

Questo volume si rivolge agli studenti dei corsi di geometria nelle facoltà scientifiche. Il testo, oltre agli argomenti più tradizionali, contiene risultati matematici poco noti e informazioni difficilmente reperibili, interessanti dal punto di vista didattico e scientifico. L'autore mantiene il giusto equilibrio tra aspetti algebrici e aspetti geometrici e discute a fondo sia la parte teorica sia quella prettamente computazionale degli argomenti. L'attenzione alla didattica è la caratteristica principale del volume: ogni definizione è ampiamente motivata con esempi e spiegazioni; continue domande dirette al lettore, esercizi consigliati e argomenti di riflessione suggeriti spingono lo studente a una lettura attiva e partecipata.

Marco Abate è professore ordinario di Geometria presso la facoltà di Scienze Matematiche, Fisiche e Naturali dell'Università di Roma Tor Vergata.

Geometria

ISBN 88-386-0722-2



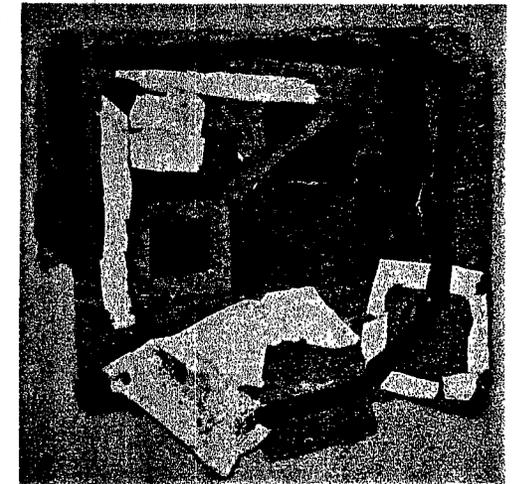
9 788838 607226

certina: Francis Davison, Bright Colours bound with Black, ca 1978/82. Collage: Private Collection.

BRIGHT

marco abate

geometria



McGraw-Hill

Indice

Prefazione	ix
1. Nozioni preliminari	
1.1 Insiemi e funzioni	1
1.2 Logica elementare	8
1.3 Numeri e operazioni	10
1.4 Prerequisiti	13
Esercizi	16
Complementi al Capitolo 1	
1C.1 L'alfabeto greco	18
2. Vettori geometrici	
2.1 Vettori applicati	19
2.2 Coordinate	26
2.3 Equazioni di rette e piani	32
Esercizi	36
Complementi al Capitolo 2	
2C.1 Tutti i triangoli sono isosceli	39
2C.2 Vettori liberi	40
3. L'eliminazione di Gauss	
3.1 Esempi e definizioni	45
3.2 Sistemi triangolari superiori	50
3.3 Il metodo d'eliminazione di Gauss	54
Esercizi	60
Complementi al Capitolo 3	
3C.1 Il principio d'induzione	63
4. Spazi vettoriali	
4.1 Spazi e sottospazi	67
4.2 Combinazioni lineari	72
4.3 Indipendenza lineare e basi	75
4.4 Esistenza delle basi	79
4.5 Somma e intersezione di sottospazi	86
4.6 Somme dirette	88
Esercizi	90
Complementi al Capitolo 4	
4C.1 Il Lemma di Zorn	93
4C.2 L'esistenza di una base	96
4C.3 Cardinalità, e dimensione degli spazi vettoriali	98
Esercizi	104

Copyright © 1996 McGraw-Hill Libri Italia srl
piazza Emilia, 5 - 20129 Milano

McGraw-Hill
A Division of The McGraw-Hill Companies 

I diritti di traduzione, di riproduzione, di memorizzazione elettronica e di adattamento totale e parziale con qualsiasi mezzo (compresi i microfilm e le copie fotostatiche) sono riservati per tutti i paesi.

Editor: Alberto Kratter Thaler
Redazione: Chiara Tartara
Produzione: Ines Farina
Grafica di copertina: Achilli & Piazza e Associati - Milano
Stampa: Arti Grafiche Murelli - Fizzonasco di Pieve Emanuele (MI)

Printed in Italy
4567890MURLLC90
1ª edizione maggio 1996
ISBN 88 386 0722-2

5. Applicazioni lineari	
5.1 Definizioni ed esempi	105
5.2 Nucleo e immagine	111
Esercizi	117
6. Sistemi lineari	
6.1 Sistemi a scala	119
6.2 La riduzione a scala	122
6.3 Tecniche di calcolo	126
6.4 Equazioni parametriche e cartesiane	130
6.5 Sottospazi affini	134
Esercizi	137
7. Matrici e applicazioni lineari	
7.1 Composizione e isomorfismi	141
7.2 Prodotto di matrici	147
7.3 Matrici invertibili	149
Esercizi	152
8. Cambiamenti di base	
8.1 Matrice di cambiamento di base	155
8.2 Matrice associata a un'applicazione lineare	159
Esercizi	165
Complementi al Capitolo 8	
8C.1 Duale e biduali di uno spazio vettoriale	168
8C.2 Trasposta di un'applicazione lineare	173
Esercizi	174
9. Determinanti	
9.1 Esistenza e unicità	175
9.2 Sviluppi di Laplace	187
9.3 Teorema di Binet	191
9.4 Teorema degli orlati	194
Esercizi	197
Complementi al Capitolo 9	
9C.1 Somme	199
10. Geometria affine	
10.1 Equazioni di rette e piani	205
10.2 Punti e rette	208
10.3 Punti e piani	211
10.4 Rette e rette	213
10.5 Rette e piani	216
10.6 Piani e piani	220
10.7 Sistemi di riferimento affini	222
10.8 Orientazione	224
Esercizi	226
Complementi al Capitolo 10	
10C.1 Affinità	230
10C.2 Geometria proiettiva	235

11. I numeri complessi	
11.1 Definizione	243
11.2 Potenze e radici	249
11.3 Il Teorema fondamentale dell'algebra	251
11.4 Rango della trasposta coniugata	253
Esercizi	254
Complementi al Capitolo 11	
11C.1 Divisibilità e radici	256
11C.2 Ideali e massimo comun divisore	259
11C.3 Fattorizzazione di polinomi	263
11C.4 Dimostrazione del Teorema fondamentale dell'algebra	266
Esercizi	268
12. Prodotti scalari ed hermitiani	
12.1 Definizioni ed esempi, caso reale	269
12.2 Definizioni ed esempi, caso complesso	279
12.3 Disuguaglianza di Cauchy-Schwarz	281
12.4 Basi ortogonali	287
12.5 Proiezioni ortogonali	293
12.6 Prodotti scalari e matrici	295
12.7 L'aggiunta	300
12.8 Endomorfismi autoaggiunti e isometrie	303
Esercizi	308
Complementi al Capitolo 12	
12C.1 Il Teorema di rappresentazione di Riesz	311
Esercizi	312
13. Geometria euclidea	
13.1 Angoli e ortogonalità	313
13.2 Distanze	319
13.3 Il prodotto vettore	323
Esercizi	327
Complementi al Capitolo 13	
13C.1 Isometrie	328
13C.2 Coniche	330
13C.3 Esempi di superfici	339
14. Autovalori e autovettori	
14.1 Definizioni ed esempi	343
14.2 Il polinomio caratteristico	349
14.3 molteplicità	353
Esercizi	357
Complementi al Capitolo 14	
14C.1 Il polinomio minimo	360
14C.2 Il Teorema di Sturm	364
Esercizi	368

15. Il Teorema spettrale	
15.1 Basi ortonormali di autovettori	369
15.2 Endomorfismi triangolabili	375
Esercizi	379
Complementi al Capitolo 15	
15C.1 Esistenza della forma canonica di Jordan	382
15C.2 Unicità della forma canonica di Jordan	384
16. Forme quadratiche reali	
16.1 Criteri di positività	389
16.2 Classificazione metrica delle forme quadratiche	395
16.3 Classificazione affine delle forme quadratiche	398
16.4 Quadriche in \mathbb{R}^n	402
Esercizi	419
Complementi al Capitolo 16	
16C.1 Il criterio di Cartesio	421
17. Conclusioni	425
Appendice: soluzione di alcuni esercizi	427
Indice analitico	433

Prefazione

Per uno studente appena iscritto a una facoltà scientifica il primo impatto con il corso di Geometria è indubbiamente duro. Lo studente si aspetta di rivedere quadrati e triangoli, o magari finalmente di imparare qualcosa su sfere e parallelepipedi, e invece nel giro di poche settimane si trova immerso in nozioni astruse tipo "spazi vettoriali di dimensione n " o peggio ancora in manipolazioni apparentemente prive di senso di tabelle di numeri chiamate per qualche misterioso motivo "matrici". Gli viene promesso che un giorno tutto ciò gli sarà utile, ma per il momento il nostro studente a volte non capisce neppure se c'è una relazione fra quanto studiato per anni e la serie di concetti astratti che gli vengono propinati spesso quasi senza motivazioni.

Eppure il corpo di argomenti che viene tradizionalmente insegnato nel primo corso di Geometria nelle università italiane non è nato nel vuoto, anzi; è stato creato pazientemente negli ultimi due secoli per rispondere a problemi ben precisi, sia applicativi sia interni alla Matematica, e diversi di questi problemi sono di interesse immediato anche per uno studente del primo anno. Per esempio, lo studio dei sistemi lineari conduce in maniera naturale agli spazi vettoriali e alle applicazioni lineari; lo sforzo di astrazione viene poi ripagato da risultati quali il Teorema di Rouché-Capelli o il Teorema della dimensione che, affiancati da appropriate tecniche di calcolo del genere dell'eliminazione di Gauss, permettono di risolvere efficacemente problemi che lo stesso studente può essersi posto sui sistemi lineari. E una volta entrati all'interno della teoria diventa facile trovare ulteriori motivazioni per proseguire.

Dopo un simile preambolo non è difficile immaginare qual è una delle caratteristiche principali di questo testo: l'abbondanza di motivazioni ed esempi. Ogni nuovo concetto è collegato a qualcosa che lo studente già conosce, e che ne giustifica l'introduzione; e numerosi esempi vengono discussi per dare la possibilità di vedere in concreto il significato delle nuove nozioni e i possibili fenomeni che si possono presentare. Una certa enfasi è posta anche sulle tecniche di calcolo, con l'obiettivo di fornire prima possibile allo studente i mezzi per operare autonomamente con i nuovi concetti; ma tutte le tecniche sono rigorosamente dimostrate e inquadrare nel contesto, in modo da evitare che divengano delle semplici macchinette da utilizzare anche a sproposito senza capirne davvero il funzionamento.

Un altro problema che si deve affrontare studiando argomenti matematici (di qualunque livello, non solo al primo anno d'università) è che la semplice lettura del testo

non basta. Un ragionamento matematico è per sua stessa natura una successione logica di argomenti; viene quindi naturale seguire i singoli passaggi senza porsi problemi (del resto, è tutto logico) per poi ritrovarsi in fondo al capitolo e accorgersi: primo, di non saper risolvere un solo esercizio; secondo, di non aver minimamente capito perché siano stati fatti certi ragionamenti piuttosto che altri; terzo, di non essere assolutamente in grado di ricostruirli. Per cercare di prevenire il problema questo testo è stato scritto con uno stile che invita a una lettura attiva. In punti strategici viene chiesto allo studente di verificare subito d'aver capito il motivo di certi passaggi; alcuni dettagli semplici ma significativi di alcune dimostrazioni sono lasciati per esercizio (ma le soluzioni sono tutte in Appendice); e sono state preferite dimostrazioni che mostrassero chiaramente cosa stava accadendo ad altre magari (ma non necessariamente) più brevi o eleganti ma più oscure. Inoltre, per invogliare ulteriormente lo studente a interagire con ciò che legge, il testo si rivolge direttamente in seconda persona al lettore. Infine, ciascun capitolo è corredato da numerosi esercizi per permettere di verificare la comprensione sia teorica che pratica di quanto studiato.

Gli argomenti presentati in questo testo coprono le necessità fondamentali di un corso di Geometria per il primo anno dei corsi di laurea in Ingegneria, Matematica e Fisica: vettori applicati, sistemi lineari, spazi vettoriali, applicazioni lineari, matrici, determinanti, geometria affine ed euclidea del piano e dello spazio, numeri complessi, prodotti scalari, autovalori e autovettori, forme quadratiche, coniche e quadriche. Buona parte dei capitoli sono corredati di Complementi contenenti materiale aggiuntivo, utile per permettere a ciascun docente di adattare il testo alle esigenze specifiche del proprio corso (per esempio, il polinomio minimo, la forma canonica di Jordan o un'introduzione alla geometria proiettiva), per fornire informazioni di cultura generale allo studente (per esempio, un'esposizione elementare del Lemma di Zorn, o la dimostrazione del Criterio di Cartesio), o semplicemente per illustrare l'uso di tecniche specifiche (per esempio, il principio d'induzione o le sommatorie).

Questo libro nasce dalle note per i corsi di Geometria da me tenuti presso le Facoltà di Ingegneria delle Università di Roma Tor Vergata, Pisa e Ancona. Diverse persone ne hanno influenzato la stesura, indirettamente o direttamente, ed è un piacere avere la possibilità di ringraziarle qui: in ordine alfabetico, Silvana Abeasis, Welleda Baldoni, Marilù Chiofalo, Franco Conti, Tullio Franzoni, Laura Geatti, Mario Landucci, Adele Manzella, Giorgio Patrizio e Roberto Tauraso. Un ringraziamento particolare va a Chiara Tartara e allo staff della McGraw-Hill Libri Italia, che hanno reso la realizzazione di questa opera il più scorrevole possibile. E infine soprattutto un saluto speciale ai miei studenti di questi anni (in particolare all'indimenticato *Abate Funs Club*); senza la loro stima, il puntiglioso controllo su quanto propinavo loro e i continui appropriati commenti nulla di tutto ciò sarebbe stato possibile.

Marco Abate

Nozioni preliminari

Per cominciare lo studio di un nuovo argomento è utile fissare un punto di partenza. Bisogna stabilire un linguaggio comune, definire i termini di uso frequente e convenire in anticipo su cosa sia supposto noto. Per i nostri scopi, un buon punto di partenza è la teoria ingenua¹ degli insiemi. Probabilmente l'avrai già incontrata nei tuoi studi precedenti; per ogni evenienza questo capitolo comincia richiamandone i concetti di base. Si prosegue con una breve discussione dei numeri e relative operazioni, sempre con l'intento di ripassare alcuni concetti fondamentali e fissare una volta per tutte la terminologia. Il capitolo si conclude con un veloce riepilogo dei concetti matematici che si suppone tu abbia imparato alle superiori.

1.1 Insiemi e funzioni

Cominciamo introducendo brevemente alcune notazioni e un po' di terminologia (ovvero qualche simbolo e non pochi nomi). Se A è un insieme, scriveremo $a \in A$ per indicare che l'elemento a appartiene all'insieme A , e $a \notin A$ per indicare che a non appartiene ad A . Se B è un altro insieme i cui elementi appartengono tutti anche ad A , diremo che B è un *sottoinsieme* di A (o che è *contenuto* in A), e scriveremo $B \subseteq A$ (oppure $A \supseteq B$, che si legge " A contiene B "). Se inoltre B è effettivamente diverso da A — cioè A contiene degli elementi che non appartengono a B — diremo che B è un *sottoinsieme proprio* di A , e scriveremo $B \subset A$ (o $A \supset B$). L'*insieme vuoto*, cioè l'insieme privo di elementi, sarà indicato con \emptyset .

Una situazione che capiterà spesso sarà quella di dover considerare "il sottoinsieme B degli elementi dell'insieme A che godono della proprietà tale". In simboli, questa definizione sarà abbreviata in

$$B = \{a \in A \mid a \text{ gode della proprietà tale}\}.$$

Per esempio, se indichiamo con \mathbb{N} l'insieme dei numeri naturali, il sottoinsieme B dei

¹ Così chiamata per distinguerla dalla ben più complicata e interessante teoria assiomatica degli insiemi, di cui parleremo brevemente nei Complementi al Capitolo 4.

multipli di 3 può essere rappresentato da

$$B = \{n \in \mathbb{N} \mid n \text{ è divisibile per } 3\},$$

o anche da

$$B = \{n \in \mathbb{N} \mid n = 3k \text{ per qualche } k \in \mathbb{N}\}.$$

Siano A e B due insiemi qualunque. Allora

- l'intersezione $A \cap B$ di A e B è l'insieme contenente solo gli elementi che stanno sia in A che in B ; se $A \cap B = \emptyset$, diremo che gli insiemi A e B sono *disgiunti*.
- l'unione $A \cup B$ di A e B è l'insieme che contiene tutti gli elementi di A assieme a tutti gli elementi di B ;
- la differenza $A \setminus B$ di A e B è l'insieme che contiene esattamente quegli elementi di A che non stanno in B ;
- il prodotto cartesiano $A \times B$ di A e B è l'insieme delle coppie ordinate (a, b) dove a è un qualunque elemento di A e b è un qualunque elemento di B .

Se esiste un concetto che caratterizza la matematica moderna è il concetto di funzione. Praticamente tutta la matematica (e la Geometria non è da meno) ne fa un uso continuo. Non ti sorprenderà quindi scoprire che il resto di questo primo paragrafo è dedicato a una discussione delle funzioni e delle loro proprietà.

Definizione 1.1 Una funzione (o applicazione) fra due insiemi A e B è una legge che associa a ciascun elemento di A uno e un solo elemento di B . L'insieme di partenza A è il *dominio* della funzione; l'insieme di arrivo B il *codominio*. In simboli, una funzione f di dominio A e codominio B verrà indicata con $f: A \rightarrow B$. Se la funzione f manda l'elemento $a \in A$ nell'elemento $b \in B$, scriveremo² $b = f(a)$, e diremo che b è *immagine di a tramite f* . L'insieme degli elementi di B che sono immagine tramite f di elementi di A è l'*immagine di f* , e viene indicata con $\text{Im } f$ oppure con $f(A)$; in simboli,

$$f(A) = \{b \in B \mid b = f(a) \text{ per qualche } a \in A\} = \{f(a) \in B \mid a \in A\}.$$

Più in generale, se $A_1 \subseteq A$ è un sottoinsieme di A , l'*immagine di A_1 tramite f* è l'insieme $f(A_1) \subseteq B$ delle immagini degli elementi di A_1 . Viceversa, se $B_1 \subseteq B$, l'insieme degli elementi di A la cui immagine tramite f appartiene a B_1 si chiama *immagine inversa $f^{-1}(B_1)$ di B_1 tramite f* ; in simboli,

$$f^{-1}(B_1) = \{a \in A \mid f(a) \in B_1\} \subseteq A.$$

Chiaramente, $f^{-1}(B) = A$ per qualunque funzione $f: A \rightarrow B$ (perché?).

² Oppure $a \mapsto b$, o semplicemente $a \mapsto b$, se il contesto individua chiaramente di quale funzione si tratta.

Osservazione 1.1 Possiamo pensare la funzione $f: A \rightarrow B$ come una specie di scatola nera, con un ingresso e un'uscita. Ogni volta che in ingresso entra un elemento del dominio, la scatola nera — la funzione — lo elabora e poi emette dall'uscita un elemento del codominio. Non è importante la natura degli elementi del dominio e del codominio (possono essere numeri, rette, patate, cavalleggeri prussiani o qualsiasi altra cosa) né il tipo di processi digestivi che avvengono all'interno della scatola. Siano somme, prodotti, classifiche o formine da sabbia, tutto è ammissibile, purché il procedimento usato sia sempre lo stesso: ogni volta che in ingresso infiliamo la stessa patata, in uscita dobbiamo ottenere sempre la stessa cipolla — a ogni elemento del dominio viene associato uno e un solo elemento del codominio, appunto.

Questa analogia ci permette di dire quando due funzioni sono uguali. Per i nostri scopi, due scatole nere che producono lo stesso oggetto quando in ingresso ricevono lo stesso elemento sono indistinguibili: non potendo vedere come sono fatte dentro, se si comportano nello stesso modo per noi sono la stessa scatola. Dunque:

Definizione 1.2 Due funzioni $f: A \rightarrow B$ e $g: A \rightarrow B$ (con lo stesso dominio e lo stesso codominio, s'intende) sono *uguali*, e scriveremo $f = g$ oppure $f \equiv g$, se e solo se $f(a) = g(a)$ per ogni $a \in A$.

Vediamo ora qualche esempio.

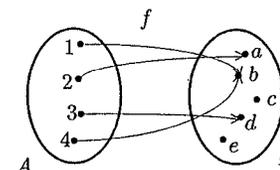


Figura 1.1 Rappresentazione grafica di una funzione.

ESEMPIO 1.1 Se A è un insieme, la funzione $\text{id}_A: A \rightarrow A$ che associa a ogni elemento di A se stesso (in simboli, $\text{id}_A(a) = a$ per ogni $a \in A$) si chiama l'*identità* di A .

ESEMPIO 1.2 Siano A e B insiemi, e $b_0 \in B$ un elemento fissato. La legge $f: A \rightarrow B$ che associa b_0 a ogni elemento di A (in simboli, $f(a) = b_0$ per ogni $a \in A$) è una funzione, detta *funzione costante* di valore b_0 .

ESEMPIO 1.3 La legge $f: \mathbb{N} \rightarrow \mathbb{N}$ data da $f(n) = 3n$ è una funzione.

ESEMPIO 1.4 Siano A e B gli insiemi $A = \{1, 2, 3, 4\}$ e $B = \{a, b, c, d, e\}$. La legge $f: A \rightarrow B$ data da $f(1) = b$, $f(2) = a$, $f(3) = d$, $f(4) = b$ è una funzione. Possiamo rappresentarla con un disegno come in Figura 1.1.

Ogni funzione in cui sia il dominio che il codominio sono costituiti da un numero finito di punti può venire visualizzata con un disegno di questo genere: l'unica cosa

importante è che da *ogni* punto del dominio deve partire *una e una sola* freccia. Per intenderci, i due disegni della Figura 1.2 *non* rappresentano funzioni: il primo perché a un elemento del dominio vengono associati due elementi del codominio, e il secondo perché a un elemento del dominio non viene associato alcun elemento del codominio.

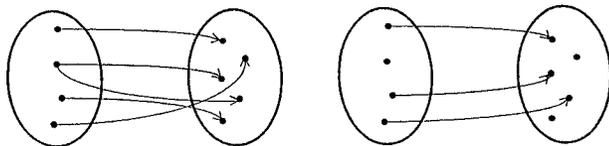


Figura 1.2 Leggi che non rappresentano funzioni.

ESEMPIO 1.5 Sia $f: A \rightarrow B$ la funzione definita nell'Esempio precedente, e poniamo $A_1 = \{2, 4\}$ e $B_1 = \{b, d, e\}$. Allora si ha $f(A_1) = \{a, b\}$ e $f^{-1}(B_1) = \{1, 3, 4\}$.

ESEMPIO 1.6 La classifica della quinta giornata del girone d'andata del campionato di serie A del 1992/93 è una funzione che associa a ogni squadra di serie A un numero naturale: il suo punteggio. Da questo punto di vista, il campionato consiste nel cambiare funzione ogni domenica.

ESEMPIO 1.7 Il prezzo al litro della benzina super dal benzinaio all'angolo è una funzione che associa a ogni giorno degli ultimi trent'anni un numero — ed è una funzione sfortunatamente non decrescente.

ESEMPIO 1.8 La legge che associa a ogni giorno dell'anno 1995 il prezzo al litro della benzina super in Italia *non* è una funzione. Infatti, il prezzo cambia da benzinaio a benzinaio, per cui non è possibile associare a ciascun giorno dell'anno un unico prezzo.

ESEMPIO 1.9 Un tipo particolare di funzioni è costituito dai polinomi. Un polinomio (a coefficienti reali, in una variabile³) è una funzione $p: \mathbb{R} \rightarrow \mathbb{R}$ della forma

$$p(t) = a_n t^n + a_{n-1} t^{n-1} + \dots + a_0, \quad (1.1)$$

dove \mathbb{R} indica l'insieme dei numeri reali, $n \in \mathbb{N}$ è un numero naturale detto *grado* del polinomio, e $a_0, \dots, a_n \in \mathbb{R}$ sono numeri reali, i *coefficienti* del polinomio. In particolare, a_0 è detto *termine noto*, e $a_n \neq 0$ *coefficiente direttivo*. I polinomi di grado zero sono esattamente le costanti. Il significato della scrittura (1.1) è il seguente: il polinomio p associa a ogni numero reale t il numero reale ottenuto prendendo t , elevandolo alla potenza n , moltiplicandolo per a_n , riprendendo t per moltiplicarlo per a_{n-1} dopo averlo elevato alla potenza $n-1$ in modo da poter sommare il risultato

³ Una volta vista la definizione dovrebbe essere chiaro come costruire polinomi in più variabili, e vedremo fra poco che ne esistono anche con altri tipi di coefficienti.

a quanto già avevamo, e così via, fino a sommare a_0 ottenendo finalmente il valore finale $p(t)$. Per esempio, se p è il polinomio dato da $p(t) = 2t^2 - t + 3\pi$, allora

$$p(-2) = 2(-2)^2 - (-2) + 3\pi = 10 + 3\pi.$$

Per noi, un polinomio sarà *sempre* una funzione, e *mai* una misteriosa combinazione formale di lettere e numeri. Indicheremo con $\mathbb{R}[t]$ l'insieme di tutti i polinomi (a coefficienti reali, in una variabile), e con $\mathbb{R}_n[t]$ l'insieme dei polinomi di grado minore o uguale a $n \in \mathbb{N}$. Una discussione alquanto più approfondita dei polinomi e delle loro proprietà è contenuta nei Complementi al Capitolo 11.

ESEMPIO 1.10 Le funzioni trigonometriche seno e coseno sono funzioni da \mathbb{R} in \mathbb{R} , come pure l'esponenziale; il logaritmo invece è una funzione con l'insieme dei numeri reali positivi come dominio ed \mathbb{R} come codominio.

A questo punto dovrete essere convinto (mi auguro) dell'esistenza di miriadi di funzioni diverse; possiamo quindi tornare a parlare di concetti generali.

Definizione 1.3 Sia $f: A \rightarrow B$ una funzione da A a B . Se A_1 è un sottoinsieme di A , la funzione f chiaramente ci determina una legge che a ogni elemento di A_1 associa un elemento di B (è la stessa legge di prima), e quindi una funzione da A_1 a B . Questa funzione si chiama *restrizione* di f ad A_1 , e si indica con $f|_{A_1}$. Se capiterà, scriveremo $f(A_1)$ e non $f|_{A_1}(A_1)$ per indicare l'immagine di f ristretta ad A_1 , in modo da non complicare troppo le formule.

ESEMPIO 1.11 La funzione $f = p|_{\mathbb{N}}: \mathbb{N} \rightarrow \mathbb{R}$ che associa a ogni numero naturale n il suo successore $n+1$ è la restrizione a \mathbb{N} del polinomio $p: \mathbb{R} \rightarrow \mathbb{R}$ dato da $p(t) = t+1$.

Ovviamente, non tutte le funzioni godono delle stesse proprietà. Per esempio, non sempre tutti gli elementi del codominio di una funzione sono immagine di elementi del dominio; in generale, l'immagine è un sottoinsieme proprio del codominio. Le funzioni per cui ciò non accade meritano un nome speciale.

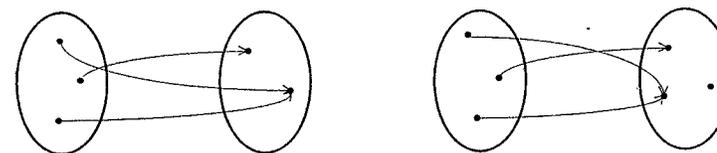


Figura 1.3 Funzioni surgettive e non surgettive.

Definizione 1.4 Se la funzione $f: A \rightarrow B$ è tale che ogni elemento del codominio arriva da uno del dominio — cioè $\text{Im } f = B$ — diremo che f è *surgettiva* (o *suriettiva*).

In termini della rappresentazione grafica introdotta nell'Esempio 1.4, una funzione è surgettiva se ogni elemento del codominio è raggiunto da almeno una freccia: nella Figura 1.3 la funzione a sinistra è surgettiva, mentre quella a destra non lo è.

Un'altra cosa che spesso accade è che a due elementi diversi del dominio la funzione f associi la stessa immagine. Di nuovo, le funzioni per cui questo non è vero hanno un nome.

Definizione 1.5 Se la funzione $f: A \rightarrow B$ associa elementi diversi del codominio a elementi diversi del dominio — cioè se $a_1 \neq a_2$ implica $f(a_1) \neq f(a_2)$ — diremo che la funzione f è *iniettiva*.

In termini della solita rappresentazione grafica, una funzione è iniettiva se su ogni elemento del codominio arriva al più una freccia (ma può anche non arrivarne alcuna): nella Figura 1.4 la funzione a sinistra è iniettiva mentre quella a destra non lo è.

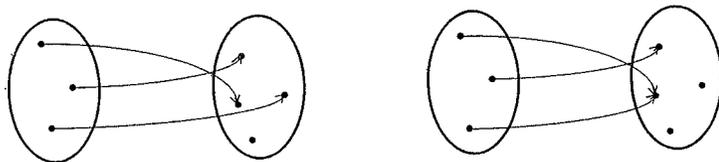


Figura 1.4 Funzioni iniettive e non iniettive.

Definizione 1.6 Una funzione $f: A \rightarrow B$ sia iniettiva che surgettiva verrà detta *bigettiva* (o *biiettiva*, o *biunivoca*). Una funzione bigettiva associa a ciascun elemento del codominio uno e un solo elemento del dominio⁴; se $b \in B$, esiste un unico $a \in A$ tale che $f(a) = b$. Questo ci permette di definire una funzione da B ad A , la *funzione inversa*, $f^{-1}: B \rightarrow A$, ponendo $f^{-1}(b) = a$, dove $a \in A$ è quell'unico elemento tale che $f(a) = b$. A volte, invece di funzione bigettiva diremo funzione *invertibile*.

Usando di nuovo la rappresentazione dell'Esempio 1.4, una funzione è bigettiva se su ogni elemento del codominio arriva esattamente una freccia, e in tal caso la funzione inversa si ottiene invertendo il senso delle frecce. Per intenderci, nessuna delle funzioni nelle Figure 1.3 e 1.4 era bigettiva, mentre la Figura 1.5 ce ne mostra una invertibile (a sinistra) assieme alla sua inversa (a destra).

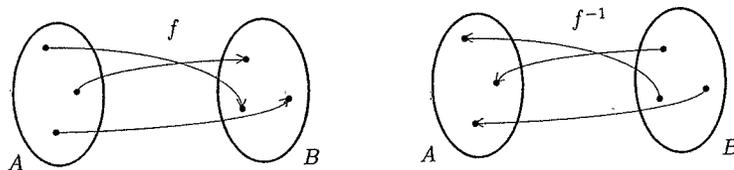


Figura 1.5 Una funzione bigettiva e la sua inversa.

⁴ In un certo senso, il dominio e il codominio si trovano ad avere lo stesso numero di elementi: in A ce ne sono tanti quanti in B . Ne ripareremo nei Complementi al Capitolo 4.

Osservazione 1.2 Attenzione a non confondere i concetti di funzione inversa e di immagine inversa. La funzione inversa f^{-1} associa a ogni elemento del codominio di f un elemento del dominio di f , ed esiste soltanto quando la funzione f è bigettiva. L'immagine inversa, invece, associa a un sottoinsieme del codominio un sottoinsieme del dominio — per cui *non* è una funzione definita sul codominio — ed esiste sempre, anche quando la funzione f non è bigettiva.

Osservazione 1.3 Attenzione anche a non confondere i concetti di funzione e di funzione iniettiva. Una funzione $f: A \rightarrow B$ associa sempre a ogni elemento di A uno e un solo elemento di B ; per una funzione iniettiva invece ogni elemento di B è immagine di al più un elemento di A , che è un concetto ben diverso. In una funzione qualunque, da ogni elemento del dominio parte esattamente una freccia; in una funzione iniettiva, su ogni elemento del codominio arriva al più una freccia.

Concludiamo questa breve carrellata sulle funzioni con un'ultima definizione.

Definizione 1.7 Supponiamo di avere due funzioni $f: A \rightarrow B$ e $g: B \rightarrow C$, dove il codominio di f coincide col dominio di g . In tal caso possiamo definire una nuova funzione, la *composizione* $g \circ f: A \rightarrow C$ delle funzioni f e g , tramite la formula

$$\forall a \in A \quad (g \circ f)(a) = g(f(a)). \quad (1.2)$$

La $g \circ f$ è effettivamente una funzione: infatti, a ciascun $a \in A$ associa un unico elemento di B , e a quest'ultimo un unico elemento di C , per cui è una legge che a ciascun punto di A associa uno e un sol punto di C .

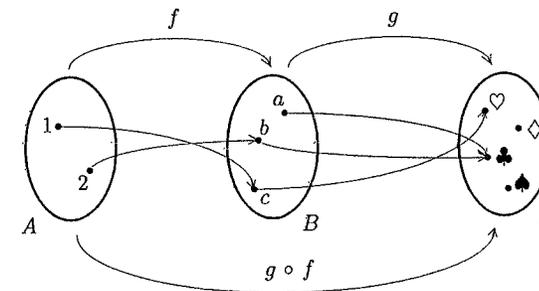


Figura 1.6 Composizione di funzioni.

ESEMPIO 1.12 Prendiamo i tre insiemi $A = \{1, 2\}$, $B = \{a, b, c\}$, $C = \{\heartsuit, \diamondsuit, \clubsuit, \spadesuit\}$ e le funzioni $f: A \rightarrow B$ e $g: B \rightarrow C$ date da $f(1) = c$, $f(2) = b$ e da $g(a) = \clubsuit = g(b)$, e $g(c) = \heartsuit$. Allora la composizione di f e g è la funzione $g \circ f: A \rightarrow C$ data da $(g \circ f)(1) = \heartsuit$ e $(g \circ f)(2) = \clubsuit$ (vedi la Figura 1.6)

ESEMPIO 1.13 Siano $f: \mathbb{N} \rightarrow \mathbb{R}$ e $g: \mathbb{R} \rightarrow \mathbb{R}$ le funzioni date da $f(n) = 2n + 3$ e da $g(x) = 2^x$. Allora la composizione di f e g è la funzione $g \circ f: \mathbb{N} \rightarrow \mathbb{R}$ data da

$$(g \circ f)(n) = g(f(n)) = g(2n + 3) = 2^{2n+3}.$$

Nota che l'immagine di $g \circ f$ è contenuta in \mathbb{N} , anche se l'immagine di g è ben più grande (è uguale a \mathbb{R}^+).

1.2 Logica elementare

Fare della matematica vuol dire giungere a determinate conclusioni facendo certi ragionamenti⁵. Capita dunque spesso di usare frasi come "implica", "per ogni" o simili; per risparmiare tempo sono stati inventati dei simboli che le rappresentano.

Invece di "implica" capiterà di scrivere " \implies ", mentre "è equivalente a" sarà talvolta sostituito da " \iff ". Può succedere che "tale che" sia sostituito da due punti ":" o da una barra verticale "|". Invece di "per ogni" a volte scriveremo " \forall ", e al posto di "esiste" a volte scriveremo " \exists ". Un esempio per chiarire: se P è l'insieme dei numeri pari, e \mathbb{N} l'insieme dei numeri naturali, la frase "per ogni numero pari a esiste un numero naturale m tale che $a = 2m$ " diventa

$$\forall a \in P \quad \exists m \in \mathbb{N}: a = 2m.$$

Capiterà talvolta anche di scrivere " $\exists!$ ", che significa "esiste unico". Per esempio, se volessimo sottolineare che esiste un *unico* numero naturale m tale che $a = 2m$ potremmo scrivere

$$\forall a \in P \quad \exists! m \in \mathbb{N}: a = 2m.$$

ESEMPIO 1.14 Proviamo a scrivere con questi simboli le definizioni di funzione e di funzione bigettiva. La legge $f: A \rightarrow B$ è una funzione se

$$\forall a \in A \quad \exists! b \in B: b = f(a);$$

una funzione $f: A \rightarrow B$ è bigettiva se e solo se

$$\forall b \in B \quad \exists! a \in A: b = f(a).$$

Osservazione 1.4 Non si possono scambiare impunemente i simboli \forall ed \exists ; la formula

$$\forall a \in P \quad \exists m \in \mathbb{N}: a = 2m$$

⁵ Lo studio della struttura di questi ragionamenti è compito della logica matematica.

significa una cosa ben diversa dalla formula

$$\exists m \in \mathbb{N}: \forall a \in P a = 2m.$$

La prima formula vuol dire "per ogni numero pari a esiste un numero naturale m (che dipende da a) tale che $a = 2m$ ", che è ovviamente vero. La seconda formula invece vuol dire che "esiste un numero naturale m (uno solo, ben determinato) tale che per ogni numero pari a si ha $a = 2m$ ", ovvero ogni numero pari è il doppio del nostro m , lo stesso qualunque sia il numero pari considerato, affermazione chiaramente falsa.

Capiterà più volte in seguito di dover negare una frase che comincia con "per ogni" o con "esiste". La negazione esatta di "per ogni a succede questo" è "non per ogni a succede questo", ovvero "esiste un a per cui non succede questo", che è ben diverso dal dire "per ogni a non succede questo". Analogamente, la negazione esatta di "esiste un b per cui succede questo" è "non esiste un b per cui succede questo", cioè "per ogni b non succede questo".

ESEMPIO 1.15 Vogliamo negare la frase "tutti i gatti sono verdi" (che è un modo più corretto grammaticalmente di dire "per ogni gatto succede che il gatto è verde"). Come abbiamo appena osservato, la negazione esatta è "non tutti i gatti sono verdi", ovvero "esiste almeno un gatto che non è verde". La frase "nessun gatto è verde" (cioè "non esiste un gatto verde") pur essendo vera *non* è la negazione del nostro enunciato originale "tutti i gatti sono verdi"; è un'affermazione molto più forte, e molto più difficile da verificare (per far vedere che "tutti i gatti sono verdi" è falsa basta trovare *un solo* gatto non verde; per far vedere che "nessun gatto è verde" è vera devi controllare il colore di tutti i gatti sulla terra).

ESEMPIO 1.16 Qual è la negazione della frase "esiste vita sugli altri pianeti del sistema solare"? La risposta esatta è "non esiste vita sugli altri pianeti del sistema solare", cioè "tutti gli altri pianeti del sistema solare sono privi di vita". L'affermazione "esiste un pianeta del sistema solare privo di vita" non esclude che ci sia un altro pianeta ancora su cui esiste la vita, per cui non è la risposta esatta.

Come forse già sai, una parte notevole della matematica consiste nel decidere se certe affermazioni sono vere o false. In alcuni casi, per stabilirlo basta un esempio; in altri, invece, anche diecimila esempi sono inutili, ed è necessario un ragionamento che copra in una volta sola tutti i casi possibili (in altre parole, è necessaria una *dimostrazione*). Orbene, un tipico problema dello studente novizio è esattamente capire quando è necessaria una dimostrazione, e quando invece è sufficiente un esempio.

L'idea di fondo è che la dimostrazione è legata al "per ogni", mentre l'esempio all'"esiste". Per vedere se l'affermazione "per ogni a succede questo" è vera, devi *dimostrarlo* con un ragionamento valido per ogni valore di a . Invece, per vedere se l'affermazione "esiste un b per cui succede questo" è vera ti basta trovare un singolo esempio (un singolo valore di b) per cui è vero.

ESEMPIO 1.17 Supponiamo di voler vedere se l'affermazione "ogni *marine* americano possiede una divisa verde" è vera. In questo caso gli esempi sono inutili: anche

dopo aver controllato gli indumenti di migliaia di *marine* non potremmo ancora escludere l'esistenza di un *marine* senza divise verdi. Ci serve un ragionamento generale; possiamo per esempio dire che il regolamento militare prescrive senza eccezioni che ogni *marine* abbia una divisa verde, e così *dimostrare* la verità della nostra affermazione senza bisogno di esempi.

ESEMPIO 1.18 Adesso vogliamo invece stabilire la verità dell'affermazione "talvolta piove di domenica". In questo caso basta un esempio; è sufficiente una domenica di pioggia per verificare che l'affermazione è corretta.

Riassumendo: quando ti viene chiesto di decidere se l'affermazione "per ogni a succede questo" è vera o falsa, tu hai due possibilità: se ritieni che sia vera, devi dimostrarlo per qualunque valore di a ; se invece ritieni che sia falsa (cioè che sia vero che "esiste un a per cui questo non succede"); ti basta trovare un esempio in cui è falsa. Analogamente, per far vedere che l'affermazione "esiste un b per cui succede questo" è vera basta trovare un esempio, cioè un b specifico per cui "questo" succede; se invece ritieni sia falsa, devi dimostrare che per ogni valore di b "questo" non accade.

Un'altra frase che compare spesso in matematica è "se succede A allora capita anche B ", che si abbrevia⁶ in " A implica B " o addirittura in " $A \implies B$ ". È importante rendersi conto che una frase del genere non dice nulla su B quando A non si verifica. Un esempio per chiarire: anche se la frase "se si gioca di martedì allora il Pontedera è in testa alla classifica del campionato di serie A" fosse vera, non sapremmo nulla sulla effettiva posizione in classifica del Pontedera, in quanto le partite si giocano la domenica. In particolare, la frase "se succede A allora capita anche B " è falsa se e solo se contemporaneamente A è vera e B è falsa (si gioca di martedì e il Pontedera non è in testa alla classifica). Dunque la negazione di " A implica B " è " A non implica B ", cioè "capita che A sia vera e B sia falsa".

1.3 Numeri e operazioni

Per i nostri scopi alcuni insiemi sono più importanti di altri, e quindi si meritano un nome e un simbolo specifici. Prima di tutto abbiamo l'insieme \mathbb{N} dei *numeri naturali*:

$$0, 1, 2, 3, \dots$$

Poi troviamo l'insieme \mathbb{Z} dei *numeri interi* (o *numeri relativi*):

$$0, +1, -1, +2, -2, \dots;$$

Subito dopo viene l'insieme \mathbb{Q} dei *numeri razionali*, ovvero le frazioni

$$0, 1, -1, 1/2, -1/2, 2, -2, 1/3, -1/3, 2/3, -2/3, \dots$$

⁶ Si dice anche che A è *condizione sufficiente* perché accada B , e che B è *condizione necessaria* perché succeda A .

Come spero ricorderai, i numeri razionali corrispondono ai numeri decimali periodici. L'insieme di tutti i numeri decimali, periodici e no, è l'insieme dei *numeri reali*, e verrà indicato con \mathbb{R} . Non ci interessa qui una definizione formalmente corretta dei numeri reali; è importante però che tu abbia ben presente due cose. Prima di tutto, numeri come $\sqrt{2}$, e , π e simili sono tutti numeri reali. In secondo luogo, una volta fissata un'origine e un'unità di misura i numeri reali si possono mettere in corrispondenza biunivoca con i punti di una retta: a ogni numero reale corrisponde uno e un sol punto della retta, e viceversa.

Ognuno di questi insiemi numerici è un'estensione del precedente (nel senso che lo contiene come sottoinsieme proprio); per l'esattezza, si ha

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}.$$

Più avanti (nel Capitolo 11) incontreremo i *numeri complessi*, un altro insieme numerico (che verrà indicato con \mathbb{C}) contenente propriamente i numeri reali. Infine, l'insieme dei numeri naturali (rispettivamente dei numeri interi, razionali, reali) non nulli verrà indicato con \mathbb{N}^* (rispettivamente \mathbb{Z}^* , \mathbb{Q}^* , \mathbb{R}^*), e l'insieme dei numeri razionali (rispettivamente dei numeri reali) positivi con \mathbb{Q}^+ (rispettivamente \mathbb{R}^+).

Se a e b sono due numeri (reali, per esempio), sappiamo stabilire qual è il più grande e qual è il più piccolo. Scriveremo $a \leq b$ (oppure $b \geq a$) se a è *minore o uguale* a b (se b è *maggiore o uguale* ad a), che è vero anche in caso sia $a = b$. Se invece vogliamo dire che a è *strettamente minore* di b (ovvero che b è *strettamente maggiore* di a) scriveremo $a < b$ (rispettivamente, $a > b$). Infine, indicheremo con $\max\{a, b\}$ il *massimo* (cioè il più grande) fra a e b , e con $\min\{a, b\}$ il *minimo* (cioè il più piccolo) fra a e b .

In tutti i nostri insiemi numerici sono definite due operazioni, la *somma* (+) e il *prodotto* (\cdot), che godono di alcune delle seguenti proprietà:

$$(i) \quad \forall a, b, c \quad (a + b) + c = a + (b + c)$$

(proprietà associativa della somma);

$$(ii) \quad \exists a_0: \forall a \quad a + a_0 = a_0 + a = a$$

(esistenza dell'elemento neutro per la somma), e di solito si scrive 0 al posto di a_0 per indicare l'elemento neutro;

$$(iii) \quad \forall a \exists a': \quad a + a' = a' + a = a_0$$

(esistenza dell'opposto), e di solito si scrive $-a$ invece di a' per indicare l'opposto;

$$(iv) \quad \forall a, b \quad a + b = b + a$$

(proprietà commutativa della somma);

$$(v) \quad \forall a, b, c \quad a \cdot (b + c) = a \cdot b + a \cdot c \quad \text{e} \quad (b + c) \cdot a = b \cdot a + c \cdot a$$

(proprietà distributiva della somma rispetto al prodotto);

$$(vi) \quad \forall a, b, c \quad (a \cdot b) \cdot c = a \cdot (b \cdot c)$$

(proprietà associativa del prodotto);

$$(vii) \quad \exists a_1: \forall a \quad a \cdot a_1 = a_1 \cdot a = a$$

(esistenza dell'elemento neutro per il prodotto), e di solito si scrive 1 al posto di a_1 per indicare l'elemento neutro;

$$(viii) \quad \forall a \neq a_0 \exists \bar{a}: \quad a \cdot \bar{a} = \bar{a} \cdot a = a_1$$

(esistenza dell'inverso), e di solito si scrive a^{-1} invece di \bar{a} per indicare l'inverso;

$$(ix) \quad \forall a, b \quad a \cdot b = b \cdot a$$

(proprietà commutativa del prodotto).

A queste proprietà è associata una serie di definizioni.

Definizione 1.8 Un gruppo è un insieme G in cui è definita un'operazione (che può essere indicata sia con il segno $+$ che con il segno \cdot piuttosto che con simboli più esoterici come \ast o \circ) che soddisfa le proprietà (i)–(iii); se vale anche la (iv), G è detto gruppo commutativo.

Per esempio, \mathbb{Z} con la somma o \mathbb{R}^* col prodotto (ma non \mathbb{R} col prodotto, o \mathbb{N} con la somma; perché?) sono gruppi commutativi. Vedremo più avanti un esempio naturale di gruppo non commutativo (l'insieme delle matrici quadrate invertibili col prodotto righe per colonne).

Definizione 1.9 Un anello è un insieme A su cui sono definite due operazioni (di solito indicate con $+$ e \cdot) che soddisfano le proprietà (i)–(vii). Se vale anche la (ix), l'insieme A è un anello commutativo; se vale anche la (viii) — ma non la (ix) — A è un corpo; se infine valgono tutte le proprietà (i)–(ix), A viene chiamato campo.

Per esempio, \mathbb{Z} è un anello commutativo ma non un campo, mentre \mathbb{Q} e \mathbb{R} (e, quando lo incontrerai, anche \mathbb{C}) sono campi; vedremo nel seguito che le matrici quadrate sono un esempio di anello non commutativo (ma non un corpo).

Osservazione 1.5 Se \mathbb{K} è un campo, possiamo considerare l'insieme $\mathbb{K}[t]$ dei polinomi a coefficienti in \mathbb{K} (e in una variabile). La definizione è la stessa di quella per i polinomi a coefficienti reali vista nell'Esempio 1.9; infatti per farla funzionare ci basta saper effettuare somme e moltiplicazioni — cioè ci basta essere in un campo qualunque.

In questo libro ci serviranno in maniera particolare i polinomi a coefficienti reali e i polinomi a coefficienti complessi; ma anche i polinomi a coefficienti razionali (o interi) hanno una teoria ricca e interessante.

L'*Algebra* è lo studio di gruppi, anelli, campi e altri oggetti simili. In questo libro ci interesseremo soltanto all'*Algebra Lineare*, quella parte (relativamente limitata) dell'*Algebra* che tratta dei cosiddetti spazi vettoriali, oggetti che introdurremo più oltre e che hanno una stretta relazione con i campi.

1.4 Prerequisiti

Benché la matematica contemporanea sia sensibilmente diversa da quella contenuta nei programmi delle scuole superiori, diversi degli argomenti che hai studiato sono utili anche per la comprensione di questo libro, e devi averli ben presenti. Per facilitare l'eventuale ripasso, questo paragrafo contiene l'elenco di tutto ciò che è necessario tu sappia prima di proseguire con la lettura.

(A) *Algebra*. Prima di tutto devi conoscere e saper utilizzare i numeri: naturali, interi, razionali e reali. Quindi nessuna esitazione è ammessa nel fare i conti con le frazioni, e nel saper riconoscere un numero decimale⁷ periodico. Non è necessario che tu conosca una definizione formalmente corretta dei numeri reali, ma devi saperli identificare come numeri decimali (periodici e no), e ricordarti che possono essere messi in corrispondenza biunivoca con i punti di una retta. Conviene anche che tu abbia presente i concetti di numero primo; di divisione col resto fra numeri interi; di massimo comun divisore e di minimo comune multiplo di numeri interi, coi relativi metodi di calcolo (l'algoritmo di Euclide).

Sicuramente devi avere una buona dimestichezza col calcolo algebrico elementare (utilizzo di lettere per indicare costanti e variabili, operazioni con lettere e numeri, prodotti notevoli, operazioni con potenze anche non intere, eccetera). In particolare, devi avere ben presenti le formule

$$(x+y)^2 = x^2 + 2xy + y^2, \quad (x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3, \quad (x+y)(x-y) = x^2 - y^2.$$

Devi saper sommare e moltiplicare polinomi, e devi conoscere un qualche metodo per effettuare la divisione con resto fra polinomi. Mi raccomando, tieni ben distinto il concetto di polinomio (che, come abbiamo visto nell'Esempio 1.9, è una funzione) da quello di equazione (che invece è un problema che si vuole risolvere; se ci pensi bene, trovare le soluzioni di un'equazione del tipo $p(x) = 0$, dove $p(x)$ è un polinomio, equivale a trovare l'immagine inversa di $\{0\}$ tramite la funzione p).

Devi saper risolvere le equazioni di primo grado e di secondo grado. Una certa dimestichezza con i sistemi lineari potrà essere utile ma non è strettamente necessaria: verranno trattati ampiamente in questo libro. Ti deve essere chiaro che le proporzioni sono soltanto un modo diverso per scrivere frazioni, e si risolvono come equazioni di

⁷ In questo libro utilizzeremo il punto decimale, e non la virgola decimale. Per esempio, $1/2$ sarà scritto come 0.5 e non come 0,5.

primo grado (o secondo grado, se l'obiettivo è trovare il medio proporzionale fra due numeri). La formula risolutiva dell'equazione di secondo grado $ax^2 + bx + c = 0$ è

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

La quantità $\Delta = b^2 - 4ac$ è detta *discriminante* dell'equazione. Se $\Delta > 0$ l'equazione ha due soluzioni (chiamate anche radici) distinte; se $\Delta = 0$ l'equazione ha un'unica soluzione (talvolta si dice che ha due soluzioni coincidenti, o che ha "una soluzione di molteplicità due"); se $\Delta < 0$ l'equazione non ha soluzioni (o meglio, non ha soluzioni reali; vedremo che ha due soluzioni complesse).

(B) *Geometria euclidea*. La cosa principale che devi aver assorbito dalla geometria euclidea è una certa abitudine a ragionare in termini di assiomi, teoremi, dimostrazioni, ipotesi e tesi; la capacità di riconoscere e condurre ragionamenti logicamente corretti è probabilmente la cosa più importante (e utile per la vita di tutti i giorni) che puoi apprendere dalla matematica.

Ci serviranno anche alcuni concetti generali e alcuni teoremi specifici della geometria del piano. Come concetti: parallelismo e perpendicolarità fra rette e fra segmenti; definizione di triangoli isosceli, equilateri, rettangoli e di parallelogrammi, rettangoli e quadrati; congruenza di figure (possibilmente collegata ai concetti di moto rigido, rotazione, traslazione e simmetria rispetto a punti e a rette); similitudine fra figure (possibilmente collegata al concetto di omotetia); angoli, lunghezze e aree (comprese le formule per il calcolo dell'area di triangoli, rettangoli e parallelogrammi); proporzionalità fra grandezze; segmenti orientati e vettori (rappresentanti grandezze dotate di lunghezza, direzione e verso).

Devi conoscere i criteri di congruenza dei triangoli, e i teoremi enunciati qui di seguito:

Teorema 1.1 *La somma degli angoli interni di un triangolo è 180° .*

Teorema 1.2 (Pitagora) *L'area del quadrato costruito sull'ipotenusa di un triangolo rettangolo è uguale alla somma delle aree dei quadrati costruiti sui cateti.*

Teorema 1.3 (Talete) *Tre rette parallele tagliano rispettivamente su due rette trasversali coppie di segmenti di lunghezza proporzionale.*

Teorema 1.4 *Un quadrilatero è un parallelogramma se e soltanto se ha due lati opposti paralleli e congruenti.*

Ci servirà invece molto poco della geometria euclidea dello spazio. Alcuni concetti (parallelismo, congruenza, moti rigidi) sono identici a quelli visti nel piano; sarà utile avere un'idea della forma dei parallelepipedi e delle sfere, e in un caso potrà essere d'aiuto conoscere come si calcola il volume di un parallelepipedo.

(C) *Geometria analitica*. Di nuovo, la cosa fondamentale è una certa dimestichezza con i concetti di base: assi cartesiani, coordinate cartesiane (ovvero come identificare ciascun punto del piano con una coppia ordinata di numeri reali, e ciascun punto

dello spazio con una terna ordinata di numeri reali), grafici di funzioni (alla funzione f a valori reali viene associato il sottoinsieme del piano composto dai punti di coordinate $(x, f(x))$, dove x varia nel dominio della funzione f).

È necessario sapere che il grafico di polinomi di primo grado (cioè di funzioni della forma $f(x) = ax + b$ che forse sei più abituato a scrivere come $y = ax + b$) è sempre una retta, e viceversa che ogni retta (tranne quelle parallele all'asse delle ordinate) è il grafico di un polinomio di primo grado. Può essere utile saper scrivere l'equazione della retta passante per due punti dati, e del segmento che li collega. Devi essere in grado di calcolare, usando le coordinate cartesiane, la distanza (euclidea) fra due punti nel piano e nello spazio.

Devi sapere che relazione c'è fra il risolvere un'equazione della forma $f(x) = 0$ e l'intersezione fra il grafico di f e l'asse delle ascisse. Devi sapere che l'insieme delle soluzioni di equazioni del tipo $f(x, y) = 0$ si può rappresentare come un sottoinsieme del piano cartesiano; in particolare, devi ricordarti che le rette del piano sono soluzione di equazioni del tipo $ax + by + c = 0$. Sarebbe utile anche sapere che i piani dello spazio si possono rappresentare come soluzione di equazioni del tipo $ax + by + cz + d = 0$.

Devi saper tracciare il grafico di polinomi di secondo grado, e saper riconoscere una parabola. Sarebbe anche utile conoscere l'equazione di iperboli ed ellissi.

(D) *Trigonometria*. Ti sarai forse chiesto perché non si è parlato di circonferenze e cerchi quando abbiamo trattato la geometria euclidea; ne parliamo qui. Devi conoscere le formule per la lunghezza di una circonferenza e per l'area del cerchio; devi essere abituato a considerare π un numero (reale, irrazionale, decimale non periodico) come tutti gli altri, ricordandoti che π non è uguale a 3.14, ma è circa 3.14.

Un *radiante* è quell'angolo che in una circonferenza di raggio unitario sottende un arco di lunghezza unitaria. Quindi per misurare l'ampiezza in radianti di un angolo basta misurare la lunghezza dell'arco sotteso: un angolo giro corrisponde a 2π radianti, un angolo piatto a π radianti, un angolo retto a $\pi/2$ radianti, e così via. In questo libro misureremo gli angoli quali esclusivamente in radianti, e quasi mai in gradi.

Devi conoscere definizioni e grafici delle funzioni trigonometriche (seno, coseno, tangente) e delle loro inverse (arcoseno, arcocoseno e arcotangente). Per il resto è sufficiente che ricordi la formula fondamentale⁸

$$\sin^2 x + \cos^2 x = 1;$$

le formule di periodicità

$$\begin{aligned} \sin(x + 2\pi) &= \sin x, & \sin(x + \pi) &= -\sin x, & \sin(x + \pi/2) &= \cos x, \\ \cos(x + 2\pi) &= \cos x, & \cos(x + \pi) &= -\cos x, & \cos(x + \pi/2) &= -\sin x, \\ \sin(-x) &= -\sin x, & \cos(-x) &= \cos x; \end{aligned}$$

e le formule di addizione

$$\sin(x + y) = \sin x \cos y + \cos x \sin y, \quad \cos(x + y) = \cos x \cos y - \sin x \sin y,$$

⁸ In questo libro la funzione seno sarà indicata con "sin", e non con "sen".

da cui seguono le formule di duplicazione

$$\sin(2x) = 2 \sin x \cos x, \quad \cos(2x) = \cos^2 x - \sin^2 x.$$

Da queste si ricavano tutte le altre formule trigonometriche, ricordandosi ovviamente che $\tan x = \sin x / \cos x$.

(E) *Analisi*. Sotto questa denominazione raccolgo sia dei prerequisiti veri e propri sia alcuni Teoremi che forse già conosci ma che sicuramente studierai nei corsi di Analisi, e che saranno utili in alcuni punti specifici di questo libro.

I prerequisiti sono pochissimi: sapere cosa vuol dire a^x quando a è un numero reale positivo e x un numero reale qualunque, e conoscere definizione e grafico delle funzioni esponenziale e logaritmo. Utilizzeremo soltanto i logaritmi naturali (indicati con "log"), cioè quelli in base e , dove e è il numero di Nepero che vale circa 2.71.

Infine, può essere utile per capire meglio alcuni esempi saper calcolare la derivata di polinomi e di funzioni razionali, e sapere che la derivata della somma è uguale alla somma delle derivate (ma la derivata del prodotto *non* è uguale al prodotto delle derivate...). A questo proposito, nei Complementi al Capitolo 16 avremo bisogno di sapere che se la derivata di un polinomio è positiva (o negativa) in un punto allora il polinomio è crescente (o decrescente) nei dintorni di quel punto. Se non lo conoscevi già, questo risultato ti sarà abbondantemente illustrato nei corsi di Analisi, come pure i seguenti tre Teoremi:

Teorema 1.5 (dei valori intermedi) *Sia $f: [a, b] \rightarrow \mathbb{R}$ una funzione continua definita sull'intervallo chiuso $[a, b] \subset \mathbb{R}$ e a valori reali. Supponiamo che $f(a) < 0$ e che $f(b) > 0$. Allora esiste almeno un punto x_0 compreso fra a e b tale che $f(x_0) = 0$.*

Teorema 1.6 (Weierstrass) *Ogni successione limitata di numeri reali ammette una sottosuccessione convergente.*

Teorema 1.7 *Sia $f: [a, b] \rightarrow \mathbb{R}$ una funzione continua definita sull'intervallo $[a, b]$ e a valori reali. Allora esiste un punto $x_0 \in [a, b]$ in cui la funzione f assume il suo valore massimo.*

Esercizi

1.1 Consideriamo gli insiemi $A = \{1, 2, 4, 5, 6\}$, $B = \{2, 4, 6, 8\}$ e $C = \{3, 6, 7\}$. Quali sono gli elementi di $A \cap B$? E di $A \cup C$? E di $B \setminus C$? E di $C \setminus B$? E di $A \times C$? E di $(A \cup B) \cap C$?

1.2 Dimostra che le operazioni di intersezione e unione fra insiemi sono associative e commutative.

1.3 Dimostra che

$$\begin{aligned} (A \cup B) \cap C &= (A \cap C) \cup (B \cap C), & (A \cap B) \cup C &= (A \cup C) \cap (B \cup C), \\ A \setminus (B \cup C) &= (A \setminus B) \cap (A \setminus C), & A \setminus (B \cap C) &= (A \setminus B) \cup (A \setminus C), \\ A \times (B \cup C) &= (A \times B) \cup (A \times C), & A \times (B \cap C) &= (A \times B) \cap (A \times C), \end{aligned}$$

quali che siano gli insiemi A , B e C .

1.4 Siano $f: A \rightarrow B$, $g: B \rightarrow C$ e $h: C \rightarrow D$ tre funzioni, con dominio e codominio scelti in modo da poter considerare le composizioni $g \circ f$, $h \circ (g \circ f)$, $h \circ g$ e $(h \circ g) \circ f$. Dimostra che $h \circ (g \circ f): A \rightarrow D$ e $(h \circ g) \circ f: A \rightarrow D$ sono la stessa funzione (si dice che la composizione è associativa).

1.5 Siano $f: A \rightarrow B$ e $g: B' \rightarrow C$ due funzioni tali che l'immagine $\text{Im } f$ di f sia contenuta nel dominio B' di g . Convinciti che la formula (1.2) definisce ancora una funzione $g \circ f: A \rightarrow C$.

1.6 Se $f: A \rightarrow B$ è una funzione bigettiva, e $f^{-1}: B \rightarrow A$ la sua inversa, dimostra che $f \circ f^{-1} = \text{id}_B$ e $f^{-1} \circ f = \text{id}_A$, dove id_A e id_B sono le funzioni identità di A e B , definite come nell'Esempio 1.1.

1.7 Dimostra che una funzione $f: A \rightarrow B$ è bigettiva se e solo se esistono due funzioni $g_1, g_2: B \rightarrow A$ tali che $f \circ g_1 = \text{id}_B$ e $g_2 \circ f = \text{id}_A$, e in tal caso $g_1 = g_2$.

1.8 Sia $f: A \rightarrow B$ una funzione, e $A_1, A_2 \subseteq A$ due sottoinsiemi di A . Dimostra che si ha $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$ ma che $f(A_1 \cap A_2) \subseteq f(A_1) \cap f(A_2)$. Trova un esempio in cui $f(A_1 \cap A_2) \neq f(A_1) \cap f(A_2)$.

1.9 Sia $f: A \rightarrow B$ una funzione *iniettiva*, e $A_1, A_2 \subseteq A$ due sottoinsiemi di A . Dimostra che $f(A_1 \cap A_2) = f(A_1) \cap f(A_2)$.

1.10 Sia $f: A \rightarrow B$ una funzione, e $A_1, A_2 \subseteq A$ due sottoinsiemi di A . Dimostra che $f(A_1) \setminus f(A_2) \subseteq f(A_1 \setminus A_2)$. Trova un esempio in cui $f(A_1) \setminus f(A_2) \neq f(A_1 \setminus A_2)$.

1.11 Sia $f: A \rightarrow B$ una funzione, e $B_1, B_2 \subseteq B$ due sottoinsiemi di B . Dimostra che $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$, $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$ e anche che $f^{-1}(B_1 \setminus B_2) = f^{-1}(B_1) \setminus f^{-1}(B_2)$.

1.12 Sia $f: A \rightarrow B$ una funzione, e $A_1 \subseteq A$ un sottoinsieme di A . Dimostra che $A_1 \subseteq f^{-1}(f(A_1))$, e trova un esempio in cui non vale l'uguaglianza.

1.13 Sia $f: A \rightarrow B$ una funzione, e $B_1 \subseteq B$ un sottoinsieme di B . Dimostra che $f(f^{-1}(B_1)) \subseteq B_1$, e trova un esempio in cui non vale l'uguaglianza.

1.14 Sia $p: \mathbb{R} \rightarrow \mathbb{R}$ un polinomio di secondo grado, $p(t) = at^2 + bt + c$. Per quali valori di a , b e c il polinomio è iniettivo? E per quali valori è surgettivo?

1.15 Siano $f: A \rightarrow B$ e $g: B \rightarrow C$ due funzioni tali che $g \circ f$ sia iniettiva. Dimostra che f è iniettiva, e trova un esempio in cui $g \circ f$ è iniettiva ma g non lo è.

1.16 Siano $f: A \rightarrow B$ e $g: B \rightarrow C$ due funzioni tali che $g \circ f$ sia surgettiva. Dimostra che g è surgettiva, e trova un esempio in cui $g \circ f$ è surgettiva ma f non lo è.

1.17 Cosa devi fare per dimostrare che l'affermazione "per ogni giorno della settimana esiste un minuto in cui Michele dorme" è vera? E per dimostrare che l'affermazione "esiste un giorno della settimana in cui Michele studia per ogni minuto della giornata" è falsa? E per verificare se la frase "se Luigi si sveglia presto la mattina allora ha fame prima di mezzogiorno" è vera o falsa?

1.18 Se $p \in \mathbb{R}[t]$ è un polinomio, indicheremo con $\deg p$ il suo grado. Per convenzione diremo che il polinomio costante 0 ha grado $-\infty$. Dimostra che

$$\deg(p_1 p_2) = (\deg p_1) + (\deg p_2) \quad \text{e} \quad \deg(p_1 + p_2) \leq \max\{\deg p_1, \deg p_2\}$$

per ogni coppia di polinomi $p_1, p_2 \in \mathbb{R}[t]$, dove s'intende che $n + (-\infty) = -\infty$ e $\max\{n, -\infty\} = n$ per ogni $n \in \mathbb{N}$. Infine, trova due polinomi p_1, p_2 tali che si abbia $\deg(p_1 + p_2) = -\infty < \max\{\deg p_1, \deg p_2\}$.

1.19 Di che proprietà godono le usuali somma e moltiplicazione dell'insieme $\mathbb{R}[t]$ dei polinomi a coefficienti reali in una variabile?

1.20 Sia A un insieme, e indichiamo con $\mathcal{P}(A)$ la famiglia di tutti i sottoinsiemi di A . Se $A_1, A_2 \in \mathcal{P}(A)$ poniamo $A_1 \oplus A_2 = (A_1 \setminus A_2) \cup (A_2 \setminus A_1)$ e $A_1 \otimes A_2 = A_1 \cap A_2$. Dimostra che $\mathcal{P}(A)$ con queste operazioni è un anello commutativo ma non un campo, con inoltre la proprietà che $A_1 \otimes A_1 = A_1$ per ogni $A_1 \in \mathcal{P}(A)$. (Suggerimento: nota che $A_1 \oplus A_2 = (A_1 \cup A_2) \setminus (A_1 \cap A_2)$, per cui $A_1 \oplus A_1 = \emptyset$.)

1.21 Sia $n > 0$ un numero naturale e poniamo $\mathbb{Z}_n = \{0, 1, \dots, n-1\}$. Se $a, b \in \mathbb{Z}_n$, definiamo $a \oplus b$ come il resto della divisione di $a + b$ per n . Dimostra che \mathbb{Z}_n con la "somma" \oplus è un gruppo commutativo. (Suggerimento: se $x \in \mathbb{N}$, indichiamo con $[x]$ il resto della divisione di x per n . Dimostra prima di tutto che $[x + y] = [x] \oplus [y]$ per ogni $x, y \in \mathbb{N}$.)

1.22 Sia $\mathbb{Q}(\sqrt{3}) = \{x = a + b\sqrt{3} \in \mathbb{R} \mid a, b \in \mathbb{Q}\}$ l'insieme dei numeri reali della forma $a + b\sqrt{3}$, con $a, b \in \mathbb{Q}$. Dimostra che $\mathbb{Q}(\sqrt{3})$, con la somma e il prodotto usuali, è un campo.

1.23 Sia $(G, +)$ un gruppo commutativo. Dimostra che in G esiste un unico elemento neutro, e che ogni $g \in G$ ammette un unico opposto.

COMPLEMENTI

1C.1 L'alfabeto greco

In matematica l'uso delle lettere greche è praticamente continuo; quindi è consigliabile saperle scrivere e riconoscere. Per aiutarti, qui di seguito troverai l'elenco alfabetico delle lettere greche maiuscole e minuscole, con relativo nome.

A	α	alfa	I	ι	iota	R	ρ	rho
B	β	beta	K	κ	kappa	Σ	σ	sigma
Γ	γ	gamma	Λ	λ	lambda	T	τ	tau
Δ	δ	delta	M	μ	mu	Υ	υ	upsilon
E	ϵ, ε	epsilon	N	ν	nu	Φ	ϕ, φ	phi
Z	ζ	zeta	Ξ	ξ	xi	X	χ	chi
H	η	eta	O	\omicron	omicron	Ψ	ψ	psi
Θ	θ, ϑ	theta	Π	π	pi	Ω	ω	omega

Vettori geometrici

Storicamente, l'algebra lineare è nata in seguito all'osservazione che due argomenti apparentemente diversi — i vettori applicati e i sistemi lineari — in realtà potevano essere trattati con strumenti sorprendentemente simili; erano entrambi applicazioni della teoria astratta degli spazi vettoriali. In questo capitolo studieremo i vettori applicati nel piano e nello spazio col duplice obiettivo di evidenziare gli aspetti che hanno portato allo sviluppo della teoria generale (teoria che inizieremo ad affrontare nel Capitolo 4) e di fornire contemporaneamente esempi concreti che possano facilitare la comprensione degli argomenti più astratti. Come vedrai, l'idea di fondo è la vecchia intuizione di Cartesio, rivisitata e ripresentata con uno stile più consono alla matematica moderna: l'introduzione di un sistema di coordinate permette di sviluppare tecniche algebriche per risolvere efficacemente problemi geometrici (e, talvolta, viceversa).

2.1 Vettori applicati

Indichiamo con \mathcal{A}^1 (rispettivamente, $\mathcal{A}^2, \mathcal{A}^3$) la retta (rispettivamente, il piano, lo spazio) euclidea usuale¹. Nel seguito parleremo principalmente del piano, ma tutto quanto detto si applicherà pari pari allo spazio, e sarà spesso banale per la retta.

Come già accennato nel Paragrafo 1.3, fissando sulla retta \mathcal{A}^1 un punto O e un'unità di misura, cioè un segmento \overline{OA} , otteniamo un'applicazione bigettiva fra i punti della retta e i numeri reali; in un certo senso, possiamo identificare la retta \mathcal{A}^1 e l'insieme \mathbb{R} . In particolare, questo introduce su \mathcal{A}^1 una somma e un prodotto che lo rendono un campo.

Vogliamo vedere se e quanto sia possibile ripetere questa costruzione nel caso del piano e dello spazio. Procediamo un passo per volta, e cominciamo col fissare un punto $O \in \mathcal{A}^2$ (oppure $O \in \mathcal{A}^3$). Adesso ogni punto del piano (o dello spazio) può

¹ Non è questo il luogo adatto per dare una definizione formalmente corretta di piano (retta, spazio) euclideo; per noi è sufficiente sapere che sono insiemi in cui valgono tutte le proprietà della geometria euclidea che hai studiato alle superiori.

venire considerato non soltanto per conto suo ma anche in relazione al punto fissato O ; e questo ci porta al concetto di vettore applicato.

Definizione 2.1 Un vettore applicato in O è un segmento orientato con primo estremo il punto $O \in \mathcal{A}^2$ (rispettivamente, $O \in \mathcal{A}^3$) e secondo estremo un altro punto $A \in \mathcal{A}^2$ (rispettivamente, $A \in \mathcal{A}^3$). Questo vettore sarà disegnato come una freccia che parte da O e giunge ad A , e indicato con \vec{OA} (Figura 2.1).

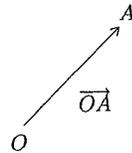


Figura 2.1 Il vettore \vec{OA} .

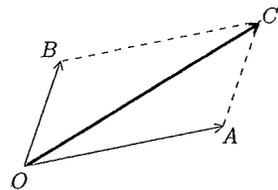
Indicheremo con \mathcal{V}_O^2 (o con \mathcal{V}_O^3 nello spazio) l'insieme dei vettori applicati in O ; il punto O è l'origine di \mathcal{V}_O^2 (o di \mathcal{V}_O^3). Nota che possiamo definire una funzione bigettiva $\Phi_O: \mathcal{A}^2 \rightarrow \mathcal{V}_O^2$ (e una analoga da \mathcal{A}^3 a \mathcal{V}_O^3) ponendo

$$\forall A \in \mathcal{A}^2 \quad \Phi_O(A) = \vec{OA};$$

al punto A viene associato il vettore applicato in O che termina in A . In particolare, all'origine O viene associato il vettore \vec{OO} , chiamato (per motivi ovvi) il vettore nullo.

Osservazione 2.1 A volte, quando l'origine sarà fissata, identificheremo i punti del piano con i vettori applicati nell'origine: semplicemente, staremo usando Φ_O senza dirlo esplicitamente.

In Fisica, i vettori servono per vari scopi; a indicare la velocità istantanea di un corpo, per esempio, o le forze che agiscono su di esso. Ora, quando applichiamo due forze diverse contemporaneamente su uno stesso oggetto questi non le sente come distinte e si muove in una direzione intermedia, come se si applicasse prima una forza e poi, finito l'effetto, l'altra. E il risultato è rappresentabile da una terza, nuova, forza, ottenuta applicando una regola molto precisa nota come regola del parallelogramma. Questo ci suggerisce di introdurre una somma nel nostro insieme \mathcal{V}_O^2 .



$$\vec{OC} = \vec{OA} + \vec{OB}$$

Figura 2.2 La somma di due vettori non allineati.

N la somma di vettori N

Definizione 2.2 Presi due vettori applicati \vec{OA} e \vec{OB} , la loro somma $\vec{OA} + \vec{OB}$ è il vettore applicato \vec{OC} , dove C è il quarto vertice del parallelogramma individuato da O, A e B (Figura 2.2). In altri termini, il punto C si trova come secondo estremo del vettore applicato in B parallelo, congruente e con lo stesso verso del vettore \vec{OA} . Quest'ultima caratterizzazione funziona bene anche quando i tre punti O, A e B sono allineati (Figura 2.3).

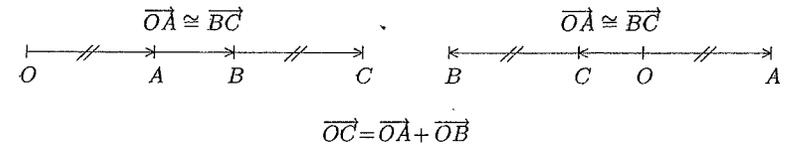


Figura 2.3 La somma di due vettori allineati.

Siccome l'abbiamo chiamata somma, si spera che questa operazione goda almeno di alcune delle proprietà elencate nel capitolo precedente. Infatti

Proposizione 2.1 \mathcal{V}_O^2 con la somma descritta nella Definizione 2.2 è un gruppo commutativo.

Dimostrazione. Dobbiamo verificare le proprietà (i)–(iv) del Paragrafo 1.3; cominciamo con quelle più semplici. L'elemento neutro è il vettore nullo: infatti, dalla definizione segue subito (controlla) che

$$\forall \vec{OA} \in \mathcal{V}_O^2 \quad \vec{OA} + \vec{OO} = \vec{OA} = \vec{OO} + \vec{OA}.$$

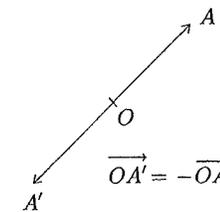


Figura 2.4 L'opposto di un vettore.

Anche l'opposto si trova facilmente: l'opposto del vettore \vec{OA} è proprio (esercizio) il vettore $\vec{OA'}$, dove A' è il simmetrico di A rispetto a O sulla retta passante per O e A (Figura 2.4).

La proprietà commutativa è pure di facile dimostrazione. Se O, A e B non sono allineati, $\vec{OA} + \vec{OB} = \vec{OB} + \vec{OA}$ segue dal fatto che i punti O, A e B e i punti O, B e A identificano lo stesso parallelogramma. Se invece i tre punti in questione sono allineati, la commutatività segue dalle proprietà della congruenza fra segmenti orientati (esercizio).

Rimane da dimostrare la proprietà associativa. Consideriamo allora tre vettori applicati \vec{OA}_1, \vec{OA}_2 e \vec{OA}_3 ; supponiamo per il momento che fra i punti O, A_1, A_2 e A_3 non ce ne siano tre allineati (Figura 2.5).

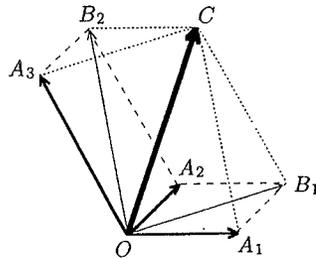


Figura 2.5 La proprietà associativa.

Poniamo $\vec{OB}_1 = \vec{OA}_1 + \vec{OA}_2, \vec{OB}_2 = \vec{OA}_2 + \vec{OA}_3$ e

$$\vec{OC} = \vec{OB}_1 + \vec{OA}_3 = (\vec{OA}_1 + \vec{OA}_2) + \vec{OA}_3.$$

Dobbiamo dimostrare che

$$\vec{OA}_1 + (\vec{OA}_2 + \vec{OA}_3) = \vec{OA}_1 + \vec{OB}_2 = \vec{OC}.$$

Prima di tutto, siccome OB_1CA_3 è un parallelogramma, vediamo subito che il segmento \vec{B}_1C è parallelo e congruente al segmento \vec{OA}_3 . Ma anche $OA_3B_2A_2$ è un parallelogramma; quindi il segmento \vec{OA}_3 è parallelo e congruente al segmento $\vec{A_2B_2}$. Dunque \vec{B}_1C è parallelo e congruente a $\vec{A_2B_2}$; ne segue (Teorema 1.4) che $A_2B_1CB_2$ (avendo i due lati opposti paralleli e congruenti) è un parallelogramma, per cui $\vec{B_2C}$ è parallelo e congruente ad $\vec{A_2B_1}$.

Ora notiamo che anche $OA_1B_1A_2$ è un parallelogramma, per cui $\vec{A_2B_1}$ è parallelo e congruente a $\vec{OA_1}$. Quindi $\vec{B_2C}$ e $\vec{OA_1}$ sono paralleli e congruenti; questo vuol dire (ancora grazie al Teorema 1.4) che OA_1CB_2 è un parallelogramma, ovvero che $\vec{OC} = \vec{OA_1} + \vec{OB_2}$, come volevamo dimostrare.

Rimane da controllare la proprietà associativa nel caso in cui tre dei punti O, A_1, A_2 e A_3 siano allineati; l'idea è praticamente la stessa, ed è un esercizio per te (Esercizio 2.1). □

Osservazione 2.2 Abbiamo evidenziato il risultato precedente isolandolo dal resto del testo e chiamandolo "Proposizione". Questo è il metodo tipico della matematica per identificare i fatti più importanti, con una vera e propria graduatoria in proposito. Un "Teorema" è un fatto veramente significativo; una "Proposizione" invece è sempre interessante, ma meno di un Teorema. Un "Lemma" è un risultato che serve solo per la dimostrazione di una Proposizione o un Teorema. Un "Corollario" invece è la conseguenza (più o meno) immediata di un Teorema o di una Proposizione; una specie di caso particolare. Infine, il simbolo □ indica la fine di una dimostrazione.

Osservazione 2.3 Forse ti sarai chiesto perché sia necessario dimostrare l'associatività (o altre proprietà analoghe) con un lungo ragionamento invece di limitarsi a fare un disegno e verificare che è vero. Il punto è che i disegni sono imprecisi, e possono ingannare: funzionano bene come guida, ma se non sorretti da argomentazioni accurate possono anche portare a conclusioni false. Un esempio di questo genere di errore è contenuto nei Complementi a questo capitolo, dove viene "dimostrato" che tutti i triangoli sono isosceli.

La somma non è l'unica operazione che possiamo fare con i vettori applicati; possiamo anche moltiplicarli per un numero reale².

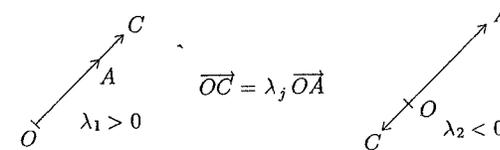


Figura 2.6 Il prodotto di un vettore per un numero reale.

Definizione 2.3 Se $\lambda \in \mathbb{R}$ e $\vec{OA} \in \mathcal{V}_0^2$, allora il prodotto di λ per \vec{OA} è il vettore $\vec{OC} = \lambda \vec{OA}$, dove C è il punto sulla retta passante per O e A tale che il rapporto fra la lunghezza del segmento \vec{OC} e quella del segmento \vec{OA} sia esattamente $|\lambda|$. Inoltre, C è sulla semiretta OA se λ è positivo, e sulla semiretta opposta se λ è negativo. In parole povere, $\lambda \vec{OA}$ ha la stessa direzione di \vec{OA} , lunghezza moltiplicata per $|\lambda|$ e verso uguale od opposto a seconda del segno di λ (Figura 2.6).

Chiaramente, $0 \vec{OA} = \vec{OO}$, $2 \vec{OA} = \vec{OA} + \vec{OA}$, $(-3) \vec{OA} = (-\vec{OA}) + (-\vec{OA}) + (-\vec{OA})$, e così via, qualunque sia il vettore \vec{OA} .

² Forse ti chiederai perché non tentiamo di definire su \mathcal{V}_0^2 o \mathcal{V}_0^3 un prodotto in modo da ottenere un campo come abbiamo fatto su \mathcal{A}^1 . Il fatto è che, mentre su \mathcal{V}_0^2 (come vedremo nel Capitolo 11 studiando i numeri complessi) questo si può fare, su \mathcal{V}_0^3 un tale prodotto non esiste. La dimostrazione di questo risultato a prima vista sorprendente è, come puoi immaginare, piuttosto difficile.

Le proprietà algebriche di questo prodotto per un numero reale sono contenute nella seguente

Proposizione 2.2 Siano λ e μ due numeri reali, e \vec{OA}, \vec{OB} due vettori applicati. Allora:

- (i) $\lambda(\vec{OA} + \vec{OB}) = \lambda\vec{OA} + \lambda\vec{OB}$;
- (ii) $(\lambda + \mu)\vec{OA} = \lambda\vec{OA} + \mu\vec{OA}$;
- (iii) $(\lambda\mu)\vec{OA} = \lambda(\mu\vec{OA})$;
- (iv) $1\vec{OA} = \vec{OA}$ e $0\vec{OA} = \vec{OO}$.

Dimostrazione. Di nuovo, le proprietà (ii)-(iv) sono ovvie (vero?); l'unica che richiede qualche parola è (i): Se O, A e B sono allineati, è di nuovo tutto banale (verificalo); supponiamo quindi che non lo siano, e cominciamo col prendere $\lambda > 0$, per cui ci troviamo nella situazione rappresentata nella Figura 2.7.

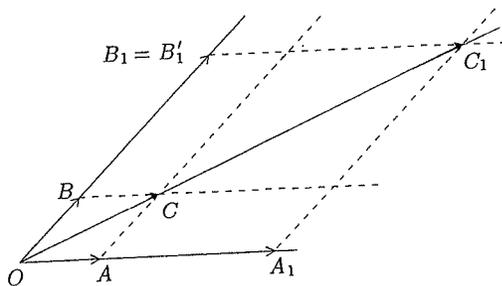


Figura 2.7 La proprietà distributiva.

Poniamo $\vec{OA}_1 = \lambda\vec{OA}$, $\vec{OB}_1 = \lambda\vec{OB}$, e $\vec{OC} = \vec{OA} + \vec{OB}$; dobbiamo dimostrare che

$$\lambda\vec{OC} = \vec{OA}_1 + \vec{OB}_1.$$

Consideriamo le rette passanti per A e A_1 parallele al segmento \vec{OB} ; queste intersecano la retta passante per O e C rispettivamente nei punti C e C_1 . Per il Teorema di Talete (Teorema 1.3), il rapporto fra le lunghezze dei segmenti \vec{OA}_1 e \vec{OA} (che è per definizione λ) è uguale al rapporto fra le lunghezze dei segmenti \vec{OC}_1 e \vec{OC} ; quindi il vettore \vec{OC}_1 è proprio $\lambda\vec{OC}$.

Tracciamo ora le rette passanti per C e C_1 parallele al segmento \vec{OA} . La retta per C interseca la retta passante per O e B nel punto B , in quanto $OACB$ è un parallelogramma; la retta per C_1 invece interseca la retta passante per O e B in un punto che chiameremo B_1 . Sempre il Teorema di Talete ci dice che il rapporto fra le lunghezze dei segmenti \vec{OC}_1 e \vec{OC} (che abbiamo visto essere λ) è uguale al rapporto fra le lunghezze dei segmenti \vec{OB}_1 e \vec{OB} ; quindi il vettore \vec{OB}_1 è proprio $\lambda\vec{OB}$,

cioè $B_1 = B_1$. Quindi $OA_1C_1B_1$ è un parallelogramma e dunque $\vec{OC}_1 = \vec{OA}_1 + \vec{OB}_1$, come volevamo dimostrare. \square

Il caso $\lambda < 0$ è analogo, e lo lasciamo come esercizio (Esercizio 2.2).

Come vedremo nel Paragrafo 4.1, il contenuto delle Proposizioni 2.1 e 2.2 si può riassumere dicendo che \mathcal{V}_O^2 è un esempio di spazio vettoriale su \mathbb{R} .

Vi è anche un altro modo di considerare le operazioni appena introdotte. Fissiamo un vettore $\vec{OA} \in \mathcal{V}_O^2$. Tramite la somma con \vec{OA} possiamo definire un'applicazione $\tau_{\vec{OA}}: \mathcal{A}^2 \rightarrow \mathcal{A}^2$ in questo modo: al punto $B \in \mathcal{A}^2$ associamo il punto $C = \tau_{\vec{OA}}(B) \in \mathcal{A}^2$ tale che $\vec{OC} = \vec{OA} + \vec{OB}$ (vedi Figura 2.8). In formule,

$$\tau_{\vec{OA}}(B) = \Phi_O^{-1}(\vec{OA} + \Phi_O(B)).$$

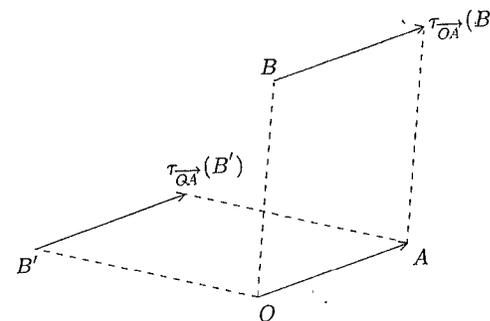


Figura 2.8 La somma come traslazione del piano.

Guardando la Figura 2.8, ti renderai conto che $\tau_{\vec{OA}}$ non è altro che una traslazione: la somma di un vettore corrisponde a una traslazione del piano.

Interpretiamo in modo analogo il prodotto per uno scalare. Fissato $\lambda \in \mathbb{R}$, possiamo definire un'applicazione $\sigma_\lambda: \mathcal{A}^2 \rightarrow \mathcal{A}^2$ associando al punto $A \in \mathcal{A}^2$ il punto $B = \sigma_\lambda(A) \in \mathcal{A}^2$ tale che $\vec{OB} = \lambda\vec{OA}$. In formule,

$$\sigma_\lambda(A) = \Phi_O^{-1}(\lambda\Phi_O(A)).$$

Questa volta σ_λ è una omotetia (o similitudine) di centro O e scala³ (o ragione) $|\lambda|$, seguita eventualmente da una simmetria rispetto a O se λ è negativo: il prodotto per un numero reale corrisponde a una omotetia del piano.

³ Questo è il motivo per cui in questo contesto i numeri reali spesso vengono detti *scalari*, e il prodotto un prodotto per scalari.

2.2 Coordinate

Prendiamo ora un singolo vettore $\vec{i} = \overrightarrow{OA_1} \in \mathcal{V}_O^2$ e consideriamo la retta r_1 passante per O e A_1 . Tutti i vettori di questa retta (cioè, tutti i vettori applicati in O il cui secondo estremo appartiene alla retta) sono della forma $t\vec{i} = t\overrightarrow{OA_1}$ per un appropriato $t \in \mathbb{R}$ — ovvero sono un multiplo di \vec{i} . In questo modo abbiamo definito una coordinata sulla retta r_1 : fissate un'origine O e un vettore unitario (il vettore \vec{i}) a ogni punto P della retta (ovvero a ogni vettore \overrightarrow{OP} di r) possiamo associare uno e un solo numero reale t tale che

$$\overrightarrow{OP} = t\vec{i};$$

il numero reale t è la *coordinata* di \overrightarrow{OP} rispetto a \vec{i} .

Ora, ci troviamo sul piano; quindi non tutti i vettori di \mathcal{V}_O^2 stanno sulla retta r . Prendiamone allora uno che non vi appartiene, $\vec{j} = \overrightarrow{OA_2}$. Tutti i vettori della retta r_2 per O e A_2 (o, come diremo, tutti i vettori *proporzionali* a \vec{j}) si scrivono come multipli di \vec{j} . In questo modo possiamo esprimere in termini di \vec{i} e \vec{j} tutti i vettori delle due rette r_1 ed r_2 . E gli altri vettori del piano? La risposta è contenuta nella

Proposizione 2.3 Siano $\vec{i} = \overrightarrow{OA_1}$ e $\vec{j} = \overrightarrow{OA_2}$ due vettori non proporzionali di \mathcal{V}_O^2 . Allora per ogni vettore $\overrightarrow{OP} \in \mathcal{V}_O^2$ esistono due numeri reali $x_1, x_2 \in \mathbb{R}$ tali che

$$\overrightarrow{OP} = x_1\vec{i} + x_2\vec{j}. \quad (2.1)$$

Inoltre, x_1 e x_2 sono unici, nel senso che se x'_1 e x'_2 sono due altri numeri reali tali che $\overrightarrow{OP} = x'_1\vec{i} + x'_2\vec{j}$, allora $x'_1 = x_1$ e $x'_2 = x_2$.

Dimostrazione. Cominciamo con l'esistenza (vedi Figura 2.9). Indichiamo con r_1 (rispettivamente, r_2) la retta passante per O e A_1 (rispettivamente, A_2). Tracciamo la parallela a r_2 passante per P ; questa interseca r_1 in un punto P_1 . Analogamente, tracciamo la parallela a r_1 passante per P ; interseca r_2 in un punto P_2 . Per costruzione, OP_1PP_2 è un parallelogramma; quindi

$$\overrightarrow{OP} = \overrightarrow{OP_1} + \overrightarrow{OP_2}.$$

D'altra parte, P_1 è sulla retta per O e A_1 ; quindi esiste un $x_1 \in \mathbb{R}$ tale che $\overrightarrow{OP_1} = x_1\vec{i}$. Analogamente troviamo un $x_2 \in \mathbb{R}$ tale che $\overrightarrow{OP_2} = x_2\vec{j}$, e l'esistenza è fatta.

Per l'unicità, da

$$x'_1\vec{i} + x'_2\vec{j} = \overrightarrow{OP} = x_1\vec{i} + x_2\vec{j}$$

otteniamo

$$(x'_1 - x_1)\vec{i} = (x_2 - x'_2)\vec{j}.$$

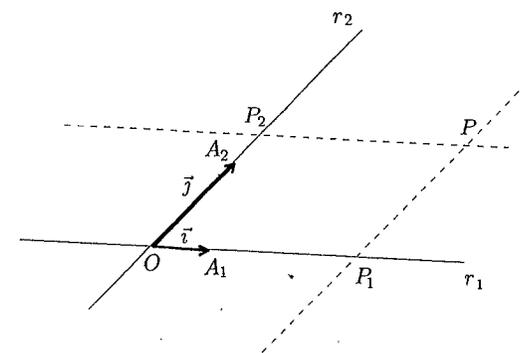


Figura 2.9 Coordinate in \mathcal{V}_O^2 .

Il vettore a primo membro è un multiplo di \vec{i} , per cui appartiene a r_1 . Il vettore a secondo membro, invece, appartiene a r_2 ; siccome sono uguali, devono stare nell'intersezione delle due rette, che è il solo vettore nullo \overrightarrow{OO} . Dunque

$$(x'_1 - x_1)\vec{i} = \overrightarrow{OO} = (x_2 - x'_2)\vec{j},$$

che può succedere se e solo se (Esercizio 2.4) $x'_1 - x_1 = 0$ e $x_2 - x'_2 = 0$, come volevamo. \square

Dunque una volta scelti due vettori non proporzionali \vec{i} e \vec{j} di \mathcal{V}_O^2 (o, come diremo, una volta fissata una base $\mathcal{B} = \{\vec{i}, \vec{j}\}$ di \mathcal{V}_O^2), a ogni vettore di \mathcal{V}_O^2 possiamo associare in modo unico una coppia di numeri reali, le sue *coordinate* rispetto alla base \mathcal{B} . In modo più formale, abbiamo definito un'applicazione $F_{\mathcal{B}}: \mathcal{V}_O^2 \rightarrow \mathbb{R}^2$ (dove $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ è l'insieme delle coppie ordinate di numeri reali) che associa a ogni vettore \overrightarrow{OP} la coppia

$$F_{\mathcal{B}}(\overrightarrow{OP}) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

dove x_1 e x_2 sono gli unici numeri reali che verificano (2.1). Nota che, per motivi che vedremo più avanti⁴, gli elementi di \mathbb{R}^2 verranno scritti per colonna, e non per riga.

ESEMPIO 2.1 Sia $\mathcal{B} = \{\vec{i}, \vec{j}\}$ una base di \mathcal{V}_O^2 , e prendiamo il vettore $\overrightarrow{OC} = 2\vec{i} - 4\vec{j}$. Le coordinate di \overrightarrow{OC} sono, per definizione, 2 e -4 ; in altre parole, $F_{\mathcal{B}}(\overrightarrow{OC}) = \begin{pmatrix} 2 \\ -4 \end{pmatrix}$.

Sia $\overrightarrow{OD} = -\vec{i} + 4\vec{j}$; le sue coordinate sono $\begin{pmatrix} -1 \\ 4 \end{pmatrix}$. Per trovare le coordinate di $\overrightarrow{OC} + \overrightarrow{OD}$

⁴ Essenzialmente, per poter scrivere prima una matrice e poi un vettore quando dovremo moltiplicarli righe per colonne.

calcoliamo:

$$\begin{aligned}\overrightarrow{OC} + \overrightarrow{OD} &= (2\vec{i} - 4\vec{j}) + (-\vec{i} + 4\vec{j}) = (2\vec{i} - \vec{i}) + (-4\vec{j} + 4\vec{j}) \\ &= (2-1)\vec{i} + (-4+4)\vec{j} = 1\vec{i} + 0\vec{j} = \vec{i}\end{aligned}$$

(dove abbiamo usato le Proposizioni 2.1 e 2.2), per cui

$$F_B(\overrightarrow{OC} + \overrightarrow{OD}) = \begin{vmatrix} 1 \\ 0 \end{vmatrix}.$$

→ Osservazione 2.4 La funzione F_B (o, se preferisci, le coordinate di un vettore) dipende dalla base scelta: cambiando base, le coordinate cambiano. Per esempio, siano $\vec{i}, \vec{j} \in \mathcal{V}_O^2$ due vettori non proporzionali, e consideriamo $\mathcal{B} = \{\vec{i}, \vec{j}\}$ e $\mathcal{B}' = \{2\vec{i}, \vec{j}\}$; siccome $2\vec{i}$ e \vec{j} continuano a essere non proporzionali, sia \mathcal{B} che \mathcal{B}' sono basi di \mathcal{V}_O^2 .

Ora prendiamo il vettore $\overrightarrow{OP} = 4\vec{i} - 3\vec{j}$. Le coordinate di \overrightarrow{OP} rispetto alla base \mathcal{B} sono chiaramente 4 e -3; invece, le sue coordinate rispetto alla base \mathcal{B}' sono 2 e -3 (perché?). Quindi

$$F_B(\overrightarrow{OP}) = \begin{vmatrix} 4 \\ -3 \end{vmatrix} \neq \begin{vmatrix} 2 \\ -3 \end{vmatrix} = F_{B'}(\overrightarrow{OP}).$$

Sulla retta, a ogni numero reale corrispondeva uno e un solo punto; la funzione coordinata era bigettiva. La stessa cosa accade qui:

Proposizione 2.4 Sia $\mathcal{B} = \{\vec{i}, \vec{j}\}$ una base di \mathcal{V}_O^2 . Allora l'applicazione $F_B: \mathcal{V}_O^2 \rightarrow \mathbb{R}^2$ è bigettiva.

Dimostrazione. Prima di tutto, è surgettiva; infatti, dati due qualunque numeri reali $x_1, x_2 \in \mathbb{R}$ abbiamo

$$\begin{vmatrix} x_1 \\ x_2 \end{vmatrix} = F_B(x_1 \vec{i} + x_2 \vec{j}) \in \text{Im } F_B.$$

Infine, è iniettiva; infatti

$$F_B(\overrightarrow{OP}) = \begin{vmatrix} y_1 \\ y_2 \end{vmatrix} = F_B(\overrightarrow{OQ})$$

implica che

$$\overrightarrow{OP} = y_1 \vec{i} + y_2 \vec{j} = \overrightarrow{OQ},$$

e ci siamo. \square

Dunque ci sono tanti vettori in \mathcal{V}_O^2 (e quindi — ricordi Φ_O ? — tanti punti in \mathcal{A}^2) quanti elementi in \mathbb{R}^2 ; come insieme, \mathbb{R}^2 ha la stessa struttura di \mathcal{V}_O^2 . Ora, in \mathcal{V}_O^2 abbiamo definito delle operazioni (grazie alle quali abbiamo potuto introdurre l'applicazione F_B); è naturale quindi cercare di scoprire come F_B si comporta rispetto a queste operazioni. Cominciamo col prodotto per scalari. Se $F_B(\overrightarrow{OP}) = \begin{vmatrix} x_1 \\ x_2 \end{vmatrix}$, vuol

dire che $\overrightarrow{OP} = x_1 \vec{i} + x_2 \vec{j}$; quindi $\lambda \overrightarrow{OP} = (\lambda x_1) \vec{i} + (\lambda x_2) \vec{j}$, per cui le coordinate di $\lambda \overrightarrow{OP}$ sono

$$F_B(\lambda \overrightarrow{OP}) = \begin{vmatrix} \lambda x_1 \\ \lambda x_2 \end{vmatrix}. \quad (2.2)$$

Vediamo invece la somma. Se $\overrightarrow{OP} = x_1 \vec{i} + x_2 \vec{j}$ e $\overrightarrow{OQ} = y_1 \vec{i} + y_2 \vec{j}$, allora (confronta con l'Esempio 2.1)

$$\overrightarrow{OP} + \overrightarrow{OQ} = (x_1 + y_1) \vec{i} + (x_2 + y_2) \vec{j},$$

per cui

$$F_B(\overrightarrow{OP} + \overrightarrow{OQ}) = \begin{vmatrix} x_1 + y_1 \\ x_2 + y_2 \end{vmatrix}. \quad (2.3)$$

A questo punto viene (o dovrebbe venire) naturale un'idea: possiamo definire su \mathbb{R}^2 una somma e un prodotto per uno scalare ponendo

$$\begin{vmatrix} x_1 \\ x_2 \end{vmatrix} + \begin{vmatrix} y_1 \\ y_2 \end{vmatrix} = \begin{vmatrix} x_1 + y_1 \\ x_2 + y_2 \end{vmatrix} \quad \text{e} \quad \lambda \begin{vmatrix} x_1 \\ x_2 \end{vmatrix} = \begin{vmatrix} \lambda x_1 \\ \lambda x_2 \end{vmatrix}.$$

È ora immediato (anche per l'associatività!) verificare che queste operazioni soddisfano le proprietà elencate nelle Proposizioni 2.1 e 2.2, dove l'elemento neutro e l'opposto ora sono dati da

$$O = \begin{vmatrix} 0 \\ 0 \end{vmatrix} \quad \text{e} \quad - \begin{vmatrix} x_1 \\ x_2 \end{vmatrix} = \begin{vmatrix} -x_1 \\ -x_2 \end{vmatrix}.$$

Inoltre, le (2.2) e (2.3) divengono

$$F_B(\overrightarrow{OP} + \overrightarrow{OQ}) = F_B(\overrightarrow{OP}) + F_B(\overrightarrow{OQ}), \quad F_B(\lambda \overrightarrow{OP}) = \lambda F_B(\overrightarrow{OP}). \quad (2.4)$$

In altri termini, F_B trasforma le operazioni di \mathcal{V}_O^2 nelle operazioni di \mathbb{R}^2 , e viceversa. Siccome F_B è anche bigettiva, da un certo punto di vista questo vuol dire che \mathcal{V}_O^2 ed \mathbb{R}^2 hanno la stessa struttura algebrica; diremo che sono *isomorfi*, e che F_B è un *isomorfismo*.

