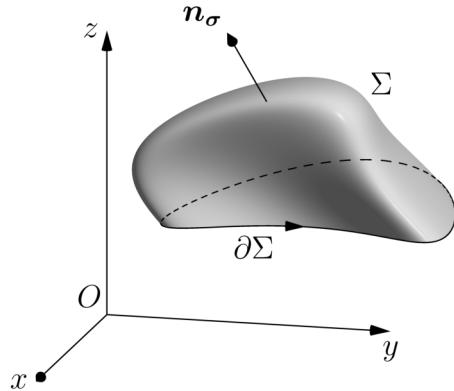


FIGURA 8. Il teorema di Stokes afferma che la circuitazione di un campo vettoriale lungo la frontiera $\partial \Sigma$ è uguale al flusso del rotore dello stesso campo attraverso Σ .

e chiusa γ : chiamiamo $\Sigma \subseteq \mathcal{S}$ l'immagine di \mathcal{A} attraverso σ , e $\partial \Sigma$ l'immagine di $\partial \mathcal{A}$ attraverso la stessa funzione, parametrizzata come $\partial \sigma(t) := \sigma(\gamma(t))$: assumendo γ orientata positivamente, si dice che $\partial \Sigma$ è anch'essa orientata positivamente. Il risultato è il seguente.

TEOREMA 5.5 (Stokes). *Sia σ una superficie di classe C^2 . Allora*

$$\int_{\sigma} \langle \operatorname{rot} \mathbf{a}, \mathbf{n}_{\sigma} \rangle d\sigma = \int_{\partial \sigma} \omega.$$

Il teorema stabilisce che *il flusso del rotore di \mathbf{a} è uguale alla sua circuitazione lungo $\partial \Sigma$, ovvero all'integrale di \mathbf{a} lungo il bordo di Σ .* Indicando con $\mathbf{u}_{\partial \Sigma}$ il vettore tangente alla frontiera $\partial \Sigma$,

$$\int_{\sigma} \langle \nabla \wedge \mathbf{a}, \mathbf{n}_{\sigma} \rangle d\sigma = \int_{\partial \sigma} \langle \mathbf{a}, \mathbf{u}_{\partial \Sigma} \rangle d.s.$$

5.3.1. *La formula di Gauss–Green come caso particolare.* La formula di Gauss–Green può essere reinterpretata come un caso particolare del teorema di Stokes. Supponiamo infatti che la

superficie in questione sia una superficie piatta nel piano, per cui $\mathbf{n}_\sigma \equiv \mathbf{e}_3$. Allora il teorema di Stokes si può scrivere

$$\int_{\sigma} (\partial_x N - \partial_y M) d\sigma = \int_{\partial\sigma} \omega ds$$

che è proprio il teorema di Gauss–Green.