Riassumiamo cos'è successo. Siamo partiti dal piano euclideo \mathcal{A}^2 ; fissando un punto O , abbiamo trovato \mathcal{V}_O^2 , con la relativa struttura algebrica; scegliendo due vettori non proporzionali \vec{i} e \vec{j} siamo arrivati a \mathbb{R}^2 . Il tutto tramite applicazioni bigettive; quindi \mathbb{R}^2 è un *modello numerico* del piano euclideo \mathcal{A}^2 . Il vantaggio di questo procedimento è che (come si è visto per esempio nel caso della proprietà associativa) lavorare e fare i conti in \mathbb{R}^2 è molto più facile che in \mathcal{A}^2 ; possiamo fare somme e moltiplicazioni con la calcolatrice invece di dover usare riga e compasso. Inoltre — e questo è un punto fondamentale — il modello *non è unico*: possiamo scegliere l'origine O e i vettori \vec{i} e \vec{j} in modo tale da adattarli al problema che stiamo affrontando rendendo i conti il più semplici possibili. Per esempio, se il nostro problema riguarda una circonferenza è probabile che ci convenga scegliere come origine il centro della circonferenza, piuttosto che un punto a caso. Bisogna farsi furbi, e adattare il modello al problema che si vuole affrontare.

ESEMPIO 2.2 Vogliamo dimostrare il seguente teorema di geometria euclidea (vedi la Figura 2.10): “Sia ABC un triangolo. Indichiamo con H il punto medio del segmento \overline{AB} , e con K il punto medio del segmento \overline{AC} . Sia L il quarto vertice del parallelogramma individuato da A, H e K . Allora L è il punto medio del segmento \overline{BC} .” Ovviamente potremmo procedere con i soliti mezzi della geometria euclidea; vediamo invece come un’appropriata scelta di un’origine e di una base rendano questo risultato quasi banale.

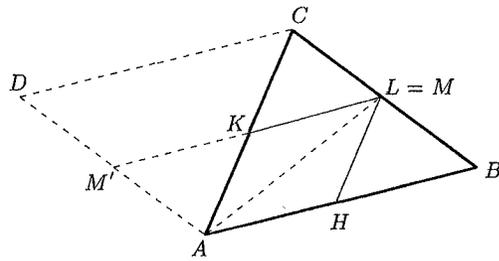


Figura 2.10 Un teorema sui triangoli.

Dovendo studiare un triangolo, è naturale scegliere come origine il punto A , e come base $B = \{\overline{AB}, \overline{AC}\}$. Rispetto a B , il punto A ha coordinate $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$, il punto B coordinate $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, e il punto C coordinate $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Il punto H , essendo il punto medio del segmento \overline{AB} , ha coordinate $\begin{bmatrix} 1/2 \\ 0 \end{bmatrix}$; analogamente, il punto K ha coordinate $\begin{bmatrix} 0 \\ 1/2 \end{bmatrix}$. Siccome, per costruzione, $\overline{AL} = \overline{AH} + \overline{AK}$, il punto L ha coordinate $\begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$. Vogliamo ora le coordinate del punto medio M del segmento \overline{BC} . Il vettore \overline{BC} è parallelo e congruente al vettore $\overline{AD} = \overline{AC} - \overline{AB}$, per cui il punto D ha coordinate

$$F_B(\overline{AC}) - F_B(\overline{AB}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix},$$

e il punto medio M' del segmento \overline{AD} ha coordinate $\begin{bmatrix} -1/2 \\ 1/2 \end{bmatrix}$. Siccome per costruzione \overline{BM} è parallelo e congruente ad $\overline{AM'}$, troviamo che $AM'MB$ è un parallelogramma, per cui $\overline{AM} = \overline{AM'} + \overline{AB}$ e le coordinate del punto M sono

$$F_B(\overline{AM'}) + F_B(\overline{AB}) = \begin{bmatrix} -1/2 \\ 1/2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}.$$

Dunque M ed L hanno le stesse coordinate, per cui sono lo stesso punto.

Riassumiamo quanto detto finora con una definizione.

Definizione 2.4 L’insieme formato da un punto $O \in A^2$ e da due vettori non proporzionali $\vec{i}, \vec{j} \in \mathcal{V}_O^2$ si chiama (sistema di) riferimento affine $RA(O, \vec{i}, \vec{j})$ del piano. Se $\vec{i} = \overline{OA_1}$ e $\vec{j} = \overline{OA_2}$ scriveremo anche $RA(O, A_1, A_2)$ invece di $RA(O, \vec{i}, \vec{j})$. Infine, le coordinate di un punto $P \in A^2$ rispetto a un sistema di riferimento affine $RA(O, \vec{i}, \vec{j})$ sono semplicemente le coordinate di \overline{OP} rispetto alla base $B = \{\vec{i}, \vec{j}\}$.

Nello spazio \mathcal{V}_O^3 , invece, due vettori non bastano; i vettori della forma $x_1 \vec{i} + x_2 \vec{j}$ sono tutti e soli quelli del piano contenente O, A_1 e A_2 — piano che chiameremo piano generato o span dei vettori \vec{i} e \vec{j} . In simboli,

$$\text{Span}(\vec{i}, \vec{j}) = \{x_1 \vec{i} + x_2 \vec{j} \in \mathcal{V}_O^3 \mid x_1, x_2 \in \mathbb{R}\}.$$

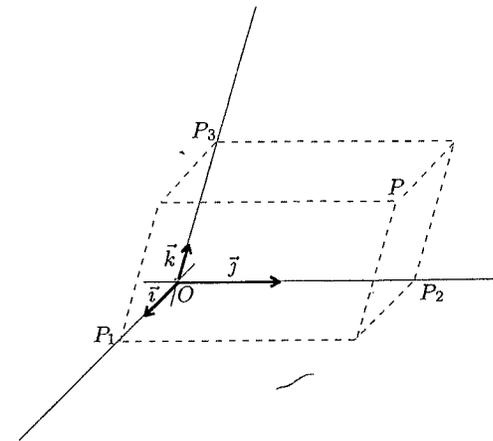


Figura 2.11 Le coordinate nello spazio.

Prendendo invece un terzo vettore $\vec{k} = \overline{OA_3}$ che non appartenga al piano generato da \vec{i} e \vec{j} si dimostra in maniera analoga a prima (vedi Figura 2.11) che qualunque vettore di \mathcal{V}_O^3 si scrive in modo unico come $\overline{OP} = x_1 \vec{i} + x_2 \vec{j} + x_3 \vec{k}$.

Dunque stavolta una base B di \mathcal{V}_O^3 è composta da tre vettori $\{\vec{i}, \vec{j}, \vec{k}\}$ non complanari; le coordinate relative a una base sono una terna (e non più una coppia) di numeri reali; e un sistema di riferimento affine verrà indicato con $RA(O, \vec{i}, \vec{j}, \vec{k})$ o con $RA(O, A_1, A_2, A_3)$. Infine, in questo caso possiamo introdurre lo spazio \mathbb{R}^3 delle terne di numeri reali e costruire un isomorfismo $F_B: \mathcal{V}_O^3 \rightarrow \mathbb{R}^3$ esattamente come prima; i dettagli sono un esercizio per te (Esercizio 2.5).

2.3 Equazioni di rette e piani

Gli elementi fondamentali della geometria del piano \mathcal{A}^2 sono i punti e le rette. Se vogliamo descrivere una retta in \mathcal{A}^2 non abbiamo molta scelta: dobbiamo dare due dei suoi punti. Il che è certo meglio che elencarli tutti, ma può essere scomodo (per esempio, trovare l'intersezione di due rette in questi termini non è certo un'operazione semplice).

Vediamo se la situazione migliora in \mathcal{V}_O^2 . In questo caso, possiamo fornire una procedura operativa — o, se preferisci, una formula — per descrivere tutti i punti della retta a partire da due di essi. Se la retta r in questione passa per l'origine O , è tutto molto facile: preso un altro punto $Q \in r$, un punto P del piano appartiene alla retta r se e solo se

$$\overrightarrow{OP} = t\overrightarrow{OQ}$$

per qualche $t \in \mathbb{R}$.

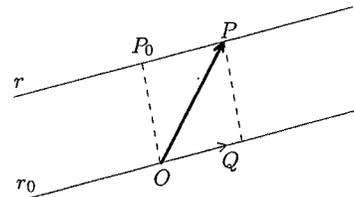


Figura 2.12 Equazione vettoriale di una retta.

Se la retta r invece è qualunque, prendiamo un punto $P_0 \in r$, e sia r_0 la retta parallela a r passante per O (Figura 2.12). Allora un punto $P \in \mathcal{A}^2$ appartiene alla retta r se e solo se il segmento $\overline{P_0P}$ è parallelo alla retta r_0 , cioè se e solo se (perché?) il vettore $\overrightarrow{PP_0} - \overrightarrow{OP_0}$ appartiene alla retta r_0 . Preso un punto $Q \in r_0$, questo vuol dire che P appartiene a r se e solo se $\overrightarrow{PP_0} - \overrightarrow{OP_0} = t\overrightarrow{OQ}$, cioè se e solo se

$$\overrightarrow{OP} = \overrightarrow{OP_0} + t\overrightarrow{OQ} \quad (2.5)$$

(equazione vettoriale di una retta) per un qualche $t \in \mathbb{R}$. Il vettore \overrightarrow{OQ} è detto *vettore direttore* della retta; chiaramente non è unico, ma due vettori direttori di r sono sempre proporzionali (perché?).

A questo punto è immediato scrivere l'equazione vettoriale della retta r passante per due punti P_1 e P_2 . Infatti, il vettore $\overrightarrow{OQ} = \overrightarrow{OP_2} - \overrightarrow{OP_1}$ è parallelo a r (perché?), per cui possiamo prenderlo come vettore direttore; quindi un punto P del piano appartiene alla retta per P_1 e P_2 se e solo se

$$\overrightarrow{OP} = \overrightarrow{OP_1} + t(\overrightarrow{OP_2} - \overrightarrow{OP_1}). \quad (2.6)$$

Osservazione 2.5 Dunque al variare di $t \in \mathbb{R}$ l'equazione (2.5) — o, se preferisci, l'equazione (2.6) — descrive tutti i punti di una retta nel piano. Ma lo stesso ragionamento che ci ha portato a queste equazioni funziona pari pari nello spazio. Quindi (2.5) e (2.6) descrivono anche tutti i punti di una retta nello spazio; basta considerare ovunque vettori di \mathcal{V}_O^3 .

In maniera analoga possiamo ottenere l'equazione vettoriale di un piano π nello spazio \mathcal{V}_O^3 . Prendiamo un punto $P_0 \in \pi$, e sia π_0 il piano parallelo a π passante per l'origine. Scegliamo una base $\{\vec{i}, \vec{j}\}$ di π_0 ; allora un punto $P \in \mathcal{A}^3$ appartiene al piano π se e solo se

$$\overrightarrow{OP} = \overrightarrow{OP_0} + s\vec{i} + t\vec{j} \quad (2.7)$$

(equazione vettoriale del piano) per qualche $s, t \in \mathbb{R}$. I vettori \vec{i} e \vec{j} , che non sono per nulla univocamente determinati, si chiamano *vettori di giacitura* del piano π .

Osservazione 2.6 Il nome "equazione vettoriale" (e più oltre quello di "equazione parametrica") è lievemente fuorviante. Qui non c'è nessuna incognita da trovare; al contrario, i numeri s e t che appaiono in queste formule (detti *parametri*) possono assumere qualunque valore. Le formule (2.5)–(2.7) sono *funzioni*, non equazioni: per esempio, la (2.5) descrive la funzione $f: \mathbb{R} \rightarrow \mathcal{V}_O^2$ data da

$$f(t) = \overrightarrow{OP_0} + t\overrightarrow{OQ}.$$

Il senso delle equazioni vettoriali e parametriche è che l'oggetto (la retta, il piano) viene descritto come l'*immagine della funzione* indicata: per esempio, la (2.5) ci dice che la retta cercata è $f(\mathbb{R})$.

Torniamo alle rette, e vediamo se l'equazione vettoriale ci aiuta a trovare l'intersezione di due rette. Sia r una retta passante per P_0 con vettore direttore \overrightarrow{OQ} , e r' una retta passante per P'_0 con vettore direttore \overrightarrow{OQ}' . Le due rette si intersecano nel punto X se e solo se esistono due numeri reali $t, t' \in \mathbb{R}$ tali che

$$\overrightarrow{OP_0} + t\overrightarrow{OQ} = \overrightarrow{OX} = \overrightarrow{OP'_0} + t'\overrightarrow{OQ}'$$

che può succedere se e solo se esistono $t, t' \in \mathbb{R}$ tali che

$$\overrightarrow{OP'_0} - \overrightarrow{OP_0} = t\overrightarrow{OQ} - t'\overrightarrow{OQ}'. \quad (2.8)$$

In altri termini, abbiamo dimostrato che *le rette r ed r' si intersecano se e solo se il vettore $\overrightarrow{OP'_0} - \overrightarrow{OP_0}$ appartiene al piano generato dai vettori direttori \overrightarrow{OQ} e \overrightarrow{OQ}'* . Questo risultato vale automaticamente sia nel piano \mathcal{V}_O^2 che nello spazio \mathcal{V}_O^3 , in quanto segue dall'equazione vettoriale di una retta, che è la stessa in entrambi i casi.

Abbiamo quindi trovato un criterio che ci dice esattamente quando due rette si intersecano. Nel piano, se i due vettori direttori \overrightarrow{OQ} e \overrightarrow{OQ}' non sono proporzionali

(ovvero, se le rette non sono *parallele*) allora formano una base di \mathcal{V}_O^3 , per cui (ricordati la Proposizione 2.3) la condizione è sempre soddisfatta — e infatti due rette non parallele nel piano s'intersecano sempre. Nello spazio, invece, due vettori non bastano per generare tutto \mathcal{V}_O^3 — e difatti esistono le rette *sghembe*, che né sono parallele né si intersecano.

Rimane però ancora aperto il problema di determinare una procedura efficace per trovare il punto d'intersezione di due rette. Per risolverlo, passiamo in \mathbb{R}^2 , ovvero fissiamo un sistema di riferimento affine $RA(O, \vec{i}, \vec{j})$. In questo sistema di riferimento, chiamiamo $\begin{vmatrix} x_0 \\ y_0 \end{vmatrix}$ le coordinate del punto P_0 , $\begin{vmatrix} l \\ m \end{vmatrix}$ le coordinate del punto Q , e $\begin{vmatrix} x \\ y \end{vmatrix}$ le coordinate del punto generico P — ovvero scriviamo $\overrightarrow{OP_0} = x_0 \vec{i} + y_0 \vec{j}$ e così via. Quindi applicando F_B all'equazione vettoriale (2.5) si ha

$$\begin{vmatrix} x \\ y \end{vmatrix} = \begin{vmatrix} x_0 \\ y_0 \end{vmatrix} + t \begin{vmatrix} l \\ m \end{vmatrix},$$

cioè P appartiene alla retta se e solo se le sue coordinate sono date da

$$\begin{cases} x = x_0 + tl, \\ y = y_0 + tm, \end{cases}$$

(equazioni parametriche di una retta nel piano) per qualche $t \in \mathbb{R}$.

Osservazione 2.7 Le equazioni vettoriali e parametriche non sono l'unico modo di descrivere una retta; come vedremo nel Paragrafo 6.4 (e forse già sai), lo si può fare anche con le cosiddette "equazioni cartesiane".

Nello spazio, scelto un riferimento affine $RA(O, \vec{i}, \vec{j}, \vec{k})$, applicando F_B a (2.5) si ha

$$\begin{vmatrix} x \\ y \\ z \end{vmatrix} = \begin{vmatrix} x_0 \\ y_0 \\ z_0 \end{vmatrix} + t \begin{vmatrix} l \\ m \\ n \end{vmatrix},$$

cioè P appartiene alla retta se e solo se le sue coordinate sono date da

$$\begin{cases} x = x_0 + tl, \\ y = y_0 + tm, \\ z = z_0 + tn, \end{cases}$$

(equazioni parametriche di una retta nello spazio) per qualche $t \in \mathbb{R}$.

Analogamente otteniamo le equazioni parametriche di un piano:

$$\begin{vmatrix} x \\ y \\ z \end{vmatrix} = \begin{vmatrix} x_0 \\ y_0 \\ z_0 \end{vmatrix} + s \begin{vmatrix} l \\ m \\ n \end{vmatrix} + t \begin{vmatrix} l' \\ m' \\ n' \end{vmatrix},$$

cioè P appartiene al piano se e solo se le sue coordinate sono date da

$$\begin{cases} x = x_0 + sl + tl', \\ y = y_0 + sm + tm', \\ z = z_0 + sn + tn', \end{cases}$$

(equazioni parametriche di un piano nello spazio) per qualche $s, t \in \mathbb{R}$.

A questo punto possiamo affrontare numericamente il problema di come trovare l'intersezione di due rette. Per semplicità supporremo di essere nel piano, ma lo stesso metodo può essere applicato nello spazio. Prendiamo di nuovo una retta r passante per P_0 con vettore direttore \overrightarrow{OQ} , e una retta r' passante per P'_0 con vettore direttore $\overrightarrow{OQ'}$. Fissiamo un sistema di riferimento affine $RA(O, \vec{i}, \vec{j})$ rispetto al quale P_0 abbia coordinate $\begin{vmatrix} x_0 \\ y_0 \end{vmatrix}$, P'_0 coordinate $\begin{vmatrix} x'_0 \\ y'_0 \end{vmatrix}$, Q coordinate $\begin{vmatrix} l \\ m \end{vmatrix}$ e Q' coordinate $\begin{vmatrix} l' \\ m' \end{vmatrix}$. Allora applicando F_B all'equazione (2.8) troviamo che le due rette si intersecano se e solo se il sistema lineare

$$\begin{cases} lt - l't' = x'_0 - x_0, \\ mt - m't' = y'_0 - y_0, \end{cases} \quad (2.9)$$

nelle incognite t e t' ammette soluzione. Ma questo è un sistema che nei casi concreti possiamo risolvere anche a mano; e se (t_0, t'_0) è la soluzione del sistema, allora il punto d'intersezione X ha coordinate

$$\begin{vmatrix} x_0 + lt_0 \\ y_0 + mt_0 \end{vmatrix} = \begin{vmatrix} x'_0 + l't'_0 \\ y'_0 + m't'_0 \end{vmatrix}$$

(perché?). In altri termini, il punto X è dato da

$$\overrightarrow{OX} = \overrightarrow{OP_0} + t_0 \overrightarrow{OQ} = \overrightarrow{OP'_0} + t'_0 \overrightarrow{OQ'}.$$

ESEMPIO 2.3 Fissiamo un sistema di riferimento affine $RA(O, \vec{i}, \vec{j})$ in \mathcal{A}^2 . Vogliamo trovare il punto d'intersezione fra la retta r passante per il punto di coordinate $\begin{vmatrix} 1 \\ 1 \end{vmatrix}$ con vettore direttore $\begin{vmatrix} 2 \\ 1 \end{vmatrix}$ e la retta r' passante per il punto $\begin{vmatrix} 2 \\ 0 \end{vmatrix}$ con vettore direttore $\begin{vmatrix} 0 \\ 1 \end{vmatrix}$. Per quanto abbiamo visto dobbiamo risolvere il sistema

$$\begin{cases} 2t - 0t' = 2 - 1, \\ 1t - 1t' = 0 - 1, \end{cases} \quad \text{ovvero} \quad \begin{cases} 2t = 1, \\ t - t' = -1, \end{cases}$$

che ha ovviamente come unica soluzione $t = 1/2$ e $t' = 3/2$. Quindi le coordinate del punto d'intersezione X sono

$$\begin{vmatrix} 1 + 2 \cdot (1/2) \\ 1 + (1/2) \end{vmatrix} = \begin{vmatrix} 2 \\ 3/2 \end{vmatrix} = \begin{vmatrix} 2 + 0 \cdot (3/2) \\ 0 + 1 \cdot (3/2) \end{vmatrix}.$$

per cui

$$\overrightarrow{OX} = 2\vec{i} + \frac{3}{2}\vec{j}.$$

Dunque il passare da \mathcal{A}^2 (o \mathcal{A}^3) a \mathbb{R}^2 (o \mathbb{R}^3) tramite la scelta di un sistema di riferimento affine trasforma problemi geometrici relativi a rette e piani in problemi relativi alla soluzione di sistemi lineari. Nel prossimo capitolo inizieremo a studiare sistematicamente i sistemi lineari; assieme a quanto visto finora questo ci porterà a identificare una struttura generale, quella di spazio vettoriale, di cui \mathcal{V}_0^2 , \mathbb{R}^2 e simili non sono altro che un caso particolare. Soltanto nel Capitolo 10 torneremo a occuparci diffusamente di rette e piani nello spazio; a quel punto avremo tutti gli strumenti necessari per risolvere efficacemente qualunque problema affine⁵ ci venga in mente riguardo rette e piani.

Esercizi

2.1 Dimostra che $(\overrightarrow{OA_1} + \overrightarrow{OA_2}) + \overrightarrow{OA_3} = \overrightarrow{OA_1} + (\overrightarrow{OA_2} + \overrightarrow{OA_3})$ quando tre dei punti O, A_1, A_2, A_3 sono allineati.

2.2 Dimostra la Proposizione 2.2.(i) nel caso $\lambda < 0$.

2.3 Definisci la somma di vettori e il prodotto per scalari in \mathcal{V}_0^3 , e dimostra le Proposizioni 2.1 e 2.2 per vettori di \mathcal{V}_0^3 .

2.4 Siano $\overrightarrow{OA} \in \mathcal{V}_0^2$ un vettore non nullo, e $\lambda \in \mathbb{R}$. Dimostra che $\lambda \overrightarrow{OA} = \overrightarrow{OO}$ se e solo se $\lambda = 0$. (Suggerimento: moltiplica per λ^{-1} .)

2.5 Sia $\mathcal{B} = \{\vec{i}, \vec{j}, \vec{k}\}$ una terna di vettori non complanari in \mathcal{V}_0^3 . Seguendo la traccia di quanto fatto nel piano, definisci un'applicazione $F_{\mathcal{B}}: \mathcal{V}_0^3 \rightarrow \mathbb{R}^3$ e una somma e un prodotto per scalari su \mathbb{R}^3 in modo che $F_{\mathcal{B}}$ risulti un'applicazione bigettiva tale che valga l'equivalente di (2.4) per tutti i vettori $\overrightarrow{OP}, \overrightarrow{OQ} \in \mathcal{V}_0^3$ e tutti i numeri reali $\lambda \in \mathbb{R}$.

2.6 Sia $\mathcal{B} = \{\vec{i}, \vec{j}\}$ una base di \mathcal{V}_0^2 . Dimostra che

$$F_{\mathcal{B}}(\vec{i}) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{e} \quad F_{\mathcal{B}}(\vec{j}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

2.7 Sia $\mathcal{B} = \{\vec{i}, \vec{j}, \vec{k}\}$ una base di \mathcal{V}_0^3 . Dati i vettori $\overrightarrow{OA} = 2\vec{i} - 3\vec{k}$, $\overrightarrow{OB} = 2\vec{j} + 3\vec{k}$ e $\overrightarrow{OC} = -\vec{i} - \vec{j} - \vec{k}$, trova le coordinate di $\overrightarrow{OD}_1 = \overrightarrow{OA} - \overrightarrow{OB} - 2\overrightarrow{OC}$ e $\overrightarrow{OD}_2 = \overrightarrow{OB} - 7\overrightarrow{OC}$.

2.8 Sia $\mathcal{B} = \{\vec{i}, \vec{j}\}$ una base di \mathcal{V}_0^2 , e considera i vettori $\overrightarrow{OA} = 2\vec{i} + 3\vec{j}$ e $\overrightarrow{OB} = -2\vec{i} + \vec{j}$. Verifica che \overrightarrow{OA} e \overrightarrow{OB} non sono proporzionali, e trova le coordinate di $\overrightarrow{OC} = \vec{i} + \vec{j}$ rispetto alla base $\mathcal{B}' = \{\overrightarrow{OA}, \overrightarrow{OB}\}$.

⁵ Cioè che non coinvolge né distanze né angoli, concetti che affronteremo solo a partire dal Capitolo 12.

2.9 Sia $\mathcal{B} = \{\vec{i}, \vec{j}\}$ una base di \mathcal{V}_0^2 , e considera i vettori $\overrightarrow{OA}_a = 2\vec{i} + a\vec{j}$ e $\overrightarrow{OB} = \vec{i} - \vec{j}$, dove $a \in \mathbb{R}$.

(i) Trova per quali valori di a i due vettori non sono proporzionali.

(ii) Calcola le coordinate di $\overrightarrow{OC} = \vec{i} + 2\vec{j}$ rispetto alla base $\mathcal{B}_a = \{\overrightarrow{OA}_a, \overrightarrow{OB}\}$ per i valori di a trovati in (i).

2.10 Dimostra che l'equazione vettoriale del piano π passante per tre punti non allineati P_1, P_2 e P_3 è

$$\overrightarrow{OP} = \overrightarrow{OP_1} + s(\overrightarrow{OP_2} - \overrightarrow{OP_1}) + t(\overrightarrow{OP_3} - \overrightarrow{OP_1}).$$

2.11 Fissato un sistema di riferimento affine $RA(O, \vec{i}, \vec{j}, \vec{k})$ nello spazio, consideriamo i vettori $\overrightarrow{OA} = \vec{i} - 3\vec{k}$ e $\overrightarrow{OB} = \vec{j} + 3\vec{k}$. Sia $\pi = \text{Span}(\overrightarrow{OA}, \overrightarrow{OB})$ il piano generato da questi due vettori. Dimostra che $\vec{i} + 2\vec{j} + 3\vec{k} \in \pi$ mentre $\vec{i} - 2\vec{j} + 3\vec{k} \notin \pi$.

2.12 Fissato un sistema di riferimento affine $RA(O, \vec{i}, \vec{j}, \vec{k})$ nello spazio, consideriamo i vettori $\overrightarrow{OA} = \vec{i} + \vec{j} + 3\vec{k}$ e $\overrightarrow{OB} = 2\vec{i} - \vec{j} - 3\vec{k}$. Sia $\pi = \text{Span}(\overrightarrow{OA}, \overrightarrow{OB})$ il piano generato da questi due vettori. Trova per quali valori di $a \in \mathbb{R}$ (se ne esistono) la retta r di equazione parametrica

$$\begin{cases} x = 8 + 4t, \\ y = 10 + 5t, \\ z = 2a + at, \end{cases}$$

è contenuta nel piano π .

2.13 Fissato un sistema di riferimento affine $RA(O, \vec{i}, \vec{j})$ nel piano, sia r la retta di equazioni parametriche

$$\begin{cases} x \\ y \end{cases} = \begin{vmatrix} -1 \\ 2 \end{vmatrix} + t \begin{vmatrix} 2 \\ -4 \end{vmatrix}.$$

Dimostra che $O \in r$.

2.14 Fissato un sistema di riferimento affine $RA(O, \vec{i}, \vec{j})$ nel piano, sia r_a la retta di equazioni parametriche

$$\begin{cases} x \\ y \end{cases} = \begin{vmatrix} 2 \\ -1 \end{vmatrix} + t \begin{vmatrix} 1 \\ a \end{vmatrix},$$

dove $a \in \mathbb{R}$. Trova per quali valori di a (se ne esistono) si ha $O \in r_a$.

2.15 Fissato un sistema di riferimento affine $RA(O, \vec{i}, \vec{j}, \vec{k})$ nello spazio, trova l'equazione parametrica della retta r passante per i punti di coordinate

$$\text{Verifica se } \begin{vmatrix} 3 \\ -e \\ -\sqrt{3} \end{vmatrix} \in r.$$

2.16 Fissato un sistema di riferimento affine $RA(O, \vec{i}, \vec{j}, \vec{k})$ nello spazio, trova l'equazione parametrica della retta r passante per i punti di coordinate $\begin{vmatrix} -1 \\ \pi \\ 0 \end{vmatrix}$ e $\begin{vmatrix} 7 \\ 52 \\ 1 \end{vmatrix}$.

2.17 Fissato un sistema di riferimento affine $RA(O, \vec{i}, \vec{j})$ nel piano, considera le rette r_0 ed r_1 di equazioni parametriche

$$r_0: \begin{cases} x = 1 + 3t, \\ y = 2 + t, \end{cases} \quad \text{e} \quad r_1: \begin{cases} x = -1 + t', \\ y = -1 + 2t'. \end{cases}$$

Dimostra che r_1 ed r_2 si intersecano, e trova il punto d'intersezione.

2.18 Fissato un sistema di riferimento affine $RA(O, \vec{i}, \vec{j})$ nel piano, sia r_1 la retta passante per i punti di coordinate $\begin{vmatrix} -1 \\ 1 \end{vmatrix}$ e $\begin{vmatrix} 0 \\ 1 \end{vmatrix}$, ed r_2 la retta per i punti di coordinate $\begin{vmatrix} 2 \\ 2 \end{vmatrix}$ e $\begin{vmatrix} 1 \\ 0 \end{vmatrix}$. Dimostra che r_1 ed r_2 si intersecano, e trova il punto d'intersezione.

2.19 Fissato un sistema di riferimento affine $RA(O, \vec{i}, \vec{j})$ nel piano, considera la retta r di equazioni parametriche

$$r: \begin{cases} x = 1 + 3t, \\ y = 2 + t, \end{cases}$$

e la retta s_a passante per i punti di coordinate $\begin{vmatrix} 0 \\ a \end{vmatrix}$ e $\begin{vmatrix} 2 \\ -1 \end{vmatrix}$, dove $a \in \mathbb{R}$. Trova per quali valori di a le rette r ed s_a si intersecano, e determina le coordinate del punto d'intersezione (quando esiste).

2.20 Fissato un sistema di riferimento affine $RA(O, \vec{i}, \vec{j}, \vec{k})$ nello spazio, sia r la retta passante per i punti di coordinate $\begin{vmatrix} 3 \\ 0 \\ 4 \end{vmatrix}$ e $\begin{vmatrix} -1 \\ 2 \\ -2 \end{vmatrix}$, e sia r' la retta passante per i punti di coordinate $\begin{vmatrix} 2 \\ 2 \\ 5 \end{vmatrix}$ e $\begin{vmatrix} 0 \\ 0 \\ -3 \end{vmatrix}$. Dimostra che le due rette si intersecano, e trova il punto d'intersezione.

2.21 Fissato un sistema di riferimento affine $RA(O, \vec{i}, \vec{j}, \vec{k})$ nello spazio, considera le rette r ed s_a di equazioni parametriche

$$r: \begin{cases} x = -1 + t, \\ y = 1, \\ z = 1 + t, \end{cases} \quad s_a: \begin{cases} x = -1 + at', \\ y = 1 + (a+1)t', \\ z = 2 - at', \end{cases}$$

dove $a \in \mathbb{R}$. Trova per quali valori di a le rette r ed s_a si intersecano, e determina le coordinate del punto d'intersezione (quando esiste).

2.22 Sia ABC un triangolo nel piano. Dimostra che le tre mediane (cioè le rette che congiungono un vertice del triangolo col punto medio del lato opposto) si intersecano in un punto (che si chiama *baricentro* del triangolo). (Suggerimento: scegli un sistema di riferimento affine opportuno e fai i conti in \mathbb{R}^2 .)

2.23 Sia $ABCD$ un tetraedro nello spazio (cioè la "piramide" a facce triangolari con vertici i quattro punti dati). Dimostra che le quattro rette che congiungono i vertici del tetraedro con i baricentri (vedi l'Esercizio precedente) delle facce opposte si intersecano in un punto (detto *baricentro* del tetraedro).

COMPLEMENTI

2C.1 Tutti i triangoli sono isosceli

Come promesso nell'Osservazione 2.3, ecco una dimostrazione (ovviamente sbagliata) del fatto che tutti i triangoli sono isosceli: mai fidarsi neppure dei disegni migliori...

Sia ABC un triangolo qualunque. Chiamato D il punto medio del segmento \overline{BC} , alziamo su D la perpendicolare DE a \overline{BC} . Tracciamo anche la bisettrice dell'angolo $B\hat{A}C$. Dobbiamo considerare due casi.

Se la bisettrice non interseca DE , vuol dire che sono parallele, e quindi la bisettrice è perpendicolare al segmento \overline{BC} , e lo interseca nel punto M . Questo vuol dire che i triangoli AMC e AMB hanno un lato in comune (\overline{AM}) e due angoli congruenti ($\widehat{AMC} \cong \widehat{AMB}$ e $\widehat{MAC} \cong \widehat{MAB}$); quindi i due triangoli sono congruenti. In particolare, $\overline{AB} \cong \overline{AC}$, cioè il triangolo ABC è isoscele.

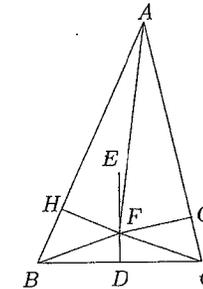


Figura 2.13 Tutti i triangoli sono isosceli.

Altrimenti, sia F il punto d'intersezione fra DE e la bisettrice. Tracciamo \overline{FB} , \overline{FC} e da F tiriamo le perpendicolari \overline{FG} ed \overline{FH} ai segmenti \overline{AC} e \overline{AB} (Figura 2.13).

⁶ Il simbolo \cong indica la congruenza di angoli (o segmenti).

Allora i triangoli AFG e AFH sono congruenti, in quanto hanno il lato \overline{AF} in comune, e gli angoli \widehat{FAG} e \widehat{GAF} sono rispettivamente uguali agli angoli \widehat{FAH} e \widehat{AHF} . Quindi $\overline{AH} \cong \overline{AG}$ e $\overline{FH} \cong \overline{FG}$.

Poi, anche i triangoli BDF e CDF sono congruenti, in quanto $\overline{BD} \cong \overline{DC}$, \overline{DF} è in comune, e gli angoli in D sono congruenti. Quindi $\overline{FB} \cong \overline{FC}$. Ora, i triangoli FHB e FGC sono triangoli rettangoli. Quindi l'area del quadrato su \overline{FB} è uguale alla somma delle aree dei quadrati su \overline{FH} e \overline{HB} ; analogamente, l'area del quadrato su \overline{FC} è uguale alla somma delle aree dei quadrati su \overline{FG} e \overline{GC} . Ma $\overline{FB} \cong \overline{FC}$ e $\overline{FH} \cong \overline{FG}$; quindi l'area del quadrato su \overline{HB} è uguale all'area del quadrato su \overline{GC} . Ma questo vuol dire che $\overline{HB} \cong \overline{GC}$. Avevamo già dimostrato che $\overline{AH} \cong \overline{AG}$; ne segue che $\overline{AB} \cong \overline{AC}$, per cui il triangolo ABC è isoscele. \square

Ovviamente, questa dimostrazione è sbagliata. Dov'è l'errore?

2C.2 Vettori liberi

In questo capitolo abbiamo considerato sempre vettori applicati in un punto fissato (l'origine). Ora, se prendiamo due vettori \overrightarrow{OA} e $\overrightarrow{OA'}$ applicati in punti diversi ma paralleli, congruenti e con lo stesso verso, chiaramente questi due vettori hanno qualcosa in comune; in un certo senso, a parte il fatto puramente incidentale di essere applicati in origini distinte, sono lo stesso vettore. In termini più formali, diremo che sono entrambi rappresentanti dello stesso vettore libero; ma per spiegare cosa significa questa frase ci serve una piccola digressione.

Definizione 2C.1 Sia A un insieme. Una relazione (binaria) su A è un sottoinsieme \mathcal{R} di $A \times A$. Se la coppia (a_1, a_2) appartiene a \mathcal{R} , diremo che a_1 è in relazione \mathcal{R} con a_2 e scriveremo $a_1 \mathcal{R} a_2$.

In pratica, quando si definisce una relazione su un insieme, prima si dice quando due elementi sono in relazione e poi si costruisce l'insieme \mathcal{R} come l'insieme delle coppie di elementi in relazione.

ESEMPIO 2C.1 Se $A = \mathbb{N}$, possiamo considerare la relazione "multiplo di". Allora $m \mathcal{R} n$ vuol dire "m è multiplo di n" e l'insieme \mathcal{R} è dato da

$$\mathcal{R} = \{(m, n) \in \mathbb{N} \times \mathbb{N} \mid m \text{ è multiplo di } n\}.$$

ESEMPIO 2C.2 Se A è l'insieme degli abitanti di Gorgonzola, possiamo considerare la relazione \mathcal{R} "genitore di". Allora $(a_1, a_2) \in \mathcal{R}$ se e solo se a_1 è il padre (o la madre) di a_2 .

Definizione 2C.2 Una relazione d'equivalenza su un insieme A è una relazione binaria su A , di solito indicata con \sim , che soddisfa le seguenti proprietà:

- (i) *Riflessività*: $a \sim a$ per ogni $a \in A$.
- (ii) *Simmetria*: $a_1 \sim a_2$ implica $a_2 \sim a_1$ per ogni $a_1, a_2 \in A$.
- (iii) *Transitività*: $a_1 \sim a_2$ e $a_2 \sim a_3$ implicano $a_1 \sim a_3$ per ogni $a_1, a_2, a_3 \in A$.

Se ci pensi un attimo, le proprietà di una relazione d'equivalenza sono esattamente le stesse che ha la relazione d'uguaglianza fra numeri: un numero è uguale a se stesso, se un numero è uguale a un altro allora il secondo è uguale al primo, e se un numero è uguale a un secondo che a sua volta è uguale a un terzo allora anche il primo è uguale al terzo.

ESEMPIO 2C.3 Sia A l'insieme degli abitanti di Gorgonzola. Allora la relazione "essere parente di" è una relazione d'equivalenza, come puoi verificare facilmente. Invece la relazione "essere genitore di" non è una relazione d'equivalenza (non è né riflessiva, né simmetrica, né transitiva).

ESEMPIO 2C.4 La relazione "essere multiplo di" su \mathbb{N} è riflessiva e transitiva (controlla), ma non simmetrica.

ESEMPIO 2C.5 La relazione di parallelismo fra rette è una relazione d'equivalenza; la relazione di incidenza fra rette no (è riflessiva e simmetrica, ma non transitiva).

Quando si ha una relazione d'equivalenza, viene naturale mettere assieme tutti gli elementi equivalenti in un unico sottoinsieme.

Definizione 2C.3 Sia A un insieme su cui è definita una relazione d'equivalenza \sim . Se $a \in A$, la classe d'equivalenza $[a] \subseteq A$ di a è il sottoinsieme di tutti gli elementi di A equivalenti ad a . Se $b \in [a]$, diremo che b è un rappresentante della classe d'equivalenza $[a]$. L'insieme di tutte le classi d'equivalenza si chiama *insieme quoziente* e viene indicato con A/\sim . L'applicazione $\pi: A \rightarrow A/\sim$ che a ogni elemento $a \in A$ associa la sua classe d'equivalenza $\pi(a) = [a]$ si chiama *applicazione quoziente*.

Le classi d'equivalenza dividono l'insieme A in sottoinsiemi a due a due disgiunti:

Proposizione 2C.1 Sia A un insieme su cui è definita una relazione d'equivalenza \sim , e $a, b \in A$. Allora o $[a] = [b]$ (che accade se e solo se $a \sim b$), oppure $[a] \cap [b] = \emptyset$. In particolare, A è l'unione delle classi d'equivalenza, che sono a due a due disgiunte.

Dimostrazione. Prima di tutto notiamo che $a \in [a]$, grazie alla riflessività. La simmetria poi ci dice che se $b \in [a]$ (che accade se e solo se $b \sim a$) allora $a \in [b]$ (in quanto $a \sim b$). Inoltre, se $b \in [a]$ allora $[b] \subseteq [a]$: infatti, se $c \in [b]$ vuol dire che $c \sim b$ e, per la transitività, $c \sim a$. Analogamente segue che $[a] \subseteq [b]$, per cui $b \sim a$ implica $[b] = [a]$.

Supponiamo esista $c \in [a] \cap [b]$. Allora $c \sim a$ e $c \sim b$; la simmetria ci dice $b \sim c$, e la transitività $b \sim a$, per cui possiamo concludere che $[b] = [a]$. Quindi o $[a] \cap [b] = \emptyset$ oppure $[a] = [b]$, come desiderato. In particolare, le classi d'equivalenza sono a due a due disgiunte; siccome $a \in [a]$, l'unione di tutte le classi d'equivalenza contiene tutti gli elementi di A , e abbiamo finito. \square

Cosa c'entra tutto ciò con i vettori? Il fatto è che i vettori liberi possono essere definiti come classi d'equivalenza di vettori applicati.

Definizione 2C.4 Indichiamo⁷ con $\tilde{\mathcal{V}}^2$ l'insieme di tutti i vettori applicati del piano, qualunque sia la loro origine. Diremo che due vettori applicati \overrightarrow{OA} e $\overrightarrow{O'A'}$ sono *equivalenti*, e scriveremo $\overrightarrow{OA} \sim \overrightarrow{O'A'}$, se sono paralleli, congruenti e hanno lo stesso verso. In altre parole, $\overrightarrow{OA} \sim \overrightarrow{O'A'}$ se e solo se $OO'A'A$ è un parallelogramma (vedi la Figura 2.14), cioè se e solo se $\overrightarrow{OA'} = \overrightarrow{OA} + \overrightarrow{OO'}$.

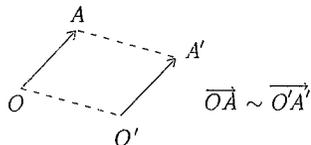


Figura 2.14 Vettori applicati equivalenti.

Verifichiamo velocemente che si tratta di una relazione d'equivalenza. Chiaramente, ogni vettore applicato è equivalente a se stesso, per cui \sim è riflessiva. Se $\overrightarrow{OA} \sim \overrightarrow{O'A'}$, allora $OO'A'A$ è un parallelogramma: dunque anche $O'OAA'$ è un parallelogramma (lo stesso), e $\overrightarrow{O'A'} \sim \overrightarrow{OA}$. Infine, supponiamo che $\overrightarrow{OA} \sim \overrightarrow{O'A'}$ e che $\overrightarrow{O'A'} \sim \overrightarrow{O''A''}$ (Figura 2.15). Allora $OO'A'A$ e $O'O'A''A''$ sono parallelogrammi, per cui anche $OO''A''A$ lo è, e $\overrightarrow{OA} \sim \overrightarrow{O''A''}$, come volevamo.

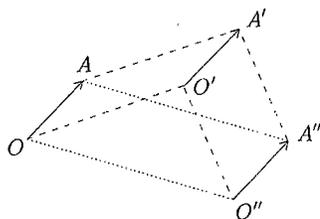


Figura 2.15 La proprietà transitiva.

I vettori liberi sono esattamente le classi d'equivalenza rispetto a questa relazione.

Definizione 2C.5 Un *vettore libero* del piano è una classe d'equivalenza per la relazione \sim sull'insieme $\tilde{\mathcal{V}}^2$. L'insieme quoziente $\tilde{\mathcal{V}}^2/\sim$ di tutti i vettori liberi verrà indicato con \mathcal{V}^2 .

⁷ Per semplicità, in questo paragrafo lavoreremo solo con vettori del piano, ma potrai verificare facilmente che tutto quanto diremo vale anche per vettori dello spazio.

Siccome due vettori applicati sono equivalenti se e solo se hanno uguale lunghezza, direzione e verso, dovrebbe esserti chiaro che questa definizione coincide col concetto intuitivo di vettore libero.

Osservazione 2C.1 Se $v \in \mathcal{V}^2$ è un vettore libero e $O \in \mathcal{A}^2$ è un punto del piano, allora esiste un unico punto $A \in \mathcal{A}^2$ tale che $v = [\overrightarrow{OA}]$, cioè esiste un unico vettore applicato in O che rappresenta v . Infatti, se $\overrightarrow{O'B}$ è un qualunque rappresentante di v , allora esiste un unico punto A tale che $OO'BA$ sia un parallelogramma: A è dato da $\overrightarrow{OA} = \overrightarrow{OB} - \overrightarrow{OO'}$. Quindi ciascuna classe d'equivalenza interseca ciascun \mathcal{V}_O^2 in esattamente uno e un solo vettore applicato.

Anche sui vettori liberi possiamo definire le operazioni di somma e di prodotto per uno scalare, utilizzando in maniera appropriata dei rappresentanti.

Definizione 2C.6 Siano $v, w \in \mathcal{V}^2$ due vettori liberi; e $\lambda \in \mathbb{R}$. Scegliamo un punto $O \in \mathcal{A}^2$, e sia \overrightarrow{OA} (rispettivamente, \overrightarrow{OB}) l'unico (vedi l'Osservazione 2C.1) rappresentante di v (rispettivamente, w) applicato in O . Allora definiamo la *somma* di v e w ponendo $v+w = [\overrightarrow{OA} + \overrightarrow{OB}]$, e il *prodotto per scalari* ponendo $\lambda v = [\lambda \overrightarrow{OA}]$.

In questa definizione abbiamo scelto un punto arbitrario O : dobbiamo quindi verificare che cambiando punto la classe d'equivalenza ottenuta non cambia.

Cominciamo con la somma. Prendiamo quattro vettori applicati $\overrightarrow{OA}, \overrightarrow{OB} \in \mathcal{V}_O^2$ e $\overrightarrow{O'A'}, \overrightarrow{O'B'} \in \mathcal{V}_{O'}^2$, e poniamo $\overrightarrow{OC} = \overrightarrow{OA} + \overrightarrow{OB}$ e $\overrightarrow{O'C'} = \overrightarrow{O'A'} + \overrightarrow{O'B'}$. Supponiamo (Figura 2.16) che $\overrightarrow{OA} \sim \overrightarrow{O'A'}$ e $\overrightarrow{OB} \sim \overrightarrow{O'B'}$; dobbiamo dimostrare che $\overrightarrow{OC} \sim \overrightarrow{O'C'}$. Per ipotesi, $OO'B'B$ è un parallelogramma; quindi $\overrightarrow{OO'}$ è parallelo e congruente a $\overrightarrow{BB'}$. Poi, \overrightarrow{BC} è parallelo e congruente a \overrightarrow{OA} , che a sua volta è parallelo e congruente a $\overrightarrow{O'A'}$, che a sua volta è parallelo e congruente a $\overrightarrow{B'C'}$; quindi $BB'C'C$ è un parallelogramma, per cui $\overrightarrow{BB'}$ è parallelo e congruente a $\overrightarrow{CC'}$. Mettendo insieme il tutto troviamo che $\overrightarrow{OO'}$ è parallelo e congruente a $\overrightarrow{CC'}$, per cui $OO'C'C$ è un parallelogramma e $\overrightarrow{OC} \sim \overrightarrow{O'C'}$, come volevamo.

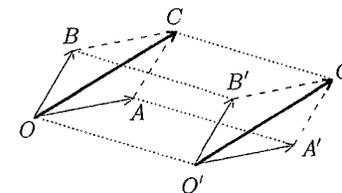


Figura 2.16 La somma di vettori liberi.

Il caso del prodotto è più semplice. Prendiamo due vettori applicati equivalenti \overrightarrow{OA} e $\overrightarrow{O'A'}$, e poniamo $\overrightarrow{OB} = \lambda \overrightarrow{OA}$ e $\overrightarrow{O'B'} = \lambda \overrightarrow{O'A'}$; dobbiamo dimostrare

che $\overrightarrow{OB} \sim \overrightarrow{O'B'}$. Grazie alla definizione di prodotto per scalari, \overrightarrow{OB} e $\overrightarrow{O'B'}$ hanno la stessa lunghezza (quella di \overrightarrow{OA} moltiplicata per $|\lambda|$), la stessa direzione (quella di \overrightarrow{OA}) e lo stesso verso (quello di \overrightarrow{OA} se $\lambda > 0$, quello opposto se $\lambda < 0$), per cui sono equivalenti.

Dunque abbiamo definito le operazioni di somma e prodotto per scalari anche sull'insieme dei vettori liberi. Potremmo dimostrare direttamente che queste operazioni hanno le solite proprietà, ma è più istruttivo procedere in modo lievemente diverso.

Fissiamo un punto $O \in \mathcal{A}^2$, e consideriamo l'applicazione $\pi: \mathcal{V}_O^2 \rightarrow \mathcal{V}^2$ che a ogni vettore applicato in O associa il vettore libero corrispondente (in altre parole, π è la restrizione a \mathcal{V}_O^2 dell'applicazione quoziente introdotta nella Definizione 2C.3). Prima di tutto, l'Osservazione 2C.1 ci assicura che π è bigettiva. Inoltre, noi abbiamo definito somma e prodotto per scalari in modo che

$$\pi(\overrightarrow{OA}) + \pi(\overrightarrow{OB}) = \pi(\overrightarrow{OA} + \overrightarrow{OB}) \quad \text{e} \quad \lambda \pi(\overrightarrow{OA}) = \pi(\lambda \overrightarrow{OA})$$

per ogni $\overrightarrow{OA}, \overrightarrow{OB} \in \mathcal{V}_O^2$ e $\lambda \in \mathbb{R}$. Quindi π è un'applicazione bigettiva che conserva le operazioni, proprio come l'applicazione $F_B: \mathcal{V}_O^2 \rightarrow \mathbb{R}^2$ introdotta nel Paragrafo 2.2. In particolare possiamo concludere che somma e prodotto per scalari in \mathcal{V}^2 hanno le stesse proprietà che avevano in \mathcal{V}_O^2 , senza bisogno di dimostrarlo nuovamente. Possiamo quindi parlare di basi e coordinate anche per i vettori liberi come per i vettori applicati. Nel Capitolo 4 riassumeremo tutto ciò dicendo che \mathcal{V}^2 è un altro esempio di spazio vettoriale.

L'eliminazione di Gauss

Nel capitolo precedente abbiamo visto come l'introduzione di un sistema di coordinate permetta di ricondurre la soluzione di problemi geometrici alla soluzione di sistemi lineari (cioè composti da equazioni di primo grado). In questo capitolo iniziamo a studiare i sistemi lineari quadrati, descrivendo in dettaglio il metodo di eliminazione di Gauss, una delle procedure più efficienti per la loro risoluzione; l'esame dei sistemi lineari qualunque è rimandato al Capitolo 6, quando avremo sviluppato gli strumenti teorici necessari per una comprensione effettiva e completa dell'argomento. La discussione comprenderà anche qualche osservazione sui diversi modi in cui è possibile dimostrare un teorema. Infine, i Complementi a questo capitolo forniscono una breve introduzione al principio d'induzione.

3.1 Esempi e definizioni

Studiando i sistemi lineari, i problemi principali che vogliamo risolvere sono tre: quando un sistema lineare ammette soluzioni? E se le ammette, quante sono? E come si trovano? Per avere un'idea del tipo di situazioni che si possono incontrare, cominciamo con l'esaminare un esempio semplicissimo.

ESEMPIO 3.1 Consideriamo il caso di un'equazione di primo grado in una sola incognita: $ax = b$. Se $a \neq 0$, l'equazione ammette un'unica soluzione $x = b/a$; se $a = 0$ ma $b \neq 0$, l'equazione non ha alcuna soluzione; se $a = 0$ e $b = 0$, l'equazione ha infinite soluzioni, in quanto qualunque $x \in \mathbb{R}$ la soddisfa.

Dunque in questo caso le soluzioni o non esistono o sono infinite oppure ce n'è una sola; non capita mai che siano in numero finito maggiore di uno. Sarebbe interessante sapere se questo accade per ogni sistema lineare. In teoria, potrebbe dipendere dall'aver considerato una sola equazione; magari ci sono sistemi di due equazioni che ammettono esattamente due soluzioni, sistemi di diciotto equazioni che ammettono esattamente diciotto soluzioni, e così via. Proviamo allora a esaminare qualche esempio di sistema lineare con più di un'equazione (e più di un'incognita), e vediamo che cosa succede.

ESEMPIO 3.2 Consideriamo il seguente sistema lineare di tre equazioni in tre incognite:

$$\begin{cases} x_1 - 2x_2 + x_3 = 1, \\ x_1 - x_2 + 4x_3 = 2, \\ x_1 - 2x_2 + 2x_3 = 3. \end{cases} \quad (3.1)$$

Per risolverlo, cominciamo col notare che la prima e la terza equazione differiscono solo per il coefficiente di x_3 e il termine noto. Quindi sottraendo la prima equazione alla terza troviamo subito il valore di x_3 :

$$\begin{cases} x_1 - 2x_2 + x_3 = 1, \\ x_1 - x_2 + 4x_3 = 2, \\ x_3 = 2. \end{cases}$$

Già che ci siamo, possiamo sottrarre la prima equazione anche alla seconda, ottenendo

$$\begin{cases} x_1 - 2x_2 + x_3 = 1, \\ x_2 + 3x_3 = 1, \\ x_3 = 2. \end{cases}$$

Così la terza equazione ci fornisce il valore di x_3 , che sostituito nella seconda ci dà il valore di x_2 , e sostituendo entrambi nella prima otteniamo il valore di x_1 :

$$\begin{cases} x_1 = -11, \\ x_2 = -5, \\ x_3 = 2. \end{cases}$$

Una veloce verifica mettendo questi valori in (3.1) conferma che abbiamo risolto il sistema, che quindi ammette un'unica soluzione (e non tre).

ESEMPIO 3.3 Ora proviamo con un sistema lievemente diverso:

$$\begin{cases} x_1 - 2x_2 + x_3 = 1, \\ x_1 - 2x_2 + 4x_3 = 2, \\ x_1 - 2x_2 + 2x_3 = 3, \end{cases} \quad (3.2)$$

dove è cambiato solo il coefficiente di x_2 nella seconda equazione. Procedendo come prima, cioè sottraendo la prima equazione alla seconda e alla terza, questa volta otteniamo

$$\begin{cases} x_1 - 2x_2 + x_3 = 1, \\ 3x_3 = 1, \\ x_3 = 2. \end{cases}$$

Ma allora la seconda equazione darebbe $x_3 = 1/3$, mentre la terza darebbe $x_3 = 2$, e non possono certo essere soddisfatte contemporaneamente. In altre parole, questo sistema non ha soluzione.

ESEMPIO 3.4 Facciamo ancora una piccola modifica, cambiando stavolta il termine noto nella seconda equazione, e consideriamo il sistema

$$\begin{cases} x_1 - 2x_2 + x_3 = 1, \\ x_1 - 2x_2 + 4x_3 = 7, \\ x_1 - 2x_2 + 2x_3 = 3. \end{cases} \quad (3.3)$$

Sottraendo nuovamente la prima equazione alla seconda e alla terza otteniamo

$$\begin{cases} x_1 - 2x_2 + x_3 = 1, \\ 3x_3 = 6, \\ x_3 = 2. \end{cases}$$

Questa volta la seconda e la terza equazione non sono più in contrasto fra loro; entrambe ci dicono che $x_3 = 2$. Inoltre, contrariamente a quanto accadeva nell'Esempio 3.2, la seconda equazione non fornisce alcuna condizione sull'incognita x_2 ; in altri termini, x_2 è *libera*, può assumere qualsiasi valore. Sostituendo quindi $x_3 = 2$ nella prima equazione e ponendo $x_2 = t$ (per ricordarci che il valore di x_2 non è fissato ma è libero) otteniamo

$$\begin{cases} x_1 = -1 + 2t, \\ x_2 = t, \\ x_3 = 2. \end{cases}$$

Stavolta il sistema ha infinite soluzioni, una per ogni valore di $t \in \mathbb{R}$. Possiamo anche essere più precisi; infatti, utilizzando le notazioni introdotte nel capitolo precedente, le soluzioni possono essere descritte da

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

che è l'equazione parametrica di una retta nello spazio. Quindi in un senso molto preciso possiamo dire che il sistema (3.3) ammette una *retta* di soluzioni.

Osservazione 3.1 Vale la pena notare esplicitamente che per passare da un sistema con soluzione unica a uno senza soluzione (o con infinite soluzioni) abbiamo modificato i coefficienti, mentre per passare da un sistema senza soluzioni a uno con infinite soluzioni abbiamo toccato i termini noti. Non è un caso, e ci ritorneremo.

Dunque da questi esempi sembrerebbe che, anche per sistemi con più di un'equazione, le soluzioni o non esistono o sono infinite oppure ce n'è una sola. Potrebbero però esistere altri esempi in cui succedono cose diverse; per sapere come stanno davvero le cose dobbiamo studiare la situazione in generale. Cominciamo introducendo un po' di terminologia.

Osservazione 3.2 Quanto faremo da qui in poi (a meno che non venga espressamente detto il contrario) dipende solo dalle proprietà formali della somma e del prodotto

di numeri reali descritte nel Paragrafo 1.3, e quindi vale in qualunque campo. Per semplicità, lavoreremo soltanto in \mathbb{R} , ma ricordati che potrai applicare quanto vedremo anche, per esempio, ai numeri complessi.

Definizione 3.1 La forma generale di un sistema lineare di m equazioni in n incognite è

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1, \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m. \end{cases} \quad (3.4)$$

Il numero n delle incognite è anche detto *ordine* del sistema. I numeri a_{11}, \dots, a_{mn} sono i *coefficienti* del sistema, e solitamente sono raccolti in una tabella di numeri, la *matrice dei coefficienti*

$$A = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{vmatrix}. \quad (3.5)$$

I numeri b_1, \dots, b_m sono i *termini noti* del sistema, e usualmente sono raccolti in una colonna (che, per motivi che vedremo più avanti, si chiama *vettore dei termini noti*):

$$b = \begin{vmatrix} b_1 \\ \vdots \\ b_m \end{vmatrix}.$$

Analogamente, il *vettore delle incognite* è

$$x = \begin{vmatrix} x_1 \\ \vdots \\ x_n \end{vmatrix}.$$

La tabella di numeri

$$A' = \begin{vmatrix} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{vmatrix}$$

è detta *matrice completa* del sistema. Per brevità (e per motivi molto più profondi che discuteremo in seguito), spesso e volentieri scriveremo

$$Ax = b$$

per indicare il sistema (3.4).

Definizione 3.2 Una *soluzione* del sistema (3.4) è una n -upla (v_1, \dots, v_n) di numeri che sostituiti ordinatamente alle incognite x_1, \dots, x_n soddisfano tutte le equazioni del sistema. Diremo che il sistema (3.4) è *compatibile* se ammette almeno una soluzione.

Definizione 3.3 In generale, una *matrice* con m righe ed n colonne è una tabella rettangolare di numeri con m righe ed n colonne, come quella in (3.5). L'insieme delle matrici con m righe ed n colonne (o, come talvolta diremo, delle matrici $m \times n$) a coefficienti reali sarà indicato con $M_{m,n}(\mathbb{R})$. Se i coefficienti sono complessi scriveremo $M_{m,n}(\mathbb{C})$, e analogamente per altri campi. A volte la matrice (3.5) sarà indicata con $A = (a_{ij})$; il numero a_{ij} è l'*elemento* (o *coefficiente*) di posto (i, j) della matrice A , dove il primo indice i è l'*indice di riga*, mentre il secondo indice j è l'*indice di colonna*.

Definizione 3.4 Sia $A \in M_{m,n}(\mathbb{R})$ una matrice $m \times n$. Indicheremo con A_i la sua riga i -esima (per $i = 1, \dots, m$), e con A^j la sua j -esima colonna (per $j = 1, \dots, n$):

$$A_i = |a_{i1} \quad a_{i2} \quad \cdots \quad a_{in}|, \quad A^j = \begin{vmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{vmatrix}.$$

Una *matrice quadrata* è proprio una matrice quadrata, cioè con tante righe quante colonne. Una matrice quadrata $n \times n$ sarà detta di *ordine* n . La *diagonale principale* di una matrice quadrata è la diagonale che va dall'angolo in alto a sinistra a quello in basso a destra; è composta dagli elementi a_{ii} , in cui l'indice di riga è uguale all'indice di colonna. Una matrice quadrata A è detta *diagonale* se tutti gli elementi al di fuori della diagonale principale sono nulli; è detta *triangolare superiore* se gli elementi al di sotto (ovvero a sinistra) della diagonale principale sono tutti nulli; *triangolare inferiore* se invece sono zero gli elementi sopra la diagonale principale. In tutti e tre i casi, gli elementi della diagonale principale possono essere sia nulli che non nulli, non è importante.

ESEMPIO 3.5 La matrice

$$A = \begin{vmatrix} 2 & 3 & 1 \\ \sqrt{2} & 0 & -6/7 \end{vmatrix}$$

è una matrice con 2 righe e 3 colonne a coefficienti reali, cioè $A \in M_{2,3}(\mathbb{R})$. I suoi elementi sono dati da $a_{11} = 2$, $a_{12} = 3$, $a_{13} = 1$, $a_{21} = \sqrt{2}$, $a_{22} = 0$, e $a_{23} = -6/7$. Le sue colonne sono

$$A^1 = \begin{vmatrix} 2 \\ \sqrt{2} \end{vmatrix}, \quad A^2 = \begin{vmatrix} 3 \\ 0 \end{vmatrix}, \quad \text{e} \quad A^3 = \begin{vmatrix} 1 \\ -6/7 \end{vmatrix};$$

le sue righe sono

$$A_1 = |2 \quad 3 \quad 1| \quad \text{e} \quad A_2 = |\sqrt{2} \quad 0 \quad -6/7|.$$

ESEMPIO 3.6 La matrice

$$A = \begin{vmatrix} 0 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 3\pi^2 \end{vmatrix}$$

è diagonale, mentre la matrice

$$B = \begin{vmatrix} 4 & 0 & \pi \\ 0 & 6\sqrt{2} & 27 \\ 0 & 0 & 0 \end{vmatrix}$$

è triangolare superiore, e la matrice

$$C = \begin{vmatrix} 0 & 0 & 0 \\ 2/3 & 0 & 0 \\ e+1 & 2/3 & 0 \end{vmatrix}$$

è triangolare inferiore.

3.2 Sistemi triangolari superiori

In questo paragrafo ci occuperemo di un tipo particolarmente semplice di sistema lineare.

Definizione 3.5 Un sistema lineare *quadrato* (cioè con tante equazioni quante incognite) si dice *triangolare superiore* se la sua matrice dei coefficienti è triangolare superiore.

Negli Esempi 3.2–3.4 abbiamo studiato alcuni sistemi lineari quadrati riducendoli prima a forma triangolare superiore e poi risolvendoli (quando possibile) a occhio. Come vedremo, potremo sempre procedere in questo modo; per questo motivo dimostriamo ora un risultato che ci dice esattamente quando un sistema triangolare superiore ammette soluzione unica.

Proposizione 3.1 *Un sistema lineare triangolare superiore $Ax = b$ di n equazioni in n incognite ammette una e una sola soluzione se e solo se tutti gli elementi della diagonale principale della matrice A dei coefficienti sono diversi da zero.*

Dimostrazione. Supponiamo prima di tutto che tutti gli elementi a_{11}, \dots, a_{nn} della diagonale principale di A siano non nulli. In questo caso, l'ultima equazione del sistema è della forma $a_{nn}x_n = b_n$, con $a_{nn} \neq 0$; quindi ammette l'unica soluzione $v_n = b_n/a_{nn}$. Sostituendo v_n nella penultima equazione otteniamo

$$a_{n-1,n-1}x_{n-1} = b_{n-1} - a_{n-1,n}v_n,$$

che di nuovo ammette un'unica soluzione v_{n-1} . Procedendo in questo modo giungiamo fino alla prima equazione, che ora è diventata

$$a_{11}x_1 = b_1 - a_{12}v_2 - \dots - a_{1n}v_n;$$

dunque di nuovo troviamo un'unica soluzione v_1 . In conclusione, quindi, la n -upla (v_1, \dots, v_n) così determinata è l'unica soluzione del sistema.

Osservazione 3.3 Questa dimostrazione è *costruttiva*: fornisce un metodo effettivo — un *algoritmo* — per il calcolo della soluzione. Questo algoritmo si chiama *risoluzione all'indietro* di un sistema triangolare superiore.

Osservazione 3.4 In questa prima parte della dimostrazione, abbiamo proceduto per *dimostrazione diretta*: abbiamo supposto vera l'ipotesi e, tramite ragionamenti, abbiamo dedotto la tesi. Questo è il modo più naturale per dimostrare affermazioni del tipo “ A implica B ”, ma non l'unico; ve ne sono perlomeno altri due. La *dimostrazione per assurdo* procede supponendo vere contemporaneamente l'ipotesi “ A ” e la negazione “non B ” della tesi per poi giungere, tramite ragionamenti, a una contraddizione, cioè a qualcosa di sicuramente falso. Quindi “ A ” e “non B ” non possono essere contemporaneamente veri, per cui se “ A ” è vera anche “ B ” dev'esserlo¹. Invece, la *dimostrazione inversa* suppone vera la negazione “non B ” della tesi e, tramite ragionamenti, deduce la negazione “non A ” dell'ipotesi. Anche questa è una dimostrazione dell'affermazione “ A implica B ”: infatti, se “non B ” implica “non A ”, allora se “ A ” e “non B ” fossero contemporaneamente vere lo sarebbero anche “ A ” e “non A ”, che è impossibile.

Dunque vi sono almeno tre modi di procedere in una dimostrazione; quale scegliere dipende dalla situazione specifica, e anche dal gusto di chi dimostra. Ritornando alla nostra Proposizione, utilizzeremo una dimostrazione inversa.

Dobbiamo dimostrare che se il nostro sistema $Ax = b$ ammette una e una sola soluzione allora tutti gli elementi sulla diagonale principale di A sono diversi da zero. Supponiamo che ci sia almeno un coefficiente nullo sulla diagonale principale (la negazione della tesi); dobbiamo dimostrare che il sistema non ammette un'unica soluzione, cioè o ne ammette più di una o non ne ammette neanche una (la negazione dell'ipotesi).

Se il nostro sistema non ammette soluzioni, abbiamo finito. Supponiamo allora che una soluzione (v_1, \dots, v_n) ci sia; ne vogliamo costruire un'altra, diversa.

Sia k il minimo indice per cui $a_{kk} = 0$; in particolare, $a_{11}, \dots, a_{k-1,k-1} \neq 0$. Sostituiamo v_{k+1}, \dots, v_n al posto di x_{k+1}, \dots, x_n nella k -esima equazione. Ricordando che $a_{kk} = 0$ troviamo

$$0x_k = b_k - a_{k,k+1}v_{k+1} - \dots - a_{kn}v_n. \quad (3.6)$$

Siccome abbiamo supposto che il nostro sistema ha almeno una soluzione, dev'essere per forza (ricorda l'Esempio 3.1)

$$b_k - a_{k,k+1}v_{k+1} - \dots - a_{kn}v_n = 0;$$

quindi qualunque valore di x_k è soluzione di (3.6). Prendiamo allora $w \neq v_k$, e sostituiamo w, v_{k+1}, \dots, v_n al posto di x_k, x_{k+1}, \dots, x_n nelle prime $k-1$ equazioni. Otteniamo così un sistema triangolare superiore di ordine $k-1$ i cui coefficienti sulla

¹ Una dimostrazione per assurdo si riconosce subito: contiene sempre una frase del tipo “supponiamo, per assurdo, che la tesi sia falsa”.

diagonale principale sono tutti non nulli; per quanto già visto, questo sistema ammette un'unica soluzione (u_1, \dots, u_{k-1}) . Ma allora le n -uple $(v_1, \dots, v_{k-1}, v_k, v_{k+1}, \dots, v_n)$ e $(u_1, \dots, u_{k-1}, v_1, v_{k+1}, \dots, v_n)$ sono due soluzioni distinte del nostro sistema, come volevamo. \square

Osservazione 3.5 La prima parte della Proposizione 3.1 può essere dimostrata anche usando una tecnica particolare di dimostrazione diretta, detta *dimostrazione per induzione*. L'idea è che per dimostrare che una qualche affermazione " A_n " dipendente da un numero naturale n è vera per tutti gli $n \in \mathbb{N}$, basta dimostrare che è vera per $n = 1$ (o 0, a seconda dei casi) e poi dimostrare che se è vera per un qualunque $n \in \mathbb{N}$ (questa si chiama *ipotesi induttiva*) allora è vera anche per $n + 1$. Infatti, in questo modo sapendo che è vera per 1 dimostriamo che è vera per 2, e poi per 3, e poi per 4, e così via.

Per esempio, la dimostrazione per induzione della prima parte della Proposizione 3.1 funziona come segue. Il numero naturale in questione è l'ordine del sistema — e diremo "procediamo per induzione sull'ordine del sistema". Quando il sistema ha ordine 1, è una sola equazione in un'incognita, con coefficiente non nullo: quindi in questo caso vi è un'unica soluzione (cioè l'affermazione è vera).

Supponiamo allora l'affermazione vera per tutti i sistemi triangolari superiori di ordine n (ipotesi induttiva), e consideriamo un sistema triangolare superiore di ordine $n + 1$. L'ultima equazione sarà del tipo $a_{n+1, n+1} x_{n+1} = b_{n+1}$ con $a_{n+1, n+1} \neq 0$: quindi ammette un'unica soluzione $v_{n+1} = b_{n+1}/a_{n+1, n+1}$. Sostituiamo v_{n+1} al posto di x_{n+1} nelle restanti n equazioni; otteniamo un sistema triangolare superiore di ordine n con coefficienti non nulli sulla diagonale principale. Per l'ipotesi induttiva, questo sistema ammette un'unica soluzione (v_1, \dots, v_n) : quindi $(v_1, \dots, v_n, v_{n+1})$ è l'unica soluzione del nostro sistema.

Dunque il *principio d'induzione* (che è il nome del ragionamento che si applica nelle dimostrazioni per induzione) ci ha permesso di rimpiazzare la verifica per sistemi di ordine qualunque con una semplice verifica per sistemi di ordine 1 e un ragionamento che ci fa passare dall'ordine n all'ordine $n + 1$. In questo caso semplice abbiamo solo risparmiato un "e così via"; ma in situazioni più complicate (e ne vedremo) la dimostrazione verrà drasticamente semplificata. Per altre informazioni sul principio d'induzione vedi i Complementi a questo capitolo.

Vediamo ora degli esempi di risoluzione all'indietro di sistemi triangolari superiori.

ESEMPIO 3.7 Consideriamo il seguente sistema di tre equazioni in tre incognite:

$$\begin{cases} 3x - y + 2z = 3, \\ -y + 2z = 0, \\ 4z = -4. \end{cases}$$

La matrice dei coefficienti e il vettore dei termini noti sono

$$A = \begin{bmatrix} 3 & -1 & 2 \\ 0 & -1 & 2 \\ 0 & 0 & 4 \end{bmatrix}, \quad b = \begin{bmatrix} 3 \\ 0 \\ -4 \end{bmatrix}.$$

Dunque il sistema è triangolare superiore con elementi non nulli sulla diagonale principale: per la Proposizione 3.1, il sistema ammette una e una sola soluzione, e possiamo trovarla risolvendo all'indietro. Infatti l'ultima equazione dice che $z = -1$: sostituendo nella seconda troviamo $-y = 2$, cioè $y = -2$; sostituendo nella prima troviamo $3x = 3$, cioè $x = 1$. Quindi la soluzione è

$$\begin{cases} x = 1, \\ y = -2, \\ z = -1. \end{cases}$$

ESEMPIO 3.8 Consideriamo il seguente sistema di tre equazioni in tre incognite:

$$\begin{cases} 2x + 4y - 2z = 8, \\ 3z = 6, \\ 2z = -2. \end{cases}$$

La matrice dei coefficienti e il vettore dei termini noti sono

$$A = \begin{bmatrix} 2 & 4 & -2 \\ 0 & 0 & 3 \\ 0 & 0 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 8 \\ 6 \\ -2 \end{bmatrix}.$$

Il sistema è triangolare superiore, ma stavolta $a_{22} = 0$: quindi la Proposizione 3.1 e l'Esercizio 3.3 ci dicono che il sistema o non ha soluzioni oppure ne ha infinite. Risolvendo all'indietro vediamo che l'ultima equazione ci dà $z = -1$, e sostituendo nella seconda otteniamo $-3 = 6$: impossibile. Quindi questo sistema non ha soluzioni.

ESEMPIO 3.9 Consideriamo il seguente sistema di tre equazioni in tre incognite:

$$\begin{cases} 2x + 4y - 2z = 8, \\ 3z = -3, \\ 2z = -2. \end{cases}$$

La matrice dei coefficienti e il vettore dei termini noti sono

$$A = \begin{bmatrix} 2 & 4 & -2 \\ 0 & 0 & 3 \\ 0 & 0 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 8 \\ -3 \\ -2 \end{bmatrix}.$$

Dunque il sistema è di nuovo triangolare superiore con $a_{22} = 0$: quindi o non ha soluzioni o ne ha infinite. Risolvendo all'indietro troviamo $z = -1$: sostituendo nella seconda otteniamo $-3 = -3$, che è sempre vera e quindi non ci fornisce ulteriori informazioni. Sostituendo il valore di z nella prima equazione troviamo $x = 3 - 2y$. Dunque stavolta abbiamo una soluzione per ogni valore di y — ovvero (come nell'Esempio 3.4) una retta di soluzioni, di equazioni parametriche

$$\begin{cases} x = 3 - 2t, \\ y = t, \\ z = -1. \end{cases}$$

3.3 Il metodo d'eliminazione di Gauss

Dunque ora sappiamo perfettamente come risolvere i sistemi (quadrati) triangolari superiori. Il nostro prossimo obiettivo è trovare un modo per trasformare un qualunque sistema lineare quadrato in un sistema triangolare superiore che abbia esattamente le stesse soluzioni; vedremo nel Paragrafo 6.2 come comportarci per sistemi non quadrati (ma non sarà molto diverso).

Definizione 3.6 Due sistemi lineari (anche non quadrati) dello stesso ordine si dicono *equivalenti* se hanno esattamente le stesse soluzioni.

Chiaramente, due sistemi ottenuti l'uno dall'altro semplicemente cambiando l'ordine delle equazioni sono equivalenti. Lo scambiare due equazioni è la prima *operazione elementare* che possiamo effettuare su un sistema senza modificare le soluzioni.

Definizione 3.7 Date due equazioni

$$a_1x_1 + \dots + a_nx_n = a \quad \text{e} \quad b_1x_1 + \dots + b_nx_n = b,$$

si dice *combinazione lineare* delle due equazioni di coefficienti $h, k \in \mathbb{R}$ l'equazione

$$h(a_1x_1 + \dots + a_nx_n) + k(b_1x_1 + \dots + b_nx_n) = ha + kb.$$

I coefficienti di questa nuova equazione sono $ha_j + kb_j$, per $j = 1, \dots, n$.

Facendo combinazioni lineari di equazioni di un sistema le soluzioni non cambiano:

Lemma 3.2 Sia $Ax = b$ un sistema lineare contenente le equazioni

$$a_1x_1 + \dots + a_nx_n = a$$

e

$$b_1x_1 + \dots + b_nx_n = b. \quad (3.7)$$

Sia $\tilde{A}x = \tilde{b}$ il sistema lineare ottenuto sostituendo in $Ax = b$ l'equazione

$$h(a_1x_1 + \dots + a_nx_n) + k(b_1x_1 + \dots + b_nx_n) = ha + kb \quad (3.8)$$

al posto dell'equazione (3.7), dove $h, k \in \mathbb{R}$ sono due numeri reali con $k \neq 0$. Allora i sistemi $Ax = b$ e $\tilde{A}x = \tilde{b}$ sono equivalenti.

Dimostrazione. Sia (v_1, \dots, v_n) una soluzione di $Ax = b$. Questo vuol dire che i numeri v_1, \dots, v_n soddisfano tutte le equazioni del sistema; in particolare

$$\begin{cases} a_1v_1 + \dots + a_nv_n = a, \\ b_1v_1 + \dots + b_nv_n = b, \end{cases} \quad (3.9)$$

per cui (v_1, \dots, v_n) è chiaramente anche una soluzione di (3.8) — e quindi anche di $\tilde{A}x = \tilde{b}$, in quanto le altre equazioni non sono cambiate.

Viceversa, sia (v_1, \dots, v_n) una soluzione di $\tilde{A}x = \tilde{b}$. Allora la prima riga di (3.9) è ancora verificata; sostituendola in (3.8) troviamo

$$k(b_1v_1 + \dots + b_nv_n) = kb,$$

e, essendo $k \neq 0$, vediamo che v_1, \dots, v_n soddisfano anche (3.7) — e quindi sono una soluzione di $Ax = b$. \square

Dunque la sostituzione di un'equazione di un sistema con una sua combinazione lineare con un'altra equazione del sistema non cambia le soluzioni; è la seconda *operazione elementare* che possiamo effettuare su un sistema lineare per ottenere sistemi equivalenti.

Il *metodo di eliminazione di Gauss* è una procedura che trasforma, tramite operazioni elementari, un qualunque sistema lineare quadrato in un sistema triangolare superiore equivalente. Siccome quest'ultimo può venire risolto facilmente con una risoluzione all'indietro, questo metodo fornisce una procedura efficiente per risolvere i sistemi lineari quadrati.

Il metodo di eliminazione di Gauss consiste in diversi passi (per l'esattezza un meno dell'ordine del sistema). Il passo i -esimo annulla gli elementi sotto la diagonale principale della colonna i -esima della matrice dei coefficienti, producendo nel contempo un numero reale p_i , l' i -esimo *pivot* del sistema relativo alla data eliminazione di Gauss — e tutto senza cambiare le soluzioni del sistema.

Ma vediamo i dettagli. Partiamo da un sistema lineare quadrato $Ax = b$ di ordine n , dove

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \quad \text{e} \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

Consideriamo la prima colonna della matrice A . Se contiene solo zeri, poniamo $p_1 = 0$ e passiamo a considerare la seconda colonna. Se invece contiene qualche elemento non nullo, scambiamo se necessario la prima equazione con una delle successive in modo che a essere diverso da zero sia a_{11} , il primo elemento della diagonale principale. Poniamo $p_1 = a_{11}$, e sommiamo all'equazione j -esima (per $j = 2, \dots, n$) la prima equazione moltiplicata per $-a_{j1}/p_1$. In questo modo otteniamo un sistema equivalente (grazie al Lemma 3.2) in cui gli elementi sotto la diagonale principale della prima colonna della matrice dei coefficienti sono tutti nulli.

I passi successivi sono molto simili. Supponiamo di aver già trattato le prime $i-1$ colonne del sistema, e consideriamo la colonna i -esima (cominciamo cioè il passo i -esimo). Se A^i contiene solo zeri dalla riga i -esima compresa in giù poniamo $p_i = 0$ e passiamo alla colonna successiva. Altrimenti, a meno di scambiare l' i -esima equazione con una sottostante, possiamo supporre che a_{ii} sia diverso da zero. Poniamo $p_i = a_{ii}$, e sommiamo alla j -esima equazione (per $j = i+1, \dots, n$) l'equazione i -esima moltiplicata per $-a_{ji}/p_i$. In questo modo otteniamo un sistema equivalente in cui gli elementi sotto la diagonale principale delle prime i colonne della matrice dei coefficienti sono tutti nulli.

Procedendo in questo modo giungiamo all'ultima colonna; ponendo infine $p_n = a_{nn}$ abbiamo ottenuto un sistema triangolare superiore, equivalente al sistema di partenza, con sulla diagonale principale gli n pivot p_1, \dots, p_n .

Per capire meglio come funziona il metodo d'eliminazione di Gauss vediamo in azione su alcuni esempi.

ESEMPIO 3.10 Consideriamo il sistema quadrato di ordine 4

$$\begin{cases} x + 3y + z - w = 1, \\ 3x + 9y + 4z + w = 1, \\ 2x + y + 5z + 2w = 0, \\ y - z - w = 2. \end{cases} \quad (3.10)$$

La matrice dei coefficienti e il vettore dei termini noti sono dati da

$$A = \begin{pmatrix} 1 & 3 & 1 & -1 \\ 3 & 9 & 4 & 1 \\ 2 & 1 & 5 & 2 \\ 0 & 1 & -1 & -1 \end{pmatrix} \quad \text{e} \quad b = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 2 \end{pmatrix}.$$

Per semplicità di scrittura, eseguiremo le operazioni previste dall'eliminazione di Gauss direttamente sulle righe della matrice e del vettore dei termini noti, senza riscrivere ogni volta le equazioni. Siccome il primo elemento in alto a sinistra della matrice A è diverso da zero, troviamo subito $p_1 = 1$ senza bisogno di scambi di righe. Poi dobbiamo sottrarre alla seconda riga 3 volte la prima, alla terza 2 volte la prima e lasciare la quarta invariata. Quindi abbiamo effettuato la trasformazione seguente:

$$\left| \begin{array}{cccc|c} 1 & 3 & 1 & -1 & 1 \\ 3 & 9 & 4 & 1 & 1 \\ 2 & 1 & 5 & 2 & 0 \\ 0 & 1 & -1 & -1 & 2 \end{array} \right| \longrightarrow \left| \begin{array}{cccc|c} 1 & 3 & 1 & -1 & 1 \\ 0 & 0 & 1 & 4 & -2 \\ 0 & -5 & 3 & 4 & -2 \\ 0 & 1 & -1 & -1 & 2 \end{array} \right|,$$

dove per semplicità il vettore dei termini noti è scritto subito a destra della matrice dei coefficienti. Ora, nella seconda colonna sotto la seconda riga ci sono dei termini non nulli ma l'elemento sulla diagonale principale è zero; quindi dobbiamo effettuare uno scambio di righe. Qui abbiamo una *scelta*: possiamo portare al secondo posto la terza riga oppure la quarta. Supponiamo di portarci la terza; otteniamo

$$\left| \begin{array}{cccc|c} 1 & 3 & 1 & -1 & 1 \\ 0 & 0 & 1 & 4 & -2 \\ 0 & -5 & 3 & 4 & -2 \\ 0 & 1 & -1 & -1 & 2 \end{array} \right| \longrightarrow \left| \begin{array}{cccc|c} 1 & 3 & 1 & -1 & 1 \\ 0 & -5 & 3 & 4 & -2 \\ 0 & 0 & 1 & 4 & -2 \\ 0 & 1 & -1 & -1 & 2 \end{array} \right|,$$

e il secondo pivot è $p_2 = -5$. A questo punto dobbiamo lasciare la terza riga invariata e sommare alla quarta $1/5$ della seconda, in modo da ottenere

$$\left| \begin{array}{cccc|c} 1 & 3 & 1 & -1 & 1 \\ 0 & -5 & 3 & 4 & -2 \\ 0 & 0 & 1 & 4 & -2 \\ 0 & 1 & -1 & -1 & 2 \end{array} \right| \longrightarrow \left| \begin{array}{cccc|c} 1 & 3 & 1 & -1 & 1 \\ 0 & -5 & 3 & 4 & -2 \\ 0 & 0 & 1 & 4 & -2 \\ 0 & 0 & -2/5 & -1/5 & 8/5 \end{array} \right|$$

Siamo al passo 3. Non servono altri scambi di righe; quindi il terzo pivot è $p_3 = 1$, e dobbiamo semplicemente sommare $2/5$ della terza riga alla quarta, ottenendo

$$\left| \begin{array}{cccc|c} 1 & 3 & 1 & -1 & 1 \\ 0 & -5 & 3 & 4 & -2 \\ 0 & 0 & 1 & 4 & -2 \\ 0 & 0 & -2/5 & -1/5 & 8/5 \end{array} \right| \longrightarrow \left| \begin{array}{cccc|c} 1 & 3 & 1 & -1 & 1 \\ 0 & -5 & 3 & 4 & -2 \\ 0 & 0 & 1 & 4 & -2 \\ 0 & 0 & 0 & 7/5 & 4/5 \end{array} \right|.$$

Il quarto pivot è $p_4 = 7/5$. In conclusione, il nostro sistema è equivalente al sistema triangolare superiore

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

Risolviendo all'indietro otteniamo

$$\begin{cases} x = 11, \\ y = -12/7, \\ z = -30/7, \\ w = 4/7, \end{cases}$$

che è la soluzione (unica) del sistema (3.10) di partenza.

ESEMPIO 3.11 Proviamo a vedere cosa sarebbe successo effettuando al secondo passo l'altro scambio di righe possibile. Scambiando la quarta riga con la seconda otteniamo

$$\left| \begin{array}{cccc|c} 1 & 3 & 1 & -1 & 1 \\ 0 & 0 & 1 & 4 & -2 \\ 0 & -5 & 3 & 4 & -2 \\ 0 & 1 & -1 & -1 & 2 \end{array} \right| \longrightarrow \left| \begin{array}{cccc|c} 1 & 3 & 1 & -1 & 1 \\ 0 & 1 & -1 & -1 & 2 \\ 0 & -5 & 3 & 4 & -2 \\ 0 & 0 & 1 & 4 & -2 \end{array} \right|,$$

e il secondo pivot stavolta è $p_2 = 1$. A questo punto dobbiamo sommare 5 volte la seconda riga alla terza e lasciare la quarta invariata, in modo da ottenere

$$\left| \begin{array}{cccc|c} 1 & 3 & 1 & -1 & 1 \\ 0 & 1 & -1 & -1 & 2 \\ 0 & -5 & 3 & 4 & -2 \\ 0 & 0 & 1 & 4 & -2 \end{array} \right| \longrightarrow \left| \begin{array}{cccc|c} 1 & 3 & 1 & -1 & 1 \\ 0 & 1 & -1 & -1 & 2 \\ 0 & 0 & -2 & -1 & 8 \\ 0 & 0 & 1 & 4 & -2 \end{array} \right|.$$

Siamo al passo 3. Non servono altri scambi di righe; quindi il terzo pivot è $p_3 = -2$, e dobbiamo semplicemente sommare $1/2$ della terza riga alla quarta, ottenendo

$$\left| \begin{array}{cccc|c} 1 & 3 & 1 & -1 & 1 \\ 0 & 1 & -1 & -1 & 2 \\ 0 & 0 & -2 & -1 & 8 \\ 0 & 0 & 1 & 4 & -2 \end{array} \right| \longrightarrow \left| \begin{array}{cccc|c} 1 & 3 & 1 & -1 & 1 \\ 0 & 1 & -1 & -1 & 2 \\ 0 & 0 & -2 & -1 & 8 \\ 0 & 0 & 0 & 7/2 & 2 \end{array} \right|$$

Il quarto pivot è $p_4 = 7/2$. Dunque il nostro sistema è equivalente anche al sistema

$$\begin{cases} x + 3y + z - w = 1, \\ y - z - w = 2, \\ -2z - w = 8, \\ \frac{7}{2}w = 2, \end{cases}$$

e infatti risolvendo all'indietro otteniamo di nuovo

$$\begin{cases} x = 11, \\ y = -12/7, \\ z = -30/7, \\ w = 4/7. \end{cases}$$

Osservazione 3.6 Nell'eliminazione di Gauss il Lemma 3.2 non viene usato a piena potenza. Infatti, ogni volta che un'equazione viene sostituita da una combinazione lineare il coefficiente dell'equazione sostituita (il k del Lemma 3.2) è sempre uguale a 1. Se il nostro obiettivo è semplicemente risolvere il sistema, non c'è motivo di limitarsi in questo modo: possiamo usare combinazioni lineari con $k \neq 0$ qualunque, e risolvere comunque il sistema (vedi il prossimo Esempio). In seguito, però, dovremo applicare il metodo di eliminazione di Gauss anche per altri scopi (per esempio per calcolare il determinante di una matrice quadrata), e in quelle situazioni sarà fondamentale usare $k = 1$ e non un k qualunque.

ESEMPIO 3.12 Riprendiamo di nuovo il sistema (3.10), e iniziamo l'eliminazione di Gauss come visto nell'Esempio 3.10. Procediamo come prima fino a ricavare il secondo pivot $p_2 = -5$. A questo punto invece di sommare alla quarta riga $1/5$ della seconda, ci discostiamo dall'eliminazione di Gauss usuale sostituendo alla quarta riga una combinazione lineare fatta dalla seconda riga più cinque volte la quarta (quindi $k = 5$ nel Lemma 3.2). Otteniamo:

$$\left| \begin{array}{cccc|c} 1 & 3 & 1 & -1 & 1 \\ 0 & -5 & 3 & 4 & -2 \\ 0 & 0 & 1 & 4 & -2 \\ 0 & 1 & -1 & -1 & 2 \end{array} \right| \longrightarrow \left| \begin{array}{cccc|c} 1 & 3 & 1 & -1 & 1 \\ 0 & -5 & 3 & 4 & -2 \\ 0 & 0 & 1 & 4 & -2 \\ 0 & 0 & -2 & -1 & 8 \end{array} \right|$$

Per concludere ora basta sommare il doppio della terza riga alla quarta ottenendo

$$\left| \begin{array}{cccc|c} 1 & 3 & 1 & -1 & 1 \\ 0 & -5 & 3 & 4 & -2 \\ 0 & 0 & 1 & 4 & -2 \\ 0 & 0 & -2 & -1 & 8 \end{array} \right| \longrightarrow \left| \begin{array}{cccc|c} 1 & 3 & 1 & -1 & 1 \\ 0 & -5 & 3 & 4 & -2 \\ 0 & 0 & 1 & 4 & -2 \\ 0 & 0 & 0 & 7 & 4 \end{array} \right|$$

Dunque il sistema (3.10) è equivalente anche al sistema triangolare superiore

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

Procedendo in questo modo abbiamo semplificato i calcoli e ottenuto comunque un sistema triangolare superiore equivalente; ma, come vedremo nel Capitolo 9, abbiamo modificato delle caratteristiche importanti della matrice dei coefficienti (quali, per esempio, il determinante), e comunque si rischia di sbagliare (vedi l'Esercizio 3.18).

Dopo aver visto come si applica in pratica il metodo di eliminazione di Gauss, possiamo discuterlo. Prima di tutto, è chiaro che i pivot e le varie operazioni effettuate dipendono solo dalla matrice dei coefficienti; in particolare, quindi, potremo applicare l'eliminazione di Gauss semplicemente a una matrice quadrata e parlare di pivot di una matrice quadrata.

In secondo luogo, il sistema triangolare superiore equivalente che troviamo alla fine dell'eliminazione di Gauss ha sulla diagonale principale esattamente i pivot. Quindi la Proposizione 3.1 ci fornisce un criterio per determinare quando un sistema lineare quadrato ammette un'unica soluzione:

Teorema 3.3 *Un sistema lineare quadrato ammette un'unica soluzione se e solo se i pivot della sua matrice dei coefficienti sono tutti non nulli.*

Dimostrazione. Infatti il sistema ammette un'unica soluzione se e solo se il sistema triangolare superiore a esso equivalente ottenuto tramite un'eliminazione di Gauss ammette un'unica soluzione, per cui la tesi segue dalla Proposizione 3.1. \square

In particolare, quindi, l'unicità della soluzione dipende solo dalla matrice dei coefficienti, come avevamo suggerito nell'Osservazione 3.1.

Osservazione 3.7 Gli esempi precedenti mostrano che i pivot di una matrice quadrata non sono univocamente determinati; i pivot cambiano a seconda di quali scambi di riga vengono effettuati. Comunque, non sono completamente arbitrari; infatti, una conseguenza del Teorema 3.3 è che se i pivot ottenuti con un'eliminazione di Gauss sono tutti non nulli allora anche i pivot ottenuti con un'altra eliminazione di Gauss lo sono (perché?). In realtà, come vedremo nel Capitolo 9, vale qualcosa di più: il valore assoluto del prodotto di tutti i pivot è indipendente dall'eliminazione di Gauss effettuata (risulterà essere il valore assoluto del determinante della matrice).

Questo ci porta alla seguente definizione.

Definizione 3.8 Una matrice quadrata è detta *non singolare* se tutti i suoi pivot (rispetto a una qualunque eliminazione di Gauss) sono non nulli; è detta *singolare* altrimenti.

Quindi il Teorema 3.3 si può anche esprimere dicendo che *un sistema quadrato ammette soluzione unica se e solo se la matrice dei coefficienti è non singolare.*

Osservazione 3.8 Un'altra conseguenza del metodo d'eliminazione di Gauss è che ogni sistema lineare quadrato è equivalente a un sistema triangolare superiore. Siccome sappiamo (Proposizione 3.1 ed Esercizio 3.3) che un sistema triangolare superiore o ha soluzione unica, o non ne ha o ne ha infinite, questo accade per qualunque sistema lineare quadrato, rispondendo così al problema che ci eravamo posti all'inizio. Nel Capitolo 6 saremo anche più precisi, descrivendo alla perfezione la struttura dell'insieme delle soluzioni.

Esercizi

3.1 Studia (cioè vedi se ha soluzioni, e in tal caso trovale) il sistema

$$\begin{cases} x_1 - 2x_2 + x_3 = 1, \\ x_1 - x_2 + 4x_3 = 7, \\ x_1 - 2x_2 + 2x_3 = 3. \end{cases}$$

3.2 La seconda parte della dimostrazione della Proposizione 3.1 si poteva condurre anche per assurdo. Supponiamo che il sistema abbia un'unica soluzione (v_1, \dots, v_n) , e supponiamo, per assurdo, che ci sia almeno un coefficiente nullo sulla diagonale principale. Scegliamo k di nuovo in modo che $a_{kk} = 0$ ma $a_{11}, \dots, a_{k-1, k-1} \neq 0$. Sostituendo v_{k+1}, \dots, v_n al posto di x_{k+1}, \dots, x_n nella k -esima equazione, troviamo almeno due soluzioni (in realtà infinite): v_k e $w \neq v_k$. Allora procedendo come prima troviamo almeno due soluzioni distinte del nostro sistema, mentre avevamo supposto ne avesse solo una, contraddizione. Completa i dettagli di questa dimostrazione.

3.3 Dimostra che se un sistema triangolare superiore ammette due soluzioni distinte allora ne ammette infinite.

3.4 Risolvi con l'eliminazione di Gauss i seguenti sistemi:

$$\begin{cases} 2x_2 - 4x_3 + x_4 = 1, \\ x_1 - 3x_2 - x_3 + x_4 = 0, \\ x_1 - x_2 + 4x_3 - 2x_4 = -1, \\ 2x_1 - 2x_2 - x_3 + 2x_4 = 0; \end{cases} \quad \begin{cases} 5x_1 + 3x_2 - 2x_3 = 1, \\ x_2 - 2x_3 = -2, \\ x_2 + 2x_3 = 1. \end{cases}$$

3.5 Risolvi con l'eliminazione di Gauss i seguenti sistemi:

$$\begin{cases} x + y + z + t = 1, \\ 8x + 4y + 2z + t = 5, \\ 27x + 9y + 3z + t = 14, \\ 64x + 16y + 4z + t = 30; \end{cases} \quad \begin{cases} 9x + y + z + t = 3, \\ 8x + 2y + z + t = 1, \\ 6x + 7y + 3t = 1, \\ 6x + 5y + t = 1. \end{cases}$$

3.6 Studia i seguenti sistemi lineari:

$$\begin{cases} 19x - y + 5z + t = 3, \\ 18x + 5z + t = 1, \\ 6x + 9y + t = 1, \\ 12x + 18y + 3t = 3, \end{cases} \quad \begin{cases} 14x + 3z + t = 3, \\ 13x + y + 3z + t = 1, \\ 6x + 8y + 2t = 1, \\ 6x + 7y + t = 1. \end{cases}$$

3.7 Studia i seguenti sistemi al variare del parametro $k \in \mathbb{R}$:

$$\begin{cases} 2x_1 + kx_2 = 2, \\ kx_1 + 2x_2 = k, \\ kx_2 + kx_3 = k; \end{cases} \quad \begin{cases} 3x + 2y + kz = 11, \\ 2x - 6y - 3z = 0, \\ kx + 4y + 2x = 7. \end{cases}$$

3.8 Determina se le seguenti matrici sono singolari o meno:

$$\begin{vmatrix} 1 & 0 & 3 & 7 & 4 \\ -1 & 1 & 0 & 2 & -1 \\ 2 & 2 & 0 & 2 & 2 \\ 3 & 1 & 4 & -1 & 7 \\ 0 & 1 & 3 & 0 & 3 \end{vmatrix}, \quad \begin{vmatrix} 1 & -1 & 2 & -2 \\ 2 & -2 & 1 & -1 \\ 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 \end{vmatrix}$$

3.9 Studia il sistema seguente al variare del parametro $a \in \mathbb{R}$:

$$\begin{cases} y + z = a, \\ 2x + 3y + 7z = 5, \\ x - 3y - z = -2. \end{cases}$$

3.10 Studia il sistema seguente al variare dei parametri $h, b \in \mathbb{R}$:

$$\begin{cases} x + 2y + hz = 1, \\ 2x + hy + 8z = -1, \\ 4x + 7y + z = b. \end{cases}$$

3.11 Scrivi un sistema lineare di tre equazioni in tre incognite, con un'equazione dipendente da un parametro k , in modo che se $k \neq 0$ allora il sistema ammette come unica soluzione la terna $(1, 0, 1)$, mentre se $k = 0$ il sistema ammette una retta di soluzioni.

3.12 Trova che relazione deve intercorrere fra $a, b, c \in \mathbb{R}^3$ perché il sistema

$$\begin{cases} 2x + y = a, \\ 2x + z = b, \\ 4x + y + z = c, \end{cases}$$

ammetta soluzione.

3.13 Calcola i pivot della matrice

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

effettuando tutti gli scambi di riga possibili, e verifica che il valore assoluto del prodotto dei pivot non cambia. Quali combinazioni di scambi di righe danno come prodotto dei pivot l'opposto di quello ottenuto effettuando un'eliminazione di Gauss senza scambi di righe?

3.14 Trova per quali valori del parametro $k \in \mathbb{R}$ la matrice

$$\begin{vmatrix} 1 & -2 & 3 & 1 \\ -1 & 0 & -1 & 0 \\ 2 & 1 & 3 & 0 \\ -2 & -4 & -4 & k \end{vmatrix}$$

è singolare.

3.15 Fissato un sistema di riferimento affine $RA(O, \vec{i}, \vec{j}, \vec{k})$ nello spazio, sia r la retta

passante per i punti di coordinate $\begin{vmatrix} 2 \\ 0 \\ 1 \end{vmatrix}$ e $\begin{vmatrix} -2 \\ 6 \\ 3 \end{vmatrix}$, e π il piano di coordinate parametriche

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

Dimostra che r e π si intersecano, e trova il punto d'intersezione.

3.16 Studia il seguente sistema lineare:

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

3.17 Sia $Ax = b$ un sistema lineare quadrato con matrice A non singolare. Dimostra che il sistema è equivalente a un sistema lineare quadrato con matrice dei coefficienti diagonale.

3.18 Sia $A \in M_{3,3}(\mathbb{R})$ la matrice

$$A = \begin{vmatrix} \beta & 0 & -1 \\ 1 & 1 & -1 \\ -1 & 1 & -1 \end{vmatrix},$$

dove $\beta \in \mathbb{R}$. Dimostra prima di tutto che A è non singolare qualunque sia $\beta \in \mathbb{R}$; in particolare, il sistema $Ax = O$ con termini noti tutti nulli ha sempre come unica soluzione la terna banale $(0, 0, 0)$. Poi considera la seguente "eliminazione di Gauss generalizzata" (nel senso dell'Osservazione 3.6 e dell'Esempio 3.12):

$$\begin{vmatrix} \beta & 0 & -1 \\ 1 & 1 & -1 \\ -1 & 1 & -1 \end{vmatrix} \longrightarrow \begin{vmatrix} \beta & 0 & -1 \\ 0 & \beta & -\beta + 1 \\ 0 & \beta & -\beta - 1 \end{vmatrix} \longrightarrow \begin{vmatrix} \beta & 0 & -1 \\ 0 & \beta & -\beta + 1 \\ 0 & 0 & -2 \end{vmatrix} = B,$$

dove nel primo passo abbiamo sostituito alla riga A_2 la combinazione lineare $\beta A_2 - A_1$ e alla riga A_3 la combinazione lineare $\beta A_3 + A_1$, mentre nel secondo passo abbiamo sottratto la seconda riga dalla terza. Per il Lemma 3.2, il sistema $Bx = O$ è equivalente al sistema $Ax = O$; in particolare, dovrebbe avere sempre come unica soluzione la terna $(0, 0, 0)$. Sfortunatamente, per $\beta = 0$ tutte le terne della forma $(x_1, x_2, 0)$ sono soluzione di $Bx = O$. Dov'è l'errore?

COMPLEMENTI

3C.1 Il principio d'induzione

Nell'Osservazione 3.5 abbiamo introdotto un metodo molto comodo per dimostrare affermazioni sui numeri naturali, il *principio d'induzione*. In questo paragrafo lo esamineremo più in dettaglio, descrivendone alcune forme equivalenti e dando qualche esempio d'applicazione.

Così come l'abbiamo incontrato noi, il principio d'induzione può venire espresso in questo modo:

Principio d'induzione 1 Sia \mathcal{A}_n un'affermazione dipendente da un numero naturale n . Supponiamo che

- (a) \mathcal{A}_0 sia vera;
 - (b) se \mathcal{A}_{n-1} è vera, allora anche \mathcal{A}_n lo è.
- Allora \mathcal{A}_n è vera per ogni numero naturale n .

Questa però non è l'unica forma in cui si usa il principio d'induzione. Un'altra possibilità è la seguente:

Principio d'induzione 2 Sia \mathcal{A}_n un'affermazione dipendente da un numero naturale n . Supponiamo che

- (a) \mathcal{A}_0 sia vera;
 - (b) se \mathcal{A}_k è vera per ogni $0 \leq k \leq n$, allora anche \mathcal{A}_{n+1} è vera.
- Allora \mathcal{A}_n è vera per ogni numero naturale n .

Vi è anche un terzo principio che può essere utile:

Principio del buon ordinamento Ogni sottoinsieme non vuoto di \mathbb{N} possiede un elemento minimo.

Cosa c'entra quest'ultima affermazione col principio d'induzione? Ebbene: sono equivalenti, come dimostrato nel

Teorema 3C.1 Le due forme del principio d'induzione e il principio del buon ordinamento sono tutte affermazioni equivalenti.

Dimostrazione. Cominciamo col dimostrare che il principio del buon ordinamento implica la prima forma del principio d'induzione. Sia allora \mathcal{A}_n un'affermazione che soddisfa le ipotesi del principio d'induzione 1, e consideriamo l'insieme

$$T = \{n \in \mathbb{N} \mid \mathcal{A}_n \text{ è falsa}\};$$

vogliamo dimostrare che $T = \emptyset$. Supponiamo, per assurdo, che T sia non vuoto; allora per il principio del buon ordinamento deve avere un elemento minimo m . Siccome sappiamo per ipotesi che \mathcal{A}_0 è vera, lo zero non appartiene a T ; quindi $m > 0$. Dunque $m - 1$ è un numero naturale e $m - 1 \notin T$ (in quanto m era l'elemento *minimo* di T), cioè \mathcal{A}_{m-1} è vera. Ma allora anche \mathcal{A}_m è vera, e quindi $m \notin T$, contraddizione. Dunque siamo costretti a concludere che $T = \emptyset$, cioè che \mathcal{A}_n è vera per ogni $n \in \mathbb{N}$.

Ora dimostriamo che la prima forma del principio d'induzione implica la seconda forma del principio d'induzione. Supponiamo allora vero il principio d'induzione 1, e sia \mathcal{A}_n un'affermazione che soddisfa le ipotesi del principio d'induzione 2. Indichiamo con \mathcal{B}_n l'affermazione " \mathcal{A}_k è vera per $0 \leq k \leq n$ ". Chiaramente, \mathcal{B}_n soddisfa le ipotesi della prima forma del principio d'induzione, per cui \mathcal{B}_n è vera per ogni $n \in \mathbb{N}$. Ma allora in particolare \mathcal{A}_n è vera per ogni $n \in \mathbb{N}$, e quindi abbiamo dimostrato la seconda forma del principio d'induzione.

Infine, dimostriamo che il principio d'induzione 2 implica il principio del buon ordinamento. Supponiamo vero il secondo principio d'induzione, e sia $T \subseteq \mathbb{N}$ un sottoinsieme dei numeri naturali privo di minimo; dobbiamo dimostrare che $T = \emptyset$. Indichiamo con \mathcal{A}_n l'affermazione " $n \notin T$ ". Chiaramente, \mathcal{A}_0 è vera, perché altrimenti 0 sarebbe l'elemento minimo di T . Supponiamo che \mathcal{A}_k sia vera per $0 \leq k \leq n$; questo vuol dire che $0, \dots, n \notin T$. Ma allora $n + 1 \notin T$, in quanto altrimenti $n + 1$ sarebbe l'elemento minimo di T , e quindi \mathcal{A}_{n+1} è vera. Per il principio d'induzione 2, l'affermazione \mathcal{A}_n è vera per ogni $n \in \mathbb{N}$, cioè $n \notin T$ per ogni $n \in \mathbb{N}$, e quindi $T = \emptyset$. \square

In questo libro useremo più volte il principio d'induzione, nelle tre forme appena indicate oppure in altre ancora. Infatti, vi sono ulteriori varianti del principio d'induzione, tutte più o meno ovvie. Qui di seguito ne troverai un'altra; la verifica che è equivalente a quelle già viste e l'ideazione di altre varianti sono lasciate per esercizio.

Principio d'induzione 3 Sia \mathcal{A}_n un'affermazione dipendente da un numero naturale n , e fissiamo $s \in \mathbb{N}$. Supponiamo che

(a) \mathcal{A}_s sia vera;

(b) se \mathcal{A}_n (con $n \geq s$) è vera, allora anche \mathcal{A}_{n+1} lo è.

Allora \mathcal{A}_n è vera per ogni numero naturale $n \geq s$.

Osservazione 3C.1 Tanto per evitare equivoci: il principio d'induzione è una proprietà dell'insieme \mathbb{N} dei numeri naturali che noi *assumiamo* vera senza dimostrazione². Il Teorema 3C.1 non dimostra il principio d'induzione, ma si limita a far vedere come diversi modi di enunciare questo principio sono equivalenti.

² Infatti, è uno degli assiomi che possono venire utilizzati per definire i numeri naturali.

Vediamo ora alcune applicazioni classiche del principio d'induzione.

ESEMPIO 3C.1 Vogliamo dimostrare che "La somma degli angoli interni di un poligono convesso di $n + 3$ lati è $(n + 1)\pi$ radianti". Utilizziamo il principio d'induzione 1. L'affermazione \mathcal{A}_0 è un teorema classico della geometria euclidea (Teorema 1.1). Supponiamo allora \mathcal{A}_{n-1} vera (cioè che la somma degli angoli interni di un poligono convesso di esattamente $n + 2$ lati sia $n\pi$ radianti) e consideriamo un poligono convesso P di $n + 3$ lati. Collegando due vertici consecutivi con una diagonale, dividiamo P in un triangolo (la cui somma degli angoli interni è π) e un poligono convesso P' di $n + 2$ lati. Per ipotesi induttiva (cioè per aver supposto \mathcal{A}_{n-1} vera) la somma degli angoli interni di P' è $n\pi$; quindi la somma degli angoli interni di P è $\pi + n\pi = (n + 1)\pi$. Dunque abbiamo dimostrato \mathcal{A}_n supponendo vera \mathcal{A}_{n-1} . Siccome \mathcal{A}_0 è vera, per il principio d'induzione abbiamo dimostrato il nostro Teorema.

ESEMPIO 3C.2 Vogliamo trovare la somma $s_n = 1 + 2 + \dots + n$ dei primi n numeri naturali. Il modo più semplice per trovarla è notare che scrivendo

$$s_n = 1 + \dots + n,$$

$$s_n = n + \dots + 1,$$

e sommando membro a membro a sinistra otteniamo $2s_n$ mentre a destra otteniamo $n + 1$ ripetuto n volte, cioè $n(n + 1)$. Dunque $2s_n = n(n + 1)$, e quindi

$$1 + \dots + n = \frac{n(n + 1)}{2}. \quad (3C.1)$$

Volendo invece dimostrare la formula (3C.1) per induzione, dobbiamo procedere come segue: prima di tutto dobbiamo verificarla per $n = 1$ (e infatti $1 = 1(1 + 1)/2$). Poi supponiamo sia vera per n ; allora

$$s_n = s_{n-1} + n = \frac{(n-1)n}{2} + n = \frac{(n-1)n + 2n}{2} = \frac{n(n+1)}{2},$$

per cui abbiamo dimostrato la formula anche per n , e quindi è sempre vera.

Osservazione 3C.2 Vi è una differenza fondamentale fra le due dimostrazioni nel precedente Esempio. Nella prima dimostrazione non avevamo idea di quanto facesse s_n , per cui abbiamo proceduto alla cieca sperando di trovare un risultato. Nella seconda dimostrazione, invece, avevamo un'idea su quanto valesse s_n , avevamo cioè una *congettura*: la formula (3C.1). Il principio d'induzione ci ha permesso di dimostrare questa congettura, ma non suggerisce come trovarla.

ESEMPIO 3C.3 Vogliamo calcolare la somma $q_n = 1^2 + 2^2 + \dots + n^2$ dei quadrati dei primi n numeri naturali. Stavolta la prima tecnica vista nel precedente Esempio non funziona (controlla); quindi ci rimane la seconda. Ma per procedere per induzione occorre una congettura. Ora, la somma dei primi n numeri naturali era espressa da un polinomio di secondo grado in n ; forse la somma dei primi n quadrati sarà

espressa da un polinomio di terzo grado. In altre parole, q_n potrebbe essere della forma $an^3 + bn^2 + cn + d$, per opportuni coefficienti $a, b, c, d \in \mathbb{R}$. Per trovare i valori di questi coefficienti vediamo cosa succede per n piccolo. Per $n = 1$ abbiamo $q_1 = 1$, per cui dev'essere $a + b + c + d = 1$. Per $n = 2$ abbiamo $q_2 = 5$, per cui dev'essere $8a + 4b + 2c + d = 5$. Controllando cosa succede per $n = 3$ e $n = 4$ giungiamo al sistema

$$\begin{cases} a + b + c + d = 1, \\ 8a + 4b + 2c + d = 5, \\ 27a + 9b + 3c + d = 14, \\ 64a + 16b + 4c + d = 30; \end{cases}$$

risolvendolo con i metodi di questo capitolo (vedi l'Esercizio 3.5) troviamo $a = 1/3$, $b = 1/2$, $c = 1/6$ e $d = 0$. Dunque la nostra congettura (dimostrata finora solo per $n \leq 4$) è

$$q_n = \frac{2n^3 + 3n^2 + n}{6} = \frac{n(n+1)(2n+1)}{6}. \quad (3C.2)$$

Vogliamo dimostrarla per induzione. Per $n = 1$ l'abbiamo già vista; supponiamola vera per $n - 1$. Allora

$$\begin{aligned} q_n &= q_{n-1} + n^2 = \frac{(n-1)n(2n-1)}{6} + n^2 \\ &= \frac{2n^3 - 3n^2 + n + 6n^2}{6} = \frac{2n^3 + 3n^2 + n}{6}, \end{aligned}$$

ed è fatta.

Infine, le dimostrazioni per induzione talvolta nascondono qualche pericolo, come mostra l'Esempio seguente.

ESEMPIO 3C.4 Vogliamo dimostrare che tutte le ragazze della Terra sono brune. Cominciamo col dimostrare che la seguente affermazione \mathcal{A}_n è vera: "Sia A un qualunque insieme di n ragazze. Allora tutte le ragazze in A hanno lo stesso colore di capelli". Per $n = 1$, in A c'è una sola ragazza, per cui l'affermazione è ovviamente vera. Supponiamo che la nostra affermazione sia vera per ogni insieme di n ragazze, e consideriamo un insieme A di $n+1$ ragazze. Se da A togliamo una ragazza, otteniamo un insieme A' di n ragazze; per l'ipotesi induttiva, tutte le ragazze di A' hanno lo stesso colore di capelli. Ora mettiamo la ragazza che abbiamo escluso prima al posto di un'altra di A' ; otteniamo un altro insieme di n ragazze, di nuovo quindi tutte con lo stesso colore di capelli, colore che chiaramente è lo stesso di prima. Di conseguenza tutte le ragazze dell'insieme A hanno lo stesso colore di capelli, e quindi abbiamo dimostrato la nostra affermazione per $n+1$ ragazze. Il principio d'induzione ci dice allora che \mathcal{A}_n è vera per ogni n ; in particolare, se n_0 è il numero (finito, disgraziatamente) delle ragazze di tutta la Terra, \mathcal{A}_{n_0} è vera. Questo vuol dire che tutte le ragazze della Terra hanno lo stesso colore di capelli; siccome mia moglie è bruna ed è una ragazza, ne segue che tutte le ragazze della Terra sono brune. E le bionde? Evidentemente non sono ragazze... Dov'è l'errore?

Spazi vettoriali

È finalmente giunto il momento di introdurre ufficialmente gli spazi vettoriali, la struttura principe dell'Algebra Lineare, di cui \mathcal{V}_O^2 , \mathcal{V}_O^3 e l'insieme delle soluzioni di un sistema lineare omogeneo (cioè con tutti i termini noti uguali a zero) sono solo i primi esempi. Parleremo di spazi e sottospazi vettoriali, di dipendenza e indipendenza lineare, di basi e dimensione, di intersezione e somma, senza trascurare alcune conseguenze riguardanti i sistemi lineari. Infine, i Complementi contengono un'introduzione elementare al Lemma di Zorn e alla teoria dei cardinali infiniti, due degli argomenti più utili della teoria degli insiemi, con applicazioni agli spazi vettoriali di dimensione infinita.

4.1 Spazi e sottospazi

Come visto nel capitolo precedente, le soluzioni di un sistema lineare con n incognite sono liste di n numeri reali: quei numeri che, sostituiti nell'ordine alle incognite, soddisfano tutte le equazioni del sistema. L'insieme delle liste di n numeri reali verrà indicato con \mathbb{R}^n : in simboli,

$$\mathbb{R}^n = \left\{ v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \mid v_1, \dots, v_n \in \mathbb{R} \right\}.$$

Ricordando quanto visto nel Paragrafo 2.2, dove a ogni punto del piano veniva associata una coppia di numeri reali e a ogni punto dello spazio una terna di numeri reali, l'insieme \mathbb{R}^n può venire considerato come uno spazio geometrico n -dimensionale¹, in cui ogni punto è rappresentato da una n -upla di numeri reali. Gli elementi (che d'ora in poi chiameremo anche *vettori*) di \mathbb{R}^n verranno scritti quasi sempre come colonne piuttosto che come righe di numeri, per motivi che vedremo più avanti (essenzialmente per poterli moltiplicare a sinistra per una matrice). Infine, è chiaro che l'insieme delle

¹ Ma vedi il Paragrafo 4.4 per il significato della parola "dimensione" in questo contesto.

n -uple di numeri razionali verrà indicato con \mathbb{Q}^n , e più in generale l'insieme delle n -uple di elementi di un campo \mathbb{K} verrà indicato con \mathbb{K}^n .

Osservazione 4.1 A questo punto forse ti starai chiedendo perché mai occuparsi di spazi n -dimensionali in un testo di Geometria, visto che lo spazio in cui viviamo è (almeno a prima vista) tridimensionale. Come prima risposta si potrebbe obiettare che in realtà il nostro spazio è per lo meno a quattro dimensioni, in quanto bisogna considerare anche il tempo²; ma il motivo vero, e più profondo, è un altro. Certo, il nostro spazio è tridimensionale; ma per rappresentare sensatamente il comportamento di un qualunque oggetto tre numeri non bastano. Prendiamo per esempio un banalissimo punto materiale in movimento: per descriverlo abbiamo bisogno sia della sua posizione che della sua velocità, per un totale di 6 coordinate — e quindi siamo in \mathbb{R}^6 . Per descrivere una squadra di calcio di punti materiali allora dobbiamo lavorare almeno in \mathbb{R}^{66} — o in \mathbb{R}^{72} se vogliamo tenere presente anche il pallone. E se poi al posto dei punti materiali consideriamo i calciatori in carne e ossa, con tutte quelle giunture e articolazioni indipendenti, finiamo difilati in \mathbb{R}^N con N molto, molto grande... Ma anche lasciando perdere il calcio, nel corso dei tuoi studi incontrerai dozzine di situazioni che richiederanno necessariamente l'uso di \mathbb{R}^n con $n > 3$. Insomma, non si tratta affatto di una generalizzazione inutile.

Fra le varie operazioni che si possono definire su \mathbb{R}^n due ci interessano in modo particolare: la somma, e il prodotto per un numero reale.

Definizione 4.1 La *somma* di due vettori di \mathbb{R}^n è semplicemente la somma componente per componente:

$$\forall v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n \quad \forall w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \in \mathbb{R}^n \quad v + w = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} v_1 + w_1 \\ \vdots \\ v_n + w_n \end{pmatrix}.$$

Anche il prodotto³ per un numero reale (o, come diremo, *prodotto per uno scalare*) è definito componente per componente:

$$\forall \lambda \in \mathbb{R} \quad \forall v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n \quad \lambda v = \lambda \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} \lambda v_1 \\ \vdots \\ \lambda v_n \end{pmatrix};$$

non abbiamo fatto altro che applicare a \mathbb{R}^n le definizioni viste per \mathbb{R}^2 ed \mathbb{R}^3 .

Ispirati da quanto scritto nel Paragrafo 1.3, vediamo quali proprietà formali hanno la somma e il prodotto per scalari. Prima di tutto, \mathbb{R}^n con la somma è un *gruppo*

² E le teorie fisiche più recenti suggeriscono che in realtà di dimensioni ne occorrono almeno undici.

³ Come già accennato nella Nota 2 del Capitolo 2, per $n \geq 3$ non è possibile definire su \mathbb{R}^n un prodotto che, assieme alla somma appena introdotta, lo renda un campo.

commutativo (come sicuramente avrai cura di verificare), il cui elemento neutro è il *vettore nullo*

$$O = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},$$

e dove l'opposto di un vettore è dato da

$$-\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} -v_1 \\ \vdots \\ -v_n \end{pmatrix}.$$

Il prodotto per scalari invece soddisfa le seguenti proprietà:

$$\begin{aligned} \forall \lambda \in \mathbb{R} \quad \forall v, w \in \mathbb{R}^n \quad \lambda(v + w) &= \lambda v + \lambda w; \\ \forall \lambda, \mu \in \mathbb{R} \quad \forall v \in \mathbb{R}^n \quad (\lambda + \mu)v &= \lambda v + \mu v; \\ \forall \lambda, \mu \in \mathbb{R} \quad \forall v \in \mathbb{R}^n \quad (\lambda\mu)v &= \lambda(\mu v); \\ \forall v \in \mathbb{R}^n \quad 1v &= v \quad \text{e} \quad 0v = O \end{aligned}$$

(sono tutte ovvie, ma tu dimostralcelle lo stesso, in modo da essere certo di averle capite).

L'insieme \mathbb{R}^n non è l'unico insieme che conosciamo su cui sono definite due operazioni che soddisfano queste proprietà: abbiamo già incontrato \mathcal{V}_0^2 e \mathcal{V}_0^3 , ma anche i polinomi in una (o più) variabili con le usuali operazioni di somma e prodotto per un numero reale le soddisfano. E allora tanto vale introdurre la seguente definizione.

Definizione 4.2 Uno *spazio vettoriale* (o *spazio lineare*) su \mathbb{R} è un insieme V su cui sono definite due operazioni: una *somma* (che a due elementi di V associa un elemento di V) e un *prodotto per scalari* (che a un elemento di \mathbb{R} e un elemento di V associa un elemento di V), soddisfacenti le proprietà seguenti:

$$\begin{aligned} \forall u, v, w \in V \quad (u + v) + w &= u + (v + w); && \text{ass.} \\ \exists O \in V : \forall v \in V \quad v + O &= O + v = v; && \text{el. neutro} \\ \forall v \in V \exists -v \in V : v + (-v) &= (-v) + v = O; && \text{opposto} \\ \forall v, w \in V \quad v + w &= w + v; && \text{comm.} \\ \forall \lambda \in \mathbb{R} \quad \forall v, w \in V \quad \lambda(v + w) &= \lambda v + \lambda w; && \text{distrib.} \\ \forall \lambda, \mu \in \mathbb{R} \quad \forall v \in V \quad (\lambda + \mu)v &= \lambda v + \mu v; && \\ \forall \lambda, \mu \in \mathbb{R} \quad \forall v \in V \quad (\lambda\mu)v &= \lambda(\mu v); && \text{ass.} \\ \forall v \in V \quad 1v &= v \quad \text{e} \quad 0v = O. && \end{aligned}$$

In particolare, V con la somma è un gruppo commutativo. Inoltre, se al posto di \mathbb{R} sostituiamo ovunque un altro campo \mathbb{K} , abbiamo la definizione di spazio vettoriale sul campo \mathbb{K} . Gli elementi di uno spazio vettoriale sono detti *vettori*, e l'elemento neutro per la somma O si chiama *vettore nullo*.

ESEMPIO 4.1 Abbiamo già visto che \mathcal{V}_O^2 (Proposizioni 2.1 e 2.2), \mathcal{V}_O^3 (Esercizio 2.3), \mathcal{V}^2 (Paragrafo 2C.2) e $\mathbb{R}[t]$ (Esercizio 1.19) sono esempi di spazi vettoriali. Anche l'insieme $M_{m,n}(\mathbb{R})$ delle matrici $m \times n$ è uno spazio vettoriale su \mathbb{R} , con le operazioni

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

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

Se queste definizioni ti fanno sospettare che $M_{m,n}(\mathbb{R})$ e \mathbb{R}^{mn} siano molto simili — semplicemente, uno è scritto in tabelle rettangolari e l'altro in colonna —, hai perfettamente ragione: come vedremo nel Capitolo 7, questi due spazi sono isomorfi, in maniera anche più stringente di quanto non lo fossero \mathcal{V}_O^2 ed \mathbb{R}^2 .

ESEMPIO 4.2 Sia A un insieme qualunque, e V l'insieme di tutte le funzioni da A in \mathbb{R} . Definiamo una somma e un prodotto per scalari su V in questo modo: se $f, g \in V$ e $\lambda \in \mathbb{R}$, allora $f + g, \lambda f: A \rightarrow \mathbb{R}$ sono definite da

$$\forall a \in A \quad (f + g)(a) = f(a) + g(a) \quad \text{e} \quad (\lambda f)(a) = \lambda f(a).$$

Si verifica immediatamente (esercizio) che V con queste operazioni è uno spazio vettoriale. Se $A = \mathbb{R}$ possiamo considerare il sottoinsieme di V delle funzioni continue, o quello delle funzioni derivabili: con le stesse operazioni, sono anch'essi esempi di spazi vettoriali.

Negli spazi vettoriali ci sono dei sottoinsiemi piuttosto importanti, che sono l'analogo delle rette e dei piani per l'origine in \mathcal{V}_O^3 .

Definizione 4.3 Un *sottospazio vettoriale* di uno spazio vettoriale V è un sottoinsieme $W \subseteq V$ chiuso rispetto alla somma e al prodotto per scalari, ovvero tale che

$$\forall w, w_1, w_2 \in W \quad \forall \lambda \in \mathbb{R} \quad w_1 + w_2 \in W \quad \text{e} \quad \lambda w \in W.$$

Osservazione 4.2 Il vettore nullo O è contenuto in qualunque sottospazio vettoriale W : infatti, se w è un qualunque elemento di W , allora $O = 0w \in W$. Analogamente, se $v \in W$ allora $-v = (-1)v \in W$, per cui contiene anche l'opposto. In particolare, se L è un sottoinsieme di uno spazio vettoriale che non contiene il vettore nullo, L non può essere un sottospazio vettoriale.

Osservazione 4.3 Un *sottospazio vettoriale*, considerato con la somma e il prodotto per scalari, è uno spazio vettoriale a tutti gli effetti (perché?), contenuto dentro uno spazio vettoriale più grande.

Ovviamente, uno spazio vettoriale V è sempre sottospazio di se stesso; all'estremo opposto, l'insieme $\{O\}$ costituito dal solo vettore nullo è anch'esso sempre un sottospazio. Altri esempi meno banali sono descritti qui di seguito:

ESEMPIO 4.3 Fissiamo un riferimento affine in A^2 , in modo da identificare il piano euclideo con \mathbb{R}^2 . Sappiamo che $\{O\}$ ed \mathbb{R}^2 sono sottospazi; vediamo quali sono gli altri. Sia V un sottospazio di \mathbb{R}^2 ; se $V \neq \{O\}$, deve contenere un vettore $v \neq O$. Allora, essendo chiuso rispetto al prodotto per scalari, V deve contenere tutti i multipli di v , cioè tutti i vettori della forma λv con $\lambda \in \mathbb{R}$. Ma ora l'insieme $\mathbb{R}v = \{\lambda v \mid \lambda \in \mathbb{R}\}$ (che geometricamente è una retta per l'origine) è un sottospazio di \mathbb{R}^2 : infatti

$$\lambda v + \mu v = (\lambda + \mu)v \quad \text{e} \quad \lambda(\mu v) = (\lambda\mu)v,$$

per cui è chiuso rispetto alla somma e al prodotto per scalari.

Quindi le rette per l'origine sono sottospazi di \mathbb{R}^2 . Supponiamo ora che V sia ancora più grosso, cioè che oltre alla retta $\mathbb{R}v$ contenga anche un altro vettore $w \notin \mathbb{R}v$. Identificando v e w con vettori applicati nell'origine, vediamo subito che ogni altro vettore di \mathbb{R}^2 si può scrivere come somma di un multiplo di v e di un multiplo di w (grazie alla Proposizione 2.3). Essendo V chiuso rispetto alla somma e al prodotto per scalari, questo vuol dire che $V = \mathbb{R}^2$. In conclusione, i sottospazi di \mathbb{R}^2 sono tutti e soli $\{O\}$, le rette per l'origine ed \mathbb{R}^2 stesso.

L'idea geometrica suggerita da questo Esempio è che i sottospazi vettoriali di \mathbb{R}^n sono degli \mathbb{R}^m con $m \leq n$ (ovvero rette, piani, eccetera).

L'esempio di sottospazio forse più importante per noi ci viene fornito dai sistemi lineari.

Definizione 4.4 Un sistema lineare della forma $Ax = O$ (ovvero in cui i termini noti sono tutti nulli) è detto *omogeneo*. Se $Bx = b$ è un sistema lineare qualunque, il sistema $Bx = O$ è il *sistema omogeneo associato* al sistema $Bx = b$.

Chiaramente il vettore nullo O è soluzione di un sistema lineare omogeneo, ma potrebbe non essere il solo, come capita per esempio nel sistema

$$\begin{cases} 2x_1 + x_2 = 0, \\ -4x_1 - 2x_2 = 0. \end{cases}$$

Proposizione 4.1 Sia $W \subseteq \mathbb{R}^n$ l'insieme delle soluzioni di un sistema lineare omogeneo in n incognite $Ax = O$. Allora W è un sottospazio vettoriale di \mathbb{R}^n .

Dimostrazione. Il nostro sistema è della forma

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = 0, \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = 0. \end{cases}$$

Siano v e w due soluzioni del sistema; dobbiamo dimostrare che anche $v + w$ e λv (dove λ è un qualunque numero reale) lo sono. Ma infatti, dire che v e w sono soluzioni del

sistema equivale a dire che

$$\begin{cases} a_{11}v_1 + \cdots + a_{1n}v_n = 0, \\ \vdots \\ a_{m1}v_1 + \cdots + a_{mn}v_n = 0, \end{cases}$$

e che

$$\begin{cases} a_{11}w_1 + \cdots + a_{1n}w_n = 0, \\ \vdots \\ a_{m1}w_1 + \cdots + a_{mn}w_n = 0. \end{cases}$$

Sommando otteniamo

$$\begin{cases} a_{11}(v_1 + w_1) + \cdots + a_{1n}(v_n + w_n) = 0, \\ \vdots \\ a_{m1}(v_1 + w_1) + \cdots + a_{mn}(v_n + w_n) = 0, \end{cases}$$

cioè $v + w$ è soluzione del sistema. Analogamente, moltiplicando per λ otteniamo

$$\begin{cases} a_{11}(\lambda v_1) + \cdots + a_{1n}(\lambda v_n) = 0, \\ \vdots \\ a_{m1}(\lambda v_1) + \cdots + a_{mn}(\lambda v_n) = 0, \end{cases}$$

cioè λv è soluzione del sistema. \square

Osservazione 4.4 L'insieme delle soluzioni di un sistema lineare non omogeneo *non* è un sottospazio vettoriale, in quanto (Osservazione 4.2) non contiene il vettore nullo.

4.2 Combinazioni lineari

Introduciamo ora un concetto fondamentale.

Definizione 4.5 Sia V uno spazio vettoriale; la *combinazione lineare* di k vettori $v_1, \dots, v_k \in V$ con *coefficienti* (o *pesi*) $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ è il vettore

$$\alpha_1 v_1 + \cdots + \alpha_k v_k \in V.$$

Lo *span* dei (o *sottospazio generato dai*) vettori v_1, \dots, v_k è l'insieme di tutte le possibili combinazioni lineari di v_1, \dots, v_k . In simboli,

$$\text{Span}(v_1, \dots, v_k) = \{ \alpha_1 v_1 + \cdots + \alpha_k v_k \mid \alpha_1, \dots, \alpha_k \in \mathbb{R} \}.$$

Per giustificare il nome dobbiamo dimostrare che lo span è effettivamente un sottospazio:

Proposizione 4.2 Sia V uno spazio vettoriale, e v_1, \dots, v_k vettori di V . Allora $\text{Span}(v_1, \dots, v_k)$ è un sottospazio di V .

Dimostrazione. Per dimostrare che è chiuso rispetto alla somma dobbiamo far vedere che la somma di due combinazioni lineari di v_1, \dots, v_k è ancora una combinazione lineare di v_1, \dots, v_k . Infatti se $\alpha_1 v_1 + \cdots + \alpha_k v_k, \beta_1 v_1 + \cdots + \beta_k v_k \in \text{Span}(v_1, \dots, v_k)$ sono due combinazioni lineari si ha

$$\begin{aligned} (\alpha_1 v_1 + \cdots + \alpha_k v_k) + (\beta_1 v_1 + \cdots + \beta_k v_k) \\ = (\alpha_1 + \beta_1)v_1 + \cdots + (\alpha_k + \beta_k)v_k \in \text{Span}(v_1, \dots, v_k), \end{aligned}$$

come voluto. Analogamente, se $\lambda \in \mathbb{R}$ si ha

$$\lambda(\alpha_1 v_1 + \cdots + \alpha_k v_k) = (\lambda \alpha_1)v_1 + \cdots + (\lambda \alpha_k)v_k \in \text{Span}(v_1, \dots, v_k),$$

per cui $\text{Span}(v_1, \dots, v_k)$ è effettivamente un sottospazio. \square

ESEMPIO 4.4 Se $v \in V$ è un vettore non nullo, allora $\text{Span}(v) = \mathbb{R}v = \{\lambda v \mid \lambda \in \mathbb{R}\}$ è la retta composta da tutti i multipli di v .

ESEMPIO 4.5 Prendiamo $V = \mathbb{R}^3$ e

$$v_1 = \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix} \quad \text{e} \quad v_2 = \begin{pmatrix} -1 \\ -1 \\ 0 \end{pmatrix}.$$

Allora

$$\text{Span}(v_1, v_2) = \left\{ \alpha_1 \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} -1 \\ -1 \\ 0 \end{pmatrix} \mid \alpha_1, \alpha_2 \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} 3\alpha_1 - \alpha_2 \\ \alpha_1 - \alpha_2 \\ 0 \end{pmatrix} \mid \alpha_1, \alpha_2 \in \mathbb{R} \right\}.$$

Geometricamente, è il piano $z = 0$, cioè l'insieme di tutti i vettori della forma $\begin{pmatrix} b_1 \\ b_2 \\ 0 \end{pmatrix}$.

Infatti è facile verificare che per ogni $b_1, b_2 \in \mathbb{R}$ esistono $\alpha_1, \alpha_2 \in \mathbb{R}$ tali che

$$\begin{cases} 3\alpha_1 - \alpha_2 = b_1, \\ \alpha_1 - \alpha_2 = b_2, \end{cases} \quad (4.1)$$

per cui

$$\begin{pmatrix} b_1 \\ b_2 \\ 0 \end{pmatrix} = \alpha_1 v_1 + \alpha_2 v_2 \in \text{Span}(v_1, v_2);$$

per l'esattezza, risolvendo il sistema (4.1) si vede che basta prendere $\alpha_1 = (b_1 - b_2)/2$ e $\alpha_2 = (b_1 - 3b_2)/2$.

ESEMPIO 4.6 Prendiamo di nuovo $V = \mathbb{R}^3$, ma stavolta consideriamo i vettori

$$v_1 = \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} -1 \\ -1 \\ 0 \end{pmatrix} \quad \text{e} \quad v_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Di primo acchito troviamo

$$\text{Span}(v_1, v_2, v_3) = \left\{ \begin{pmatrix} 3\alpha_1 - \alpha_2 + \alpha_3 \\ \alpha_1 - \alpha_2 \\ 0 \end{pmatrix} \mid \alpha_1, \alpha_2, \alpha_3 \in \mathbb{R} \right\}.$$

Ma ora, ponendo $\beta_1 = \alpha_1 + \alpha_3/2$ e $\beta_2 = \alpha_2 + \alpha_3/2$, otteniamo

$$\begin{pmatrix} 3\alpha_1 - \alpha_2 + \alpha_3 \\ \alpha_1 - \alpha_2 \\ 0 \end{pmatrix} = \beta_1 \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix} + \beta_2 \begin{pmatrix} -1 \\ -1 \\ 0 \end{pmatrix},$$

per cui

$$\text{Span}(v_1, v_2, v_3) = \text{Span}(v_1, v_2).$$

Osservazione 4.5 A volte, se $S = \{v_1, \dots, v_k\}$ è un sottoinsieme di uno spazio vettoriale V scriveremo $\text{Span}(S)$ invece di $\text{Span}(v_1, \dots, v_k)$. Questo invita a definire lo span anche di sottoinsiemi infiniti di V . Se S è un sottoinsieme qualunque (finito o infinito, non importa) di V , il suo span è l'insieme $\text{Span}(S)$ delle combinazioni lineari finite di elementi di S . In simboli,

$$\text{Span}(S) = \{\alpha_1 v_1 + \dots + \alpha_p v_p \in V \mid p \in \mathbb{N}, v_1, \dots, v_p \in S, \alpha_1, \dots, \alpha_p \in \mathbb{R}\};$$

si dimostra facilmente (Esercizio 4.9) che $\text{Span}(S)$ è sempre un sottospazio vettoriale di V . Comunque, useremo molto raramente questo concetto quando S è infinito.

Ricordo che un sistema lineare $Ax = b$ è *compatibile* se ammette soluzioni. Con la terminologia introdotta possiamo dare un criterio (non molto operativo per il momento ma che sarà utile in seguito) per decidere quando un sistema è compatibile:

Proposizione 4.3 Sia $Ax = b$ un sistema lineare in m equazioni ed n incognite, e siano $A^1, \dots, A^n \in \mathbb{R}^m$ le colonne della matrice dei coefficienti. Allora il sistema è compatibile se e solo se

$$b \in \text{Span}(A^1, \dots, A^n).$$

Dimostrazione. Infatti il sistema si può scrivere come

$$x_1 A^1 + \dots + x_n A^n = b \quad (4.2)$$

(convincitene prima di proseguire; è importante). Quindi il sistema è compatibile se e solo se b è combinazione lineare delle colonne della matrice dei coefficienti, cioè se e solo se $b \in \text{Span}(A^1, \dots, A^n)$. \square

4.3 Indipendenza lineare e basi

Nel Paragrafo 2.2 avevamo parlato di basi di \mathcal{V}_O^2 e \mathcal{V}_O^3 ; in questo paragrafo vedremo come recuperare questo concetto fondamentale nel contesto degli spazi vettoriali.

Definizione 4.6 Un insieme di vettori $v_1, \dots, v_k \in V$ di uno spazio vettoriale V si dice *linearmente dipendente* se esistono $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ non tutti nulli tali che

$$\alpha_1 v_1 + \dots + \alpha_k v_k = O. \quad (4.3)$$

Viceversa, v_1, \dots, v_k si dicono *linearmente indipendenti* se non sono linearmente dipendenti, ovvero se $\alpha_1 v_1 + \dots + \alpha_k v_k = O$ implica $\alpha_1 = \dots = \alpha_k = 0$. Un'equazione della forma (4.3) con gli α_j non tutti nulli sarà chiamata *relazione di dipendenza lineare* fra v_1, \dots, v_k .

ESEMPIO 4.7 Se fra i vettori $v_1, \dots, v_k \in V$ vi è il vettore nullo, allora v_1, \dots, v_k sono linearmente dipendenti. Infatti se $v_i = O$ allora

$$0v_1 + \dots + 0v_{i-1} + 1v_i + 0v_{i+1} + \dots + 0v_k = O$$

è una relazione di dipendenza lineare fra v_1, \dots, v_k .

ESEMPIO 4.8 Se fra i vettori $v_1, \dots, v_k \in V$ ve ne sono due uguali, allora v_1, \dots, v_k sono automaticamente linearmente dipendenti. Infatti se si ha $v_i = v_j$ con $i < j$ allora

$$0v_1 + \dots + 0v_{i-1} + v_i + 0v_{i+1} + \dots + 0v_{j-1} + (-1)v_j + 0v_{j+1} + \dots + 0v_k = O$$

è una relazione di dipendenza lineare fra v_1, \dots, v_k .

ESEMPIO 4.9 I vettori v_1 e $v_2 \in \mathbb{R}^3$ dell'Esempio 4.5 sono linearmente indipendenti. Infatti, la relazione $\alpha_1 v_1 + \alpha_2 v_2 = O$ vale se e solo se

$$\begin{cases} 3\alpha_1 - \alpha_2 = 0, \\ \alpha_1 - \alpha_2 = 0, \end{cases}$$

e questo sistema ha come unica soluzione $\alpha_1 = \alpha_2 = 0$.

ESEMPIO 4.10 I vettori $v_1, v_2, v_3 \in \mathbb{R}^3$ dell'Esempio 4.6 sono invece linearmente dipendenti; infatti,

$$\frac{1}{2}v_1 + \frac{1}{2}v_2 - v_3 = O.$$

In generale, per vedere se i vettori $v_1, \dots, v_k \in \mathbb{R}^n$ sono linearmente indipendenti si forma la matrice

$$A = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_k \\ | & & | \end{bmatrix} \in M_{n,k}(\mathbb{R})$$

e si studia il sistema $Ax = O$: i vettori sono dipendenti se e solo se il sistema ammette soluzioni diverse dalla soluzione banale $x = O$. Questo è il contenuto della seguente

Proposizione 4.4 Il sistema omogeneo $Ax = O$ in n incognite ha soluzione unica (quella banale $x = O$) se e solo se le colonne A^1, \dots, A^n sono linearmente indipendenti.

Dimostrazione. Per (4.2), una soluzione del sistema è data da $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ tali che

$$\alpha_1 A^1 + \dots + \alpha_n A^n = O.$$

Quindi esiste una soluzione diversa da $x = O$ se e solo se le colonne di A sono linearmente dipendenti. \square

Osservazione 4.6 Dire che i vettori $v_1, \dots, v_k \in V$ sono linearmente dipendenti equivale a dire che è possibile ricavare uno di loro come combinazione lineare degli altri. Infatti, se abbiamo $\alpha_1 v_1 + \dots + \alpha_k v_k = O$ con, per esempio, $\alpha_1 \neq 0$, allora

$$v_1 = -\frac{\alpha_2}{\alpha_1} v_2 - \dots - \frac{\alpha_k}{\alpha_1} v_k.$$

In particolare, quindi

$$\text{Span}(v_1, \dots, v_k) = \text{Span}(v_2, \dots, v_k)$$

(perché? Confronta l'Esempio 4.6), ovvero il vettore v_1 in realtà non serviva per descrivere il sottospazio generato da $\{v_1, \dots, v_k\}$; gli altri erano sufficienti. Questo ci porta alla nozione di base di uno spazio vettoriale.

Definizione 4.7 Sia V uno spazio vettoriale. Un insieme $\mathcal{B} = \{v_1, \dots, v_n\}$ di vettori di V è una base di V se

- (a) $V = \text{Span}(v_1, \dots, v_n)$, cioè v_1, \dots, v_n sono un sistema di generatori di V ;
- (b) v_1, \dots, v_n sono linearmente indipendenti.

ESEMPIO 4.11 I vettori $e_1, \dots, e_n \in \mathbb{R}^n$, dove

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix},$$

sono una base di \mathbb{R}^n , la base canonica.

ESEMPIO 4.12 Sia

$$V = \left\{ \begin{pmatrix} b_1 \\ b_2 \\ 0 \end{pmatrix} \mid b_1, b_2 \in \mathbb{R} \right\}$$

il piano $z = 0$ in \mathbb{R}^3 . Allora i vettori $\{v_1, v_2\}$ dell'Esempio 4.5 sono una base di V ; infatti, abbiamo visto nell'Esempio 4.5 che sono un sistema di generatori di V , e nell'Esempio 4.9 che sono linearmente indipendenti. Non sono l'unica base di V ; anche $\{e_1, e_2\} \subset \mathbb{R}^3$ lo è, e non è difficile trovarne infinite altre.

ESEMPIO 4.13 Consideriamo lo spazio vettoriale (Esercizio 4.5) $V = \mathbb{R}_3[t]$ dei polinomi in una variabile a coefficienti reali di grado minore o uguale a 3. Allora i polinomi

$$p_0 = 1, p_1 = t, p_2 = t^2, p_3 = t^3$$

compongono una base di V . Infatti, sono un sistema di generatori: un generico polinomio

$$p = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3$$

si scrive come combinazione lineare di p_0, \dots, p_3 nel modo più naturale, cioè

$$p = \alpha_0 p_0 + \alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3.$$

Sono anche linearmente indipendenti: infatti un polinomio è identicamente nullo se e solo se tutti i suoi coefficienti sono zero.

Ovviamente, lo stesso ragionamento ci dice che

$$p_0 = 1, p_1 = t, \dots, p_n = t^n$$

è una base di $\mathbb{R}_n[t]$, lo spazio dei polinomi in una variabile a coefficienti reali di grado minore o uguale a n .

ESEMPIO 4.14 Se i vettori v_1, \dots, v_k di uno spazio vettoriale V sono linearmente indipendenti, allora sono una base del sottospazio $\text{Span}(v_1, \dots, v_k) \subseteq V$.

L'importanza delle basi consiste nel fatto che se $\{v_1, \dots, v_n\}$ è una base di V allora ogni elemento di V può essere scritto in un solo modo come combinazione lineare dei vettori v_1, \dots, v_n .

Proposizione 4.5 Siano v_1, \dots, v_k vettori linearmente indipendenti di uno spazio vettoriale V . Supponiamo che $\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_k \in \mathbb{R}$ siano tali che

$$\alpha_1 v_1 + \dots + \alpha_k v_k = \beta_1 v_1 + \dots + \beta_k v_k. \quad (4.4)$$

Allora $\alpha_j = \beta_j$ per $j = 1, \dots, k$.

Dimostrazione. La (4.4) implica che

$$(\alpha_1 - \beta_1)v_1 + \dots + (\alpha_k - \beta_k)v_k = 0;$$

la tesi segue dall'indipendenza lineare dei v_j . \square

Sia $\{v_1, \dots, v_n\}$ una base di uno spazio vettoriale V , e prendiamo $v \in V$. Siccome la base è un sistema di generatori di V , esistono $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ tali che $v = \alpha_1 v_1 + \dots + \alpha_n v_n$; siccome i v_j sono linearmente indipendenti, per la Proposizione 4.5 gli $\alpha_1, \dots, \alpha_n$ sono univocamente determinati.

Definizione 4.8 Sia $\mathcal{B} = \{v_1, \dots, v_n\}$ una base di uno spazio vettoriale V , e $v \in V$. Gli n numeri reali $(\alpha_1, \dots, \alpha_n)$ tali che $v = \alpha_1 v_1 + \dots + \alpha_n v_n$ sono le coordinate di v rispetto alla base \mathcal{B} .

In particolare, quindi, *uno spazio vettoriale⁴ è completamente determinato una volta che se ne conosce una base*: infatti ogni suo elemento si ottiene in uno e un solo modo come combinazione lineare degli elementi della base, con le coordinate come coefficienti.

ESEMPIO 4.15 Se $x \in \mathbb{R}^n$ possiamo scrivere

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1 e_1 + \cdots + x_n e_n,$$

per cui (x_1, \dots, x_n) sono le coordinate di x rispetto alla base canonica.

ESEMPIO 4.16 Fissiamo un riferimento affine $RA(O, \vec{i}, \vec{j})$ di \mathcal{V}_O^2 . Allora $\mathcal{B} = \{\vec{i}, \vec{j}\}$ è una base di \mathcal{V}_O^2 ; infatti la Proposizione 2.3 ci dice che sono un sistema di generatori, e l'indipendenza lineare segue dal fatto che non sono proporzionali (Esercizio 4.19). Dunque le coordinate che avevamo introdotto nel Paragrafo 2.2 non sono altro che un caso particolare delle coordinate appena definite.

ESEMPIO 4.17 Sia V lo spazio vettoriale dell'Esempio 4.12, e prendiamo

$$w = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} \in V.$$

Allora w ha (perché?) coordinate $(2, 0)$ rispetto alla base $\mathcal{B}_1 = \{e_1, e_2\}$. Per trovare le coordinate di w rispetto alla base $\mathcal{B}_2 = \{v_1, v_2\}$, dobbiamo risolvere il sistema $w = \alpha_1 v_1 + \alpha_2 v_2$, cioè

$$\begin{cases} 3\alpha_1 - \alpha_2 = 2, \\ \alpha_1 - \alpha_2 = 0. \end{cases}$$

Si vede subito che la soluzione è $\alpha_1 = \alpha_2 = 1$, per cui le coordinate di w rispetto a \mathcal{B}_2 sono $(1, 1)$.

Osservazione 4.7 Per stabilire se $v_1, \dots, v_k \in \mathbb{R}^n$ sono un sistema di generatori di \mathbb{R}^n , si considera di nuovo la matrice

$$A = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_k \\ | & & | \end{pmatrix} \in M_{n,k}(\mathbb{R});$$

i vettori v_1, \dots, v_k sono un sistema di generatori se e solo se il sistema $Ax = b$ ha soluzione qualunque sia $b \in \mathbb{R}^n$. Ovviamente questa è una tecnica un po' scomoda; un approccio migliore è descritto nell'Esercizio 4.24, ma per un sistema davvero efficace si rimanda al Paragrafo 6.3.

⁴ Per esempio, l'insieme delle soluzioni di un sistema lineare omogeneo.

4.4 Esistenza delle basi

A questo punto una domanda sorge spontanea: ogni spazio vettoriale ha una base (finita)? Risposta: no; non è neanche detto che abbia un sistema finito (cioè composto da un numero finito di vettori) di generatori⁵; per esempio, $\mathbb{R}[t]$ non ce l'ha (Esercizio 4.26). Come vedremo, questa è l'unica difficoltà: non appena esiste un sistema finito di generatori, esiste anche una base finita.

Definizione 4.9 Sia $\mathcal{A} \subseteq V$ un sottoinsieme di uno spazio vettoriale V . Diremo che $\mathcal{B} \subseteq \mathcal{A}$ è un sottoinsieme *massimale in \mathcal{A} di vettori linearmente indipendenti* se prima di tutto gli elementi di \mathcal{B} sono linearmente indipendenti, e inoltre aggiungendo a \mathcal{B} un qualunque altro elemento di \mathcal{A} si ottiene un insieme di vettori linearmente dipendenti.

ESEMPIO 4.18 Sia V il solito spazio vettoriale dell'Esempio 4.12, e

$$\mathcal{A} = \left\{ v_1, v_2, \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}, e_1 \right\} \subset V.$$

Allora $\mathcal{B}_1 = \{v_1, v_2\}$ o $\mathcal{B}_2 = \{e_1, v_1\}$ sono esempi di sottoinsiemi massimali in \mathcal{A} di vettori linearmente indipendenti.

Una base (se esiste) è un tipico caso di insieme massimale di vettori linearmente indipendenti (vedi anche l'Esercizio 4.20):

Proposizione 4.6 Sia $\mathcal{B} = \{v_1, \dots, v_n\}$ una base di uno spazio vettoriale V . Allora \mathcal{B} è un insieme massimale in V di vettori linearmente indipendenti.

Dimostrazione. Sappiamo per definizione che i vettori v_1, \dots, v_n sono linearmente indipendenti; dobbiamo dimostrare che, qualunque sia $v \in V$, i vettori v, v_1, \dots, v_n sono linearmente dipendenti. Ma infatti \mathcal{B} è un sistema di generatori; quindi esistono $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ tali che

$$v = \alpha_1 v_1 + \cdots + \alpha_n v_n.$$

Quindi si ha

$$v - \alpha_1 v_1 - \cdots - \alpha_n v_n = 0,$$

che è una relazione di dipendenza lineare fra v, v_1, \dots, v_n (il coefficiente di v è 1, che è non nullo). \square

Viceversa, un sottoinsieme massimale di vettori linearmente indipendenti in un sistema di generatori è una base. Per dimostrarlo, ci serve un piccolo lemma:

⁵ Attenzione: è possibile dare una definizione di base composta anche da infiniti elementi, e quindi dimostrare che ogni spazio vettoriale ha una base in questo senso; vedi i Complementi a questo capitolo. In questo libro però considereremo quasi esclusivamente basi finite.

Lemma 4.7 Sia $\mathcal{B} = \{v_1, \dots, v_r\}$ un sottoinsieme di uno spazio vettoriale V . Supponiamo che $\text{Span}(\mathcal{B})$ contenga un sistema di generatori di V . Allora $V = \text{Span}(\mathcal{B})$, cioè anche $\{v_1, \dots, v_r\}$ è un sistema di generatori di V .

Dimostrazione. Sia \mathcal{A} un sistema di generatori di V contenuto in $\text{Span}(\mathcal{B})$, e sia v un qualunque elemento di V . Essendo \mathcal{A} un sistema di generatori di V , possiamo scrivere v come combinazione lineare di elementi di \mathcal{A} . Siccome $\mathcal{A} \subset \text{Span}(\mathcal{B})$, possiamo scrivere gli elementi di \mathcal{A} come combinazione lineare di elementi di \mathcal{B} ; mettendo il tutto insieme, possiamo scrivere v come combinazione lineare di elementi di \mathcal{B} , per cui $v \in \text{Span}(\mathcal{B})$. Essendo v generico, abbiamo fatto vedere che \mathcal{B} è un sistema di generatori di V , come richiesto. \square

Teorema 4.8 Sia $\mathcal{A} = \{v_1, \dots, v_k\}$ un sistema di generatori di uno spazio vettoriale V . Sia $\mathcal{B} \subseteq \mathcal{A}$ un sottoinsieme massimale in \mathcal{A} di vettori linearmente indipendenti. Allora \mathcal{B} è una base di V .

Dimostrazione. Scriviamo $\mathcal{B} = \{v_1, \dots, v_r\}$ con $r \leq k$ (cosa sempre possibile, a meno di cambiare l'ordine in cui abbiamo elencato gli elementi di \mathcal{A}). Essendo gli elementi di \mathcal{B} linearmente indipendenti per ipotesi, dobbiamo solo dimostrare che $V = \text{Span}(\mathcal{B})$, sapendo che \mathcal{A} è un sistema di generatori.

Grazie al Lemma 4.7, è sufficiente far vedere che $\mathcal{A} \subset \text{Span}(\mathcal{B})$. Sia w un qualunque elemento di \mathcal{A} che non stia già in \mathcal{B} . Per l'ipotesi di massimalità, v_1, \dots, v_r, w sono linearmente dipendenti; quindi esistono $\alpha_1, \dots, \alpha_r, \beta \in \mathbb{R}$ non tutti nulli tali che

$$\alpha_1 v_1 + \dots + \alpha_r v_r + \beta w = 0.$$

Ora, se fosse $\beta = 0$ avremmo scritto una relazione di dipendenza lineare fra i vettori v_1, \dots, v_r che sappiamo essere linearmente indipendenti; quindi dev'essere per forza $\beta \neq 0$. Ma allora

$$w = -\frac{\alpha_1}{\beta} v_1 - \dots - \frac{\alpha_r}{\beta} v_r \in \text{Span}(\mathcal{B}),$$

e, essendo w generico, otteniamo $\mathcal{A} \subset \text{Span}(\mathcal{B})$. \square

Siccome ogni insieme finito di generatori contiene (perché?) un insieme massimale di vettori linearmente indipendenti, abbiamo dimostrato il

Corollario 4.9 Ogni spazio vettoriale contenente un sistema finito di generatori ammette una base.

Rimane il problema di come trovare una base di un dato spazio vettoriale. Il Teorema 4.8 suggerisce una prima procedura. Sia $\mathcal{A} = \{v_1, \dots, v_k\}$ un sistema di generatori di V . Se $V = \{0\}$ abbiamo finito; altrimenti, \mathcal{A} contiene un elemento non nullo, diciamo v_1 . Se v_1 e v_j sono linearmente dipendenti qualunque sia $v_j \in \mathcal{A}$, allora $\{v_1\}$ è un sottoinsieme massimale in \mathcal{A} di vettori linearmente indipendenti, per cui è la base cercata. Altrimenti, ci sarà un altro vettore di \mathcal{A} , diciamo v_2 , tale che v_1

e v_2 sono linearmente indipendenti. Se $\{v_1, v_2\}$ è un sottoinsieme massimale in \mathcal{A} di vettori linearmente indipendenti, abbiamo trovato la nostra base; altrimenti ci sarà un terzo vettore di \mathcal{A} , diciamo v_3 , tale che v_1, v_2 e v_3 sono linearmente indipendenti. Si procede così fino a trovare la base voluta. Chiaramente, non è un metodo molto efficace; tecniche migliori le studieremo nel Paragrafo 6.3.

ESEMPIO 4.19 Consideriamo lo spazio vettoriale

$$V = \text{Span} \left(v_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 2 \end{pmatrix}, v_2 = \begin{pmatrix} 2 \\ 1 \\ 1 \\ 0 \end{pmatrix}, v_3 = \begin{pmatrix} 5 \\ 4 \\ 2 \\ 2 \end{pmatrix} \right) \subset \mathbb{R}^4;$$

vogliamo una base di V . Prima di tutto, $V \neq \{0\}$; quindi possiamo cominciare a costruire la nostra base prendendo v_1 . Verifichiamo se v_1 e v_2 sono linearmente indipendenti; il sistema

$$\alpha_1 \begin{pmatrix} 1 \\ 2 \\ 0 \\ 2 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

ammette solo la soluzione $\alpha_1 = \alpha_2 = 0$ (come si vede facilmente considerando per esempio le prime due equazioni), per cui v_1 e v_2 sono linearmente indipendenti. Rimane da vedere cosa succede aggiungendo v_3 . Il sistema

$$\alpha_1 \begin{pmatrix} 1 \\ 2 \\ 0 \\ 2 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ 1 \\ 1 \\ 0 \end{pmatrix} + \alpha_3 \begin{pmatrix} 5 \\ 4 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

ammette la soluzione $\alpha_1 = 1, \alpha_2 = 2$ e $\alpha_3 = -1$ (come si vede risolvendo il sistema composto dalle prime tre equazioni e poi sostituendo il risultato nella quarta), per cui v_1, v_2 e v_3 sono linearmente dipendenti. Questo vuol dire che $\{v_1, v_2\}$ è un sottoinsieme massimale di vettori linearmente indipendenti in un sistema di generatori, e quindi una base di V .

Ora, è facile vedere che ogni spazio vettoriale ha un'infinità di basi distinte (esercizio: se $\{v_1, \dots, v_n\}$ è una base di V , anche $\{2v_1, v_2, \dots, v_n\}$ lo è). Ma due diverse basi di uno stesso spazio vettoriale hanno una caratteristica fondamentale in comune: hanno sempre lo stesso numero di elementi. Questa è una conseguenza del seguente (e importante) *Teorema del completamento*:

Teorema 4.10 (del completamento) Sia $\mathcal{B} = \{v_1, \dots, v_n\}$ una base di uno spazio vettoriale V , e siano $w_1, \dots, w_p \in V$ (con $p \leq n$) vettori linearmente indipendenti. Allora esistono $n - p$ vettori di \mathcal{B} che insieme a w_1, \dots, w_p formano una base di V .

Dimostrazione. Procediamo per induzione su p . Sia $p = 1$; siccome \mathcal{B} è una base di V , esistono $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ tali che

$$w_1 = \alpha_1 v_1 + \dots + \alpha_n v_n. \quad (4.5)$$

Nel caso $p = 1$, l'ipotesi d'indipendenza lineare si riduce a $w_1 \neq O$; quindi almeno uno degli α_j è non nullo. A meno dell'ordine possiamo supporre $\alpha_1 \neq 0$ ricavando

$$v_1 = \frac{1}{\alpha_1} w_1 - \frac{\alpha_2}{\alpha_1} v_2 - \dots - \frac{\alpha_n}{\alpha_1} v_n \in \text{Span}(w_1, v_2, \dots, v_n).$$

Dunque $\mathcal{B} \subset \text{Span}(w_1, v_2, \dots, v_n)$ per cui, grazie al Lemma 4.7, $\{w_1, v_2, \dots, v_n\}$ è un sistema di generatori di V ; rimane da dimostrare che sono linearmente indipendenti.

Supponiamo che si abbia $\beta_1 w_1 + \beta_2 v_2 + \dots + \beta_n v_n = O$ per certi $\beta_1, \dots, \beta_n \in \mathbb{R}$; dobbiamo dimostrare che $\beta_1 = \dots = \beta_n = 0$. Usando la (4.5) otteniamo

$$\begin{aligned} O &= \beta_1(\alpha_1 v_1 + \dots + \alpha_n v_n) + \beta_2 v_2 + \dots + \beta_n v_n \\ &= (\beta_1 \alpha_1) v_1 + (\beta_1 \alpha_2 + \beta_2) v_2 + \dots + (\beta_1 \alpha_n + \beta_n) v_n. \end{aligned}$$

Siccome v_1, \dots, v_n sono linearmente indipendenti per ipotesi, i coefficienti di questa combinazione lineare devono essere tutti nulli; essendo $\alpha_1 \neq 0$ ne segue che $\beta_1 = 0$ e quindi $\beta_2 = \dots = \beta_n = 0$. In conclusione w_1, v_2, \dots, v_n sono linearmente indipendenti, e il caso $p = 1$ è dimostrato.

Supponiamo ora la tesi vera per $p - 1$ vettori, e dimostriamola per p vettori. Per ipotesi, possiamo trovare $n - (p - 1)$ elementi di \mathcal{B} (che a meno dell'ordine possiamo supporre essere v_p, \dots, v_n) tali che $\{w_1, \dots, w_{p-1}, v_p, \dots, v_n\}$ sia una base di V . Procediamo come prima. Devono esistere $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ tali che

$$w_p = \alpha_1 w_1 + \dots + \alpha_{p-1} w_{p-1} + \alpha_p v_p + \dots + \alpha_n v_n.$$

Non tutti gli α_j con $j = p, \dots, n$ possono essere nulli, perché altrimenti avremmo scritto una relazione di dipendenza lineare fra w_1, \dots, w_p ; a meno dell'ordine possiamo supporre che $\alpha_p \neq 0$. Dunque

$$v_p = \frac{1}{\alpha_p} w_p - \frac{\alpha_1}{\alpha_p} w_1 - \dots - \frac{\alpha_{p-1}}{\alpha_p} w_{p-1} - \frac{\alpha_{p+1}}{\alpha_p} v_{p+1} - \dots - \frac{\alpha_n}{\alpha_p} v_n$$

appartiene a $\text{Span}(w_1, \dots, w_p, v_{p+1}, \dots, v_n)$. Di nuovo, questo implica che i vettori $\{w_1, \dots, w_p, v_{p+1}, \dots, v_n\}$ formano un sistema di generatori, grazie al Lemma 4.7; rimane da dimostrare che sono anche linearmente indipendenti.

Siano $\beta_1, \dots, \beta_n \in \mathbb{R}$ tali che

$$\beta_1 w_1 + \dots + \beta_p w_p + \beta_{p+1} v_{p+1} + \dots + \beta_n v_n = O.$$

Procedendo come nel caso $p = 1$ troviamo

$$\begin{aligned} O &= \beta_1 w_1 + \dots + \beta_{p-1} w_{p-1} + \beta_p (\alpha_1 w_1 + \dots + \alpha_{p-1} w_{p-1} + \alpha_p v_p + \dots + \alpha_n v_n) \\ &\quad + \beta_{p+1} v_{p+1} + \dots + \beta_n v_n \\ &= \gamma_1 w_1 + \dots + \gamma_{p-1} w_{p-1} + (\beta_p \alpha_p) v_p + \gamma_{p+1} v_{p+1} + \dots + \gamma_n v_n, \end{aligned}$$

dove abbiamo posto $\gamma_j = \beta_j + \beta_p \alpha_j$. Per ipotesi $w_1, \dots, w_{p-1}, v_p, \dots, v_n$ sono linearmente indipendenti; quindi (essendo $\alpha_p \neq 0$) deduciamo $\beta_p = 0$, e dunque tutti i β_j sono zero. Di conseguenza $w_1, \dots, w_p, v_{p+1}, \dots, v_n$ sono linearmente indipendenti, come volevamo. \square

Quando applichiamo questo teorema diremo che *completiamo* $\{w_1, \dots, w_p\}$ a una base di V .

Osservazione 4.8 Se w_1, \dots, w_p sono vettori linearmente indipendenti di \mathbb{R}^n , per completarli a una base basta aggiungerci $n - p$ vettori della base canonica. Il procedimento consiste nell'estrarre una base dall'insieme $\{w_1, \dots, w_p, e_1, \dots, e_n\}$ come indicato nel prossimo Esempio o (meglio) con le tecniche che vedremo nel Paragrafo 6.3.

ESEMPIO 4.20 Prendiamo i vettori

$$w_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad w_2 = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \in \mathbb{R}^4;$$

vogliamo dimostrare che sono linearmente indipendenti e poi completare $\{w_1, w_2\}$ a una base di \mathbb{R}^4 . Per vedere che sono linearmente indipendenti dobbiamo risolvere il sistema $\alpha_1 w_1 + \alpha_2 w_2 = O$, cioè

$$\begin{cases} \alpha_1 + \alpha_2 = 0, \\ 2\alpha_2 = 0, \\ \alpha_1 + 3\alpha_2 = 0, \\ 4\alpha_2 = 0, \end{cases}$$

che chiaramente ammette come unica soluzione $\alpha_1 = \alpha_2 = 0$, per cui i nostri due vettori sono linearmente indipendenti. Per completare $\{w_1, w_2\}$ a una base di \mathbb{R}^4 procediamo come nella dimostrazione del Teorema 4.10. Cominciamo con lo scrivere w_1 come combinazione lineare dei vettori della base canonica di \mathbb{R}^4 :

$$w_1 = 1e_1 + 0e_2 + 1e_3 + 0e_4. \quad (4.6)$$

La dimostrazione del Teorema 4.10 ci dice che possiamo mettere w_1 al posto di uno qualunque dei vettori della base canonica che compaiono con coefficiente non nullo in (4.6), ottenendo ancora una base di \mathbb{R}^4 . Per esempio, mettendo w_1 al posto di e_1 otteniamo la base $\mathcal{B}_1 = \{w_1, e_2, e_3, e_4\}$. Adesso scriviamo w_2 come combinazione lineare dei vettori di \mathcal{B}_1 :

$$w_2 = 1w_1 + 2e_2 + 2e_3 + 4e_4. \quad (4.7)$$

La dimostrazione del Teorema 4.10 ci dice che adesso possiamo mettere w_2 al posto di uno qualunque dei vettori di \mathcal{B}_1 (w_1 escluso, ovviamente) che compaiono con un coefficiente non nullo in (4.7). Per esempio sostituendo w_2 a e_3 otteniamo la base $\{w_1, w_2, e_2, e_4\}$, che completa $\{w_1, w_2\}$ come volevamo.

E ora una conseguenza fondamentale del precedente teorema:

Corollario 4.11 Siano $B = \{v_1, \dots, v_k\}$ e $C = \{w_1, \dots, w_h\}$ due basi di uno spazio vettoriale V . Allora $h = k$.

Dimostrazione. Supponiamo, per assurdo, $k > h$. Allora il Teorema 4.10 applicato con $n = k$ e $p = h$ ci dice che esistono $k - h$ elementi di B che aggiunti a C formano una base. Ma C era già una base, e quindi un insieme *massimale* di vettori linearmente indipendenti (Proposizione 4.6), contraddizione. Un analogo ragionamento esclude la possibilità che h sia maggiore di k ; quindi $h = k$. \square

Dunque il numero di elementi di una base, cioè il numero di vettori necessari e sufficienti per descrivere un dato spazio vettoriale, è una caratteristica ben definita dello spazio vettoriale: la sua dimensione.

Definizione 4.10 Se $\{v_1, \dots, v_n\}$ è una base di uno spazio vettoriale V , il numero n (che per il Corollario 4.11 non dipende dalla base scelta), si chiama *dimensione* di V ; scriveremo $n = \dim V$ (oppure $n = \dim_{\mathbb{K}} V$, se è necessario specificare che V è uno spazio vettoriale sul campo \mathbb{K}). Lo spazio vettoriale $\{O\}$ composto dal solo vettore nullo ha dimensione zero. Diremo che uno spazio vettoriale privo di sistemi di generatori finiti è di *dimensione infinita*; in questo libro ci occuperemo quasi esclusivamente di spazi vettoriali di dimensione finita.

Per calcolare la dimensione di uno spazio vettoriale basta trovarne una base. Per esempio, grazie alla base canonica (Esempio 4.11) vediamo subito che \mathbb{R}^n ha dimensione n ; in particolare, una retta ha dimensione 1, un piano ha dimensione 2 e così via, coerentemente con la nozione intuitiva di dimensione.

Osservazione 4.9 In un certo senso, la dimensione di uno spazio vettoriale V è il numero dei parametri necessari per descriverlo; infatti, scelta una base tutti gli elementi di V si scrivono come combinazione lineare degli $n = \dim V$ fissati elementi della base con coefficienti le n generiche coordinate. Per chiarire meglio questo concetto vediamo un esempio.

ESEMPIO 4.21 Consideriamo l'insieme

$$V = \{A \in M_{2,2}(\mathbb{R}) \mid a_{11} + a_{22} = 0\};$$

è facile verificare (esercizio) che V è un sottospazio di $M_{2,2}(\mathbb{R})$. Il generico elemento di V è della forma

$$A = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}, \quad (4.8)$$

per cui dipende da tre parametri. Questo suggerisce l'idea che V abbia dimensione 3; per dimostrarlo, cerchiamone una base. Se vogliamo che i parametri a, b, c siano le coordinate rispetto a una base, la matrice generica $A \in V$ si deve poter scrivere

come $A = aM_1 + bM_2 + cM_3$, dove $\{M_1, M_2, M_3\}$ è la base che stiamo cercando. Guardando (4.8) è chiaro che questo è possibile solo prendendo

$$M_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad M_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Per costruzione, $\{M_1, M_2, M_3\}$ è un sistema di generatori di V ; rimane da vedere che sono anche linearmente indipendenti. Ma infatti la relazione di dipendenza lineare $\alpha_1 M_1 + \alpha_2 M_2 + \alpha_3 M_3 = O$ implica immediatamente (perché?) $\alpha_1 = \alpha_2 = \alpha_3 = 0$, e ci siamo.

Qualche altro corollario del Teorema 4.10:

Corollario 4.12 Sia V uno spazio vettoriale di dimensione n . Allora ogni n -upla di vettori linearmente indipendenti di V è una base.

Dimostrazione. Basta applicare il Teorema 4.10 con $p = n$. \square

Corollario 4.13 Sia V uno spazio vettoriale di dimensione n , e $w_1, \dots, w_p \in V$. Se $p > n$, allora w_1, \dots, w_p sono linearmente dipendenti.

Dimostrazione. Se w_1, \dots, w_n sono linearmente dipendenti, abbiamo finito. Altrimenti, per il Corollario 4.12 sono una base di V e, in particolare, un insieme massimale in V di vettori linearmente indipendenti (Proposizione 4.6); ma in tal caso i vettori $w_1, \dots, w_n, w_{n+1}, \dots, w_p$ sono necessariamente linearmente dipendenti. \square

Osservazione 4.10 Dunque la dimensione di uno spazio vettoriale è il *massimo* numero di vettori linearmente indipendenti; analogamente si può vedere (Esercizio 4.22) che è il *minimo* numero di vettori in un sistema di generatori.

Concludiamo questo paragrafo confrontando la dimensione di un sottospazio con quella dello spazio che lo contiene. Il risultato è prevedibile:

Proposizione 4.14 Sia V uno spazio vettoriale di dimensione n , e W un suo sottospazio. Allora:

- (i) W ha dimensione finita minore o uguale a n ;
- (ii) $\dim W = \dim V$ se e soltanto se $W = V$.

Dimostrazione. (i) Se $W = \{O\}$ abbiamo finito. Altrimenti, sia w_1 un elemento non nullo di W . Se $\{w_1\}$ non è un insieme massimale in W di vettori linearmente indipendenti, troviamo un $w_2 \in W$ tale che w_1 e w_2 sono linearmente indipendenti. Se $\{w_1, w_2\}$ non è un insieme massimale in W di vettori linearmente indipendenti, troviamo un $w_3 \in W$ tale che w_1, w_2 e w_3 sono linearmente indipendenti.

Continuando in questo modo, prima o poi dovremo fermarci, in quanto in V (che è più grosso di W) non possono esistere più di n elementi linearmente indipendenti fra loro (Corollario 4.13). Quindi troveremo un insieme $\{w_1, \dots, w_m\}$ (con $m \leq n$)

massimale in W di vettori linearmente indipendenti, ovvero una base di W (Esercizio 4.20).

(ii) Se $\dim W = n$, esiste una base di W composta da n elementi. Ma allora questa è anche una base di V , per il Corollario 4.12, e quindi $W = V$. Il viceversa è ovvio. \square

Osservazione 4.11 Se avessimo saputo a priori che il sottospazio W aveva un sistema finito di generatori — e quindi una base finita — il Corollario 4.13 ci permetterebbe subito di concludere che $\dim W \leq \dim V$. Tutto il punto della dimostrazione della Proposizione 4.14 è esattamente far vedere che ogni sottospazio di uno spazio vettoriale di dimensione finita ammette un sistema finito di generatori.

4.5 Somma e intersezione di sottospazi

Sia V uno spazio vettoriale, e U e W suoi sottospazi. Possiamo allora costruire altri sottospazi di V :

- (a) il sottospazio *intersezione* $U \cap W$;
- (b) il sottospazio *somma* $U + W = \{u + w \mid u \in U, w \in W\}$.

Ovviamente in entrambi i casi bisogna controllare che si tratti effettivamente di sottospazi. La verifica per l'intersezione è un esercizio per te; vediamo invece il caso della somma. Siano $u_1 + w_1, u_2 + w_2 \in U + W$; allora

$$(u_1 + w_1) + (u_2 + w_2) = (u_1 + u_2) + (w_1 + w_2) \in U + W$$

e

$$\lambda(u_1 + w_1) = (\lambda u_1) + (\lambda w_1) \in U + W$$

qualunque sia $\lambda \in \mathbb{R}$, per cui $U + W$ è un sottospazio.

Osservazione 4.12 Vale la pena notare che, in generale, l'unione insiemistica $U \cup W$ non è un sottospazio. Per esempio, se $V = \mathbb{R}^2$ e U e W sono due distinte rette per l'origine, $U \cup W$ non è chiuso rispetto alla somma, per cui non è un sottospazio di \mathbb{R}^2 .

Un problema naturale a questo punto è se sia o meno possibile calcolare la dimensione di $U + W$ e $U \cap W$ a partire dalle dimensioni di U e W . Per rispondere ci serve il seguente

Lemma 4.15 Siano U, W sottospazi di uno spazio vettoriale V , B un sistema di generatori di U e C un sistema di generatori di W . Allora $B \cup C$ è un sistema di generatori di $U + W$.

Dimostrazione. Ogni elemento di U (rispettivamente, W) si scrive come combinazione lineare di vettori di B (rispettivamente C). Quindi il generico elemento di $U + W$ si scrive come somma di una combinazione lineare di vettori di B e di una combinazione lineare di vettori di C — ovvero come combinazione lineare di vettori di $B \cup C$. \square

Osservazione 4.13 Se B è una base di U e C è una base di W , non è detto che $B \cup C$ sia una base di $U + W$, in quanto i vettori di $B \cup C$ potrebbero essere linearmente dipendenti (vedi l'Esercizio 4.31).

La differenza fra $\dim U + \dim W$ e $\dim(U + W)$ è data dalla dimensione dell'intersezione. Questo è il contenuto dell'importante Teorema di Grassmann:

Teorema 4.16 (Grassmann) Siano U e W sottospazi di uno spazio vettoriale di dimensione finita V . Allora

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

Dimostrazione. Sia $\{v_1, \dots, v_p\}$ una base di $U \cap W$, dove $p = \dim(U \cap W)$. Usando il Teorema 4.10, possiamo completarla in U a una base $\{v_1, \dots, v_p, u_{p+1}, \dots, u_r\}$ di U con $r = \dim U$, e possiamo completarla in W a una base $\{v_1, \dots, v_p, w_{p+1}, \dots, w_s\}$ di W dove $s = \dim W$. Allora ci basta dimostrare che

$$B = \{v_1, \dots, v_p, u_{p+1}, \dots, u_r, w_{p+1}, \dots, w_s\}$$

è una base di $U + W$; infatti, in tal caso avremo $\dim(U + W) = p + (r - p) + (s - p) = r + s - p$

$$\dim(U + W) = p + (r - p) + (s - p) = r + s - p = \dim U + \dim W - \dim(U \cap W),$$

come voluto.

Il Lemma 4.15 ci assicura già che B è un sistema di generatori di $U + W$; dobbiamo solo dimostrare che sono linearmente indipendenti. Supponiamo allora che

$$\alpha_1 v_1 + \dots + \alpha_p v_p + \beta_{p+1} u_{p+1} + \dots + \beta_r u_r + \gamma_{p+1} w_{p+1} + \dots + \gamma_s w_s = 0; \quad (4.9)$$

dobbiamo dimostrare che tutti gli α_i , β_j e γ_k sono zero. Consideriamo i tre vettori

$$\begin{aligned} v &= \alpha_1 v_1 + \dots + \alpha_p v_p \in U \cap W, \\ u &= \beta_{p+1} u_{p+1} + \dots + \beta_r u_r \in U, \\ w &= \gamma_{p+1} w_{p+1} + \dots + \gamma_s w_s \in W. \end{aligned}$$

$$w = -u - v$$

La nostra ipotesi è $v + u + w = 0$; questo implica che $w = -u - v \in U$, in quanto $u \in U$ e $v \in U \cap W \subseteq U$. Ne segue che $w \in U \cap W$ per cui si deve poter scrivere come combinazione lineare dei v_j , diciamo $w = \delta_1 v_1 + \dots + \delta_p v_p$. Ricordando com'era stato definito w troviamo

$$\gamma_{p+1} w_{p+1} + \dots + \gamma_s w_s = \delta_1 v_1 + \dots + \delta_p v_p$$

essendo $v_1, \dots, v_p, w_{p+1}, \dots, w_s$ linearmente indipendenti, questo implica

$$\gamma_{p+1} = \dots = \gamma_s = \delta_1 = \dots = \delta_p = 0.$$

A questo punto è fatta: infatti ora (4.9) ci dice che

$$\alpha_1 v_1 + \cdots + \alpha_p v_p + \beta_{p+1} u_{p+1} + \cdots + \beta_r u_r = 0,$$

e l'indipendenza lineare di $v_1, \dots, v_p, u_{p+1}, \dots, u_r$ ci permette di concludere che

$$\alpha_1 = \cdots = \alpha_p = \beta_{p+1} = \cdots = \beta_r = 0,$$

come desiderato. \square

Osservazione 4.14 Per trovare una base di $U + W$, basta unire una base di U a una di W , e poi estrarre un insieme massimale di vettori linearmente indipendenti. Vedremo nel Paragrafo 6.3 un metodo generale per farlo — assieme anche a un modo per trovare una base di $U \cap W$.

4.6 Somme dirette

Se U e W sono sottospazi di uno spazio vettoriale V tali che $U \cap W = \{O\}$, il Teorema di Grassmann ci dice che

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

In questo caso scriveremo $U \oplus W$ e diremo che $U \oplus W$ è la *somma diretta* di U e W .

Definizione 4.11 Due sottospazi U e W di uno spazio vettoriale V si dicono *supplementari* (e l'uno un *supplementare* dell'altro) se $V = U \oplus W$.

Questa è una situazione importante, perché quando accade ogni vettore di V si scrive in modo unico come somma di un vettore di U e di uno di W :

Proposizione 4.17 Siano U e W due sottospazi di uno spazio vettoriale V tali che $U \cap W = \{O\}$, e prendiamo $u_1, u_2 \in U$ e $w_1, w_2 \in W$. Allora $u_1 + w_1 = u_2 + w_2$ se e solo se $u_1 = u_2$ e $w_1 = w_2$.

Dimostrazione. Infatti $u_1 + w_1 = u_2 + w_2$ implica $u_1 - u_2 = w_2 - w_1 \in U \cap W$; la tesi segue allora da $U \cap W = \{O\}$. \square

I supplementari esistono sempre:

Proposizione 4.18 Ogni sottospazio U di uno spazio vettoriale V di dimensione finita⁶ ha un supplementare.

Dimostrazione. Sia $\{u_1, \dots, u_p\}$ una base di U ; possiamo completarla a una base $\{u_1, \dots, u_p, w_{p+1}, \dots, w_n\}$ di V e porre $W = \text{Span}(w_{p+1}, \dots, w_n)$. Chiaramente si

⁶ E non solo: vedi l'Esercizio 4C.3.

ha $V = U + W$; dobbiamo solo dimostrare che $U \cap W = \{O\}$. Ma infatti $\dim U = p$ e $\dim W = n - p$ per costruzione; quindi il Teorema 4.16 ci dice che

$$\dim(U \cap W) = \dim U + \dim W - \dim(U + W) = p + (n - p) - n = 0,$$

per cui $U \cap W = \{O\}$. \square

Osservazione 4.15 La dimostrazione della Proposizione 4.18 fornisce anche un metodo per trovare un supplementare di un sottospazio U . Si prende una base $\{u_1, \dots, u_p\}$ di U e la si completa a una base $\{u_1, \dots, u_p, w_{p+1}, \dots, w_n\}$ di V : il supplementare cercato è $W = \text{Span}(w_{p+1}, \dots, w_n)$. Come conseguenza, il supplementare di U si guarda bene dall'essere unico (perché?).

Concludiamo il capitolo definendo il concetto di somma diretta di più sottospazi.

Definizione 4.12 Siano V_1, \dots, V_k sottospazi di uno spazio vettoriale V . Indichiamo con $V_1 + \cdots + V_k$ il *sottospazio somma*

$$V_1 + \cdots + V_k = \{v_1 + \cdots + v_k \in V \mid v_1 \in V_1, \dots, v_k \in V_k\};$$

nota che $V_1 + \cdots + V_k = (V_1 + \cdots + V_{k-1}) + V_k$ (perché?). Diremo che V è *somma diretta* di V_1, \dots, V_k , e scriveremo $V = V_1 \oplus \cdots \oplus V_k$, se ogni elemento non nullo di V si scrive in maniera unica come somma di elementi di V_1, \dots, V_k .

Guardando la Proposizione 4.17 si vede che questa definizione effettivamente generalizza quella già data per due sottospazi. Possiamo comunque riscriverla anche in termini di intersezioni:

Proposizione 4.19 Siano V_1, \dots, V_k sottospazi di uno stesso spazio vettoriale di dimensione finita V . Allora:

(i) $V = V_1 \oplus \cdots \oplus V_k$ se e solo se $V = V_1 + \cdots + V_k$ e

$$V_i \cap (V_1 + \cdots + V_{i-1} + V_{i+1} + \cdots + V_k) = \{O\} \quad (4.10)$$

per $i = 1, \dots, k$;
(ii) se (4.10) vale, allora $\dim(V_1 + \cdots + V_k) = \dim V_1 + \cdots + \dim V_k$.

Dimostrazione. (i) Ogni elemento non nullo di V si scrive come somma di elementi di V_1, \dots, V_k se e solo se $V = V_1 + \cdots + V_k$; quindi dobbiamo far vedere che la scrittura è unica se e solo se (4.10) vale.

Supponiamo che (4.10) sia verificata, e prendiamo $v_i, w_i \in V_i$ (per $i = 1, \dots, k$) tali che

$$v_1 + \cdots + v_k = w_1 + \cdots + w_k.$$

Allora per $i = 1, \dots, k$ il vettore

$$v_i - w_i = (w_1 - v_1) + \cdots + (w_{i-1} - v_{i-1}) + (w_{i+1} - v_{i+1}) + \cdots + (w_k - v_k)$$

appartiene a $V_i \cap (V_1 + \dots + V_{i-1} + V_{i+1} + \dots + V_k) = \{O\}$, per cui $w_i = v_i$ e la scrittura è unica, come volevamo.

Viceversa, se $v \in V_i \cap (V_1 + \dots + V_{i-1} + V_{i+1} + \dots + V_k)$ possiamo scrivere

$$v_1 + \dots + v_{i-1} + v_{i+1} + \dots + v_k = v \in V_i$$

per opportuni $v_1 \in V_1, \dots, v_k \in V_k$; quindi abbiamo scritto v in due modi diversi come somma di elementi di V_1, \dots, V_k , per cui $v = O$.

(ii) Per induzione su k . Per $k = 1$ non c'è nulla da dimostrare; sia vero per $k - 1$. Allora (4.10), il Teorema di Grassmann e l'ipotesi induttiva ci dicono che

$$\begin{aligned} \dim(V_1 + \dots + V_k) &= \dim((V_1 + \dots + V_{k-1}) + V_k) = \dim(V_1 + \dots + V_{k-1}) + \dim V_k \\ &= \dim V_1 + \dots + \dim V_{k-1} + \dim V_k. \end{aligned}$$

□

Esercizi

4.1 Sia V uno spazio vettoriale. Dimostra che $0v = O$ per ogni $v \in V$ utilizzando soltanto le altre proprietà della definizione di spazio vettoriale. (*Suggerimento*: parti da $0 + 0 = 0$.)

4.2 Sia V uno spazio vettoriale. Dimostra che $(-1)v + v = O$ e che $(-1)v = -v$ per ogni $v \in V$. Come conseguenza, dimostra che se $v, v_1, v_2 \in V$ sono tali che $v_1 + v = O = v_2 + v$, allora $v_1 = v_2$; in altre parole, l'opposto di un vettore è univocamente determinato.

4.3 Sia V uno spazio vettoriale sul campo \mathbb{K} , e prendiamo $v \in V$ e $\lambda \in \mathbb{K}$. Dimostra che $\lambda v = O$ se e solo se $\lambda = 0$ oppure $v = O$.

4.4 Fissiamo un riferimento affine in \mathcal{A}^3 , in modo da identificare lo spazio euclideo con \mathbb{R}^3 . Dimostra che $\{O\}$, le rette per l'origine, i piani per l'origine ed \mathbb{R}^3 sono tutti e soli i sottospazi di \mathbb{R}^3 .

4.5 Dimostra che $\mathbb{R}_n[t]$, l'insieme dei polinomi a coefficienti reali in una variabile di grado minore o uguale a n , è un sottospazio vettoriale di $\mathbb{R}[t]$.

4.6 Dimostra che l'insieme delle matrici diagonali (rispettivamente, triangolari superiori o triangolari inferiori) è un sottospazio vettoriale dello spazio delle matrici quadrate.

4.7 Per quali polinomi $p \in \mathbb{R}[t]$ il grafico $\Gamma = \{(t, p(t)) \mid t \in \mathbb{R}\}$ è un sottospazio vettoriale di \mathbb{R}^2 ?

4.8 Sia $V_c = \{p \in \mathbb{R}_7[t] \mid p(1) = c\}$. Per quali valori di $c \in \mathbb{R}$ l'insieme V_c è un sottospazio vettoriale di $\mathbb{R}_7[t]$?

4.9 Sia $S \subseteq V$ un sottoinsieme qualunque (anche infinito) di uno spazio vettoriale V . Dimostra che $\text{Span}(S)$ è un sottospazio di V .

4.10 Sia $S \subseteq V$ un sottoinsieme qualunque di uno spazio vettoriale. Dimostra che $\text{Span}(S)$ è il più piccolo sottospazio di V che contiene S , nel senso che se W è un sottospazio di V che contiene S allora $W \supseteq \text{Span}(S)$.

4.11 Dimostra che

$$\left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}, \quad \left\{ \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix} \right\} \quad \text{e} \quad \left\{ \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\}$$

sono basi di \mathbb{R}^2 .

4.12 Trova le coordinate dei vettori $\begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 7 \\ -7 \end{pmatrix}$ e $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ rispetto alle basi di \mathbb{R}^2 del precedente esercizio.

4.13 Considera

$$\mathcal{B}_1 = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\} \quad \text{e} \quad \mathcal{B}_2 = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} \right\};$$

dimostra che \mathcal{B}_1 è una base di \mathbb{R}^3 mentre \mathcal{B}_2 non lo è.

4.14 Trova le coordinate del vettore $\begin{pmatrix} 2 \\ 1 \\ 7 \end{pmatrix} \in \mathbb{R}^3$ rispetto alla base \mathcal{B}_1 del precedente esercizio.

4.15 Per quali valori di $t \in \mathbb{R}$ l'insieme $\left\{ \begin{pmatrix} 2 \\ t \end{pmatrix}, \begin{pmatrix} t \\ 2 \end{pmatrix} \right\}$ è una base di \mathbb{R}^2 ?

4.16 Dimostra che $\left\{ \begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix} \right\}$ è una base di \mathbb{R}^2 se e solo se $ad - bc \neq 0$.

4.17 Prendiamo i polinomi $p_1(t) = 1 + t$, $p_2(t) = 1 + 2t + t^2$, $p_3(t) = t - t^2 \in \mathbb{R}_2[t]$. Dimostra che $\{p_1, p_2, p_3\}$ è una base di $\mathbb{R}_2[t]$, e trova le coordinate di $q_1(t) = 2 - t + t^2$ e $q_2(t) = 3 + t^2$ rispetto a questa base.

4.18 Sia $E_{ij} \in M_{m,n}(\mathbb{R})$ la matrice che ha 1 al posto (i, j) e 0 altrove. Dimostra che $\{E_{11}, \dots, E_{mn}\}$ è una base di $M_{m,n}(\mathbb{R})$; in particolare, $\dim M_{m,n}(\mathbb{R}) = mn$.

4.19 Dimostra che due vettori di \mathcal{V}_2^3 sono linearmente indipendenti se e solo se non sono proporzionali, e che tre vettori di \mathcal{V}_3^3 sono linearmente indipendenti se e solo se non sono complanari.

4.20 Sia \mathcal{B} un sottoinsieme finito massimale in uno spazio vettoriale V di vettori linearmente indipendenti. Dimostra che \mathcal{B} è una base di V .

4.21 Sia V uno spazio vettoriale di dimensione n , e $\mathcal{B} = \{v_1, \dots, v_n\}$ un sistema di generatori di V . Dimostra che \mathcal{B} è una base di V .

4.22 Sia V uno spazio vettoriale di dimensione n , e prendiamo $w_1, \dots, w_p \in V$. Dimostra che se $p < n$ allora $\{w_1, \dots, w_p\}$ non è un sistema di generatori di V .

4.23 Dimostra che un sistema lineare omogeneo con più incognite che equazioni ha sempre soluzioni non banali.

4.24 Sia V uno spazio vettoriale di dimensione n , e $\mathcal{A} = \{v_1, \dots, v_k\} \subset V$. Dimostra che \mathcal{A} è un sistema di generatori di V se e solo se $k \geq n$ e \mathcal{A} contiene n vettori linearmente indipendenti.

4.25 Sia \mathbb{K} un campo. Dimostra che \mathbb{K} è uno spazio vettoriale su se stesso di dimensione 1.

4.26 Dimostra che $\mathbb{R}[t]$ ha dimensione infinita.

4.27 Siano V e W due spazi vettoriali. Definisci una somma e un prodotto per scalari sul prodotto cartesiano $V \times W$ in modo da renderlo uno spazio vettoriale. Se $\dim V = m$ e $\dim W = n$, calcola la dimensione di $V \times W$.

4.28 Dimostra che $U + W$ è il più piccolo sottospazio di V contenente sia U che W .

4.29 Sia V uno spazio vettoriale, e U, W due sottospazi. Dimostra che $U \cup W$ è un sottospazio se e solo se $U \subseteq W$ oppure $W \subseteq U$. Nel primo caso, $U + W = U \cup W = W$, mentre nel secondo $U + W = U \cup W = U$.

4.30 Trova due sottoinsiemi di \mathbb{R}^2 , il primo chiuso rispetto alla somma ma non rispetto al prodotto per scalari, e il secondo chiuso rispetto al prodotto per scalari ma non rispetto alla somma.

4.31 Trova quattro vettori $u_1, u_2, w_1, w_2 \in \mathbb{R}^3$ tali che u_1 e u_2 siano linearmente indipendenti — per cui sono una base di $U = \text{Span}(u_1, u_2)$ —, w_1 e w_2 siano linearmente indipendenti — per cui sono una base di $W = \text{Span}(w_1, w_2)$ — ma presi tutti e quattro insieme non sono linearmente indipendenti — per cui non formano una base di $U + W$.

4.32 Siano U e W sottospazi di uno spazio vettoriale V , con $V = U \oplus W$. Dimostra che se B è una base di U e C una base di W , allora $B \cup C$ è una base di V .

4.33 Sia $v_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \in \mathbb{R}^2$; trova quattro supplementari distinti del sottospazio $\mathbb{R}v_0$ in \mathbb{R}^2 .

4.34 Siano

$$U = \text{Span} \left(\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right) \quad \text{e} \quad W = \text{Span} \left(\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right)$$

due sottospazi di \mathbb{R}^3 . Dimostra che $\mathbb{R}^3 = U \oplus W$.

4.35 Sia U un sottospazio di uno spazio vettoriale di dimensione finita V . Siano W_1 e W_2 altri sottospazi di V tali che $U \cap W_1 = \{O\}$, $U + W_2 = V$ e

$$\dim W_1 = \dim W_2 = \dim V - \dim U.$$

Dimostra che sia W_1 che W_2 sono supplementari di U .

4.36 Dimostra che ogni spazio vettoriale di dimensione n si può scrivere come somma diretta di n sottospazi di dimensione uno.

4.37 Sia V uno spazio vettoriale di dimensione finita. Una *bandiera* di V è una successione V_0, \dots, V_n di sottospazi di V tali che $V_0 \subset V_1 \subset \dots \subset V_n$. Dimostra che $\dim V_j = j$ per $j = 1, \dots, n$ (per cui $V_0 = \{O\}$ e $V_n = V$), e che ogni sottospazio di V è contenuto in una bandiera.

4.38 Indichiamo con $\mathbb{R}^{\mathbb{N}}$ l'insieme delle successioni (a_0, a_1, a_2, \dots) di elementi di \mathbb{R} . Definisci una somma e un prodotto per scalari in $\mathbb{R}^{\mathbb{N}}$ in modo da renderlo uno spazio vettoriale.

4.39 Indichiamo con ℓ^∞ il sottoinsieme di $\mathbb{R}^{\mathbb{N}}$ delle successioni limitate: la successione $(a_0, a_1, \dots) \in \mathbb{R}^{\mathbb{N}}$ appartiene a ℓ^∞ se e solo se esiste $M > 0$ tale che $|a_j| < M$ per ogni $j \in \mathbb{N}$. Dimostra che ℓ^∞ è un sottospazio vettoriale di $\mathbb{R}^{\mathbb{N}}$.

COMPLEMENTI

4C.1 Il Lemma di Zorn

Lo scopo di questi Complementi è introdurre due strumenti matematici di notevole importanza, il *Lemma di Zorn* e la teoria dei cardinali infiniti, e descriverne alcune applicazioni alla teoria degli spazi vettoriali di dimensione infinita. Lungo la strada, introdurremo delle nozioni e dei teoremi (spesso senza dimostrazione) assolutamente fondamentali nella teoria degli insiemi.

Cominciamo col descrivere il tipo di ambiente in cui vive il Lemma di Zorn.

Definizione 4C.1 Sia A un insieme qualunque; una *relazione d'ordine* \leq su A è una relazione binaria (vedi la Definizione 2C.1) definita su A con le proprietà seguenti:

- Riflessività*: $x \leq x$ per ogni $x \in A$;
- Antisimmetria*: se $x, y \in A$ sono tali che $x \leq y$ e $y \leq x$, allora $x = y$;
- Transitività*: se $x, y, z \in A$ sono tali che $x \leq y$ e $y \leq z$, allora $x \leq z$.

La coppia (A, \leq) è detta *insieme ordinato*.

Esempi di insiemi ordinati sono i numeri naturali (interi, razionali, reali) con la solita relazione d'ordine. Un altro esempio, con caratteristiche diverse, è $(\mathcal{P}(A), \subseteq)$, dove A è un insieme qualunque, e

$$\mathcal{P}(A) = \{B \mid B \text{ sottoinsieme di } A\}$$

è l'insieme delle parti di A .

Definizione 4C.2 Un insieme ordinato (A, \leq) è detto *totalmente ordinato* se due qualunque elementi di A sono confrontabili, cioè se per ogni $x, y \in A$ si ha $x \leq y$ oppure $y \leq x$.

Intuitivamente, gli elementi di un insieme totalmente ordinato sono disposti in fila uno dietro l'altro; per esempio, i numeri reali con l'ordinamento canonico sono totalmente ordinati. Invece l'insieme $\mathcal{P}(A)$ non è totalmente ordinato rispetto all'inclusione: se $A = \{0, 1\}$, allora $\{0\}$ non è contenuto né contiene $\{1\}$, per cui non sono confrontabili.

I sottoinsiemi totalmente ordinati di un insieme ordinato hanno un nome:

Definizione 4C.3 Sia (A, \leq) un insieme ordinato. Un sottoinsieme totalmente ordinato di A è detto *catena*.

Per esempio, $\{\emptyset, \{0\}, \{0, 1\}\}$ è una catena di $\mathcal{P}(\{0, 1\})$ rispetto all'inclusione.

Definizione 4C.4 Sia (A, \leq) un insieme ordinato, e $B \subseteq A$. Diremo che $x \in B$ è il *massimo* (*minimo*) di B se $y \leq x$ (rispettivamente, $x \leq y$) per ogni $y \in B$. Chiaramente, non è detto che il massimo o il minimo esistano (si prenda per esempio $A = \mathbb{R}$ e $B = (0, 1)$ con la solita relazione d'ordine). Un insieme ordinato (A, \leq) è detto *bene ordinato* se ogni sottoinsieme di A ammette minimo.

Per esempio, \mathbb{N} con l'ordinamento canonico è bene ordinato (vedi i Complementi al Capitolo 3), mentre \mathbb{Q} non lo è (l'insieme dei numeri razionali x tali che $x^2 \geq 2$ non ha minimo, come pure l'insieme dei numeri razionali strettamente positivi).

Osservazione 4C.1 Ogni insieme bene ordinato è totalmente ordinato: se $x, y \in A$, il sottoinsieme $B = \{x, y\}$ di A deve avere minimo, per cui $x \leq y$ oppure $y \leq x$.

La maggior parte degli assiomi della teoria assiomatica degli insiemi sono accettati come naturali senza discussione da parte di tutti i matematici. La principale eccezione è il famigerato

Assioma di scelta Sia $\{A_i\}_{i \in I}$ una famiglia di insiemi non vuoti a due a due disgiunti, dipendente da un indice $i \in I$. Allora esiste una funzione $f: I \rightarrow \bigcup_{i \in I} A_i$ tale che $f(i) \in A_i$ per ogni $i \in I$.

In altri termini, l'Assioma di scelta dice che data una famiglia di insiemi non vuoti a due a due disgiunti possiamo scegliere simultaneamente un elemento da ogni insieme della famiglia. Nota che l'Assioma di scelta ci dice che possiamo farlo, ma non *come* possiamo farlo: questa è una delle principali obiezioni sollevate dai suoi oppositori (un'altra è che l'Assioma di scelta afferma che possiamo eseguire un'infinità di operazioni simultaneamente, cosa che non sembra essere umanamente possibile).

Nei primi anni sessanta è stato dimostrato che la teoria degli insiemi se è priva di contraddizioni senza l'Assioma di scelta, lo è anche con l'Assioma di scelta⁷ (proprio come per la geometria euclidea con o senza l'assioma delle parallele), per cui accettare

⁷ Non è possibile dimostrare che la teoria degli insiemi è non contraddittoria; questo è uno dei fatti che circolano attorno al famoso Teorema di Gödel.

o meno l'Assioma di scelta è più che altro questione di gusto personale. Come vedremo anche in questi Complementi, l'Assioma di scelta serve a dimostrare alcuni risultati molto importanti, per cui noi l'accetteremo senza batter ciglio. Questa decisione è ulteriormente confortata dal fatto che l'Assioma di scelta è equivalente al seguente assioma, molto naturale:

Assioma del prodotto non vuoto Sia $\{A_i\}_{i \in I}$ una famiglia di insiemi non vuoti. Allora il prodotto cartesiano $\prod_{i \in I} A_i$ di tutti gli insiemi della famiglia è non vuoto.

L'Assioma di scelta entra anche nel contesto degli insiemi ordinati, in quanto è equivalente al seguente

Assioma del buon ordinamento Su ogni insieme esiste un buon ordinamento.

Per valutare la non ovvietà di questo assioma, prova a definire un buon ordinamento sull'insieme dei numeri reali⁸.

Il Lemma di Zorn è un'ulteriore incarnazione dell'Assioma di scelta. Per enunciarlo ci serve ancora qualche definizione.

Definizione 4C.5 Sia (A, \leq) un insieme ordinato, e B un sottoinsieme di A . Un elemento $x \in A$ è un *maggiorante* di B se $y \leq x$ per ogni $y \in B$. Un *elemento massimale* di B è un elemento $x \in B$ tale che l'unico elemento y di B tale che $x \leq y$ è x stesso.

Chiaramente, il massimo è un elemento massimale e, se B è totalmente ordinato, è anche l'unico elemento massimale; ma se B non è totalmente ordinato, possono esistere più elementi massimali.

ESEMPIO 4C.1 Se $A = \mathcal{P}(\{0, 1\})$ e $B = \{\emptyset, \{0\}, \{1\}\}$, allora sia $\{0\}$ che $\{1\}$ sono elementi massimali di B rispetto all'inclusione, mentre $\{0, 1\}$ è l'unico maggiorante di B in A .

Definizione 4C.6 Diremo che un insieme ordinato (A, \leq) è *induttivo* se ogni catena di A ha un maggiorante in A .

Allora il Lemma di Zorn è:

Lemma di Zorn Ogni insieme ordinato induttivo ha elementi massimali.

Lo schema tipico di applicazione del Lemma di Zorn è il seguente: vogliamo trovare un oggetto che sia massimale rispetto a una qualche relazione d'ordine. Allora basta limitarsi a controllare le catene: se queste ammettono maggioranti, il Lemma di Zorn ci fornisce automaticamente l'esistenza dell'elemento cercato (ma non ci dice

⁸ Non disperarti se non lo trovi: nessuno ne ha mai descritto uno esplicitamente.

minimamente come trovarlo; le dimostrazioni che dipendono dall'Assioma di scelta sono quanto di meno costruttivo ci sia in matematica).

Esempi di applicazioni del Lemma di Zorn verranno dati fra poco e, come vedrai, una volta capita la tecnica non vi è nulla di veramente difficile (è solo un po' ostico al primo approccio). Talvolta può convenire usare un'altra versione del Lemma di Zorn, del tutto equivalente:

Principio di Hausdorff Sia A un insieme, e sia E un sottoinsieme di $\mathcal{P}(A)$. Supponiamo che l'unione di ogni catena di (E, \subseteq) sia contenuta in E . Allora (E, \subseteq) possiede elementi massimali.

Il Principio di Hausdorff segue subito dal Lemma di Zorn: basta applicare quest'ultimo a (E, \subseteq) , notando che l'unione di una catena si presta mirabilmente a fare da maggiorante. Viceversa, si può dimostrare che il Principio di Hausdorff implica il Lemma di Zorn, come pure che entrambi sono equivalenti agli assiomi (di scelta, del prodotto non vuoto, del buon ordinamento) di cui abbiamo parlato in questo paragrafo, ma si tratta di dimostrazioni altamente non ovvie.

4C.2 L'esistenza di una base

Il nostro primo obiettivo è usare il Lemma di Zorn per dimostrare l'esistenza di una base in uno spazio vettoriale qualunque. Cominciamo col dare una definizione di base che funziona anche in spazi vettoriali privi di sistemi di generatori finiti.

Definizione 4C.7 Sia V uno spazio vettoriale su un campo \mathbb{K} , e \mathcal{B} un sottoinsieme di V . Diremo che \mathcal{B} è *libero* se ogni suo sottoinsieme *finito* è linearmente indipendente; diremo che è un *sistema di generatori* se $V = \text{Span}(\mathcal{B})$, cioè se per ogni $v \in V$ esistono $n \in \mathbb{N}$, $v_1, \dots, v_n \in \mathcal{B}$ e $\lambda_1, \dots, \lambda_n \in \mathbb{K}$ tali che $v = \lambda_1 v_1 + \dots + \lambda_n v_n$. Infine una *base* di V non è altro che un sistema di generatori libero.

Considereremo $\mathcal{P}(V)$ ordinato per inclusione, come al solito; allora una caratterizzazione delle basi di V è data dalla

Proposizione 4C.1 Sia V uno spazio vettoriale, e \mathcal{B} un sottoinsieme di V . Allora \mathcal{B} è una base di V se e solo se è un sottoinsieme libero massimale.

Dimostrazione. Sia \mathcal{B} un sottoinsieme libero massimale. Questo vuol dire che non esiste nessun sottoinsieme di V contenente propriamente \mathcal{B} che sia ancora libero. In altre parole, se $v \in V$ non appartiene a \mathcal{B} , allora $\mathcal{B} \cup \{v\}$ non è libero. Quindi esistono $v_1, \dots, v_n \in \mathcal{B}$ e $\lambda, \lambda_1, \dots, \lambda_n \in \mathbb{K}$ non tutti nulli tali che

$$\lambda v + \lambda_1 v_1 + \dots + \lambda_n v_n = 0.$$

Essendo \mathcal{B} libero, λ dev'essere diverso da zero. Quindi

$$v = -\frac{\lambda_1}{\lambda} v_1 - \dots - \frac{\lambda_n}{\lambda} v_n,$$

e v appartiene al sottospazio generato da \mathcal{B} . Siccome v era generico, questo vuol dire che \mathcal{B} è una base di V .

Viceversa, sia \mathcal{B} una base. Allora \mathcal{B} è libero; inoltre, siccome \mathcal{B} è un sistema di generatori, $\mathcal{B} \cup \{v\}$ non è libero, dove v è un qualunque elemento di V non appartenente a \mathcal{B} . Quindi \mathcal{B} è un sottoinsieme libero massimale, e ci siamo. \square

Siamo quindi nella situazione ideale per l'applicazione del Lemma di Zorn (o, se preferisci, del Principio di Hausdorff): trovare una base equivale a trovare un elemento massimale della famiglia dei sottoinsiemi liberi di V (nota che di basi ve ne sono miriadi, coerentemente col fatto che l'ordinamento per inclusione si guarda bene dall'essere totale). Quindi ci basterà verificare che l'unione di una catena di sottoinsiemi liberi, tutti contenuti uno dentro l'altro, è ancora un sottoinsieme libero.

Teorema 4C.2 Ogni spazio vettoriale ammette una base.

Dimostrazione. Sia V uno spazio vettoriale, e indichiamo con \mathcal{L} la famiglia dei sottoinsiemi liberi di V , ordinato per inclusione. Grazie alla Proposizione 4C.1, per trovare una base ci basta trovare un elemento massimale di \mathcal{L} ; grazie al Lemma di Zorn (o, se preferisci, al Principio di Hausdorff) ci basta dimostrare che l'unione di una qualunque catena di sottoinsiemi liberi è ancora in \mathcal{L} — ovvero è ancora libera.

Sia allora $\{\mathcal{B}_\alpha\}_{\alpha \in A}$ una catena in \mathcal{L} (ovvero ogni \mathcal{B}_α è un sottoinsieme libero di V e ogni volta che si scelgono $\alpha, \beta \in A$ abbiamo $\mathcal{B}_\alpha \subseteq \mathcal{B}_\beta$ o $\mathcal{B}_\beta \subseteq \mathcal{B}_\alpha$), e sia $\mathcal{B} = \bigcup_{\alpha \in A} \mathcal{B}_\alpha$.

Supponiamo vi siano $v_1, \dots, v_n \in \mathcal{B}$ tali che

$$\lambda_1 v_1 + \dots + \lambda_n v_n = 0 \quad (4C.1)$$

per certi $\lambda_1, \dots, \lambda_n \in \mathbb{K}$. Siccome \mathcal{B} è l'unione dei \mathcal{B}_α , esistono $\alpha_1, \dots, \alpha_n \in A$ tali che $v_j \in \mathcal{B}_{\alpha_j}$ per $j = 1, \dots, n$. Ma $\{\mathcal{B}_\alpha\}_{\alpha \in A}$ è una catena; quindi uno degli \mathcal{B}_{α_j} (che possiamo supporre, per semplicità, essere \mathcal{B}_{α_n}) contiene tutti gli altri. In particolare, quindi, $v_1, \dots, v_n \in \mathcal{B}_{\alpha_n}$. Ora, \mathcal{B}_{α_n} è libero; quindi (4C.1) implica

$$\lambda_1 = \dots = \lambda_n = 0.$$

Quindi ogni sottoinsieme finito di \mathcal{B} è linearmente indipendente, ovvero \mathcal{B} è libero, e la dimostrazione è conclusa. \square

In modo analogo, possiamo dimostrare il Teorema di completamento.

Teorema 4C.3 Sia V uno spazio vettoriale, e \mathcal{B} un sottoinsieme libero di V . Allora esiste una base di V contenente \mathcal{B} .

Dimostrazione. Sia \mathcal{L} la famiglia dei sottoinsiemi liberi di V contenenti \mathcal{B} . \mathcal{L} è non vuota, in quanto contiene \mathcal{B} ; inoltre, un elemento massimale di \mathcal{L} è chiaramente un sottoinsieme libero massimale contenente \mathcal{B} , ovvero una base di V come da noi cercata. Siccome l'unione di una catena di elementi di \mathcal{L} ovviamente contiene \mathcal{B} , ed è libera per quanto visto nella dimostrazione del Teorema 4C.2, il Lemma di Zorn ci assicura l'esistenza di un elemento massimale di \mathcal{L} , e abbiamo finito. \square

4C.3 Cardinalità, e dimensione degli spazi vettoriali

Per gli spazi vettoriali di dimensione finita vi è un altro fatto fondamentale riguardante le basi: due basi di uno stesso spazio vettoriale hanno sempre lo stesso numero di elementi. Per poterlo dimostrare in generale dobbiamo chiarire quando due insiemi hanno lo stesso numero di elementi (in modo da poter poi definire la dimensione di un qualunque spazio vettoriale). L'idea è che due insiemi hanno lo stesso numero di elementi quando è possibile accoppiare a due a due tutti gli elementi dei due insiemi.

Definizione 4C.8 Diremo che due insiemi X e Y hanno la stessa cardinalità (ovvero che sono *equipotenti*) se esiste una applicazione bigettiva $f: X \rightarrow Y$, e scriveremo $|X| = |Y|$. Più in generale, diremo che X ha *cardinalità minore* (maggiore) di Y se esiste una funzione iniettiva (rispettivamente, surgettiva) $f: X \rightarrow Y$, e scriveremo $|X| \leq |Y|$ (rispettivamente, $|X| \geq |Y|$).

Perché sia lecito utilizzare i simboli di minore, maggiore e uguaglianza bisogna che siano verificate alcune proprietà. Prima di tutto, la relazione d'equipotenza dev'essere una relazione d'equivalenza (vedi la Definizione 2C.2), che è ovvio. In secondo luogo, $|X| \leq |Y|$ dev'essere equivalente a $|Y| \geq |X|$ (vedi l'Esercizio 4C.4). Infine, \leq dev'essere una relazione d'ordine sulla classe⁹ delle cardinalità. Infatti, è chiaramente riflessiva e transitiva (Esercizio 4C.5), ed è antisimmetrica grazie al

Teorema di Cantor-Bernstein Siano X e Y due insiemi tali che esistano due applicazioni $f: X \rightarrow Y$ e $g: Y \rightarrow X$ entrambe iniettive. Allora esiste una applicazione $h: X \rightarrow Y$ bigettiva.

Se X e Y sono due insiemi tali che $|X| \leq |Y|$ ma $|X| \neq |Y|$, scriveremo $|X| < |Y|$ oppure $|Y| > |X|$.

Osservazione 4C.2 Una conseguenza dell'Esercizio 4C.1 è che \leq è un ordine totale sulla classe delle cardinalità.

Ora, quando si contano gli elementi di un insieme, la prima cosa che si vorrebbe sapere è se prima o poi si finisce, ovvero se l'insieme è finito.

Definizione 4C.9 Diremo che un insieme X è *infinito* se è equipotente a un suo sottoinsieme proprio; diremo che è *finito* altrimenti.

⁹ Utilizziamo il termine classe al posto di quello d'insieme, perché il concetto di "insieme di tutte le cardinalità" non è ben definito (soffre dello stesso tipo di problemi del concetto di "insieme di tutti gli insiemi", che se esistesse sarebbe elemento di se stesso, dando vita a un circolo vizioso di proporzioni cosmiche). Per questo motivo, parlare di relazioni d'ordine fra cardinalità sarebbe lievemente scorretto, ma è comunque sufficientemente suggestivo da valerne la pena. Tra parentesi, un elemento della classe della cardinalità viene chiamato *cardinale*.

Per verificare che questa definizione cattura la nostra intuizione del finito, dobbiamo controllare che gli insiemi che noi usualmente pensiamo essere infiniti lo sono anche in questo senso, mentre quelli finiti non lo sono. Infatti, $\{1, \dots, n\}$ è finito per ogni $n \in \mathbb{N}$ (esercizio); d'altro canto, \mathbb{N} è infinito (è equipotente al sottoinsieme dei numeri pari, tramite l'applicazione bigettiva $f(n) = 2n$).

La cardinalità di \mathbb{N} è usualmente indicata con \aleph_0 (letto "alef zero"), che è un cardinale particolarmente interessante:

Proposizione 4C.4 \aleph_0 è il più piccolo cardinale infinito.

Dimostrazione. Sia A un insieme infinito: dobbiamo costruire un'applicazione iniettiva $f: \mathbb{N} \rightarrow A$. Per ogni $n \in \mathbb{N}$ indichiamo con \mathbf{n} l'insieme¹⁰ $\{0, \dots, n\}$, e poniamo

$$\mathcal{F}_n = \{g: \mathbf{n} \rightarrow A \mid g \text{ iniettiva}\}.$$

Siccome A è infinito, è facile dimostrare per induzione che ogni \mathcal{F}_n è non vuoto. Con l'Assioma di scelta peschiamo un $f_n \in \mathcal{F}_n$ per ogni $n \in \mathbb{N}$, e definiamo $f: \mathbb{N} \rightarrow A$ ponendo

$$f(0) = f_0(0) \quad \text{e} \quad f(n) = f_n(k)$$

dove $k = \min\{i \mid f_n(i) \neq f(j) \text{ per } 1 \leq j \leq n-1\}$. È chiaro che la f così definita è iniettiva. \square

Come con ogni nuova teoria che si rispetti, dobbiamo essere in grado di provare i nuovi concetti sui vecchi oggetti, e confrontare, per esempio, la cardinalità degli insiemi numerici con quella di \mathbb{N} . I risultati sono:

Proposizione 4C.5 Valgono le seguenti relazioni fra cardinalità:

- (i) $|\mathbb{Z}| = \aleph_0$.
- (ii) $|\mathbb{Q}| = \aleph_0$.
- (iii) $|\mathbb{R}| > \aleph_0$.
- (iv) $|\mathbb{R} \times \mathbb{R}| = |\mathbb{R}|$, ovvero vi sono tanti punti nel piano quanti ve ne sono in una retta.
- (v) $|\mathbb{R}| = |\mathcal{P}(\mathbb{N})|$.

Dimostrazione. (i) Un'applicazione bigettiva da \mathbb{Z} su \mathbb{N} è data da

$$f(i) = \begin{cases} 2i & \text{se } i \geq 0; \\ -2i - 1 & \text{se } i < 0. \end{cases}$$

(ii) Chiaramente, $|\mathbb{Q}| \geq \aleph_0$, in quanto \mathbb{N} è contenuto in \mathbb{Q} ; quindi è sufficiente dimostrare che $|\mathbb{Q}| \leq \aleph_0$. Sia $f: \mathbb{Q} \rightarrow \mathbb{Z} \times \mathbb{N}$ data da $f(r) = (p, q)$, dove $r = p/q$ è la rappresentazione di r ridotta ai minimi termini, e $q > 0$. Chiaramente f è iniettiva, per cui $|\mathbb{Q}| \leq |\mathbb{Z} \times \mathbb{N}|$. Grazie a (i), abbiamo $|\mathbb{Z} \times \mathbb{N}| = |\mathbb{N} \times \mathbb{N}|$ (esercizio); quindi

¹⁰ Ogni insieme finito è equipotente a uno degli insiemi \mathbf{n} : vedi l'Esercizio 4C.7.

è sufficiente dimostrare che $|\mathbb{N} \times \mathbb{N}| = \aleph_0$, cioè costruire una applicazione bigettiva $f: \mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N}$ (ovvero, in termini meno formali, contare gli elementi di $\mathbb{N} \times \mathbb{N}$).

Sistemiamo gli elementi di $\mathbb{N} \times \mathbb{N}$ nella seguente matrice:

$$\begin{array}{cccc}
 (0,0) \rightarrow (0,1) & (0,2) \rightarrow (0,3) & \dots & \\
 \swarrow & \nearrow & \swarrow & \nearrow \\
 (1,0) & (1,1) & (1,2) & (1,3) \dots \\
 \downarrow & \nearrow & \swarrow & \nearrow \\
 (2,0) & (2,1) & (2,2) & (2,3) \dots \\
 \swarrow & \nearrow & \swarrow & \nearrow \\
 (3,0) & (3,1) & (3,2) & (3,3) \dots \\
 \vdots & \vdots & \vdots & \vdots
 \end{array}$$

Allora una applicazione bigettiva da \mathbb{N} a $\mathbb{N} \times \mathbb{N}$ è ottenuta contando gli elementi di $\mathbb{N} \times \mathbb{N}$ nell'ordine indicato dalle frecce.

(iii) Ovviamente $|\mathbb{R}| \geq \aleph_0$. Supponiamo, per assurdo, che esista una applicazione surgettiva $f: \mathbb{N} \rightarrow \mathbb{R}$. Per semplicità, componiamo f con l'applicazione bigettiva $g: \mathbb{R} \rightarrow (0,1)$ data da $g(x) = \frac{1}{\pi} \arctan x + \frac{1}{2}$, in modo da ottenere un'applicazione surgettiva $h: \mathbb{N} \rightarrow (0,1)$. Questo significa che abbiamo messo tutti i numeri decimali fra 0 e 1 in una lista:

$$0.a_{00}a_{01}a_{02}a_{03} \dots$$

$$0.a_{10}a_{11}a_{12}a_{13} \dots$$

$$0.a_{20}a_{21}a_{22}a_{23} \dots$$

...

dove non sono ammesse espansioni terminanti con una successione infinita di 9. Per avere una contraddizione basta trovare un numero decimale compreso fra 0 e 1 che non sia in questa lista. Ma infatti il numero $0.b_0b_1b_2b_3 \dots$, dove $b_j = 1$ se $a_{jj} \neq 1$ e $b_j = 2$ altrimenti, è diverso da tutti i numeri della lista, per cui non vi appartiene. Quindi non può esistere alcuna applicazione surgettiva da \mathbb{N} su \mathbb{R} , per cui siamo costretti a concludere che $|\mathbb{R}| > \aleph_0$.

(iv) Per quanto già detto dimostrando (iii), ci basta far vedere che esiste un'applicazione iniettiva da $(0,1) \times (0,1)$ in $(0,1)$. Di nuovo, scriviamo i punti di $(0,1)$ nella loro espansione decimale, escludendo le successioni infinite di 9. Allora l'applicazione cercata manda la coppia $(0.a_0a_1a_2a_3 \dots, 0.b_0b_1b_2b_3 \dots)$ nel numero $0.a_0b_0a_1b_1a_2b_2 \dots$.

(v) Infatti sia l'applicazione $\varphi: \mathcal{P}(\mathbb{N}) \rightarrow (0,1)$ che associa al sottoinsieme $A \subseteq \mathbb{N}$ il numero reale $\varphi(A) = 0.a_1a_2a_3 \dots$, dove $a_j = 1$ se $j \in A$, e $a_j = 0$ altrimenti, che l'applicazione $\psi: (0,1) \rightarrow \mathcal{P}(\mathbb{N})$ che associa ad $a = 0.a_1a_2a_3 \dots$ il sottoinsieme $\psi(a) = \{a_1, 10 + a_2, 20 + a_3, \dots, 10(j-1) + a_j, \dots\} \subseteq \mathbb{N}$, sono iniettive. \square

Quindi guardando fra gli insiemi noti, abbiamo finora trovato soltanto una nuova cardinalità:

Definizione 4C.10 La cardinalità di \mathbb{R} , che coincide con quella di $\mathcal{P}(\mathbb{N})$, è detta *cardinalità del continuo*. Inoltre, un insieme equipotente a \mathbb{N} è detto *numerabile*.

Nella Matematica di tutti i giorni è raro trovare altre cardinalità: come indicato anche nell'Esercizio 4C.6 la maggior parte delle costruzioni fatte partendo dagli oggetti matematici usuali produce insiemi numerabili o con la cardinalità del continuo. Esiste però una tecnica standard per costruire nuove cardinalità:

Proposizione 4C.6 Sia A un insieme infinito. Allora $|\mathcal{P}(A)| > |A|$.

Dimostrazione. Chiaramente $|\mathcal{P}(A)| \geq |A|$. Sia $f: A \rightarrow \mathcal{P}(A)$ una qualunque applicazione da A in $\mathcal{P}(A)$, e consideriamo l'insieme

$$B = \{x \in A \mid x \notin f(x)\} \subseteq A.$$

Se mostriamo che $B \notin f(A)$ otteniamo che f non è surgettiva; essendo f qualunque, questo vuol dire che $|A| < |\mathcal{P}(A)|$, e la dimostrazione è conclusa. Supponiamo allora per assurdo che $B = f(x_0)$ per qualche $x_0 \in A$. Se fosse $x_0 \in B$, per definizione di B dovremmo avere $x_0 \notin f(x_0) = B$, contraddizione. Analogamente, se fosse $x_0 \notin B$ dovremmo avere $x_0 \in f(x_0) = B$, contraddizione. Quindi B non può essere nell'immagine di f , e abbiamo finito. \square

A questo punto viene naturale chiedersi se esistono delle cardinalità intermedie fra quella di \mathbb{N} e quella di \mathbb{R} . Se indichiamo con \aleph_1 la più piccola cardinalità strettamente più grande di \aleph_0 , la soluzione più semplice a questo problema è la seguente

Ipotesi del continuo $|\mathbb{R}| = \aleph_1$.

Sorprendentemente, questa ipotesi è esattamente nella stessa posizione dell'Assioma di scelta: la teoria degli insiemi, ammesso sia coerente, rimane tale sia con l'Ipotesi del continuo che senza di essa (indipendentemente dalla presenza dell'Assioma di scelta, fra l'altro). Anche in questo caso, quindi, accettarla o meno è solo questione di gusto; comunque ne abbiamo parlato solo per completezza, in quanto non influenza il resto di questo libro.

Quello che ci interessa adesso, invece, è la cosiddetta aritmetica dei cardinali. Tra i cardinali finiti sono definiti (in quanto grazie all'Esercizio 4C.7 i cardinali finiti sono in corrispondenza biunivoca con gli elementi di \mathbb{N}) somma e prodotto; vogliamo estendere queste operazioni a cardinali qualunque.

Definizione 4C.11 Siano α e β due cardinali, X un insieme di cardinalità α e Y un insieme di cardinalità β , con $X \cap Y = \emptyset$. Allora $\alpha \times \beta$ è la cardinalità di $X \times Y$, e $\alpha + \beta$ è la cardinalità di $X \cup Y$. Più in generale, se $\{X_i\}_{i \in I}$ è una famiglia di insiemi a due a due disgiunti, porremo $\sum_{i \in I} |X_i| = |\bigcup_{i \in I} X_i|$.

Che questa sia una buona definizione (nel senso che non dipende dagli insiemi scelti ma solo dalle cardinalità) è dimostrato nell'Esercizio 4C.9; qui invece siamo interessati alle proprietà dell'aritmetica fra cardinali, che sono alquanto diverse da quelle solite, come del resto faceva sospettare la Proposizione 4C.5. Il risultato principale, equivalente (anche lui!) all'Assioma di scelta, è il

Teorema del prodotto Se α è un cardinale infinito si ha $\alpha^2 = \alpha$.

Usando questo teorema è facile dedurre altre proprietà:

Corollario 4C.7 (i) Se α è un cardinale infinito, allora $\alpha \times \beta = \max\{\alpha, \beta\}$ per ogni cardinale β .

(ii) $\alpha^n = \alpha$ per ogni $n \in \mathbb{N}$ e ogni cardinale infinito α ;

(iii) Se $\{\alpha_i\}_{i \in I}$ è una famiglia di cardinali maggiori di zero, di cui almeno uno è infinito o, in mancanza di ciò, tale che I è infinito, allora

$$\sum_{i \in I} \alpha_i = \max\{|I|, \alpha_i\}.$$

In particolare, se α è infinito, allora $\alpha + \beta = \max\{\alpha, \beta\}$ per ogni cardinale β .

Dimostrazione. (i) Sia $\gamma = \max\{\alpha, \beta\}$. Allora

$$\gamma \preceq \alpha \times \beta \preceq \gamma^2 = \gamma.$$

(ii) Segue da (i) per induzione.

(iii) Sia $\gamma = \max\{|I|, \alpha_i\}$. Allora

$$\gamma \preceq \sum_{i \in I} \alpha_i \preceq \sum_{i \in I} \gamma \preceq |I| \times \gamma \preceq \gamma^2 = \gamma.$$

□

Un'altra conseguenza di tutto questo è vitale per gli spazi vettoriali. Se A è un insieme, indicheremo con $\mathcal{P}^*(A)$ la famiglia dei sottoinsiemi finiti di A . Allora abbiamo (confronta con la Proposizione 4C.6):

Proposizione 4C.8 Sia A un insieme infinito. Allora $|\mathcal{P}^*(A)| = |A|$.

Dimostrazione. Sia $\mathcal{P}^{(n)}(A)$ la famiglia dei sottoinsiemi di A con n elementi; in particolare, $\mathcal{P}^*(A)$ è l'unione disgiunta dei $\mathcal{P}^{(n)}(A)$. Ora, $|\mathcal{P}^{(n)}(A)| \preceq |A|^n$ (esercizio), e quindi

$$|A| \preceq |\mathcal{P}^*(A)| = \sum_{n \in \mathbb{N}} |\mathcal{P}^{(n)}(A)| \preceq \sum_{n \in \mathbb{N}} |A|^n = \sum_{n \in \mathbb{N}} |A| = \aleph_0 \times |A| = |A|,$$

dove abbiamo usato il Corollario 4C.7 e la Proposizione 4C.4. □

Siamo finalmente in grado di dimostrare che anche gli spazi vettoriali di dimensione infinita hanno una dimensione:

Teorema 4C.9 Sia V uno spazio vettoriale sul campo \mathbb{K} , e $\mathcal{B}_1, \mathcal{B}_2$ due basi di V . Allora \mathcal{B}_1 e \mathcal{B}_2 sono equipotenti.

Dimostrazione. Se una delle due basi è finita, allora anche l'altra lo è (Corollario 4.13) e la tesi segue dal Corollario 4.11. Supponiamo allora che entrambe le basi siano infinite. Costruiamo una applicazione $f: \mathcal{B}_1 \rightarrow \mathcal{P}^*(\mathcal{B}_2)$ come segue: se $v \in \mathcal{B}_1$, possiamo scrivere in modo unico

$$v = \lambda_1 w_1 + \dots + \lambda_r w_r,$$

per appropriati $\lambda_1, \dots, \lambda_r \in \mathbb{K}$ e $w_1, \dots, w_r \in \mathcal{B}_2$; poniamo $f(v) = \{w_1, \dots, w_r\}$.

Ora, prendiamo $W \in \mathcal{P}^*(\mathcal{B}_2)$ composto da n vettori, e sia $\text{Span}(W)$ il sottospazio generato da W . Chiaramente, $\text{Span}(W)$ ha dimensione n , per cui $f^{-1}(W)$ contiene al più n elementi (perché?). La base \mathcal{B}_1 è l'unione degli insiemi a due a due disgiunti $f^{-1}(W)$ al variare di W in $\mathcal{P}^*(\mathcal{B}_2)$; quindi

$$|\mathcal{B}_1| = \sum_{W \in \mathcal{P}^*(\mathcal{B}_2)} |f^{-1}(W)| \preceq \sum_{W \in \mathcal{P}^*(\mathcal{B}_2)} \aleph_0 = |\mathcal{P}^*(\mathcal{B}_2)| = |\mathcal{B}_2|,$$

dove abbiamo utilizzato il Corollario 4C.7 e la Proposizione 4C.8. Scambiando il ruolo di \mathcal{B}_1 e \mathcal{B}_2 otteniamo analogamente $|\mathcal{B}_2| \preceq |\mathcal{B}_1|$, e la tesi segue dal Teorema di Cantor-Bernstein. □

In particolare, quindi, il concetto di dimensione di uno spazio vettoriale come cardinalità di una base è ben definito per ogni spazio vettoriale, e non solo per quelli di dimensione finita.

Concludiamo questi Complementi con un esempio un po' strano, la cui rilevanza sarà chiarita nel prossimo capitolo.

ESEMPIO 4C.2 Consideriamo \mathbb{R} come spazio vettoriale su \mathbb{Q} . Essendo \mathbb{Q} numerabile mentre \mathbb{R} non lo è, l'Esercizio 4C.12 implica che \mathbb{R} ha dimensione infinita su \mathbb{Q} . Ora, \mathbb{Q} stesso è un sottospazio vettoriale (su \mathbb{Q}) di \mathbb{R} , con base $\{1\}$. Per il Teorema 4C.3, esiste un sottoinsieme libero $\mathcal{B} \subset \mathbb{R}$ tale che $\{1\} \cup \mathcal{B}$ sia una base di \mathbb{R} su \mathbb{Q} . Poniamo $W = \text{Span}(\mathcal{B})$, in modo da poter scrivere $\mathbb{R} = \mathbb{Q} \oplus W$, e definiamo $T: \mathbb{R} \rightarrow \mathbb{R}$ ponendo

$$\forall r \in \mathbb{Q} \quad \forall w \in W \quad T(r+w) = w.$$

L'applicazione T è ben definita, in quanto ogni elemento di \mathbb{R} si scrive in modo unico come somma di un elemento di \mathbb{Q} e di un elemento di W . Inoltre, per ogni $x_1 = r_1 + w_1, x_2 = r_2 + w_2 \in \mathbb{R}$ si ha

$$T(x_1 + x_2) = T(r_1 + r_2 + w_1 + w_2) = w_1 + w_2 = T(r_1 + w_1) + T(r_2 + w_2) = T(x_1) + T(x_2).$$

D'altra parte, però, se $\lambda \in \mathbb{R} \setminus \mathbb{Q}$ e $r \in \mathbb{Q}$, si ha che $\lambda r \notin \mathbb{Q}$, per cui

$$T(\lambda r) \neq 0 = \lambda T(r).$$

Come vedremo, questo vuol dire che T è *additiva* ma non *omogenea* su \mathbb{R} .

Esercizi

4C.1 Siano A e B due insiemi non vuoti. Dimostra che esiste sempre una funzione $f: A \rightarrow B$ iniettiva oppure una funzione $f: A \rightarrow B$ surgettiva. (Suggerimento: Considera il seguente insieme:

$$\mathcal{L} = \{(X, g) \mid X \subseteq A, g: X \rightarrow B \text{ iniettiva}\},$$

con l'ordinamento definito da $(X, g) \leq (X', g')$ se e solo se $X \subseteq X'$ e $g'|_X = g$, e applica il Lemma di Zorn.)

4C.2 Sia V uno spazio vettoriale, \mathcal{B} una base di V e \mathcal{C} un sottoinsieme libero di V . Dimostra che esiste un sottoinsieme \mathcal{B}' di \mathcal{B} tale che $\mathcal{C} \cup \mathcal{B}'$ è una base di V .

4C.3 Dimostra che ogni sottospazio U di uno spazio vettoriale V ammette un supplementare (cioè un sottospazio W tale che $V = U \oplus W$).

4C.4 Siano X e Y due insiemi. Dimostra che $|X| \leq |Y|$ se e solo se $|Y| \geq |X|$. (Suggerimento: può essere utile l'Assioma di scelta).

4C.5 Dimostra che \leq è una relazione riflessiva e transitiva sulla classe delle cardinalità.

4C.6 Dimostra che:

- (i) L'insieme dei polinomi in una variabile a coefficienti razionali è numerabile;
- (ii) L'insieme dei polinomi in n variabili a coefficienti razionali è numerabile;
- (iii) Ogni intervallo di \mathbb{R} ha la cardinalità del continuo;
- (iv) $|\mathbb{R}^n| = |\mathbb{R}|$ per ogni $n \geq 1$;
- (v) Lo spazio delle funzioni continue da \mathbb{R} in \mathbb{R} ha la cardinalità del continuo. (Suggerimento: Conviene studiare prima la cardinalità dello spazio delle funzioni continue da \mathbb{Q} in \mathbb{R} .)

4C.7 Sia A un insieme finito. Dimostra che A è equipotente a $n = \{1, \dots, n\}$ per qualche $n \in \mathbb{N}$. In tal caso diremo che A ha cardinalità n e scriveremo $|A| = n$.

4C.8 Dimostra che $|\mathcal{P}(n)| = 2^n$.

4C.9 Siano X_1, X_2, Y_1 e Y_2 quattro insiemi tali che $|X_1| = |X_2|$, $|Y_1| = |Y_2|$ e $X_1 \cap Y_1 = X_2 \cap Y_2 = \emptyset$. Dimostra che $|X_1 \times Y_1| = |X_2 \times Y_2|$ e $|X_1 \cup Y_1| = |X_2 \cup Y_2|$.

4C.10 Dimostra che se $|X_i| = \alpha$ per ogni $i \in I$ allora $\sum_{i \in I} |X_i| = |I| \times \alpha$.

4C.11 Trova tre cardinali α, β e γ tali che $\alpha + \beta = \alpha + \gamma$ ma $\beta \neq \gamma$.

4C.12 Sia V uno spazio vettoriale di dimensione finita $n \geq 0$ sul campo \mathbb{K} . Dimostra che se \mathbb{K} è infinito allora $|V| = |\mathbb{K}|$, mentre se \mathbb{K} è finito allora $|V| = |\mathbb{K}|^n$.

4C.13 Sia V uno spazio vettoriale, W un suo sottospazio e U_1, U_2 due supplementari di W in V . Dimostra che U_1 e U_2 hanno la stessa dimensione.

Applicazioni lineari

In matematica, ogni volta che si introduce una nuova struttura su un insieme si studiano immediatamente anche le funzioni che rispettano questa struttura. Nel caso degli spazi vettoriali queste funzioni sono chiamate applicazioni lineari, e sono l'argomento di questo capitolo. Ne studieremo le proprietà principali, giungendo a definire il rango di un'applicazione lineare, un numero naturale che ne riassume le principali caratteristiche; dimostreremo inoltre il fondamentale Teorema della dimensione, che lega il rango di un'applicazione lineare alla dimensione del dominio. Vedremo anche come interpretare i sistemi lineari in termini di applicazioni lineari, e questo ci darà un criterio necessario e sufficiente per la risolubilità di un sistema lineare, il Teorema di Rouché-Capelli.

5.1 Definizioni ed esempi

Nel capitolo precedente abbiamo dimostrato che le soluzioni di un sistema lineare omogeneo formano un sottospazio vettoriale. La forma dell'insieme delle soluzioni di un sistema lineare qualunque è descritta nel prossimo risultato, noto come *Teorema di struttura per i sistemi lineari*:

Proposizione 5.1 (Teorema di struttura) Sia $v^o \in \mathbb{R}^n$ una soluzione del sistema lineare $Ax = b$ di ordine n . Allora ogni altra soluzione è della forma $v = v^o + w$, dove $w \in \mathbb{R}^n$ è una soluzione del sistema omogeneo $Ax = O$. In altre parole, se $L \subseteq \mathbb{R}^n$ è l'insieme delle soluzioni del sistema $Ax = b$ e $W \subseteq \mathbb{R}^n$ è l'insieme delle soluzioni del sistema omogeneo associato $Ax = O$, si ha

$$L = v^o + W = \{v^o + w \mid w \in W\}.$$

In particolare, v^o è l'unica soluzione del sistema se e solo se le colonne di A sono linearmente indipendenti.

Dimostrazione. Sia $v^o = (v_1^o, \dots, v_n^o)$, e prendiamo una soluzione $w = (w_1, \dots, w_n)$

del sistema omogeneo. Questo vuol dire che

$$\begin{aligned} v_1^o A^1 + \cdots + v_n^o A^n &= b, \\ w_1 A^1 + \cdots + w_n A^n &= O: \end{aligned}$$

sommando otteniamo

$$(v_1^o + w_1)A^1 + \cdots + (v_n^o + w_n)A^n = b,$$

cioè $v^o + w$ è un'altra soluzione del sistema $Ax = b$.

Viceversa, sia v un'altra soluzione di $Ax = b$; dobbiamo dimostrare che $w = v - v^o$ è una soluzione del sistema $Ax = O$. Ma infatti stavolta abbiamo

$$\begin{aligned} v_1 A^1 + \cdots + v_n A^n &= b, \\ v_1^o A^1 + \cdots + v_n^o A^n &= b, \end{aligned}$$

e sottraendo otteniamo

$$(v_1 - v_1^o)A^1 + \cdots + (v_n - v_n^o)A^n = O,$$

come voluto. L'ultima affermazione segue dalla Proposizione 4.4. \square

Osservazione 5.1 Di nuovo, mentre l'esistenza di una soluzione del sistema $Ax = b$ dipende dalla specifica relazione fra A e b (vedi Proposizione 4.3), l'unicità dipende soltanto dalla matrice dei coefficienti.

Se per un attimo ci facciamo trasportare dalle notazioni, potremmo credere di avere in mano una dimostrazione molto più veloce della Proposizione 5.1. Infatti, da $Av^o = b$ e $Aw = O$ potremmo essere portati a dedurre

$$A(v^o + w) = Av^o + Aw = b + O = b,$$

e prendere questa come dimostrazione del fatto che $v^o + w$ è una soluzione di $Ax = b$. Il problema è che, fino a ora, $Ax = b$ era soltanto una notazione abbreviata per indicare l'intero sistema lineare, per cui la prima uguaglianza in questa "dimostrazione" non sembra avere molto senso. L'obiettivo di questo capitolo è introdurre quanto ci serve per renderla corretta, e poi procedere verso lidi anche più interessanti.

Definizione 5.1 Sia $A \in M_{m,n}(\mathbb{R})$ una matrice con m righe ed n colonne, cioè

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

con $A^1, \dots, A^n \in \mathbb{R}^m$. Allora definiamo un'applicazione $L_A: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ponendo

$$L_A(x) = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix} = x_1 A^1 + \cdots + x_n A^n \in \mathbb{R}^m,$$

per ogni $x \in \mathbb{R}^n$. Nota che x è soluzione del sistema $Ax = b$ sse $L_A(x) = b$. A volte, invece di scrivere $L_A(x)$ scriveremo Ax o anche $A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

Osservazione 5.2 Perché $L_A(x)$ o Ax abbiano senso, occorre che il numero di colonne di A coincida col numero di coordinate — di righe — del vettore x .

Quest'applicazione ha due proprietà fondamentali:

$$L_A(x + y) = L_A(x) + L_A(y) \quad \text{e} \quad L_A(\lambda x) = \lambda L_A(x) \quad (5.1)$$

quali che siano $x, y \in \mathbb{R}^n$ e $\lambda \in \mathbb{R}$ (esercizio). Con questa nuova notazione è possibile dare senso (vedi l'Esercizio 5.1) alla precedente "dimostrazione", utilizzando unicamente le proprietà (5.1). Questo suggerisce una definizione.

Definizione 5.2 Una applicazione (o trasformazione) lineare fra due spazi vettoriali V e W è una funzione $T: V \rightarrow W$ tale che

- (i) $T(v_1 + v_2) = T(v_1) + T(v_2)$ per tutti i $v_1, v_2 \in V$ (diremo che T è *additiva*);
- (ii) $T(\lambda v) = \lambda T(v)$ per tutti i $\lambda \in \mathbb{R}$ e $v \in V$ (diremo che T è *omogenea*).

In altre parole, un'applicazione lineare trasforma (rispetta, conserva) le operazioni dallo spazio di partenza in quello d'arrivo. Se $V = W$, si parla di *endomorfismo* (od *operatore lineare*).

Osservazione 5.3 Perché $T: V \rightarrow W$ sia lineare deve soddisfare sia (i) che (ii). Infatti esistono funzioni non lineari che soddisfano (ii) ma non (i) — vedi l'Esempio 5.5 — come pure funzioni non lineari che soddisfano (i) ma non (ii) — vedi l'Esempio 4C.2.

Osservazione 5.4 Se $T: V \rightarrow W$ è lineare, allora $T(O) = O$: infatti, preso $v \in V$ qualunque, si ha

$$T(O) = T(0v) = 0T(v) = O.$$

Per lo stesso motivo abbiamo $T(-v) = T((-1)v) = (-1)T(v) = -T(v)$.

Vediamo ora qualche esempio.

ESEMPIO 5.1 L'applicazione identica (o identità) di uno spazio vettoriale V è l'applicazione $\text{id}_V: V \rightarrow V$ data da $\text{id}_V(v) = v$ per tutti i $v \in V$. L'applicazione identica è ovviamente lineare.

ESEMPIO 5.2 Siano V e W due spazi vettoriali; l'applicazione nulla $O: V \rightarrow W$ è definita da $O(v) = O$ (dove il primo O è l'applicazione nulla, e il secondo è il vettore nullo di W) per tutti $v \in V$. Di nuovo, O è evidentemente lineare.

ESEMPIO 5.3 Per ogni matrice $A \in M_{m,n}(\mathbb{R})$ l'applicazione associata $L_A: \mathbb{R}^n \rightarrow \mathbb{R}^m$ è lineare.

ESEMPIO 5.4 Sia $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ data da

$$T \begin{vmatrix} x \\ y \\ z \end{vmatrix} = \begin{vmatrix} 3x - 2y \\ z \\ x + z \end{vmatrix}.$$

Allora abbiamo

$$\begin{aligned} T \left(\begin{vmatrix} x_1 \\ y_1 \\ z_1 \end{vmatrix} + \begin{vmatrix} x_2 \\ y_2 \\ z_2 \end{vmatrix} \right) &= T \begin{vmatrix} x_1 + x_2 \\ y_1 + y_2 \\ z_1 + z_2 \end{vmatrix} = \begin{vmatrix} 3(x_1 + x_2) - 2(y_1 + y_2) \\ z_1 + z_2 \\ (x_1 + x_2) + (z_1 + z_2) \end{vmatrix} \\ &= \begin{vmatrix} 3x_1 - 2y_1 \\ z_1 \\ x_1 + z_1 \end{vmatrix} + \begin{vmatrix} 3x_2 - 2y_2 \\ z_2 \\ x_2 + z_2 \end{vmatrix} = T \begin{vmatrix} x_1 \\ y_1 \\ z_1 \end{vmatrix} + T \begin{vmatrix} x_2 \\ y_2 \\ z_2 \end{vmatrix}, \end{aligned}$$

e

$$T \left(\lambda \begin{vmatrix} x \\ y \\ z \end{vmatrix} \right) = T \begin{vmatrix} \lambda x \\ \lambda y \\ \lambda z \end{vmatrix} = \begin{vmatrix} 3\lambda x - 2\lambda y \\ \lambda z \\ \lambda x + \lambda z \end{vmatrix} = \lambda \begin{vmatrix} 3x - 2y \\ z \\ x + z \end{vmatrix} = \lambda T \begin{vmatrix} x \\ y \\ z \end{vmatrix},$$

per cui T è lineare. Nota che $T = L_A$, dove

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

ESEMPIO 5.5 Sia $T: \mathbb{R}^2 \rightarrow \mathbb{R}$ data da

$$T \begin{vmatrix} x \\ y \end{vmatrix} = (x^3 + y^3)^{1/3}.$$

Allora $T(\lambda v) = \lambda T(v)$ per ogni $\lambda \in \mathbb{R}$ e $v \in \mathbb{R}^2$, ma T non è additiva: infatti

$$T \left(\begin{vmatrix} 1 \\ 0 \end{vmatrix} + \begin{vmatrix} 0 \\ 1 \end{vmatrix} \right) = T \begin{vmatrix} 1 \\ 1 \end{vmatrix} = 2^{1/3} \neq 2 = 1 + 1 = T \begin{vmatrix} 1 \\ 0 \end{vmatrix} + T \begin{vmatrix} 0 \\ 1 \end{vmatrix}.$$

Osservazione 5.5 Il punto è che le coordinate di un'applicazione lineare devono essere polinomi di primo grado senza termine noto delle coordinate del vettore di partenza (vedi l'Esercizio 5.4 e l'Esercizio 7.11).

ESEMPIO 5.6 Sia $V = \mathbb{R}_2[t]$ lo spazio vettoriale dei polinomi di grado minore o uguale a 2. Allora definiamo $T: \mathbb{R}_2[t] \rightarrow \mathbb{R}_2[t]$ ponendo

$$\forall p \in \mathbb{R}_2[t] \quad [T(p)](t) = p(t+1).$$

Per esempio, se $p(t) = t^2$ allora $T(p)$ è il polinomio $(t+1)^2 = t^2 + 2t + 1$. L'applicazione T è lineare: infatti

$$\begin{aligned} [T(p_1 + p_2)](t) &= (p_1 + p_2)(t+1) = p_1(t+1) + p_2(t+1) = [T(p_1)](t) + [T(p_2)](t), \\ [T(\lambda p)](t) &= (\lambda p)(t+1) = \lambda(p(t+1)) = [(\lambda T)(p)](t) \end{aligned}$$

per ogni $t \in \mathbb{R}$, per cui $T(p_1 + p_2) \equiv T(p_1) + T(p_2)$ e $T(\lambda p) \equiv \lambda T(p)$.

ESEMPIO 5.7 Sia $\mathcal{B} = \{v_1, \dots, v_n\}$ una base di uno spazio vettoriale V . Allora ogni $v \in V$ si scrive come $v = \alpha_1 v_1 + \dots + \alpha_n v_n$, dove le coordinate $\alpha_1, \dots, \alpha_n$ sono univocamente determinate. Definiamo $F_{\mathcal{B}}: V \rightarrow \mathbb{R}^n$ ponendo

$$F_{\mathcal{B}}(v) = \begin{vmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{vmatrix};$$

in altri termini, $F_{\mathcal{B}}$ associa a ogni $v \in V$ le sue coordinate rispetto alla base \mathcal{B} (ed è la generalizzazione della $F_{\mathcal{B}}$ vista nel Paragrafo 2.2). Anche $F_{\mathcal{B}}$ è lineare: se $v = \alpha_1 v_1 + \dots + \alpha_n v_n$ e $w = \beta_1 v_1 + \dots + \beta_n v_n$, allora

$$v + w = (\alpha_1 + \beta_1)v_1 + \dots + (\alpha_n + \beta_n)v_n$$

e quindi

$$F_{\mathcal{B}}(v + w) = \begin{vmatrix} \alpha_1 + \beta_1 \\ \vdots \\ \alpha_n + \beta_n \end{vmatrix} = \begin{vmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{vmatrix} + \begin{vmatrix} \beta_1 \\ \vdots \\ \beta_n \end{vmatrix} = F_{\mathcal{B}}(v) + F_{\mathcal{B}}(w).$$

Un argomento analogo mostra che $F_{\mathcal{B}}(\lambda v) = \lambda F_{\mathcal{B}}$. Nota che $F_{\mathcal{B}}$ è bigettiva (perché?).

ESEMPIO 5.8 Siano U, W sottospazi di uno spazio vettoriale V tali che $V = U \oplus W$. Questo vuol dire che ogni elemento $v \in V$ si può scrivere in modo unico come $v = u + w$, con $u \in U$ e $w \in W$. Ponendo $P(v) = u$ definiamo un'applicazione $P: V \rightarrow U \subset V$, detta *proiezione su U lungo W* . È facile dimostrare che P è lineare; per esempio, se $\lambda \in \mathbb{R}$ e $v = u + w \in V$, allora $\lambda v = \lambda u + \lambda w$ e quindi $P(\lambda v) = \lambda P(v)$. In modo analogo si dimostra che P è additiva.

ESEMPIO 5.9 Fissiamo $a \in \mathbb{R}^n$, e definiamo $\varphi_a: \mathbb{R}^n \rightarrow \mathbb{R}$ ponendo

$$\forall x \in \mathbb{R}^n \quad \varphi_a(x) = a_1 x_1 + \dots + a_n x_n \in \mathbb{R}.$$

Per esempio, se $n = 2$ e $a = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, allora

$$\varphi_a(x) = x_1 + 2x_2.$$

Si verifica facilmente (esercizio) che φ_a è un'applicazione lineare da \mathbb{R}^n in \mathbb{R} .

ESEMPIO 5.10 La trasposizione $T: M_{m,n}(\mathbb{R}) \rightarrow M_{n,m}(\mathbb{R})$, che associa a una matrice A la trasposta A^T , è l'applicazione data da

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

Per esempio

$$A = \begin{pmatrix} 1 & -1 \\ 0 & 27 \\ -\sqrt{33} & \pi^2 \end{pmatrix} \mapsto A^T = \begin{pmatrix} 1 & 0 & -\sqrt{33} \\ -1 & 27 & \pi^2 \end{pmatrix}.$$

La trasposizione effettua una simmetria rispetto alla diagonale principale della matrice. In termini di righe, colonne ed elementi abbiamo

$$(A^T)_{ij} = a_{ji}, \quad (A^T)_i = (A^i)^T \quad \text{e} \quad (A^T)^j = (A_j)^T, \quad (5.2)$$

dove $A = (a_{ij})$ e $(A^T)_{ij}$ è l'elemento di posto (i, j) della matrice A^T . Capito questo, è evidente che la trasposizione è un'applicazione lineare, e che

$$(A^T)^T = A.$$

Cominciamo ora a studiare le applicazioni lineari in generale. La prima osservazione importante è che un'applicazione lineare è completamente determinata dai valori che assume su una base:

Proposizione 5.2 Siano V e W due spazi vettoriali, $\{v_1, \dots, v_n\}$ una base di V , e w_1, \dots, w_n vettori qualunque di W . Allora esiste un'unica applicazione lineare $T: V \rightarrow W$ tale che $T(v_j) = w_j$ per $j = 1, \dots, n$. L'applicazione T è definita da

$$T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 w_1 + \cdots + \alpha_n w_n \quad (5.3)$$

per tutti gli $\alpha_1, \dots, \alpha_n \in \mathbb{R}$.

Dimostrazione. Prima di tutto, verifichiamo che la T definita in (5.3) è lineare. Siano $v, v' \in V$, e scriviamoli come combinazione lineare della base di V :

$$v = \alpha_1 v_1 + \cdots + \alpha_n v_n \quad \text{e} \quad v' = \alpha'_1 v_1 + \cdots + \alpha'_n v_n.$$

Allora $v + v' = (\alpha_1 + \alpha'_1)v_1 + \cdots + (\alpha_n + \alpha'_n)v_n$ e

$$\begin{aligned} T(v + v') &= (\alpha_1 + \alpha'_1)w_1 + \cdots + (\alpha_n + \alpha'_n)w_n \\ &= (\alpha_1 w_1 + \cdots + \alpha_n w_n) + (\alpha'_1 w_1 + \cdots + \alpha'_n w_n) = T(v) + T(v'). \end{aligned}$$

Nello stesso modo (esercizio) si dimostra che $T(\lambda v) = \lambda T(v)$.

Quindi l'esistenza è sistemata. Per l'unicità, sia $S: V \rightarrow W$ un'altra applicazione lineare tale che $S(v_j) = w_j$ per $j = 1, \dots, n$; dobbiamo dimostrare che $S(v) = T(v)$ qualunque sia $v \in V$. Prendiamo allora $v \in V$, e scriviamolo di nuovo come combinazione lineare degli elementi della base, $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$. Allora

$$\begin{aligned} S(v) &= S(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 S(v_1) + \cdots + \alpha_n S(v_n) \\ &= \alpha_1 w_1 + \cdots + \alpha_n w_n = T(v). \end{aligned}$$

□

Corollario 5.3 Siano V e W due spazi vettoriali, $\{v_1, \dots, v_n\}$ una base di V , e $S, T: V \rightarrow W$ due applicazioni lineari. Supponiamo che si abbia $S(v_j) = T(v_j)$ per $j = 1, \dots, n$. Allora $S \equiv T$, cioè $S(v) = T(v)$ per tutti i vettori $v \in V$.

Dimostrazione. Basta usare la Proposizione 5.2 con $w_j = T(v_j)$, per $j = 1, \dots, n$. □

Due matrici diverse danno sempre origine ad applicazioni lineari diverse:

Proposizione 5.4 Siano $A, B \in M_{m,n}(\mathbb{R})$ due matrici. Allora $L_A \equiv L_B$ se e solo se $A = B$.

Dimostrazione. Chiaramente se $A = B$ allora $L_A \equiv L_B$. Viceversa, supponiamo che $L_A \equiv L_B$; siccome si ha

$$\forall j = 1, \dots, n \quad L_A(e_j) = A^j, \quad (5.4)$$

(dove e_1, \dots, e_n è la base canonica di \mathbb{R}^n) le due matrici A e B devono avere le stesse colonne, per cui sono uguali. □

5.2 Nucleo e immagine

A ogni applicazione lineare $T: V \rightarrow W$ possiamo associare due sottoinsiemi:

- il nucleo¹ $\text{Ker } T = \{v \in V \mid T(v) = O\} \subseteq V$;
- l'immagine $\text{Im } T = T(V) = \{T(v) \mid v \in V\} \subseteq W$.

Il fatto interessante è che nucleo e immagine, oltre a essere sottospazi vettoriali, caratterizzano l'iniettività e la surgettività delle applicazioni lineari:

¹ Ker è l'abbreviazione della parola inglese *kernel*, che vuol dire appunto nucleo.

Proposizione 5.5 Sia $T: V \rightarrow W$ un'applicazione lineare. Allora:

- (i) $\text{Ker } T$ è un sottospazio di V ;
- (ii) $\text{Im } T$ è un sottospazio di W ;
- (iii) T è surgettiva se e solo se $\text{Im } T = W$;
- (iv) T è iniettiva se e solo se $\text{Ker } T = \{O\}$.

Dimostrazione. (i) Facciamo vedere che il nucleo è chiuso rispetto alla somma; la chiusura rispetto al prodotto per scalari è analoga, e lasciata per esercizio. Prendiamo $v_1, v_2 \in \text{Ker } T$, cioè tali che $T(v_1) = T(v_2) = O$; dobbiamo dimostrare che $v_1 + v_2 \in \text{Ker } T$. Ma infatti

$$T(v_1 + v_2) = T(v_1) + T(v_2) = O + O = O.$$

(ii) Stavolta dimostriamo che l'immagine è chiusa rispetto al prodotto per scalari, lasciando la chiusura rispetto alla somma per esercizio. Siano $\lambda \in \mathbb{R}$ e $T(v) \in \text{Im } T$; dobbiamo far vedere che $\lambda T(v) \in \text{Im } T$. Ma infatti $\lambda T(v) = T(\lambda v) \in \text{Im } T$.

(iii) È la definizione di surgettività.

(iv) Questo è l'unico punto interessante; che distingue nettamente le applicazioni lineari dalle funzioni qualsiasi. Se T è iniettiva, $T(v) = O = T(O)$ implica $v = O$ e quindi $\text{Ker } T = \{O\}$. Viceversa, supponiamo $\text{Ker } T = \{O\}$. Prendiamo $v_1, v_2 \in V$ tali che $T(v_1) = T(v_2)$; dobbiamo dimostrare che $v_1 = v_2$. Ma infatti

$$T(v_1) = T(v_2) \implies T(v_1) - T(v_2) = O \implies T(v_1 - v_2) = O \implies v_1 - v_2 = O,$$

dove l'ultima implicazione vale perché $\text{Ker } T = \{O\}$, e ci siamo. \square

Come vedremo, calcolare il nucleo di un'applicazione lineare è più o meno equivalente a risolvere un sistema lineare omogeneo. Per il calcolo dell'immagine di un'applicazione lineare è invece utile il Lemma seguente:

Lemma 5.6 Sia $T: V \rightarrow W$ un'applicazione lineare, e $\{v_1, \dots, v_n\}$ una base di V . Allora

$$\text{Im } T = \text{Span}(T(v_1), \dots, T(v_n)).$$

Dimostrazione. Infatti $\text{Im } T$ è l'insieme degli elementi della forma $T(v)$ al variare di $v \in V$. Ma ogni elemento di V si scrive come combinazione lineare di v_1, \dots, v_n ; quindi

$$\begin{aligned} \text{Im } T &= \{T(\alpha_1 v_1 + \dots + \alpha_n v_n) \mid \alpha_1, \dots, \alpha_n \in \mathbb{R}\} \\ &= \{\alpha_1 T(v_1) + \dots + \alpha_n T(v_n) \mid \alpha_1, \dots, \alpha_n \in \mathbb{R}\} = \text{Span}(T(v_1), \dots, T(v_n)). \end{aligned}$$

\square

Osservazione 5.6 Non è detto che $\{T(v_1), \dots, T(v_n)\}$ sia una base di $\text{Im } T$; è solo un sistema di generatori (vedi l'Esercizio 5.17).

ESEMPIO 5.11 Sia $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ data da

$$T \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x - y \\ 2y - 2x \\ z \end{pmatrix}.$$

È chiaramente lineare; per essere precisi, $T = L_A$ con

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

Il nucleo è costituito dai vettori mandati in O , cioè dalle soluzioni del sistema lineare omogeneo

$$\begin{cases} x - y = 0, \\ 2y - 2x = 0, \\ z = 0; \end{cases}$$

un breve conto con le tecniche del Capitolo 3 mostra che

$$\text{Ker } T = \mathbb{R} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

D'altra parte,

$$\text{Im } T = \text{Span} \left(\begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right) = \text{Span} \left(\begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right),$$

come segue immediatamente dal Lemma 5.6 applicato usando la base canonica di \mathbb{R}^3 .

Osservazione 5.7 Se A è una matrice, spesso e volentieri scriveremo $\text{Ker } A$ invece di $\text{Ker } L_A$, e $\text{Im } A$ invece di $\text{Im } L_A$.

Una delle caratteristiche più importanti di un'applicazione lineare (che, come vedremo, è in grado da sola di dire se un sistema lineare ammette o meno soluzione) è il rango.

Definizione 5.3 Il rango $\text{rg } T$ di un'applicazione lineare $T: V \rightarrow W$ è la dimensione dell'immagine:

$$\text{rg } T = \dim \text{Im } T.$$

ESEMPIO 5.12 Nel caso delle matrici, (5.4) e il Lemma 5.6 ci dicono che

$$\forall A \in M_{m,n}(\mathbb{R}) \quad \text{Im } L_A = \text{Span}(A^1, \dots, A^n) \subseteq \mathbb{R}^m.$$

Quindi il rango di L_A (o, come spesso diremo, il rango di A) è la dimensione dello spazio generato dalle colonne. In particolare, $\text{rg } A \leq \min\{m, n\}$.

Siccome l'immagine di T è un sottospazio di W , chiaramente $\text{rg } T \leq \dim W$. D'altra parte, il Lemma 5.6 ci dice che $\text{rg } T \leq \dim V$. È possibile fare di meglio? La risposta è positiva, ed è contenuta nel fondamentale *Teorema della dimensione*:

Teorema 5.7 (della dimensione) *Sia $T: V \rightarrow W$ un'applicazione lineare. Allora*

$$\dim V = \dim \text{Ker } T + \text{rg } T.$$

Dimostrazione. Sia $\{u_1, \dots, u_r\}$ una base di $\text{Ker } T$; col Teorema 4.10 la possiamo completare a una base $\{u_1, \dots, u_r, v_{r+1}, \dots, v_n\}$ di V (se $\text{Ker } T = \{O\}$, prendiamo direttamente una base $\{v_1, \dots, v_n\}$ di V , e consideriamo $r = 0$ ed $s = n$ nel seguito). Poniamo $w_j = T(v_{r+j}) \in W$ per $j = 1, \dots, s = n - r$; se dimostriamo che $\mathcal{B} = \{w_1, \dots, w_s\}$ è una base di $\text{Im } T$ abbiamo finito. Grazie al Lemma 5.6 (e al fatto che $T(u_i) = O$ per $i = 1, \dots, r$) sappiamo già che \mathcal{B} è un sistema di generatori di $\text{Im } T$; dobbiamo solo far vedere che w_1, \dots, w_s sono linearmente indipendenti.

Supponiamo che $\alpha_1, \dots, \alpha_s \in \mathbb{R}$ siano tali che

$$\alpha_1 w_1 + \dots + \alpha_s w_s = O.$$

Allora

$$O = \alpha_1 T(v_{r+1}) + \dots + \alpha_s T(v_{r+s}) = T(\alpha_1 v_{r+1} + \dots + \alpha_s v_{r+s}),$$

per cui $\alpha_1 v_{r+1} + \dots + \alpha_s v_{r+s} \in \text{Ker } T$. Questo vuol dire che esistono $\beta_1, \dots, \beta_r \in \mathbb{R}$ tali che $\alpha_1 v_{r+1} + \dots + \alpha_s v_{r+s} = \beta_1 u_1 + \dots + \beta_r u_r$; quindi

$$\beta_1 u_1 + \dots + \beta_r u_r - \alpha_1 v_{r+1} - \dots - \alpha_s v_{r+s} = O,$$

e l'indipendenza lineare di $\{u_1, \dots, u_r, v_{r+1}, \dots, v_{r+s}\}$ implica $\alpha_1 = \dots = \alpha_s = 0$, come desiderato. \square

Osservazione 5.8 Nel caso di una matrice $A \in M_{m,n}(\mathbb{R})$ il Teorema 5.7 diventa

$$n = \dim \text{Ker } A + \text{rg } A.$$

Una prima conseguenza del Teorema della dimensione è che per vedere se un'applicazione lineare è iniettiva o surgettiva basta controllare il rango:

Corollario 5.8 *Sia $T: V \rightarrow W$ un'applicazione lineare. Allora:*

- (i) T è iniettiva se e solo se $\text{rg } T = \dim V$;
- (ii) T è surgettiva se e solo se $\text{rg } T = \dim W$;
- (iii) se $\dim V = \dim W$ (in particolare, se $V = W$), T è iniettiva se e solo se è surgettiva.

Dimostrazione. (i) L'applicazione T è iniettiva se e solo se $\text{Ker } T = \{O\}$, ovvero se e solo se $\dim \text{Ker } T = 0$; basta allora applicare il Teorema 5.7.

(ii) T è surgettiva se e solo se $\text{Im } T = W$, che succede se e solo se $\text{rg } T = \dim W$ (Proposizione 4.14). \square

(iii) Basta confrontare (i) e (ii).

Un'altra conseguenza importante è il *Teorema di Rouché-Capelli*, un criterio per decidere quando un sistema lineare ammette soluzioni:

Corollario 5.9 (Rouché-Capelli) *Sia $Ax = b$ un sistema lineare di m equazioni in n incognite, e sia A' la matrice completa del sistema. Allora il sistema ammette soluzioni se e solo se $\text{rg } A' = \text{rg } A$. Inoltre la soluzione, se esiste, è unica se e solo se $\text{rg } A = n$.*

Dimostrazione. Grazie alla Proposizione 4.3, sappiamo che la soluzione esiste se e solo se $b \in \text{Span}(A^1, \dots, A^n)$, che è equivalente a richiedere

$$\text{Span}(A^1, \dots, A^n) = \text{Span}(A^1, \dots, A^n, b). \quad (5.5)$$

Ma il primo span è l'immagine di L_A , mentre il secondo è l'immagine di $L_{A'}$; siccome si ha sempre

$$\text{Span}(A^1, \dots, A^n) \subseteq \text{Span}(A^1, \dots, A^n, b),$$

l'uguaglianza (5.5) vale se e solo se $\text{rg } A = \text{rg } A'$.

L'unicità vale se e solo se $\text{Ker } L_A = \{O\}$ (Proposizione 5.1), il che accade se e solo se $\text{rg } A = \dim \mathbb{R}^n = n$ (Teorema 5.7). \square

È quindi evidente la necessità di trovare un metodo operativo efficace per il calcolo del rango di una matrice (che non si riduca semplicemente al recuperare un insieme massimale di colonne linearmente indipendenti). Questo metodo ci verrà fornito, nel prossimo capitolo, dalla riduzione a scala di matrici qualunque; per il momento vediamo come il rango si possa calcolare anche utilizzando le righe della matrice al posto delle colonne.

Ci serve una relazione molto interessante fra la trasposizione di matrici e le applicazioni φ_a dell'Esempio 5.9 (relazione che riprenderemo nei Complementi al Capitolo 8):

Lemma 5.10 *Per ogni matrice $A \in M_{m,n}(\mathbb{R})$ e vettori $x \in \mathbb{R}^n$ e $a \in \mathbb{R}^m$ si ha*

$$\varphi_{A^T a}(x) = \varphi_a(Ax).$$

Dimostrazione. Notiamo prima di tutto che quanto abbiamo scritto ha senso. L_A manda vettori di \mathbb{R}^n in vettori di \mathbb{R}^m ; quindi $Ax \in \mathbb{R}^m$ e vi possiamo applicare φ_a , che manda vettori di \mathbb{R}^m in \mathbb{R} ; quindi $\varphi_a(Ax)$ è un ben definito numero reale. D'altra parte, L_{A^T} manda vettori di \mathbb{R}^m in vettori di \mathbb{R}^n ; quindi $\varphi_{A^T a}$ manda vettori di \mathbb{R}^n in \mathbb{R} , e di nuovo $\varphi_{A^T a}(x)$ è un ben definito numero reale. Per dimostrare che i due sono uguali, basta calcolarli. Scriviamo

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad a = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix}.$$

Quindi

$$A^T a = \begin{pmatrix} a_{11}\alpha_1 + \cdots + a_{m1}\alpha_m \\ \vdots \\ a_{1n}\alpha_1 + \cdots + a_{mn}\alpha_m \end{pmatrix}, \quad Ax = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix},$$

e

$$\begin{aligned} \varphi_{A^T a}(x) &= (a_{11}\alpha_1 + \cdots + a_{m1}\alpha_m)x_1 + \cdots + (a_{1n}\alpha_1 + \cdots + a_{mn}\alpha_m)x_n \\ &= \alpha_1(a_{11}x_1 + \cdots + a_{1n}x_n) + \cdots + \alpha_m(a_{m1}x_1 + \cdots + a_{mn}x_n) = \varphi_a(Ax). \end{aligned}$$

□

E ora possiamo dimostrare che il rango calcolato usando le righe è uguale al rango calcolato usando le colonne:

Proposizione 5.11 Sia $A \in M_{m,n}(\mathbb{R})$. Allora:

- (i) $\mathbb{R}^m = \text{Im } L_A \oplus \text{Ker } L_{A^T}$ e $\mathbb{R}^n = \text{Im } L_{A^T} \oplus \text{Ker } L_A$;
- (ii) $\text{rg } A = \text{rg } A^T$, ovvero il massimo numero di colonne linearmente indipendenti coincide col massimo numero di righe linearmente indipendenti.

Osservazione 5.9 Questo è l'unico risultato in tutta questa parte del testo che dipende da proprietà specifiche di \mathbb{R} , e non vale per un campo qualunque. La dimostrazione che faremo funziona in \mathbb{R} , \mathbb{Q} ed eventuali altri campi contenuti in \mathbb{R} ; vedremo più oltre (Proposizione 11.10) l'aspetto di questo risultato su \mathbb{C} .

Dimostrazione. Cominciamo col dimostrare che $\text{Im } L_A \cap \text{Ker } L_{A^T} = \{O\}$. Prendiamo $a \in \text{Im } L_A \cap \text{Ker } L_{A^T} \subseteq \mathbb{R}^m$; in particolare, $A^T a = O$, ed esiste $b \in \mathbb{R}^n$ tale che $Ab = a$. Allora

$$0 = \varphi_O(b) = \varphi_{A^T a}(b) = \varphi_a(Ab) = \varphi_a(a) = a_1^2 + \cdots + a_m^2,$$

dove abbiamo usato il Lemma 5.10; questo può accadere se e solo se $a_1 = \cdots = a_m = 0$ (in quanto $a_1, \dots, a_m \in \mathbb{R}$), e quindi $a = O$.

Dunque $\text{Im } L_A \cap \text{Ker } L_{A^T} = \{O\}$. Essendo $\text{Im } L_A + \text{Ker } L_{A^T} \subseteq \mathbb{R}^m$, la Proposizione 4.14 e i Teoremi 4.16 e 5.7 ci dicono che

$$m \geq \dim(\text{Im } L_A + \text{Ker } L_{A^T}) = \dim \text{Im } L_A + \dim \text{Ker } L_{A^T} = \text{rg } A + m - \text{rg } A^T, \quad (5.6)$$

per cui $\text{rg } A^T \geq \text{rg } A$. Rifacendo lo stesso ragionamento a partire da A^T , e ricordandosi che $(A^T)^T = A$, otteniamo $\text{rg } A^T \leq \text{rg } A$; mettendo tutto insieme otteniamo (ii). In particolare, quindi, (5.6) ci dice che

$$\dim(\text{Im } L_A + \text{Ker } L_{A^T}) = m,$$

per cui $\mathbb{R}^m = \text{Im } L_A \oplus \text{Ker } L_{A^T}$. L'ultima affermazione di (i) si ottiene scambiando un'altra volta A e A^T . □

Esercizi

5.1 Dimostra di nuovo la Proposizione 5.1 utilizzando L_A . (*Suggerimento:* dimostra prima di tutto che $x \in \mathbb{R}^n$ è una soluzione di $Ax = b$ se e solo se $L_A(x) = b$.)

5.2 Sia V il solito spazio vettoriale dell'Esempio 4.12. Definiamo $T, S_1, S_2: V \rightarrow V$ con

$$T \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = \begin{pmatrix} 2x - y \\ x \\ 0 \end{pmatrix}, \quad S_1 \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = \begin{pmatrix} x^2 \\ 2y \\ 0 \end{pmatrix}, \quad S_2 \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = \begin{pmatrix} x + 1 \\ y \\ 0 \end{pmatrix}.$$

Dimostra che T è lineare mentre S_1 ed S_2 non lo sono.

5.3 Sia $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ (con $n > m$) data da

$$T \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}.$$

Dimostra che T è lineare e trova, se esiste, una matrice $A \in M_{m,n}(\mathbb{R})$ tale che $T = L_A$.

5.4 Sia $T: V \rightarrow W$ una funzione fra due spazi vettoriali, e fissiamo una base \mathcal{B} di V e una base \mathcal{C} di W . Dimostra che se le coordinate (rispetto a \mathcal{C}) di $T(v)$ si ottengono dalle coordinate di v (rispetto a \mathcal{B}) tramite polinomi di primo grado senza termini noti allora la funzione T è lineare.

5.5 Siano V_1, \dots, V_k sottospazi di V tali che $V = V_1 \oplus \cdots \oplus V_k$. Per ogni $i = 1, \dots, k$ definisci una proiezione $P_i: V \rightarrow V_i \subset V$, e dimostra che è lineare.

5.6 La *traccia* $\text{tr } A$ di una matrice quadrata $A \in M_{n,n}(\mathbb{R})$ è per definizione la somma degli elementi sulla diagonale principale:

$$\text{tr } A = a_{11} + \cdots + a_{nn}.$$

Dimostra che $\text{tr}: M_{n,n}(\mathbb{R}) \rightarrow \mathbb{R}$ è lineare.

5.7 Sia $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ l'unica applicazione lineare tale che

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

e per ogni $a \in \mathbb{R}$ sia $S_a: \mathbb{R}^2 \rightarrow \mathbb{R}^3$ l'unica applicazione lineare tale che

$$S_a \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 6 \\ 4 \\ 2 \end{pmatrix} \quad \text{e} \quad S_a \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} a \\ 0 \\ 4 \end{pmatrix}.$$

Trova per quali $a \in \mathbb{R}$ si ha $\text{Im } T = \text{Im } S_a$, e calcola la dimensione di $\text{Im } T \cap \text{Im } S_a$ al variare di $a \in \mathbb{R}$.

5.8 Sia $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ un'applicazione lineare. Dimostra che esiste $a \in \mathbb{R}^n$ tale che $T \equiv \varphi_a$, dove $\varphi_a: \mathbb{R}^n \rightarrow \mathbb{R}^n$ è l'applicazione definita nell'Esempio 5.9. (Suggerimento: prendi $a_j = T(e_j)$ per $j = 1, \dots, n$, dove $\{e_1, \dots, e_n\}$ è la base canonica di \mathbb{R}^n .)

5.9 Una matrice quadrata $A \in M_{n,n}(\mathbb{R})$ è detta *simmetrica* (rispettivamente, *antisimmetrica*) se $A^T = A$ (rispettivamente, $A^T = -A$). Indichiamo con $\mathcal{S}_n(\mathbb{R})$ (rispettivamente, $\mathcal{A}_n(\mathbb{R})$) l'insieme delle matrici simmetriche (rispettivamente, antisimmetriche) di ordine n . Dimostra che $\mathcal{S}_n(\mathbb{R})$ e $\mathcal{A}_n(\mathbb{R})$ sono sottospazi di $M_{n,n}(\mathbb{R})$ e calcolane la dimensione. (Suggerimento: può essere utile l'Esempio 3C.2.)

5.10 Dimostra che $M_{n,n}(\mathbb{R}) = \mathcal{S}_n(\mathbb{R}) \oplus \mathcal{A}_n(\mathbb{R})$. (Suggerimento: usa $(A + A^T)/2$ e $(A - A^T)/2$.)

5.11 Sia $T: V \rightarrow W$ un'applicazione lineare, $V_1 \subseteq V$ un sottospazio di V e $W_1 \subseteq W$ un sottospazio di W . Dimostra che:

- (i) $T^{-1}(W_1) = \{v \in V \mid T(v) \in W_1\} \subseteq V$ è un sottospazio di V ;
- (ii) $T(V_1) = \{T(v) \mid v \in V_1\} \subseteq W$ è un sottospazio di W .

5.12 Dimostra che il sistema lineare $Ax = b$ è compatibile se e solo se $b \in \text{Im } L_A$; la soluzione, se esiste, è unica se e solo se $\text{Ker } L_A = \{O\}$.

5.13 In ciascuno dei tre casi seguenti, scopri se è possibile costruire applicazioni lineari che soddisfano le condizioni indicate, e in caso ne esistano più di una trovane almeno due distinte:

(i) $T: \mathbb{R}^4 \rightarrow \mathbb{R}^3$ surgettiva e tale che $\text{Ker } T = \mathbb{R} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$;

(ii) $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ tale che $\text{Im } T = \mathbb{R} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$;

(iii) $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ iniettiva e tale che $\text{Im } T = \text{Span} \left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix} \right)$.

5.14 Sia $T: V \rightarrow W$ un'applicazione lineare. Dimostra che:

- (i) se $\dim V > \dim W$ l'applicazione T non può essere iniettiva;
- (ii) se $\dim V < \dim W$ l'applicazione T non può essere surgettiva.

5.15 Sia $T: V \rightarrow W$ un'applicazione lineare. Dimostra che se $v_1, \dots, v_k \in V$ sono linearmente dipendenti, allora anche $T(v_1), \dots, T(v_k) \in W$ lo sono. Dimostra anche che se $T(v_1), \dots, T(v_k) \in W$ sono linearmente indipendenti, allora anche $v_1, \dots, v_k \in V$ lo sono.

5.16 Sia $T: V \rightarrow W$ un'applicazione lineare *iniettiva*. Dimostra che se i vettori $v_1, \dots, v_k \in V$ sono linearmente indipendenti allora anche $T(v_1), \dots, T(v_k) \in W$ lo sono.

5.17 Trova un'applicazione lineare $T: V \rightarrow W$ e dei vettori $v_1, v_2 \in V$ tali che v_1 e v_2 siano linearmente indipendenti e $T(v_1)$ e $T(v_2)$ non lo siano.

Sistemi lineari

È giunto il momento di affrontare un certo numero di problemi riguardanti tecniche di calcolo: come si trova il rango di una matrice, o di un'applicazione lineare? Come si risolve un sistema lineare non quadrato? Come si trovano dimensione e base di un sottospazio generato da certi vettori? Come si trova dimensione e base del nucleo di un'applicazione lineare? Come si estrae una base da un sistema di generatori? Come si trovano dimensione e base dell'unione e dell'intersezione di due sottospazi? Si tratta di problemi tutti collegati fra loro, e che possono essere risolti con l'ausilio di un unico strumento: la riduzione a scala, una versione dell'eliminazione di Gauss per matrici non quadrate. Parleremo inoltre di equazioni parametriche e cartesiane, e definiremo ufficialmente i sottospazi affini (già comparsi in realtà nel Teorema di struttura).

6.1 Sistemi a scala

L'eliminazione di Gauss trasforma un sistema lineare quadrato in un sistema equivalente triangolare superiore; cominciamo allora col vedere cosa prende il posto delle matrici triangolari superiori nel caso non quadrato.

Definizione 6.1 Una *matrice a scala* è una matrice $m \times n$ siffatta:

$$\left| \begin{array}{cccccccccccccccc} 0 & \dots & 0 & p_1 & * & \dots & * \\ \vdots & & & \vdots & 0 & 0 & \dots & 0 & p_2 & * & \dots & \dots & \dots & \dots & \dots & * \\ \vdots & & & \vdots & \vdots & \vdots & & \vdots & 0 & 0 & \dots & 0 & p_3 & * & \dots & * \\ \vdots & & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & 0 & p_r & * & \dots & * \\ \vdots & & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & 0 & 0 & \dots & 0 \\ \vdots & & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & \dots & 0 & 0 & 0 & \dots & 0 \end{array} \right|$$

dove una * indica che può esserci qualunque cosa. I numeri $p_1, \dots, p_r \in \mathbb{R}^*$, tutti non nulli, sono i pivot della matrice a scala. Un sistema a scala è un sistema lineare la cui matrice dei coefficienti è una matrice a scala.

Osservazione 6.1 I pivot di una matrice quadrata triangolare superiore descritti nel Paragrafo 3.3 non sono necessariamente gli stessi definiti qui; coincidono solo nel caso in cui la matrice sia non singolare. In questo capitolo gli unici pivot che ci riguardano sono quelli che abbiamo appena definito, e nel seguito sarà sempre chiaro di quali pivot si sta parlando.

Le proprietà principali di una matrice a scala sono contenute nel seguente

Lemma 6.1 Sia $S \in M_{m,n}(\mathbb{R})$ una matrice a scala con r pivot, e poniamo

$$V_r = \left\{ \begin{pmatrix} b_1 \\ \vdots \\ b_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^m \mid b_1, \dots, b_r \in \mathbb{R} \right\} = \text{Span}(e_1, \dots, e_r) \subset \mathbb{R}^m.$$

Indichiamo con S^{j_k} la colonna di S in cui compare il k -esimo pivot p_k , per $k = 1, \dots, r$. Allora $\text{Im } S = V_r$, $\text{rg } S = r$ e $\{S^{j_1}, \dots, S^{j_r}\}$ è una base di $\text{Im } S$.

Dimostrazione. Siccome tutte le colonne di S appartengono a V_r e generano $\text{Im } S$, è chiaro che $\text{Im } S \subseteq V_r$. È pure evidente che $\dim V_r = r$; quindi basta far vedere (perché?) che S^{j_1}, \dots, S^{j_r} sono linearmente indipendenti.

Consideriamo il sistema omogeneo $\alpha_1 S^{j_1} + \dots + \alpha_r S^{j_r} = 0$, di incognite $\alpha_1, \dots, \alpha_r$. Questo sistema ha come matrice dei coefficienti una matrice della forma

$$\begin{pmatrix} p_1 & * & \dots & * \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \dots & 0 & p_r \\ 0 & \dots & \dots & 0 \\ \vdots & & & \vdots \\ 0 & \dots & \dots & 0 \end{pmatrix}.$$

Quindi siamo ricondotti a un sistema omogeneo quadrato triangolare superiore con elementi non nulli sulla diagonale principale; per la Proposizione 3.1, l'unica soluzione possibile è $\alpha_1 = \dots = \alpha_r = 0$. \square

Corollario 6.2 Sia $S \in M_{m,n}(\mathbb{R})$ una matrice a scala di rango r . Allora il sistema $Sx = c$ ha soluzione se e solo se le ultime $m - r$ coordinate di c sono zero, e lo spazio delle soluzioni del sistema omogeneo $Sx = 0$ ha dimensione $n - r$.

Dimostrazione. Il sistema $Sx = c$ ha soluzione se e solo se $c \in \text{Im } S = V_r$. Lo spazio delle soluzioni di $Sx = 0$ è $\text{Ker } S$ che ha dimensione $n - \text{rg } S$ (Osservazione 5.8). \square

Proviamo ora a risolvere un sistema a scala.

ESEMPIO 6.1 Consideriamo il sistema

$$\begin{cases} x_2 - x_3 + 3x_4 - x_5 + \frac{1}{2}x_6 = 1, \\ 2x_4 - x_6 = 0, \\ x_5 + 4x_6 = 1. \end{cases}$$

È un sistema a scala $Sx = c$, dove

$$S = \begin{pmatrix} 0 & 1 & -1 & 3 & -1 & 1/2 \\ 0 & 0 & 0 & 2 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{pmatrix} \quad \text{e} \quad c = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

La matrice $S \in M_{3,6}(\mathbb{R})$ ha tre pivot, $p_1 = 1$ in colonna 2, $p_2 = 2$ in colonna 4, e $p_3 = 1$ in colonna 5. Dalla terza equazione possiamo ricavare x_5 in funzione di x_6 : si ha $x_5 = 1 - 4x_6$. Analogamente, la seconda equazione ci fornisce x_4 in funzione di x_6 : abbiamo $x_4 = \frac{1}{2}x_6$. Sostituendo quanto ottenuto nella prima equazione, ricaviamo anche x_2 in funzione di x_3 e x_6 : $x_2 = x_3 - 6x_6 + 2$. Infine, x_1 è libera, in quanto il sistema non impone alcuna condizione su di essa. Riassumendo, le soluzioni del nostro sistema dipendono da 3 parametri liberi, e sono date da

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_3 - 6x_6 + 2 \\ x_3 \\ x_6/2 \\ 1 - 4x_6 \\ x_6 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_6 \begin{pmatrix} 0 \\ -6 \\ 0 \\ 1/2 \\ -4 \\ 1 \end{pmatrix}. \quad (6.1)$$

Il metodo utilizzato per risolvere questo sistema a scala si chiama, per motivi ovvi, *risoluzione all'indietro*. Come si applica è probabilmente già chiaro dall'Esempio, ma per completezza vediamo una descrizione in generale. Sia $Sx = c$ un sistema a scala, dove S è una matrice con r pivot situati sulle colonne j_1, \dots, j_r (e sulle righe $1, \dots, r$, ovviamente), e supponiamo che c appartenga a V_r , in modo che il sistema sia compatibile. Allora, essendo $p_r \neq 0$ per definizione, dalla r -esima equazione possiamo ricavare x_{j_r} in funzione delle variabili successive. Sostituiamo quanto ottenuto nell'equazione precedente, e ricaviamo $x_{j_{r-1}}$ in funzione delle variabili successive (x_{j_r} esclusa). Procediamo così fino a giungere alla prima equazione, da cui ricaviamo x_{j_1} in funzione delle variabili successive (x_{j_2}, \dots, x_{j_r} escluse).

Dunque la soluzione risulta essere la somma di un vettore costante (la soluzione particolare v^o della Proposizione 5.1) e del sottospazio (di dimensione $n - r$) delle soluzioni del sistema omogeneo associato $Sx = O$, dipendente da $n - r$ parametri, le variabili che non corrispondono ai pivot, le cosiddette *variabili libere*. Le altre variabili, quelle corrispondenti ai pivot e che si esprimono in funzione delle variabili libere, si chiamano *variabili dipendenti*, e sono esattamente $r = \text{rg } S$.

Osservazione 6.2 La risoluzione all'indietro fornisce sempre le soluzioni nella forma

$$v = v^o + x_{i_1} w_1 + \dots + x_{i_{n-r}} w_{n-r},$$

dove $x_{i_1}, \dots, x_{i_{n-r}}$ sono le variabili libere, $v^o \in \mathbb{R}^n$ è una soluzione particolare, e $w_1, \dots, w_{n-r} \in \mathbb{R}^n$ sono un sistema di generatori di $\text{Ker } S$. In realtà, i w_j sono proprio una base di $\text{Ker } S$. Questo lo si può vedere in due modi: o direttamente, andando a notare dove sono messi gli 1 in questi vettori, o ricordandosi l'Esercizio 4.21, in quanto sappiamo già che $\dim \text{Ker } S = n - r$.

6.2 La riduzione a scala

Vediamo ora il caso generale. La *riduzione a scala* di matrici qualunque è un algoritmo (simile all'eliminazione di Gauss) che tramite opportune operazioni elementari trasforma una matrice $A \in M_{m,n}(\mathbb{R})$ in una matrice a scala $S \in M_{m,n}(\mathbb{R})$ in modo da trasformare qualunque sistema $Ax = b$ in un sistema a scala $Sx = c$ a esso equivalente, chiamato *riduzione a scala* di $Ax = b$.

Come l'eliminazione di Gauss, anche la riduzione a scala procede per passi successivi: per l'esattezza, in numero uguale al rango della matrice. Il passo i -esimo opera in maniera tale da: trasformare la riga i -esima della matrice dei coefficienti nella forma adatta a una matrice a scala; produrre un numero reale non nullo $p_i \in \mathbb{R}^*$, l' i -esimo pivot della matrice, situato nella colonna j_i ; annullare tutti gli elementi della matrice sotto la i -esima riga nelle colonne dalla prima alla j_i -esima inclusa — e tutto sempre senza cambiare le soluzioni del sistema.

Vediamo i dettagli. Partiamo da un sistema lineare $Ax = b$ con $A \in M_{m,n}(\mathbb{R})$ e $b \in \mathbb{R}^n$. Se A è la matrice nulla, abbiamo finito. Altrimenti, sia A^{j_1} la prima colonna non nulla di A ; scambiando se necessario la prima riga con una sottostante possiamo supporre che sia il primo elemento a_{1j_1} della colonna A^{j_1} a essere diverso da zero; poniamo $p_1 = a_{1j_1}$. Sommiamo ora alla riga h -esima (con $h = 2, \dots, n$) un adeguato multiplo della prima riga in modo da annullare tutti gli elementi della j_1 -esima colonna tranne il primo. In questo modo la nostra matrice diventa della forma

$$\begin{pmatrix} 0 & \dots & 0 & p_1 & * & \dots & * \\ \vdots & & \vdots & 0 & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & * & \dots & * \end{pmatrix},$$

con $p_1 \neq 0$.

I passi successivi sono molto simili. Supponiamo di aver già trattato le prime $i - 1$ righe; se le righe dalla i -esima in poi sono tutte nulle (oppure se la matrice ha solo $i - 1$ righe) abbiamo finito. Altrimenti, sia j_i l'indice della prima colonna che contiene un elemento non nullo in una riga sotto la $(i - 1)$ -esima. Scambiando se necessario la riga i -esima con una sottostante possiamo supporre che l'elemento diverso da zero della colonna A^{j_i} sia proprio sulla riga i -esima; poniamo p_i uguale a questo elemento. Ora sommiamo alla riga h -esima (per $h = i + 1, \dots, n$) un adeguato multiplo della i -esima riga in modo da annullare tutti gli elementi della colonna A^{j_i} sotto la riga i -esima. In questo modo arriviamo a una matrice della forma

$$\begin{pmatrix} 0 & \dots & 0 & p_1 & * & \dots & * & \dots & \dots & \dots & \dots & \dots & * \\ \vdots & & \vdots & 0 & \dots & \dots & 0 & p_2 & * & \dots & \dots & \dots & * \\ \vdots & & \vdots & \vdots & & & 0 & 0 & & & & & \vdots \\ \vdots & & \vdots & \vdots & & & \vdots & \vdots & \vdots & & 0 & p_i & * & \dots & * \\ \vdots & & \vdots & \vdots & & & \vdots & \vdots & \vdots & & 0 & 0 & * & \dots & * \\ \vdots & & \vdots & \vdots & & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & \dots & \dots & 0 & 0 & 0 & \dots & 0 & 0 & * & \dots & * \end{pmatrix},$$

con $p_1, \dots, p_i \neq 0$.

Procedendo in questo modo arriviamo prima o poi a una matrice in cui le ultime righe sono tutte nulle, oppure in cui l'ultimo pivot appartiene all'ultima riga; in entrambi i casi abbiamo ridotto a scala la nostra matrice di partenza.

Ma vediamo come funziona la riduzione a scala su un esempio.

ESEMPIO 6.2 Consideriamo il sistema

$$\begin{cases} 2x_1 - x_2 + 4x_3 + x_4 = -2, \\ -2x_1 + x_2 - 7x_3 + x_4 = -1, \\ 4x_1 - 2x_2 + 5x_3 + 4x_4 = -7, \end{cases}$$

con matrice A e vettore dei termini noti b dati da

$$A = \begin{pmatrix} 2 & -1 & 4 & 1 \\ -2 & 1 & -7 & 1 \\ 4 & -2 & 5 & 4 \end{pmatrix}, \quad b = \begin{pmatrix} -2 \\ -1 \\ -7 \end{pmatrix}.$$

La matrice A non è nulla, e il primo elemento della prima colonna è diverso da zero, per cui otteniamo $j_1 = 1$ e $p_1 = 2$. Per concludere il primo passo dobbiamo sommare la prima equazione alla seconda, e sottrarre il doppio della prima alla terza, ottenendo

$$\begin{pmatrix} 2 & -1 & 4 & 1 \\ -2 & 1 & -7 & 1 \\ 4 & -2 & 5 & 4 \end{pmatrix} \begin{pmatrix} -2 \\ -1 \\ -7 \end{pmatrix} \rightarrow \begin{pmatrix} 2 & -1 & 4 & 1 \\ 0 & 0 & -3 & 2 \\ 0 & 0 & -3 & 2 \end{pmatrix} \begin{pmatrix} -2 \\ -3 \\ -3 \end{pmatrix}.$$

Passiamo al passo 2; siccome la seconda colonna ha tutti zero dalla seconda riga in giù, mentre la terza colonna ha un termine non nullo nella seconda riga, troviamo $j_2 = 3$ e $p_2 = -3$. A questo punto dobbiamo sottrarre la seconda riga alla terza, e otteniamo

$$\left| \begin{array}{cccc|c} 2 & -1 & 4 & 1 & -2 \\ 0 & 0 & -3 & 2 & -3 \\ 0 & 0 & -3 & 2 & -3 \end{array} \right| \rightarrow \left| \begin{array}{cccc|c} 2 & -1 & 4 & 1 & -2 \\ 0 & 0 & -3 & 2 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right|$$

La terza riga è nulla, per cui la riduzione a scala termina qui. La matrice S ottenuta è una matrice a scala con due pivot, e il relativo vettore c dei termini noti appartiene a V_2 (cioè ha nulli gli elementi dal terzo in poi); quindi il Corollario 6.2 ci assicura che il sistema $Sx = c$ (e quindi il nostro sistema iniziale) è compatibile. Risolvendo all'indietro, troviamo

$$\begin{cases} x_1 = \frac{1}{2}x_2 - \frac{11}{6}x_4 - 3, \\ x_3 = \frac{2}{3}x_4 + 1, \end{cases}$$

cioè

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -3 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 1/2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -11/6 \\ 0 \\ 2/3 \\ 1 \end{pmatrix},$$

con due variabili dipendenti (x_1 e x_3 , corrispondenti alle colonne dei pivot) e due variabili libere (x_2 e x_4), coerentemente con quanto ci aspettavamo.

ESEMPIO 6.3 Consideriamo invece il sistema

$$\begin{cases} 2x_1 - x_2 + 4x_3 + x_4 = -2, \\ -2x_1 + x_2 - 7x_3 + x_4 = -1, \\ 4x_1 - 2x_2 + 5x_3 + 4x_4 = 7, \end{cases}$$

con matrice A e vettore dei termini noti b dati da

$$A = \begin{pmatrix} 2 & -1 & 4 & 1 \\ -2 & 1 & -7 & 1 \\ 4 & -2 & 5 & 4 \end{pmatrix}, \quad b = \begin{pmatrix} -2 \\ -1 \\ 7 \end{pmatrix}.$$

In questo caso la riduzione a scala ci dà

$$\left| \begin{array}{cccc|c} 2 & -1 & 4 & 1 & -2 \\ -2 & 1 & -7 & 1 & -1 \\ 4 & -2 & 5 & 4 & 7 \end{array} \right| \rightarrow \left| \begin{array}{cccc|c} 2 & -1 & 4 & 1 & -2 \\ 0 & 0 & -3 & 2 & -3 \\ 0 & 0 & -3 & 2 & 11 \end{array} \right| \rightarrow \left| \begin{array}{cccc|c} 2 & -1 & 4 & 1 & -2 \\ 0 & 0 & -3 & 2 & -3 \\ 0 & 0 & 0 & 0 & 14 \end{array} \right|$$

Dunque stavolta il vettore c non appartiene allo spazio V_2 , per cui il sistema originale non è compatibile — non ammette soluzioni.

Vediamo ora di riassumere le relazioni che intercorrono fra un sistema $Ax = b$ e una sua riduzione a scala $Sx = c$.

Teorema 6.3 Sia $Ax = b$ un sistema lineare, e $Sx = c$ una sua riduzione a scala. Allora:

- (i) lo spazio delle soluzioni di $Ax = b$ coincide con lo spazio delle soluzioni di $Sx = c$;
- (ii) $\text{Ker } A = \text{Ker } S$;
- (iii) $\text{rg } A = \text{rg } S$ (ma in generale $\text{Im } A \neq \text{Im } S$);
- (iv) siano S^{j_1}, \dots, S^{j_r} , dove $r = \text{rg } S$, le colonne corrispondenti ai pivot di S ; allora $\{A^{j_1}, \dots, A^{j_r}\}$ è una base di $\text{Im } A$.

Dimostrazione. (i) Infatti i due sistemi sono equivalenti.

(ii) Il sistema $Ax = 0$ è equivalente al sistema $Sx = 0$.

(iii) Sia n il numero di colonne di A (ed S). Allora il Teorema 5.7 ci dice che

$$\text{rg } A = n - \dim \text{Ker } A = n - \dim \text{Ker } S = \text{rg } S.$$

(iv) Essendo $\text{rg } A = \text{rg } S$, basta dimostrare (perché?) che le colonne $\{A^{j_1}, \dots, A^{j_r}\}$ sono linearmente indipendenti, cioè che il sistema

$$x_1 A^{j_1} + \dots + x_r A^{j_r} = 0$$

ammette solo la soluzione $x_1 = \dots = x_r = 0$. Ma applicando la stessa riduzione a scala che ha trasformato A in S riduciamo questo sistema nel sistema a scala equivalente

$$x_1 S^{j_1} + \dots + x_r S^{j_r} = 0,$$

che nella dimostrazione del Lemma 6.1 abbiamo visto avere solo la soluzione nulla. \square

Osservazione 6.3 Anche stavolta i pivot di una matrice non sono univocamente determinati, in quanto vi è libertà di scelta nei quali righe scambiare. Quello che però non dipende dalle scelte effettuate è il numero dei pivot ottenuti. Infatti, il numero dei pivot di una matrice a scala è uguale al suo rango; siccome, grazie al Teorema 6.3.(iii), il rango di una riduzione a scala S di una matrice A è uguale al rango di A , il numero dei pivot di una qualunque riduzione a scala di A è esattamente uguale a $\text{rg } A$.

Osservazione 6.4 Benché il rango di A sia uguale a quello di una sua riduzione a scala S , l'immagine di L_A è in generale diversa dall'immagine di L_S (controllato sull'Esempio 6.2). Questo è dovuto al fatto che la riduzione a scala opera sulle righe della matrice, mentre l'immagine dipende dalle colonne.

6.3 Tecniche di calcolo

A questo punto abbiamo finalmente tutti gli strumenti necessari per la risoluzione dei nostri problemi.

(a) Per risolvere il sistema $Ax = b$: basta applicare la riduzione a scala al nostro sistema per trasformarlo in un sistema a scala equivalente, la cui compatibilità è subito vista col Corollario 6.2, e che viene risolto facilmente con la risoluzione all'indietro.

(b) Per trovare il rango e una base dell'immagine di una matrice A : basta applicare la riduzione a scala per trasformarla in una matrice a scala dello stesso rango, e utilizzare il Teorema 6.3 e il Lemma 6.1.

(c) Per trovare dimensione e base del nucleo di un'applicazione lineare L_A : basta applicare la riduzione a scala ad A per trasformarla in una matrice a scala S e poi trovare $\text{Ker } S$; questo ci fornisce dimensione e base di $\text{Ker } A$ grazie al Teorema 6.3 e all'Osservazione 6.2.

(d) Per trovare dimensione e base del sottospazio generato da $v_1, \dots, v_k \in \mathbb{R}^n$: consideriamo la matrice

$$A = \begin{vmatrix} v_1 & \dots & v_k \end{vmatrix} \in M_{n,k}(\mathbb{R}),$$

in cui abbiamo messo i vettori v_j per colonna. Siccome $\text{Span}(v_1, \dots, v_k) = \text{Im } A$, la dimensione dello span è il rango di A , e una base la si trova applicando il Teorema 6.3.(iv).

(e) Per completare un insieme $\{v_1, \dots, v_k\}$ di vettori linearmente indipendenti di \mathbb{R}^n a una base: basta estrarre una base dall'insieme $\{v_1, \dots, v_k, e_1, \dots, e_n\}$ col metodo del punto (d), dove $\{e_1, \dots, e_n\}$ è la base canonica di \mathbb{R}^n . Infatti la riduzione a scala è fatta in modo tale da assicurare (perché?) la presenza dei vettori v_1, \dots, v_k nella base estratta.

(f) Per trovare dimensione e base di $U + W$, dove U e W sono sottospazi di \mathbb{R}^n : sia $\mathcal{B} \subset \mathbb{R}^n$ una base di U , e $\mathcal{C} \subset \mathbb{R}^n$ una base di W ; allora $U + W$ è generato da $\mathcal{B} \cup \mathcal{C}$, per cui basta applicare il punto (d) a $\mathcal{B} \cup \mathcal{C}$.

(g) Per trovare dimensione e base di $U \cap W$, dove U e W sono sottospazi di \mathbb{R}^n : sia $\{u_1, \dots, u_r\} \subset \mathbb{R}^n$ una base di U , $\{w_1, \dots, w_s\} \subset \mathbb{R}^n$ una base di W , e consideriamo la matrice¹

$$A = \begin{vmatrix} u_1 & \dots & u_r & w_1 & \dots & w_s \end{vmatrix} \in M_{n,r+s}(\mathbb{R}).$$

Sappiamo (perché?) che $U + W = \text{Im } A$; dunque i Teoremi di Grassmann e della dimensione implicano

$$\dim(U \cap W) = \dim U + \dim W - \dim(U + W) = r + s - \text{rg } A = \dim \text{Ker } A.$$

Dunque per calcolare la dimensione dell'intersezione basta applicare il punto (c) alla matrice A ; per trovare invece una base di $U \cap W$ bisogna lavorare ancora. Un vettore

¹ Se $r < s$, conviene invece mettere prima i w_j e poi gli u_i per ridurre i conti che dovremo fare in seguito.

$v \in \mathbb{R}^n$ appartiene a $U \cap W$ se e solo se esistono (unici) $x_1, \dots, x_r, y_1, \dots, y_s \in \mathbb{R}$ tali che

$$x_1 u_1 + \dots + x_r u_r = v = y_1 w_1 + \dots + y_s w_s, \quad (6.2)$$

e quindi se e solo se

$$\begin{pmatrix} -x_1 \\ \vdots \\ -x_r \\ y_1 \\ \vdots \\ y_s \end{pmatrix} \in \text{Ker } A. \quad (6.3)$$

Dunque ponendo

$$\psi(v) = \begin{pmatrix} -x_1 \\ \vdots \\ -x_r \\ y_1 \\ \vdots \\ y_s \end{pmatrix}$$

definiamo un'applicazione $\psi: U \cap W \rightarrow \text{Ker } A$ lineare, iniettiva (esercizio) e quindi surgettiva, per il Corollario 5.8.(iii). L'applicazione inversa $\psi^{-1}: \text{Ker } A \rightarrow U \cap W$ è data da

$$\psi^{-1} \begin{pmatrix} -x_1 \\ \vdots \\ -x_r \\ y_1 \\ \vdots \\ y_s \end{pmatrix} = y_1 w_1 + \dots + y_s w_s, \quad (6.4)$$

grazie a (6.2) e (6.3). Dunque se troviamo² col punto (c) una base $\{z_1, \dots, z_p\}$ di $\text{Ker } A$, allora $\{\psi^{-1}(z_1), \dots, \psi^{-1}(z_p)\}$ sarà una base di $U \cap W$ (in quanto è un sistema di generatori per il Lemma 5.6, e quindi una base per l'Esercizio 4.21).

Vediamo ora questi metodi in azione su alcuni esempi.

ESEMPIO 6.4 Vogliamo dimensione e base di $\text{Ker } L_A$ e $\text{Im } L_A$, dove

$$A = \begin{pmatrix} 1 & 0 & 3 \\ 2 & 1 & 1 \\ 3 & 1 & 4 \\ -1 & -1 & 2 \end{pmatrix} \in M_{4,3}(\mathbb{R}).$$

² In realtà, non avremo bisogno di calcolare completamente i vettori della base di $\text{Ker } A$ ma soltanto le loro ultime s coordinate, le uniche che servono per scrivere ψ^{-1} . Questa osservazione riduce lievemente i conti necessari, come vedremo nell'Esempio 6.6.

Dobbiamo prima di tutto ridurre a scala la matrice A :

$$\left| \begin{array}{ccc|c} 1 & 0 & 3 & \\ 2 & 1 & 1 & \\ 3 & 1 & 4 & \\ -1 & -1 & 2 & \end{array} \right| \rightarrow \left| \begin{array}{ccc|c} 1 & 0 & 3 & \\ 0 & 1 & -5 & \\ 0 & 1 & -5 & \\ 0 & -1 & 5 & \end{array} \right| \rightarrow \left| \begin{array}{ccc|c} 1 & 0 & 3 & \\ 0 & 1 & -5 & \\ 0 & 0 & 0 & \\ 0 & 0 & 0 & \end{array} \right| = S.$$

Dunque abbiamo trovato due pivot; questo vuol dire che $\text{rg } A = \dim \text{Im } L_A = 2$ e $\dim \text{Ker } L_A = 1$. Inoltre sappiamo anche che una base di $\text{Im } A$ è data da $\{A^1, A^2\}$. Per trovare una base del nucleo, dobbiamo semplicemente risolvere all'indietro il sistema $Sx = 0$; troviamo $x_1 = -3x_3$ e $x_2 = 5x_3$, per cui

$$\text{Ker } L_A = \mathbb{R} \begin{pmatrix} -3 \\ 5 \\ 1 \end{pmatrix}.$$

ESEMPIO 6.5 Vogliamo dimensione e base di $U = \text{Span}(u_1, u_2, u_3) \subset \mathbb{R}^4$, dove

$$u_1 = \begin{pmatrix} 1 \\ -1 \\ 3 \\ 2 \end{pmatrix}, \quad u_2 = \begin{pmatrix} -2 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \quad u_3 = \begin{pmatrix} 8 \\ -5 \\ 6 \\ 1 \end{pmatrix}.$$

La matrice A è data da

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

Applicando la riduzione a scala otteniamo

$$S = \begin{pmatrix} 1 & -2 & 8 \\ 0 & -1 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix};$$

quindi $\dim U = \text{rg } S = 2$, e una base di U è data da $\{u_1, u_2\}$.

ESEMPIO 6.6 Vogliamo dimensione e base di $U + W$ e $U \cap W$, dove $U \subset \mathbb{R}^4$ è il sottospazio dell'Esempio precedente mentre $W = \text{Span}(w_1, w_2) \subset \mathbb{R}^4$, con

$$w_1 = \begin{pmatrix} 1 \\ 0 \\ -3 \\ -3 \end{pmatrix}, \quad w_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

Una veloce riduzione a scala mostra che $\dim W = 2$, cioè che i vettori w_1 e w_2 sono linearmente indipendenti. Stavolta la matrice da considerare è

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

Effettuiamo una riduzione a scala di A , ottenendo

$$S = \begin{pmatrix} 1 & -2 & 1 & 0 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Dunque $\dim(U + W) = \text{rg } A = \text{rg } S = 3$, e

$$\dim(U \cap W) = \dim \text{Ker } A = \dim \text{Ker } S = 4 - \text{rg } S = 1.$$

Guardando la matrice S — o meglio, guardando dove sono situati i suoi pivot — vediamo subito che una base di $U + W$ è data da $\{u_1, u_2, w_2\}$; per trovare una base di $U \cap W$ dobbiamo studiare $\text{Ker } A = \text{Ker } S$. Il sistema omogeneo di matrice S è

$$\begin{cases} x_1 - 2x_2 + y_1 = 0, \\ -x_2 + y_1 + y_2 = 0, \\ 6y_2 = 0, \end{cases}$$

da cui si vede subito che $y_2 = 0$ e che y_1 è libera. Dunque una base di $\text{Ker } S$ è $\{z\}$, dove $z \in \mathbb{R}^4$ è un vettore della forma

$$z = \begin{pmatrix} * \\ * \\ 1 \\ 0 \end{pmatrix}$$

e gli asterischi servono a indicare che non ci interessano gli effettivi valori delle prime due coordinate di z , in quanto (vedi la Nota 2) in (6.4) compaiono solo le y_j . Quindi una base di $U \cap W$ è $\{\psi^{-1}(z)\}$, che si ottiene facendo una combinazione lineare della base di W usando come coefficienti gli elementi di z . In altre parole,

$$\psi^{-1}(z) = 1w_1 + 0w_2,$$

per cui una base di $U \cap W$ è data da $\{w_1\}$.

Dunque ora siamo in grado di risolvere tutti i nostri problemi su \mathbb{R}^n ; come procedere in uno spazio vettoriale V qualunque lo vedremo nel Paragrafo 8.1 (ma non sarà molto diverso).

6.4 Equazioni parametriche e cartesiane

L'obiettivo di questo paragrafo è vedere come si possono rappresentare i sottospazi vettoriali (e poi i sottospazi affini) di \mathbb{R}^n .

Ora, come si descrive un sottospazio o, più in generale, un sottoinsieme X di \mathbb{R}^n ? Ci sono essenzialmente due possibilità: tramite una parametrizzazione, o tramite equazioni.

- (a) *Parametrizzazione*: vuol dire che esiste un insieme $U \subseteq \mathbb{R}^k$ (con di solito $k \neq n$) e una applicazione (possibilmente iniettiva) $F: U \rightarrow \mathbb{R}^n$ tale che $F(U) = X$. In altri termini, X è l'immagine di una applicazione; un punto $x \in \mathbb{R}^n$ appartiene a X se e solo se $x = F(t)$ per qualche $t \in U$. L'applicazione F è detta *parametrizzazione* di X , e gli elementi di U *parametri*. Abbiamo visto esempi di parametrizzazione di rette e piani (con $U = \mathbb{R}$ o $U = \mathbb{R}^2$, rispettivamente) nel Paragrafo 2.3; più avanti vedremo come parametrizzare anche altri oggetti.
- (b) *Equazioni*: vuol dire che indichiamo quali condizioni deve soddisfare un punto di \mathbb{R}^n per appartenere a X . Formalmente, scegliamo un sottoinsieme $V \subseteq \mathbb{R}^n$ (con spesso $V = \mathbb{R}^n$) e un'applicazione $G: V \rightarrow \mathbb{R}^k$ (con $k \neq n$, di solito), e diciamo che $x \in X$ se e solo se $G(x) = O$; queste sono *equazioni cartesiane* per X . Per esempio, se prendiamo $A \in M_{m,n}(\mathbb{R})$ e $G = L_A: \mathbb{R}^n \rightarrow \mathbb{R}^m$, allora $G(x) = O$ sono equazioni che descrivono $\text{Ker } A$ — ma ne ripareremo più avanti.

ESEMPIO 6.7 Supponiamo di voler descrivere la circonferenza X di centro l'origine e raggio 1 nel piano. Una possibile parametrizzazione è data da $F: \mathbb{R} \rightarrow \mathbb{R}^2$ con $F(t) = (\cos t, \sin t)$. Se invece prendiamo la funzione $G: \mathbb{R}^2 \rightarrow \mathbb{R}$ data da $G(x) = x_1^2 + x_2^2 - 1$, allora $G(x) = 0$ è un'equazione cartesiana della circonferenza X .

Vediamo come si concretizzano queste descrizioni nel caso dei sottospazi vettoriali, cominciando con le parametrizzazioni. Sia V un sottospazio di dimensione d di \mathbb{R}^n , e fissiamo una base $\{v_1, \dots, v_d\}$ di V . Allora possiamo definire una parametrizzazione $F: \mathbb{R}^d \rightarrow \mathbb{R}^n$ di V ponendo

$$F \begin{pmatrix} t_1 \\ \vdots \\ t_d \end{pmatrix} = t_1 v_1 + \dots + t_d v_d.$$

Nota che F è data da una matrice; per l'esattezza, essendo $F(e_j) = v_j$ per $j = 1, \dots, d$, la Proposizione 5.4 ci dice che $F = L_A$, dove A è la matrice

$$A = |v_1 \ \dots \ v_d| \in M_{n,d}(\mathbb{R}).$$

Questo ci suggerisce una definizione.

Definizione 6.2 Una *parametrizzazione lineare* di un sottospazio V di \mathbb{R}^n è un'applicazione lineare della forma $L_A: \mathbb{R}^k \rightarrow \mathbb{R}^n$, dove $A \in M_{n,k}(\mathbb{R})$, tale che $V = \text{Im } A$; in altri termini, V viene descritto come span delle colonne di A . In coordinate, i

vettori $v \in V$ saranno dati dalle equazioni

$$v = t_1 A^1 + \dots + t_k A^k,$$

per cui *equazioni parametriche* del sottospazio $V \subseteq \mathbb{R}^n$ sono

$$\begin{cases} x_1 = a_{11}t_1 + \dots + a_{1k}t_k, \\ \vdots \\ x_n = a_{n1}t_1 + \dots + a_{nk}t_k, \end{cases}$$

dove abbiamo posto $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

Osservazione 6.5 È chiaro che le equazioni parametriche si guardano bene dall'essere uniche. Inoltre, $\dim V = \text{rg } A \leq k$; quindi il numero dei parametri in una data parametrizzazione potrebbe essere eccessivo, superiore alla dimensione del sottospazio (è quanto capitava nell'Esempio 4.6) se le colonne della matrice A non sono linearmente indipendenti (cioè se l'applicazione L_A non è iniettiva). Di solito, in tal caso si sostituisce ad A una matrice \tilde{A} di ugual immagine e con rango esattamente uguale a $\dim V$, ottenuta per esempio considerando solo le colonne di A che formano una base di $\text{Im } A = V$.

ESEMPIO 6.8 Vogliamo equazioni parametriche del sottospazio $U \subset \mathbb{R}^4$ dell'Esempio 6.5. Un primo insieme di equazioni parametriche l'abbiamo subito:

$$\begin{cases} x_1 = t_1 - 2t_2 + 8t_3, \\ x_2 = -t_1 + t_2 - 5t_3, \\ x_3 = 3t_1 + 6t_3, \\ x_4 = 2t_1 + t_2 + t_3. \end{cases}$$

In realtà, come abbiamo visto nell'Esempio 6.5, ci sono dei parametri superflui in quanto la dimensione di U è 2, non 3. Utilizzando la base di U che avevamo trovato in quell'Esempio otteniamo come equazioni parametriche

$$\begin{cases} x_1 = t_1 - 2t_2, \\ x_2 = -t_1 + t_2, \\ x_3 = 3t_1, \\ x_4 = 2t_1 + t_2. \end{cases}$$

Vediamo ora le equazioni cartesiane. Dato che con le parametriche ce la siamo cavata usando soltanto applicazioni lineari (o meglio, matrici), cerchiamo di fare altrettanto in questo caso. Dato un sottospazio V di \mathbb{R}^n vogliamo descriverlo come nucleo di una matrice $B \in M_{p,n}(\mathbb{R})$, cioè dire che $x \in V$ se e solo se $Bx = O$. In

coordinate, $x \in V$ se e solo se

$$\begin{cases} b_{11}x_1 + \cdots + b_{1n}x_n = 0, \\ \vdots \\ b_{p1}x_1 + \cdots + b_{pn}x_n = 0. \end{cases} \quad (6.5)$$

Definizione 6.3 Le (6.5) sono equazioni cartesiane (lineari) per il sottospazio V , che risulta essere lo spazio delle soluzioni del sistema omogeneo di matrice B .

Osservazione 6.6 In questo caso, la dimensione di V è data da $n - \text{rg } B$ (Osservazione 5.8), con $\text{rg } B \leq p$. Dunque in generale la dimensione di V potrebbe essere maggiore del numero di incognite (n) meno il numero di equazioni (p), in quanto potrebbero esservi delle equazioni inutili (come accade nel sistema omogeneo associato al sistema studiato nell'Esempio 6.2). Di solito, in tal caso si sostituisce a B una matrice \tilde{B} di ugual nucleo e con esattamente $\text{rg } B = n - \dim V$ righe, ottenuta per esempio tramite una riduzione a scala di B . In particolare, anche le equazioni cartesiane si guardano bene dall'essere uniche.

ESEMPIO 6.9 Il sottospazio $V \subset \mathbb{R}^4$ delle soluzioni del sistema

$$\begin{cases} 2x_1 - x_2 + 4x_3 + x_4 = 0, \\ -2x_1 + x_2 - 7x_3 + x_4 = 0, \\ 4x_1 - 2x_2 + 5x_3 + 4x_4 = 0, \end{cases}$$

ha come equazioni cartesiane il sistema appena scritto, cioè il sistema $Bx = 0$ con

$$B = \begin{vmatrix} 2 & -1 & 4 & 1 \\ -2 & 1 & -7 & 1 \\ 4 & -2 & 5 & 4 \end{vmatrix}.$$

Per quanto visto nell'Esempio 6.2, a B possiamo sostituire la matrice

$$\tilde{B} = \begin{vmatrix} 2 & -1 & 4 & 1 \\ 0 & 0 & -3 & 2 \end{vmatrix},$$

ottenendo

$$\begin{cases} 2x_1 - x_2 + 4x_3 + x_4 = 0, \\ -3x_3 + 2x_4 = 0, \end{cases} \quad (6.6)$$

come altre equazioni cartesiane per V .

A questo punto si pongono due problemi: è vero che ogni sottospazio di \mathbb{R}^n è descrivibile da equazioni cartesiane lineari? E in tal caso, come si passa da equazioni parametriche a equazioni cartesiane o viceversa?

Passare da equazioni cartesiane a equazioni parametriche è facile: basta effettuare una riduzione a scala e risolvere all'indietro il sistema omogeneo corrispondente.

ESEMPIO 6.10 Vogliamo equazioni parametriche per il sottospazio $V \subset \mathbb{R}^4$ dell'Esempio 6.9. Risolvendo all'indietro il sistema (6.6) troviamo

$$\begin{cases} x_1 = \frac{1}{2}x_2 - \frac{11}{6}x_4, \\ x_3 = \frac{2}{3}x_4; \end{cases}$$

quindi equazioni parametriche sono

$$\begin{cases} x_1 = \frac{1}{2}t_1 - \frac{11}{6}t_2, \\ x_2 = t_1, \\ x_3 = \frac{2}{3}t_2, \\ x_4 = t_2, \end{cases}$$

ovvero $V = \text{Im } A$ con

$$A = \begin{vmatrix} 1/2 & -11/6 \\ 1 & 0 \\ 0 & 2/3 \\ 0 & 1 \end{vmatrix}.$$

Viceversa, si può sempre passare da equazioni parametriche a equazioni cartesiane (per cui, in particolare, ogni sottospazio può venire descritto da equazioni cartesiane lineari). Sia $V = \text{Im } A \subseteq \mathbb{R}^n$, in modo che i vettori $v \in V$ siano della forma $v = At$. Per passare a equazioni cartesiane, dobbiamo trovare quali condizioni devono soddisfare le coordinate di $v \in \mathbb{R}^n$ perché il sistema $At = v$ ammetta soluzione. Effettuiamo una riduzione a scala; otteniamo il sistema equivalente $St = c$, dove S è a scala con le ultime $n - \text{rg } A$ righe nulle. Dunque il sistema $At = v$ è compatibile se e solo se le ultime $n - \text{rg } A$ componenti di c sono uguali a zero — e queste sono le equazioni cartesiane cercate.

ESEMPIO 6.11 Applichiamo questo metodo al sottospazio $U = \text{Im } A$ dell'Esempio 6.8, dove

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

Operando una riduzione a scala sul sistema $At = v$ otteniamo

$$\left[\begin{array}{cc|c} 1 & -2 & x_1 \\ -1 & 1 & x_2 \\ 3 & 0 & x_3 \\ 2 & 1 & x_4 \end{array} \right] \rightarrow \left[\begin{array}{cc|c} 1 & -2 & x_1 \\ 0 & -1 & x_2 + x_1 \\ 0 & 6 & x_3 - 3x_1 \\ 0 & 5 & x_4 - 2x_1 \end{array} \right] \rightarrow \left[\begin{array}{cc|c} 1 & -2 & x_1 \\ 0 & -1 & x_2 + x_1 \\ 0 & 0 & x_3 - 3x_1 + 6(x_2 + x_1) \\ 0 & 0 & x_4 - 2x_1 + 5(x_2 + x_1) \end{array} \right],$$

ovvero

$$\left[\begin{array}{cc|c} 1 & -2 & x_1 \\ 0 & -1 & x_1 + x_2 \\ 0 & 0 & 3x_1 + 6x_2 + x_3 \\ 0 & 0 & 3x_1 + 5x_2 + x_4 \end{array} \right]$$

per cui delle possibili equazioni cartesiane sono

$$\begin{cases} 3x_1 + 6x_2 + x_3 = 0, \\ 3x_1 + 5x_2 + x_4 = 0. \end{cases}$$

Osservazione 6.7 Un altro metodo per passare da equazioni parametriche a equazioni cartesiane verrà descritto nell'Esercizio 7.24.

Osservazione 6.8 È molto facile trovare le equazioni cartesiane di una intersezione. Infatti se $B_1x = O$ sono equazioni cartesiane del sottospazio $U \subseteq \mathbb{R}^n$, e $B_2x = O$ sono equazioni cartesiane del sottospazio $W \subseteq \mathbb{R}^n$, allora $x \in U \cap W$ se e solo se soddisfa sia $B_1x = O$ che $B_2x = O$, per cui le equazioni cartesiane di $U \cap W$ sono semplicemente

$$\begin{cases} B_1x = O, \\ B_2x = O. \end{cases}$$

6.5 Sottospazi affini

Abbiamo quindi visto come descrivere i sottospazi vettoriali di \mathbb{R}^n , e dunque in particolare l'insieme delle soluzioni di un sistema lineare omogeneo. In questo paragrafo discuteremo come descrivere l'insieme delle soluzioni di un sistema lineare non omogeneo — o meglio, l'analogo in spazi vettoriali qualunque.

Definizione 6.4 Un *sottospazio affine* L di uno spazio vettoriale V è un sottoinsieme di V della forma

$$L = v_0 + W = \{v_0 + w \mid w \in W\},$$

dove $v_0 \in V$ e W è un sottospazio vettoriale di V ; diremo che L è *parallelo* a W , e che W è il *sottospazio di giacitura* di L . La *dimensione* $\dim L$ di L è per definizione uguale alla dimensione di W .

Osservazione 6.9 Di solito, un sottospazio affine di dimensione 1 si chiama *retta*, e uno di dimensione 2 *piano*. A volte, un sottospazio affine di V di dimensione $\dim V - 1$ viene chiamato *iperpiano*.

In parole povere, un sottospazio affine è un *traslato* di un sottospazio vettoriale. Per esempio, una retta che non passa per l'origine, o un piano che non passa³ per l'origine, sono sottospazi affini. Inoltre, il Teorema di struttura (Proposizione 5.1) ci dice che l'insieme delle soluzioni di un sistema lineare $Ax = b$ è un sottospazio affine, parallelo allo spazio delle soluzioni del sistema omogeneo associato $Ax = O$.

³ O anche che ci passa; i sottospazi vettoriali sono particolari sottospazi affini, basta prendere $v_0 = O$.

Un sottospazio affine di \mathbb{R}^n può essere descritto con equazioni parametriche o cartesiane, a seconda di come viene descritto il sottospazio di giacitura W .

Definizione 6.5 Diremo che un sottospazio affine $L \subseteq \mathbb{R}^n$ è descritto da *equazioni parametriche* se

$$L = \{v = At + v_0 \in \mathbb{R}^n \mid t \in \mathbb{R}^k\} = v_0 + \text{Im } A,$$

per qualche $v_0 \in \mathbb{R}^n$ e $A \in M_{n,k}(\mathbb{R})$; il sottospazio affine L è l'immagine dell'applicazione $F: \mathbb{R}^k \rightarrow \mathbb{R}^n$ data da $F(t) = At + v_0$. In questo caso, il sottospazio di giacitura ha equazioni parametriche $w = At$.

Definizione 6.6 Diremo che un sottospazio affine $L \subseteq \mathbb{R}^n$ è descritto da *equazioni cartesiane* se

$$L = \{x \in \mathbb{R}^n \mid Bx = b\}$$

per qualche $B \in M_{p,n}(\mathbb{R})$ e $b \in \mathbb{R}^p$; il sottospazio affine L è l'insieme dei vettori di \mathbb{R}^n dove l'applicazione $G: \mathbb{R}^n \rightarrow \mathbb{R}^p$ data da $G(x) = Bx - b$ si annulla. In questo caso, il sottospazio di giacitura ha equazioni cartesiane $Bx = O$.

Di nuovo si pone il problema di passare dalle equazioni parametriche a quelle cartesiane e viceversa. Se $Bx = b$ sono equazioni cartesiane, risolvendo il sistema troviamo equazioni parametriche⁴. Viceversa:

Proposizione 6.4 Sia $L \subseteq \mathbb{R}^n$ un sottospazio affine di \mathbb{R}^n di equazioni parametriche $v = At + v_0$, e siano $Bx = O$ equazioni cartesiane del sottospazio di giacitura. Allora equazioni cartesiane di L sono date da

$$Bx = Bv_0. \quad (6.7)$$

Dimostrazione. Infatti, un vettore $x \in \mathbb{R}^n$ appartiene a L se e solo se $x - v_0 \in W$, che accade se e solo se $B(x - v_0) = O$. \square

ESEMPIO 6.12 Consideriamo il sistema lineare

$$\begin{cases} 2x_1 - x_2 + x_3 = 1, \\ \quad \quad x_2 + x_3 = 3, \\ x_1 \quad \quad + x_3 = 2, \end{cases}$$

di matrice dei coefficienti B e vettore dei termini noti b dati da

$$B = \begin{vmatrix} 2 & -1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{vmatrix}, \quad b = \begin{vmatrix} 1 \\ 3 \\ 2 \end{vmatrix}.$$

⁴ In altre parole, "risolvere" un sistema lineare vuol dire passare da equazioni cartesiane a equazioni parametriche.

L'insieme L delle soluzioni di questo sistema è un sottospazio affine di \mathbb{R}^3 ; vogliamo le sue equazioni parametriche. Effettuando una riduzione a scala troviamo che la terza equazione è inutile; risolvendo all'indietro il sistema a scala formato dalle prime due otteniamo le equazioni parametriche

$$\begin{cases} x_1 = 2 - t, \\ x_2 = 3 - t, \\ x_3 = t, \end{cases} \quad \text{cioè} \quad x = \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix},$$

Quindi L è una retta passante per $\begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}$ e parallela a $\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$.

ESEMPIO 6.13 Consideriamo il sottospazio affine $L \subseteq \mathbb{R}^4$ di equazioni parametriche

$$\begin{cases} x_1 = 3 + s - t, \\ x_2 = -1 + 2s, \\ x_3 = 1 + 2t, \\ x_4 = s + t, \end{cases}$$

cioè $v = At + v_0$, con

$$A = \begin{pmatrix} 1 & -1 \\ 2 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix} \quad \text{e} \quad v_0 = \begin{pmatrix} 3 \\ -1 \\ 1 \\ 0 \end{pmatrix}.$$

Effettuando una riduzione a scala del sistema $At = v$, troviamo

$$\left(\begin{array}{cc|c} 1 & -1 & x_1 \\ 2 & 0 & x_2 \\ 0 & 2 & x_3 \\ 1 & 1 & x_4 \end{array} \right) \longrightarrow \left(\begin{array}{cc|c} 1 & -1 & x_1 \\ 0 & 2 & x_2 - 2x_1 \\ 0 & 2 & x_3 \\ 0 & 2 & x_4 - x_1 \end{array} \right) \longrightarrow \left(\begin{array}{cc|c} 1 & -1 & x_1 \\ 0 & 2 & x_2 - 2x_1 \\ 0 & 0 & x_3 - x_2 + 2x_1 \\ 0 & 0 & x_4 - x_2 + x_1 \end{array} \right);$$

quindi la Proposizione 6.4 ci dice che le equazioni cartesiane di L sono

$$\begin{cases} 2x_1 - x_2 + x_3 = 8, \\ x_1 - x_2 + x_4 = 4. \end{cases}$$

Osservazione 6.10 Quante equazioni cartesiane servono per descrivere una retta nello spazio? Le equazioni cartesiane sono della forma $Bx = b$, con $x \in \mathbb{R}^3$ e $B \in M_{p,3}(\mathbb{R})$. Dovendo descrivere una retta, occorre che $1 = \dim \text{Ker } B = 3 - \text{rg } B$, cioè occorre che $\text{rg } B = 2$. Quindi (dopo aver eliminato le equazioni inutili) $p = 2$, cioè per descrivere una retta nello spazio servono 2 equazioni cartesiane. Analogamente, per descrivere una retta in \mathbb{R}^n servono $n - 1$ equazioni cartesiane.

Concludiamo con un'ultima definizione.

Definizione 6.7 Siano $L_1, L_2 \subseteq V$ due sottospazi affini, di sottospazi di giacitura rispettivamente W_1 e W_2 . Diremo che L_1 ed L_2 sono *paralleli* se $W_1 \subseteq W_2$ o $W_2 \subseteq W_1$.

Esercizi

6.1 Risolvi il sistema a scala

$$\begin{cases} x_1 - x_2 + x_3 - x_4 + x_5 - x_6 = 1, \\ x_3 - 2x_4 + x_5 + x_6 = 0, \\ x_5 - x_6 = 2. \end{cases}$$

6.2 Studia (cioè vedi se ammettono soluzioni, e in tal caso trovale) i sistemi

$$\begin{cases} 4x + y + z + 2v + 3w = 0, \\ 14x + 2y + 2z + 7v + 11w = 0, \\ 15x + 3y + 3z + 6v + 10w = 0; \end{cases} \quad \begin{cases} 5x + 4y + 7z = 3, \\ x + 2y + 3z = 1, \\ x - y - z = 0, \\ 3x + 3y + 5z = 2. \end{cases}$$

6.3 Studia al variare del parametro $k \in \mathbb{R}$ i seguenti sistemi

$$\begin{cases} x + (k-1)y + z = 1, \\ (2k-3)x + y + (k-1)z = 3 - k, \\ 2x + ky + kz = k, \\ kx + 2y + (2k-2)z = 4 - k; \end{cases} \quad \begin{cases} 3kx + 3y + (k+2)z = k + 2, \\ x + 2y + 2z = k + 1, \\ x + ky + kz = 1. \end{cases}$$

6.4 Studia al variare dei parametri $\alpha, \beta \in \mathbb{R}$ il sistema

$$\begin{cases} \alpha x + y + z = 1, \\ x + \alpha y + z = \beta, \\ x + y + \alpha z = \beta^2. \end{cases}$$

6.5 Se A è una matrice quadrata di ordine n , dimostra che il rango di A è uguale al numero di pivot (nel senso del Paragrafo 3.3) *non nulli* ottenuti con una qualunque eliminazione di Gauss di A .

6.6 Sia $A \in M_{3,3}(\mathbb{R})$ la matrice

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 2 & 3 & 7 \\ 1 & -3 & -1 \end{pmatrix}.$$

Trova dimensione e base di $\text{Ker } A$, $\text{Im } A$, $\text{Ker } A^T$ e $\text{Im } A^T$, e studia il sistema $Ax = b_k$ al variare del parametro $k \in \mathbb{R}$, dove

$$b_k = \begin{pmatrix} k \\ 5 \\ -2 \end{pmatrix} \in \mathbb{R}^3.$$

6.7 Trova dimensione e base del nucleo e dell'immagine delle seguenti matrici e delle loro trasposte:

$$\begin{pmatrix} 1 & 3 & 1 & 1 \\ 2 & -1 & -2 & 2 \\ 5 & 8 & 1 & 4 \end{pmatrix}, \quad \begin{pmatrix} 2 & 1 & -1 & -1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \end{pmatrix}.$$

6.8 Sia $A_\alpha \in M_{3,3}(\mathbb{R})$ la matrice

$$A_\alpha = \begin{pmatrix} \alpha & \alpha & \alpha+1 \\ -1 & 0 & 2 \\ 2 & 0 & \alpha \end{pmatrix}.$$

Determina, al variare di $\alpha \in \mathbb{R}$, le dimensioni di $\text{Ker } A_\alpha$ e $\text{Im } A_\alpha$, e trova per quali valori di α si ha $\mathbb{R}^3 = \text{Ker } A_\alpha \oplus \text{Im } A_\alpha$.

6.9 Sia $T: \mathbb{R}^3 \rightarrow \mathbb{R}^4$ data da

$$T \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} y+z \\ 2x+y-z \\ z \\ 2x+2y \end{pmatrix}.$$

Determina immagine e nucleo di T , e se U, W sono i sottospazi di \mathbb{R}^3 dati da

$$U = \text{Span} \left(\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 2 \end{pmatrix} \right) \quad \text{e} \quad W = \text{Span} \left(\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right),$$

trova dimensione e base per $T(U) \cap T(W)$.

6.10 Considera i sottospazi di \mathbb{R}^4 dati da

$$U = \{x \in \mathbb{R}^4 \mid 2x_1 + x_2 - x_4 = 0, 3x_3 - \frac{5}{2}x_4 = 0\} \quad \text{e} \quad W = \{x \in \mathbb{R}^4 \mid 2x_2 - x_3 = 0\}.$$

Trova dimensione, base, equazioni cartesiane ed equazioni parametriche per $U+W$ e $U \cap W$.

6.11 Consideriamo i seguenti due sottospazi di \mathbb{R}^3 :

$$U = \text{Span} \left(\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 2 \end{pmatrix} \right) \quad \text{e} \quad W = \text{Span} \left(\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} \right);$$

vogliamo trovare dimensione e base di $U \cap W$. Procedendo come al solito dobbiamo effettuare la seguente riduzione a scala:

$$\left| \begin{array}{cccccc} 1 & 2 & 3 & 0 & 2 & \\ 0 & 2 & 2 & 1 & -1 & \\ 1 & 1 & 2 & 1 & 0 & \end{array} \right| \longrightarrow \left| \begin{array}{cccccc} 1 & 2 & 3 & 0 & 2 & \\ 0 & 2 & 2 & 1 & -1 & \\ 0 & -1 & -1 & 1 & -2 & \end{array} \right| \longrightarrow \left| \begin{array}{cccccc} 1 & 2 & 3 & 0 & 2 & \\ 0 & 2 & 2 & 1 & -1 & \\ 0 & 0 & 0 & 3/2 & -5/2 & \end{array} \right| = S.$$

Dunque la matrice ha rango 3, per cui $\dim U \cap W = 5 - 3 = 2$. Per trovarne una base dobbiamo prima trovare una base di $\text{Ker } S$. Risolvendo all'indietro si vede che

$$\text{Ker } S = \text{Span} \left(z_1 = \begin{pmatrix} -1 \\ -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, z_2 = \begin{pmatrix} -4/3 \\ -1/3 \\ 0 \\ 5/3 \\ 1 \end{pmatrix} \right).$$

per cui $U \cap W = \text{Span}(\psi^{-1}(z_1), \psi^{-1}(z_2))$. Ma sfortunatamente si ha $\psi^{-1}(z_1) = 0$; dunque

$$U \cap W = \text{Span} \left(\psi^{-1}(z_2) = \begin{pmatrix} 2 \\ 2/3 \\ 5/3 \end{pmatrix} \right)$$

e stavolta si ottiene $\dim U \cap W = 1$. Dov'è l'errore?

6.12 È possibile trovare in \mathbb{R}^4 tre sottospazi V_1, V_2, V_3 di dimensione 2 tali che

$$V_1 \cap V_2 = V_1 \cap V_3 = V_2 \cap V_3 = \{0\}?$$

6.13 Considera i sottospazi

$$V_t = \text{Span} \left(\begin{pmatrix} 0 \\ 1 \\ t \end{pmatrix}, \begin{pmatrix} 1 \\ t \\ 1 \end{pmatrix}, \begin{pmatrix} t \\ 0 \\ 0 \end{pmatrix} \right) \subseteq \mathbb{R}^3, \quad W_t = \text{Span} \left(\begin{pmatrix} 0 \\ t \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ t \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ t \end{pmatrix} \right) \subseteq \mathbb{R}^4$$

al variare di $t \in \mathbb{R}$. Determina dimensione e base di $\bigcap_{t \in \mathbb{R}} V_t$ e di $\bigcap_{t \in \mathbb{R}} W_t$.

6.14 Quante equazioni cartesiane servono per descrivere un piano in \mathbb{R}^n ?

6.15 Determina equazioni cartesiane e parametriche per il sottospazio di \mathbb{R}^4 generato dai vettori

$$\begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 2 \\ 3 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 3 \\ 1 \\ 1 \\ -2 \end{pmatrix}.$$

6.16 Trova equazioni cartesiane della retta $s \subset \mathbb{R}^5$ di equazioni parametriche

$$\begin{cases} x_1 = t - 7, \\ x_2 = -t + 2, \\ x_3 = -t - 6, \\ x_4 = t - 5, \\ x_5 = t + 3. \end{cases}$$

6.17 Trova equazioni cartesiane del sottospazio vettoriale di \mathbb{R}^4 di equazioni parametriche

$$\begin{cases} x_1 = t_1 + 3t_2, \\ x_2 = t_2 + t_3, \\ x_3 = 2t_1 + t_3, \\ x_4 = t_1 + t_2 + t_3. \end{cases}$$

6.18 Trova equazioni parametriche e cartesiane del sottospazio di \mathbb{R}^4 dato da

$$V_t = \text{Span} \left(\begin{pmatrix} 1 \\ 1 \\ t \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ t \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ t \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2t \\ 1 \\ 0 \end{pmatrix} \right)$$

al variare di $t \in \mathbb{R}$.

6.19 Trova equazioni parametriche e cartesiane per il sottospazio W di $M_{2,3}(\mathbb{R})$ generato dalle matrici

$$\begin{vmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{vmatrix}, \quad \begin{vmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \end{vmatrix}, \quad \begin{vmatrix} 2 & 0 & 1 \\ 0 & 1 & 0 \end{vmatrix}.$$

(Suggerimento: comportati come se $M_{2,3}(\mathbb{R})$ fosse \mathbb{R}^6 .)

6.20 Siano W_1 e W_2 sottospazi di uno spazio vettoriale V , e $v_1, v_2 \in V$. Dimostra che $v_1 + W_1 = v_2 + W_2$ se e solo se $W_1 = W_2$ e $v_2 - v_1 \in W_1 = W_2$. In particolare, $v_1 + W_1$ è un sottospazio vettoriale se e solo se $v_1 \in W_1$.

6.21 Definisci i concetti di equazioni parametriche e cartesiane lineari per un sottospazio affine di uno spazio vettoriale V qualunque. (Suggerimento: utilizza applicazioni lineari da V a \mathbb{R}^n o viceversa.)

6.22 Siano $L_1 = v_1 + W_1$ ed $L_2 = v_2 + W_2$ due sottospazi affini di uno spazio vettoriale V , e poniamo $W_0 = W_1 + W_2 + \text{Span}(v_2 - v_1)$. Dimostra che $L_0 = v_1 + W_0$ è il più piccolo sottospazio affine di V contenente sia L_1 che L_2 . Il sottospazio affine L_0 viene chiamato *somma* di L_1 ed L_2 , e indicato con $L_1 + L_2$.

6.23 Siano L_1 ed L_2 due sottospazi affini di uno spazio vettoriale V . Dimostra il Teorema di Grassmann per sottospazi affini:

$$\dim(L_1 + L_2) + \dim(L_1 \cap L_2) \leq \dim L_1 + \dim L_2,$$

dove per convenzione si pone $\dim(L_1 \cap L_2) = -1$ se $L_1 \cap L_2 = \emptyset$. Dimostra che l'uguaglianza vale se e solo se $L_1 \cap L_2 \neq \emptyset$ oppure $L_1 \cap L_2 = \emptyset$ e $W_1 \cap W_2 = \{O\}$.

6.24 Dimostra che due sistemi compatibili sono equivalenti se e solo se le equazioni di ciascuno dei due si possono ottenere da quelle dell'altro tramite operazioni elementari.

Matrici e applicazioni lineari

Nel Capitolo 5 abbiamo introdotto le applicazioni lineari; in questo capitolo ne riprendiamo lo studio concentrandoci sulle operazioni che possiamo effettuare con esse. Vedremo come sommarle e moltiplicarle per uno scalare, in modo che l'insieme delle applicazioni lineari fra due spazi vettoriali dati risulti essere a sua volta uno spazio vettoriale. Studieremo la composizione di applicazioni lineari, e questo ci condurrà alla fondamentale nozione di isomorfismo fra spazi vettoriali. Vedremo anche che tutte e sole le applicazioni lineari fra \mathbb{R}^n ed \mathbb{R}^m sono quelle della forma L_A dove A è una matrice $m \times n$, e quindi definiremo il prodotto di matrici tramite la composizione delle corrispondenti applicazioni lineari. Infine ci soffermeremo sul caso delle matrici quadrate, dando diversi criteri per stabilire quando una matrice è invertibile e descrivendo un algoritmo per il calcolo dell'inversa.

7.1 Composizione e isomorfismi

Il nostro primo obiettivo è definire una struttura di spazio vettoriale sull'insieme delle applicazioni lineari fra due spazi vettoriali dati. Per farlo è sufficiente imitare quanto visto nell'Esempio 4.2.

Definizione 7.1 Siano V e W due spazi vettoriali, $S, T: V \rightarrow W$ due applicazioni lineari e $\lambda \in \mathbb{R}$. Definiamo le applicazioni $S + T: V \rightarrow W$ e $\lambda T: V \rightarrow W$ ponendo

$$(S + T)(v) = S(v) + T(v) \quad \text{e} \quad (\lambda T)(v) = \lambda(T(v))$$

per ogni $v \in V$. Le applicazioni $S + T$ e λT sono ancora lineari, come si verifica subito: per esempio

$$(S + T)(\lambda v) = S(\lambda v) + T(\lambda v) = \lambda S(v) + \lambda T(v) = \lambda(S(v) + T(v)) = \lambda(S + T)(v),$$

e così via (controllare, prego). Sia $\mathcal{L}(V, W)$ l'insieme delle applicazioni lineari da V in W ; con queste operazioni $\mathcal{L}(V, W)$ diviene uno spazio vettoriale (esercizio). In particolare, lo spazio $\mathcal{L}(V, \mathbb{R})$ delle applicazioni lineari da V in \mathbb{R} si chiama spazio

duale di V e si indica con V' (o V^*). Maggiori informazioni sul duale di uno spazio vettoriale le troverai nei Complementi al Capitolo 8.

Ora, vi è un'altra operazione naturale fra applicazioni lineari: la composizione.

Definizione 7.2 Siano $S:U \rightarrow V$ e $T:V \rightarrow W$ due applicazioni lineari; allora la *composizione* di S e T è l'applicazione $T \circ S:U \rightarrow W$ data da

$$\forall u \in U \quad (T \circ S)(u) = T(S(u)).$$

Proposizione 7.1 Siano $S:U \rightarrow V$ e $T:V \rightarrow W$ applicazioni lineari fra gli spazi vettoriali U, V e W . Allora la composizione $T \circ S:U \rightarrow W$ è lineare.

Dimostrazione. Si tratta di una semplice verifica. Presi $u_1, u_2 \in U$ si ha

$$\begin{aligned} (T \circ S)(u_1 + u_2) &= T(S(u_1 + u_2)) = T(S(u_1) + S(u_2)) \\ &= T(S(u_1)) + T(S(u_2)) = (T \circ S)(u_1) + (T \circ S)(u_2). \end{aligned}$$

In modo analogo si dimostra che T è omogenea (esercizio). \square

Osservazione 7.1 Nel caso in cui $U = V = W$, cioè quando sia T che S sono endomorfismi, risultano definiti sia $T \circ S$ che $S \circ T$. In generale, però, la composizione *non* è commutativa, cioè $T \circ S \neq S \circ T$.

ESEMPIO 7.1 Definiamo $S, T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ ponendo

$$S \begin{vmatrix} x \\ y \\ z \end{vmatrix} = \begin{vmatrix} x \\ y \\ 0 \end{vmatrix} \quad \text{e} \quad T \begin{vmatrix} x \\ y \\ z \end{vmatrix} = \begin{vmatrix} x \\ z \\ 0 \end{vmatrix}.$$

Allora

$$(S \circ T) \begin{vmatrix} x \\ y \\ z \end{vmatrix} = \begin{vmatrix} x \\ z \\ 0 \end{vmatrix} \neq \begin{vmatrix} x \\ 0 \\ 0 \end{vmatrix} = (T \circ S) \begin{vmatrix} x \\ y \\ z \end{vmatrix}.$$

Ovviamente, esistono endomorfismi che commutano; per esempio, se $S:V \rightarrow V$ è un endomorfismo qualunque, allora $S \circ \text{id}_V = \text{id}_V \circ S = S$ e $S \circ O = O \circ S = O$.

La composizione, come nel caso di funzioni qualunque, è sempre associativa:

$$R \circ (S \circ T) = (R \circ S) \circ T$$

non appena R, S e T sono applicazioni lineari tali che $S \circ T, R \circ S, R \circ (S \circ T)$ e $(R \circ S) \circ T$ siano definite. È facile verificare (esercizio) che la composizione è distributiva rispetto alla somma e al prodotto per scalari, cioè che

$$\begin{aligned} (S_1 + S_2) \circ T &= S_1 \circ T + S_2 \circ T, & S \circ (T_1 + T_2) &= S \circ T_1 + S \circ T_2, \\ (\lambda S) \circ T &= \lambda(S \circ T) = S \circ (\lambda T), \end{aligned}$$

non appena tutte le applicazioni lineari coinvolte sono definite. In particolare, quindi, $\mathcal{L}(V, V)$ con la somma e la composizione è un anello non commutativo (in cui l'elemento neutro rispetto alla somma è l'applicazione nulla O e quello rispetto alla composizione è l'identità id_V).

Definizione 7.3 Diremo che un'applicazione lineare $T:V \rightarrow W$ è *invertibile* se esiste un'applicazione lineare $S:W \rightarrow V$, l'*inversa* di T , tale che $T \circ S = \text{id}_W$ e $S \circ T = \text{id}_V$. L'inversa di T , se esiste, si indica con T^{-1} .

Osservazione 7.2 L'inversa, se esiste, è unica: se S ed S' sono due inverse di un'applicazione T , abbiamo

$$S' = S' \circ \text{id}_W = S' \circ (T \circ S) = (S' \circ T) \circ S = \text{id}_V \circ S = S.$$

Nella definizione di inversa abbiamo richiesto, per chiarezza, che l'inversa sia lineare: stiamo lavorando in spazi vettoriali con applicazioni lineari, per cui un'inversa non lineare (che non rispetta le operazioni) sarebbe del tutto inutile. In realtà, risulta che non appena $T:V \rightarrow W$ è invertibile come funzione (cioè non appena è bigettiva), allora l'inversa è automaticamente lineare:

Proposizione 7.2 Sia $T:V \rightarrow W$ un'applicazione lineare. Allora T è invertibile se e solo se è iniettiva e surgettiva.

Dimostrazione. Se T è invertibile è chiaramente bigettiva. Viceversa, supponiamo T iniettiva e surgettiva; allora esiste la funzione inversa $S:W \rightarrow V$. Dobbiamo solo dimostrare che S è lineare.

Prendiamo w_1 e $w_2 \in W$; essendo T surgettiva, esistono $v_1, v_2 \in V$ tali che si abbia $T(v_1) = w_1$ e $T(v_2) = w_2$. Allora $T(v_1 + v_2) = w_1 + w_2$; applicando S otteniamo

$$S(w_1 + w_2) = v_1 + v_2 = S(w_1) + S(w_2).$$

Analogamente si dimostra (esercizio) che $S(\lambda w_1) = \lambda S(w_1)$. \square

Corollario 7.3 Sia $T:V \rightarrow W$ un'applicazione lineare. Supponiamo che si abbia $\dim V = \dim W$. Allora T è invertibile se e solo se è iniettiva se e solo se è surgettiva.

Dimostrazione. Da un lato, T invertibile implica T iniettiva e surgettiva; poi, per il Corollario 5.8.(iii), T iniettiva (rispettivamente, surgettiva) implica T surgettiva (rispettivamente, iniettiva), e le due cose insieme implicano T invertibile. \square

In particolare, un endomorfismo $T:V \rightarrow V$ è invertibile se e solo se è iniettivo (o surgettivo); quindi per vedere se è invertibile, basta controllare se $\text{Ker } T = \{O\}$.

L'esistenza di un'applicazione $T:V \rightarrow W$ lineare invertibile fra due spazi vettoriali V e W implica $\dim V = \dim W$ (perché? Confronta con l'Esercizio 7.8). Di più: ci dice che V e W sono *essenzialmente lo stesso spazio*; qualunque cosa accada

in V possiamo, tramite T , trasferirla in W e leggerla lì; e viceversa, tramite T^{-1} . Per questo motivo introduciamo la seguente definizione:

Definizione 7.4 Due spazi vettoriali V e W sono *isomorfi* (e si scrive $V \cong W$) se esiste un'isomorfismo fra V e W , cioè un'applicazione lineare invertibile $T: V \rightarrow W$.

ESEMPIO 7.2 Lo spazio $M_{m,n}(\mathbb{R})$ è isomorfo a $M_{n,m}(\mathbb{R})$. L'isomorfismo è la trasposizione $T: M_{m,n}(\mathbb{R}) \rightarrow M_{n,m}(\mathbb{R})$; infatti il fatto che $(A^T)^T = A$ per qualunque matrice A ci dice che l'inversa della trasposizione è ancora la trasposizione (perché?).

ESEMPIO 7.3 Lo spazio $M_{m,n}(\mathbb{R})$ è anche isomorfo a \mathbb{R}^{mn} , tramite l'applicazione $T: M_{m,n}(\mathbb{R}) \rightarrow \mathbb{R}^{mn}$ che prende le colonne della matrice e le mette una sotto l'altra:

$$T \left(\begin{bmatrix} A^1 & \dots & A^n \end{bmatrix} \right) = \begin{bmatrix} A^1 \\ \vdots \\ A^n \end{bmatrix}.$$

L'applicazione T è ovviamente lineare e bigettiva, per cui è un isomorfismo.

ESEMPIO 7.4 Sia V uno spazio vettoriale di dimensione n ; prendiamo una sua base $\mathcal{B} = \{v_1, \dots, v_n\}$, e sia $F_{\mathcal{B}}: V \rightarrow \mathbb{R}^n$ l'applicazione lineare definita nell'Esempio 5.7. Allora $F_{\mathcal{B}}$ è un isomorfismo; l'inversa è l'applicazione $S: \mathbb{R}^n \rightarrow V$ data da

$$\forall a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{R}^n \quad S(a) = a_1 v_1 + \dots + a_n v_n.$$

Osservazione 7.3 Questo esempio ci dice che ogni spazio vettoriale su \mathbb{R} di dimensione n è isomorfo a \mathbb{R}^n ; quindi perché non limitarsi a \mathbb{R}^n invece di studiare gli spazi vettoriali in generale? Per almeno due motivi. Prima di tutto, l'isomorfismo sopra descritto non è, come si dice, *naturale* o *canonico*; non è univocamente definito una volta dato lo spazio V ; dipende dalla *scelta* di una base. Una volta fissata una base \mathcal{B} hai l'isomorfismo e puoi trasformare tutti i problemi su V in problemi su \mathbb{R}^n ; ma chi dice che sia la base giusta? Che con un'altra base non sia tutto più semplice? E poi, di solito, uno desidera delle risposte che non dipendono dalla base scelta; l'isomorfismo $F_{\mathcal{B}}$ può essere (e lo è, come vedremo nel prossimo paragrafo) molto utile come strumento di calcolo, ma poi alla fine i risultati li vogliamo in V , non in \mathbb{R}^n , indipendenti dalla base scelta. L'isomorfismo dell'Esempio 7.2 è invece di tutt'altro genere: non dipende da alcuna scelta, e ci permette a tutti (o quasi...) gli effetti di identificare $M_{m,n}(\mathbb{R})$ e $M_{n,m}(\mathbb{R})$ come spazi vettoriali.

Ma non basta: due spazi vettoriali isomorfi possono essere considerati lo stesso spazio *solo quando ci interessano unicamente la struttura e le operazioni di spazio vettoriale*. Se sono in gioco altre strutture, i due spazi devono rimanere accuratamente distinti. Per esempio, $\mathbb{R}_3[t]$ è isomorfo a \mathbb{R}^4 . Ma coi polinomi possiamo fare tante cose che su \mathbb{R}^4 non hanno molto senso: possiamo dire "quanto vale il polinomio $p(t)$ nel punto $t = 27$ " mentre "quanto vale il vettore $v \in \mathbb{R}^4$ nel punto $t = 27$ "

non significa proprio nulla. Un esempio anche più importante è $\mathcal{L}(V, V)$: mentre la composizione di applicazioni lineari è un'operazione assolutamente naturale e quasi inevitabile (e senza la quale non saremmo qui a porci questo problema), la "composizione" di vettori di \mathbb{R}^n è un concetto un po' forzato.

La morale di tutto ciò è che gli isomorfismi $F_{\mathcal{B}}$ sono degli ottimi e importanti strumenti di calcolo, ma non ci devono far dimenticare che una cosa sono gli spazi vettoriali in generale, e un'altra gli spazi \mathbb{R}^n (o che una cosa sono le applicazioni lineari e un'altra le matrici).

Dopo tutto questo bel discorso vediamo invece due isomorfismi naturali, cioè che non dipendono da alcuna scelta arbitraria:

Proposizione 7.4 (i) Per ogni $n, m \in \mathbb{N}$ lo spazio $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ è isomorfo allo spazio delle matrici $M_{m,n}(\mathbb{R})$. In particolare, lo spazio duale $(\mathbb{R}^n)'$ è isomorfo allo spazio dei vettori riga $M_{1,n}(\mathbb{R})$.

(ii) Sia V uno spazio vettoriale. Allora lo spazio $\mathcal{L}(\mathbb{R}, V)$ è isomorfo a V .

Dimostrazione. (i) Sia $L: M_{m,n}(\mathbb{R}) \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ l'applicazione che associa alla matrice $A \in M_{m,n}(\mathbb{R})$ l'applicazione lineare $L_A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$; l'idea è che L è l'isomorfismo cercato. Prima di tutto, dobbiamo dimostrare che L è lineare, cioè che $L_{A+B} \equiv L_A + L_B$ e che $L_{\lambda A} \equiv \lambda L_A$. Ma infatti

$$\begin{aligned} L_{A+B}(x) &= (A+B)x = x_1(A^1 + B^1) + \dots + x_n(A^n + B^n) \\ &= (x_1 A^1 + \dots + x_n A^n) + (x_1 B^1 + \dots + x_n B^n) \\ &= Ax + Bx = L_A(x) + L_B(x) = (L_A + L_B)(x) \end{aligned}$$

per tutti gli $x \in \mathbb{R}^n$, per cui $L_{A+B} \equiv L_A + L_B$. Analogamente,

$$L_{\lambda A}(x) = x_1(\lambda A^1) + \dots + x_n(\lambda A^n) = \lambda(x_1 A^1 + \dots + x_n A^n) = \lambda(L_A(x)) = (\lambda L_A)(x)$$

qualunque sia $\lambda \in \mathbb{R}$, e ci siamo.

Per dimostrare che L è un isomorfismo, vogliamo costruirne l'inversa. Noi sappiamo che un'applicazione lineare è completamente determinata dai valori che assume su una base (Proposizione 5.2); quindi ogni $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ è completamente descritta dai vettori $T(e_1), \dots, T(e_n)$. Consideriamo allora l'applicazione

$$J: \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \rightarrow M_{m,n}(\mathbb{R})$$

che associa all'applicazione lineare $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ la matrice che ha come prima colonna $T(e_1)$, come seconda colonna $T(e_2)$ e così via:

$$J(T) = \begin{bmatrix} T(e_1) & \dots & T(e_n) \end{bmatrix}.$$

Vogliamo dimostrare che J è l'inversa di L . Grazie alla Proposizione 7.2, ci basta far vedere che $L \circ J$ e $J \circ L$ sono l'identità degli spazi appropriati, senza perder tempo a dimostrare che J è lineare (anche se è facile verificarlo).

Per far vedere che $J \circ L = \text{id}_{M_{m,n}(\mathbb{R})}$ basta dimostrare che la j -esima colonna della matrice $J(L_A)$ è esattamente A^j , qualunque siano $A \in M_{m,n}(\mathbb{R})$ e $j = 1, \dots, n$. E infatti la j -esima colonna di $J(L_A)$ è proprio $L_A(e_j) = A^j$, grazie a (5.4).

Viceversa, prendiamo $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$; grazie al Corollario 5.3, per dimostrare che $(L \circ J)(T) \equiv T$ ci basta far vedere che $L_{J(T)}(e_j) = T(e_j)$ per $j = 1, \dots, n$. E infatti, di nuovo (5.4) ci dice che $L_{J(T)}(e_j)$ è la j -esima colonna di $J(T)$, ovvero, per definizione, $T(e_j)$.

(ii) L'idea della dimostrazione è la seguente: un'applicazione lineare $T \in \mathcal{L}(\mathbb{R}, V)$ è completamente determinata dal valore su una base; $\{1\}$ è una base di \mathbb{R} , e quindi $T(1)$ determina completamente T . Dunque viene naturale tentare di dimostrare che l'applicazione $\Phi: \mathcal{L}(\mathbb{R}, V) \rightarrow V$ definita da $\Phi(T) = T(1)$ è l'isomorfismo cercato.

L'applicazione Φ è chiaramente lineare (esercizio); dimostriamo che è iniettiva e surgettiva. È iniettiva: se $\Phi(T) = O$ abbiamo

$$\forall \lambda \in \mathbb{R} \quad T(\lambda) = \lambda T(1) = \lambda O = O,$$

per cui $T \equiv O$; dunque $\text{Ker } \Phi = \{O\}$, e l'iniettività di Φ segue dalla Proposizione 5.5.(iv).

Infine, è surgettiva: preso $v \in V$, definiamo $T_v \in \mathcal{L}(\mathbb{R}, V)$ ponendo $T_v(\lambda) = \lambda v$. Allora

$$\Phi(T_v) = T_v(1) = v,$$

per cui $v \in \text{Im } \Phi$, ed è fatta. \square

In particolare, tutte le applicazioni lineari da \mathbb{R}^n a \mathbb{R}^m sono del tipo L_A per un'opportuna matrice $A \in M_{m,n}(\mathbb{R})$, in quanto l'applicazione $L: M_{m,n}(\mathbb{R}) \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ definita nella dimostrazione della Proposizione 7.4 è bigettiva.

ESEMPIO 7.5 La matrice associata all'identità di \mathbb{R}^n deve avere come j -esima colonna $\text{id}_{\mathbb{R}^n}(e_j) = e_j$; quindi otteniamo la matrice identica

$$I_n = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & 1 \end{pmatrix},$$

dove gli spazi bianchi vanno riempiti di zeri; a volte scriveremo I al posto di I_n . Analogamente, l'applicazione nulla viene associata alla matrice nulla

$$O = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}.$$

Osservazione 7.4 Gli elementi della matrice identica sono spesso indicati con δ_{ij} , dove

$$\delta_{ij} = \begin{cases} 0 & \text{se } i \neq j, \\ 1 & \text{se } i = j. \end{cases}$$

Il simbolo δ_{ij} è chiamato *delta di Kronecker*.

7.2 Prodotto di matrici

Abbiamo quindi visto che le applicazioni lineari fra \mathbb{R}^n ed \mathbb{R}^m sono essenzialmente la stessa cosa delle matrici $m \times n$; quindi deve esistere per le matrici un analogo della composizione fra applicazioni lineari. Prendiamo allora due matrici $A \in M_{m,n}(\mathbb{R})$ e $B \in M_{n,p}(\mathbb{R})$, e le relative applicazioni lineari $L_A: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ed $L_B: \mathbb{R}^p \rightarrow \mathbb{R}^n$. I domini e i codomini sono messi in modo tale da permetterci di considerare la composizione $L_A \circ L_B: \mathbb{R}^p \rightarrow \mathbb{R}^m$, che è un'applicazione lineare appartenente a $\mathcal{L}(\mathbb{R}^p, \mathbb{R}^m)$. Per quanto visto, deve quindi esistere una (e una sola) matrice $C \in M_{m,p}(\mathbb{R})$ tale che $L_C = L_A \circ L_B$.

Definizione 7.5 Date due matrici $A \in M_{m,n}(\mathbb{R})$ e $B \in M_{n,p}(\mathbb{R})$, diremo che la matrice $C \in M_{m,p}(\mathbb{R})$ tale che $L_A \circ L_B = L_C$ è il *prodotto (righe per colonne)* di A e B , e scriveremo $C = AB$ (o, talvolta, $C = A \cdot B$). In particolare si ha

$$L_A \circ L_B = L_{AB}.$$

Prima di controllare se questo prodotto ha le proprietà necessarie per essere considerato tale, vediamo come si calcola esplicitamente. La colonna j -esima C^j della matrice prodotto $C = AB$ dev'essere l'immagine tramite $L_A \circ L_B$ del vettore e_j della base canonica; quindi

$$\begin{aligned} \begin{pmatrix} c_{1j} \\ \vdots \\ c_{mj} \end{pmatrix} &= C^j = L_A \circ L_B(e_j) = L_A(B^j) \\ &= \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{1j} \\ \vdots \\ b_{nj} \end{pmatrix} = \begin{pmatrix} a_{11}b_{1j} + \cdots + a_{1n}b_{nj} \\ \vdots \\ a_{m1}b_{1j} + \cdots + a_{mn}b_{nj} \end{pmatrix}, \end{aligned}$$

per $j = 1, \dots, p$. In altri termini, l'elemento di posto (i, j) della matrice AB è

$$(AB)_{ij} = c_{ij} = a_{i1}b_{1j} + \cdots + a_{in}b_{nj} = A_i B^j, \quad (7.1)$$

giustificando il nome di prodotto righe per colonne.

Osservazione 7.5 Perché AB sia definita occorre che il numero di *colonne* di A sia uguale al numero di *righe* di B , in modo che la composizione $L_A \circ L_B$ e la formula (7.1) abbiano senso.

ESEMPIO 7.6 Più che un esempio è un esercizio: verifica che i seguenti prodotti di matrici siano corretti.

$$\begin{aligned} \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 3 & 6 & 1 \\ -1 & -1 & 0 \end{pmatrix} &= \begin{pmatrix} 0 & 3 & 1 \\ 2 & 8 & 2 \end{pmatrix}; & \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}; \\ \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} &= \begin{pmatrix} 1 & 4 \\ 3 & 10 \end{pmatrix} \neq \begin{pmatrix} 7 & 10 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}. \end{aligned}$$

In particolare, quindi, il prodotto di matrici non è commutativo, e moltiplicando due matrici non nulle possiamo ottenere la matrice nulla (vedi l'Esercizio 7.23).

Osservazione 7.6 Se consideriamo un vettore $x \in \mathbb{R}^n$ come vettore colonna, si vede subito che il prodotto di una matrice $A \in M_{m,n}(\mathbb{R})$ per la colonna $x \in M_{n,1}(\mathbb{R})$ è esattamente quanto abbiamo da sempre indicato con Ax , per cui non c'è rischio di confusione.

Vediamo ora se il prodotto di matrici gode delle proprietà necessarie per giustificare il nome "prodotto". Nell'enunciato seguente, dire che delle matrici hanno "le dimensioni giuste" significa che hanno dimensioni tali da rendere possibili tutte le operazioni (somme e prodotti) considerate.

Proposizione 7.5 Siano A, B e C matrici delle dimensioni giuste, e $\lambda \in \mathbb{R}$. Allora:

- (i) $A(B + C) = AB + AC$ e $(A + B)C = AC + BC$;
- (ii) $(\lambda A)B = \lambda(AB)$;
- (iii) $(AB)C = A(BC)$;
- (iv) $AI = IA = A$ e $AO = OA = O$;
- (v) $(AB)^T = B^T A^T$.

Dimostrazione. (i), (ii) e (iii) si dimostrano tutte allo stesso modo; dimostriamo per esempio (iii), lasciando le altre due per esercizio (Esercizio 7.13). Siccome l'applicazione L che associa a una matrice A l'applicazione lineare L_A è un isomorfismo (e in particolare è bigettiva), per dimostrare che $(AB)C = A(BC)$ basta far vedere che $L_{(AB)C} \equiv L_{A(BC)}$. Questo segue subito dalla definizione di prodotto e dalle proprietà della composizione di applicazioni lineari: infatti,

$$L_{(AB)C} = L_{AB} \circ L_C = (L_A \circ L_B) \circ L_C = L_A \circ (L_B \circ L_C) = L_A \circ L_{BC} = L_{A(BC)}.$$

La (iv) segue subito dal fatto che $L_I = \text{id}$ e $L_O = O$; rimane quindi solo da dimostrare (v). Nel caso in cui A sia un vettore riga $a \in M_{1,n}(\mathbb{R})$ e B un vettore colonna $b \in M_{n,1}(\mathbb{R})$ abbiamo

$$(ab)^T = ab = a_1 b_1 + \dots + a_n b_n = b^T a^T,$$

dove la prima uguaglianza vale in quanto ab è un numero reale¹, per cui la (v) in questo caso è verificata.

Vediamo ora il caso generale; poniamo, per semplicità, $C = (AB)^T$ e $D = B^T A^T$. Allora usando (5.2), la definizione di prodotto di matrici e il caso particolare appena visto otteniamo

$$d_{ij} = (B^T)_i (A^T)^j = (B^i)^T (A_j)^T = A_j B^i = (AB)_{ji} = c_{ij},$$

per cui $D = C$ e abbiamo finito. \square

¹ Anche ba e $a^T b^T$ hanno senso, ma sono entrambe delle matrici $n \times n$, l'una trasposta dell'altra.

7.3 Matrici invertibili

Consideriamo l'insieme $M_{n,n}(\mathbb{R})$ delle matrici quadrate di ordine n . Contrariamente al caso delle matrici rettangolari qualsiasi, il prodotto di due matrici di $M_{n,n}(\mathbb{R})$ si può sempre fare ed è ancora un elemento di $M_{n,n}(\mathbb{R})$, per cui il prodotto righe per colonne è un'operazione definita su $M_{n,n}(\mathbb{R})$. La Proposizione 7.5 allora ci dice che con questo prodotto e la solita somma l'insieme $M_{n,n}(\mathbb{R})$ diviene un anello non commutativo, in cui l'elemento neutro per il prodotto è la matrice identica I_n . L'insieme $M_{n,n}(\mathbb{R})$ però non è mai un corpo, in quanto contiene sempre delle matrici prive di inversa. Per esempio, se $E_{11} \in M_{n,n}(\mathbb{R})$ è la matrice con 1 al posto $(1, 1)$ e 0 altrove, allora $E_{11}A \neq I_n$ (perché?) per ogni matrice $A \in M_{n,n}(\mathbb{R})$.

Le matrici che ammettono un inverso hanno un nome e proprietà specifiche:

Definizione 7.6 Diremo che una matrice $A \in M_{n,n}(\mathbb{R})$ è *invertibile* se esiste una matrice $B \in M_{n,n}(\mathbb{R})$ tale che $AB = BA = I_n$; la matrice B è l'*inversa* di A . L'inversa, se esiste, è unica: se B' è un'altra inversa abbiamo

$$B' = B' I_n = B'(AB) = (B'A)B = I_n B = B.$$

L'inversa di A , se esiste, si indica con A^{-1} . L'insieme delle matrici invertibili di ordine n si denota con $GL_n(\mathbb{R})$, dove GL è l'abbreviazione di "gruppo lineare".

Proposizione 7.6 Siano $A, B \in GL_n(\mathbb{R})$ due matrici invertibili. Allora A^{-1}, A^T e AB sono invertibili, e si ha

$$(A^{-1})^{-1} = A, \quad (A^T)^{-1} = (A^{-1})^T \quad \text{e} \quad (AB)^{-1} = B^{-1}A^{-1}.$$

Dimostrazione. Il fatto che

$$A^{-1}A = AA^{-1} = I$$

ci dice che A è l'inversa di A^{-1} . Trasponendo la precedente uguaglianza troviamo che $(A^{-1})^T$ è l'inversa di A^T . Infine,

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

e $B^{-1}A^{-1}$ è l'inversa di AB . \square

In particolare, $GL_n(\mathbb{R})$ è un gruppo (rispetto al prodotto, non alla somma: Esercizio 7.22) *non commutativo*; vedi l'Esempio 7.6.

Vi è ora un problema fondamentale: quando una matrice $A \in M_{n,n}(\mathbb{R})$ è invertibile? La risposta è contenuta nel seguente Teorema e nel successivo Corollario:

Teorema 7.7 Sia $A \in M_{n,n}(\mathbb{R})$. Allora le seguenti affermazioni sono equivalenti:

- (i) A è invertibile;
- (ii) L_A è invertibile;
- (iii) L_A è iniettiva;
- (iv) L_A è surgettiva;
- (v) $\text{rg } A = n$;
- (vi) Le colonne di A sono linearmente indipendenti;
- (vii) Le righe di A sono linearmente indipendenti;
- (viii) Il sistema omogeneo $Ax = O$ ha come unica soluzione $x = O$;
- (ix) Per ogni $b \in \mathbb{R}^n$ il sistema $Ax = b$ ha come unica soluzione $x = A^{-1}b$;
- (x) Tutti i pivot di A , comunque determinati, sono non nulli (cioè A è non singolare).

Osservazione 7.7 Per dimostrare l'equivalenza di queste affermazioni faremo vedere che la (i) implica la (ix), che a sua volta implica la (x), che a sua volta implica la (viii), e così via fino a tornare al punto di partenza, chiudendo il cerchio. Di conseguenza ciascuna di queste affermazioni implicherà tutte le altre, per cui saranno equivalenti.

Dimostrazione. (i) \implies (ix). Supponiamo A invertibile. Allora il vettore $v^o = A^{-1}b$ è soluzione del sistema: infatti $Av^o = AA^{-1}b = Ib = b$. Inoltre è l'unica soluzione: se v è soluzione, si ha $Av = b$ e quindi, applicando A^{-1} a entrambi i membri, $v = A^{-1}b = v^o$.

(ix) \implies (x). Teorema 3.3.

(x) \implies (viii). Di nuovo il Teorema 3.3, applicato al sistema $Ax = O$.

(viii) \implies (vi). Proposizione 4.4.

(vi) \implies (vii). Proposizione 5.11.(ii).

(vii) \implies (v). Di nuovo Proposizione 5.11.(ii).

(v) \implies (iv). Corollario 5.8.(ii).

(iv) \implies (iii). Corollario 5.8.(iii).

(iii) \implies (ii). Corollario 7.3.

(ii) \implies (i). L'inversa di L_A dev'essere della forma L_B per qualche $B \in M_{n,n}(\mathbb{R})$. È allora chiaro (perché?) che B è l'inversa di A . \square

Corollario 7.8 Sia $A \in M_{n,n}(\mathbb{R})$. Allora le seguenti affermazioni sono equivalenti:

- (i) A è invertibile;
- (ii) esiste una matrice $B_1 \in M_{n,n}(\mathbb{R})$ tale che $B_1A = I_n$;
- (iii) esiste una matrice $B_2 \in M_{n,n}(\mathbb{R})$ tale che $AB_2 = I_n$.

Dimostrazione. (i) \implies (ii), (iii). Basta prendere $B_1 = B_2 = A^{-1}$.

(ii) \implies (i). Supponiamo che $v \in \mathbb{R}^n$ sia tale che $L_A(v) = O$. Allora (ii) ci dà

$$O = L_{B_1}(O) = L_{B_1}(L_A(v)) = L_{B_1A}(v) = L_{I_n}(v) = v,$$

per cui $\text{Ker } L_A = \{O\}$, cioè L_A è iniettiva e A è invertibile.

(iii) \implies (i). Prendiamo $v \in \mathbb{R}^n$, e sia $w = L_{B_2}(v)$. Allora

$$L_A(w) = L_A(L_{B_2}(v)) = L_{AB_2}(v) = L_{I_n}(v) = v,$$

per cui L_A è surgettiva, e A è invertibile. \square

Osservazione 7.8 Nel Capitolo 9 vedremo che tutte queste affermazioni sono equivalenti anche a richiedere che il determinante della matrice sia diverso da zero.

Rimane il problema, una volta determinato che $A \in M_{n,n}(\mathbb{R})$ è invertibile, di come calcolarne l'inversa. Il caso $n = 2$ è trattato nell'Esercizio 7.19; per $n > 2$ giunge in aiuto l'eliminazione di Gauss. Prima di tutto, grazie al Corollario 7.8 la matrice A è invertibile se e solo se esiste una matrice $X \in M_{n,n}(\mathbb{R})$ tale che $AX = I_n$. Ora, dire che $AX = I_n$ equivale a dire che $AX^1 = e_1, \dots, AX^n = e_n$; quindi per trovare X dobbiamo risolvere simultaneamente i sistemi lineari quadrati

$$Ax = e_1, \quad \dots, \quad Ax = e_n, \quad (7.2)$$

che hanno tutti la stessa matrice dei coefficienti. La soluzione del primo sistema sarà la prima colonna di A^{-1} , la soluzione del secondo sistema sarà la seconda colonna di A^{-1} , e così via.

L'eliminazione di Gauss ci fornisce una tecnica ideale per risolvere un problema di questo genere; infatti, le operazioni necessarie per ridurre in forma triangolare superiore i sistemi (7.2) dipendono solo dalla matrice A , e quindi possono essere effettuate una volta sola per tutti i sistemi considerati. Vediamo questo procedimento in azione su un esempio, in cui sarà indicato anche come concludere il calcolo dell'inversa senza bisogno di risolvere all'indietro i sistemi (7.2) ma usando in modo opportuno operazioni elementari.

ESEMPIO 7.7 Vogliamo determinare l'inversa (se esiste) della matrice

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

Dobbiamo risolvere i tre sistemi $Ax = e_1$, $Ax = e_2$ e $Ax = e_3$. Applichiamo l'eliminazione di Gauss ad A e alla matrice identica (che contiene i termini noti dei sistemi che ci interessano):

$$\left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 4 & 1 & 0 & 0 & 1 & 0 \\ -2 & 2 & 1 & 0 & 0 & 1 \end{array} \right] \longrightarrow \left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -1 & -2 & -2 & 1 & 0 \\ 0 & 3 & 2 & 1 & 0 & 1 \end{array} \right] \longrightarrow \left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -1 & -2 & -2 & 1 & 0 \\ 0 & 0 & -4 & -5 & 3 & 1 \end{array} \right].$$

Quindi la matrice A è non singolare, e in particolare è invertibile. A questo punto, invece di risolvere all'indietro i tre sistemi triangolari superiori che abbiamo ottenuto,

agiamo con delle operazioni elementari in modo analogo a quanto si fa nell'eliminazione di Gauss, ma partendo dalla terza riga in basso e salendo:

$$\begin{array}{l} \left(\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -1 & -2 & -2 & 1 & 0 \\ 0 & 0 & -4 & -5 & 3 & 1 \end{array} \right) \longrightarrow \left(\begin{array}{ccc|ccc} 2 & 1 & 0 & -1/4 & 3/4 & 1/4 \\ 0 & -1 & 0 & 1/2 & -1/2 & -1/2 \\ 0 & 0 & -4 & -5 & 3 & 1 \end{array} \right) \\ \longrightarrow \left(\begin{array}{ccc|ccc} 2 & 0 & 0 & 1/4 & 1/4 & -1/4 \\ 0 & -1 & 0 & 1/2 & -1/2 & -1/2 \\ 0 & 0 & -4 & -5 & 3 & 1 \end{array} \right) = |S|T|. \end{array}$$

La matrice S così ottenuta è la matrice dei coefficienti di tre sistemi ancora equivalenti ai tre di partenza; i termini noti di questi nuovi sistemi sono contenuti nella matrice T . Quindi se dividiamo ciascuna riga di T per il corrispondente pivot otteniamo le soluzioni dei tre sistemi di partenza, e quindi la matrice inversa

$$A^{-1} = \begin{pmatrix} 1/8 & 1/8 & -1/8 \\ -1/2 & 1/2 & 1/2 \\ 5/4 & -3/4 & -1/4 \end{pmatrix}.$$

Controlla, per esercizio, che $A \cdot A^{-1} = A^{-1} \cdot A = I_3$.

L'Esercizio 9.16 descrive un altro modo per calcolare l'inversa di una matrice.

Esercizi

7.1 Siano $S:U \rightarrow V$ e $T:V \rightarrow W$ due applicazioni lineari fra spazi vettoriali. Dimostra che $\text{rg}(T \circ S) \leq \min\{\text{rg} S, \text{rg} T\}$.

7.2 Siano $T:\mathbb{R}^3 \rightarrow \mathbb{R}_2[t]$ ed $S:\mathbb{R}_2[t] \rightarrow \mathbb{R}^2$ le applicazioni lineari date da

$$T \begin{pmatrix} x \\ y \\ z \end{pmatrix} = x + (2y - x)t + zt^2 \quad \text{e} \quad S(p) = \begin{pmatrix} p(0) \\ p(1) \end{pmatrix}.$$

Calcola $S \circ T: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ e trova $\text{Ker}(S \circ T)$ e $\text{Im}(S \circ T)$.

7.3 Sia $T:V \rightarrow V$ un endomorfismo tale che $T \circ T = O$. Dimostra che $T + \text{id}_V$ è invertibile.

7.4 Scrivi un'applicazione lineare da $M_{3,3}(\mathbb{R})$ a $M_{2,2}(\mathbb{R})$. Trova un isomorfismo fra $M_{3,4}(\mathbb{R})$ e $M_{6,2}(\mathbb{R})$, se esiste.

7.5 Sia $T_a:\mathbb{R}_2[t] \rightarrow \mathbb{R}^3$ l'applicazione lineare data da

$$\forall p \in \mathbb{R}_2[t] \quad T_a(p) = \begin{pmatrix} p(0) \\ p(a) \\ p(1) \end{pmatrix},$$

dove $a \in \mathbb{R}$. Trova per quali valori di a l'applicazione T_a è un isomorfismo.

7.6 Dimostra che un sistema lineare quadrato $Ax = b$ ammette soluzione qualunque sia b se e solo se il sistema omogeneo $Ax = O$ ammette solo la soluzione $x = O$.

7.7 Un endomorfismo $P \in \mathcal{L}(V, V)$ tale che $P \circ P = P$ si chiama *proiezione*. Dimostra che se P è una proiezione allora $V = \text{Ker} P \oplus \text{Im} P$. (Suggerimento: prova a scrivere $v = (v - P(v)) + P(v)$.)

7.8 Dimostra che due spazi vettoriali di dimensione finita sono isomorfi se e solo se hanno la stessa dimensione.

7.9 Siano U e W due sottospazi di uno spazio vettoriale V tali che $U \cap W = \{O\}$. Dimostra che $U \oplus W$ è isomorfo al prodotto cartesiano $U \times W$ (vedi l'Esercizio 4.27).

7.10 Dato $\lambda \in \mathbb{R}$, quale matrice viene associata all'applicazione $T_\lambda \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ definita da $T_\lambda(x) = \lambda x$?

7.11 Sia $T:V \rightarrow W$ un'applicazione lineare, e siano \mathcal{B} una base di V e \mathcal{C} una base di W . Dimostra che per ogni $v \in V$ le coordinate di $T(v)$ rispetto alla base \mathcal{C} si esprimono come polinomi di primo grado senza termini noti nelle coordinate di v rispetto alla base \mathcal{B} . (Suggerimento: considera $F_{\mathcal{C}} \circ T \circ (F_{\mathcal{B}})^{-1}$.)

7.12 Verifica che se $a \in \mathbb{R}^n$ l'applicazione $\varphi_a:\mathbb{R}^n \rightarrow \mathbb{R}$ definita nell'Esempio 5.9 è data da $\varphi_a(x) = a^T \cdot x$. Utilizza questa formula per ridimostrare il Lemma 5.10.

7.13 Dimostra la Proposizione 7.5.(i) e (ii).

7.14 Calcola tutti i prodotti possibili fra le matrici seguenti:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} -2 & 3 \\ 1 & 7 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 4 \end{pmatrix}.$$

7.15 Data la matrice

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

trova tutte le matrici $B \in M_{2,2}(\mathbb{R})$ tali che $AB = BA$.

7.16 Data la matrice

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 0 & 0 & 1 \end{pmatrix} \in M_{3,3}(\mathbb{R}),$$

calcola la potenza n -esima di A (cioè il prodotto di A per se stessa n volte, che si indica con A^n) per ogni $n \in \mathbb{N}$.

7.17 Data la matrice

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

considera l'insieme $W = \{X \in M_{2,2}(\mathbb{R}) \mid AX = O\}$. Dimostra che W è un sottospazio vettoriale di $M_{2,2}(\mathbb{R})$, calcolane la dimensione e trovanne una base.

7.18 Trova per quali $k \in \mathbb{R}$ esiste una matrice $X \in M_{3,2}(\mathbb{R})$ tale che

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

7.19 Sia

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_{2,2}(\mathbb{R}).$$

Dimostra che A è invertibile se e solo se $ad - bc \neq 0$, e che l'inversa è data da

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

(Suggerimento: vedi l'Esercizio 4.16.)

7.20 Calcola l'inversa (se esiste) delle seguenti matrici:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 0 & 2 & 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & -1 & 2 \\ 0 & 1 & 0 & 1 \\ 2 & 3 & 2 & 0 \\ 1 & 0 & -1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 4 & 1 & -1 & 1 \\ 1 & -3 & 0 & 1 \\ 1 & 0 & 2 & 0 \\ 0 & 0 & 1 & 2 \end{pmatrix}.$$

7.21 Trova per quali $h \in \mathbb{R}$ la seguente matrice

$$\begin{pmatrix} h & 1 & 0 \\ 1 & h & h \\ 0 & 1 & 2 \end{pmatrix}$$

è invertibile, e calcolane quando possibile l'inversa.

7.22 Trova due matrici quadrate invertibili la cui somma sia una matrice non invertibile (per cui $GL_n(\mathbb{R})$ non è un gruppo rispetto alla somma).

7.23 Sia $A \in M_{n,n}(\mathbb{R})$. Dimostra che A non è invertibile se e solo se esiste una matrice $B \in M_{n,n}(\mathbb{R})$ non nulla tale che $AB = O$.

7.24 Sia V un sottospazio di \mathbb{R}^n di equazioni parametriche $v = At$, dove $t \in \mathbb{R}^k$ per cui $A \in M_{n,k}(\mathbb{R})$, e sia $B \in M_{n-\text{rg } A, n}(\mathbb{R})$ una matrice le cui righe (trasposte) sono una base di $\text{Ker } A^T$. Dimostra che $Bx = O$ sono equazioni cartesiane per V .

7.25 Siano $A, B \in M_{n,n}(\mathbb{R})$; dimostra che $\text{tr}(AB) = \text{tr}(BA)$, dove $\text{tr } A$ è la traccia della matrice A (definita nell'Esercizio 5.6), e deduci che $\text{tr}(B^{-1}AB) = \text{tr } A$ per ogni $A \in M_{n,n}(\mathbb{R})$ e $B \in GL_n(\mathbb{R})$.

7.26 Dimostra che non esistono due matrici $A, B \in M_{n,n}(\mathbb{R})$ tali che $AB - BA = I_n$. (Suggerimento: usa l'Esercizio precedente.)

7.27 Dimostra che le proiezioni definite nell'Esempio 5.8 e nell'Esercizio 5.5 sono proiezioni anche nel senso dell'Esercizio 7.7.

Cambiamenti di base

o ogni vettore $v \in V$
 ha coordinate in \mathcal{B}

In questo capitolo studieremo più in dettaglio l'isomorfismo $F_{\mathcal{B}}: V \rightarrow \mathbb{R}^n$ fra uno spazio vettoriale V di dimensione n e lo spazio \mathbb{R}^n che associa a ogni vettore $v \in V$ le sue coordinate rispetto a una base \mathcal{B} . Vedremo come muta $F_{\mathcal{B}}$ cambiando la base \mathcal{B} ; come associare una matrice a qualunque applicazione lineare $T: V \rightarrow W$ una volta fissate una base di V e una di W ; e vedremo come muta questa matrice cambiando le basi in V e W . Come conseguenza ricaveremo delle tecniche efficienti per risolvere in qualunque spazio vettoriale (di-dimensione finita) i problemi che nel Paragrafo 6.3 avevamo risolto in \mathbb{R}^n . I Complementi a questo capitolo, infine, sono dedicati allo studio dello spazio duale di uno spazio vettoriale.

8.1 Matrice di cambiamento di base

Nel Paragrafo 6.3 abbiamo risolto tutta una serie di problemi di calcolo per \mathbb{R}^n ; gli stessi problemi rimangono però ancora aperti in spazi vettoriali qualunque. In particolare, se V è uno spazio vettoriale e $v_1, \dots, v_k \in V$, come si trovano dimensione e base di $\text{Span}(v_1, \dots, v_k)$? E dimensione e base di somma e intersezione di sottospazi? Se $T: V \rightarrow W$ è un'applicazione lineare, come si trova il rango di T e una base di $\text{Im } T$? E dimensione e base di $\text{Ker } T$?

La risposta consiste nel "riconcondursi al problema precedente" (frase proverbialmente associata ai matematici). L'idea è che, come discusso nell'Osservazione 7.3, ogni spazio vettoriale V di dimensione n è isomorfo a \mathbb{R}^n ; quindi per risolvere qualunque problema su V che riguardi solo la struttura di spazio vettoriale possiamo trasferire il problema in \mathbb{R}^n , lì risolverlo e ritrasportare indietro la soluzione, in modo da avere la risposta in V . Questo trasferimento avviene sostituendo ai vettori le loro coordinate rispetto a una base fissata \mathcal{B} , cioè utilizzando l'isomorfismo $F_{\mathcal{B}}: V \rightarrow \mathbb{R}^n$ definito nell'Esempio 5.7. Un paio di esempi chiariranno le idee:

ESEMPIO 8.1 Sia $V = \mathbb{R}_2[t]$, e consideriamo i polinomi $p_1(t) = t - 1$, $p_2(t) = t^2 - t$ e $p_3(t) = 2t^2 + 4t - 6$. Vogliamo dimensione e base di $W = \text{Span}(p_1, p_2, p_3) \subseteq V$. Il procedimento sarà: fissata una base \mathcal{B} di V , troveremo la dimensione d e una base $\{v_1, \dots, v_d\}$ di $\text{Span}(F_{\mathcal{B}}(p_1), F_{\mathcal{B}}(p_2), F_{\mathcal{B}}(p_3)) \subseteq \mathbb{R}^3$. Allora, essendo $F_{\mathcal{B}}$ un iso-

morfismo, la dimensione di W sarà d , e una base sarà $\{(F_B)^{-1}(v_1), \dots, (F_B)^{-1}(v_d)\}$. Nel nostro caso, scegliamo come base $\mathcal{B} = \{1, t, t^2\}$. Allora

$$F_{\mathcal{B}}(p_1) = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \quad F_{\mathcal{B}}(p_2) = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \quad F_{\mathcal{B}}(p_3) = \begin{pmatrix} -6 \\ 4 \\ 2 \end{pmatrix}$$

sono le coordinate di p_1, p_2 e p_3 nella base \mathcal{B} . Con le tecniche viste nel Paragrafo 6.3 troviamo subito che la dimensione di $\text{Span}(F_{\mathcal{B}}(p_1), F_{\mathcal{B}}(p_2), F_{\mathcal{B}}(p_3))$ è 2, e che una sua base è $\{v_1 = F_{\mathcal{B}}(p_1), v_2 = F_{\mathcal{B}}(p_2)\}$; quindi la dimensione di $\text{Span}(p_1, p_2, p_3) \subset V$ è 2, e una base è data da $\{p_1, p_2\}$.

ESEMPIO 8.2 Sia $V = \mathbb{R}_3[t]$, e consideriamo i polinomi $q_1(t) = t^2 - 1$ e $q_2(t) = t^3 + t$. Poniamo $U = \text{Span}(q_1, q_2)$ e $W = \text{Span}(p_1, p_2, p_3)$, dove p_1, p_2 e p_3 sono i polinomi dell'esempio precedente. Vogliamo dimensione e una base di $U + W$ e di $U \cap W$. Utilizziamo la base $\mathcal{B} = \{1, t, t^2, t^3\}$; rispetto a questa base, i polinomi che ci interessano hanno coordinate

$$F_{\mathcal{B}}(p_1) = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad F_{\mathcal{B}}(p_2) = \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \quad F_{\mathcal{B}}(p_3) = \begin{pmatrix} -6 \\ 4 \\ 2 \\ 0 \end{pmatrix},$$

$$F_{\mathcal{B}}(q_1) = \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad F_{\mathcal{B}}(q_2) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

Sappiamo che una base di W è $\{p_1, p_2\}$, per cui possiamo tralasciare p_3 . Quindi, siccome grazie all'Esercizio 8.1 si ha $F_{\mathcal{B}}(U+W) = F_{\mathcal{B}}(U) + F_{\mathcal{B}}(W)$, per studiare $U+W$ e $U \cap W$ dobbiamo considerare la matrice

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

Una veloce riduzione a scala mostra che $\text{rg } A = 3$ e che una base di $\text{Im } A$ è data da $\{A^1, A^2, A^4\}$; quindi le tecniche studiate nel Paragrafo 6.3 ci dicono subito che $\dim(F_{\mathcal{B}}(U) + F_{\mathcal{B}}(W)) = 3$, $\dim(F_{\mathcal{B}}(U) \cap F_{\mathcal{B}}(W)) = 1$, e che $\{F_{\mathcal{B}}(p_1), F_{\mathcal{B}}(p_2), F_{\mathcal{B}}(q_2)\}$ è una base di $F_{\mathcal{B}}(U) + F_{\mathcal{B}}(W)$. Ne segue che $\dim U + W = 3$, che $\dim U \cap W = 1$ e che $\{p_1, p_2, q_2\}$ è una base di $U + W$. Infine sempre le tecniche del Paragrafo 6.3 mostrano che $\{F_{\mathcal{B}}(q_1)\}$ è una base di $F_{\mathcal{B}}(U) \cap F_{\mathcal{B}}(W)$, per cui $\{q_1\}$ è una base di $U \cap W$.

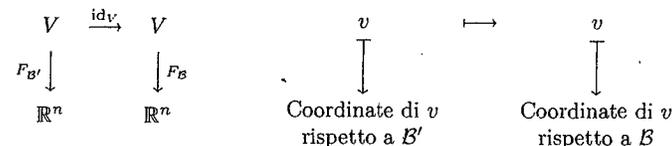
Dunque adesso sappiamo come risolvere i primi due problemi (calcolo di dimensione e base di uno span, e dell'intersezione e/o somma di sottospazi) in qualunque spazio vettoriale di dimensione finita. Ovviamente, i conti che devono essere effettuati dipendono dalla scelta di una base.

ESEMPIO 8.3 Sia $V = \mathbb{R}_2[t]$ e $p_1, p_2, p_3 \in V$ gli stessi polinomi dell'Esempio 8.1; questa volta prendiamo come base $\mathcal{B}' = \{1, t-1, 2t^2+4t-6\}$. Ovviamente, bisogna prima verificare che \mathcal{B}' sia effettivamente una base di $\mathbb{R}_2[t]$; ma è sufficiente osservare che $F_{\mathcal{B}'}(1), F_{\mathcal{B}'}(t-1)$ e $F_{\mathcal{B}'}(2t^2+4t-6)$ sono linearmente indipendenti in \mathbb{R}^3 (esercizio). Le coordinate dei tre polinomi rispetto a \mathcal{B}' sono date da

$$F_{\mathcal{B}'}(p_1) = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad F_{\mathcal{B}'}(p_2) = \begin{pmatrix} 2 \\ -3 \\ 1/2 \end{pmatrix}, \quad F_{\mathcal{B}'}(p_3) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

per cui senza bisogno di altri conti vediamo subito che $\dim \text{Span}(p_1, p_2, p_3) = 2$ e che una base è data da $\{p_1, p_3\}$.

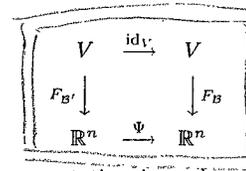
Diventa quindi interessante vedere cosa succede in generale alle coordinate di un vettore cambiando base. Sia V uno spazio vettoriale di dimensione n , e fissiamo due basi $\mathcal{B} = \{v_1, \dots, v_n\}$ e $\mathcal{B}' = \{v'_1, \dots, v'_n\}$ di V . Questo ci fornisce due isomorfismi $F_{\mathcal{B}}, F_{\mathcal{B}'}: V \rightarrow \mathbb{R}^n$, ciascuno dei quali associa a un vettore $v \in V$ le sue coordinate rispetto alla corrispondente base. Abbiamo quindi la seguente situazione:



A questo punto viene la tentazione di chiudere il quadrato tracciando una freccia nel lato inferiore; si tratta di trovare un'applicazione lineare da \mathbb{R}^n in \mathbb{R}^n che trasformi le coordinate rispetto a \mathcal{B}' in coordinate rispetto a \mathcal{B} . Ma questa applicazione esiste: si tratta di

$$\Psi = F_{\mathcal{B}} \circ (F_{\mathcal{B}'})^{-1}.$$

Il diagramma completo



è un esempio di *diagramma commutativo*: in qualunque ordine si seguono le frecce, si ottiene lo stesso risultato. In questo caso vuol dire che (partendo dall'angolo in alto a sinistra) andare prima a destra e poi scendere è lo stesso di prima scendere e poi andare a destra; infatti

$$F_{\mathcal{B}} \circ \text{id}_V = F_{\mathcal{B}} = \Psi \circ F_{\mathcal{B}'}$$

Ora, $\Psi \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$; per la Proposizione 7.4, deve esistere una matrice quadrata B tale che $\Psi = L_B$; essendo Ψ invertibile (l'inversa è $F_{\mathcal{B}'} \circ (F_{\mathcal{B}})^{-1}$), anche la matrice B è invertibile.

Dunque ogni volta che abbiamo due basi \mathcal{B} e \mathcal{B}' di uno spazio vettoriale V troviamo una matrice invertibile B che trasforma le coordinate rispetto a \mathcal{B}' nelle coordinate

rispetto a \mathcal{B} . In concreto, preso un vettore $v \in V$ indichiamo con $x = F_{\mathcal{B}}(v) \in \mathbb{R}^n$ le sue coordinate rispetto a \mathcal{B} , e con $x' = F_{\mathcal{B}'}(v) \in \mathbb{R}^n$ le sue coordinate rispetto a \mathcal{B}' . Allora si ha

$$x = Bx', \tag{8.1}$$

cioè

$$\begin{cases} x_1 = b_{11}x'_1 + \dots + b_{1n}x'_n, \\ \vdots \\ x_n = b_{n1}x'_1 + \dots + b_{nn}x'_n. \end{cases}$$

Nota che

$$x' = B^{-1}x.$$

Si tratta ora di vedere come si trova la matrice B . La formula (5.4) ci dice che la colonna h -esima di B è data da

$$B^h = L_B(e_h) = \Psi(e_h) = (F_B \circ (F_{\mathcal{B}'}^{-1}))(e_h) = F_B((F_{\mathcal{B}'}^{-1})(e_h)) = F_B(v'_h),$$

cioè è data dalle coordinate di v'_h rispetto alla base \mathcal{B} . In altre parole, la matrice B contiene per colonne le coordinate dei vettori della nuova base \mathcal{B}' rispetto alla vecchia base \mathcal{B} . In simboli,

$$B = \left| F_B(v'_1) \dots F_B(v'_n) \right|.$$

Definizione 8.1 La matrice B viene chiamata matrice di cambiamento di base (o di passaggio) da \mathcal{B} a \mathcal{B}' . Chiaramente, B^{-1} è la matrice di cambiamento di base da \mathcal{B}' a \mathcal{B} ; cioè nella direzione inversa.

Osservazione 8.1 Dire che B contiene per colonne le coordinate dei vettori della nuova base \mathcal{B}' rispetto alla vecchia base \mathcal{B} è equivalente (perché?) a scrivere

$$\left| v'_1 \dots v'_n \right| = \left| v_1 \dots v_n \right| B, \tag{8.2}$$

dove il prodotto è il solito prodotto righe per colonne.

ESEMPIO 8.4 Prendiamo $V = \mathbb{R}_2[t]$, $\mathcal{B} = \{1, t, t^2\}$ e $\mathcal{B}' = \{1, t - 1, 2t^2 + 4t - 6\}$. Allora la matrice di cambiamento di base da \mathcal{B} a \mathcal{B}' contiene per colonne le coordinate dei polinomi di \mathcal{B}' rispetto alla base \mathcal{B} , cioè

$$B = \begin{vmatrix} 1 & -1 & -6 \\ 0 & 1 & 4 \\ 0 & 0 & 2 \end{vmatrix}.$$

Le matrici di cambiamento di base compaiono, ovviamente, anche nel caso di sottospazi di \mathbb{R}^n :

ESEMPIO 8.5 Sia $V \subset \mathbb{R}^3$ lo spazio descritto nell'Esempio 4.12, e consideriamo le due basi $\mathcal{B} = \{v_1, v_2\}$ e $\mathcal{B}' = \{e_1, e_2\}$, dove i vettori $v_1, v_2 \in V$ sono dati da

$$v_1 = \begin{vmatrix} 3 \\ 1 \\ 0 \end{vmatrix} \quad \text{e} \quad v_2 = \begin{vmatrix} -1 \\ -1 \\ 0 \end{vmatrix}.$$

Allora

$$\begin{cases} e_1 = \frac{1}{2}v_1 + \frac{1}{2}v_2, \\ e_2 = -\frac{1}{2}v_1 - \frac{3}{2}v_2, \end{cases}$$

per cui la matrice di passaggio da \mathcal{B} a \mathcal{B}' è data da

$$B = \begin{vmatrix} 1/2 & -1/2 \\ 1/2 & -3/2 \end{vmatrix}.$$

Analogamente,

$$\begin{cases} v_1 = 3e_1 + e_2, \\ v_2 = -e_1 - e_2, \end{cases}$$

per cui la matrice di cambiamento di base da \mathcal{B}' a \mathcal{B} è data da

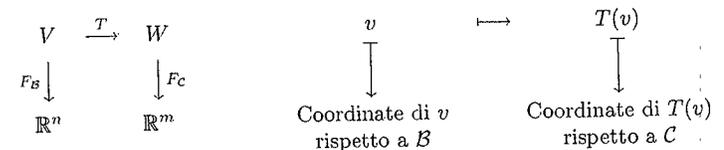
$$B' = \begin{vmatrix} 3 & -1 \\ 1 & -1 \end{vmatrix}.$$

È facile verificare che $BB' = B'B = I_2$.

8.2 Matrice associata a un'applicazione lineare

Torniamo ora ai problemi posti all'inizio; dobbiamo vedere come risolvere quelli relativi alle applicazioni lineari. Siano V, W spazi vettoriali, con $\dim V = n$ e $\dim W = m$, e $T: V \rightarrow W$ un'applicazione lineare; vogliamo associare a T un'applicazione lineare di \mathbb{R}^n in \mathbb{R}^m (ovvero una matrice) da usare per rispondere a tutti i quesiti che potremmo porci su T . L'idea è che la matrice deve fare alle coordinate dei vettori ciò che T fa ai vettori stessi.

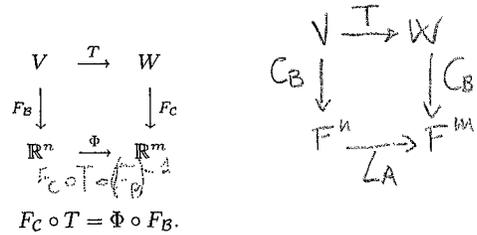
Fissiamo allora una base $\mathcal{B} = \{v_1, \dots, v_n\}$ di V e una base $\mathcal{C} = \{w_1, \dots, w_m\}$ di W . Questa volta abbiamo il seguente diagramma:



Anche questa volta possiamo chiudere il quadrato: se poniamo

$$\Phi = F_{\mathcal{C}} \circ T \circ (F_{\mathcal{B}})^{-1}, \tag{8.3}$$

il diagramma



diventa commutativo, cioè

Ora, $\Phi \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$; per la Proposizione 7.4 deve esistere una matrice $A \in M_{m,n}(\mathbb{R})$ tale che $\Phi = L_A$.

Definizione 8.2 La matrice A è la matrice associata a T (o che rappresenta T) rispetto alle basi B e C .

Dunque se per ogni $v \in V$ indichiamo con $x \in \mathbb{R}^n$ le coordinate di v rispetto a B , il vettore $y = Ax \in \mathbb{R}^m$ deve contenere le coordinate di $T(v)$ rispetto alla base C . Per calcolare esplicitamente la matrice A , basta notare che, siccome le coordinate di v_j rispetto a B sono date dal vettore e_j , abbiamo

$$A^j = L_A(e_j) = F_C \circ T \circ (F_B)^{-1}(e_j) = F_C(T(v_j)),$$

per cui la j -esima colonna A^j di A contiene le coordinate di $T(v_j)$ rispetto a C . In altri termini, la matrice A contiene per colonne le coordinate rispetto alla base di arrivo C dei trasformati secondo T dei vettori della base di partenza B . In simboli,

$$|T(v_1) \cdots T(v_n)| = |w_1 \cdots w_m| A.$$

ESEMPIO 8.6 Prendiamo $V = \mathbb{R}_2[t]$, $W = \mathbb{R}_3[t]$ e sia $T: V \rightarrow W$ l'applicazione lineare data da $p(t) \mapsto t^2 p'(t+1)$, dove l'apice indica la derivata rispetto a t . Per esempio, se $p(t) = t^2$ allora $p'(t) = 2t$, $p'(t+1) = 2(t+1)$ e

$$[T(p)](t) = t^2 p'(t+1) = 2t^2(t+1) = 2t^3 + 2t^2,$$

Prendiamo come basi le solite $B = \{1, t, t^2\}$ e $C = \{1, t, t^2, t^3\}$; essendo $\dim V = 3$ e $\dim W = 4$, la matrice A associata a T sarà una matrice 4×3 . Per trovarla, calcoliamo i trasformati dei vettori di B :

$$T(1) = 0, \quad T(t) = t^2, \quad T(t^2) = 2t^3 + 2t^2.$$

Scrivendo le coordinate dei trasformati rispetto alla base C otteniamo:

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

ESEMPIO 8.7 Prendiamo $A \in M_{m,n}(\mathbb{R})$. Allora A è la matrice che rappresenta $L_A: \mathbb{R}^n \rightarrow \mathbb{R}^m$ rispetto alle basi canoniche di \mathbb{R}^n ed \mathbb{R}^m (perché?). Se invece prendiamo basi non canoniche, L_A può essere rappresentata da una matrice diversa. Per esempio, sia $T: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ data da

$$T \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2x + 2z \\ x - y \end{pmatrix}.$$

Allora $T = L_A$, dove

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

Prendiamo come basi

$$B' = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\} \quad \text{e} \quad C' = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix} \right\}.$$

Allora

$$T \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix} = 10 \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 3 \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad T \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -2 \end{pmatrix} = 4 \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 2 \begin{pmatrix} 2 \\ 3 \end{pmatrix},$$

$$T \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 2 \\ 3 \end{pmatrix},$$

per cui la matrice associata a T rispetto alle basi B' e C' è

$$A' = \begin{pmatrix} 10 & 4 & -2 \\ -3 & -2 & 1 \end{pmatrix}.$$

A questo punto abbiamo gli strumenti necessari per rispondere a tutte le domande iniziali. Infatti, $F_C \circ T = L_{A'} \circ F_{B'}$ ci dice esattamente (perché?) che

$$\text{Im } T = (F_C)^{-1}(\text{Im } A') \quad \text{e} \quad \text{Ker } T = (F_{B'})^{-1}(\text{Ker } A');$$

quindi tutte le informazioni che cerchiamo le possiamo recuperare esaminando A' .

ESEMPIO 8.8 Siano V, W e $T: V \rightarrow W$ come nell'Esempio 8.6; vogliamo dimensioni e basi di $\text{Ker } T$ e $\text{Im } T$. Prima di tutto, si vede subito che il rango della matrice A che rappresenta T rispetto alle basi B e C è 2; quindi, essendo F_C e F_B isomorfismi,

$$\text{rg } T = \dim \text{Im } T = \dim \text{Im } A = \text{rg } A = 2 \quad \text{e} \quad \dim \text{Ker } T = \dim \text{Ker } A = 3 - 2 = 1.$$

Siccome le ultime due colonne di A sono indipendenti, esse formano una base di $\text{Im } A$; dunque i trasformati degli ultimi due elementi di B , cioè $\{t^2, 2t^3 + 2t^2\}$, compongono una base di $\text{Im } T$. Infine, la prima colonna di A è nulla, cioè $Ae_1 = 0$; quindi il primo vettore di B forma una base di $\text{Ker } T$, la base $\{1\}$.

Possiamo formalizzare quanto ottenuto con la seguente

Proposizione 8.1 Siano V e W spazi vettoriali, con $\dim V = n$ e $\dim W = m$. Fissiamo una base \mathcal{B} di V e una base \mathcal{C} di W . Allora l'applicazione $T \mapsto A$ che associa a un'applicazione lineare $T: V \rightarrow W$ la matrice che la rappresenta rispetto alle basi \mathcal{B} e \mathcal{C} è un isomorfismo fra $\mathcal{L}(V, W)$ e $M_{m,n}(\mathbb{R})$. In particolare,

$$\dim \mathcal{L}(V, W) = (\dim V)(\dim W).$$

Dimostrazione. Basta dimostrare che l'associazione $T \mapsto L_A$ è un isomorfismo fra $\mathcal{L}(V, W)$ ed $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$; infatti la Proposizione 7.4 ci dice già che $L_A \mapsto A$ è un isomorfismo fra $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ ed $M_{m,n}(\mathbb{R})$, e la composizione di isomorfismi è ancora un isomorfismo (esercizio).

Definiamo $\Phi: \mathcal{L}(V, W) \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ ponendo

$$\Phi(T) = L_A = F_C \circ T \circ (F_B)^{-1}.$$

L'applicazione Φ è chiaramente lineare; per esempio,

$$\Phi(T + S) = F_C \circ (T + S) \circ (F_B)^{-1} = F_C \circ T \circ (F_B)^{-1} + F_C \circ S \circ (F_B)^{-1} = \Phi(T) + \Phi(S).$$

Rimane da dimostrare che Φ è invertibile. L'inversa $\Psi: \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \rightarrow \mathcal{L}(V, W)$ è

$$\Psi(L_A) = (F_C)^{-1} \circ L_A \circ F_B.$$

Infatti,

$$\Psi \circ \Phi(T) = (F_C)^{-1} \circ \Phi(T) \circ F_B = (F_C)^{-1} \circ F_C \circ T \circ (F_B)^{-1} \circ F_B = T,$$

e analogamente $\Phi \circ \Psi(L_A) = L_A$. □

Ovviamente, la matrice A dipende fortemente dalle basi scelte; cambiando basi, cambia la matrice.

ESEMPIO 8.9 Siano V, W e T come nell'Esempio 8.6, ma stavolta prendiamo come base di V l'insieme $\mathcal{B}' = \{t+1, t^2+1, t^2+t\}$; lasciamo invariata la base di W . Questa volta i trasformati della base di partenza sono

$$T(t+1) = t^2, \quad T(t^2+1) = 2t^3 + 2t^2, \quad T(t^2+t) = 2t^3 + 3t^2,$$

per cui la matrice associata è

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

Vediamo di capire come la matrice associata dipende dalle basi. Siano V e W due spazi vettoriali con $\dim V = n$ e $\dim W = m$. Scegliamo due basi $\mathcal{B}, \mathcal{B}'$ di V , due basi $\mathcal{C}, \mathcal{C}'$ di W , e sia $T: V \rightarrow W$ un'applicazione lineare. Indichiamo con B la matrice di cambiamento di base da \mathcal{B} a \mathcal{B}' , con D la matrice di cambiamento di base da \mathcal{C} a \mathcal{C}' (in modo che D^{-1} sia la matrice di cambiamento di base da \mathcal{C}' a \mathcal{C}), con A la matrice che rappresenta T rispetto alle basi \mathcal{B} e \mathcal{C} , e con A' la matrice che rappresenta T rispetto alle basi \mathcal{B}' e \mathcal{C}' . Mettendo insieme quanto visto finora, otteniamo i seguenti diagrammi commutativi:

$$\begin{array}{ccccccc} V & \xrightarrow{\text{id}_V} & V & \xrightarrow{T} & W & \xrightarrow{\text{id}_W} & W \\ F_{B'} \downarrow & & F_B \downarrow & & \downarrow F_C & & \downarrow F_{C'} \\ \mathbb{R}^n & \xrightarrow{L_B} & \mathbb{R}^n & \xrightarrow{L_A} & \mathbb{R}^m & \xrightarrow{(L_D)^{-1}} & \mathbb{R}^m \end{array} \quad \begin{array}{ccc} V & \xrightarrow{T} & W \\ F_{B'} \downarrow & & \downarrow F_{C'} \\ \mathbb{R}^n & \xrightarrow{L_{A'}} & \mathbb{R}^m \end{array}$$

Ora l'applicazione in basso nel diagramma a destra, $L_{A'}$, è l'unica che può completare il diagramma; ma se mettiamo insieme le applicazioni in basso nel diagramma a sinistra ne otteniamo una seconda (nota che gli altri tre lati del diagramma a destra corrispondono esattamente ai tre lati esterni del diagramma a sinistra). L'unica possibilità è che le due applicazioni coincidano: quindi $L_{A'} = (L_D)^{-1} \circ L_A \circ L_B$, ovvero

$$A' = D^{-1}AB. \tag{8.4}$$

Questo è il modo con cui cambia la matrice associata a T al cambiare delle basi.

Questa tecnica di procedere con i diagrammi è molto efficace (abbiamo finito senza fare alcun conto) ma forse poco intuitiva. Vediamo allora di ottenere lo stesso risultato in un altro modo. Prendiamo $v \in V$, e indichiamo con $x, x' \in \mathbb{R}^n$ le sue coordinate rispetto alle basi \mathcal{B} e \mathcal{B}' . Analogamente, indichiamo con $y, y' \in \mathbb{R}^m$ le coordinate di $T(v)$ rispetto alle basi \mathcal{C} e \mathcal{C}' . Allora noi sappiamo che

$$x = Bx', \quad y = Ax, \quad y' = A'x', \quad y = Dy'.$$

Quindi

$$A'x' = y' = D^{-1}y = D^{-1}Ax = D^{-1}ABx';$$

siccome questo deve valere per ogni $x' \in \mathbb{R}^n$, otteniamo di nuovo (8.4).

ESEMPIO 8.10 Verifichiamo la teoria nell'esempio che abbiamo più volte visto in questo paragrafo (vedi gli Esempi 8.6, 8.8 e 8.9). Con le notazioni introdotte abbiamo

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{pmatrix}, \quad A' = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 2 & 3 \\ 0 & 2 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix};$$

è facile quindi verificare che $A' = D^{-1}AB$.

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

ESEMPIO 8.11 Rivediamo ora l'Esempio 8.7. In questo caso \mathcal{B} e \mathcal{C} sono le basi canoniche di \mathbb{R}^3 ed \mathbb{R}^2 , per cui

$$A = \begin{pmatrix} 2 & 0 & 2 \\ 1 & -1 & 0 \end{pmatrix}, \quad A' = \begin{pmatrix} 10 & 4 & -2 \\ -3 & -2 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & -1 \end{pmatrix}, \quad D = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix},$$

e di nuovo $A' = D^{-1}AB$.

Un caso particolarmente interessante è $V = W$, ovvero quando $T: V \rightarrow V$ è un endomorfismo. In tal caso, tanto vale prendere la stessa base \mathcal{B} in partenza e in arrivo, per cui si parlerà di *matrice associata a T* (o che *rappresenta T*) *rispetto alla base \mathcal{B}* . In particolare, se A è associata a T rispetto alla base \mathcal{B} , A' rispetto alla base \mathcal{B}' , e B è la matrice di cambiamento di base da \mathcal{B} a \mathcal{B}' , abbiamo

$$A' = B^{-1}AB. \quad (8.5)$$

Definizione 8.3 Due matrici quadrate $A, A' \in M_{n,n}(\mathbb{R})$ si dicono *simili* se esiste una matrice $B \in GL_n(\mathbb{R})$ tale che $A' = B^{-1}AB$ (e quindi in tal caso si ha anche $A = BA'B^{-1} = (B^{-1})^{-1}A'B^{-1}$).

Proposizione 8.2 Sia $T: V \rightarrow V$ un endomorfismo, \mathcal{B} una base di V e A la matrice associata a T rispetto a \mathcal{B} . Allora:

- (i) se \mathcal{B}' è un'altra base di V , la matrice A' associata a T rispetto a \mathcal{B}' è simile ad A ;
- (ii) viceversa, se A' è una matrice simile ad A , allora esiste una base \mathcal{B}' di V tale che A' rappresenta T rispetto a \mathcal{B}' .

Dimostrazione. (i) È la (8.5).

(ii) Sia $\mathcal{B} = \{v_1, \dots, v_n\}$ e $B \in GL_n(\mathbb{R})$ tale che $A' = B^{-1}AB$. Definiamo $\mathcal{B}' = \{v'_1, \dots, v'_n\}$ tramite la

$$|v'_1 \cdots v'_n| = |v_1 \cdots v_n|B,$$

ovvero

$$\begin{cases} v'_1 = b_{11}v_1 + \cdots + b_{n1}v_n, \\ \vdots \\ v'_n = b_{1n}v_1 + \cdots + b_{nn}v_n. \end{cases}$$

Se \mathcal{B}' è una base di V ci siamo: infatti in tal caso B è proprio la matrice di cambiamento di base da \mathcal{B} a \mathcal{B}' , e quindi A' rappresenta T rispetto a \mathcal{B}' .

Per dimostrare che \mathcal{B}' è una base, basta dimostrare che v'_1, \dots, v'_n sono linearmente indipendenti. Ma infatti da

$$0 = \alpha_1 v'_1 + \cdots + \alpha_n v'_n = |v'_1 \cdots v'_n| \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = |v_1 \cdots v_n| B \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

deduciamo

$$B \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = 0,$$

in quanto v_1, \dots, v_n sono linearmente indipendenti, e quindi

$$\alpha_1 = \cdots = \alpha_n = 0,$$

perché B è invertibile. \square

Definizione 8.4 La classe di similitudine \mathcal{O}_A di una matrice $A \in M_{n,n}(\mathbb{R})$ è

$$\mathcal{O}_A = \{B^{-1}AB \mid B \in GL_n(\mathbb{R})\}.$$

Dunque a ogni endomorfismo $T: V \rightarrow V$ possiamo associare in maniera unica una classe di similitudine che contiene tutte e sole le matrici che possono rappresentare T rispetto a una qualche base di V . I Capitoli 14 e 15 saranno dedicati, fra le altre cose, al problema di trovare in questa classe di similitudine una matrice particolarmente semplice (per esempio diagonale), da cui sia molto facile leggere le proprietà di T .

Infine, nota che due matrici $A, A' \in M_{n,n}(\mathbb{R})$ sono simili se e solo se il sistema lineare omogeneo $BA' - AB = 0$, dove le incognite sono le n^2 componenti della matrice B , ammette una soluzione in cui B risulta invertibile (perché?).

Esercizi

8.1 Sia $T: V \rightarrow W$ un'applicazione lineare, e U_1, U_2 due sottospazi di V . Dimostra che $T(U_1 + U_2) = T(U_1) + T(U_2)$.

8.2 Considera i tre vettori

$$v_1 = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix}.$$

Dimostra che $\mathcal{B} = \{v_1, v_2, v_3\}$ è una base di \mathbb{R}^3 , e trova la matrice di cambiamento di base dalla base \mathcal{B} alla base canonica.

8.3 Considera i polinomi $p_1(t) = t^2 - 2t$, $p_2(t) = 1 + 2t$, $p_3(t) = 2 - t^2$, $q_1(t) = -1 + t$, $q_2(t) = -1 + t - t^2$ e $q_3(t) = 2t + 2t^2$. Dimostra che $\mathcal{B} = \{p_1, p_2, p_3\}$ e $\mathcal{C} = \{q_1, q_2, q_3\}$ sono basi di $\mathbb{R}_2[t]$, e trova la matrice di cambiamento di base da \mathcal{B} a \mathcal{C} .

8.4 Considera i polinomi $p_1(t) = 3$, $p_2(t) = 2 + t^3$, $p_3(t) = t - t^2 - 4t^3$, $p_4(t) = t^2 - t^3$, e $p_5(t) = t + 2t^2$. Estrai, se possibile, da $\{p_1, \dots, p_5\}$ una base di $\mathbb{R}_3[t]$.

8.5 Sia $W \subseteq M_{2,2}(\mathbb{R})$ il sottospazio

$$W = \text{Span} \left(\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \right).$$

Calcola la dimensione di W e trovane una base.

8.6 Considera i sottospazi $U = \text{Span}(p_1, p_2)$ e $W = \text{Span}(q_2, q_3)$ di $\mathbb{R}_2[t]$, dove i polinomi p_1, p_2, q_2 e q_3 sono gli stessi dell'Esercizio 8.3. Calcola dimensione e base di $U + W$ e $U \cap W$.

8.7 Sia $S: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ l'applicazione lineare rappresentata rispetto alle basi canoniche dalla matrice

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

Trova la matrice $A' \in M_{2,3}(\mathbb{R})$ che rappresenta S rispetto alle basi

$$C_1 = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\} \subset \mathbb{R}^3 \quad \text{e} \quad C_2 = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}.$$

8.8 Siano $A, B \in M_{3,3}(\mathbb{R})$ le matrici

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 2 & 3 & 7 \\ 1 & -3 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 0 & 2 \\ -1 & 3 & -1 \\ 1 & -1 & 3 \end{pmatrix}.$$

Scrivi le matrici associate agli endomorfismi $L_A, L_B: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ rispetto alle due basi seguenti:

$$B_1 = \left\{ \begin{pmatrix} -3 \\ 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\}, \quad B_2 = \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

8.9 Sia $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ l'endomorfismo rappresentato rispetto alla base canonica dalla matrice

$$A = \begin{pmatrix} 1 & 3 \\ 3 & 5 \end{pmatrix}.$$

Trova la matrice che rappresenta T rispetto alla base

$$B = \left\{ \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 3 \\ 5 \end{pmatrix} \right\}.$$

8.10 Siano $v_1, v_2, v_3, w_1, w_2, w_3 \in \mathbb{R}^3$ i vettori

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}, \quad w_1 = \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}, \quad w_2 = \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}, \quad w_3 = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}.$$

Dimostra che $B = \{v_1, v_2, v_3\}$ è una base di \mathbb{R}^3 , che esiste un unico endomorfismo $T \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^3)$ tale che $T(v_j) = w_j$ per $j = 1, 2, 3$, e trova la matrice associata a T rispetto alla base B e rispetto alla base canonica.

8.11 Sia $\{E_{11}, E_{12}, E_{21}, E_{22}\}$ la base di $M_{2,2}(\mathbb{R})$ descritta nell'Esercizio 4.18, e sia $T: M_{2,2}(\mathbb{R}) \rightarrow M_{2,2}(\mathbb{R})$ l'endomorfismo dato da

$$T(E_{11}) = T(E_{22}) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad T(E_{12}) = T(E_{21}) = \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix}.$$

Calcola dimensione e base di $\text{Ker } T$ e $\text{Im } T$.

8.12 Sia $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ l'applicazione lineare data da

$$T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x - y \\ y - x \end{pmatrix}.$$

Trova una matrice $A \in M_{2,2}(\mathbb{R})$ che non rappresenta T rispetto ad alcuna base di \mathbb{R}^2 .

8.13 Sia $W \subset \mathbb{R}^4$ il sottospazio vettoriale di equazione cartesiana $x_1 + 2x_2 - x_3 = 0$. Trova una base B di W . Sia $T: W \rightarrow \mathbb{R}^4$ l'applicazione lineare data da

$$T \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 + 2x_3 \\ x_1 - x_3 + x_4 \\ 3x_1 + x_2 + 2x_4 \\ x_2 - x_4 \end{pmatrix}.$$

Dimostra che $\text{Im } T \subseteq W$, per cui possiamo considerare T come un endomorfismo di W ; trova la matrice che rappresenta questo endomorfismo rispetto alla base B .

8.14 Data la matrice

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

considera l'applicazione $T: M_{2,2}(\mathbb{R}) \rightarrow M_{2,2}(\mathbb{R})$ data da $T(X) = AX - XA$. Dimostra che T è lineare, calcola $\text{Ker } T$, $\text{Im } T$ e dimostra che $M_{2,2}(\mathbb{R}) = \text{Ker } T \oplus \text{Im } T$.

8.15 Data la matrice

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

considera l'endomorfismo $T: M_{2,2}(\mathbb{R}) \rightarrow M_{2,2}(\mathbb{R})$ dato da $T(X) = AX - XA$. Calcola il rango di T , e trova dimensione e base di $\text{Ker } T$ e $\text{Im } T$.

8.16 Data la matrice

$$A = \begin{pmatrix} 5 & 1 & 0 \\ 3 & 0 & 1 \\ 4 & -1 & 3 \end{pmatrix},$$

considera l'endomorfismo $T: M_{3,3}(\mathbb{R}) \rightarrow M_{3,3}(\mathbb{R})$ dato da $T(X) = XA$. Trova il rango di T e base e dimensione del nucleo di T .

8.17 Sia $T: \mathbb{R}_2[t] \rightarrow \mathbb{R}_2[t]$ l'applicazione lineare data da $T(p) = tp' - 2p$, dove l'apice indica la derivata del polinomio. Trova la matrice che rappresenta T rispetto alle basi B, C dell'Esercizio 8.3, e rispetto alla base $D = \{1, t, t^2\}$; trova poi dimensione e base di $\text{Ker } T$ e $\text{Im } T$.

8.18 Scopri se le matrici

$$\begin{vmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{vmatrix} \quad \text{e} \quad \begin{vmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{vmatrix}$$

sono simili oppure non lo sono.

8.19 Scopri se le matrici

$$\begin{vmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{vmatrix} \quad \text{e} \quad \begin{vmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 5 & 7 \end{vmatrix}$$

sono simili oppure non lo sono.

8.20 Sia $T: V \rightarrow W$ un'applicazione lineare di rango r . Dimostra che esistono una base \mathcal{B} di V e una base \mathcal{C} di W tali che la matrice associata a T rispetto a queste basi sia della forma

$$\begin{vmatrix} I_r & O \\ O & O \end{vmatrix}.$$

8.21 Trova un endomorfismo $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ tale che la matrice

$$\begin{vmatrix} 1 & 0 \\ 0 & 0 \end{vmatrix}$$

non può essere associata a T rispetto ad alcuna base di \mathbb{R}^2 .

8.22 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale V . Supponiamo che esista una base \mathcal{B} di V tale che la matrice associata a T rispetto a \mathcal{B} sia la matrice identica. Dimostra che $T = \text{id}_V$.

8.23 Sia $P: V \rightarrow V$ una proiezione (nel senso dell'Esercizio 7.7) di rango r . Dimostra che esiste una base di V rispetto a cui P è rappresentata dalla matrice

$$\begin{vmatrix} I_r & O \\ O & O \end{vmatrix}.$$

COMPLEMENTI

8C.1 Duale e biduale di uno spazio vettoriale

Nella Definizione 7.1 abbiamo introdotto il concetto di spazio duale $V' = \mathcal{L}(V, \mathbb{R})$ di uno spazio vettoriale V ; in questi Complementi vogliamo studiarlo un po' meglio. In particolare, vedremo qual è il vero significato della trasposta di una matrice.

Sia V uno spazio vettoriale di dimensione n , e fissiamo una base $\mathcal{B} = \{v_1, \dots, v_n\}$. Come già detto più volte, ogni elemento $v \in V$ si scrive in modo unico come combinazione lineare degli elementi della base, $v = x_1 v_1 + \dots + x_n v_n$. Quindi possiamo definire un'applicazione $v'_i: V \rightarrow \mathbb{R}$ che associa a ogni $v \in V$ la sua i -esima coordinata rispetto alla base \mathcal{B} . In simboli, $v'_i(v) = x_i$ per $i = 1, \dots, n$, o anche

$$F_{\mathcal{B}}(v) = \begin{pmatrix} v'_1(v) \\ \vdots \\ v'_n(v) \end{pmatrix}.$$

In particolare,

$$v'_i(v_j) = \delta_{ij}, \quad (8C.1)$$

dove δ_{ij} è il delta di Kronecker introdotto nell'Osservazione 7.4.

È facile vedere (esercizio) che ogni v'_i è lineare, per cui appartiene allo spazio duale V' . Di più, possiamo dimostrare la seguente

Proposizione 8C.1 Sia V uno spazio vettoriale di dimensione finita, e scegliamo una sua base $\mathcal{B} = \{v_1, \dots, v_n\}$. Allora $\mathcal{B}' = \{v'_1, \dots, v'_n\}$ è una base di V' . In particolare, $\dim V' = \dim V$.

Dimostrazione. Prima di tutto, v'_1, \dots, v'_n sono linearmente indipendenti. Infatti, supponiamo che si abbia

$$\alpha_1 v'_1 + \dots + \alpha_n v'_n = 0,$$

dove qui $O: V \rightarrow \mathbb{R}$ indica l'applicazione nulla. Allora (8C.1) ci dice che

$$0 = O(v_j) = (\alpha_1 v'_1 + \dots + \alpha_n v'_n)(v_j) = \alpha_j$$

per $j = 1, \dots, n$, e l'indipendenza lineare è dimostrata.

Prendiamo ora $\varphi \in V'$, e poniamo $\beta_j = \varphi(v_j)$ per $j = 1, \dots, n$. Allora

$$(\beta_1 v'_1 + \dots + \beta_n v'_n)(v_j) = \beta_j = \varphi(v_j)$$

per $j = 1, \dots, n$; dunque il Corollario 5.3 ci dà

$$\beta_1 v'_1 + \dots + \beta_n v'_n = \varphi,$$

e \mathcal{B}' è anche un sistema di generatori. \square

Definizione 8C.1 La base $\mathcal{B}' = \{v'_1, \dots, v'_n\}$ di V' è detta la base duale della base $\mathcal{B} = \{v_1, \dots, v_n\}$ di V .

ESEMPIO 8C.1 Prendiamo $V = \mathbb{R}_2[t]$. Ci sono alcuni elementi del duale particolarmente facili da definire: le valutazioni. Fissato $t_0 \in \mathbb{R}$ la valutazione in t_0 è semplicemente l'applicazione lineare $\nu_{t_0}: \mathbb{R}[t] \rightarrow \mathbb{R}$ data da

$$\nu_{t_0}(p) = p(t_0) \in \mathbb{R}.$$

Più in generale possiamo usare le derivate, e per $k \geq 0$ definire la k -valutazione $\nu_{t_0}^k$ in t_0 ponendo

$$\nu_{t_0}^k(p) = p^{(k)}(t_0) \in \mathbb{R},$$

dove $p^{(k)}$ indica la k -esima derivata del polinomio p . Per esempio, se $p(t) = 2t^2 - 2$, allora

$$\nu_1(p) = p(1) = 0, \quad \nu_3^1(p) = p'(3) = 4 \cdot 3 = 12, \quad \nu_{-\pi}^2 = p''(-\pi) = 4,$$

e così via. In particolare, se $p(t) = a_0 + a_1t + a_2t^2$, allora

$$a_0 = \nu_0(p), \quad a_1 = \nu_0^1(p) \quad \text{e} \quad a_2 = \frac{1}{2}\nu_0^2(p). \quad (8C.2)$$

Consideriamo la base $\mathcal{B} = \{p_0, p_1, p_2\}$ di V data dai soliti polinomi $p_0 = 1$, $p_1 = t$, e $p_2 = t^2$. Allora (8C.2) ci dice esattamente che la base duale di V' è data da

$$(p_0)' = \nu_0, \quad (p_1)' = \nu_0^1, \quad (p_2)' = \frac{1}{2}\nu_0^2,$$

dove stavolta l'apice indica che si tratta degli elementi della base duale, e non una derivata. Se invece come base di V prendiamo la base $\mathcal{C} = \{q_0, q_1, q_2\}$ dove $q_0 = 1 - t^2$, $q_1 = (t + t^2)/2$ e $q_2 = (t^2 - t)/2$, allora rispetto a questa base i polinomi di V si esprimono

$$a_0 + a_1t + a_2t^2 = a_0q_0 + (a_0 + a_1 + a_2)q_1 + (a_0 - a_1 + a_2)q_2,$$

per cui la base duale stavolta è data da

$$(q_0)' = \nu_0, \quad (q_1)' = \nu_1, \quad (q_2)' = \nu_{-1}.$$

In questo caso le due basi individuano due isomorfismi fra V e V' diversi. Infatti la base \mathcal{B} individua l'isomorfismo

$$a_0 + a_1t + a_2t^2 \mapsto a_0\nu_0 + a_1\nu_0^1 + \frac{1}{2}a_2\nu_0^2,$$

mentre la base \mathcal{C} individua l'isomorfismo

$$a_0 + a_1t + a_2t^2 \mapsto a_0\nu_0 + a_1\nu_1 + a_2\nu_{-1}.$$

Come conseguenza della Proposizione 8C.1, gli spazi V e V' hanno la stessa dimensione (ammesso che V abbia dimensione finita), per cui pure $V'' = (V')'$, il duale del duale, ha la stessa dimensione di V .

Definizione 8C.2 Lo spazio $V'' = (V')'$ è detto il *biduale* di V .

Abbiamo visto che V e V' sono isomorfi, ma l'isomorfismo dipende dalla scelta di una base. Invece V e V'' sono isomorfi in modo *canonico*, ovvero esiste un isomorfismo indipendente da qualunque scelta arbitraria. Infatti, definiamo un'applicazione $\Psi: V \rightarrow V''$ in questo modo: a ogni $v \in V$ associamo l'elemento $\Psi(v) = \psi_v: V' \rightarrow \mathbb{R}$ del biduale tale che $\psi_v(\varphi) = \varphi(v)$ per ogni $\varphi \in V'$ (nota che gli elementi del biduale sono applicazioni da V' in \mathbb{R} , cioè applicazioni che associano a ogni applicazione lineare da V in \mathbb{R} un numero reale).

Proposizione 8C.2 Sia V uno spazio vettoriale di dimensione finita. Allora l'applicazione $\Psi: V \rightarrow V''$ appena definita è un isomorfismo.

Dimostrazione. Verifichiamo prima di tutto che è lineare. Prendiamo $v_1, v_2 \in V$ e $\varphi \in V'$. Allora

$$\begin{aligned} \Psi(v_1 + v_2)(\varphi) &= \psi_{v_1+v_2}(\varphi) = \varphi(v_1 + v_2) \\ &= \varphi(v_1) + \varphi(v_2) = \psi_{v_1}(\varphi) + \psi_{v_2}(\varphi) = [\Psi(v_1) + \Psi(v_2)](\varphi), \end{aligned}$$

cioè $\Psi(v_1 + v_2) = \Psi(v_1) + \Psi(v_2)$ per ogni $v_1, v_2 \in V$. Analogamente si dimostra che Ψ è omogenea, per cui è lineare.

Ora, noi sappiamo già che $\dim V'' = \dim V$; quindi per dimostrare che Ψ è un isomorfismo ci basta far vedere che è iniettiva. Ma infatti, supponiamo che $\Psi(v) = 0$. Fissiamo una base $\mathcal{B} = \{v_1, \dots, v_n\}$ di V , e scriviamo $v = x_1v_1 + \dots + x_nv_n$. Allora

$$0 = \Psi(v)(v'_j) = v'_j(v) = x_j$$

per $j = 1, \dots, n$, e dunque $v = 0$. \square

Osservazione 8C.1 Se V non è di dimensione finita, questa proposizione è in generale falsa. Certo, possiamo ancora definire la $\Psi: V \rightarrow V''$ come prima, e dimostrare che è iniettiva; sfortunatamente, però, se V ha dimensione infinita la Ψ può risultare non surgettiva. La classe degli spazi vettoriali di dimensione infinita in cui Ψ è un isomorfismo (detti *spazi riflessivi*) è molto importante per lo studio delle equazioni differenziali.

A ogni sottospazio di uno spazio vettoriale è possibile associare un sottospazio dello spazio duale (e viceversa):

Definizione 8C.3 Sia S un sottoinsieme di uno spazio vettoriale V . L'*annullatore* di S è il sottoinsieme S° del duale V' dato da

$$S^\circ = \{\varphi \in V' \mid \varphi(v) = 0 \quad \forall v \in S\}.$$

Analogamente, se R è un sottoinsieme di V' il suo *annullatore* ${}^\circ R \subset V$ è il sottoinsieme ${}^\circ R$ di V dato da

$${}^\circ R = \{v \in V \mid \varphi(v) = 0 \quad \forall \varphi \in R\}.$$

Nota che ${}^\circ R$ è l'annullatore di R nel biduale, riletto in V tramite l'isomorfismo Ψ .

Proposizione 8C.3 Sia V uno spazio vettoriale, e V' il suo duale. Allora:

- (i) Se S è un sottoinsieme di V , allora S° è un sottospazio di V' .
- (ii) Se R è un sottoinsieme di V' , allora ${}^\circ R$ è un sottospazio di V .
- (iii) Se S è un sottoinsieme di V , allora $S^\circ = (\text{Span}(S))^\circ$.
- (iv) Se R è un sottoinsieme di V' , allora ${}^\circ R = {}^\circ(\text{Span}(R))$.
- (v) Se S è un sottoinsieme di V , allora ${}^\circ(S^\circ) = \text{Span}(S)$.
- (vi) Se V ha dimensione finita e U è un sottospazio di V , allora

$$\dim U^\circ = \dim V - \dim U.$$

- (vii) Se V ha dimensione finita e W è un sottospazio di V' , allora

$$\dim {}^\circ W = \dim V - \dim W.$$

Dimostrazione. (i) Siano $\varphi_1, \varphi_2 \in S^\circ$. Allora

$$\forall v \in S \quad (\varphi_1 + \varphi_2)(v) = \varphi_1(v) + \varphi_2(v) = 0 + 0 = 0,$$

per cui $\varphi_1 + \varphi_2 \in S^\circ$. Analogamente S° è chiuso rispetto al prodotto per scalari.

- (ii) Come in (i).

(iii) Siccome $S \subseteq \text{Span}(S)$, è chiaro (Esercizio 8C.2) che $(\text{Span}(S))^\circ \subseteq S^\circ$. Viceversa, se $\varphi \in S^\circ$, $v_1, \dots, v_k \in S$ e $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ si ha

$$\varphi(\alpha_1 v_1 + \dots + \alpha_k v_k) = \alpha_1 \varphi(v_1) + \dots + \alpha_k \varphi(v_k) = 0 + \dots + 0 = 0,$$

e quindi $\varphi \in (\text{Span}(S))^\circ$. Ma φ è generico; quindi $S^\circ \subseteq (\text{Span}(S))^\circ$, come richiesto.

- (iv) Come in (iii).

(v) Prendiamo $\varphi \in S^\circ$ e $v \in S$; allora $\varphi(v) = 0$, per cui $v \in {}^\circ(S^\circ)$. Dunque $S \subseteq {}^\circ(S^\circ)$; da (ii) e (iv) otteniamo che $\text{Span}(S) \subseteq {}^\circ(S^\circ)$. Viceversa, prendiamo un elemento $w \in V \setminus \text{Span}(S)$. Per l'Esercizio 8C.3, esiste $\varphi \in \text{Span}(S)^\circ$ tale che $\varphi(w) = 1$; quindi $w \notin {}^\circ(\text{Span}(S)^\circ)$, per cui ${}^\circ(\text{Span}(S)^\circ) \subseteq \text{Span}(S)$, e ci siamo.

- (vi) Sia $\{u_1, \dots, u_k\}$ una base di U ; completiamola a una base

$$\mathcal{B} = \{u_1, \dots, u_k, v_{k+1}, \dots, v_n\}$$

di V , e consideriamo la base duale $\mathcal{B}' = \{u'_1, \dots, u'_k, v'_{k+1}, \dots, v'_n\}$. Ci basta far vedere che $\{v'_{k+1}, \dots, v'_n\}$ è una base di U° . Prima di tutto, ogni elemento di U si scrive come

$$u = x_1 u_1 + \dots + x_k u_k + 0 v_{k+1} + \dots + 0 v_n,$$

per cui $v'_j(u) = 0$ per ogni $u \in U$ e $j = k+1, \dots, n$, cioè $v'_{k+1}, \dots, v'_n \in U^\circ$. Inoltre, se

$$\varphi = \alpha_1 u'_1 + \dots + \alpha_k u'_k + \alpha_{k+1} v'_{k+1} + \dots + \alpha_n v'_n \in U^\circ$$

si deve avere $0 = \varphi(u_i) = \alpha_i$ per $i = 1, \dots, k$. Quindi v'_{k+1}, \dots, v'_n formano anche un sistema di generatori di U° , e ci siamo.

- (vii) Segue subito da (vi) considerando il bidual (esercizio). \square

8C.2 Trasposta di un'applicazione lineare

Vediamo ora che relazioni ci sono fra gli spazi duali e le applicazioni lineari.

Proposizione 8C.4 Sia $T: V \rightarrow W$ un'applicazione lineare. Allora esiste un'unica applicazione lineare $T': W' \rightarrow V'$ tale che

$$\varphi(T(v)) = T'(\varphi)(v) \quad (8C.3)$$

per ogni $v \in V$ e $\varphi \in W'$.

Dimostrazione. Per ogni $\varphi \in W'$ l'applicazione $v \mapsto \varphi(T(v)) \in \mathbb{R}$ è lineare per cui è un elemento univocamente individuato di V' , elemento che chiameremo $T'(\varphi)$. Dobbiamo solo dimostrare che T' è lineare. Ma infatti

$$T'(\lambda\varphi)(v) = (\lambda\varphi)(T(v)) = \lambda[\varphi(T(v))] = \lambda T'(\varphi)(v)$$

per ogni $v \in V$, cioè $T'(\lambda\varphi) = \lambda T'(\varphi)$. Analogamente si dimostra che T' è additiva. \square

Definizione 8C.4 L'applicazione $T': W' \rightarrow V'$ così definita si dice *trasposta* (o *duale*) dell'applicazione lineare $T: V \rightarrow W$.

Osservazione 8C.2 Il motivo del nome è che la matrice associata a T' è esattamente la trasposta della matrice associata a T . Infatti, fissiamo una base \mathcal{B} di V e una base \mathcal{C} di W , e indichiamo con \mathcal{B}' e \mathcal{C}' le basi duali di V' e W' . Prendiamo $\varphi \in W'$, e indichiamo con $a = F_{\mathcal{C}'}(\varphi) \in \mathbb{R}^m$ le sue coordinate rispetto alla base \mathcal{C}' . Se $w \in W$ ha coordinate $y = F_{\mathcal{C}}(w) \in \mathbb{R}^m$, allora è facile verificare (esercizio) che

$$\varphi(w) = \varphi_a(y),$$

dove $\varphi_a: \mathbb{R}^m \rightarrow \mathbb{R}$ è la solita funzione definita nell'Esempio 5.9. Indichiamo con $A \in M_{m,n}(\mathbb{R})$ la matrice che rappresenta T rispetto a \mathcal{B} e \mathcal{C} , e con $A' \in M_{n,m}(\mathbb{R})$ la matrice che rappresenta T' rispetto a \mathcal{C}' e \mathcal{B}' . Se indichiamo con $x = F_{\mathcal{B}'}(v) \in \mathbb{R}^n$ le coordinate di un vettore $v \in V$ allora la (8C.3) è equivalente a

$$\varphi_a(Ax) = \varphi_{A'a}(x).$$

Ma noi conosciamo già una matrice che soddisfa questa uguaglianza: la matrice trasposta A^T (vedi il Lemma 5.10). Siccome l'unica matrice che la può soddisfare è quella che rappresenta T' rispetto a \mathcal{C}' e \mathcal{B}' , ne deduciamo che $A' = A^T$, come voluto.

Mettendo insieme annullatori e applicazione trasposta otteniamo la

Proposizione 8C.5 Sia $T: V \rightarrow W$ un'applicazione lineare fra spazi vettoriali di dimensione finita. Allora si ha

$$\text{Ker } T = {}^\circ(\text{Im } T'), \quad \text{Ker } T' = (\text{Im } T)^\circ, \quad \text{Im } T = {}^\circ(\text{Ker } T'), \quad \text{Im } T' = (\text{Ker } T)^\circ.$$

Dimostrazione. Sia $v \in \text{Ker } T$, e $\varphi = T'(\psi) \in \text{Im } T'$. Allora

$$\varphi(v) = T'(\psi)(v) = \psi(T(v)) = \psi(O) = 0,$$

per cui $\text{Ker } T \subseteq {}^\circ(\text{Im } T')$. Viceversa, se $v \in {}^\circ(\text{Im } T')$ si deve avere

$$0 = T'(\psi)(v) = \psi(T(v))$$

per ogni $\psi \in W'$, e questo implica (Esercizio 8C.1) che $T(v) = O$. Essendo v generico, questo vuol dire che ${}^\circ(\text{Im } T') \subseteq \text{Ker } T$, e la prima uguaglianza è dimostrata.

La seconda si dimostra nello stesso modo (esercizio); vediamo la terza. Prendiamo $w = T(v) \in \text{Im } T$, e $\psi \in \text{Ker } T'$. Allora

$$\psi(w) = \psi(T(v)) = T'(\psi)(v) = O(v) = 0,$$

per cui $\text{Im } T \subseteq {}^\circ(\text{Ker } T')$. Ma ora

$$\begin{aligned} \dim \text{Im } T &= \dim V - \dim \text{Ker } T = \dim V - \dim {}^\circ(\text{Im } T') \\ &= \dim \text{Im } T' = \dim W - \dim \text{Ker } T' = \dim {}^\circ(\text{Ker } T'), \end{aligned} \quad (8C.4)$$

dove abbiamo usato il Teorema della dimensione e la Proposizione 8C.3, per cui anche la terza uguaglianza è dimostrata. La quarta si verifica analogamente, e ci siamo. \square

Osservazione 8C.3 Una conseguenza di (8C.4) è che un'applicazione lineare e la sua trasposta hanno lo stesso rango.

Esercizi

8C.1 Sia V uno spazio vettoriale di dimensione finita. Dimostra che $\varphi(v) = 0$ per ogni $\varphi \in V'$ se e solo se $v = O$.

8C.2 Siano $U_1, U_2 \subseteq V$ sottospazi di uno spazio vettoriale V . Dimostra che se si ha $U_1 \subseteq U_2$ allora $U_2^\circ \subseteq U_1^\circ$; che $(U_1 + U_2)^\circ = U_1^\circ \cap U_2^\circ$; e che $(U_1 \cap U_2)^\circ = U_1^\circ + U_2^\circ$.

8C.3 Sia U un sottospazio vettoriale di uno spazio vettoriale V , e prendiamo un vettore $w \in V \setminus U$. Dimostra che esiste $\varphi \in U^\circ$ tale che $\varphi(w) = 1$. (*Suggerimento:* completa una base di $U \oplus \mathbb{R}w$ a una base di V .)

8C.4 Sia $T: V \rightarrow W$ un'applicazione lineare fra spazi vettoriali di dimensione finita, e consideriamo la sua trasposta $T': W' \rightarrow V'$ e la trasposta della trasposta $T'': V'' \rightarrow W''$. Dimostra che $T'' = \Psi_W \circ T \circ \Psi_V^{-1}$, dove $\Psi_V: V \rightarrow V''$ e $\Psi_W: W \rightarrow W''$ sono gli isomorfismi canonici. In altri termini, se identifichiamo V e W coi loro biduali, allora la trasposta della trasposta coincide con l'applicazione originaria.

Determinanti

In questo capitolo introduciamo uno strumento fondamentale per lo studio dell'Algebra Lineare: il determinante di una matrice quadrata, una funzione che dice esattamente quando le colonne (o le righe) di una matrice quadrata sono linearmente indipendenti. Invece di definirlo direttamente con una formula, elencheremo le proprietà che vorremmo possedesse, e faremo vedere che esiste al più una funzione che soddisfa le nostre richieste. L'esistenza sarà poi dimostrata tramite gli sviluppi di Laplace, uno dei due metodi principali per il calcolo del determinante (l'altro è un'applicazione dell'eliminazione di Gauss). Faremo anche vedere che il determinante del prodotto di matrici è uguale al prodotto dei determinanti (Teorema di Binet); in particolare, due matrici simili hanno sempre lo stesso determinante, per cui possiamo definire il determinante di un endomorfismo come il determinante di una qualunque delle matrici che lo rappresentano. Come conseguenza del Teorema di Binet otterremo anche il Teorema di Cramer sulle soluzioni di un sistema lineare quadrato non singolare. Vedremo anche, nel Teorema degli orlati, come calcolare il rango di matrici qualunque usando il determinante di sottomatrici quadrate. Infine, i Complementi di questo capitolo sono dedicati a un'introduzione all'uso del simbolo di sommatoria.

9.1 Esistenza e unicità

Nell'Esercizio 7.19 abbiamo visto che una matrice

$$A = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \in M_{2,2}(\mathbb{R})$$

è invertibile se e solo se $a_{11}a_{22} - a_{12}a_{21} \neq 0$. Questo vuol dire che le righe A_1 e A_2 (o le colonne A^1 e A^2) di A sono linearmente indipendenti se e solo se $a_{11}a_{22} - a_{12}a_{21} \neq 0$. Abbiamo dunque trovato una funzione $d: \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ tale che i vettori $v, w \in \mathbb{R}^2$ sono linearmente indipendenti se e solo se $d(v, w) \neq 0$: la funzione

$$d \left(\begin{vmatrix} v_1 & w_1 \\ v_2 & w_2 \end{vmatrix} \right) = v_1 w_2 - v_2 w_1.$$

Osservazione 9.1 Esiste un'altra interpretazione della funzione $d: \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$. Prendiamo due vettori $v, w \in \mathbb{R}^2$, di coordinate rispettivamente (v_1, v_2) e (w_1, w_2) , e consideriamo il parallelogramma P di vertici O, v, w e $v+w$. Per semplicità, supponiamo che tutte le coordinate di v e w siano positive (vedi la Figura 9.1).

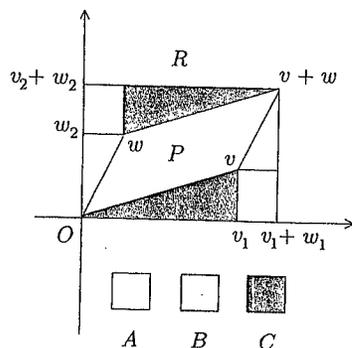


Figura 9.1 Il determinante come area orientata.

Se indichiamo con R il rettangolo di vertici opposti O e $v+w$, vediamo dalla figura che R può venire suddiviso in sette parti: un parallelogramma (P), due rettangoli congruenti (B) e due coppie di triangoli congruenti (A e C). Quindi

$$\text{Area}(P) = \text{Area}(R) - 2 \text{Area}(A) - 2 \text{Area}(B) - 2 \text{Area}(C).$$

Siccome si ha

$$\begin{aligned} \text{Area}(R) &= (v_1 + w_1)(v_2 + w_2), & \text{Area}(A) &= w_1 w_2 / 2, \\ \text{Area}(B) &= v_2 w_1, & \text{Area}(C) &= v_1 v_2 / 2, \end{aligned}$$

ne segue che

$$\text{Area}(P) = v_1 w_2 - v_2 w_1 = d(v, w).$$

La funzione d viene così a coincidere con l'area del parallelogramma generato da v e w . O meglio, con l'area orientata¹ del parallelogramma: $d(v, w) = \text{Area}(P)$ se per andare da v a w si ruota in senso antiorario, come nella Figura 9.1; $d(v, w) = -\text{Area}(P)$ altrimenti (verifica che questo è effettivamente quanto accade). In particolare, ritroviamo nuovamente che d si annulla se e solo se v e w sono linearmente dipendenti: infatti in tal caso il "parallelogramma" da essi generato ha area nulla — è un segmento.

¹ Ripareremo dell'orientazione nel Capitolo 10.

Il principale obiettivo di questo paragrafo è scoprire se esiste qualcosa di analogo in dimensione n , cioè una funzione che ci dica esattamente quando n vettori di \mathbb{R}^n sono linearmente indipendenti².

Per trovare questa funzione (che poi non è altro che il *determinante*), cominciamo col fare un elenco delle proprietà che desideriamo abbia; l'idea è che partendo da queste proprietà saremo poi in grado di calcolare la funzione cercata. Questo procedimento è tipico della matematica: non sapendo se l'oggetto desiderato esiste o meno, si elencano tutte le qualità che si vorrebbe avesse, per poi usarle come mattoni da cui partire per costruire l'oggetto agognato. In questo caso, ovviamente, le proprietà che indicheremo sono state selezionate con la certezza — fornitaci da più di un secolo di esperienza — che ci porteranno alla giusta conclusione; ma altre scelte apparentemente altrettanto valide potrebbero condurre a risultati infausti (vedi l'Esempio 9.4).

Un'osservazione preliminare. Invece di considerare n vettori di \mathbb{R}^n separatamente, possiamo metterli in una matrice A e cercare una funzione $d: M_{n,n}(\mathbb{R}) \rightarrow \mathbb{R}$ tale che $d(A) = 0$ se e solo se le righe³ di A sono linearmente dipendenti. Quindi considereremo d sia come funzione della matrice A che come funzione delle sue n righe A_1, \dots, A_n .

Il caso più eclatante in cui le righe di A sono linearmente dipendenti è quando due righe sono uguali. Quindi la prima proprietà che possiamo desiderare per la nostra funzione è:

$$(A) \quad d(A_1, \dots, A_n) = 0 \quad \text{se } A_i = A_j \text{ per qualche } i \neq j.$$

In altri termini, d si annulla se due righe sono uguali.

Ora, se le n righe A_1, \dots, A_n sono linearmente dipendenti, chiaramente lo sono anche $\lambda A_1, A_2, \dots, A_n$ qualunque sia $\lambda \in \mathbb{R}$. Viceversa, se A_1, \dots, A_n sono linearmente indipendenti anche $\lambda A_1, A_2, \dots, A_n$ lo sono; però, se λ diventa sempre più piccolo in valore assoluto allora i vettori $\lambda A_1, A_2, \dots, A_n$ sono sempre più pericolosamente vicini a essere dipendenti — diventandolo esattamente nel momento in cui λ raggiunge il valore 0. Un modo abbastanza naturale di rappresentare questo comportamento con la nostra funzione d (e del tutto coerente con l'interpretazione di d come area del parallelogramma vista per $n = 2$) è richiedere che $d(\lambda A_1, A_2, \dots, A_n)$ sia uguale a λ volte $d(A_1, \dots, A_n)$. Siccome è evidente che lo stesso ragionamento si poteva applicare a qualsiasi riga, e non solo alla prima, la nostra seconda richiesta allora è

$$(B) \quad d(\dots, \lambda A_i, \dots) = \lambda d(\dots, A_i, \dots) \quad \text{per qualsiasi } \lambda \in \mathbb{R} \text{ e } i = 1, \dots, n.$$

In altri termini, d è omogeneo in ogni riga.

Dunque abbiamo richiesto che d si comporti bene rispetto al prodotto per scalari. E rispetto alla somma? Nei capitoli precedenti abbiamo studiato a lungo le applicazioni lineari; quindi ci farebbe piacere che anche il determinante fosse in qual modo lineare,

² E vedremo anche come usarla per stabilire quando p vettori di \mathbb{R}^n , con $p \leq n$, sono linearmente indipendenti.

³ Utilizzeremo le righe invece delle colonne di A per mantenere una certa compatibilità con l'eliminazione di Gauss.

e non si comportasse invece in strane esotiche maniere. Possiamo quindi provare ad aggiungere alla nostra lista di desideri la seguente proprietà:

$$(C) \quad d(\dots, A'_i + A''_i, \dots) = d(\dots, A'_i, \dots) + d(\dots, A''_i, \dots) \quad \text{per qualsiasi } i = 1, \dots, n.$$

In altri termini, richiediamo che d sia *additivo in ogni riga*. Le proprietà (B) e (C) si possono riassumere dicendo che d è *lineare in ogni riga*.

Vi è una funzione banale che soddisfa automaticamente queste proprietà: la funzione identicamente nulla. Per escludere questo caso possiamo richiedere che d sia non zero su una matrice fissata. Per semplicità (e perché è quanto succede in dimensione 2) esprimiamo ciò aggiungendo alla nostra lista anche la proprietà

$$(D) \quad d(I_n) = 1,$$

cioè chiediamo che d della matrice identica sia uguale a 1.

Osservazione 9.2 Certo, ci potrebbe far piacere che la funzione d , oltre a essere additiva in ogni riga, ci prepari la colazione ogni mattina; sfortunatamente, saremmo costretti a concludere che una tale funzione non esiste. Questo è un rischio che corriamo anche con le nostre più miti richieste: una funzione che soddisfi (A)–(D) potrebbe non esserci (vedi l'Esempio 9.4 e l'Esercizio 9.2). Ma qualche motivo di speranza c'è, in quanto almeno per $n = 2$ una tale funzione l'abbiamo già trovata.

Per capire se stiamo procedendo nella direzione giusta, vediamo che conseguenze si possono trarre dalle proprietà (A)–(C).

Proposizione 9.1 *Supponiamo che la funzione $d: M_{n,n}(\mathbb{R}) \rightarrow \mathbb{R}$ goda delle proprietà (A)–(C). Allora:*

- (i) *Se A ha una riga nulla allora $d(A) = 0$.*
 (ii) *Il valore di d non cambia sommando a una riga un multiplo di un'altra, ovvero*

$$d(\dots, A_i + \lambda A_j, \dots, A_j, \dots) = d(\dots, A_i, \dots, A_j, \dots)$$

per ogni $i \neq j$ e $\lambda \in \mathbb{R}$.

- (iii) *Il valore di d cambia segno se si scambiano due righe qualunque, ovvero*

$$d(\dots, A_i, \dots, A_j, \dots) = -d(\dots, A_j, \dots, A_i, \dots),$$

per ogni $i \neq j$.

- (iv) *Sia S una matrice triangolare superiore ottenuta da A effettuando un'eliminazione di Gauss (o una riduzione a scala) con σ scambi di righe. Allora*

$$d(A) = (-1)^\sigma d(S).$$

- (v) *Se le righe di A sono linearmente dipendenti allora $d(A) = 0$.*

Osservazione 9.3 Il numero $(-1)^\sigma$ è uguale a -1 se σ è dispari, ed è uguale a $+1$ se σ è pari.

Dimostrazione. (i) Basta applicare la proprietà (B) con $\lambda = 0$, dove A_i è la riga nulla.

- (ii) Utilizzando nell'ordine le proprietà (C), (B) e (A) otteniamo

$$\begin{aligned} d(\dots, A_i + \lambda A_j, \dots, A_j, \dots) &= d(\dots, A_i, \dots, A_j, \dots) + d(\dots, \lambda A_j, \dots, A_j, \dots) \\ &= d(\dots, A_i, \dots, A_j, \dots) + \lambda d(\dots, A_j, \dots, A_j, \dots) \\ &= d(\dots, A_i, \dots, A_j, \dots). \end{aligned}$$

- (iii) Questa volta utilizziamo nell'ordine le proprietà (A), (C) e di nuovo (A):

$$\begin{aligned} 0 &= d(\dots, A_i + A_j, \dots, A_i + A_j, \dots) \\ &= d(\dots, A_i, \dots, A_i + A_j, \dots) + d(\dots, A_j, \dots, A_i + A_j, \dots) \\ &= d(\dots, A_i, \dots, A_i, \dots) + d(\dots, A_i, \dots, A_j, \dots) \\ &\quad + d(\dots, A_j, \dots, A_i, \dots) + d(\dots, A_j, \dots, A_j, \dots) \\ &= d(\dots, A_i, \dots, A_j, \dots) + d(\dots, A_j, \dots, A_i, \dots). \end{aligned}$$

(iv) La matrice S viene ottenuta a partire dalla matrice A tramite due tipi di operazioni: lo scambio di due righe, e la sostituzione di una riga A_i con una combinazione lineare del tipo $A_i + \lambda A_j$, con $j \neq i$. Per il punto (iii), a ogni scambio di righe il valore di d viene moltiplicato per -1 (cioè cambiato di segno). La seconda operazione, invece, lascia d immutato, grazie al punto (ii). Quindi $d(S) = \pm d(A)$, dove il segno dipende dal numero di scambi di righe esattamente come asserito.

(v) Se le righe di A sono linearmente dipendenti, qualunque riduzione a scala S di A ha almeno una riga nulla. Per (i), dunque, $d(S) = 0$, e il punto (iv) ci assicura che in tal caso $d(A) = 0$. \square

Osservazione 9.4 La Proposizione 9.1 ci dice che le operazioni elementari usate nell'eliminazione di Gauss (o nella riduzione a scala) possono al massimo cambiare di segno il valore di d . Questo non è più vero se effettuiamo altre operazioni elementari: se alla riga A_i sostituiamo la combinazione lineare $\lambda A_i + \mu A_j$ con $i \neq j$, allora il valore di d viene moltiplicato per λ , grazie alla Proposizione 9.1.(ii) e alla proprietà (B). Questo è il pericolo insito nell'"eliminazione di Gauss generalizzata" di cui si era parlato nell'Osservazione 3.6.

Dunque la nostra funzione d , se esiste, si annulla sulle righe linearmente dipendenti. Non è ancora quanto volevamo; *a priori*, potrebbe annullarsi anche su righe linearmente indipendenti. Sarà la condizione (D) ad assicurarci che questo non accade; ma prima, vediamo se è possibile calcolare d con le informazioni che disponiamo, almeno in alcuni casi semplici.

ESEMPIO 9.1 Sia $d: M_{2,2}(\mathbb{R}) \rightarrow \mathbb{R}$ una funzione soddisfacente le proprietà (A)–(C), e prendiamo una matrice

$$A = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \in M_{2,2}(\mathbb{R}).$$

Le righe di A possono venire scritte usando la base canonica $\{e_1, e_2\}$ di \mathbb{R}^2 come segue:

$$A_1 = a_{11}e_1^T + a_{12}e_2^T \quad \text{e} \quad A_2 = a_{21}e_1^T + a_{22}e_2^T$$

(convincitene prima di proseguire), dove $e_1^T = |1 \ 0|$ ed $e_2^T = |0 \ 1|$. Allora, utilizzando nell'ordine le proprietà (B) e (C) sulla prima riga, le stesse sulla seconda riga, la proprietà (A) e la Proposizione 9.1.(iii), otteniamo

$$\begin{aligned} d(A) &= d(a_{11}e_1^T + a_{12}e_2^T, a_{21}e_1^T + a_{22}e_2^T) \\ &= a_{11}d(e_1^T, a_{21}e_1^T + a_{22}e_2^T) + a_{12}d(e_2^T, a_{21}e_1^T + a_{22}e_2^T) \\ &= a_{11}a_{21}d(e_1^T, e_1^T) + a_{11}a_{22}d(e_1^T, e_2^T) + a_{12}a_{21}d(e_2^T, e_1^T) + a_{12}a_{22}d(e_2^T, e_2^T) \\ &= a_{11}a_{22}d(e_1^T, e_2^T) + a_{12}a_{21}d(e_2^T, e_1^T) \\ &= (a_{11}a_{22} - a_{12}a_{21})d(e_1^T, e_2^T) \\ &= (a_{11}a_{22} - a_{12}a_{21})d(I_2). \end{aligned}$$

In particolare, se vale anche (D) otteniamo

$$d(A) = a_{11}a_{22} - a_{12}a_{21}. \quad (9.1)$$

Abbiamo quindi ricavato una formula per calcolare d su qualunque matrice 2×2 , utilizzando soltanto le proprietà (A)-(D) e le loro conseguenze. In altre parole, per $n = 2$ se una tale funzione d esiste è necessariamente unica, in quanto dev'essere data dalla formula (9.1). Viceversa, la formula (9.1) effettivamente definisce una funzione d che soddisfa le proprietà (A)-(D), come avrai cura di verificare immediatamente (e, del resto, è proprio la funzione con cui avevamo iniziato questo paragrafo). Dunque per $n = 2$ esiste una e una sola funzione che soddisfa le proprietà (A)-(D), il *determinante* delle matrici 2×2 .

Osservazione 9.5 Il procedimento utilizzato per ricavare la formula (9.1) può essere applicato anche per $n > 2$, e condurre così a una dimostrazione dell'esistenza e unicità del determinante in generale tramite quello che viene detto il gruppo delle *permutazioni* (o *sostituzioni*). Sfortunatamente, la formula che si ottiene è piuttosto complicata da scrivere e praticamente inutile per il calcolo effettivo del determinante; quindi qui preferiamo procedere in un altro modo. Per avere un'idea del tipo di fenomeni che si devono affrontare con questo approccio, risolvi l'Esercizio 9.1.

Continuiamo con esempi del calcolo della funzione d su matrici semplici.

ESEMPIO 9.2 Sia

$$A = \begin{vmatrix} a_{11} & & & \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{nn} \end{vmatrix} \in M_{n,n}(\mathbb{R})$$

una matrice diagonale (dove gli spazi bianchi sono da riempirsi di zeri). Allora utilizzando ripetutamente la proprietà (B) otteniamo

$$\begin{aligned} d(A) &= d(a_{11}e_1^T, a_{22}e_2^T, \dots, a_{nn}e_n^T) \\ &= a_{11}d(e_1^T, a_{22}e_2^T, \dots, a_{nn}e_n^T) \\ &= a_{11}a_{22}d(e_1^T, e_2^T, \dots, a_{nn}e_n^T) \\ &\quad \vdots \\ &= a_{11}a_{22} \cdots a_{nn}d(e_1^T, e_2^T, \dots, e_n^T) \\ &= a_{11}a_{22} \cdots a_{nn}d(I_n). \end{aligned}$$

In particolare, se vale anche (D) si ha

$$d(A) = a_{11}a_{22} \cdots a_{nn}.$$

ESEMPIO 9.3 Sia

$$A = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{nn} \end{vmatrix} \in M_{n,n}(\mathbb{R})$$

una matrice triangolare superiore; vogliamo trovare $d(A)$. Scriviamo $A_1 = A'_1 + A''_1$, dove

$$A'_1 = |a_{11} \ 0 \ \cdots \ 0| \quad \text{e} \quad A''_1 = |0 \ a_{12} \ \cdots \ a_{1n}|.$$

Allora

$$\begin{aligned} d(A) &= d(A_1, A_2, \dots, A_n) = d(A'_1, A_2, \dots, A_n) + d(A''_1, A_2, \dots, A_n) \\ &= d(A'_1, A_2, \dots, A_n), \end{aligned}$$

grazie alla Proposizione 9.1.(v), in quanto le righe A''_1, A_2, \dots, A_n sono linearmente dipendenti (perché?). Analogamente possiamo scrivere $A_2 = A'_2 + A''_2$, dove

$$A'_2 = |0 \ a_{22} \ 0 \ \cdots \ 0| \quad \text{e} \quad A''_2 = |0 \ 0 \ a_{23} \ \cdots \ a_{2n}|.$$

Allora

$$\begin{aligned} d(A) &= d(A'_1, A_2, \dots, A_n) = d(A'_1, A'_2, \dots, A_n) + d(A'_1, A''_2, \dots, A_n) \\ &= d(A'_1, A'_2, \dots, A_n), \end{aligned}$$

in quanto di nuovo le righe A'_1, A''_2, \dots, A_n sono linearmente dipendenti (perché?). Procedendo in questo modo arriviamo a

$$d(A) = d(A'_1, A'_2, \dots, A'_n) = a_{11}a_{22} \cdots a_{nn}d(I_n),$$

grazie all'Esempio precedente. In particolare, se vale anche la (D) otteniamo

$$d(A) = d(A'_1, A'_2, \dots, A'_n) = a_{11}a_{22} \cdots a_{nn} :$$

il determinante di una matrice triangolare superiore è uguale al prodotto degli elementi sulla diagonale principale.

Questo è quanto ci basta per calcolare d applicato a una matrice qualunque:

Proposizione 9.2 Supponiamo che la funzione $d: M_{n,n}(\mathbb{R}) \rightarrow \mathbb{R}$ goda delle proprietà (A)-(C). Presa una matrice $A \in M_{n,n}(\mathbb{R})$, siano $p_1, \dots, p_n \in \mathbb{R}$ i pivot ottenuti tramite un'eliminazione di Gauss effettuata con σ scambi di righe. Allora

$$d(A) = (-1)^\sigma p_1 \cdots p_n d(I_n). \quad (9.2)$$

Dimostrazione. Abbiamo già visto che $d(A) = (-1)^\sigma d(S)$, dove S è la matrice triangolare superiore prodotta dall'eliminazione di Gauss considerata; quindi per concludere basta utilizzare l'Esempio 9.3. \square

Corollario 9.3 Sia $d: M_{n,n}(\mathbb{R}) \rightarrow \mathbb{R}$ una funzione che gode delle proprietà (A)-(C), e inoltre tale che $d(I_n) \neq 0$. Allora $d(A) = 0$ se e solo se le righe di A sono linearmente dipendenti.

Dimostrazione. Se le righe di A sono linearmente dipendenti, sappiamo già che si deve avere $d(A) = 0$. Viceversa, se sono linearmente indipendenti il rango di A dev'essere n , per cui i pivot sono tutti non nulli e dunque (9.2) ci dice che $d(A) \neq 0$. \square

Corollario 9.4 Per ogni $n \in \mathbb{N}^*$ esiste al più un'unica funzione $d: M_{n,n}(\mathbb{R}) \rightarrow \mathbb{R}$ che gode delle proprietà (A)-(D).

Dimostrazione. Infatti, se d esiste si deve calcolare tramite (9.2). \square

Osservazione 9.6 La formula (9.2) non basta per dimostrare che la nostra funzione d esiste. Il problema è che, come abbiamo detto più volte, i pivot di una matrice non sono univocamente determinati; quindi chi ci dice che il loro prodotto lo sia? Quello che potrebbe succedere è che effettuando un'eliminazione di Gauss la formula (9.2) dia un valore per $d(A)$, mentre effettuando una diversa eliminazione di Gauss la formula dia un valore *diversa* per $d(A)$. Se ciò accadesse, saremmo costretti a concludere che una tale funzione d non può esistere, in quanto una funzione non può assumere due valori diversi calcolata sullo stesso elemento. In altri termini, se ciò accadesse saremmo costretti a concludere che le proprietà (A)-(D) sono *contraddittorie*: non possono verificarsi tutte contemporaneamente.

ESEMPIO 9.4 Se alle proprietà (A)-(D) avessimo aggiunto la

$$(E) \quad d(\dots, A_i, \dots, A_j, \dots) = d(\dots, A_j, \dots, A_i, \dots) \quad \text{se } i \neq j,$$

(proprietà molto ragionevole, in quanto scambiare due righe non influisce sulla dipendenza o indipendenza lineare), avremmo scoperto che nessuna funzione può soddisfare contemporaneamente tutte le proprietà (A)-(E). Infatti, la Proposizione 9.1.(iii) — che è conseguenza solo delle proprietà (A)-(C) — assieme alla (E) darebbe $d(A) \equiv 0$, contraddicendo platealmente la (D).

Dunque la (9.2) non basta per dimostrare l'esistenza della funzione d . Per superare questo problema, nel prossimo Teorema daremo una formula per calcolare d che non dipende da nessuna scelta arbitraria. Ma prima ci serve una definizione.

Definizione 9.1 Sia $A \in M_{n,n}(\mathbb{R})$. La matrice che si ottiene cancellando la riga i e la colonna j di A si chiama *minore* (i, j) di A , e si indica con $A_{ij} \in M_{n-1, n-1}(\mathbb{R})$.

ESEMPIO 9.5 Sia

$$A = \begin{vmatrix} 1 & 2 & -3 \\ 4 & 6 & 1 \\ 3 & 3 & -2 \end{vmatrix}.$$

Allora

$$A_{11} = \begin{vmatrix} 6 & 1 \\ 3 & -2 \end{vmatrix}, \quad A_{23} = \begin{vmatrix} 1 & 2 \\ 3 & 3 \end{vmatrix},$$

e così via.

Teorema 9.5 Per ogni $n \in \mathbb{N}^*$ esiste un'unica funzione $d_n: M_{n,n}(\mathbb{R}) \rightarrow \mathbb{R}$ che soddisfa le proprietà (A)-(D). Questa funzione è data da $d_1(A) = a_{11}$ se $n = 1$, e da

$$\begin{aligned} d_n(A) &= a_{11}d_{n-1}(A_{11}) - a_{21}d_{n-1}(A_{21}) + \cdots + (-1)^{n+1}a_{n1}d_{n-1}(A_{n1}) \\ &= \sum_{i=1}^n (-1)^{i+1} a_{i1} d_{n-1}(A_{i1}) \end{aligned} \quad (9.3)$$

se $n > 1$.

Osservazione 9.7 Per informazioni sul simbolo di sommatoria \sum vedi i Complementi a questo capitolo.

Osservazione 9.8 La formula (9.3) è un esempio di *formula ricorsiva* (o *induttiva*); la funzione d_n , che dipende dal parametro n , viene calcolata in termini di d_{n-1} , la quale a sua volta viene espressa in termini di d_{n-2} , e così via. Eseguendo un numero sufficiente di passaggi si riconduce il calcolo di d_n al calcolo di d_2 che sappiamo perfettamente come eseguire, grazie all'Esempio 9.1; e quindi, risalendo pazientemente, otteniamo anche un valore unico per d_n . In altri termini, (9.3) definisce effettivamente una funzione $d_n: M_{n,n}(\mathbb{R}) \rightarrow \mathbb{R}$; tutto quello che dovremo fare è verificare che soddisfi le proprietà (A)-(D) qualunque sia $n \in \mathbb{N}^*$.

Dimostrazione. Procediamo per induzione su n . Per $n = 2$ sappiamo già che esiste un'unica funzione $d_2: M_{2,2}(\mathbb{R}) \rightarrow \mathbb{R}$ che soddisfa le proprietà (A)-(D), la funzione

$$d_2(A) = a_{11}a_{22} - a_{12}a_{21}. \quad (9.4)$$

Siccome in questo caso

$$A_{11} = |a_{22}|, \quad A_{12} = |a_{21}|, \quad A_{21} = |a_{12}|, \quad A_{22} = |a_{11}|,$$

ponendo $d_1(|a|) = a$ per ogni matrice $|a| \in M_{1,1}(\mathbb{R})$, vediamo che le formule (9.4) e (9.3) coincidono.

Supponiamo allora che esista un'unica funzione $d_{n-1}: M_{n-1,n-1}(\mathbb{R}) \rightarrow \mathbb{R}$ che soddisfi (A)-(D), e definiamo $d_n: M_{n,n}(\mathbb{R}) \rightarrow \mathbb{R}$ tramite la (9.3). Dobbiamo dimostrare che anche questa d_n soddisfa le proprietà (A)-(D).

Cominciamo dalla più facile, la (D). Come si verifica subito, il minore (1,1) della matrice identica I_n è esattamente la matrice identica I_{n-1} . Inoltre, siccome gli elementi della prima colonna di I_n successivi al primo sono tutti zero, soltanto il primo addendo della formula (9.3) è non nullo, e otteniamo

$$d_n(I_n) = 1 \cdot d_{n-1}(I_{n-1}) + 0 + \dots + 0 = d_{n-1}(I_{n-1}) = 1,$$

perché, per ipotesi induttiva, d_{n-1} soddisfa la proprietà (D).

Passiamo a (C). Supponiamo che la riga in cui avviene la somma sia la riga j_0 . Consideriamo tre matrici:

$$B = \begin{vmatrix} A_1 \\ \vdots \\ A'_{j_0} + A''_{j_0} \\ \vdots \\ A_n \end{vmatrix}, \quad A' = \begin{vmatrix} A_1 \\ \vdots \\ A'_{j_0} \\ \vdots \\ A_n \end{vmatrix}, \quad A'' = \begin{vmatrix} A_1 \\ \vdots \\ A''_{j_0} \\ \vdots \\ A_n \end{vmatrix};$$

vogliamo dimostrare che $d_n(B) = d_n(A') + d_n(A'')$. Per far ciò dobbiamo descrivere gli elementi della prima colonna e i minori $(i, 1)$ di B in termini dei corrispondenti oggetti di A' e A'' . Prima di tutto,

$$b_{i1} = \begin{cases} a_{i1} & \text{se } i \neq j_0; \\ a'_{j_0 1} + a''_{j_0 1} & \text{se } i = j_0. \end{cases}$$

Per i minori, invece, avviene un fenomeno complementare. Infatti il minore $B_{j_0 1}$ è uguale al minore $A'_{j_0 1}$ (che a sua volta è uguale al minore $A''_{j_0 1}$), mentre i minori B_{i1} con $i \neq j_0$ differiscono dai minori A'_{i1} e A''_{i1} solo nella riga proveniente dalla riga j_0 , che è rimasta una somma. Ricordando che, per ipotesi induttiva, d_{n-1} gode della proprietà (C) otteniamo

$$b_{i1} d_{n-1}(B_{i1}) = \begin{cases} a_{i1} [d_{n-1}(A'_{i1}) + d_{n-1}(A''_{i1})] & \text{se } i \neq j_0; \\ [a'_{j_0 1} + a''_{j_0 1}] d_{n-1}(A'_{j_0 1}) = [a'_{j_0 1} + a''_{j_0 1}] d_{n-1}(A''_{j_0 1}) & \text{se } i = j_0. \end{cases}$$

Dunque abbiamo

$$\begin{aligned} d_n(B) &= b_{11} d_{n-1}(B_{11}) + \dots + (-1)^{j_0+1} b_{j_0 1} d_{n-1}(B_{j_0 1}) \\ &\quad + \dots + (-1)^{n+1} b_{n1} d_{n-1}(B_{n1}) \\ &= a_{11} [d_{n-1}(A'_{11}) + d_{n-1}(A''_{11})] + \dots + (-1)^{j_0+1} [a'_{j_0 1} + a''_{j_0 1}] d_{n-1}(A'_{j_0 1}) \\ &\quad + \dots + (-1)^{n+1} a_{n1} [d_{n-1}(A'_{n1}) + d_{n-1}(A''_{n1})] \\ &= [a_{11} d_{n-1}(A'_{11}) + \dots + (-1)^{j_0+1} a'_{j_0 1} d_{n-1}(A'_{j_0 1}) \\ &\quad + \dots + (-1)^{n+1} a_{n1} d_{n-1}(A'_{n1})] \\ &\quad + [a_{11} d_{n-1}(A''_{11}) + \dots + (-1)^{j_0+1} a''_{j_0 1} d_{n-1}(A''_{j_0 1}) \\ &\quad + \dots + (-1)^{n+1} a_{n1} d_{n-1}(A''_{n1})] \\ &= d_n(A') + d_n(A''), \end{aligned}$$

e ci siamo.

La dimostrazione di (B) è analoga. Supponiamo che la moltiplicazione per λ avvenga nella riga i_0 ; questo ci porta a considerare due matrici:

$$A = \begin{vmatrix} A_1 \\ \vdots \\ A_{i_0} \\ \vdots \\ A_n \end{vmatrix}, \quad B = \begin{vmatrix} A_1 \\ \vdots \\ \lambda A_{i_0} \\ \vdots \\ A_n \end{vmatrix};$$

vogliamo dimostrare che $d_n(B) = \lambda d_n(A)$. Per far ciò dobbiamo di nuovo descrivere gli elementi della prima colonna e i minori $(i, 1)$ di B in termini dei corrispondenti oggetti di A . Prima di tutto,

$$b_{i1} = \begin{cases} a_{i1} & \text{se } i \neq i_0; \\ \lambda a_{i_0 1} & \text{se } i = i_0. \end{cases}$$

Per i minori avviene nuovamente un fenomeno complementare. Infatti il minore $B_{i_0 1}$ è uguale al minore $A_{i_0 1}$, mentre i minori B_{i1} con $i \neq i_0$ differiscono dai minori A_{i1} solo nella riga proveniente dalla riga i_0 , che è ancora una riga moltiplicata per λ . Ricordando che, grazie all'ipotesi induttiva, d_{n-1} gode della proprietà (B) otteniamo subito

$$b_{i1} d_{n-1}(B_{i1}) = \begin{cases} a_{i1} d_{n-1}(A_{i1}) = \lambda a_{i_0 1} d_{n-1}(A_{i1}) & \text{se } i \neq i_0; \\ \lambda a_{i_0 1} d_{n-1}(A_{i_0 1}) & \text{se } i = i_0. \end{cases}$$

Dunque

$$\begin{aligned} d_n(B) &= b_{11} d_{n-1}(B_{11}) + \dots + (-1)^{i_0+1} b_{i_0 1} d_{n-1}(B_{i_0 1}) \\ &\quad + \dots + (-1)^{n+1} b_{n1} d_{n-1}(B_{n1}) \\ &= \lambda a_{11} d_{n-1}(A_{11}) + \dots + (-1)^{i_0+1} \lambda a_{i_0 1} d_{n-1}(A_{i_0 1}) \\ &\quad + \dots + (-1)^{n+1} \lambda a_{n1} d_{n-1}(A_{n1}) \\ &= \lambda d_n(A). \end{aligned}$$

e ci siamo.

Rimane da dimostrare la proprietà (A) — e finalmente vedremo a cosa serve l'alternarsi dei segni nella formula (9.3). Prendiamo una matrice A con due righe uguali, la i_0 e la j_0 ; in altri termini, supponiamo $A_{i_0} = A_{j_0}$. Dobbiamo dimostrare che $d_n(A) = 0$. Se $i \neq i_0, j_0$ anche il minore A_{i1} ha due righe uguali; per ipotesi induttiva, $d_{n-1}(A_{i1}) = 0$. Quindi la (9.3) si riduce a

$$d_n(A) = (-1)^{i_0+1} a_{i_01} d_{n-1}(A_{i_01}) + (-1)^{j_0+1} a_{j_01} d_{n-1}(A_{j_01}).$$

Prima di tutto, essendo le righe i_0 e j_0 uguali, abbiamo $a_{i_01} = a_{j_01}$. Inoltre, A_{i_01} e A_{j_01} hanno le stesse righe, disposte in ordine diverso. Per l'esattezza, si ha una situazione di questo genere:

$$A_{i_01} = \begin{vmatrix} \vdots \\ A'_{i_0-1} \\ A'_{i_0+1} \\ \vdots \\ A'_{j_0-1} \\ A'_{j_0} = A'_{i_0} \\ A'_{j_0+1} \\ \vdots \end{vmatrix}, \quad A_{j_01} = \begin{vmatrix} \vdots \\ A'_{i_0-1} \\ A'_{i_0} = A'_{j_0} \\ A'_{i_0+1} \\ \vdots \\ A'_{j_0-1} \\ A'_{j_0+1} \\ \vdots \end{vmatrix},$$

dove A'_i non è altro che la riga A_i di A privata del primo elemento. Dunque, per trasformare A_{j_01} in A_{i_01} basta scambiare A'_{i_0} con, nell'ordine, A'_{i_0+1} , A'_{i_0+2} e così via fino a A'_{j_0-1} . In questo modo effettuiamo $j_0 - i_0 - 1$ scambi di righe (in quanto siamo partiti da $i_0 + 1$ e siamo arrivati a $i_0 + (j_0 - i_0) - 1$). Dunque la Proposizione 9.1.(iii), che per ipotesi induttiva vale per d_{n-1} , ci dà $d_{n-1}(A_{i_01}) = (-1)^{j_0-i_0-1} d_{n-1}(A_{j_01})$, per cui

$$\begin{aligned} d_n(A) &= (-1)^{i_0+1} a_{i_01} d_{n-1}(A_{i_01}) + (-1)^{j_0+1} a_{j_01} d_{n-1}(A_{j_01}) \\ &= (-1)^{i_0+1} (-1)^{j_0-i_0-1} a_{j_01} d_{n-1}(A_{j_01}) + (-1)^{j_0+1} a_{j_01} d_{n-1}(A_{j_01}) \\ &= (-1)^{j_0} (a_{j_01} d_{n-1}(A_{j_01}) - a_{j_01} d_{n-1}(A_{j_01})) = 0. \end{aligned}$$

□

Definizione 9.2 La funzione definita dall'equazione (9.3) si chiama *determinante*, e viene indicata con $\det: M_{n,n}(\mathbb{R}) \rightarrow \mathbb{R}$.

ESEMPIO 9.6 Sia

$$A = \begin{vmatrix} 3 & 4 & 2 & 1 \\ 0 & 7 & 6 & 3 \\ 1 & 1 & 2 & 4 \\ 3 & 1 & -2 & 4 \end{vmatrix} \in M_{4,4}(\mathbb{R});$$

vogliamo calcolare il determinante $\det A$. Utilizzando la formula (9.3) si ha

$$\begin{aligned} \det(A) &= 3 \det \begin{vmatrix} 7 & 6 & 3 \\ 1 & 2 & 4 \\ 1 & -2 & 4 \end{vmatrix} - 0 \det \begin{vmatrix} 4 & 2 & 1 \\ 1 & 2 & 4 \\ 1 & -2 & 4 \end{vmatrix} + \det \begin{vmatrix} 4 & 2 & 1 \\ 7 & 6 & 3 \\ 1 & -2 & 4 \end{vmatrix} - 3 \det \begin{vmatrix} 4 & 2 & 1 \\ 7 & 6 & 3 \\ 1 & 2 & 4 \end{vmatrix} \\ &= 3 \left[7 \det \begin{vmatrix} 2 & 4 \\ -2 & 4 \end{vmatrix} - \det \begin{vmatrix} 6 & 3 \\ -2 & 4 \end{vmatrix} + \det \begin{vmatrix} 6 & 3 \\ 2 & 4 \end{vmatrix} \right] \\ &\quad + \left[4 \det \begin{vmatrix} 6 & 3 \\ -2 & 4 \end{vmatrix} - 7 \det \begin{vmatrix} 2 & 1 \\ -2 & 4 \end{vmatrix} + \det \begin{vmatrix} 2 & 1 \\ 6 & 3 \end{vmatrix} \right] \\ &\quad - 3 \left[4 \det \begin{vmatrix} 6 & 3 \\ 2 & 4 \end{vmatrix} - 7 \det \begin{vmatrix} 2 & 1 \\ 2 & 4 \end{vmatrix} + \det \begin{vmatrix} 2 & 1 \\ 6 & 3 \end{vmatrix} \right] \\ &= 3[7 \cdot 16 - 30 + 18] + [4 \cdot 30 - 7 \cdot 10 + 0] - 3[4 \cdot 18 - 7 \cdot 6 + 0] \\ &= 300 + 50 - 90 = 260. \end{aligned}$$

Altrimenti, possiamo effettuare una eliminazione di Gauss (con ulteriori scambi di righe se necessario) su A e applicare la Proposizione 9.3. In tal caso otteniamo

$$\begin{aligned} \begin{vmatrix} 3 & 4 & 2 & 1 \\ 0 & 7 & 6 & 3 \\ 1 & 1 & 2 & 4 \\ 3 & 1 & -2 & 4 \end{vmatrix} &\rightarrow \begin{vmatrix} 1 & 1 & 2 & 4 \\ 0 & 7 & 6 & 3 \\ 3 & 4 & 2 & 1 \\ 3 & 1 & -2 & 4 \end{vmatrix} \rightarrow \begin{vmatrix} 1 & 1 & 2 & 4 \\ 0 & 7 & 6 & 3 \\ 0 & 1 & -4 & -11 \\ 0 & -2 & -8 & -8 \end{vmatrix} \\ &\rightarrow \begin{vmatrix} 1 & 1 & 2 & 4 \\ 0 & 1 & -4 & -11 \\ 0 & 7 & 6 & 3 \\ 0 & -2 & -8 & -8 \end{vmatrix} \rightarrow \begin{vmatrix} 1 & 1 & 2 & 4 \\ 0 & 1 & -4 & -11 \\ 0 & 0 & 34 & 80 \\ 0 & 0 & -16 & -30 \end{vmatrix} \rightarrow \begin{vmatrix} 1 & 1 & 2 & 4 \\ 0 & 1 & -4 & -11 \\ 0 & 0 & 34 & 80 \\ 0 & 0 & 0 & 130/17 \end{vmatrix} \end{aligned}$$

Abbiamo effettuato due scambi di righe, per cui di nuovo

$$\det(A) = (-1)^2 1 \cdot 1 \cdot 34 \cdot (130/17) = 260.$$

9.2 Sviluppi di Laplace

Dunque abbiamo dimostrato l'esistenza e l'unicità della funzione determinante, e abbiamo dato due modi per calcolarla: tramite l'eliminazione di Gauss, o tramite la formula (9.3). Vediamo ora qualche conseguenza di quanto ottenuto finora.

Prima di tutto, è chiaro (esercizio) che la dimostrazione del Teorema 9.5 può essere adattata in modo da dimostrare anche che

$$\det(A) = (-1)^{1+j} a_{1j} \det(A_{1j}) + \dots + (-1)^{n+j} a_{nj} \det(A_{nj}) = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det(A_{ij})$$

qualunque sia $1 \leq j \leq n$. Questa formula si chiama *sviluppo di Laplace del determinante lungo la colonna j -esima*. Vi è uno sviluppo analogo per righe:

Proposizione 9.6 Sia $A \in M_{n,n}(\mathbb{R})$ e $1 \leq i \leq n$. Allora

$$\det(A) = (-1)^{i+1} a_{i1} \det(A_{i1}) + \cdots + (-1)^{i+n} a_{in} \det(A_{in}) = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det(A_{ij}),$$

(sviluppo di Laplace del determinante lungo la riga i -esima).

Dimostrazione. Sia $A_i = a_{i1}e_1^T + \cdots + a_{in}e_n^T$ la riga i -esima. Allora

$$\begin{aligned} \det(A) &= \det(A_1, \dots, A_i, \dots, A_n) \\ &= a_{i1} \det(A_1, \dots, e_1^T, \dots, A_n) + \cdots + a_{in} \det(A_1, \dots, e_n^T, \dots, A_n) \\ &= a_{i1} \det(B_1) + \cdots + a_{in} \det(B_n), \end{aligned} \quad (9.5)$$

dove

$$B_j = \begin{pmatrix} A_1 \\ \vdots \\ A_{i-1} \\ e_j^T \\ A_{i+1} \\ \vdots \\ A_n \end{pmatrix} \in M_{n,n}(\mathbb{R}).$$

Ora, se nella matrice B_j sottraiamo opportuni multipli della i -esima riga (che è e_j^T) a tutte le altre righe, otteniamo una matrice C_j di uguale determinante e la cui j -esima colonna è e_i . Per l'esattezza,

$$C_j = \begin{array}{c|ccc|ccc} a_{11} & \cdots & a_{1,j-1} & 0 & a_{1,j+1} & \cdots & a_{1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{i-1,1} & \cdots & a_{i-1,j-1} & 0 & a_{i-1,j+1} & \cdots & a_{i-1,n} \\ \hline 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \hline a_{i+1,1} & \cdots & a_{i+1,j-1} & 0 & a_{i+1,j+1} & \cdots & a_{i+1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{n,j-1} & 0 & a_{n,j+1} & \cdots & a_{nn} \end{array}.$$

Calcolando ora $\det(C_j)$ con lo sviluppo di Laplace lungo la colonna j -esima, solo un addendo è diverso da zero, per cui

$$\det(B_j) = \det(C_j) = (-1)^{i+j} \det(A_{ij}). \quad (9.6)$$

Mettendo insieme la (9.5) e la (9.6) abbiamo la tesi. \square

Dunque per calcolare il determinante possiamo sviluppare sia per righe che per colonne. Del resto, avevamo già visto che il numero di righe linearmente indipendenti di una matrice è uguale al numero di colonne linearmente indipendenti; quindi

è ragionevole che una funzione creata per misurare quanto le righe sono linearmente indipendenti dia risultati analoghi anche per le colonne. Anzi, i risultati sono esattamente uguali:

Corollario 9.7 Sia $A \in M_{n,n}(\mathbb{R})$. Allora

$$\det(A^T) = \det(A).$$

Dimostrazione. Procediamo per induzione su n . Per $n = 2$ basta guardare la formula. Supponiamo sia vero per $n-1$; per dimostrarlo per n , basta osservare (esercizio) che la trasposta del minore A_{1i} è esattamente il minore $(i, 1)$ della trasposta A^T , e che l'elemento di posto $(i, 1)$ di A^T è a_{1i} . Allora se calcoliamo il determinante di A^T sviluppando lungo la prima colonna si ha

$$\begin{aligned} \det A^T &= \sum_{i=1}^n (-1)^{i+1} a_{1i} \det((A^T)_{i1}) = \sum_{i=1}^n (-1)^{i+1} a_{1i} \det((A_{1i})^T) \\ &= \sum_{i=1}^n (-1)^{i+1} a_{1i} \det(A_{1i}) = \det A, \end{aligned}$$

grazie all'ipotesi induttiva e alla Proposizione 9.6. \square

Osservazione 9.9 Quale metodo conviene utilizzare per calcolare un determinante dipende dalla situazione. In generale (cioè per matrici senza particolari caratteristiche) l'eliminazione di Gauss è il metodo più efficiente, quello che richiede il minor numero di conti. Altre volte, invece, può convenire usare gli sviluppi di Laplace. Per esempio questo è il caso per matrici con molti zeri (le cosiddette *matrici sparse*): una scelta saggia delle righe o colonne lungo cui sviluppare può ridurre notevolmente il numero dei conti necessari.

ESEMPIO 9.7 Vogliamo il determinante di

$$A = \begin{pmatrix} 0 & 0 & 0 & 27 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 4 & 1 \\ 2 & 0 & 1 & 0 & \sqrt{2} \end{pmatrix}.$$

La terza colonna ha un solo termine non nullo, al posto $(5, 3)$; quindi sviluppando lungo la terza colonna otteniamo

$$\det A = (-1)^{5+3} \det \begin{pmatrix} 0 & 0 & 27 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 3 & 4 & 1 \end{pmatrix}.$$

A questo punto possiamo sviluppare lungo la prima colonna (ma anche lungo la quarta, o la seconda riga, o la terza riga) ottenendo

$$\det A = -\det \begin{vmatrix} 0 & 27 & 0 \\ -1 & 0 & 0 \\ 3 & 4 & 1 \end{vmatrix}.$$

Infine, uno sviluppo lungo la prima riga fornisce

$$\det A = -(-1)^{1+2} 27 \det \begin{vmatrix} -1 & 0 \\ 3 & 1 \end{vmatrix} = -27,$$

risultato ottenuto essenzialmente senza aver dovuto fare alcun conto.

ESEMPIO 9.8 Con un po' di allenamento s'impara anche a semplificare il calcolo del determinante con qualche saggia applicazione della Proposizione 9.1.(ii) e dell'Esercizio 9.4. L'idea è di creare più zeri possibile, in modo da semplificare drasticamente gli sviluppi di Laplace. Prendiamo per esempio

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

Sottraendo la terza colonna dalla prima otteniamo

$$\det A = \det \begin{vmatrix} 3 & 3 & 0 & 2 \\ 0 & 0 & 1 & 0 \\ -3 & -1 & -1 & 0 \\ 3 & 3 & 0 & 4 \end{vmatrix}.$$

A questo punto possiamo sviluppare lungo la seconda riga ottenendo

$$\det A = -\det \begin{vmatrix} 3 & 3 & 2 \\ -3 & -1 & 0 \\ 3 & 3 & 4 \end{vmatrix}.$$

Sommando la seconda riga alla prima e alla terza otteniamo

$$\det A = -\det \begin{vmatrix} 0 & 2 & 2 \\ -3 & -1 & 0 \\ 0 & 2 & 4 \end{vmatrix},$$

e quindi, sviluppando lungo la prima colonna,

$$\det A = -3 \det \begin{vmatrix} 2 & 2 \\ 2 & 4 \end{vmatrix} = -12.$$

Ovviamente, questo era solo uno dei possibili modi di procedere; tentane altri, per verificare che otterrai sempre $\det A = -12$.

Osservazione 9.10 Un buon modo per ricordarsi i segni negli sviluppi di Laplace è di visualizzarli in una scacchiera infinita:

$$\begin{vmatrix} + & - & + & - & \dots \\ - & + & - & + & \dots \\ + & - & + & - & \dots \\ - & + & - & + & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}.$$

Quindi partendo dall'angolo in alto a sinistra (segno +) si procede in orizzontale o verticale di un elemento alla volta, cambiando segno a ogni passo, fino a giungere all'elemento (e quindi al segno) corrispondente al minore desiderato.

Un'altra conseguenza dell'esistenza e unicità del determinante è il fatto già più volte annunciato che il prodotto dei pivot di una matrice quadrata è, a meno del segno, indipendente dall'eliminazione di Gauss usata:

Corollario 9.8 Siano $p_1, \dots, p_n \in \mathbb{R}$ i pivot ottenuti tramite un'eliminazione di Gauss di una matrice quadrata $A \in M_{n,n}(\mathbb{R})$. Allora

$$|p_1 \cdots p_n| = |\det A|.$$

In particolare, il modulo del prodotto dei pivot non dipende dall'eliminazione di Gauss effettuata, ma solo dalla matrice A .

Dimostrazione. Segue subito dall'esistenza e unicità del determinante e dalla Proposizione 9.2. \square

Osservazione 9.11 Abbiamo visto nell'Osservazione 9.1 che il determinante delle matrici 2×2 misura l'area orientata del parallelogramma generato dalle righe della matrice. Quando (nel Capitolo 13) parleremo del prodotto vettore in \mathbb{R}^3 vedremo che il determinante delle matrici 3×3 misura il volume orientato del solido (un parallelepipedo inclinato) generato dalle righe della matrice. In realtà, questo accade in ogni dimensione: con le appropriate definizioni (che non daremo) si può dimostrare che il determinante delle matrici $n \times n$ misura il volume n -dimensionale del "solido" n -dimensionale generato dalle righe della matrice. Per questo a volte si dice che il determinante è l'elemento di volume n -dimensionale.

9.3 Teorema di Binet

Abbiamo visto che il determinante è lineare in ciascuna riga e dunque, grazie al Corollario 9.7, in ciascuna colonna (perché?). Viene quindi naturale chiedersi come si comporta rispetto alle operazioni proprie dello spazio delle matrici $M_{n,n}(\mathbb{R})$.

Rispetto alla somma di matrici si comporta molto male, come illustrato negli Esercizi 9.11 e 9.12: il determinante della somma di matrici ha poco a che vedere con la somma dei determinanti. Le cose migliorano rispetto al prodotto per scalari, in

quanto (Esercizio 9.13) si ha $\det(\lambda A) = \lambda^n \det A$ per ogni matrice $A \in M_{n,n}(\mathbb{R})$. Ma il determinante rivela tutta la sua potenza in connessione con il prodotto di matrici, come mostra il fondamentale *Teorema di Binet*:

Teorema 9.9 (Binet) *Se $A, B \in M_{n,n}(\mathbb{R})$ allora*

$$\det(AB) = (\det A)(\det B). \quad (9.7)$$

Dimostrazione. Cominciamo supponendo che $\det B = 0$; in particolare, grazie al Corollario 9.3, le righe (e quindi le colonne) di B sono linearmente dipendenti. Allora esiste un vettore $v \in \mathbb{R}^n$ non nullo tale che $Bv = 0$; quindi $(AB)v = 0$, cioè anche la matrice AB è singolare. Ne segue che anche le righe della matrice AB sono linearmente dipendenti, per cui $\det(AB) = 0$, e (9.7) in questo caso è verificata⁴.

Supponiamo allora che $\det B \neq 0$. Dunque possiamo considerare la funzione $f: M_{n,n}(\mathbb{R}) \rightarrow \mathbb{R}$ definita da

$$f(A) = \frac{\det(AB)}{\det(B)}.$$

Se dimostriamo che f soddisfa le proprietà (A)-(D), l'unicità del determinante implicherà $f(A) = \det(A)$, e avremo la tesi.

Cominciamo, al solito, col dimostrare che f soddisfa (D). Si ha

$$f(I_n) = \frac{\det(I_n B)}{\det(B)} = \frac{\det(B)}{\det(B)} = 1,$$

e la prima è verificata.

Passiamo a (C). Supponiamo che la riga j_0 -esima di A sia somma di due vettori, cioè che

$$A_{j_0} = A'_{j_0} + A''_{j_0}.$$

Allora la riga j_0 -esima di AB è pure somma di due righe,

$$(AB)_{j_0} = A'_{j_0} B + A''_{j_0} B.$$

Siccome il determinante è additivo rispetto alle righe, $\det(AB)$ si decompone in una somma; quindi altrettanto succede per $f(A)$, e la proprietà (C) è verificata.

La dimostrazione della proprietà (B) è del tutto analoga, per non dire più semplice (esercizio). Rimane la proprietà (A). Se la matrice A ha due righe uguali, anche la matrice AB le ha; quindi $\det(AB) = 0$ e $f(A) = 0$, come dovevamo dimostrare. \square

⁴ In questo ragionamento abbiamo abbondantemente utilizzato il Teorema 7.7.

Corollario 9.10 *Sia $A \in M_{n,n}(\mathbb{R})$. Allora A è invertibile se e solo se $\det(A) \neq 0$, e in tal caso*

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

Dimostrazione. La matrice A è invertibile se e solo se le sue righe sono linearmente indipendenti se e solo se $\det A \neq 0$. Infine il Teorema 9.9 ci dice che

$$\det(A) \det(A^{-1}) = \det(AA^{-1}) = \det(I_n) = 1,$$

per cui $\det(A^{-1}) = (\det A)^{-1}$. \square

Il determinante ci fornisce anche un altro metodo per risolvere un sistema quadrato di matrice non singolare, il *Teorema di Cramer*:

Corollario 9.11 (Cramer) *Sia $Ax = b$ un sistema quadrato di ordine n con matrice non singolare. Allora l'unica soluzione $v = (v_1, \dots, v_n) \in \mathbb{R}^n$ del sistema è data da*

$$\forall i = 1, \dots, n \quad v_i = \frac{\det B_i}{\det A},$$

dove B_i è la matrice ottenuta sostituendo in A alla colonna i -esima la colonna b dei termini noti, cioè

$$B_i = |A^1 \dots A^{i-1} b A^{i+1} \dots A^n|.$$

Dimostrazione. Indichiamo con X_i la matrice ottenuta sostituendo nella matrice identica alla colonna i -esima il vettore v , cioè

$$X_i = |e_1 \dots e_{i-1} v e_{i+1} \dots e_n|.$$

Allora si ha $Av = b$ se e solo se

$$\forall i = 1, \dots, n \quad AX_i = B_i.$$

Prendendo il determinante di entrambi i membri e applicando il Teorema di Binet troviamo

$$\det B_i = \det(AX_i) = (\det A)v_i,$$

in quanto $\det X_i = v_i$, e ci siamo. \square

ESEMPIO 9.9 Vogliamo trovare la soluzione del sistema

$$\begin{cases} x + 2z = 1, \\ -x + 3y + z = 0, \\ x + 7z = 1, \end{cases}$$

applicando il Teorema di Cramer appena dimostrato. Si calcola subito $\det A = 15$, per cui la matrice A è invertibile. Poi

$$B_1 = \begin{vmatrix} 1 & 0 & 2 \\ 0 & 3 & 1 \\ 1 & 0 & 7 \end{vmatrix}, \quad B_2 = \begin{vmatrix} 1 & 1 & 2 \\ -1 & 0 & 1 \\ 1 & 1 & 7 \end{vmatrix}, \quad B_3 = \begin{vmatrix} 1 & 0 & 1 \\ -1 & 3 & 0 \\ 1 & 0 & 1 \end{vmatrix};$$

quindi $\det B_1 = 15$, $\det B_2 = 5$ e $\det B_3 = 0$, e la soluzione è data da

$$v = \begin{vmatrix} 1 \\ 1/3 \\ 0 \end{vmatrix}.$$

Vi è un'altra conseguenza del Teorema di Binet che sarà di fondamentale importanza in seguito.

Corollario 9.12 Se $A \in M_{n,n}(\mathbb{R})$ e $B \in GL_n(\mathbb{R})$ si ha

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

Dimostrazione. Infatti

$$\det(B^{-1}AB) = \det(B^{-1}) \det(A) \det(B) = (\det B)^{-1} (\det B) \det A = \det A. \quad \square$$

Dunque due matrici che appartengono alla stessa classe di similitudine (vedi la Definizione 8.4) hanno lo stesso determinante. Questo ci permette di dare la seguente definizione:

Definizione 9.3 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale V di dimensione n . Allora il *determinante* di T è dato da $\det A$, dove $A \in M_{n,n}(\mathbb{R})$ è la matrice che rappresenta T rispetto a una qualsiasi base di V .

Per la Proposizione 8.2 e il Corollario 9.12 il determinante di un endomorfismo non dipende dalla base scelta, per cui è ben definito. Vedremo più avanti come assieme al rango sia una delle caratteristiche più importanti di un endomorfismo.

9.4 Teorema degli orlati

Avevamo anticipato nella Nota 2 che è possibile utilizzare il determinante per vedere quando p vettori di \mathbb{R}^n (con $p \leq n$) sono linearmente dipendenti. In realtà, possiamo fare qualcosa di più: possiamo usare il determinante per trovare il rango di una matrice qualunque, anche rettangolare. Questo è il contenuto del *Teorema degli orlati*, che enunceremo fra poco.

Definizione 9.4 Sia $A \in M_{m,n}(\mathbb{R})$; una *sottomatrice* (quadrata) di ordine p di A è una matrice quadrata A' ottenuta considerando solo p righe e colonne fissate di A .

Se alla sottomatrice A' aggiungiamo un'altra riga e un'altra colonna di A diremo che stiamo *orlando* la sottomatrice.

ESEMPIO 9.10 Se

$$A = \begin{vmatrix} 1 & 0 & 2 & 4 \\ 0 & 3 & 1 & 4 \\ 1 & 0 & 7 & 3 \end{vmatrix},$$

allora una sottomatrice di ordine 2 ottenuta considerando solo la prima e la seconda riga, e la seconda e la terza colonna, è

$$A' = \begin{vmatrix} 0 & 2 \\ 3 & 1 \end{vmatrix}.$$

Possiamo orlare A' aggiungendo la terza riga e, per esempio, la prima colonna; questo ci fornisce la sottomatrice di ordine 3

$$A'' = \begin{vmatrix} 1 & 0 & 2 \\ 0 & 3 & 1 \\ 1 & 0 & 7 \end{vmatrix}.$$

Teorema 9.13 (degli orlati) Sia $A \in M_{m,n}(\mathbb{R})$ una matrice. Allora $\text{rg } A = r$ se e solo se esiste una sottomatrice A' di ordine r di A non singolare, e tutte le sottomatrici di ordine $r+1$ di A ottenute orlando A' hanno determinante nullo.

Dimostrazione. Supponiamo $\text{rg } A = r$. Allora tutte le $(r+1)$ -uple di colonne di A sono linearmente dipendenti; in particolare, quindi, qualunque sottomatrice di ordine $r+1$ ha determinante nullo. D'altra parte, esiste una sottomatrice quadrata di A di ordine r con determinante non nullo. Infatti, siccome $\text{rg } A = r$ esistono r colonne di A linearmente indipendenti. La matrice formata da queste colonne ha rango r ; quindi deve avere anche r righe linearmente indipendenti. Considerando queste righe e queste colonne otteniamo una sottomatrice quadrata di ordine r e rango r , e quindi di determinante non nullo.

Viceversa, supponiamo che esista una sottomatrice A' di ordine r di A non singolare, e tale che tutte le sottomatrici di ordine $r+1$ ottenute orlando A' abbiano determinante nullo. Supponiamo, per assurdo, che il rango di A non sia r : vogliamo giungere a una contraddizione. Prima di tutto, l'esistenza di una sottomatrice di ordine r non singolare ci assicura che A contiene almeno r colonne linearmente indipendenti. In particolare, il rango di A non può essere inferiore a r ; dovendo essere diverso da r , è per lo meno uguale a $r+1$.

Siccome $\text{rg } A \geq r+1$, possiamo trovare una $r+1$ -esima colonna da aggiungere alle r colonne di A' in modo da ottenere ancora un insieme di colonne linearmente indipendenti. La matrice formata da queste colonne ha rango $r+1$; quindi possiamo trovare una riga che aggiunta alle r righe di A' formi un insieme di righe linearmente indipendenti. In questo modo abbiamo identificato una sottomatrice non singolare di ordine $r+1$ ottenuta orlando A' , contro l'ipotesi. La contraddizione deriva dall'aver supposto $\text{rg } A \neq r$; quindi si deve necessariamente avere $\text{rg } A = r$, e abbiamo finito. \square

ESEMPIO 9.11 Vogliamo trovare il rango della matrice

$$A = \begin{vmatrix} 1 & 0 & 1 & 1 \\ 3 & 1 & -4 & 2 \\ 0 & 7 & 7 & -7 \\ 2 & 1 & 3 & 1 \\ 7 & 2 & 9 & 5 \end{vmatrix}.$$

Siccome non è la matrice nulla, il rango è almeno 1; considerando la prima riga e la prima colonna, troviamo una sottomatrice di ordine 1 non singolare, $\begin{vmatrix} 1 \\ 1 \end{vmatrix}$. La sottomatrice di ordine 2 nell'angolo in alto a sinistra ha determinante

$$\det \begin{vmatrix} 1 & 0 \\ 3 & 1 \end{vmatrix} = 1 \neq 0;$$

quindi il rango è almeno 2. La sottomatrice di ordine 3 nell'angolo in alto a sinistra ha determinante

$$\det \begin{vmatrix} 1 & 0 & 1 \\ 3 & 1 & -4 \\ 0 & 7 & 7 \end{vmatrix} = 56 \neq 0,$$

per cui il rango è almeno 3. Per stabilire se il rango è 3 oppure 4, dobbiamo orlare questa sottomatrice e calcolare i determinanti degli orlati. Ci sono solo due modi per orlarla: aggiungendo la quarta riga e la quarta colonna, o aggiungendo la quinta riga e la quarta colonna. Nel primo caso otteniamo

$$\begin{aligned} \det \begin{vmatrix} 1 & 0 & 1 & 1 \\ 3 & 1 & -4 & 2 \\ 0 & 7 & 7 & -7 \\ 2 & 1 & 3 & 1 \end{vmatrix} &= \det \begin{vmatrix} 1 & 0 & 1 & 2 \\ 3 & 1 & -4 & -2 \\ 0 & 7 & 7 & 0 \\ 2 & 1 & 3 & 4 \end{vmatrix} = \det \begin{vmatrix} 1 & 0 & 1 & 2 \\ 3 & 1 & -5 & -2 \\ 0 & 7 & 0 & 0 \\ 2 & 1 & 2 & 4 \end{vmatrix} \\ &= -7 \det \begin{vmatrix} 1 & 1 & 2 \\ 3 & -5 & -2 \\ 2 & 2 & 4 \end{vmatrix} = 0. \end{aligned}$$

Analogamente,

$$\det \begin{vmatrix} 1 & 0 & 1 & 1 \\ 3 & 1 & -4 & 2 \\ 0 & 7 & 7 & -7 \\ 7 & 2 & 9 & 5 \end{vmatrix} = 0,$$

per cui $\text{rg } A = 3$.

Chiaramente il teorema degli orlati non è il metodo più efficace per calcolare il rango di una matrice (la riduzione a scala è molto più efficiente), ma può comunque essere utile in alcune situazioni particolari (per esempio matrici con molti zeri, in cui è facile calcolare determinanti; ma vedi anche il Capitolo 10).

Esercizi

9.1 Dimostra, utilizzando il metodo dell'Esempio 9.1, che esiste una e una sola funzione $d: M_{3,3}(\mathbb{R}) \rightarrow \mathbb{R}$ che soddisfa le proprietà (A)-(D).

9.2 Dimostra che non esiste alcuna matrice $A \in M_{2,2}(\mathbb{R})$ tale che $A^T = A$, $A^2 = I$, $\det A > 0$ e $|a_{11} + a_{22}| \neq 2$.

9.3 Calcola il determinante delle seguenti matrici:

$$\begin{vmatrix} 0 & 2 & 7 \\ 0 & 0 & \pi \\ 3 & 0 & 0 \end{vmatrix}, \quad \begin{vmatrix} -2 & 1 & -2 \\ 1 & -2 & 1 \\ -2 & 1 & -3 \end{vmatrix}, \quad \begin{vmatrix} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 3 \\ 3 & 3 & 4 & 4 \\ 5 & 6 & 7 & 8 \end{vmatrix}.$$

9.4 Dimostra l'equivalente per le colonne della Proposizione 9.1.(i)-(iii). Più precisamente, dimostra che se una matrice ha una colonna nulla il suo determinante è zero; che scambiando due colonne il determinante cambia di segno; che sommando a una colonna un multiplo di un'altra colonna il valore del determinante non cambia.

9.5 Siano $x, y, z \in \mathbb{R}$ tali che

$$\det \begin{vmatrix} x & y & z \\ 1 & 1 & 2 \\ 0 & 1 & 3 \end{vmatrix} = 1.$$

Calcola i determinanti delle seguenti matrici:

$$\begin{vmatrix} 3x & 3y & 3z \\ 1/2 & 1/2 & 1 \\ 0 & 1 & 3 \end{vmatrix}, \quad \begin{vmatrix} x & y & z \\ 1 & 2 & 5 \\ 2x & 1+2y & 3+2z \end{vmatrix}, \quad \begin{vmatrix} x & y & z \\ 2+2x & 2+2y & 4+2z \\ 1 & 0 & -1 \end{vmatrix}.$$

9.6 Calcola, con più tecniche, i determinanti delle seguenti matrici:

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 0 & 5 \\ 3 & 0 & 1 & -1 \\ 2 & 1 & 0 & 2 \end{vmatrix}, \quad \begin{vmatrix} 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 \end{vmatrix}.$$

9.7 Siano $a, b, c \in \mathbb{R}$. Dimostra che

$$\det \begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{vmatrix} = (b-a)(c-a)(c-b).$$

9.8 Siano $a, b, c \in \mathbb{R}$. Calcola i determinanti delle seguenti matrici:

$$\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^3 & b^3 & c^3 \end{vmatrix}, \quad \begin{vmatrix} 1 & 1 & 1 \\ a^2 & b^2 & c^2 \\ a^4 & b^4 & c^4 \end{vmatrix}.$$

9.9 Siano $a_1, a_2, a_3, a_4 \in \mathbb{R}$. Calcola il determinante della matrice

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ a_1 & a_2 & a_3 & a_4 \\ a_1^2 & a_2^2 & a_3^2 & a_4^2 \\ a_1^3 & a_2^3 & a_3^3 & a_4^3 \end{vmatrix},$$

e trova una generalizzazione per il caso di $n \geq 4$ numeri. (Suggerimento: moltiplica ogni riga per a_i e sottraila dalla successiva.)

9.10 Sia $A \in M_{n,n}(\mathbb{R})$ una matrice della forma

$$A = \begin{vmatrix} B & C \\ O & D \end{vmatrix},$$

con $B \in M_{h,h}(\mathbb{R})$, $C \in M_{h,n-h}(\mathbb{R})$, $D \in M_{n-h,n-h}(\mathbb{R})$ e $h < n$. Dimostra che $\det A = (\det B)(\det D)$. (Suggerimento: procedi per induzione su h .)

9.11 Trova due matrici $A, B \in M_{2,2}(\mathbb{R})$ tali che $\det A + \det B \neq 0 = \det(A + B)$.

9.12 Trova due matrici $C, D \in M_{2,2}(\mathbb{R})$ tali che $\det(C + D) = \det(C) + \det(D)$.

9.13 Se $A \in M_{n,n}(\mathbb{R})$ e $\lambda \in \mathbb{R}$, allora dimostra che $\det(\lambda A) = \lambda^n \det(A)$.

9.14 Sia $A \in M_{n,n}(\mathbb{R})$ una matrice antisimmetrica, cioè tale che $A^T = -A$. Dimostra che se n è dispari allora $\det A = 0$.

9.15 Sia $A \in M_{n,n}(\mathbb{R})$ e $1 \leq i, j \leq n$. Dimostra che

$$\sum_{k=1}^n (-1)^{j+k} a_{ik} \det(A_{jk}) = \begin{cases} 0 & \text{se } i \neq j, \\ \det A & \text{se } i = j. \end{cases}$$

(Suggerimento: se $i \neq j$ è il determinante di una matrice con due righe uguali.)

9.16 Se $A \in M_{n,n}(\mathbb{R})$, la matrice cofattore di A è la matrice $\text{Cof}(A)$ che al posto (i, j) ha il numero $(-1)^{j+i} \det(A_{ji})$. Dimostra che $A \cdot \text{Cof}(A) = (\det A)I_n$. In particolare, se A è invertibile dimostra che $A^{-1} = (\det A)^{-1} \text{Cof}(A)$. (Suggerimento: usa l'Esercizio precedente.)

9.17 Sia $A \in M_{n,n}(\mathbb{R})$. Dimostra che $\det(\text{Cof}(A)) = (\det A)^{n-1}$. (Suggerimento: usa l'Esercizio precedente e il Corollario 9.10.)

9.18 Sia

$$A = \begin{vmatrix} 1 & 1/2 & 1/3 & 1/4 \\ 1/2 & 1/3 & 1/4 & 1/5 \\ 1/3 & 1/4 & 1/5 & 1/6 \\ 1/4 & 1/5 & 1/6 & 1/7 \end{vmatrix}.$$

Dimostra che A è invertibile e calcolane l'inversa.

9.19 Risolvi usando il Teorema di Cramer i seguenti sistemi lineari:

$$\begin{cases} x + 7y + 3z = 2, \\ -x + 2z = -1, \\ 3x + y + z = 1; \end{cases} \quad \begin{cases} x + 2y + 3z = -1, \\ y + 2z = -2, \\ 2x + 3y + 4z = 0. \end{cases}$$

9.20 Risolvi usando il Teorema di Cramer il seguente sistema:

$$\begin{cases} 888445x + 887112y = 1, \\ 887112x + 885781y = 0. \end{cases}$$

(Nota per i patiti delle calcolatrici tascabili tutt'fare: se la tua calcolatrice è in grado di risolvere direttamente i sistemi lineari, prova con questo. I risultati potrebbero sorprenderti...)

9.21 Sia $A \in M_{n,n}(\mathbb{R})$, e sia $T: M_{n,n}(\mathbb{R}) \rightarrow M_{n,n}(\mathbb{R})$ dato da $T(X) = AX - XA$. Dimostra che $\det T = 0$.

9.22 Sia $A \in M_{n,n}(\mathbb{R})$, e sia $T: M_{n,n}(\mathbb{R}) \rightarrow M_{n,n}(\mathbb{R})$ data da $T(X) = AX$. Dimostra che $\det T = (\det A)^n$.

9.23 Sia $T: \mathbb{R}_3[t] \rightarrow \mathbb{R}_3[t]$ l'endomorfismo $T(p) = p'$. Calcola $\det T$.

9.24 Calcola il rango delle seguenti matrici:

$$\begin{vmatrix} 1 & -1 & 0 & -2 \\ 2 & -1 & 1 & -1 \\ 3 & 1 & 4 & 6 \end{vmatrix}, \quad \begin{vmatrix} 0 & -1 & 2 \\ 1 & 2 & -1 \\ 3 & 3 & 3 \\ 4 & 6 & 0 \end{vmatrix}.$$

9.25 Calcola, al variare di $t \in \mathbb{R}$, il rango della seguente matrice

$$\begin{vmatrix} t+1 & 1 & 1 \\ t+2 & t^2+1 & 2 \\ 2t & 1 & 1 \end{vmatrix}.$$

COMPLEMENTI

9C.1 Sommatorie

In questo testo ci troveremo spesso a dover usare somme del tipo $a_1 + \dots + a_n$, dove a_j è un numero (o un vettore, o una matrice, o altro ancora) dipendente da un indice j che di solito è un numero intero variabile in un certo insieme, per esempio da 1 a n . Per trattare questo tipo di somme è stata introdotta una notazione abbreviata molto comoda: la *sommatoria* \sum .

Nella sua versione più semplice, \sum si utilizza come segue:

$$\sum_{j=1}^n a_j = \sum_{1 \leq j \leq n} a_j = a_1 + \cdots + a_n.$$

Ma sono possibili anche costruzioni più complesse, come

$$\sum_{\substack{1 \leq j \leq 2n \\ j \text{ pari}}} a_j = a_0 + a_2 + a_4 + \cdots + a_{2n}.$$

Questa stessa somma poteva anche essere scritta in modo più compatto come

$$\sum_{h=0}^n a_{2h} = \sum_{0 \leq h \leq n} a_{2h}.$$

In questo paragrafo cercheremo di descrivere brevemente come utilizzare al meglio il simbolo di sommatoria, indicandone le principali proprietà e studiando qualche esempio. Cominciamo con la definizione generale del simbolo di sommatoria.

Definizione 9C.1 Se $R(j)$ è una condizione coinvolgente il numero intero j (per esempio “ $-3 \leq j \leq 3$ ”, oppure “ $0 \leq j \leq r$ e j pari”), allora il simbolo

$$\sum_{R(j)} a_j$$

indica la somma di tutti gli a_j per cui l'intero j soddisfa la condizione $R(j)$. Per convenzione, se nessun intero j soddisfa la condizione $R(j)$ si assume che la sommatoria valga zero. A volte scriveremo anche $\sum_{R(j)} a_j$; un'altra possibilità, se $R(j)$ è della forma $\alpha \leq j \leq \beta$, è

$$\sum_{j=\alpha}^{\beta} a_j.$$

L'indice j è chiamato *indice muto* della sommatoria. L'indice muto serve soltanto all'interno del simbolo di sommatoria, e non ha alcun significato intrinseco — vedi la proprietà (B) più oltre. Tipici indici muti sono h, i, j, k, l , talvolta accentati o con indici a loro volta.

Osservazione 9C.1 In questo libro utilizzeremo il simbolo di sommatoria solo per indicare somme finite. In altre parole, la condizione $R(j)$ sarà sempre soddisfatta soltanto da un numero finito di interi j . Il simbolo di sommatoria serve anche per indicare somme infinite (ovvero le *serie*), ma la loro definizione e uso sono argomenti di Analisi.

Le principali proprietà del simbolo di sommatoria sono quattro.

(A) *Proprietà distributiva*, che riguarda il prodotto di sommatorie:

$$\left(\sum_{R(h)} a_h \right) \left(\sum_{S(k)} b_k \right) = \sum_{R(h)} \left(\sum_{S(k)} a_h b_k \right). \quad (9C.1)$$

Per esempio,

$$\begin{aligned} \left(\sum_{h=1}^2 a_h \right) \left(\sum_{k=1}^3 b_k \right) &= (a_1 + a_2)(b_1 + b_2 + b_3) \\ &= (a_1 b_1 + a_1 b_2 + a_1 b_3) + (a_2 b_1 + a_2 b_2 + a_2 b_3) \\ &= \sum_{h=1}^2 \left(\sum_{k=1}^3 a_h b_k \right). \end{aligned}$$

Di solito, le parentesi nelle “somme multiple” come quella sul lato destro di (9C.1) vengono tralasciate: invece di $\sum_{R(h)} \left(\sum_{S(k)} a_{hk} \right)$ si scrive $\sum_{R(h)} \sum_{S(k)} a_{hk}$.

(B) *Cambiamento di variabile*. L'indice muto di una sommatoria è, appunto, muto, per cui non si lamenta se gli viene cambiato il nome. In altri termini, si ha

$$\sum_{R(h)} a_h = \sum_{R(k)} a_k = \sum_{R(s)} a_s, \quad (9C.2)$$

e così via, dove k, s e simili sono tutti indici non altrimenti presenti nella condizione R o nei termini a_k (per intenderci: se $a_k = n^k$, oppure se $R(k)$ è la relazione “ $1 \leq k \leq n$ ”, non possiamo usare n come indice muto). Ma si può fare di più: data una condizione $R(h)$, indichiamo con A l'insieme degli $h \in \mathbb{Z}$ che soddisfano questa condizione. Una *permutazione degli indici* è una funzione $p: \mathbb{Z} \rightarrow \mathbb{Z}$ tale che la restrizione $p|_{p^{-1}(A)}: p^{-1}(A) \rightarrow A$ è bigettiva. In altre parole, per ogni $h \in A$ deve esistere uno e un solo $k \in \mathbb{Z}$ tale che $p(k) = h$. Allora se p è una permutazione degli indici per la condizione $R(h)$ si ha

$$\sum_{R(h)} a_h = \sum_{R(p(k))} a_{p(k)}. \quad (9C.3)$$

Infatti, su entrambi i lati di (9C.3) sono presenti gli stessi addendi. I casi più comuni di permutazione degli indici sono del tipo $p(k) = \pm k \pm c$, dove c è un intero fissato che non dipende da k . Per esempio, se $p(k) = k - 1$ abbiamo

$$\sum_{1 \leq h \leq n} a_h = \sum_{1 \leq k-1 \leq n} a_{k-1} = \sum_{2 \leq k \leq n+1} a_{k-1} = \sum_{2 \leq h \leq n+1} a_{h-1},$$

dove nell'ultimo passaggio abbiamo utilizzato (9C.2).

(C) *Proprietà commutativa.* Quando effettuiamo una somma, l'ordine in cui consideriamo gli addendi è irrilevante. Quindi

$$\sum_{R(h)} \sum_{S(k)} a_{hk} = \sum_{S(k)} \sum_{R(h)} a_{hk}. \quad (9C.4)$$

Per esempio,

$$\sum_{R(h)} \sum_{1 \leq k \leq 2} a_{hk} = \sum_{R(h)} (a_{h1} + a_{h2}) = \sum_{R(h)} a_{h1} + \sum_{R(h)} a_{h2} = \sum_{1 \leq k \leq 2} \sum_{R(h)} a_{hk}.$$

Potrebbe capitare anche un caso più generale, in cui la condizione $S(k)$ dipenda anche da h — e quindi sarà scritta $S(h, k)$. In tal caso la (9C.4) diventa qualcosa del tipo

$$\sum_{R(h)} \sum_{S(h, k)} a_{hk} = \sum_{S'(k)} \sum_{R'(h, k)} a_{hk}, \quad (9C.5)$$

dove $S'(k)$ è la condizione “esiste un intero h tale che sia $R(h)$ che $S(h, k)$ sono verificate”, e $R'(h, k)$ è la condizione “sia $R(h)$ che $S(h, k)$ sono verificate”.

Un esempio per chiarire. Se $R(h)$ è la solita condizione $1 \leq h \leq n$, ed $S(h, k)$ è la condizione $1 \leq k \leq h$ — ovvero stiamo sommando gli elementi in basso a sinistra rispetto alla diagonale principale della matrice (a_{hk}) — allora la condizione $S'(k)$ è “esiste un intero h tale che $1 \leq h \leq n$ e $1 \leq k \leq h$ ”, cioè semplicemente “ $1 \leq k \leq n$ ”, mentre la condizione $R'(h, k)$ è “devono valere sia $1 \leq h \leq n$ che $1 \leq k \leq h$ ”, cioè “ $k \leq h \leq n$ ”. Quindi

$$\sum_{1 \leq h \leq n} \sum_{1 \leq k \leq h} a_{hk} = \sum_{1 \leq k \leq n} \sum_{k \leq h \leq n} a_{hk}. \quad (9C.6)$$

(D) *Manipolazione del dominio.* Se $R(h)$ e $S(h)$ sono due condizioni, abbiamo

$$\sum_{R(h)} a_h + \sum_{S(h)} a_h = \sum_{R(h) \circ S(h)} a_h + \sum_{R(h) \text{ e } S(h)} a_h, \quad (9C.7)$$

dove “ $R(h)$ o $S(h)$ ” vuol dire che h soddisfa $R(h)$ oppure soddisfa $S(h)$, mentre “ $R(h)$ e $S(h)$ ” vuol dire che h soddisfa sia $R(h)$ che $S(h)$. Per esempio, se $m \leq n$ si ha

$$\sum_{1 \leq k \leq m} a_k + \sum_{m \leq k \leq n} a_k = \left(\sum_{1 \leq k \leq n} a_k \right) + a_m.$$

Se poi non capita mai che $R(h)$ ed $S(h)$ siano simultaneamente verificate, allora la seconda sommatoria sul lato destro di (9C.7) semplicemente scompare.

Vediamo ora qualche esempio di applicazione di queste regole.

ESEMPIO 9C.1 *La somma di una progressione aritmetica.* Prendiamo $a, b \in \mathbb{R}$ ed $n \in \mathbb{N}$. Allora:

$$\begin{aligned} a + (a+b) + \cdots + (a+nb) &= \sum_{0 \leq h \leq n} (a+hb) \\ &= \sum_{0 \leq n-h \leq n} (a+(n-h)b) = \sum_{0 \leq n-h \leq n} ((2a+nb) - (a+hb)) \\ &= \sum_{0 \leq h \leq n} (2a+nb) - \sum_{0 \leq h \leq n} (a+hb) \\ &= (n+1)(2a+nb) - \sum_{0 \leq h \leq n} (a+hb), \end{aligned}$$

grazie a: (9C.3) assieme a (9C.2); il fatto che le condizioni “ $0 \leq n-h \leq n$ ” e “ $0 \leq h \leq n$ ” sono verificate dagli stessi interi h ; e il fatto che la prima sommatoria nella penultima riga è semplicemente la somma di $n+1$ addendi che non dipendono da h . Ma ora se confrontiamo la prima e l'ultima riga e dividiamo per 2 otteniamo la formula cercata:

$$\sum_{0 \leq h \leq n} (a+hb) = a(n+1) + \frac{bn(n+1)}{2}.$$

ESEMPIO 9C.2 *La somma di una progressione geometrica.* Prendiamo $a, x \in \mathbb{R}$ con $x \neq 1$, ed $n \in \mathbb{N}$. Allora

$$\begin{aligned} a + ax + \cdots + ax^n &= \sum_{0 \leq k \leq n} ax^k \\ &= a + \sum_{1 \leq k \leq n} ax^k \\ &= a + x \sum_{1 \leq k \leq n} ax^{k-1} \\ &= a + x \sum_{0 \leq k \leq n-1} ax^k \\ &= a + x \sum_{0 \leq k \leq n} ax^k - ax^{n+1}, \end{aligned}$$

grazie a, nell'ordine: (9C.7); un caso molto particolare di (9C.1); (9C.3); (9C.7). Confrontando la prima riga con l'ultima otteniamo

$$(1-x) \sum_{0 \leq k \leq n} ax^k = a(1-x^{n+1}),$$

e quindi la formula cercata:

$$\sum_{0 \leq k \leq n} ax^k = a \frac{1 - x^{n+1}}{1 - x}.$$

ESEMPIO 9C.3 Sia

$$S_1 = \sum_{0 \leq h \leq n} \sum_{0 \leq k \leq h} a_h a_k = \sum_{0 \leq k \leq n} \sum_{k \leq h \leq n} a_h a_k = \sum_{0 \leq h \leq n} \sum_{h \leq k \leq n} a_h a_k,$$

dove abbiamo usato (9C.6) e (9C.2). Se chiamiamo S_2 quest'ultima sommatoria, abbiamo $S_1 = S_2$ e quindi

$$\begin{aligned} 2S_1 &= S_1 + S_2 = \sum_{0 \leq h \leq n} \left(\sum_{0 \leq k \leq h} a_h a_k + \sum_{h \leq k \leq n} a_h a_k \right) \\ &= \sum_{0 \leq h \leq n} \left(\sum_{0 \leq k \leq n} a_h a_k + a_h a_h \right) = \sum_{0 \leq h \leq n} \sum_{0 \leq k \leq n} a_h a_k + \sum_{0 \leq h \leq n} a_h a_h \\ &= \left(\sum_{0 \leq h \leq n} a_h \right) \left(\sum_{0 \leq k \leq n} a_k \right) + \left(\sum_{0 \leq h \leq n} a_h^2 \right) \\ &= \left(\sum_{0 \leq h \leq n} a_h \right)^2 + \left(\sum_{0 \leq h \leq n} a_h^2 \right), \end{aligned}$$

grazie a, nell'ordine, (9C.7), (9C.1), e (9C.2). Abbiamo dunque ricavato la seguente identità:

$$\sum_{0 \leq h \leq n} \sum_{0 \leq k \leq h} a_h a_k = \frac{1}{2} \left\{ \left(\sum_{0 \leq h \leq n} a_h \right)^2 + \left(\sum_{0 \leq h \leq n} a_h^2 \right) \right\}.$$

ESEMPIO 9C.4 Un altro simbolo che si utilizza spesso assieme alle sommatorie è il delta di Kronecker δ_{ij} introdotto nell'Osservazione 7.4. Ricordo che δ_{ij} è dato da

$$\delta_{ij} = \begin{cases} 0 & \text{se } i \neq j, \\ 1 & \text{se } i = j. \end{cases}$$

In particolare si ha

$$\sum_{j=1}^n b_j \delta_{ij} = b_i,$$

e quindi otteniamo formule del tipo

$$\sum_{i=1}^n \sum_{j=1}^n a_i b_j \delta_{ij} = \sum_{i=1}^n a_i \sum_{j=1}^n b_j \delta_{ij} = \sum_{i=1}^n a_i b_i.$$

In questo capitolo torniamo a occuparci della geometria del piano e dello spazio. Vedremo come identificare le posizioni reciproche di punti, rette e piani usando sia equazioni parametriche che equazioni cartesiane: si tratterà semplicemente di applicare in questa situazione particolare la teoria sviluppata finora (in particolare sarà utile il Teorema di Rouché-Capelli). Discuteremo anche i cambiamenti di coordinate affini e introdurremo il concetto di orientazione. Infine, i Complementi a questo capitolo sono dedicati alle affinità e ad alcuni cenni di geometria proiettiva.

10.1 Equazioni di rette e piani

Vogliamo applicare quanto visto finora allo studio della geometria affine del piano e dello spazio. La *geometria affine*¹ si occupa delle posizioni reciproche di rette e piani (e triangoli, e pentagoni, e così via) nello spazio \mathcal{A}^3 e nel piano \mathcal{A}^2 . Il nostro obiettivo è dare criteri operativi per risolvere problemi del genere di "decidere quando due rette sono sghembe", "trovare il piano passante per tre punti non allineati", o "scoprire se quella retta e quel piano sono paralleli".

Fissiamo una volta per tutte un sistema di riferimento affine $RA(O, A_1, A_2, A_3)$ in \mathcal{A}^3 , in modo da poter identificare \mathcal{A}^3 con \mathbb{R}^3 tramite le coordinate (come indicato nel Paragrafo 2.2). In questo modo punti, rette e piani di \mathcal{A}^3 diventano sottospazi affini di \mathbb{R}^3 di dimensione 0, 1 o 2 rispettivamente. Per semplicità tipografica in questo capitolo le coordinate dei punti di \mathbb{R}^2 ed \mathbb{R}^3 verranno talvolta scritte per riga invece che per colonna: $(1, 2) \in \mathbb{R}^2$, o $(x_0, y_0, z_0) \in \mathbb{R}^3$ e così via.

Come visto nel Paragrafo 6.5, un sottospazio affine L di \mathbb{R}^3 può venire rappresentato tramite equazioni cartesiane o tramite equazioni parametriche. Per comodità di lettura, riportiamo qui le definizioni adattate alla situazione attuale.

¹ L'aggettivo "affine" serve a distinguerla dalla geometria "metrica" di rette e piani che affronteremo nel Capitolo 13, e dalla geometria "proiettiva" discussa nei Complementi a questo capitolo. Vedi anche il Paragrafo 10C.1 per un'ulteriore precisazione del termine "affine".

Definizione 10.1 Se L è lo spazio delle soluzioni del sistema

$$Ax = b, \quad (10.1)$$

dove $A \in M_{p,3}(\mathbb{R})$ e $b \in \mathbb{R}^p$, diremo che (10.1) è un'equazione cartesiana per L . La dimensione di L è $3 - \text{rg } A$, e il sottospazio di giacitura di L è lo spazio V delle soluzioni del sistema omogeneo $Ax = O$.

Come già discusso nei Paragrafi 6.4 e 6.5, di solito si scelgono le equazioni cartesiane in modo che $\text{rg } A = p$, cioè in modo che non vi siano equazioni inutili (il che equivale a richiedere che L_A sia surgettiva, o che il sistema $Ax = b$ abbia soluzione qualunque sia $b \in \mathbb{R}^p$). Quindi un piano nello spazio sarà descritto da un'equazione e una retta da due. Ma vediamo in dettaglio.

(a) Equazione cartesiana di un piano nello spazio: L'equazione

$$\boxed{ax + by + cz = d} \quad (10.2)$$

rappresenta un piano in \mathbb{R}^3 , ammesso che la matrice $A = |a \ b \ c|$ abbia rango 1 (il che equivale semplicemente a richiedere che almeno uno dei coefficienti a, b, c sia diverso da zero). I coefficienti a, b e c si dicono *parametri di giacitura* del piano. I parametri di giacitura non sono univocamente determinati; l'equazione

$$(\lambda a)x + (\lambda b)y + (\lambda c)z = (\lambda d)$$

identifica lo stesso piano qualunque sia $\lambda \in \mathbb{R}$ con $\lambda \neq 0$ (vedi anche l'Esercizio 10.1).

(b) Equazioni cartesiane di una retta nello spazio: Il sistema

$$\boxed{\begin{cases} ax + by + cz = d, \\ a'x + b'y + c'z = d', \end{cases}} \quad (10.3)$$

rappresenta una retta in \mathbb{R}^3 , sempre che la matrice

$$A = \begin{vmatrix} a & b & c \\ a' & b' & c' \end{vmatrix}$$

abbia rango 2. In questo modo abbiamo rappresentato la retta come intersezione di due piani: il piano $ax + by + cz = d$, e il piano $a'x + b'y + c'z = d'$. Come vedremo nel Paragrafo 10.6, la condizione $\text{rg } A = 2$ equivale a richiedere che i due piani non siano paralleli. Inutile dire che la matrice A si guarda bene dall'essere unica: una retta è intersezione di infinite coppie distinte di piani.

Osservazione 10.1 Il numero minimo di equazioni necessarie per descrivere un sottospazio affine è dato dalla dimensione dello spazio ambiente meno la dimensione del sottospazio. Dunque per descrivere una retta (dimensione 1) nel piano (dimensione 2) basta un'equazione (e non ne servono due, come accadeva nello spazio). Quindi nel piano \mathbb{R}^2 l'equazione

$$ax + by = c \quad (10.4)$$

rappresenta (ammesso che almeno uno dei due coefficienti a e b sia diverso da zero) una retta. In particolare, *l'insieme rappresentato da delle equazioni cartesiane dipende dallo spazio ambiente*. In \mathbb{R}^2 , l'equazione (10.4) rappresenta una retta; in \mathbb{R}^3 , la stessa equazione rappresenta un piano.

Passiamo ora alle equazioni parametriche.

Definizione 10.2 Se L è l'insieme dei punti $P \in \mathbb{R}^3$ della forma

$$P = P_0 + Bt, \quad (10.5)$$

al variare di $t \in \mathbb{R}^q$, dove $P_0 \in \mathbb{R}^3$ e $B \in M_{3,q}(\mathbb{R})$, diremo che (10.5) è un'equazione parametrica per L . Il sottospazio L ha dimensione $\text{rg } B$, e il sottospazio vettoriale V di giacitura di L ha equazione parametrica $P = Bt$.

Anche le equazioni parametriche vengono di solito scelte in modo che $\text{rg } B = q$, cioè in modo che non ci siano variabili inutili (il che equivale a richiedere che L_B sia iniettiva, cioè che ogni punto del sottospazio affine corrisponda a uno e un solo valore dei parametri). Dunque questa volta il numero dei parametri dipende dalla dimensione del sottospazio, mentre il numero delle equazioni (tre in \mathbb{R}^3 , due in \mathbb{R}^2) dipende dall'ambiente. Ma vediamo in dettaglio.

(a) Equazioni parametriche di una retta nello spazio: le equazioni

$$\boxed{\begin{cases} x = x_0 + lt, \\ y = y_0 + mt, \\ z = z_0 + nt, \end{cases}} \quad \text{oppure} \quad \boxed{P = P_0 + t \begin{vmatrix} l \\ m \\ n \end{vmatrix}} \quad (10.6)$$

al variare di $t \in \mathbb{R}$ rappresentano una retta nello spazio, ammesso che il vettore

$$v = \begin{vmatrix} l \\ m \\ n \end{vmatrix}$$

detto vettore direttore della retta, non sia il vettore nullo. Le tre coordinate l, m ed n del vettore direttore si chiamano *parametri direttori* della retta. Chiaramente, non sono univocamente determinati; i vettori v e λv sono vettori direttori della stessa retta, qualunque sia $\lambda \in \mathbb{R}$ con $\lambda \neq 0$. Anche il punto $P_0 = (x_0, y_0, z_0)$ si guarda bene dall'essere univocamente determinato: può essere sostituito da un qualunque altro punto della retta (vedi l'Esercizio 10.3).

(b) Equazioni parametriche di un piano nello spazio: le equazioni

$$\boxed{\begin{cases} x = x_0 + ls + l't, \\ y = y_0 + ms + m't, \\ z = z_0 + ns + n't, \end{cases}} \quad \text{oppure} \quad \boxed{P = P_0 + s \begin{vmatrix} l \\ m \\ n \end{vmatrix} + t \begin{vmatrix} l' \\ m' \\ n' \end{vmatrix}} \quad (10.7)$$

al variare di $s, t \in \mathbb{R}$ rappresentano un piano nello spazio, ammesso che la matrice

$$B = \begin{vmatrix} l & l' \\ m & m' \\ n & n' \end{vmatrix}$$

abbia rango 2 (ovvero che abbia colonne linearmente indipendenti). I due vettori

$$v = \begin{vmatrix} l \\ m \\ n \end{vmatrix} \quad \text{e} \quad v' = \begin{vmatrix} l' \\ m' \\ n' \end{vmatrix},$$

che formano una base del sottospazio di giacitura del piano, sono detti *vettori di giacitura* del piano. Non sono unici: possono essere sostituiti da una qualunque altra base del sottospazio di giacitura, e P_0 può essere un qualunque punto del piano.

Osservazione 10.2 Le equazioni parametriche di una retta nel piano non sono molto diverse da quelle di una retta nello spazio; semplicemente, sono due invece di tre. Per la precisione, sono della forma

$$\begin{cases} x = x_0 + lt, \\ y = y_0 + mt, \end{cases} \quad \text{oppure} \quad P = P_0 + t \begin{vmatrix} l \\ m \end{vmatrix}$$

al variare di $t \in \mathbb{R}$, sempre supponendo che il vettore

$$v = \begin{vmatrix} l \\ m \end{vmatrix},$$

detto *vettore direttore* della retta, non sia il vettore nullo.

Tecniche per passare da equazioni cartesiane a equazioni parametriche e viceversa nel caso di rette e piani sono descritte negli Esercizi 10.4–10.7.

10.2 Punti e rette

Dopo aver rivisto come rappresentare punti, rette e piani, passiamo a discutere in che posizioni reciproche si possono trovare.

(A) *Una retta e un punto.* Iniziamo nel piano. Dato un punto $P_0 = (x_0, y_0) \in \mathbb{R}^2$, vogliamo tutte le rette che passano per quel punto. Le equazioni parametriche sono immediate: se $P = P_0 + tv$ sono equazioni parametriche per la retta r , ponendo $t = 0$ vediamo che P_0 appartiene alla retta. Quindi le rette che passano per P_0 sono tutte e sole quelle di equazioni

$$\begin{cases} x = x_0 + tl, \\ y = y_0 + tm, \end{cases} \quad (10.8)$$

qualunque siano $l, m \in \mathbb{R}$.

Anche le equazioni cartesiane sono recuperate facilmente. Vogliamo che P_0 appartenga alla retta, cioè che le sue coordinate siano soluzione dell'equazione $ax + by = c$. Quindi $ax_0 + by_0 = c$, per cui l'equazione cartesiana della generica retta per P_0 è

$$a(x - x_0) + b(y - y_0) = 0. \quad (10.9)$$

Definizione 10.3 L'insieme delle rette del piano passanti per P_0 si chiama *fascio di rette di centro P_0* .

Passiamo ora allo spazio. Dato un punto $P_0 = (x_0, y_0, z_0) \in \mathbb{R}^3$, vogliamo tutte le rette che passano per quel punto. Le equazioni parametriche si ottengono esattamente come prima, e sono

$$\begin{cases} x = x_0 + tl, \\ y = y_0 + tm, \\ z = z_0 + tn, \end{cases} \quad (10.10)$$

qualunque siano $l, m, n \in \mathbb{R}$.

Solito discorso anche per le equazioni cartesiane. Questa volta troviamo

$$\begin{cases} a(x - x_0) + b(y - y_0) + c(z - z_0) = 0, \\ a'(x - x_0) + b'(y - y_0) + c'(z - z_0) = 0. \end{cases} \quad (10.11)$$

Definizione 10.4 L'insieme delle rette dello spazio passanti per P_0 si chiama *stella di rette di centro P_0* .

(B) *Una retta e due punti.* Com'è noto, per due punti distinti passa una e una sola retta; vogliamo trovarla. Cominciamo di nuovo nel piano. Vogliamo la retta che passa per i punti distinti $P_0 = (x_0, y_0)$ e $P_1 = (x_1, y_1) \in \mathbb{R}^2$. Usando le equazioni parametriche dobbiamo trovare $l, m \in \mathbb{R}$ tali che il sistema nell'incognita t

$$\begin{cases} x_1 = x_0 + tl, \\ y_1 = y_0 + tm, \end{cases} \quad (10.12)$$

abbia soluzione. In questo caso l ed m si trovano facilmente a occhio (vero?), ma visto che utilizzeremo lo stesso procedimento molte volte in futuro vediamo di risolvere il problema con tutti i crismi. Abbiamo un sistema lineare di cui vogliamo soltanto sapere se ha soluzione. È quindi naturale utilizzare il Teorema di Rouché-Capelli (Corollario 5.9), che dice che il sistema (10.12) ha soluzione se e solo se

$$1 = \operatorname{rg} \begin{vmatrix} l & l \\ m & m \end{vmatrix} = \operatorname{rg} \begin{vmatrix} l & x_1 - x_0 \\ m & y_1 - y_0 \end{vmatrix}.$$

Questo accade se e solo se la matrice sulla destra è singolare, cioè se e solo se

$$l(y_1 - y_0) - m(x_1 - x_0) = 0 \quad (10.13)$$

(è il determinante della matrice). Una soluzione di questa equazione (tutte le altre sono dei multipli, per cui ci forniscono la stessa retta) è $l = x_1 - x_0$ e $m = y_1 - y_0$.

Dunque la retta passante per P_0 e P_1 ha equazioni parametriche

$$\begin{cases} x = x_0 + (x_1 - x_0)t, \\ y = y_0 + (y_1 - y_0)t. \end{cases} \quad (10.14)$$

Osservazione 10.3 La tecnica utilizzata per risolvere questo problema può essere riassunta come segue: la condizione posta (passaggio per il punto P_1) si traduce in un sistema dipendente da parametri (l ed m) che deve avere soluzione; il Teorema di Rouché-Capelli fornisce allora una condizione necessaria e sufficiente in termini di ranghi di matrici; a sua volta questa condizione può essere espressa in termini di determinanti, e quindi in un certo numero di equazioni lineari nei parametri. La soluzione di queste equazioni ci fornisce la soluzione del nostro problema.

Vediamo ora le equazioni cartesiane della retta per due punti. Richiedendo che la generica retta del fascio (10.9) passi per P_1 otteniamo la condizione

$$a(x_1 - x_0) + b(y_1 - y_0) = 0;$$

quindi risolvendo quest'equazione troviamo che la retta passante per P_0 e P_1 ha equazione cartesiana

$$(y_1 - y_0)(x - x_0) - (x_1 - x_0)(y - y_0) = 0. \quad (10.15)$$

Passiamo ora ad A^3 ; vogliamo la retta passante per i punti $P_0 = (x_0, y_0, z_0)$ e $P_1 = (x_1, y_1, z_1) \in \mathbb{R}^3$. In termini di equazioni parametriche, vogliamo che il sistema

$$\begin{cases} x_1 = x_0 + tl, \\ y_1 = y_0 + tm, \\ z_1 = z_0 + tn, \end{cases}$$

abbia soluzione. Per il Teorema di Rouché-Capelli questo accade se e solo se

$$\operatorname{rg} \begin{vmatrix} l & x_1 - x_0 \\ m & y_1 - y_0 \\ n & z_1 - z_0 \end{vmatrix} = 1,$$

che, grazie al Teorema degli orlati, avviene se (ma vedi anche l'Esercizio 10.15)

$$\begin{cases} l(y_1 - y_0) - m(x_1 - x_0) = 0, \\ l(z_1 - z_0) - n(x_1 - x_0) = 0. \end{cases} \quad (10.16)$$

Questo sistema ha soluzione $l = x_1 - x_0$, $m = y_1 - y_0$ e $n = z_1 - z_0$ (o multipli), per cui la retta per P_0 e P_1 ha equazioni parametriche

$$\begin{cases} x = x_0 + (x_1 - x_0)t, \\ y = y_0 + (y_1 - y_0)t, \\ z = z_0 + (z_1 - z_0)t. \end{cases} \quad (10.17)$$

Per le equazioni cartesiane conviene procedere in modo lievemente diverso, evitando di ricorrere al sistema (10.11) in quanto dipendente da troppi parametri. Noi sappiamo che un punto $P = (x, y, z) \in \mathbb{R}^3$ appartiene alla retta passante per P_0 e P_1 se e solo se il vettore $P - P_0$ è un multiplo del vettore $P_1 - P_0$; questo è un altro modo di esprimere il sistema (10.17). In termini di matrici, questo accade se e solo se

$$\operatorname{rg} \begin{vmatrix} x - x_0 & x_1 - x_0 \\ y - y_0 & y_1 - y_0 \\ z - z_0 & z_1 - z_0 \end{vmatrix} = 1.$$

Utilizzando il Teorema degli orlati (Teorema 9.13) a partire da una coordinata non nulla di $P_1 - P_0$ otteniamo le due equazioni cartesiane cercate. Per esempio, se $x_1 \neq x_0$ le equazioni sono

$$\begin{cases} (y_1 - y_0)(x - x_0) - (x_1 - x_0)(y - y_0) = 0, \\ (z_1 - z_0)(x - x_0) - (x_1 - x_0)(z - z_0) = 0. \end{cases} \quad (10.18)$$

ESEMPIO 10.1 Vogliamo equazioni parametriche e cartesiane della retta passante per i punti $P_0 = (1, 0, 1)$ e $P_1 = (1, 1, 2)$. Le equazioni parametriche sono date da (10.17):

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

Per le equazioni cartesiane, non possiamo utilizzare (10.18) in quanto $x_1 = 1 = x_0$. Ma $y_1 - y_0 = 1 \neq 0$; quindi il Teorema degli orlati ci fornisce come equazioni cartesiane

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

10.3 Punti e piani

Passiamo a vedere le relazioni fra piani e punti — ovviamente, solo nello spazio.

(C) *Un piano e un punto.* Dato un punto $P_0 = (x_0, y_0, z_0) \in \mathbb{R}^3$ vogliamo tutti i piani che passano per quel punto. Si procede esattamente come per le rette, per cui troviamo

$$\begin{cases} x = x_0 + sl + tl', \\ y = y_0 + sm + tm', \\ z = z_0 + sn + tn', \end{cases} \quad (10.19)$$

come equazioni parametriche, e

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0 \quad (10.20)$$

come equazione cartesiana.

Definizione 10.5 L'insieme dei piani passanti per P_0 si chiama *stella di piani di centro P_0* .

(D) *Un piano e due punti.* Dati due punti $P_0 = (x_0, y_0, z_0)$ e $P_1 = (x_1, y_1, z_1) \in \mathbb{R}^3$ con $P_1 \neq P_0$, vogliamo tutti i piani che contengono questi due punti². Per le equazioni parametriche, conviene ragionare come segue: il piano π contiene sia P_0 che P_1 se e solo se contiene P_0 ed è parallelo al vettore $P_1 - P_0$. Dunque possiamo prendere $P_1 - P_0$ come uno dei vettori di giacitura del piano, e le equazioni parametriche sono

$$\begin{cases} x = x_0 + s(x_1 - x_0) + tl, \\ y = y_0 + s(y_1 - y_0) + tm, \\ z = z_0 + s(z_1 - z_0) + tn, \end{cases}$$

per tutti gli $l, m, n \in \mathbb{R}$ tali che

$$\operatorname{rg} \begin{vmatrix} x_1 - x_0 & l \\ y_1 - y_0 & m \\ z_1 - z_0 & n \end{vmatrix} = 2.$$

I parametri di giacitura di questo piano devono soddisfare (perché?) l'equazione

$$a(x_1 - x_0) + b(y_1 - y_0) + c(z_1 - z_0) = 0.$$

Per esempio, se $z_1 \neq z_0$ possiamo ricavare c in funzione di a e b e, dopo aver moltiplicato per $z_1 - z_0$, otteniamo come equazione cartesiana la

$$a(z_1 - z_0)(x - x_0) + b(z_1 - z_0)(y - y_0) - (a(x_1 - x_0) + b(y_1 - y_0))(z - z_0) = 0.$$

ESEMPIO 10.2 Vogliamo l'equazione cartesiana di tutti i piani contenenti i punti $P_0 = (-1, 0, 1)$ e $P_1 = (1, 1, 1)$. I parametri di giacitura di questi piani devono soddisfare l'equazione

$$2a + b = 0,$$

e quindi come equazione cartesiana troviamo

$$a(x + 1) - 2ay + c(z - 1) = 0.$$

(E) *Un piano e tre punti.* Dati tre punti non allineati P_0, P_1 e $P_2 = (x_2, y_2, z_2) \in \mathbb{R}^3$ vogliamo il piano che passa per questi tre punti. Prima di tutto notiamo che i tre punti non sono allineati se e solo se (perché?) i vettori $P_1 - P_0$ e $P_2 - P_0$ sono linearmente indipendenti (vedi anche l'Esercizio 10.9). Dunque possiamo prenderli entrambi come vettori di giacitura, e otteniamo le equazioni parametriche cercate:

$$\begin{cases} x = x_0 + s(x_1 - x_0) + t(x_2 - x_0), \\ y = y_0 + s(y_1 - y_0) + t(y_2 - y_0), \\ z = z_0 + s(z_1 - z_0) + t(z_2 - z_0). \end{cases}$$

² Ovvero che contengono la retta passante per P_0 e P_1 ; vedi (H) più oltre.

Per le equazioni cartesiane, i parametri di giacitura sono le soluzioni del sistema

$$\begin{cases} a(x_1 - x_0) + b(y_1 - y_0) + c(z_1 - z_0) = 0, \\ a(x_2 - x_0) + b(y_2 - y_0) + c(z_2 - z_0) = 0. \end{cases}$$

Ma c'è un modo più elegante per ottenere le equazioni cartesiane. Un punto $P \in \mathbb{R}^3$ appartiene al piano passante per P_0, P_1 e P_2 se e solo se il vettore $P - P_0$ appartiene al sottospazio di giacitura, che è generato da $P_1 - P_0$ e $P_2 - P_0$. Quindi $P = (x, y, z)$ appartiene al piano se e solo se i tre vettori $P - P_0, P_1 - P_0$ e $P_2 - P_0$ sono linearmente dipendenti, cioè se e solo se

$$\det \begin{vmatrix} x - x_0 & x_1 - x_0 & x_2 - x_0 \\ y - y_0 & y_1 - y_0 & y_2 - y_0 \\ z - z_0 & z_1 - z_0 & z_2 - z_0 \end{vmatrix} = 0.$$

Sviluppando il determinante troviamo l'equazione cartesiana cercata.

ESEMPIO 10.3 Vogliamo l'equazione cartesiana del piano passante per $P_0 = (1, 0, 0)$, $P_1 = (0, 0, 1)$ e $P_2 = (1, 1, 1)$. Per quanto visto l'equazione è

$$\det \begin{vmatrix} x - 1 & -1 & 0 \\ y & 0 & 1 \\ z & 1 & 1 \end{vmatrix} = 0,$$

cioè $-x + y - z = -1$.

10.4 Rette e rette

Vediamo ora le posizioni reciproche di due rette.

(F) *Due rette.* Per fissare la terminologia ci serve una definizione.

Definizione 10.6 Diremo che due rette r_0 ed r_1 (sia nel piano che nello spazio) sono *incidenti* se si intersecano in un punto; *parallele* se non si intersecano ma hanno vettori direttori paralleli (cioè linearmente dipendenti); *coincidenti* se sono la stessa retta; e infine *sghembe* se non sono né incidenti né parallele (e dunque in particolare hanno vettori direttori linearmente indipendenti).

Vogliamo stabilire quando due rette sono incidenti, parallele o sghembe. Cominciamo supponendo che entrambe siano date in forma parametrica, la retta r_0 dalle equazioni $P = P_0 + tv_0$, e la retta r_1 dalle equazioni $P = P_1 + sv_1$. Se siamo nel piano, P_0 e P_1 saranno punti del piano, e v_0, v_1 vettori di \mathbb{R}^2 ; nello spazio, saranno invece rispettivamente punti dello spazio e vettori di \mathbb{R}^3 . Le due rette sono parallele (o coincidenti) se e solo se i vettori direttori sono linearmente dipendenti, cioè

$$\operatorname{rg} |v_0 \ v_1| = 1$$

(condizione di parallelismo).

Vediamo ora quando le due rette si intersecano (in uno o più punti). Se $P \in r_0 \cap r_1$, il punto P deve potersi scrivere sia come $P_0 + tv_0$ per un determinato valore di t , sia come $P_1 + sv_1$ per un determinato valore di s . Quindi le due rette r_0 ed r_1 si intersecano se e solo se il sistema $P_0 + tv_0 = P_1 + sv_1$ ammette soluzione. Applicando il Teorema di Rouché-Capelli otteniamo

$$\operatorname{rg}[v_0 \ v_1 \ P_0 - P_1] = \begin{cases} \operatorname{rg}[v_0 \ v_1] & \text{le rette si intersecano;} \\ \operatorname{rg}[v_0 \ v_1] + 1 & \text{le rette non si intersecano.} \end{cases}$$

Supponiamo che $\operatorname{rg}[v_0 \ v_1] = 1$. Allora la condizione $\operatorname{rg}[v_0 \ v_1 \ P_0 - P_1] = 1$ implica che le due rette si intersecano, necessariamente in una retta, per cui sono *coincidenti*. Se invece $\operatorname{rg}[v_0 \ v_1 \ P_0 - P_1] = 2$, non si intersecano per cui sono *parallele*.

Supponiamo che $\operatorname{rg}[v_0 \ v_1] = 2$. Allora la condizione $\operatorname{rg}[v_0 \ v_1 \ P_0 - P_1] = 2$ implica che le due rette si intersecano, necessariamente (perché?) in un punto, per cui sono *incidenti*. Se invece $\operatorname{rg}[v_0 \ v_1 \ P_0 - P_1] = 3$ le due rette non sono né parallele né intersecantisi, per cui sono *sghembe*.

Quanto abbiamo ottenuto può venire riassunto dalla tabella seguente:

	$\operatorname{rg}[v_0 \ v_1 \ P_0 - P_1] = 1$	$\operatorname{rg}[v_0 \ v_1 \ P_0 - P_1] = 2$	$\operatorname{rg}[v_0 \ v_1 \ P_0 - P_1] = 3$
$\operatorname{rg}[v_0 \ v_1] = 1$	Rette coincidenti	Rette parallele	Impossibile
$\operatorname{rg}[v_0 \ v_1] = 2$	Impossibile	Rette incidenti	Rette sghembe

Osservazione 10.4 Nel piano la matrice $[v_0 \ v_1 \ P_0 - P_1]$ ha solo due righe, e quindi non può mai avere rango 3. Per questo motivo nel piano due rette sono sempre parallele o incidenti (o coincidenti).

Passiamo ora al caso in cui le due rette siano descritte da equazioni cartesiane, cominciando con il piano. Supponiamo che la retta r_0 abbia equazione $a_0x + b_0y = c_0$, e che la retta r_1 abbia equazione $a_1x + b_1y = c_1$. Le due rette (sono distinte e) si intersecano se e solo se il sistema

$$\begin{cases} a_0x + b_0y = c_0, \\ a_1x + b_1y = c_1, \end{cases}$$

ammette un'unica soluzione. Questo a sua volta accade, come sappiamo, se e solo se

$$\det \begin{vmatrix} a_0 & b_0 \\ a_1 & b_1 \end{vmatrix} \neq 0.$$

Dunque le due rette sono parallele (o coincidenti) se e solo se

$$a_0b_1 - a_1b_0 = 0$$

(condizione di parallelismo).

Infine, le due rette sono coincidenti se e solo se sono parallele e s'intersecano, che accade se e solo se

$$\operatorname{rg} \begin{vmatrix} a_0 & b_0 & c_0 \\ a_1 & b_1 & c_1 \end{vmatrix} = 1,$$

ovvero se e solo se le due equazioni sono una multiplo dell'altra.

Nello spazio, invece, la retta r_0 ha equazione $A_0x = b_0$, e la retta r_1 ha equazione $A_1x = b_1$, dove A_0 e A_1 sono matrici 2×3 di rango 2, e b_0 e b_1 sono vettori di \mathbb{R}^2 . I sottospazi di giacitura delle due rette hanno rispettivamente equazione $A_0x = O$ e $A_1x = O$. Le due rette sono parallele (o coincidenti) se e solo se questi due sistemi omogenei hanno le stesse soluzioni. Siccome entrambi hanno una retta di soluzioni, le due rette sono parallele se e solo se il sistema di quattro equazioni in tre incognite

$$\begin{cases} A_0x = O, \\ A_1x = O, \end{cases}$$

ammette soluzioni diverse da $x = O$, che accade se e solo se

$$\operatorname{rg} \begin{vmatrix} A_0 \\ A_1 \end{vmatrix} = 2$$

(condizione di parallelismo; stiamo usando il Teorema 5.7).

D'altro canto le due rette si intersecano se e solo se il sistema

$$\begin{cases} A_0x = b_0, \\ A_1x = b_1, \end{cases} \tag{10.21}$$

ammette soluzioni, il che accade se e solo se

$$\operatorname{rg} \begin{vmatrix} A_0 & b_0 \\ A_1 & b_1 \end{vmatrix} = \operatorname{rg} \begin{vmatrix} A_0 \\ A_1 \end{vmatrix};$$

se i ranghi di queste matrici sono diversi le rette sono parallele o sghembe.

Anche stavolta possiamo riassumere quanto trovato in una tabella:

	$\operatorname{rg} \begin{vmatrix} A_0 & b_0 \\ A_1 & b_1 \end{vmatrix} = 2$	$\operatorname{rg} \begin{vmatrix} A_0 & b_0 \\ A_1 & b_1 \end{vmatrix} = 3$	$\operatorname{rg} \begin{vmatrix} A_0 & b_0 \\ A_1 & b_1 \end{vmatrix} = 4$
$\operatorname{rg} \begin{vmatrix} A_0 \\ A_1 \end{vmatrix} = 2$	Rette coincidenti	Rette parallele	Impossibile
$\operatorname{rg} \begin{vmatrix} A_0 \\ A_1 \end{vmatrix} = 3$	Impossibile	Rette incidenti	Rette sghembe

ESEMPIO 10.4 Date le rette r_0 ed r_1 di equazione cartesiana

$$r_0 : \begin{cases} x - y + z = 1, \\ x + y - z = 1, \end{cases} \quad r_1 : \begin{cases} x + y + z = 0, \\ 3x + y + z = 2, \end{cases}$$

vogliamo stabilire in che posizione reciproca si trovano. Siccome

$$\operatorname{rg} \begin{vmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \\ 3 & 1 & 1 \end{vmatrix} = 3 = \operatorname{rg} \begin{vmatrix} 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 0 \\ 3 & 1 & 1 & 2 \end{vmatrix},$$

le due rette sono incidenti.

10.5 Rette e piani

Puntiamo ora i riflettori sulle relazioni fra piani e rette.

(G) *Una retta e un piano.* Consideriamo una retta r di equazioni parametriche $P = P_0 + sv_0$ e un piano π di equazioni parametriche $P = P_1 + tv_1 + t'v_2$. Abbiamo due possibilità:

— Se $\det|v_0 \ v_1 \ v_2| \neq 0$, allora il sistema

$$P_0 + sv_0 = P_1 + tv_1 + t'v_2 \quad (10.22)$$

nelle incognite s, t e t' ammette un'unica soluzione. Questo vuol dire che la retta e il piano sono *incidenti*: si intersecano in un punto.

— Se invece $\det|v_0 \ v_1 \ v_2| = 0$, il sistema (10.22) o non ha soluzioni — nel qual caso la retta e il piano sono *paralleli* —, oppure ha una retta di soluzioni — nel qual caso la retta r è *contenuta* nel piano. In particolare, retta e piano sono paralleli se e solo se

$$\operatorname{rg}|v_0 \ v_1 \ v_2 \ P_0 - P_1| = 3 \neq 2 = \operatorname{rg}|v_0 \ v_1 \ v_2|.$$

Supponiamo invece che la retta r abbia equazioni cartesiane $A_0x = b_0$, e che il piano π abbia equazioni cartesiane $\alpha^T x = d$, dove $\alpha \in \mathbb{R}^3$ è il vettore dei parametri di giacitura del piano. I due sono paralleli se e solo se il sistema quadrato

$$\begin{cases} A_0x = 0, \\ \alpha^T x = 0, \end{cases}$$

ammette soluzioni diverse da $x = O$ (perché?), il che accade se e solo se

$$\det \begin{vmatrix} A_0 \\ \alpha^T \end{vmatrix} = 0.$$

In questo caso, la retta è contenuta nel piano se e solo se

$$\operatorname{rg} \begin{vmatrix} A_0 & b_0 \\ \alpha^T & d \end{vmatrix} = 2.$$

Invece

$$\det \begin{vmatrix} A_0 \\ \alpha^T \end{vmatrix} \neq 0$$

se e solo se il piano e la retta sono incidenti.

ESEMPIO 10.5 Consideriamo la retta r e il piano π di equazioni cartesiane

$$r: \begin{cases} x + y + z = 1, \\ x - y + z = 0, \end{cases} \quad \pi: 3x + y + 3z = 2.$$

Siccome

$$\operatorname{rg} \begin{vmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 3 & 1 & 3 \end{vmatrix} = 2 = \operatorname{rg} \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 0 \\ 3 & 1 & 3 & 2 \end{vmatrix},$$

la retta r è contenuta nel piano π .

Vi è un altro caso che vale la pena considerare. Supponiamo che il piano π abbia equazione cartesiana $ax + by + cz = d$, e che la retta r abbia equazioni parametriche $P = P_0 + tv_0$. In tal caso, la retta e il piano sono paralleli se e solo se il vettore direttore $v_0 = (l, m, n)$ della retta appartiene al sottospazio di giacitura del piano, cioè se e solo se

$$al + bm + cn = 0 \quad (10.23)$$

(condizione di parallelismo). In questo caso, la retta è contenuta nel piano se e solo se inoltre $P_0 \in \pi$, il che accade se e solo se

$$d = ax_0 + by_0 + cz_0.$$

(H) *Un piano e una retta.* Data una retta r , vogliamo trovare tutti i piani che la contengono. Cominciamo come al solito supponendo che la retta r abbia equazioni parametriche $P = P_0 + tv_0$, con $P_0 = (x_0, y_0, z_0)$ e $v_0 = (l_0, m_0, n_0)$. Se un piano contiene r , il vettore v_0 può venire preso come un vettore di giacitura del piano. Quindi tutti i piani contenenti r sono descritti dalle equazioni parametriche

$$\begin{vmatrix} x \\ y \\ z \end{vmatrix} = \begin{vmatrix} x_0 \\ y_0 \\ z_0 \end{vmatrix} + t \begin{vmatrix} l_0 \\ m_0 \\ n_0 \end{vmatrix} + s \begin{vmatrix} l \\ m \\ n \end{vmatrix} \quad (10.24)$$

al variare di $l, m, n \in \mathbb{R}$ in modo che si abbia

$$\operatorname{rg} \begin{vmatrix} l_0 & l \\ m_0 & m \\ n_0 & n \end{vmatrix} = 2.$$

Supponiamo invece che la retta r sia data da equazioni cartesiane,

$$\begin{cases} ax + by + cz = d, \\ a'x + b'y + c'z = d'. \end{cases}$$

Per quanto visto in (G), il piano di equazione cartesiana $a''x + b''y + c''z = d''$ contiene la retta r se e solo se

$$\operatorname{rg} \begin{pmatrix} a & b & c & d \\ a' & b' & c' & d' \\ a'' & b'' & c'' & d'' \end{pmatrix} = 2,$$

cioè se e solo se esistono $\lambda, \mu \in \mathbb{R}$ non entrambi nulli tali che

$$\begin{pmatrix} a'' \\ b'' \\ c'' \\ d'' \end{pmatrix} = \lambda \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} + \mu \begin{pmatrix} a' \\ b' \\ c' \\ d' \end{pmatrix}.$$

Dunque tutti i piani contenenti r sono descritti dall'equazione cartesiana

$$\lambda(ax + by + cz) + \mu(a'x + b'y + c'z) = \lambda d + \mu d'$$

al variare di $(\lambda, \mu) \in \mathbb{R}^2$ con $(\lambda, \mu) \neq (0, 0)$.

Definizione 10.7 L'insieme dei piani passanti per la retta r si chiama *fascio di piani di asse r* .

(I) *Un piano, una retta e un punto.* Vogliamo trovare il piano contenente una retta data r e passante per un punto $P_1 = (x_1, y_1, z_1)$ non appartenente alla retta. Supponiamo che la retta abbia equazioni parametriche $P = P_0 + tv_0$, con $P_0 = (x_0, y_0, z_0)$ e $v_0 = (l_0, m_0, n_0)$. Ricordando la forma delle equazioni parametriche dei piani del fascio di asse r , vediamo subito che dobbiamo trovare $l, m, n \in \mathbb{R}$ tali che il sistema

$$P_1 = P_0 + tv_0 + sv \quad (10.25)$$

abbia soluzione, dove $v = (l, m, n)$. Chiaramente, $v = P_1 - P_0$ è una possibilità (la soluzione è $t = 0, s = 1$). Viceversa, supponiamo che (10.25) sia soddisfatta per qualche $s_0, t_0 \in \mathbb{R}$. Siccome per ipotesi P_1 non appartiene alla retta r , la differenza $P_1 - P_0$ non può essere un multiplo di v_0 , per cui $s_0 \neq 0$. Dividendo per s_0 troviamo

$$v = \lambda(P_1 - P_0) + \mu v_0$$

per qualche $\lambda, \mu \in \mathbb{R}$ con $\lambda = 1/s_0 \neq 0$. Inoltre, qualunque scelta di $\lambda \in \mathbb{R}^*$ e $\mu \in \mathbb{R}$ ci fornisce lo stesso piano, in quanto

$$\operatorname{Span}(v_0, \lambda(P_1 - P_0) + \mu v_0) = \operatorname{Span}(v_0, P_1 - P_0).$$

In conclusione, esiste un unico piano contenente r e P_0 , di equazioni parametriche

$$P = P_0 + tv_0 + s(P_1 - P_0).$$

Supponiamo invece che la retta r sia data dalle equazioni cartesiane

$$\begin{cases} ax + by + cz = d, \\ a'x + b'y + c'z = d'. \end{cases}$$

L'unico piano del fascio di asse r che contiene il punto P_1 è quello per cui i parametri λ e μ del fascio soddisfano l'equazione

$$(ax_1 + by_1 + cz_1 - d)\lambda + (a'x_1 + b'y_1 + c'z_1 - d')\mu = 0$$

(nota che, siccome $P_1 \notin r$, i coefficienti di questa equazione non sono entrambi nulli). Quindi il piano ha equazione cartesiana

$$(a'x_1 + b'y_1 + c'z_1 - d')(ax + by + cz - d) - (ax_1 + by_1 + cz_1 - d)(a'x + b'y + c'z - d') = 0.$$

(L) *Un piano e due rette.* Vogliamo vedere quando (e se) esiste un piano contenente due rette diverse r_0 ed r_1 . Come vedremo, questo piano esiste sempre tranne nel caso in cui le rette sono sghembe.

Supponiamo che le equazioni parametriche delle rette siano $P = P_0 + tv_0$ e $P = P_1 + sv_1$ rispettivamente, e distinguiamo due casi.

— $\operatorname{rg}(v_0, v_1) = 2$: le rette sono incidenti o sghembe. Perché un piano π contenga sia r_0 che r_1 deve contenere sia P_0 che P_1 , e il sottospazio di giacitura deve contenere sia v_0 che v_1 . Siccome v_0 e v_1 sono per ipotesi linearmente indipendenti, possiamo prenderli come vettori di giacitura di π , che ha quindi

$$P = P_0 + sv_0 + tv_1 \quad (10.26)$$

come equazioni parametriche. Questo piano contiene anche P_1 (e quindi entrambe le rette) se e solo se il sistema $P_1 = P_0 + sv_0 + tv_1$ ha soluzione, il che accade se e solo se $\operatorname{rg}(P_0 - P_1, v_0, v_1) = \operatorname{rg}(v_0, v_1)$, per cui le rette sono incidenti. Dunque se v_0 e v_1 sono linearmente indipendenti esiste un piano che contiene r_0 ed r_1 se e solo se le due rette sono incidenti, e in tal caso il piano ha (10.26) come equazioni parametriche.

$\operatorname{rg}(v_0, v_1) = 1$: le rette sono parallele (e distinte, per ipotesi). In questo caso basta prendere il piano π che contiene r_0 e P_1 , di equazioni parametriche

$$P = P_0 + sv_0 + t(P_1 - P_0);$$

infatti π contiene automaticamente anche r_1 (perché?).

Vediamo ora il caso in cui r_0 ed r_1 siano date da equazioni cartesiane

$$r_0 : \begin{cases} a_0x + b_0y + c_0z = d_0, \\ a'_0x + b'_0y + c'_0z = d'_0, \end{cases} \quad r_1 : \begin{cases} a_1x + b_1y + c_1z = d_1, \\ a'_1x + b'_1y + c'_1z = d'_1. \end{cases}$$

Il piano che le contiene entrambe (se esiste) dev'essere un piano comune ai due fasci di asse rispettivamente r_0 ed r_1 . Questo accade se e solo se esistono $\lambda_0, \mu_0, \lambda_1, \mu_1 \in \mathbb{R}$ non tutti nulli tali che le due equazioni

$$\begin{aligned} \lambda_0(a_0x + b_0y + c_0z) + \mu_0(a'_0x + b'_0y + c'_0z) &= \lambda_0d_0 + \mu_0d'_0, \\ \lambda_1(a_1x + b_1y + c_1z) + \mu_1(a'_1x + b'_1y + c'_1z) &= \lambda_1d_1 + \mu_1d'_1, \end{aligned} \quad (10.27)$$

siano uguali (vedi l'Esercizio 10.1). Confrontando i coefficienti rispettivamente di x , y , z e i termini noti, vediamo che a sua volta questo succede se e solo se il sistema

$$\begin{vmatrix} a_0 & a'_0 & -a_1 & -a'_1 & \lambda_0 \\ b_0 & b'_0 & -b_1 & -b'_1 & \mu_0 \\ c_0 & c'_0 & -c_1 & -c'_1 & \lambda_1 \\ d_0 & d'_0 & -d_1 & -d'_1 & \mu_1 \end{vmatrix} = \begin{vmatrix} 0 \\ 0 \\ 0 \\ 0 \end{vmatrix} \quad (10.28)$$

ammette soluzioni diverse dalla soluzione nulla. Questo accade se e solo se la matrice dei coefficienti ha rango strettamente minore di 4, il che — ricordando (F) — equivale a dire che le rette non sono sghembe. Infine, il piano cercato ha come equazione cartesiana la seconda delle (10.27), dove λ_0 , μ_0 , λ_1 e μ_1 sono soluzione del sistema (10.28).

ESEMPIO 10.6 Vogliamo l'equazione cartesiana del piano contenente le rette incidenti dell'Esempio 10.4. Risolvendo il sistema

$$\begin{vmatrix} 1 & 1 & -1 & -3 & \lambda_0 \\ -1 & 1 & -1 & -1 & \mu_0 \\ 1 & -1 & -1 & -1 & \lambda_1 \\ 1 & 1 & 0 & -2 & \mu_1 \end{vmatrix} = \begin{vmatrix} 0 \\ 0 \\ 0 \\ 0 \end{vmatrix}$$

troviamo (per esempio) $\lambda_1 = -1$ e $\mu_1 = 1$, per cui il piano cercato ha equazione cartesiana $x = 1$.

10.6 Piani e piani

Concludiamo con le posizioni reciproche di due piani.

(M) *Due piani*. Una definizione è di rigore.

Definizione 10.8 Diremo che due piani π_0 e π_1 sono *incidenti* se si intersecano in una retta; *paralleli* se non si intersecano o coincidono (cioè sono lo stesso piano).

Supponiamo che i due piani π_0 e π_1 abbiano equazioni parametriche rispettivamente $P = P_0 + sv_0 + tw_0$ e $P = P_1 + s'v_1 + t'w_1$. Ci sono due possibilità: se

$$\text{rg}\{v_0 \ w_0 \ v_1 \ w_1\} = 2$$

i due piani hanno lo stesso sottospazio di giacitura, e quindi sono *paralleli*. In particolare, sono coincidenti (perché?) se e solo se

$$\text{rg}\{v_0 \ w_0 \ v_1 \ w_1 \ P_0 - P_1\} = 2.$$

Se invece

$$\text{rg}\{v_0 \ w_0 \ v_1 \ w_1\} = 3,$$

i due piani sono *incidenti*; infatti in tal caso il sistema

$$P_0 + sv_0 + tw_0 = P_1 + s'v_1 + t'w_1$$

ha sempre una retta di soluzioni.

Supponiamo ora che i due piani abbiano equazioni cartesiane $a_0x + b_0y + c_0z = d_0$ e $a_1x + b_1y + c_1z = d_1$ rispettivamente. Anche stavolta ci sono due possibilità. Se

$$\text{rg} \begin{vmatrix} a_0 & b_0 & c_0 \\ a_1 & b_1 & c_1 \end{vmatrix} = 1,$$

(condizione di parallelismo) i due piani hanno lo stesso sottospazio di giacitura, e quindi sono paralleli. In particolare, sono coincidenti se e solo se

$$\text{rg} \begin{vmatrix} a_0 & b_0 & c_0 & d_0 \\ a_1 & b_1 & c_1 & d_1 \end{vmatrix} = 1.$$

Se invece

$$\text{rg} \begin{vmatrix} a_0 & b_0 & c_0 \\ a_1 & b_1 & c_1 \end{vmatrix} = 2,$$

il sistema composto dalle equazioni dei due piani ammette una retta di soluzioni, per cui i piani sono incidenti.

Quanto visto finora ovviamente non esaurisce le questioni di geometria affine che è possibile porsi (e risolvere), ma descrive in maniera abbastanza esauritiva il tipo di tecniche che vengono utilizzate. Concludiamo questo paragrafo con un esempio di come si possa risolvere un altro problema di questo genere.

ESEMPIO 10.7 Vogliamo equazioni parametriche e cartesiane del piano passante per il punto $P_0 = (1, 0, 1)$ e parallelo alle rette

$$r_1 : \begin{cases} x = t, \\ y = 2t, \\ z = 1 + 3t, \end{cases} \quad r_2 : \begin{cases} 2x + z = 1, \\ 2y - z = 3. \end{cases}$$

La retta r_2 ha equazioni parametriche

$$\begin{cases} x = -2s, \\ y = 2 + 2s, \\ z = 1 + 4s; \end{cases}$$

quindi il piano cercato ha equazioni parametriche

$$\begin{cases} x = 1 + t - 2s, \\ y = 2t + 2s, \\ z = 1 + 3t + 4s, \end{cases}$$

o dunque equazione cartesiana $2(x-1) - 10y + 6(z-1) = 0$, ovvero

$$x - 5y + 3z = 4.$$

10.7 Sistemi di riferimento affini

Tutto il discorso fatto finora si basa sull'aver fissato una volta per tutte un sistema di riferimento affine $RA(O, A_1, A_2, A_3)$ in \mathcal{A}^3 . Ma che succede se cambiamo sistema di riferimento? Ovviamente, cambieranno le coordinate. Come abbiamo già accennato nel Paragrafo 2.2, a volte può essere utile scegliere un sistema di riferimento costruito *ad hoc* per il problema da esaminare; sorge quindi spontaneo il problema di come mutano le coordinate passando da un sistema di riferimento a un altro. Abbiamo già esaminato la questione nel Paragrafo 8.1 nel caso degli spazi vettoriali, in cui cambiava la base ma l'origine rimaneva la stessa. Cambiando sistema di riferimento affine, invece, muoviamo anche l'origine, per cui la situazione è lievemente diversa.

Ricordiamo brevemente le definizioni.

Definizione 10.9 Un sistema di riferimento affine $RA(O, A_1, A_2, A_3)$ in \mathcal{A}^3 consiste in una quaterna ordinata di punti non complanari $O, A_1, A_2, A_3 \in \mathcal{A}^3$. Il primo punto O è l'origine del sistema di riferimento. Siccome i quattro punti sono non complanari, i tre vettori $\overrightarrow{OA_1}, \overrightarrow{OA_2}$ e $\overrightarrow{OA_3}$ formano una base dello spazio vettoriale \mathcal{V}_O^3 . Le coordinate associate a questa base si chiamano *coordinate affini* rispetto al sistema di riferimento affine $RA(O, A_1, A_2, A_3)$.

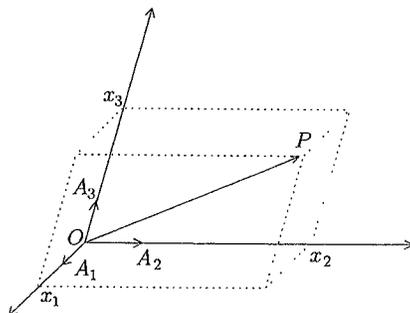


Figura 10.1 Le coordinate affini.

Per l'esattezza, un punto $P \in \mathcal{A}^3$ ha coordinate affini $(x_1, x_2, x_3) \in \mathbb{R}^3$ se e solo se (vedi la Figura 10.1)

$$\overrightarrow{OP} = x_1 \overrightarrow{OA_1} + x_2 \overrightarrow{OA_2} + x_3 \overrightarrow{OA_3}.$$

Siano $\mathcal{R} = RA(O, A_1, A_2, A_3)$ e $\mathcal{R}' = RA(O', A'_1, A'_2, A'_3)$ due sistemi di riferimento affini in \mathcal{A}^3 . Prendiamo un punto $P \in \mathcal{A}^3$; vogliamo vedere che relazione intercorre fra le sue coordinate rispetto a \mathcal{R} e quelle rispetto a \mathcal{R}' .

Siano (c_1, c_2, c_3) le coordinate di O' rispetto a \mathcal{R} (ovviamente, O' ha coordinate $(0, 0, 0)$ rispetto a \mathcal{R}'), cioè

$$\overrightarrow{OO'} = c_1 \overrightarrow{OA_1} + c_2 \overrightarrow{OA_2} + c_3 \overrightarrow{OA_3} = \begin{vmatrix} \overrightarrow{OA_1} & \overrightarrow{OA_2} & \overrightarrow{OA_3} \end{vmatrix} c,$$

dove

$$c = \begin{vmatrix} c_1 \\ c_2 \\ c_3 \end{vmatrix} \in \mathbb{R}^3,$$

e il prodotto è il solito prodotto righe per colonne.

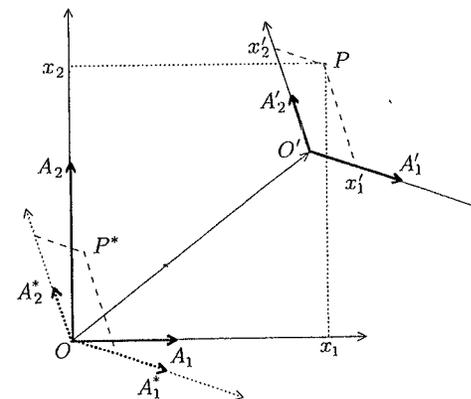


Figura 10.2 Cambiamento di coordinate affini in \mathcal{A}^2 .

Consideriamo ora il sistema di riferimento affine $\mathcal{R}^* = RA(O, A_1^*, A_2^*, A_3^*)$ ottenuto traslando l'origine di \mathcal{R}' in O (vedi la Figura 10.2, dove per semplicità è raffigurata la situazione in \mathcal{A}^2). I punti A_j^* sono ottenuti traslando i punti A_j' tramite il vettore $-\overrightarrow{OO'}$, cioè

$$\overrightarrow{OA_j^*} = \overrightarrow{OA_j'} - \overrightarrow{OO'},$$

per $j = 1, 2, 3$. Siccome i quattro punti O', A_1', A_2' e A_3' non erano complanari, anche i punti O, A_1^*, A_2^* e A_3^* non lo sono, in quanto loro traslati; ne segue che i vettori $\{\overrightarrow{OA_1^*}, \overrightarrow{OA_2^*}, \overrightarrow{OA_3^*}\}$ sono una base dello spazio vettoriale \mathcal{V}_O^3 . Quanto visto nel Paragrafo 8.1 ci fornisce allora una matrice $B \in GL_3(\mathbb{R})$ — la matrice di cambiamento di base — tale che

$$\begin{vmatrix} \overrightarrow{OA_1^*} & \overrightarrow{OA_2^*} & \overrightarrow{OA_3^*} \end{vmatrix} = \begin{vmatrix} \overrightarrow{OA_1} & \overrightarrow{OA_2} & \overrightarrow{OA_3} \end{vmatrix} B.$$

Prendiamo allora il nostro punto $P \in \mathcal{A}^3$, di coordinate $x = (x_1, x_2, x_3)$ rispetto a \mathcal{R} , e $x' = (x'_1, x'_2, x'_3)$ rispetto a \mathcal{R}' . Il punto P viene traslato nel punto P^* tramite

il vettore $-\overrightarrow{OO'}$, in modo che

$$\overrightarrow{OP^*} = \overrightarrow{OP} - \overrightarrow{OO'}.$$

Dunque le coordinate di P^* rispetto a \mathcal{R}^* sono le stesse di P rispetto a \mathcal{R}' , per cui

$$\begin{aligned} \left| \overrightarrow{OA_1} \quad \overrightarrow{OA_2} \quad \overrightarrow{OA_3} \right| x &= x_1 \overrightarrow{OA_1} + x_2 \overrightarrow{OA_2} + x_3 \overrightarrow{OA_3} = \overrightarrow{OP} = \overrightarrow{OO'} + \overrightarrow{OP^*} \\ &= \overrightarrow{OO'} + x'_1 \overrightarrow{OA_1^*} + x'_2 \overrightarrow{OA_2^*} + x'_3 \overrightarrow{OA_3^*} \\ &= \left| \overrightarrow{OA_1} \quad \overrightarrow{OA_2} \quad \overrightarrow{OA_3} \right| c + \left| \overrightarrow{OA_1^*} \quad \overrightarrow{OA_2^*} \quad \overrightarrow{OA_3^*} \right| x' \\ &= \left| \overrightarrow{OA_1} \quad \overrightarrow{OA_2} \quad \overrightarrow{OA_3} \right| (c + Bx'). \end{aligned}$$

Quindi

$$x = Bx' + c, \quad (10.29)$$

cioè

$$\begin{cases} x_1 = b_{11}x'_1 + b_{12}x'_2 + b_{13}x'_3 + c_1, \\ x_2 = b_{21}x'_1 + b_{22}x'_2 + b_{23}x'_3 + c_2, \\ x_3 = b_{31}x'_1 + b_{32}x'_2 + b_{33}x'_3 + c_3. \end{cases} \quad (10.30)$$

Dunque un cambiamento di coordinate affini consiste in un'applicazione lineare seguita da una traslazione (combinazione che rivedremo nei Complementi a questo capitolo quando parleremo di affinità). Per completezza, riportiamo qui anche il cambiamento inverso, che è dato da

$$x' = B^{-1}x - B^{-1}c.$$

Osservazione 10.5 Fin qui abbiamo ragionato in \mathcal{A}^3 , ma ovviamente la stessa argomentazione si applica in \mathcal{A}^2 , con una coordinata in meno. La relazione (10.29) rimane identica, con $B \in GL_2(\mathbb{R})$ e $c \in \mathbb{R}^2$, mentre il sistema (10.30) perde una riga diventando

$$\begin{cases} x_1 = b_{11}x'_1 + b_{12}x'_2 + c_1, \\ x_2 = b_{21}x'_1 + b_{22}x'_2 + c_2. \end{cases}$$

10.8 Orientazione

La formula (10.29) ci permette di introdurre un'ulteriore definizione.

Definizione 10.10 Due sistemi di riferimento affine $\mathcal{R} = RA(O, A_1, A_2, A_3)$ e $\mathcal{R}' = RA(O', A'_1, A'_2, A'_3)$ hanno la *stessa orientazione* (diremo anche che sono *equiversi*) se la matrice B del cambiamento di coordinate affini da \mathcal{R} a \mathcal{R}' ha determinante positivo; diremo che \mathcal{R} e \mathcal{R}' hanno *orientazione opposta* (sono *contraversi*) se invece $\det B < 0$. La stessa definizione si applica nel caso del piano.

Per capire meglio di cosa stiamo parlando, fissiamo un sistema di riferimento affine $\mathcal{R}_1 = RA(O, A_1, A_2)$ in \mathcal{A}^2 . Rispetto a \mathcal{R} , la base $\{v_1 = \overrightarrow{OA_1}, v_2 = \overrightarrow{OA_2}\}$ di \mathcal{V}_O^2 ha come coordinate i vettori della base canonica $\{e_1, e_2\}$ di \mathbb{R}^2 . Nel disegno usuale, in cui le ascisse vengono messe in orizzontale dirette verso destra e le ordinate in verticale dirette verso l'alto, per passare da e_1 a e_2 bisogna effettuare una rotazione in senso antiorario.

Consideriamo invece il sistema di riferimento affine $\mathcal{R}_2 = RA(O, A_2, A_1)$, con la stessa origine³ ma con i vettori di base $\{v_1 = \overrightarrow{OA_2}, v_2 = \overrightarrow{OA_1}\}$ scambiati. Stavolta per passare dal primo al secondo vettore di base bisogna ruotare in senso orario. La matrice del cambiamento di coordinate è

$$B = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}$$

che ha determinante -1 ; quindi i due sistemi \mathcal{R}_1 ed \mathcal{R}_2 hanno orientazione opposta.

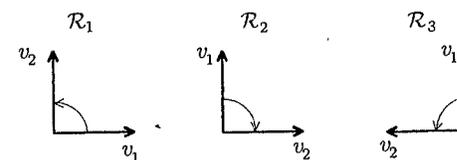


Figura 10.3 Orientazione in \mathcal{A}^2 .

Se ora in \mathcal{R}_2 sostituiamo A_1 col suo opposto rispetto a O otteniamo (vedi la Figura 10.3) un sistema di riferimento affine \mathcal{R}_3 con base $\{v_1 = \overrightarrow{OA_2}, v_2 = -\overrightarrow{OA_1}\}$. Le matrici di cambiamento di base B_1 da \mathcal{R}_1 a \mathcal{R}_3 e B_2 da \mathcal{R}_2 a \mathcal{R}_3 sono date da

$$B_1 = \begin{vmatrix} 0 & -1 \\ 1 & 0 \end{vmatrix} \quad \text{e} \quad B_2 = \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix};$$

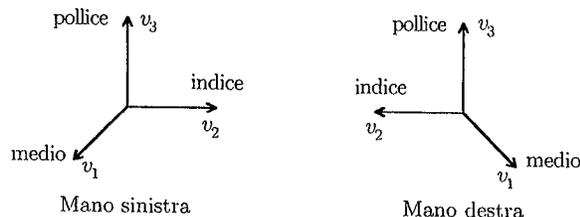
dunque \mathcal{R}_2 e \mathcal{R}_3 hanno orientazione opposta, mentre \mathcal{R}_1 e \mathcal{R}_3 hanno la stessa orientazione. E infatti per passare dal primo al secondo vettore di base in \mathcal{R}_3 bisogna ruotare di nuovo in senso antiorario.

Riassumendo, in \mathcal{A}^2 il succo del concetto di orientazione è: due sistemi di riferimento affine hanno la stessa orientazione se e solo se per passare dal primo al secondo vettore di base si effettua una rotazione che ha lo stesso senso in entrambi i casi⁴.

In \mathcal{A}^3 la situazione è analoga, anche se espressa in maniera lievemente diversa. Nei disegni usuali, la base canonica $\{e_1, e_2, e_3\}$ di \mathbb{R}^3 viene rappresentata nella stessa

³ O anche un'altra: è evidente che l'orientazione non dipende dall'origine scelta.

⁴ Sia chiaro che noi non abbiamo dimostrato questa affermazione; per farlo dovremmo definire rigorosamente il concetto di "senso di rotazione". Qui ci siamo limitati a far vedere come sia plausibile identificare orientazione e senso di rotazione.

Figura 10.4 Orientazione in \mathcal{A}^3 .

posizione in cui si trovano, nell'ordine, il medio, l'indice e il pollice della mano sinistra se cerchiamo di tenerli perpendicolari (vedi la Figura 10.4).

Allora due sistemi di riferimento affine in \mathcal{A}^3 hanno la stessa orientazione se e solo se i tre vettori della base dei due sistemi sono disposti come medio, indice e pollice (nell'ordine) della stessa mano, destra o sinistra che sia.

Esercizi

10.1 Dimostra che due equazioni cartesiane $ax + by + cz = d$ e $a'x + b'y + c'z = d'$ descrivono lo stesso piano se e solo se esiste $\lambda \in \mathbb{R}$ non nullo tale che $a' = \lambda a$, $b' = \lambda b$, $c' = \lambda c$ e $d' = \lambda d$.

10.2 Che insieme viene descritto da equazioni cartesiane del tipo $Ax = b$, dove $A \in GL_3(\mathbb{R})$, $x \in \mathbb{R}^3$ e $b \in \mathbb{R}^3$?

10.3 Dimostra che le equazioni parametriche

$$\begin{cases} x = 1 + t, \\ y = -2 - t, \\ z = 3 + 2t, \end{cases} \quad \text{e} \quad \begin{cases} x = 4 - 2s, \\ y = -5 + 2s, \\ z = 9 - 4s, \end{cases}$$

descrivono la stessa retta.

10.4 Se $ax + by = c$ è l'equazione cartesiana della retta r nel piano, dimostra che $\begin{vmatrix} -b \\ a \end{vmatrix}$ è un vettore direttore per r .

10.5 Se

$$\begin{cases} x = x_0 + lt, \\ y = y_0 + mt, \end{cases}$$

sono equazioni parametriche per una retta r nel piano, dimostra che

$$mx - ly = mx_0 - ly_0$$

è un'equazione cartesiana per r .

10.6 Se

$$\begin{cases} ax + by + cz = d, \\ a'x + b'y + c'z = d', \end{cases}$$

sono equazioni cartesiane per una retta r nello spazio, dimostra che

$$l = \det \begin{vmatrix} b & c \\ b' & c' \end{vmatrix}, \quad m = -\det \begin{vmatrix} a & c \\ a' & c' \end{vmatrix}, \quad n = \det \begin{vmatrix} a & b \\ a' & b' \end{vmatrix}$$

sono parametri direttori per r .

10.7 Se

$$\begin{cases} x = x_0 + sl + tl', \\ y = y_0 + sm + tm', \\ z = z_0 + sn + tn', \end{cases}$$

sono equazioni parametriche per un piano π , dimostra che

$$a = \det \begin{vmatrix} m & m' \\ n & n' \end{vmatrix}, \quad b = -\det \begin{vmatrix} l & l' \\ n & n' \end{vmatrix}, \quad c = \det \begin{vmatrix} l & l' \\ m & m' \end{vmatrix}$$

sono parametri di giacitura per π .

10.8 Dimostra che tre punti del piano $P_0 = (x_0, y_0)$, $P_1 = (x_1, y_1)$ e $P_2 = (x_2, y_2)$ sono allineati se e solo se

$$\det \begin{vmatrix} x_1 - x_0 & x_2 - x_0 \\ y_1 - y_0 & y_2 - y_0 \end{vmatrix} = 0.$$

(Suggerimento: i tre punti sono allineati se e solo se esistono $v \in \mathbb{R}^2$ e $t_1, t_2 \in \mathbb{R}$ tali che $P_j - P_0 = t_j v$ per $j = 1, 2$.)

10.9 Dimostra che i punti $P_0 = (x_0, y_0, z_0)$, $P_1 = (x_1, y_1, z_1)$ e $P_2 = (x_2, y_2, z_2)$ dello spazio sono allineati se e solo se

$$\operatorname{rg} \begin{vmatrix} x_1 - x_0 & x_2 - x_0 \\ y_1 - y_0 & y_2 - y_0 \\ z_1 - z_0 & z_2 - z_0 \end{vmatrix} \leq 1.$$

10.10 Dimostra che i punti $P_0 = (x_0, y_0, z_0)$, $P_1 = (x_1, y_1, z_1)$, $P_2 = (x_2, y_2, z_2)$ e $P_3 = (x_3, y_3, z_3)$ dello spazio sono complanari se e solo se

$$\det \begin{vmatrix} x_1 - x_0 & x_2 - x_0 & x_3 - x_0 \\ y_1 - y_0 & y_2 - y_0 & y_3 - y_0 \\ z_1 - z_0 & z_2 - z_0 & z_3 - z_0 \end{vmatrix} = 0.$$

10.11 Verifica che le definizioni di parallelismo fra rette e/o piani date nelle Definizioni 10.6, 10.8 e nel Paragrafo 10.5 sono casi particolari della Definizione 6.7 di parallelismo fra sottospazi affini.

10.12 L'insieme delle rette del piano parallele a una retta data r si chiama *fascio improprio di rette di asse r* . Se la retta r ha equazioni parametriche $P = P_0 + tv_0$, dimostra che le rette di equazioni parametriche $P = Q + tv_0$ al variare di $Q \in \mathbb{R}^2$ sono tutte e sole le rette del fascio improprio di asse r .

10.13 Se la retta r del piano ha equazioni cartesiane $ax + by = c$, dimostra che le rette di equazioni cartesiane $ax + by = k$ al variare di $k \in \mathbb{R}$ sono tutte e sole le rette del fascio improprio di asse r .

10.14 Dimostra che il sistema

$$\begin{cases} m(x - x_0) = l(y - y_0), \\ n(y - y_0) = m(z - z_0), \end{cases}$$

al variare di $l, m, n \in \mathbb{R}$ descrive tutte le rette della stella di centro $P_0 = (x_0, y_0, z_0)$ escluse quelle parallele all'asse z e all'asse x .

10.15 Nel calcolo delle equazioni parametriche di una retta per due punti nello spazio, se $x_1 = x_0$ il sistema (10.16) ci dice solo $l = 0$, lasciando m ed n liberi. Eppure anche in questo caso la retta passante per i due punti è unica, ed è data ancora da (10.17). Cosa sta succedendo, e perché?

10.16 Dimostra che la retta passante per i punti $P_0 = (\pi, 1, 3)$ e $P_1 = (\pi, 0, 2)$ dello spazio ha equazioni parametriche

$$\begin{cases} x = \pi, \\ y = 1 - t, \\ z = 3 - t, \end{cases}$$

ed equazioni cartesiane

$$\begin{cases} x = \pi, \\ y - z = -2. \end{cases}$$

10.17 Trova equazioni parametriche e cartesiane del piano in \mathcal{A}^3 passante per i punti $P_0 = (0, 1, 0)$, $P_1 = (-1, 0, 0)$ e $P_2 = (0, 0, 1)$.

10.18 Perché nelle tabelle del Paragrafo 10.4 due caselle descrivono situazioni impossibili?

10.19 Date le rette di equazioni parametriche

$$r_0: \begin{cases} x = 3 + t, \\ y = -1 - t, \\ z = 2 + 2t, \end{cases} \quad r_1: \begin{cases} x = 2 - t, \\ y = 1 + t, \\ z = 2 + t, \end{cases}$$

decidi se sono parallele, incidenti o sghembe.

10.20 Siano r_1 ed r_2 due rette del piano di equazioni cartesiane rispettivamente $a_1x + b_1y = c_1$ e $a_2x + b_2y = c_2$. Dimostra che l'insieme delle rette di equazione cartesiana

$$\lambda(a_1x + b_1y) + \mu(a_2x + b_2y) = \lambda c_1 + \mu c_2$$

al variare di $\lambda, \mu \in \mathbb{R}$, con $(\lambda, \mu) \neq (0, 0)$, è il fascio di centro il punto d'intersezione di r_1 ed r_2 se le due rette s'intersecano, o il fascio improprio di asse r_1 (o r_2) se le due rette sono parallele.

10.21 Date le rette di equazioni cartesiane

$$r_0: \begin{cases} x + y = 0, \\ x + 2y + z = 0, \end{cases} \quad r_1: \begin{cases} x - y = -1, \\ x + 5y + z = 0, \end{cases}$$

stabilisci se sono parallele, incidenti, sghembe o coincidenti.

10.22 Dati il piano π e la retta r di equazioni cartesiane

$$\pi: 2x + 3y - z = 4, \quad r: \begin{cases} x + y = 0, \\ x - z = 1, \end{cases}$$

determina se sono paralleli o incidenti e trova l'eventuale punto di intersezione.

10.23 L'insieme dei piani paralleli a un piano dato π si chiama *fascio improprio di giacitura* π . Supponiamo che π abbia equazioni parametriche $P = P_0 + sv_1 + tv_2$ ed equazione cartesiana $ax + by + cz = d$. Dimostra che i piani del fascio improprio di giacitura π sono descritti dalle equazioni parametriche $P = Q + sv_1 + tv_2$ al variare di $Q \in \mathbb{R}^3$, e dall'equazione cartesiana $ax + by + cz = k$ al variare di $k \in \mathbb{R}$.

10.24 Siano π_1 e π_2 due piani di equazioni cartesiane $a_1x + b_1y + c_1z = d_1$ e $a_2x + b_2y + c_2z = d_2$ rispettivamente. Dimostra che l'insieme dei piani di equazione cartesiana

$$\lambda(a_1x + b_1y + c_1z) + \mu(a_2x + b_2y + c_2z) = \lambda d_1 + \mu d_2$$

al variare di $\lambda, \mu \in \mathbb{R}$, con $(\lambda, \mu) \neq (0, 0)$, è il fascio di asse la retta d'intersezione di π_1 e π_2 se i due piani s'intersecano, o il fascio improprio di giacitura π_1 (o π_2) se i due piani sono paralleli.

10.25 Trova equazioni parametriche e cartesiane del piano in \mathcal{A}^3 passante per il punto $P_1 = (1, 0, 1)$ e contenente la retta di equazioni cartesiane

$$\begin{cases} x + y + z = 3, \\ x - y - z = 0. \end{cases}$$

10.26 Trova le equazioni cartesiane e parametriche del piano (se esiste) che contiene le rette di equazioni cartesiane

$$\begin{cases} 3x - y + z = 0, \\ x + y + z = 0, \end{cases} \quad \text{e} \quad \begin{cases} -x - y + 2z = 0, \\ x - z = 0. \end{cases}$$

10.27 Trova equazioni parametriche e cartesiane della retta passante per il punto P_0 di coordinate $(0, 7, 1)$ e parallela alla retta di equazioni

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

10.28 Sia fissato un sistema di riferimento affine \mathcal{R}_0 in \mathcal{A}^3 , e considera i punti di coordinate affini (rispetto a \mathcal{R}_0)

$$O' = \begin{vmatrix} 1 \\ 0 \\ 1 \end{vmatrix}, \quad A'_1 = \begin{vmatrix} 0 \\ 1 \\ 1 \end{vmatrix}, \quad A'_2 = \begin{vmatrix} 1 \\ 1 \\ 0 \end{vmatrix}, \quad A'_3 = \begin{vmatrix} 1 \\ 1 \\ 1 \end{vmatrix},$$

$$O'' = \begin{vmatrix} 0 \\ 2 \\ 0 \end{vmatrix}, \quad A''_1 = \begin{vmatrix} -1 \\ 1 \\ -1 \end{vmatrix}, \quad A''_2 = \begin{vmatrix} 1 \\ 0 \\ 2 \end{vmatrix}, \quad A''_3 = \begin{vmatrix} 1 \\ 0 \\ 0 \end{vmatrix}.$$

Dimostra che $\mathcal{R}' = RA(O', A'_1, A'_2, A'_3)$ e $\mathcal{R}'' = RA(O'', A''_1, A''_2, A''_3)$ sono sistemi di riferimento affini, trova la formula di cambiamento di coordinate affini da \mathcal{R}' a \mathcal{R}'' , e determina quali dei tre sistemi \mathcal{R}_0 , \mathcal{R}' ed \mathcal{R}'' hanno la stessa orientazione.

10.29 Diremo che due basi di uno spazio vettoriale V hanno la *stessa orientazione* se la matrice di cambiamento di base ha determinante positivo; che hanno *orientazione opposta* altrimenti. Dimostra che "avere la stessa orientazione" è una relazione d'equivalenza sull'insieme delle basi di V , e che l'insieme quoziente è composto da due soli elementi (vedi il Paragrafo 2C.2 per le necessarie definizioni).

COMPLEMENTI

10C.1 Affinità

Possiamo ora descrivere meglio lo scopo della geometria affine: consiste nello studio delle cosiddette *proprietà affini*, cioè di quelle proprietà dei sottoinsiemi di \mathcal{A}^3 (o \mathcal{A}^2 , o \mathcal{A}^1) che non dipendono dal sistema di riferimento affine usato per descriverli. Per esempio, "essere una retta", oppure "essere parallela a un dato piano" o "intersecare una data retta" sono tutte proprietà affini; "avere la terza coordinata nulla", invece, non è una proprietà affine (come puoi agevolmente verificare).

Una domanda abbastanza naturale a questo punto, stimolata anche dal fatto che finora ci siamo sistematicamente rifiutati di parlarne, è: la distanza fra due punti è una proprietà affine oppure no? Per rispondere, consideriamo il caso di due punti P_1 e P_2 in \mathcal{A}^1 , cioè sulla retta. Fissiamo un sistema di riferimento affine $\mathcal{R} = RA(O, A_1)$ in \mathcal{A}^1 in modo da poter scrivere

$$\overrightarrow{OP_1} = x_1 \overrightarrow{OA_1} \quad \text{e} \quad \overrightarrow{OP_2} = x_2 \overrightarrow{OA_1}.$$

Dunque in questo sistema di riferimento la "distanza" fra P_1 e P_2 è data da $|x_2 - x_1|$. In un altro sistema di riferimento affine $\mathcal{R}' = RA(O', A'_1)$ invece la "distanza" fra P_1 e P_2 sarà data da $|x'_2 - x'_1|$, dove stavolta x'_1 e x'_2 sono le coordinate affini di P_1 e P_2 rispetto a \mathcal{R}' . Per quanto visto nel Paragrafo 10.7 (adattato al caso di \mathcal{A}^1)

esistono $b \in \mathbb{R}^*$ e $c \in \mathbb{R}$ tali che

$$x_1 = bx'_1 + c \quad \text{e} \quad x_2 = bx'_2 + c.$$

Quindi non appena $|b| \neq 1$ le "distanze" $|x_2 - x_1|$ e $|x'_2 - x'_1|$ risultano diverse, per cui la distanza fra due punti *non* è una proprietà affine.

Questo risultato suggerisce però un'altra possibilità. Consideriamo anche un terzo punto P_3 , di coordinata x_3 rispetto a \mathcal{R} e x'_3 rispetto a \mathcal{R}' . Siccome abbiamo di nuovo $x_3 = bx'_3 + c$, vediamo subito che

$$\frac{x_3 - x_1}{x_2 - x_1} = \frac{x'_3 - x'_1}{x'_2 - x'_1}$$

non appena $P_2 \neq P_1$. Dunque il rapporto $(x_3 - x_1)/(x_2 - x_1)$ è una *proprietà affine* dei tre punti P_1, P_2 e P_3 presi in quest'ordine; diamogli un nome.

Definizione 10C.1 Siano $P_1, P_2, P_3 \in \mathcal{A}^1$ tre punti della retta. Il *rapporto semplice* $(P_1, P_2, P_3) \in \mathbb{R} \cup \{\infty\}$ di P_1, P_2 e P_3 (presi in quest'ordine e con $P_1 \neq P_2$) è

$$(P_1, P_2, P_3) = \frac{x_3 - x_1}{x_2 - x_1}, \quad (10C.1)$$

dove x_j è la coordinata di P_j (per $j = 1, 2, 3$) rispetto a un qualsiasi sistema di riferimento affine su \mathcal{A}^1 (infatti abbiamo appena visto che il quoziente (10C.1) non dipende dal sistema di riferimento in cui viene calcolato). Se invece $P_1 = P_2 \neq P_3$ poniamo $(P_1, P_2, P_3) = \infty$. Infine, il rapporto semplice (P_1, P_2, P_3) non è definito se $P_1 = P_2 = P_3$.

Un modo per capire il significato geometrico del rapporto semplice (e il motivo per cui non dipende dal sistema di riferimento) è contenuto nel seguente

Lemma 10C.1 Siano $O, P_1, P_2, P_3 \in \mathcal{A}^1$ quattro punti della retta, con $P_1 \neq P_2$. Allora si ha $(P_1, P_2, P_3) = t \in \mathbb{R}$ se e solo se

$$\overrightarrow{OP_3} = \overrightarrow{OP_1} + t(\overrightarrow{OP_2} - \overrightarrow{OP_1}).$$

Dimostrazione. Siccome $P_1 \neq P_2$, esiste un unico $t \in \mathbb{R}$ tale che

$$\overrightarrow{OP_3} - \overrightarrow{OP_1} = t(\overrightarrow{OP_2} - \overrightarrow{OP_1}); \quad (10C.2)$$

dobbiamo verificare che $t = (P_1, P_2, P_3)$. Fissiamo un sistema di riferimento affine $RA(O, A_1)$ rispetto a cui P_j ha coordinata x_j , per $j = 1, 2, 3$. Allora (10C.2)

diventa

$$(x_3 - x_1)\overrightarrow{OA_1} = t(x_2 - x_1)\overrightarrow{OA_1},$$

ed è fatta. \square

Studiando gli spazi vettoriali ci siamo soffermati a lungo sulle applicazioni lineari, cioè su quelle funzioni che conservano la struttura di spazio vettoriale (ovvero le operazioni di somma e prodotto per scalari). Analogamente ora siamo interessati alle funzioni che conservano le proprietà affini, come l'appartenenza o il rapporto semplice. Non esamineremo però la situazione nella sua massima generalità (in particolare, non considereremo il caso di spazi affini di dimensione maggiore di tre, o di applicazioni fra spazi affini di dimensione diversa), limitandoci invece a quanto descritto nelle seguenti definizioni:

Definizione 10C.2 Una *retta affine* (piano affine, spazio affine) è semplicemente una copia di \mathcal{A}^1 (rispettivamente, di \mathcal{A}^2 o \mathcal{A}^3). Per esempio, una qualunque retta contenuta in \mathcal{A}^3 è una retta affine.

Definizione 10C.3 Un' *affinità* fra due rette affini r ed r' è una funzione bigettiva $\varphi: r \rightarrow r'$ che conserva i rapporti semplici, cioè tale che

$$(\varphi(P_1), \varphi(P_2), \varphi(P_3)) = (P_1, P_2, P_3)$$

per qualunque tripla di punti $P_1, P_2, P_3 \in r$ non tutti coincidenti. Un' *affinità* fra due piani affini π e π' è una funzione bigettiva $\varphi: \pi \rightarrow \pi'$ che manda rette in rette e tale che ristretta a ogni retta $r \subset \pi$ sia un' *affinità* fra r e la sua immagine $\varphi(r)$. In particolare φ conserva i rapporti semplici di punti allineati. Infine, un' *affinità* fra due spazi affini Σ e Σ' è una funzione bigettiva $\varphi: \Sigma \rightarrow \Sigma'$ che manda piani in piani e tale che ristretta a ogni piano $\pi \subset \Sigma$ sia un' *affinità* fra π e la sua immagine $\varphi(\pi)$. In particolare, φ manda rette in rette e conserva i rapporti semplici di punti allineati.

Dalla definizione risulta chiaro che le affinità conservano alcune delle proprietà affini (l'appartenenza, il rapporto semplice, l'essere una retta); vedremo fra un istante che le conservano proprio tutte, giustificando quindi il nostro interesse per loro.

Cominciamo con le affinità fra rette.

Proposizione 10C.2 Sia $\varphi: r \rightarrow r'$ un' *affinità* fra le rette affini r ed r' . Fissiamo un sistema di riferimento affine $\mathcal{R} = RA(O, A_1)$ su r , e uno $\mathcal{R}' = RA(O', A'_1)$ su r' . Allora esistono $b \in \mathbb{R}^*$ e $c \in \mathbb{R}$ tali che

$$\forall P \in r \quad y = bx + c, \quad (10C.3)$$

dove x è la coordinata di P rispetto a \mathcal{R} e y è la coordinata di $\varphi(P)$ rispetto a \mathcal{R}' . In particolare, la funzione φ conserva tutte le proprietà affini. Viceversa, ogni $\varphi: r \rightarrow r'$ che in coordinate è descritta da (10C.3) con $b \neq 0$ è un' *affinità*.

Dimostrazione. Sia $c \in \mathbb{R}$ la coordinata di $\varphi(O)$ rispetto a \mathcal{R}' , e $d \in \mathbb{R}$ la coordinata di $\varphi(A_1)$ rispetto a \mathcal{R}' . Essendo φ iniettiva, il numero $b = d - c$ è diverso da zero.

Prendiamo ora $P \in r$ qualunque; siccome φ conserva i rapporti semplici, si ha

$$(O, A_1, P) = (\varphi(O), \varphi(A_1), \varphi(P)),$$

ovvero, usando il riferimento \mathcal{R} in r e il riferimento \mathcal{R}' in r' ,

$$\frac{x-0}{1-0} = \frac{y-c}{d-c},$$

cioè $y = bx + c$, come voluto. In particolare, siccome (10C.3) ha la stessa espressione di un cambiamento di coordinate affini, e le proprietà affini sono esattamente quelle conservate da tali cambiamenti di coordinate, ne segue che φ conserva tutte le proprietà affini.

Infine, una $\varphi: r \rightarrow r'$ data in coordinate da (10C.3) è chiaramente bigettiva (in quanto $b \neq 0$) e conserva i rapporti semplici (esercizio), per cui è un' *affinità*. \square

Osservazione 10C.1 Attenzione a non confondere affinità e cambiamenti di coordinate, nonostante abbiano la stessa espressione. In un' *affinità*, i sistemi di riferimento sono fissi e sono i punti che si muovono: il punto P di coordinata x va a finire nel punto $\varphi(P)$ di coordinata y . Invece, nei cambiamenti di coordinate i punti sono fermi e cambia il sistema di riferimento: il punto P ha coordinata x rispetto a \mathcal{R} , e coordinata x' rispetto a \mathcal{R}' .

Passiamo ora alle affinità del piano. Ci serve un lemma:

Lemma 10C.3 Sia $\varphi: \pi \rightarrow \pi'$ un' *affinità* fra piani affini, e $O, A_1, A_2 \in \pi$ tre punti non allineati. Allora anche $\varphi(O), \varphi(A_1), \varphi(A_2) \in \pi'$ sono non allineati.

Dimostrazione. Supponiamo per assurdo che $\varphi(O), \varphi(A_1)$ e $\varphi(A_2)$ siano allineati, e consideriamo la retta r passante per O e A_1 . Ora, φ manda rette in rette; quindi $\varphi(r)$ è la retta che contiene $\varphi(O)$ e $\varphi(A_1)$. In particolare, $\varphi(A_2) \in \varphi(r)$, per cui esiste $P \in r$ tale che $\varphi(P) = \varphi(A_2)$. Ma O, A_1 e A_2 non sono allineati; quindi $A_2 \notin r$, e dunque $P \neq A_2$ mentre per ipotesi φ è iniettiva, contraddizione. \square

Proposizione 10C.4 Sia $\varphi: \pi \rightarrow \pi'$ un' *affinità* fra i piani affini π e π' . Fissiamo un sistema di riferimento affine $\mathcal{R} = RA(O, A_1, A_2)$ su π , e uno $\mathcal{R}' = RA(O', A'_1, A'_2)$ su π' . Allora esistono $B \in GL_2(\mathbb{R})$ e $c \in \mathbb{R}^2$ tali che

$$\forall P \in \pi \quad y = Bx + c, \quad (10C.4)$$

dove $x = (x_1, x_2)$ sono le coordinate di P rispetto a \mathcal{R} e $y = (y_1, y_2)$ le coordinate di $\varphi(P)$ rispetto a \mathcal{R}' . In particolare, φ conserva tutte le proprietà affini. Viceversa, ogni $\varphi: \pi \rightarrow \pi'$ che in coordinate è descritta da (10C.4) con $B \in GL_2(\mathbb{R})$ è un' *affinità*.

Dimostrazione. Siano (c_1, c_2) le coordinate di $\varphi(O)$ rispetto a \mathcal{R}' , e (d_{1j}, d_{2j}) le coordinate di $\varphi(A_j)$ rispetto a \mathcal{R}' , per $j = 1, 2$.

Un punto $P_1 \in \pi$ di coordinate $(x_1, 0)$ sta sulla retta passante per O e A_1 ; quindi $\varphi(P_1)$ deve stare sulla retta passante per $\varphi(O)$ e $\varphi(A_1)$. Siccome il rapporto semplice fra O, A_1 e P_1 è proprio x_1 , si deve avere $(\varphi(O), \varphi(A_1), \varphi(P_1)) = x_1$. Il Lemma 10C.1 allora ci dice che $\varphi(P_1)$ deve avere coordinate

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + x_1 \begin{bmatrix} d_{11} - c_1 \\ d_{21} - c_2 \end{bmatrix}.$$

Analogamente un punto di coordinate $(0, x_2)$ ha per immagine un punto di coordinate

$$\begin{vmatrix} c_1 \\ c_2 \end{vmatrix} + x_1 \begin{vmatrix} d_{12} - c_1 \\ d_{22} - c_2 \end{vmatrix}.$$

Ora, l'affinità φ manda rette in rette per ipotesi, e conserva il parallelismo in quanto bigettiva; dunque in particolare manda parallelogrammi in parallelogrammi. Se $P \in \pi$ ha coordinate (x_1, x_2) allora è il quarto vertice del parallelogramma determinato da O e dai punti P_1 di coordinate $(x_1, 0)$ e P_2 di coordinate $(0, x_2)$. Ne segue che $\varphi(P)$ è il quarto vertice del parallelogramma di vertici $\varphi(O)$, $\varphi(P_1)$ e $\varphi(P_2)$, cioè

$$\overrightarrow{\varphi(O)\varphi(P)} = \overrightarrow{\varphi(O)\varphi(P_1)} + \overrightarrow{\varphi(O)\varphi(P_2)}.$$

In coordinate questo diventa (perché?)

$$\begin{vmatrix} y_1 \\ y_2 \end{vmatrix} - \begin{vmatrix} c_1 \\ c_2 \end{vmatrix} = \left\{ \begin{vmatrix} c_1 \\ c_2 \end{vmatrix} + x_1 \begin{vmatrix} d_{11} - c_1 \\ d_{21} - c_2 \end{vmatrix} - \begin{vmatrix} c_1 \\ c_2 \end{vmatrix} \right\} + \left\{ \begin{vmatrix} c_1 \\ c_2 \end{vmatrix} + x_2 \begin{vmatrix} d_{12} - c_1 \\ d_{22} - c_2 \end{vmatrix} - \begin{vmatrix} c_1 \\ c_2 \end{vmatrix} \right\},$$

cioè

$$\begin{vmatrix} y_1 \\ y_2 \end{vmatrix} = \begin{vmatrix} d_{11} - c_1 & d_{12} - c_1 \\ d_{21} - c_2 & d_{22} - c_2 \end{vmatrix} \begin{vmatrix} x_1 \\ x_2 \end{vmatrix} + \begin{vmatrix} c_1 \\ c_2 \end{vmatrix}.$$

Ponendo

$$B = \begin{vmatrix} d_{11} - c_1 & d_{12} - c_1 \\ d_{21} - c_2 & d_{22} - c_2 \end{vmatrix},$$

abbiamo dimostrato (10C.4), perché l'Esercizio 10.8 implica $B \in GL_2(\mathbb{R})$ in quanto $\varphi(O)$, $\varphi(A_1)$ e $\varphi(A_2)$ non sono allineati grazie al Lemma 10C.3.

Il fatto che φ conserva tutte le proprietà affini si dimostra esattamente come nel caso delle affinità fra rette. Infine, se $\varphi: \pi \rightarrow \pi'$ è data in coordinate da (10C.4), allora φ è bigettiva perché $B \in GL_2(\mathbb{R})$; manda ovviamente rette in rette; e ristretta a qualunque retta si esprime in coordinate nella forma (10C.3). Quindi φ è un'affinità fra i due piani. \square

Con considerazioni analoghe si risolve anche il caso dello spazio:

Proposizione 10C.5 Sia $\varphi: \Sigma \rightarrow \Sigma'$ un'affinità fra gli spazi affini Σ e Σ' . Fissiamo un sistema di riferimento affine $\mathcal{R} = RA(O, A_1, A_2, A_3)$ su Σ , e un secondo $\mathcal{R}' = RA(O', A'_1, A'_2, A'_3)$ su Σ' . Allora esistono $B \in GL_3(\mathbb{R})$ e $c \in \mathbb{R}^3$ tali che

$$\forall P \in \Sigma \quad y = Bx + c, \quad (10C.5)$$

dove $x = (x_1, x_2, x_3)$ sono le coordinate di P rispetto a \mathcal{R} , e $y = (y_1, y_2, y_3)$ sono le coordinate di $\varphi(P)$ rispetto a \mathcal{R}' . In particolare, φ conserva tutte le proprietà affini. Viceversa, ogni $\varphi: \Sigma \rightarrow \Sigma'$ che in coordinate è descritta da (10C.5) con $B \in GL_3(\mathbb{R})$ è un'affinità.

Dimostrazione. Si procede come per la Proposizione 10C.4, usando parallelepipedi al posto dei parallelogrammi. \square

10C.2 Geometria proiettiva

In questo paragrafo vogliamo introdurre i concetti di base della geometria proiettiva, con l'obiettivo di mostrare come nascono in modo naturale dalla geometria affine.

La geometria proiettiva, come dice il nome, ha a che fare con le proiezioni; cominciamo allora a studiare queste ultime, nel caso del piano affine.

Definizione 10C.4 Siano r_1 e $r_2 \subset \mathcal{A}^2$ due rette distinte del piano, e $r_0 \subset \mathcal{A}^2$ una terza retta, incidente sia r_1 che r_2 . La *proiezione parallela* di direzione r_0 da r_1 a r_2 è l'applicazione $\psi_{r_0}: r_1 \rightarrow r_2$ che a ogni punto $P \in r_1$ associa l'intersezione $\psi_{r_0}(P) \in r_2$ di r_2 con la retta passante per P e parallela a r_0 .

Definizione 10C.5 Siano r_1 e $r_2 \subset \mathcal{A}^2$ due rette distinte del piano, e $P_0 \in \mathcal{A}^2$ un punto che non appartiene né a r_1 né a r_2 . La *proiezione centrale* di centro P_0 da r_1 a r_2 è l'applicazione $\psi_{P_0}: r_1 \rightarrow r_2$ che a ogni punto⁵ $P \in r_1$ associa l'intersezione $\psi_{P_0}(P) \in r_2$ di r_2 con la retta passante per P_0 e P .

Vediamo come si esprimono in coordinate queste proiezioni. Fissiamo un sistema di riferimento affine $\mathcal{R} = RA(O, \vec{i}, \vec{j})$ in \mathcal{A}^2 , un sistema di riferimento affine $\mathcal{R}_1 = RA(O_1, A_1)$ in r_1 e un altro $\mathcal{R}_2 = RA(O_2, A_2)$ in r_2 . Per $j = 1, 2$, indichiamo con

$$O_j \doteq \begin{vmatrix} c_{1j} \\ c_{2j} \end{vmatrix}, \quad A_j = \begin{vmatrix} d_{1j} \\ d_{2j} \end{vmatrix}$$

le coordinate rispetto a \mathcal{R} dei punti che compongono il riferimento \mathcal{R}_j . In particolare, un vettore direttore di r_j ha coordinate

$$\begin{vmatrix} e_{1j} \\ e_{2j} \end{vmatrix} = \begin{vmatrix} d_{1j} - c_{1j} \\ d_{2j} - c_{2j} \end{vmatrix},$$

per cui un punto $P \in r_1$ di coordinata x rispetto a \mathcal{R}_1 avrà coordinate

$$\begin{vmatrix} c_{11} + e_{11}x \\ c_{21} + e_{21}x \end{vmatrix}$$

rispetto a \mathcal{R} , e un punto $Q \in r_2$ di coordinata y rispetto a \mathcal{R}_2 avrà coordinate

$$\begin{vmatrix} c_{12} + e_{12}y \\ c_{22} + e_{22}y \end{vmatrix}$$

rispetto a \mathcal{R} .

Iniziamo con lo studiare la proiezione parallela ψ_{r_0} . Indichiamo con

$$v = \begin{vmatrix} v_1 \\ v_2 \end{vmatrix}$$

⁵ O meglio, a quasi ogni punto; come vedremo, c'è un problema, e per risolverlo saremo costretti a considerare la geometria proiettiva.

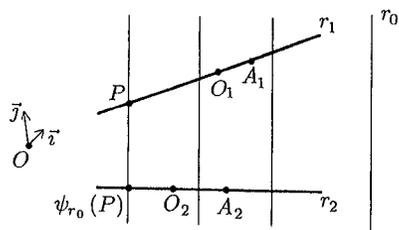


Figura 10.5 La proiezione parallela.

le coordinate rispetto a \mathcal{R} di un vettore direttore di r_0 . Preso (vedi la Figura 10.5) un punto $P \in r_1$ di coordinata x (rispetto a \mathcal{R}_1), i punti della retta per P parallela a r_0 hanno coordinate

$$\begin{cases} c_{11} + e_{11}x + tv_1 \\ c_{21} + e_{21}x + tv_2 \end{cases}$$

al variare di $t \in \mathbb{R}$, per cui la coordinata y (rispetto a \mathcal{R}_2) del punto $\psi_{r_0}(P)$ deve soddisfare il sistema

$$\begin{cases} c_{11} + e_{11}x + tv_1 = c_{12} + e_{12}y, \\ c_{21} + e_{21}x + tv_2 = c_{22} + e_{22}y, \end{cases}$$

ovvero

$$\begin{cases} e_{12}y - v_1t = (c_{11} - c_{12}) + e_{11}x, \\ e_{22}y - v_2t = (c_{21} - c_{22}) + e_{21}x. \end{cases}$$

Siccome a noi interessa solo il valore di y , possiamo applicare il Teorema di Cramer (Corollario 9.11) ottenendo

$$y = \frac{(v_1e_{21} - v_2e_{11})x + [v_1(c_{21} - c_{22}) - v_2(c_{11} - c_{12})]}{v_1e_{22} - v_2e_{12}};$$

nota che $v_1e_{21} - v_2e_{11}$, $v_1e_{22} - v_2e_{12} \neq 0$ in quanto r_0 non è parallela né a r_1 né a r_2 . Dunque in coordinate la proiezione parallela è data da una formula del tipo

$$x \mapsto y = \alpha x + \beta$$

con $\alpha \neq 0$, per cui (Proposizione 10C.2) è un'affinità fra le due rette.

Consideriamo invece (vedi la Figura 10.6) la proiezione centrale ψ_{P_0} di centro il punto P_0 di coordinate

$$P_0 = \begin{vmatrix} x_1 \\ x_2 \end{vmatrix}$$

rispetto a \mathcal{R} . Preso un punto $P \in r_1$ di coordinata x (rispetto a \mathcal{R}_1), i punti della retta per P e P_0 hanno coordinate

$$\begin{vmatrix} x_1 + t(c_{11} + e_{11}x - x_1) \\ x_2 + t(c_{21} + e_{21}x - x_2) \end{vmatrix}$$

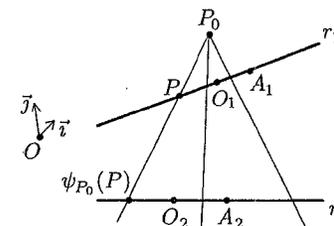


Figura 10.6 La proiezione centrale.

al variare di $t \in \mathbb{R}$, per cui la coordinata y (rispetto a \mathcal{R}_2) del punto $\psi_{P_0}(P)$ deve soddisfare il sistema

$$\begin{cases} e_{12}y - (c_{11} - x_1 + e_{11}x)t = x_1 - c_{12}, \\ e_{22}y - (c_{21} - x_2 + e_{21}x)t = x_2 - c_{22}. \end{cases}$$

Siccome a noi interessa solo il valore di y , possiamo di nuovo applicare il Teorema di Cramer ottenendo

$$y = \frac{[e_{11}(x_2 - c_{22}) - e_{21}(x_1 - c_{12})]x + [(x_1 - c_{12})(x_2 - c_{21}) - (x_1 - c_{11})(x_2 - c_{22})]}{(e_{11}e_{22} - e_{12}e_{21})x + [e_{12}(x_2 - c_{21}) - e_{22}(x_1 - c_{11})]}.$$

Dunque la proiezione centrale è data in coordinate da

$$x \mapsto y = \frac{\alpha x + \beta}{\gamma x + \delta}, \quad (10C.6)$$

per cui non è un'affinità⁶; vogliamo studiarne le proprietà, nel caso più interessante in cui r_1 ed r_2 non siano parallele.

Prima di tutto, la presenza di un denominatore in (10C.6) ci avvisa che la proiezione centrale non è definita su tutta la retta r_1 . Geometricamente era prevedibile: siccome r_1 ed r_2 non sono parallele, esiste un punto $P_1^\infty \in r_1$ (di coordinata $x = -\delta/\gamma$; esercizio) tale che la retta passante per lui e P_0 sia parallela a r_2 , per cui $\psi_{P_0}(P_1^\infty)$ non esiste.

Dunque la proiezione centrale è definita soltanto su $r_1 \setminus \{P_1^\infty\}$; vediamo se è almeno iniettiva. Di nuovo, geometricamente è immediato: supponiamo che $P_1, P_2 \in r_1$ siano tali che $\psi_{P_0}(P_1) = Q = \psi_{P_0}(P_2)$. Siccome $P_0 \neq Q$, la retta s passante per P_0 e Q è unica, e interseca r_1 sia in P_1 che in P_2 . Se P_1 fosse diverso da P_2 , questo implicherebbe $s = r_1$, e quindi $P_0 \in r_1$ contro l'ipotesi; quindi $P_1 = P_2$ e ψ_{P_0} è iniettiva.

Rimane da vedere se ψ_{P_0} è surgettiva. Siccome r_1 ed r_2 non sono parallele, esiste un punto $P_2^\infty \in r_2$ (di coordinata $y = \alpha/\gamma$; esercizio) tale che la retta per P_0 e P_2^∞ sia parallela a r_1 ; quindi P_2^∞ non può essere immagine di alcun punto di r_1 .

⁶ A meno che si abbia $\gamma = 0$, che accade se e solo se r_1 ed r_2 sono parallele.

Tutto ciò è piuttosto fastidioso. Siamo partiti da una tranquilla operazione geometrica, e ci siamo trovati costretti a tener presente tutta una serie di eccezioni. Per vedere se è possibile evitarlo, esaminiamo più accuratamente la situazione. Prendiamo un punto $P \in r_1 \setminus \{P_1^\infty\}$, e facciamolo avvicinare a P_1^∞ ; l'immagine $\psi_{P_0}(P)$ si allontana sempre di più su r_2 , fino a fuggire all'infinito quando P raggiunge P_1^∞ . Analogamente, se $Q \in r_2$ l'intersezione fra r_1 e la retta per P_0 e Q tende ad allontanarsi all'infinito ogni volta che Q si avvicina a P_2^∞ . Questo suggerisce di eliminare tutte le eccezioni aggiungendo un "punto all'infinito" ∞_j a ciascuna delle rette r_j per poi definire $\psi_{P_0}(P_1^\infty) = \infty_2$ e $\psi_{P_0}(\infty_1) = P_2^\infty$. In questo modo otteniamo una vera applicazione bigettiva da $\hat{r}_1 = r_1 \cup \{\infty_1\}$ a $\hat{r}_2 = r_2 \cup \{\infty_2\}$.

Dunque ci siamo trovati a dover mettere un punto in più a ciascuna retta, un punto che si raggiunge andando all'infinito (in una qualunque delle due direzioni); in un certo senso, è come considerare la nostra retta una sorta di circonferenza di raggio infinito. Per capire meglio la struttura dell'oggetto che abbiamo costruito, esaminiamo (per esempio) r_1 supponendo per semplicità che \mathcal{R} sia stato scelto in modo che la retta abbia equazione $y = 1$.

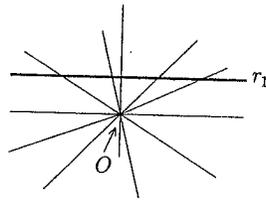


Figura 10.7 Le coordinate omogenee.

Ogni punto $P \in r_1$ determina una e una sola retta del fascio \mathcal{F} di centro l'origine O ; viceversa, ogni retta di \mathcal{F} (tranne guarda caso, una sola, quella parallela a r_1) interseca r_1 esattamente in un punto. Dunque abbiamo una corrispondenza biunivoca fra i punti di r_1 e le rette del fascio \mathcal{F} tranne una; associando quest'ultima al punto all'infinito ∞_1 otteniamo una perfetta corrispondenza biunivoca fra \hat{r}_1 ed \mathcal{F} (vedi la Figura 10.7). Ora, una retta del fascio ha equazione parametrica

$$\begin{cases} x = tv_1, \\ y = tv_2, \end{cases} \quad (10C.7)$$

con $(v_1, v_2) \neq (0, 0)$; due equazioni

$$\begin{cases} x = tv_1, \\ y = tv_2, \end{cases} \quad \begin{cases} x = tv'_1, \\ y = tv'_2, \end{cases}$$

descrivono la stessa retta se e solo se $v'_1 = \lambda v_1$ e $v'_2 = \lambda v_2$ per qualche $\lambda \neq 0$. Inoltre, la retta di equazione (10C.7) interseca r_1 nel punto $(v_1/v_2, 1)$ tranne quando $v_2 = 0$. caso in cui abbiamo la retta $y = 0$ che "interseca" r_1 nel punto all'infinito ∞_1 .

Possiamo esprimere queste osservazioni in modo lievemente diverso. Diremo che due vettori $v, w \in \mathbb{R}^2 \setminus \{O\}$ sono equivalenti, e scriveremo $v \sim w$, se esiste un $\lambda \in \mathbb{R}^*$ tale che $w = \lambda v$ (e quindi $v = \lambda^{-1}w$). In altre parole, $v \sim w$ se e solo se appartengono alla stessa retta per l'origine. Si verifica facilmente che \sim è una relazione d'equivalenza (vedi il Paragrafo 2C.2), e chiaramente c'è una corrispondenza biunivoca fra il fascio \mathcal{F} e l'insieme quoziente $\mathbb{R}^2 \setminus \{O\} / \sim$ (e quindi fra questo insieme ed \hat{r}_1).

Definizione 10C.6 L'insieme quoziente $\mathbb{P}^1(\mathbb{R}) = \mathbb{R}^2 \setminus \{O\} / \sim$ si chiama *retta proiettiva*. Se $\pi: \mathbb{R}^2 \setminus \{O\} \rightarrow \mathbb{P}^1(\mathbb{R})$ è l'applicazione quoziente e $(v_1, v_2) \in \mathbb{R}^2 \setminus \{O\}$, scriveremo $[v_1: v_2]$ al posto di $\pi(v_1, v_2)$; la coppia $(v_1, v_2) \neq (0, 0)$ contiene le *coordinate omogenee* del punto $[v_1: v_2] \in \mathbb{P}^1(\mathbb{R})$. Chiaramente, $[v_1: v_2] = [\lambda v_1: \lambda v_2]$ per ogni $\lambda \in \mathbb{R}^*$. Se $v_2 \neq 0$, il rapporto v_1/v_2 viene detto *coordinate non omogenea* del punto *proprio* $[v_1: v_2]$; il punto $[1: 0]$ (l'unico per cui $v_2 = 0$) è detto *punto all'infinito* (o *punto improprio*) della retta proiettiva, e verrà indicato con ∞ .

Osservazione 10C.2 Sia $f: \mathbb{R} \rightarrow \mathbb{P}^1(\mathbb{R})$ l'applicazione data da $f(x) = [x: 1]$. Si vede subito che f è iniettiva e che $f(\mathbb{R}) = \mathbb{P}^1(\mathbb{R}) \setminus \{\infty\}$. Quindi in un certo senso la retta proiettiva può venire considerata come la retta reale a cui si è aggiunto un punto.

Nella corrispondenza biunivoca fra \hat{r}_1 e $\mathbb{P}^1(\mathbb{R})$ ottenuta passando tramite il fascio \mathcal{F} al punto $(x, 1) \in r_1$ viene associato il punto $[x: 1]$, e al punto all'infinito il punto di coordinate omogenee $[1: 0]$; quindi $\mathbb{P}^1(\mathbb{R})$ riproduce esattamente la struttura dell'insieme \hat{r}_1 . In particolare, la coordinata affine x non è altro che la coordinata non omogenea della retta proiettiva; se vogliamo considerare tutti i punti di \hat{r}_1 allo stesso modo, senza fastidiose eccezioni, conviene quindi utilizzare coordinate omogenee.

Torniamo ora alla nostra proiezione centrale $\psi_{P_0}: \hat{r}_1 \rightarrow \hat{r}_2$, che abbiamo visto essere espressa in coordinate non omogenee dalla (10C.6). Se indichiamo con $[x_1: x_2]$ le coordinate omogenee in \hat{r}_1 e con $[y_1: y_2]$ le coordinate omogenee in \hat{r}_2 , la (10C.6) può venire scritta

$$\frac{y_1}{y_2} = \frac{\alpha(x_1/x_2) + \beta}{\gamma(x_1/x_2) + \delta} = \frac{\alpha x_1 + \beta x_2}{\gamma x_1 + \delta x_2},$$

ovvero

$$\psi_{P_0}([x_1: x_2]) = [\alpha x_1 + \beta x_2: \gamma x_1 + \delta x_2], \quad (10C.8)$$

e questa espressione dà il risultato corretto (controlla) anche per i punti all'infinito! Inoltre, se consideriamo la matrice (non singolare; esercizio)

$$A = \begin{vmatrix} \alpha & \beta \\ \gamma & \delta \end{vmatrix},$$

vediamo che

$$\forall v \in \mathbb{R}^2 \setminus \{O\} \quad \psi_{P_0}(\pi(v)) = \pi(L_A(v)),$$

dove $\pi: \mathbb{R}^2 \setminus \{O\} \rightarrow \mathbb{P}^1(\mathbb{R})$ è l'applicazione quoziente. In altre parole, la proiezione centrale è ottenibile tramite un'applicazione lineare invertibile di \mathbb{R}^2 .

Definizione 10C.7 Una *proiettività* $\psi: \mathbb{P}^1(\mathbb{R}) \rightarrow \mathbb{P}^1(\mathbb{R})$ della retta proiettiva è un'applicazione della forma (10C.8) con $\alpha\delta - \beta\gamma \neq 0$, cioè tale che $A \in GL_2(\mathbb{R})$.

Osservazione 10C.3 Le affinità possono essere considerate come particolari proiettività: quelle della forma (10C.8) con $\gamma = 0$, cioè tali da mandare il punto all'infinito in se stesso.

Dunque abbiamo completamente chiarito la struttura di una proiezione centrale: è una proiettività. Vogliamo ora esaminare brevemente che proprietà geometriche possiede. Abbiamo visto che le affinità conservano i rapporti semplici; e le proiettività? Sia $\psi: \mathbb{P}^1(\mathbb{R}) \rightarrow \mathbb{P}^1(\mathbb{R})$ una proiettività, e consideriamo tre punti propri P_1, P_2 e P_3 , di coordinate non omogenee x_1, x_2 e x_3 rispettivamente. Supponiamo per il momento anche che le immagini dei P_j siano punti propri, in modo da poter utilizzare (10C.6). Allora un conto veloce mostra che

$$(\psi(P_1), \psi(P_2), \psi(P_3)) = \frac{\gamma x_2 + \delta}{\gamma x_3 + \delta} (P_1, P_2, P_3).$$

Dunque il rapporto semplice non è conservato, ma viene moltiplicato per un fattore. Viene dunque la tentazione di prendere un quarto punto P_4 e considerare qualcosa del tipo

$$\frac{(P_1, P_3, P_4)}{(P_2, P_3, P_4)} = \frac{x_4 - x_2}{x_3 - x_2} \frac{x_4 - x_1}{x_3 - x_1}. \quad (10C.9)$$

Se riscriviamo questa formula in coordinate omogenee otteniamo la seguente

Definizione 10C.8 Siano $P_1, P_2, P_3, P_4 \in \mathbb{P}^1(\mathbb{R})$ quattro punti distinti. Supponiamo che le coordinate omogenee di P_j siano $v_j \in \mathbb{R}^2 \setminus \{O\}$, per $j = 1, \dots, 4$. Allora il *birapporto* $[P_1, P_2, P_3, P_4]$ dei quattro punti (nell'ordine) è

$$[P_1, P_2, P_3, P_4] = \frac{\det|v_2 v_4|}{\det|v_2 v_3|} \frac{\det|v_1 v_4|}{\det|v_1 v_3|}. \quad (10C.10)$$

Poniamo poi

$$\begin{aligned} [P_1, P_1, P_3, P_4] &= [P_1, P_2, P_3, P_3] = 1, \\ [P_1, P_2, P_1, P_4] &= [P_1, P_2, P_3, P_2] = 0, \\ [P_1, P_2, P_3, P_1] &= [P_1, P_2, P_2, P_4] = \infty; \end{aligned}$$

in questo modo il birapporto è definito per ogni quaterna di punti in cui non ve ne siano tre uguali.

Osservazione 10C.4 Utilizzare le coordinate omogenee per definire degli oggetti sulla retta proiettiva richiede una certa attenzione. Il punto è che se $v_0 \in \mathbb{R}^2 \setminus \{O\}$ ci dà coordinate omogenee per il punto $P_0 \in \mathbb{P}^1(\mathbb{R})$ e definiamo qualcosa utilizzando le coordinate di v_0 , questo qualcosa non deve cambiare se moltiplichiamo tutte le coordinate per un numero $\lambda \neq 0$, in quanto anche il vettore λv_0 fornisce coordinate omogenee

dello stesso punto P_0 . Nel caso in questione, se per $j = 1, \dots, 4$ prendiamo coordinate omogenee v_j dei punti P_j e scegliamo dei numeri $\lambda_j \neq 0$, la formula (10C.10) deve dare lo stesso risultato sia usando v_j che usando $\lambda_j v_j$ come coordinate omogenee di P_j . Fortunatamente, abbiamo

$$\det|\lambda_h v_h \lambda_k v_k| = \lambda_h \lambda_k \det|v_h v_k|,$$

e quindi (10C.10) fornisce effettivamente lo stesso risultato qualsiasi valore assumano i $\lambda_j \in \mathbb{R}^*$. Invece, il "rapporto semplice" $\det|v_1 v_3| / \det|v_1 v_2|$ non è ben definito su $\mathbb{P}^1(\mathbb{R})$ in quanto

$$\frac{\det|\lambda_1 v_1 \lambda_3 v_3|}{\det|\lambda_1 v_1 \lambda_2 v_2|} = \frac{\lambda_3}{\lambda_2} \frac{\det|v_1 v_3|}{\det|v_1 v_2|} \neq \frac{\det|v_1 v_3|}{\det|v_1 v_2|}$$

non appena $\lambda_2 \neq \lambda_3$.

La (10C.10) scritta in coordinate non omogenee si riduce alla (10C.9), per cui è molto ragionevole che il birapporto sia conservato da qualsiasi proiettività. Infatti la (10C.8) ci dice che se v_j sono le coordinate omogenee di P_j , allora le coordinate omogenee di $\psi(P_j)$ sono date da Av_j . Siccome

$$\det|Av_h Av_k| = (\det A) \det|v_h v_k|$$

(esercizio), il birapporto è conservato.

Le proprietà del birapporto sono molte e mirabili; per esempio, non è difficile dimostrare (esercizio sulla falsariga della Proposizione 10C.2) che un'applicazione bigettiva $\psi: \mathbb{P}^1(\mathbb{R}) \rightarrow \mathbb{P}^1(\mathbb{R})$ è una proiettività se e solo se conserva il birapporto. Ma noi ora preme procedere in una direzione diversa, per giungere a definire il piano proiettivo.

Lo studio delle proiezioni centrali ci ha portato ad aggiungere a ogni retta del piano un punto all'infinito, in modo da trasformarle tutte in rette proiettive. Ma è davvero necessario aggiungere un punto diverso per ciascuna retta? Esaminiamo nuovamente la corrispondenza fra il fascio \mathcal{F} e la retta r_1 che ci aveva portato a identificare la struttura proiettiva di \hat{r}_1 . Se prendiamo una retta r'_1 parallela a r_1 e ripetiamo il procedimento, vediamo che la retta di \mathcal{F} associata al punto all'infinito di r'_1 è la stessa che veniva associata al punto all'infinito di r_1 . Viene dunque naturale pensare che il punto all'infinito di r'_1 sia lo stesso di quello di r_1 ; in un certo senso stiamo dicendo che due rette parallele si "intersecano" all'infinito.

Questa idea viene rafforzata ripetendo per il piano la costruzione che ci aveva portato a definire $\mathbb{P}^1(\mathbb{R})$. Identifichiamo il piano \mathbb{R}^2 col piano affine π_0 di \mathbb{R}^3 di equazione cartesiana $z = 1$. A ogni punto di π_0 corrisponde una e una sola retta della stella S di centro l'origine; rimangono fuori soltanto le rette del piano $z = 0$, che per analogia col caso della retta corrispondono ai punti all'infinito del piano π_0 . Ma se andiamo all'infinito lungo due rette parallele r_1 ed r_2 di π_0 , in S finiamo in entrambi i casi sulla stessa retta, l'unica parallela a r_1 ed r_2 . Quindi di nuovo viene naturale assegnare lo stesso punto all'infinito a rette parallele.

Formalizzando questa costruzione otteniamo il piano proiettivo.

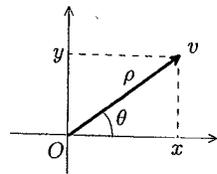


Figura 11.1 Le coordinate polari.

(vedi la Figura 11.1); in particolare, se $x \neq 0$ si ha $\tan \theta = y/x$. La coppia (ρ, θ) è chiamata coordinate polari del vettore (o punto del piano) $v \in \mathbb{R}^2$.

Osservazione 11.1 Il modulo dell'origine è ovviamente 0, ma il suo argomento non è univocamente determinato, per cui le coordinate polari non sono definite nell'origine. Inoltre, se all'argomento di $z = (x, y) \in \mathbb{R}^2$ sommiamo un multiplo intero di 2π , otteniamo un numero θ' che soddisfa ancora le condizioni $x = \rho \cos \theta'$ e $y = \rho \sin \theta'$; con un lieve abuso di linguaggio, chiameremo "argomento di z " anche un tale θ' .

ESEMPIO 11.1 Prendiamo $v = (2, -2)$. Allora v ha modulo $\rho = \sqrt{2^2 + (-2)^2} = 2\sqrt{2}$ e argomento $\theta = 7\pi/4$ (in quanto si deve avere $\tan \theta = -1$ e $\cos \theta > 0$).

Dunque un vettore v_1 di coordinate polari $(\rho_1, \theta_1) \in \mathbb{R}^+ \times [0, 2\pi)$ ha coordinate cartesiane

$$(x_1, y_1) = (\rho_1 \cos \theta_1, \rho_1 \sin \theta_1) = \rho_1(\cos \theta_1, \sin \theta_1).$$

In un certo senso, le coordinate polari suddividono il vettore in una componente scalare, il modulo, e in una componente di rotazione, l'argomento. Questo suggerisce un'idea. Per far agire (per esempio, con un prodotto) il vettore v_1 su un altro vettore v_2 , possiamo ruotare v_2 di un angolo θ_1 e moltiplicarlo per ρ_1 , in modo da tenere presenti entrambe le componenti di v_1 (vedi la Figura 11.2). Se v_2 ha coordinate polari (ρ_2, θ_2) potremmo quindi definire il prodotto fra v_1 e v_2 ponendo

$$\begin{aligned} v_1 \cdot v_2 &= \rho_1 \rho_2 (\cos(\theta_1 + \theta_2), \sin(\theta_1 + \theta_2)) \\ &= \rho_1 \rho_2 (\cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2, \sin \theta_1 \cos \theta_2 + \cos \theta_1 \sin \theta_2). \end{aligned} \quad (11.1)$$

In coordinate cartesiane, (11.1) si scrive

$$(x_1, y_1) \cdot (x_2, y_2) = (x_1 x_2 - y_1 y_2, x_1 y_2 + x_2 y_1). \quad (11.2)$$

Questo è esattamente il prodotto che cercavamo.

Proposizione 11.1 $(\mathbb{R}^2, +, \cdot)$ è un campo.

Dimostrazione. Che \mathbb{R}^2 con la somma usuale sia un gruppo commutativo è ormai più che ben noto. Le proprietà distributive si verificano immediatamente: infatti basta (perché?) notare il fatto, evidente guardando la (11.2), che le applicazioni $v \mapsto v \cdot v_2$ e $v \mapsto v_1 \cdot v$ sono lineari, qualsiasi siano $v_1, v_2 \in \mathbb{R}^2$.

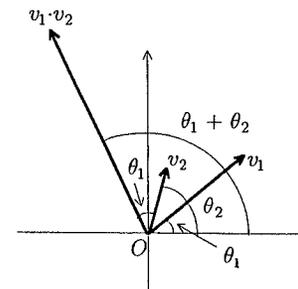


Figura 11.2 Il prodotto in coordinate polari.

Il vettore $(1, 0)$ è chiaramente elemento neutro per il prodotto. L'associatività segue dall'associatività degli ordinari somma e prodotto, grazie alla (11.1). La possiamo anche verificare direttamente con la (11.2) come segue:

$$\begin{aligned} [(x_1, y_1) \cdot (x_2, y_2)] \cdot (x_3, y_3) &= (x_1 x_2 - y_1 y_2, x_1 y_2 + x_2 y_1) \cdot (x_3, y_3) \\ &= ((x_1 x_2 - y_1 y_2)x_3 - (x_1 y_2 + x_2 y_1)y_3, (x_1 x_2 - y_1 y_2)y_3 + x_3(x_1 y_2 + x_2 y_1)) \\ &= (x_1(x_2 x_3 - y_2 y_3) - y_1(x_2 y_3 + x_3 y_2), x_1(x_2 y_3 + x_3 y_2) + y_1(x_2 x_3 - y_2 y_3)) \\ &= (x_1, y_1) \cdot (x_2 x_3 - y_2 y_3, x_2 y_3 + x_3 y_2) \\ &= (x_1, y_1) \cdot [(x_2, y_2) \cdot (x_3, y_3)]. \end{aligned}$$

L'inverso di (x, y) deve avere modulo inverso e argomento opposto, come si vede subito da (11.1); siccome $\cos(-\theta) = \cos \theta$ e $\sin(-\theta) = -\sin \theta$, otteniamo che l'inverso è

$$(x, y)^{-1} = \left(\frac{x}{x^2 + y^2}, \frac{-y}{x^2 + y^2} \right),$$

come si verifica subito anche con (11.2). Infine la commutatività del prodotto è evidente. \square

Definizione 11.2 Il campo \mathbb{C} dei numeri complessi è \mathbb{R}^2 con la somma usuale e questo prodotto.

Da un punto di vista geometrico è naturale identificare \mathbb{R} con l'asse delle ascisse, cioè considerare l'applicazione $T: \mathbb{R} \rightarrow \mathbb{C}$ data da $T(x) = (x, 0)$. Guardando (11.2) risulta evidente che questa T è un *isomorfismo* fra \mathbb{R} e l'asse delle ascisse, nel senso che conserva tutte le operazioni:

$$\forall x, y \in \mathbb{R} \quad T(x + y) = T(x) + T(y) \quad \text{e} \quad T(xy) = T(x) \cdot T(y).$$

Per questo motivo d'ora in poi considereremo \mathbb{R} direttamente come un sottoinsieme di \mathbb{C} , identificandolo con l'asse delle ascisse tramite T . In particolare, scriveremo spesso x al posto di $(x, 0)$.

Definizione 10C.9 Due vettori $v, w \in \mathbb{R}^3 \setminus \{O\}$ si dicono *equivalenti*, e scriveremo $v \sim w$, se esiste $\lambda \neq 0$ tale che $w = \lambda v$. L'insieme quoziente $\mathbb{P}^2(\mathbb{R}) = \mathbb{R}^3 \setminus \{O\} / \sim$ si chiama *piano proiettivo*. Se $\pi: \mathbb{R}^3 \setminus \{O\} \rightarrow \mathbb{P}^2(\mathbb{R})$ è l'applicazione quoziente e $(v_1, v_2, v_3) \in \mathbb{R}^3 \setminus \{O\}$, scriveremo $[v_1: v_2: v_3]$ al posto di $\pi(v_1, v_2, v_3)$; la tripla di numeri (v_1, v_2, v_3) contiene le *coordinate omogenee* del punto $[v_1: v_2: v_3] \in \mathbb{P}^1(\mathbb{R})$. Se $v_3 \neq 0$, i rapporti $(v_1/v_3, v_2/v_3)$ vengono detti *coordinate non omogenee* del punto proprio $[v_1: v_2: v_3]$. I punti con $v_3 = 0$ sono chiamati *punti all'infinito* (o *punti impropri*) del piano proiettivo.

Osservazione 10C.5 Possiamo immergere \mathbb{R}^2 in $\mathbb{P}^2(\mathbb{R})$ tramite l'applicazione iniettiva $f: \mathbb{R}^2 \rightarrow \mathbb{P}^2(\mathbb{R})$ data da $f(x_1, x_2) = [x_1: x_2: 1]$. L'immagine di f è costituita dai punti propri di $\mathbb{P}^2(\mathbb{R})$, che quindi formano una copia di \mathbb{R}^2 . Poi abbiamo un'altra applicazione iniettiva $g: \mathbb{P}^1(\mathbb{R}) \rightarrow \mathbb{P}^2(\mathbb{R})$ data da $g([v_1: v_2]) = [v_1: v_2: 0]$; l'immagine di g sono i punti impropri del piano proiettivo, che quindi compongono una retta proiettiva. Per questo motivo, l'insieme dei punti impropri di $\mathbb{P}^2(\mathbb{R})$ è detto *retta all'infinito*. In particolare, $\mathbb{P}^2(\mathbb{R})$ è l'unione di un piano affine e di una retta proiettiva, o anche (Osservazione 10C.2) di un piano affine, una retta affine e un punto.

Identifichiamo allora il nostro piano \mathbb{R}^2 con l'insieme dei punti propri di $\mathbb{P}^2(\mathbb{R})$, e sia r_1 una retta di equazione cartesiana $ax_1 + bx_2 + c = 0$. Questo vuol dire che il punto proprio $[v_1: v_2: v_3] \in \mathbb{P}^2(\mathbb{R})$ appartiene a r_1 se e solo se $a(v_1/v_3) + b(v_2/v_3) + c = 0$, cioè se e solo se

$$av_1 + bv_2 + cv_3 = 0.$$

Questa equazione è soddisfatta anche da uno (e da uno solo) punto improprio: il punto $[-b: a: 0]$, che è quindi il punto all'infinito di r_1 . Se r'_1 è una retta parallela a r_1 , questa ha equazione cartesiana $ax_1 + bx_2 + c' = 0$ per qualche $c' \in \mathbb{R}$ (Esercizio 10.13), per cui il suo punto all'infinito è ancora $[-b: a: 0]$, confermando l'idea informale che due rette parallele debbano avere lo stesso punto all'infinito. In particolare, *due rette del piano proiettivo si intersecano sempre* (eventualmente in un punto all'infinito).

Questa breve introduzione alla geometria proiettiva termina qui. Ci siamo limitati a scalfire la superficie dell'argomento, sia perché le proprietà geometriche del piano proiettivo sono tante e interessanti, ma soprattutto perché nell'ultimo secolo lo studio degli spazi proiettivi di dimensione qualunque (definiti come nella Definizione 10C.9 sostituendo \mathbb{R}^n — o addirittura uno spazio vettoriale qualsiasi — allo spazio \mathbb{R}^3) ha assunto un'importanza inimmaginabile prima, diventando l'ambiente principe della Geometria Algebrica. Ma tutto ciò decisamente esula da quanto possiamo trattare in questo testo.

I numeri complessi

Come già accennato nella Nota 2 del Capitolo 2 e nella Nota 3 del Capitolo 4, per $n \geq 3$ non è possibile definire un prodotto che, assieme alla somma usuale, renda \mathbb{R}^n un campo. Su \mathbb{R}^2 invece si può fare, e si ottiene il campo \mathbb{C} dei numeri complessi, di cui studieremo le prime proprietà in questo capitolo. Dopo aver definito in maniera geometrica il prodotto e descritto alcune altre operazioni tipiche dei numeri complessi (modulo, parte reale e immaginaria, coniugio), ci concentreremo sulla proprietà algebrica che maggiormente distingue i numeri complessi dai numeri reali: ogni numero complesso non nullo possiede esattamente n radici n -esime distinte. Questo è un caso particolare del celebre Teorema fondamentale dell'algebra, che verrà presentato nel Paragrafo 11.3. Il capitolo si conclude con la versione complessa dell'unico risultato dei Capitoli 3–9 che non vale in un campo qualunque, la Proposizione 5.11. I Complementi a questo capitolo contengono un'introduzione alla teoria algebrica dei polinomi a coefficienti in un campo, teoria che sarà utilizzata nei Complementi al Capitolo 14, e una dimostrazione del Teorema fondamentale dell'algebra.

11.1 Definizione

Come annunciato, il primo obiettivo di questo capitolo è definire un prodotto in \mathbb{R}^2 . Per far ciò dobbiamo esaminare più in dettaglio la struttura dei vettori del piano.

Definizione 11.1 Sia $v = (x, y) \in \mathbb{R}^2$ un vettore non nullo (considerato come un vettore applicato nell'origine). Allora il *modulo* (o *lunghezza*) di v è il numero positivo

$$\rho = |v| = \sqrt{x^2 + y^2}.$$

l'*argomento* di v è invece quell'angolo $\theta \in [0, 2\pi)$, misurato in radianti, tale che

$$x = \rho \cos \theta \quad \text{e} \quad y = \rho \sin \theta,$$

$$\cos \theta = \frac{x}{\sqrt{x^2 + y^2}} \quad \text{e} \quad \sin \theta = \frac{y}{\sqrt{x^2 + y^2}}$$

Il campo \mathbb{C} , essendo \mathbb{R}^2 sotto mentite spoglie, è uno spazio vettoriale di dimensione 2 su \mathbb{R} . La base canonica di \mathbb{C} su \mathbb{R} è data da $\{(1, 0), (0, 1)\}$. Il vettore $(1, 0)$ è l'unità moltiplicativa di \mathbb{C} , e come appena detto verrà indicato semplicemente con 1. Il vettore $(0, 1)$, detto *unità immaginaria*, verrà invece indicato sempre con i . Nota che

$$i^2 = i \cdot i = (0, 1) \cdot (0, 1) = (-1, 0) = -1$$

(infatti se ruotiamo i di $\pi/2$ radianti andiamo a cadere esattamente su -1), cioè i è una soluzione dell'equazione $x^2 + 1 = 0$, che su \mathbb{R} non ne aveva nessuna¹!

Utilizzando le due unità 1 e i possiamo scrivere i numeri complessi in una forma più comoda per i conti. Infatti un generico numero complesso (a, b) può venire ora espresso come

$$(a, b) = (a, 0) + (0, b) = a(1, 0) + b(0, 1) = a \cdot 1 + b \cdot i = a + ib,$$

che è il modo con cui spesso vengono presentati i numeri complessi. Questa notazione è comoda perché per fare calcoli con essa basta utilizzare le solite proprietà (associativa, distributiva, eccetera) e ricordarsi che $i^2 = -1$. Per esempio, ritroviamo le formule per il prodotto e l'inverso come segue:

$$(a + ib)(c + id) = ac + iad + ibc + i^2bd = (ac - bd) + i(ad + bc); \quad (11.3)$$

$$\frac{1}{a + ib} = \frac{a - ib}{(a + ib)(a - ib)} = \frac{a - ib}{a^2 - iab + iab - i^2b^2} = \frac{a - ib}{a^2 + b^2}. \quad (11.4)$$

ESEMPIO 11.2 Più che un esempio è un esercizio. Controlla che le seguenti operazioni siano esatte:

$$\begin{aligned} (3 + 7i)(2 - 2i) &= (6 + 14) + i(-6 + 14) = 20 + 8i; \\ (1 + i)(2 - \pi i)^{-1} &= (1 + i) \frac{2 + \pi i}{4 + \pi^2} = \frac{2 - \pi}{4 + \pi^2} + i \frac{2 + \pi}{4 + \pi^2}; \\ 2i + \frac{1}{1 + i} &= 2i + \frac{1 - i}{2} = \frac{1}{2}(1 + 3i); \\ \left(\frac{-1 + i\sqrt{3}}{2}\right)^2 &= \frac{-1 - i\sqrt{3}}{2}. \end{aligned}$$

Definizione 11.3 Se $z = a + ib \in \mathbb{C}$ è un numero complesso, la *parte reale* $\operatorname{Re} z$ di z è il numero reale a , mentre la *parte immaginaria* $\operatorname{Im} z$ di z è il numero reale b . Se $\operatorname{Re}(z) = 0$ (cioè se $z = ib$), il numero complesso z si dice *immaginario puro*. Infine, il *modulo* di z è il numero reale $|z| = \sqrt{a^2 + b^2}$.

La formula dell'inverso invita a considerare anche un'altra definizione.

¹ Un'altra soluzione è $-i$; ne ripareremo.

Definizione 11.4 Il *complesso coniugato* di un numero complesso $z = a + ib \in \mathbb{C}$ è il numero complesso $\bar{z} = a - ib$. L'applicazione $\bar{\cdot} : \mathbb{C} \rightarrow \mathbb{C}$ che associa a un numero complesso il suo coniugato si chiama *coniugio*.

Le proprietà del coniugio e del modulo sono riassunte nelle prossime due Proposizioni.

Proposizione 11.2 Siano $z_0, z_1, z_2 \in \mathbb{C}$ numeri complessi. Allora:

- (i) $\overline{\bar{z}_0} = z_0$;
- (ii) $z_0 + \bar{z}_0 = 2 \operatorname{Re} z_0$;
- (iii) $z_0 - \bar{z}_0 = 2i \operatorname{Im} z_0$;
- (iv) $\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2$;
- (v) $\overline{z_1 z_2} = \bar{z}_1 \cdot \bar{z}_2$;
- (vi) $\overline{1/z_0} = 1/\bar{z}_0$ se $z_0 \neq 0$;
- (vii) $\bar{z}_0 = z_0$ se e solo se $z_0 \in \mathbb{R}$;
- (viii) $\bar{z}_0 = -z_0$ se e solo se z_0 è immaginario puro;
- (ix) $z_0 \bar{z}_0 = |z_0|^2 \geq 0$, e $z_0 \bar{z}_0 = 0$ se e solo se $z_0 = 0$;
- (x) $z_0^{-1} = \bar{z}_0 / |z_0|^2$.

Dimostrazione. (i), (ii), (iii) e (iv) seguono immediatamente dalla definizione. La (v) si ottiene notando che il coniugio consiste nel cambiare segno alla parte immaginaria (cioè alla y), e guardando quindi la (11.2).

La (vi) è conseguenza della (v), mentre (vii) e (viii) sono di nuovo ovvie. Per la (ix), se $z_0 = a + ib$ si ha

$$z_0 \bar{z}_0 = (a + ib)(a - ib) = a^2 + b^2 + i(-ab + ba) = a^2 + b^2,$$

non ci siamo. Infine la (x) segue dalla (11.4). □

Proposizione 11.3 Siano $z_0, z_1, z_2 \in \mathbb{C}$ numeri complessi. Allora:

- (i) $|z_0| \geq 0$, e $|z_0| = 0$ se e solo se $z_0 = 0$;
- (ii) $|\bar{z}_0| = |z_0|$;
- (iii) $|\operatorname{Re} z_0| \leq |z_0|$, $|\operatorname{Im} z_0| \leq |z_0|$ e $|z_0| \leq |\operatorname{Re} z_0| + |\operatorname{Im} z_0|$;
- (iv) $|z_1 z_2| = |z_1| \cdot |z_2|$;
- (v) $|1/z_0| = 1/|z_0|$ se $z_0 \neq 0$;
- (vi) $|z_1 + z_2| \leq |z_1| + |z_2|$ (*disuguaglianza triangolare*);
- (vii) $|z_1 + z_2| \geq ||z_1| - |z_2||$.

Dimostrazione. La (i) segue dalla Proposizione 11.2.(ix), mentre la (ii) è ovvia. Se $z_0 = a + ib$, la (iii) dice semplicemente che

$$|a| \leq \sqrt{a^2 + b^2}, \quad |b| \leq \sqrt{a^2 + b^2}, \quad \sqrt{a^2 + b^2} \leq |a| + |b|,$$

che si verificano subito elevando al quadrato.

Per la (iv) basta notare che

$$|z_1 z_2|^2 = (z_1 z_2)(\overline{z_1 z_2}) = (z_1 \overline{z_1})(z_2 \overline{z_2}) = |z_1|^2 |z_2|^2.$$

La (v) segue subito da (iv); rimangono la (vi) e la (vii).

Scriviamo $z_j = x_j + iy_j$ per $j = 1, 2$. Elevando al quadrato, vediamo che (vi) e (vii) valgono se e solo se si ha

$$||z_1| - |z_2||^2 \leq |z_1 + z_2|^2 \leq (|z_1| + |z_2|)^2,$$

cioè se e solo se

$$\begin{aligned} (x_1^2 + y_1^2) + (x_2^2 + y_2^2) - 2\sqrt{(x_1^2 + y_1^2)(x_2^2 + y_2^2)} &\leq (x_1 + x_2)^2 + (y_1 + y_2)^2 \\ &\leq (x_1^2 + y_1^2) + (x_2^2 + y_2^2) + 2\sqrt{(x_1^2 + y_1^2)(x_2^2 + y_2^2)}. \end{aligned}$$

Sviluppando i conti vediamo che questo accade se e solo se

$$-\sqrt{(x_1^2 + y_1^2)(x_2^2 + y_2^2)} \leq x_1 x_2 + y_1 y_2 \leq \sqrt{(x_1^2 + y_1^2)(x_2^2 + y_2^2)},$$

che equivale a

$$|x_1 x_2 + y_1 y_2| \leq \sqrt{(x_1^2 + y_1^2)(x_2^2 + y_2^2)}.$$

A sua volta, elevando al quadrato vediamo che questo succede se e solo se

$$x_1^2 x_2^2 + y_1^2 y_2^2 + 2x_1 x_2 y_1 y_2 \leq x_1^2 x_2^2 + x_1^2 y_2^2 + y_1^2 x_2^2 + y_1^2 y_2^2,$$

che equivale a

$$0 \leq (x_1 y_2 - x_2 y_1)^2,$$

che (finalmente!) è vera. E quindi abbiamo dimostrato (vi) e (vii). \square

Osservazione 11.2 Le definizioni sono state scelte in modo che $|w_2 - w_1|$ sia esattamente la distanza euclidea fra i punti w_1 e w_2 del piano (come puoi facilmente verificare). Quindi se nella Proposizione 11.3.(vi) prendiamo $z_1 = w_3 - w_2$ e $z_2 = w_2 - w_1$ otteniamo

$$|w_3 - w_1| \leq |w_3 - w_2| + |w_2 - w_1|,$$

per cui nel triangolo di vertici w_1 , w_2 e w_3 la lunghezza di un lato è minore o uguale alla somma delle lunghezze degli altri due. Per questo motivo la Proposizione 11.3.(vi) si chiama *disuguaglianza triangolare*.

11.2 Potenze e radici

Abbiamo visto che la relazione fra coordinate polari e coordinate cartesiane è espressa dalla formula

$$(x, y) = \rho(\cos \theta, \sin \theta).$$

Analogamente, un numero complesso $z = x + iy$ può venire scritto in forma trigonometrica come

$$z = |z|(\cos \theta + i \sin \theta). \quad (11.5)$$

Il prodotto e il quoziente si scrivono bene in forma trigonometrica: se $z_j \in \mathbb{C}$ ha coordinate polari (ρ_j, θ_j) per $j = 1, 2$, allora

$$\begin{aligned} z_1 z_2 &= \rho_1 \rho_2 (\cos(\theta_1 + \theta_2) + i \sin(\theta_1 + \theta_2)), \\ \frac{z_1}{z_2} &= \frac{\rho_1}{\rho_2} (\cos(\theta_1 - \theta_2) + i \sin(\theta_1 - \theta_2)). \end{aligned} \quad (11.6)$$

Una conseguenza importante della (11.6) è la *formula di De Moivre*:

Lemma 11.4 Sia $z = \rho(\cos \theta + i \sin \theta) \in \mathbb{C}$. Allora

$$\forall n \in \mathbb{N} \quad z^n = \rho^n [\cos(n\theta) + i \sin(n\theta)].$$

Dimostrazione. Procediamo per induzione su n . Se $n = 1$ la formula è ovvia. Sia vera per $n - 1$; allora z^{n-1} ha modulo ρ^{n-1} e argomento $(n - 1)\theta$, per cui la (11.6) implica

$$\begin{aligned} z^n &= z \cdot z^{n-1} = \rho \rho^{n-1} [\cos(\theta + (n - 1)\theta) + i \sin(\theta + (n - 1)\theta)] \\ &= \rho^n [\cos(n\theta) + i \sin(n\theta)]. \end{aligned}$$

\square

In particolare, quindi

$$\forall n \in \mathbb{N} \quad (\cos \theta + i \sin \theta)^n = \cos(n\theta) + i \sin(n\theta). \quad (11.7)$$

Definizione 11.5 Sia $w \in \mathbb{C}$ e $n \geq 1$. Allora una *radice n-esima* di w è un numero complesso $z \in \mathbb{C}$ tale che $z^n = w$.

Noi sappiamo che ogni numero reale x ha due radici n -esime reali se n è pari e x è positivo, una sola se n è dispari (e x è qualunque), e nessuna se n è pari e x è negativo. Sui complessi, la situazione cambia drasticamente. Abbiamo già trovato due radici quadrate di -1 , cioè i e $-i$; quindi in particolare (perché?) sui complessi ogni numero reale diverso da zero ha esattamente due radici quadrate distinte. Ma vale un risultato molto più forte, conseguenza della formula di De Moivre:

Proposizione 11.5 Sia $w \in \mathbb{C}^*$ un numero complesso diverso da zero, e $n \geq 1$. Allora esistono esattamente n radici distinte n -esime z_0, \dots, z_{n-1} di w . Per la precisione, se $w = \rho(\cos \theta + i \sin \theta)$ allora

$$z_k = \rho^{1/n} \left[\cos \left(\frac{\theta + 2k\pi}{n} \right) + i \sin \left(\frac{\theta + 2k\pi}{n} \right) \right] \quad (11.8)$$

per $k = 0, \dots, n-1$, dove $\rho^{1/n}$ indica l'unica radice n -esima reale positiva di ρ .

Dimostrazione. Prima di tutto, grazie alla formula di De Moivre abbiamo

$$(z_k)^n = \rho [\cos(\theta + 2k\pi) + i \sin(\theta + 2k\pi)] = \rho(\cos \theta + i \sin \theta) = w$$

per $k = 0, \dots, n-1$, per cui z_0, \dots, z_{n-1} sono effettivamente radici n -esime di w .

Viceversa, se $z = r(\cos \psi + i \sin \psi)$ è tale che $z^n = w$, si deve avere

$$r^n = \rho, \quad \cos(n\psi) = \cos \theta \quad \text{e} \quad \sin(n\psi) = \sin \theta,$$

cioè $r = \rho^{1/n}$ e $n\psi = \theta + 2k\pi$ per qualche $k \in \mathbb{Z}$, per cui z è della forma (11.8). Siccome (esercizio) $z_{k+\alpha n} = z_k$ per ogni $k, \alpha \in \mathbb{Z}$, rimane soltanto da dimostrare che i numeri z_0, \dots, z_{n-1} sono tutti diversi. Supponiamo che esistano $0 \leq h, k \leq n-1$ tali che $z_h = z_k$. Allora

$$\cos \left(\frac{\theta + 2h\pi}{n} \right) = \cos \left(\frac{\theta + 2k\pi}{n} \right), \quad \sin \left(\frac{\theta + 2h\pi}{n} \right) = \sin \left(\frac{\theta + 2k\pi}{n} \right),$$

e quindi deve esistere $l \in \mathbb{Z}$ tale che

$$\frac{\theta + 2h\pi}{n} = \frac{\theta + 2k\pi}{n} + 2l\pi,$$

cioè $h - k$ dev'essere un multiplo intero di n . Ma h e k sono entrambi compresi fra 0 ed $n-1$; quindi necessariamente $h = k$. \square

Osservazione 11.3 Tutte le radici n -esime di w hanno lo stesso modulo, $|w|^{1/n}$; quindi stanno tutte sulla circonferenza di raggio $|w|^{1/n}$ (vedi la Figura 11.3).

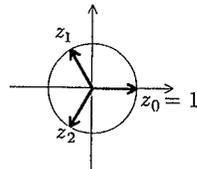


Figura 11.3 Le radici terze dell'unità.

ESEMPIO 11.3 Calcoliamo le radici terze di $1 = 1(\cos 0 + i \sin 0)$. La formula (11.8) ci dà

$$\begin{aligned} z_0 &= \cos 0 + i \sin 0 = 1, \\ z_1 &= \cos \frac{2\pi}{3} + i \sin \frac{2\pi}{3} = \frac{1}{2}(-1 + i\sqrt{3}), \\ z_2 &= \cos \frac{4\pi}{3} + i \sin \frac{4\pi}{3} = -\frac{1}{2}(1 + i\sqrt{3}); \end{aligned}$$

confronta con l'Esercizio 11.2. Nota che

$$z_3 = \cos \frac{6\pi}{3} + i \sin \frac{6\pi}{3} = \cos(2\pi) + i \sin(2\pi) = 1 = z_0,$$

come previsto. Le radici terze dell'unità sono rappresentate nella Figura 11.3.

11.3 Il Teorema fondamentale dell'algebra

Un altro modo di esprimere la Proposizione 11.5 è dire che il polinomio $z^n - c$ ammette sempre esattamente n radici qualunque sia $c \in \mathbb{C}^*$. La proprietà di gran lunga più importante dei numeri complessi è che questo vale per qualsiasi polinomio, e non solo per $z^n - c$: un polinomio di grado n ammette sempre esattamente n radici. Ma per spiegare bene di cosa si tratta, premettiamo un paio di definizioni e lemmi.

Definizione 11.6 Se $p \in \mathbb{C}[z]$ è un polinomio in una variabile a coefficienti complessi, una radice (o zero) di p è un numero complesso $z_0 \in \mathbb{C}$ tale che $p(z_0) = 0$.

Lemma 11.6 Sia $p \in \mathbb{C}[z]$ un polinomio in una variabile a coefficienti complessi di grado n , e $z_0 \in \mathbb{C}$. Allora z_0 è una radice di p se e solo se esiste un polinomio $q \in \mathbb{C}[z]$ di grado $n-1$ tale che

$$p(z) = (z - z_0)q(z).$$

Dimostrazione. Se effettuiamo la divisione² fra $p(z)$ e il polinomio $z - z_0$, troviamo un quoziente $q(z)$ di grado $n-1$ e un resto r di grado zero (cioè una costante) tali che

$$p(z) = (z - z_0)q(z) + r.$$

In particolare, $r = p(z_0)$ e quindi $p(z_0) = 0$ se e solo se $r = 0$. \square

Definizione 11.7 Sia $p \in \mathbb{C}[z]$ un polinomio a coefficienti complessi, e $z_0 \in \mathbb{C}$ una sua radice. Diremo che z_0 è una radice di molteplicità $m \in \mathbb{N}$ se il polinomio $(z - z_0)^m$ divide $p(z)$, mentre $(z - z_0)^{m+1}$ non lo divide.

² La divisione fra polinomi a coefficienti complessi si effettua esattamente come la divisione fra polinomi a coefficienti reali. Per ulteriori informazioni leggi i Complementi a questo capitolo.

ESEMPIO 11.4 Prendiamo il polinomio $p(z) = z^3 + (1-2i)z^2 - (1+2i)z - 1$. Siccome

$$p(z) = (z-i)^2(z+1),$$

vediamo che i è radice di p di molteplicità 2, mentre -1 è radice di p di molteplicità 1.

Definizione 11.8 Diremo che un polinomio $p \in \mathbb{C}[z]$ ha n radici, contate con la relativa molteplicità, se le radici di p sono z_1, \dots, z_k , di molteplicità rispettivamente m_1, \dots, m_k , e si ha $m_1 + \dots + m_k = n$.

Dunque il polinomio $p(z) = z^n - c$ ha n radici di molteplicità 1 se $c \neq 0$, e una sola radice di molteplicità n se $c = 0$; in entrambi i casi ha esattamente n radici, contate con la relativa molteplicità. Bene; il risultato basilare più importante della teoria dei numeri complessi è che questo accade per ogni polinomio di grado n :

Teorema 11.7 Sia $p \in \mathbb{C}[z]$ un polinomio di grado n . Allora p ha esattamente n radici, contate con la relativa molteplicità.

Questo è il *Teorema fondamentale dell'algebra*, che è il motivo per cui spesso i numeri complessi sono più utili dei numeri reali: tutti i polinomi hanno il numero giusto di radici complesse, non una di più, e soprattutto non una di meno. Vedremo più oltre come questo sia utilissimo per esempio nel diagonalizzare matrici.

Esistono molte dimostrazioni di questo Teorema, ma nessuna completamente elementare. Vi sono dimostrazioni puramente algebriche, molto lunghe e complicate: oppure dimostrazioni brevissime, che dipendono però da risultati (anche profondi) di altre teorie. Nel Paragrafo 11C.4 ne descriveremo una che cerca di mediare fra comprensibilità e numero di prerequisiti, richiedendo solo due risultati di Analisi: il fatto che i polinomi sono funzioni continue, e il fatto che una successione limitata di numeri reali ammette sempre una sottosuccessione convergente (Teorema di Weierstrass).

Concludiamo questo Paragrafo con due risultati riguardanti polinomi a coefficienti reali:

Proposizione 11.8 Sia $p \in \mathbb{R}[z]$ un polinomio a coefficienti reali. Allora $z_0 \in \mathbb{C}$ è radice di p se e solo se \bar{z}_0 lo è, e hanno la stessa molteplicità. In particolare, un polinomio a coefficienti reali ha necessariamente un numero pari (che può essere anche zero) di radici complesse non reali.

Dimostrazione. Sia $p(z) = a_n z^n + a_{n-1} z^{n-1} + \dots + a_0$, con $a_0, \dots, a_n \in \mathbb{R}$. Allora

$$p(\bar{z}) = a_n \bar{z}^n + a_{n-1} \bar{z}^{n-1} + \dots + a_0 = \overline{a_n z^n + a_{n-1} z^{n-1} + \dots + a_0} = \overline{p(z)},$$

in quanto $\overline{a_j} = a_j$ per $j = 0, \dots, n$, per cui z_0 è radice se e solo se \bar{z}_0 lo è.

Se z_0 è radice, allora possiamo scrivere

$$p(z) = (z - z_0)(z - \bar{z}_0)q(z);$$

siccome $(z - z_0)(z - \bar{z}_0) = z^2 - 2\operatorname{Re}(z_0)z + |z_0|^2$ è un polinomio a coefficienti reali, anche $q(z)$ lo è; quindi le radici non reali di q sono a coppie complesse coniugate. In particolare, z_0 e \bar{z}_0 devono avere la stessa molteplicità, e ci siamo. \square

Corollario 11.9 Ogni polinomio a coefficienti reali di grado dispari ha almeno una radice reale.

Dimostrazione. Infatti deve avere un numero dispari, contando le molteplicità, di radici complesse (per il Teorema 11.7), ma solo un numero pari (per la Proposizione 11.8) di radici non reali. \square

11.4 Rango della trasposta coniugata

Ci rimane un ultimo affare da sbrigare con i numeri complessi. Come già detto più volte, tutti i risultati su spazi vettoriali e matrici che abbiamo visto nei Capitoli 3-9 valgono qualunque sia il campo di scalari; in particolare, valgono sui complessi. O meglio, tutti tranne uno: la Proposizione 5.11, nella cui dimostrazione abbiamo usato in maniera essenziale il fatto che una somma di quadrati è zero se e solo se tutti i numeri coinvolti sono zero. Questo sui complessi non è vero: basti pensare a $i^2 + 1^2$.

In questo paragrafo vogliamo vedere che aspetto assume la Proposizione 5.11 sul campo dei complessi. Ci servono un paio di definizioni.

Definizione 11.9 Sia $A \in M_{m,n}(\mathbb{C})$ una matrice a coefficienti complessi. La *matrice coniugata* di A è la matrice $\bar{A} \in M_{m,n}(\mathbb{C})$ i cui elementi sono i coniugati degli elementi di A . La *matrice trasposta coniugata* (o *aggiunta*) $A^H \in M_{n,m}(\mathbb{C})$ di A è la trasposta della coniugata di A ; in simboli $A^H = \bar{A}^T = \overline{A^T}$.

Chiaramente,

$$(A^H)^H = A \quad \text{e} \quad (AB)^H = B^H A^H;$$

inoltre, si ha $\operatorname{rg} A = \operatorname{rg} \bar{A}$ (Esercizio 11.13).

Poi, per ogni $a \in \mathbb{C}^m$ definiamo l'applicazione lineare $\psi_a: \mathbb{C}^m \rightarrow \mathbb{C}$ ponendo

$$\forall x \in \mathbb{C}^m \quad \psi_a(x) = \bar{a}_1 x_1 + \dots + \bar{a}_m x_m.$$

Se consideriamo a e x come matrici $m \times 1$ possiamo anche scrivere $\psi_a(x) = a^H \cdot x$. Di conseguenza se $A \in M_{m,n}(\mathbb{C})$, $a \in \mathbb{C}^m$ e $x \in \mathbb{C}^n$ abbiamo

$$\psi_{A^H a}(x) = (A^H a)^H x = a^H (Ax) = \psi_a(Ax). \quad (11.9)$$

Osservazione 11.4 \mathbb{C}^n , come spazio vettoriale su \mathbb{C} , ammette una base canonica formata, proprio come per \mathbb{R}^n , dai vettori $\{e_1, \dots, e_n\}$, dove le componenti di e_j sono tutte zero tranne la j -esima che è uguale a 1.

Allora la versione complessa della Proposizione 5.11 è la seguente:

Proposizione 11.10 Sia $A \in M_{m,n}(\mathbb{C})$. Allora:

- (i) $\mathbb{C}^m = \text{Im } L_A \oplus \text{Ker } L_{A^H}$ e $\mathbb{C}^n = \text{Im } L_{A^H} \oplus \text{Ker } L_A$;
 (ii) $\text{rg } A = \text{rg } A^T = \text{rg } A^H$.

Dimostrazione. Prima di tutto notiamo che (Esercizio 11.13) $\text{rg } A^T = \text{rg } A^H$; quindi ci basta dimostrare che $\text{Im } L_A \cap \text{Ker } L_{A^H} = \{O\}$, e poi procedere esattamente come nella dimostrazione della Proposizione 5.11.

Prendiamo $a \in \text{Im } L_A \cap \text{Ker } L_{A^H}$, cioè tale che $A^H a = O$ ed esista $b \in \mathbb{C}^n$ con $Ab = a$. Allora la (11.9) ci dice che

$$0 = \psi_O(b) = \psi_{A^H a}(b) = \psi_a(Ab) = \psi_a(a) = |a_1|^2 + \dots + |a_m|^2.$$

Dunque abbiamo di nuovo una somma di numeri reali positivi uguale a zero; quindi necessariamente $|a_1| = \dots = |a_m| = 0$, per cui $a = O$, e abbiamo finito. \square

Osservazione 11.5 Per matrici A a coefficienti complessi, *non* è detto che l'intersezione fra l'immagine di L_A e il nucleo della trasposta sia il solo vettore nullo. Per esempio, se

$$A = \begin{vmatrix} i & 1 \\ 1 & -i \end{vmatrix} = A^T,$$

allora

$$\text{Im } L_A = \text{Ker } L_{A^T} = \mathbb{C} \begin{vmatrix} i \\ 1 \end{vmatrix}.$$

Esercizi

11.1 Calcola le coordinate polari dei seguenti punti: $(-3, 0)$, $(0, 2\pi)$, $(2\sqrt{3}, 2)$, $(6, -6)$ e $(-2, 2\sqrt{3})$.

11.2 Esprimi nella forma $a + bi$ i seguenti numeri complessi (ovvero, fai i conti):

$$\left(\frac{1+i\sqrt{3}}{2}\right)^3; \quad (-i)^4; \quad (2+2i)(-2+2i); \quad \frac{1+3i}{1-3i}.$$

11.3 Dimostra la formula del "modulo al quadrato del binomio":

$$|z+w|^2 = |z|^2 + 2\text{Re}(z\bar{w}) + |w|^2.$$

11.4 Utilizza (11.7) per ritrovare le formule trigonometriche che esprimono seno e coseno del doppio e del triplo di un angolo.

11.5 Trova le quattro radici quarte di 1, e le due radici quadrate di i .

11.6 Trova le radici quadrate di $1 - 4i\sqrt{3}$.

11.7 Determina modulo e argomento delle radici cubiche di $(1+i)(1-i)^{-1}$.

11.8 Determina i numeri complessi $z \in \mathbb{C}$ tali che $z^4 = \bar{z}$.

11.9 Determina, al variare di $k \in \mathbb{R}$, le soluzioni dell'equazione $ikz^3 = |z|^3$.

11.10 Trova le radici complesse del polinomio $p_1(z) = z^3 + i \in \mathbb{C}[z]$. Trova, se esiste, un polinomio $p_2 \in \mathbb{C}[z]$ tale che $p_1 p_2$ abbia tutti i coefficienti reali.

11.11 Determina per quali $\lambda \in \mathbb{C}$ l'equazione

$$(\lambda^8 - 1)z^3 + (\lambda^4 - 1)z + \lambda^2 - i = 0$$

non ammette soluzioni.

11.12 Trova un polinomio a coefficienti reali di quinto grado con solo una radice reale, e un polinomio a coefficienti reali di quarto grado senza nessuna radice reale.

11.13 Sia $A \in M_{m,n}(\mathbb{C})$. Dimostra che $\text{rg } \bar{A} = \text{rg } A$.

11.14 Sia $A \in M_{n,n}(\mathbb{C})$. Dimostra che $\det A^H = \det \bar{A} = \overline{\det A}$. Deduci che se $A^H = A$ allora $\det A$ è un numero reale.

11.15 Dimostra che l'insieme delle matrici quadrate a coefficienti complessi $M_{n,n}(\mathbb{C})$ ha dimensione n^2 come spazio vettoriale su \mathbb{C} , ma dimensione $2n^2$ se considerato come spazio vettoriale su \mathbb{R} . Dimostra che l'insieme $\{A \in M_{n,n}(\mathbb{C}) \mid A^H = A\}$ è uno spazio vettoriale su \mathbb{R} di dimensione n^2 .

11.16 Se $z = a + ib \in \mathbb{C}$, sia $m_z: \mathbb{C} \rightarrow \mathbb{C}$ l'applicazione lineare data da $m_z(w) = wz$. Trova la matrice che rappresenta m_z rispetto alla base $\{1, i\}$ di \mathbb{C} considerato come spazio vettoriale su \mathbb{R} .

11.17 Dimostra che il coniugio è lineare se consideriamo \mathbb{C} come spazio vettoriale su \mathbb{R} , mentre non lo è se consideriamo \mathbb{C} come spazio vettoriale su se stesso.

11.18 Trova rango, nucleo e immagine delle seguenti matrici a coefficienti complessi:

$$\begin{vmatrix} 1 & i & -1 \\ 1 & -i & -1 \\ 1 & 1 & 1 \end{vmatrix}, \quad \begin{vmatrix} 1 & i & 0 \\ i & -1 & 1 \\ 2 & 2i & 0 \end{vmatrix}, \quad \begin{vmatrix} 1 & i & 1 \\ i & -1 & i \\ 1 & i & 1 \end{vmatrix}.$$

11.19 Studia il sistema lineare (in cui le incognite sono numeri complessi)

$$\begin{cases} x + 2y + kz = 0, \\ ky - z = 1, \\ kx + z = a, \end{cases}$$

al variare di $a, k \in \mathbb{C}$.

11.20 Se $z = x + iy \in \mathbb{C}$, si definisce l'*esponenziale complesso* $e^z \in \mathbb{C}$ con la formula

$$e^z = e^x(\cos y + i \sin y).$$

Dimostra che $e^z \cdot e^w = e^{z+w}$ per ogni $z, w \in \mathbb{C}$, e che $e^{i\pi} + 1 = 0$.

11.21 Sia $p(t) = a_n t^n + \dots + a_0 \in \mathbb{Z}[t]$ un polinomio a coefficienti interi, e sia $r = a/b \in \mathbb{Q}$ una radice razionale di p , dove $a, b \in \mathbb{Z}$ sono relativamente primi. Dimostra che a divide a_0 , e che b divide a_n .

11.22 Trova le radici dei polinomi a coefficienti interi $p_1(t) = t^4 - 2t^3 - 5t^2 + 4t + 6$ e $p_2(t) = t^5 - 5t^3 - 2t^2 + 6t + 4$. (Suggerimento: usa l'Esercizio precedente.)

COMPLEMENTI

11C.1 Divisibilità e radici

Questi Complementi sono dedicati ai polinomi a coefficienti in un campo qualunque; partendo dalle definizioni arriveremo a dimostrare risultati piuttosto profondi e utili.

Fissiamo un campo \mathbb{K} ; nella maggior parte delle applicazioni \mathbb{K} sarà il campo reale \mathbb{R} o il campo complesso \mathbb{C} , ma non ci costa nessuna fatica in più mantenere questa generalità. Ricordiamo la definizione già vista nell'Esempio 1.9 e nell'Osservazione 1.5.

Definizione 11C.1 L'insieme $\mathbb{K}[t]$ dei polinomi in una variabile a coefficienti in \mathbb{K} è l'insieme delle funzioni $p: \mathbb{K} \rightarrow \mathbb{K}$ tali che esistano un numero naturale $n \geq 0$ e coefficienti $a_0, \dots, a_n \in \mathbb{K}$ con $a_n \neq 0$ in modo che

$$\forall t \in \mathbb{K} \quad p(t) = a_n t^n + \dots + a_0. \quad (11C.1)$$

Il numero n è detto *grado* del polinomio p , e si indica con $\deg p$ (dove "deg" è l'abbreviazione di *degree*, la parola inglese per "grado"); i numeri a_0, \dots, a_n sono i *coefficienti* di p , il numero a_0 è il *termine noto* e $a_n \neq 0$ il *coefficiente direttivo*. Per convenzione, si dice che il polinomio nullo $p \equiv 0$ ha grado $-\infty$.

Di solito, quando si scrive un polinomio come in (11C.1) si suppone che a_n sia diverso da zero, in modo da vedere bene il grado del polinomio. Ma ovviamente nulla ci impedisce di scrivere anche

$$p(t) = 0t^m + \dots + 0t^{n+1} + a_n t^n + \dots + a_0 \quad (11C.2)$$

per ogni $m > n$ (è chiaramente la stessa funzione); con questo trucco possiamo facilmente definire la somma di due polinomi.

Siano $p, q \in \mathbb{K}[t]$ due polinomi, e scriviamo

$$\begin{aligned} p(t) &= a_n t^n + \dots + a_0, \\ q(t) &= b_n t^n + \dots + b_0, \end{aligned}$$

dove n è il massimo fra il grado di p e il grado di q , e qualcuno dei coefficienti può essere zero, come in (11C.2). Allora il polinomio *somma* $p + q \in \mathbb{K}[t]$ è definito

semplicemente come

$$(p + q)(t) = (a_n + b_n)t^n + \dots + (a_0 + b_0);$$

abbiamo effettuato la somma delle funzioni p e q .

Se $\lambda \in \mathbb{K}$ è uno scalare, possiamo definire il polinomio λp semplicemente ponendo

$$(\lambda p)(t) = (\lambda a_n)t^n + \dots + (\lambda a_0).$$

È facile verificare (vedi l'Esercizio 1.19) che con queste operazioni $\mathbb{K}[t]$ diventa uno spazio vettoriale su \mathbb{K} . Possiamo però anche definire il *prodotto* $pq \in \mathbb{K}[t]$ di due polinomi p e q semplicemente moltiplicando le funzioni: se $p(t) = a_n t^n + \dots + a_0$ e $q(t) = b_m t^m + \dots + b_0$, allora

$$(pq)(t) = (a_n t^n + \dots + a_0)(b_m t^m + \dots + b_0) = c_{m+n} t^{m+n} + \dots + c_0,$$

dove

$$c_j = a_0 b_j + a_1 b_{j-1} + \dots + a_{j-1} b_1 + a_j b_0 = \sum_{h+k=j} a_h b_k$$

per $j = 0, \dots, n + m$, e si intende che $a_h = 0 = b_k$ se $h > n$ o $k > m$.

Con questo prodotto e la somma sopra definita, è facile vedere (Esercizio 1.19) che $\mathbb{K}[t]$ è un anello ma non un campo: se $\deg p > 0$, non c'è modo di moltiplicare p per un altro polinomio in modo da ottenere una costante. Infatti, dalla definizione è evidente che

$$\deg(pq) = \deg p + \deg q, \quad (11C.3)$$

per cui $\deg(pq) \geq \deg p$ se q non è il polinomio nullo, e quindi pq non può essere una costante (in quanto le costanti sono esattamente i polinomi di grado zero).

Anche gli interi \mathbb{Z} non sono un campo; e quindi alle elementari ci hanno insegnato a fare le divisioni col resto³. Con i polinomi si può fare qualcosa di analogo, come probabilmente saprai già:

Teorema 11C.1 Siano $p_0, p_1 \in \mathbb{K}[t]$ due polinomi sul campo \mathbb{K} , con p_1 diverso dal polinomio nullo. Allora esistono e sono unici due polinomi $q, r \in \mathbb{K}[t]$ con $\deg r < \deg p_1$ tali che

$$p_0 = qp_1 + r. \quad (11C.4)$$

Dimostrazione. Sia $n = \deg p_0$, $m = \deg p_1 \geq 0$ e scriviamo

$$p_0(t) = a_n t^n + \dots + a_0, \quad p_1(t) = b_m t^m + \dots + b_0,$$

con $a_n, b_m \neq 0$. Procediamo per induzione su n . Se $n = 0$, cioè $p_0 = a_0$, vi sono due casi: se $m > 0$ basta prendere $q = 0$ e $r = a_0$; se invece $m = 0$, cioè $p_1 = b_0 \neq 0$, basta prendere $q = a_0/b_0$ ed $r = 0$.

³ In un campo, due numeri non nulli sono sempre divisibili fra loro, senza resto.

Sia vero allora per tutti i polinomi di grado minore di n , e prendiamo p_0 di grado n . Se $n < m$, basta prendere $q = 0$ ed $r = p_0$. Se invece $n \geq m$, poniamo

$$p_2(t) = p_0(t) - a_n b_m^{-1} t^{n-m} p_1(t).$$

Chiaramente $\deg p_2 < \deg p_0 = n$; quindi per ipotesi induttiva esistono due polinomi q_1 ed r con $\deg r < \deg p_1$ tali che $p_2 = q_1 p_1 + r$. Allora

$$p_0(t) = p_2(t) + a_n b_m^{-1} t^{n-m} p_1(t) = [q_1(t) + a_n b_m^{-1} t^{n-m}] p_1(t) + r(t),$$

per cui $q(t) = q_1(t) + a_n b_m^{-1} t^{n-m}$ ed $r(t)$ sono come volevamo.

Per l'unicità, supponiamo che $q_1, q_2, r_1, r_2 \in \mathbb{K}[t]$ siano tali che

$$q_1 p_1 + r_1 = q_2 p_1 + r_2$$

con $\deg r_1, \deg r_2 < \deg p_1$. Allora

$$(q_1 - q_2) p_1 = r_2 - r_1;$$

ma il primo membro, se non è zero, ha necessariamente grado maggiore o uguale al grado di p_1 , mentre il secondo membro ha grado strettamente minore del grado di p_1 . Dunque l'uguaglianza è possibile se e solo se $q_2 = q_1$ e $r_2 = r_1$, come volevamo. \square

Definizione 11C.2 I polinomi q ed r in (11C.4) si dicono rispettivamente *quoziente* e *resto* della divisione di p_0 per p_1 . Se $r = 0$ diremo che p_1 divide p_0 .

In un campo c'è un elemento molto particolare, lo zero. È quindi naturale esaminare che relazione c'è fra un polinomio e gli elementi che vengono mandati in zero.

Definizione 11C.3 Sia $p \in \mathbb{K}[t]$ un polinomio non identicamente nullo. Allora una *radice* (o *zero*) di p è un elemento $\alpha \in \mathbb{K}$ tale che $p(\alpha) = 0$.

Corollario 11C.2 Sia $p \in \mathbb{K}[t]$ un polinomio non nullo. Allora $\alpha \in \mathbb{K}$ è una radice di p se e solo se $t - \alpha$ divide p , cioè se e solo se esiste un polinomio $q \in \mathbb{K}[t]$ tale che

$$p(t) = (t - \alpha)q(t).$$

Più in generale, $\alpha_1, \dots, \alpha_h \in \mathbb{K}$ sono radici distinte di p se e solo se $(t - \alpha_1) \cdots (t - \alpha_h)$ divide p .

Dimostrazione. Per induzione su h . Cominciamo con $h = 1$. Se $t - \alpha_1$ divide p si ha subito $p(\alpha_1) = (\alpha_1 - \alpha_1)q(\alpha_1) = 0$. Viceversa, supponiamo che α_1 sia una radice di p , e scriviamo

$$p(t) = (t - \alpha_1)q(t) + r(t),$$

dove r è un polinomio di grado minore di 1, cioè una costante r_0 . Ma allora

$$0 = p(\alpha_1) = (\alpha_1 - \alpha_1)q(\alpha_1) + r(\alpha_1) = r_0,$$

e quindi $t - \alpha_1$ divide p .

Sia allora vero per $h - 1$ radici, per cui possiamo scrivere

$$p(t) = (t - \alpha_1) \cdots (t - \alpha_{h-1})q(t).$$

Siccome

$$p(\alpha_h) = (\alpha_h - \alpha_1) \cdots (\alpha_h - \alpha_{h-1})q(\alpha_h),$$

il numero α_h è una radice (distinta dalle precedenti) di p se e solo se è radice di q , cioè se e solo se $q(t) = (t - \alpha_h)q_1(t)$, e ci siamo. \square

Corollario 11C.3 Sia $p \in \mathbb{K}[t]$ un polinomio non nullo di grado n . Allora p ha al più n radici distinte.

Dimostrazione. Infatti se p avesse $n + 1$ radici distinte $\alpha_1, \dots, \alpha_{n+1} \in \mathbb{K}$, il Corollario 11C.2 ci darebbe un polinomio non nullo $q \in \mathbb{K}[t]$ tale che

$$p(t) = (t - \alpha_1) \cdots (t - \alpha_{n+1})q(t),$$

per cui il grado di p sarebbe almeno $n + 1$, contraddizione. \square

Corollario 11C.4 Siano $p, q \in \mathbb{K}[t]$ due polinomi tali che $p(t) = q(t)$ per ogni $t \in \mathbb{K}$, e supponiamo che \mathbb{K} contenga infiniti elementi⁴. Allora p e q hanno lo stesso grado e gli stessi coefficienti — cioè sono lo stesso polinomio.

Dimostrazione. Infatti il polinomio $p - q$ ha necessariamente infinite radici, e quindi per il Corollario 11C.3 è il polinomio nullo. \square

Dunque due polinomi (intesi come funzioni) coincidono se e solo se si scrivono nello stesso modo; abbiamo utilizzato questo fatto nell'Esempio 4.13.

11C.2 Ideali e massimo comun divisore

Nello studio dei numeri interi un altro concetto molto importante è quello di massimo comun divisore. Esiste anche il massimo comun divisore di polinomi; per definirlo seguiremo una strada indiretta ma che fornirà altri risultati importanti.

Cominciamo introducendo un concetto che, in un certo senso, è parente di quello di sottospazio vettoriale.

Definizione 11C.4 Un sottoinsieme $\mathcal{J} \subseteq \mathbb{K}[t]$ si dice *ideale* di $\mathbb{K}[t]$ se è un sottospazio vettoriale di $\mathbb{K}[t]$ tale che $qp \in \mathcal{J}$ non appena $p \in \mathcal{J}$, qualunque sia il polinomio $q \in \mathbb{K}[t]$. In altri termini, \mathcal{J} è chiuso non solo rispetto al prodotto per scalari ma rispetto al prodotto per polinomi qualunque.

⁴ Esistono dei campi con solo un numero finito di elementi, che però a noi non interessano.

ESEMPIO 11C.1 Chiaramente sia $\mathbb{K}[t]$ che l'insieme costituito dal solo polinomio nullo sono ideali, detti *ideali banali*. Un ideale non banale viene detto *ideale proprio*.

ESEMPIO 11C.2 Se $p \in \mathbb{K}[t]$ è un polinomio non identicamente nullo, indicheremo con $\langle p \rangle$ l'insieme dei multipli di p . In simboli,

$$\langle p \rangle = \{qp \in \mathbb{K}[t] \mid q \in \mathbb{K}[t]\}.$$

È facile vedere (esercizio) che è un ideale di $\mathbb{K}[t]$, l'ideale generato da p . Se $\deg p > 0$, si tratta chiaramente di un ideale proprio, in quanto contiene $p = 1p$ e non contiene le costanti. Più in generale, se $p_1, \dots, p_k \in \mathbb{K}[t]$ sono dei polinomi, allora l'ideale generato da p_1, \dots, p_k è l'insieme

$$\langle p_1, \dots, p_k \rangle = \{q_1 p_1 + \dots + q_k p_k \mid q_1, \dots, q_k \in \mathbb{K}[t]\}.$$

Di nuovo, è facile verificare (esercizio) che $\langle p_1, \dots, p_k \rangle$ è sempre un ideale; ma contrariamente al caso di un solo generatore, talvolta potrebbe essere banale. Per esempio, se $p_1 = t - 1$ e $p_2 = t + 2$, allora $1 = p_1 + p_2 \in \langle p_1, p_2 \rangle$, e quindi $q = q1 \in \langle p_1, p_2 \rangle$ per ogni $q \in \mathbb{K}[t]$, cioè $\langle p_1, p_2 \rangle = \mathbb{K}[t]$. Vedremo che questo accade se e solo se p_1 e p_2 sono relativamente primi.

Definizione 11C.5 Sia \mathcal{J} un ideale di $\mathbb{K}[t]$. Diremo che p_1, \dots, p_k generano \mathcal{J} (e che \mathcal{J} è generato da p_1, \dots, p_k) se $\mathcal{J} = \langle p_1, \dots, p_k \rangle$.

Dimostriamo fra poco che ogni ideale di $\mathbb{K}[t]$ è generato da un singolo polinomio, oltretutto di tipo particolare:

Definizione 11C.6 Un polinomio $p \in \mathbb{K}[t]$ è detto *monico* se ha coefficiente direttivo uguale a 1.

Ci serve un lemma preliminare.

Lemma 11C.5 Siano $p, \tilde{p} \in \mathbb{K}[t]$ polinomi monici. Supponiamo che p divida \tilde{p} e inoltre che $\deg p = \deg \tilde{p}$, oppure che \tilde{p} divida p . Allora $\tilde{p} = p$.

Dimostrazione. Siccome se p divide \tilde{p} e \tilde{p} divide p deve per forza essere $\deg p = \deg \tilde{p}$, possiamo in entrambi i casi supporre che $\deg p = \deg \tilde{p} = n$, e scrivere

$$p(t) = t^n + a_{n-1}t^{n-1} + \dots + a_0, \quad \tilde{p}(t) = t^n + \tilde{a}_{n-1}t^{n-1} + \dots + \tilde{a}_0.$$

Se p divide \tilde{p} , deve esistere un polinomio q tale che $\tilde{p} = qp$. Essendo $\deg p = \deg \tilde{p}$, la (11C.3) implica che $\deg q = 0$, cioè che q è una costante c_0 . In particolare,

$$t^n + \tilde{a}_{n-1}t^{n-1} + \dots + \tilde{a}_0 = \tilde{p} = c_0 p = c_0 t^n + c_0 a_{n-1} t^{n-1} + \dots + c_0 a_0;$$

ma allora $c_0 = 1$, e dunque $\tilde{p} = p$. \square

Teorema 11C.6 Sia $\mathcal{J} \neq \{0\}$ un ideale di $\mathbb{K}[t]$. Allora esiste un unico polinomio monico p che genera \mathcal{J} , e p è il polinomio monico di grado minimo contenuto in \mathcal{J} .

Dimostrazione. Sia $p \in \mathcal{J}$ un polinomio non nullo di grado minimo fra quelli contenuti in \mathcal{J} ; a meno di moltiplicarlo per una costante possiamo supporre che p sia monico. Vogliamo dimostrare che $\mathcal{J} = \langle p \rangle$. Infatti, sia $p_1 \in \mathcal{J}$; per il Teorema 11C.1, possiamo scrivere

$$p_1 = qp + r,$$

con $\deg r < \deg p$. Ma, essendo \mathcal{J} un ideale, abbiamo $qp \in \mathcal{J}$ e quindi $r = p_1 - qp \in \mathcal{J}$. Poiché $\deg r < \deg p$ e p era un polinomio di grado minimo fra quelli non nulli in \mathcal{J} , ne segue che $r = 0$. Dunque $p_1 = qp$ per cui, essendo $p_1 \in \mathcal{J}$ generico, deve aversi $\mathcal{J} \subseteq \langle p \rangle$. Siccome l'altra inclusione è ovvia (perché?) abbiamo dimostrato che p è un generatore di \mathcal{J} .

Rimane da far vedere che è l'unico generatore monico. Sia \tilde{p} un altro generatore monico. Allora dev'essere $p = q\tilde{p}$ per un qualche polinomio q ; in particolare, \tilde{p} divide p . Ma per lo stesso motivo p divide \tilde{p} ; quindi il Lemma 11C.5 ci dà $p = \tilde{p}$. \square

A questo punto possiamo definire di massimo comun divisore fra polinomi:

Definizione 11C.7 Siano $p_1, \dots, p_k \in \mathbb{K}[t]$ polinomi non nulli. Il *massimo comun divisore* di p_1, \dots, p_k è un polinomio monico $q \in \mathbb{K}[t]$ che divide p_1, \dots, p_k , e tale che se $\tilde{p} \in \mathbb{K}[t]$ è un altro polinomio che divide p_1, \dots, p_k allora \tilde{p} divide anche q . Indicheremo con $(p_1, \dots, p_k) \in \mathbb{K}[t]$ il massimo comun divisore di p_1, \dots, p_k .

Ovviamente si pone il problema dell'esistenza del massimo comun divisore. La risposta è contenuta nella

Proposizione 11C.7 Siano $p_1, \dots, p_k \in \mathbb{K}[t]$ polinomi non nulli. Allora il massimo comun divisore di p_1, \dots, p_k è il generatore monico dell'ideale $\langle p_1, \dots, p_k \rangle$.

Dimostrazione. Sia q il generatore monico di $\mathcal{J} = \langle p_1, \dots, p_k \rangle$. Siccome $p_i \in \mathcal{J}$ per $i = 1, \dots, k$, abbiamo subito che q divide p_1, \dots, p_k . Sia ora \tilde{p} un polinomio che divide tutti i p_i , e scriviamo $p_i = \tilde{q}_i \tilde{p}$ per $i = 1, \dots, k$. Siccome $q \in \mathcal{J}$, esistono polinomi $q_i \in \mathbb{K}[t]$ tali che $q = q_1 p_1 + \dots + q_k p_k$; quindi

$$q = (q_1 \tilde{q}_1 + \dots + q_k \tilde{q}_k) \tilde{p},$$

e \tilde{p} divide anche q .

Abbiamo quindi dimostrato che q è un massimo comun divisore di p_1, \dots, p_k ; supponiamo che ce ne sia un altro, \tilde{q} . Allora q deve dividere \tilde{q} , in quanto \tilde{q} divide p_1, \dots, p_k . Analogamente \tilde{q} deve dividere q , per cui il Lemma 11C.5 ci dà $\tilde{q} = q$. \square

Osservazione 11C.1 Per definizione, il massimo comun divisore (p_1, p_2) di due polinomi p_1, p_2 appartiene all'ideale $\langle p_1, p_2 \rangle$; quindi esistono due polinomi $q_1, q_2 \in \mathbb{K}[t]$ tali che $q_1 p_1 + q_2 p_2 = (p_1, p_2)$. In particolare, se $(p_1, p_2) = 1$ troviamo $q_1, q_2 \in \mathbb{K}[t]$ tali che

$$q_1 p_1 + q_2 p_2 = 1.$$

Definizione 11C.8 I polinomi $p_1, \dots, p_k \in \mathbb{K}[t]$ sono detti *relativamente primi* se $(p_1, \dots, p_k) = 1$ o, equivalentemente, se $\langle p_1, \dots, p_k \rangle = \mathbb{K}[t]$.

Vi è un procedimento semplice per trovare il massimo comun divisore di due polinomi, detto *algoritmo di Euclide*. Prendiamo due polinomi non nulli $p_1, p_2 \in \mathbb{K}[t]$; a meno dell'ordine, possiamo supporre che $\deg p_2 \leq \deg p_1$. Per il Teorema 11C.1, esistono due polinomi q_1 e $p_3 \in \mathbb{K}[t]$ con $\deg p_3 < \deg p_2$ tali che

$$p_1 = q_1 p_2 + p_3.$$

Applicando il Teorema 11C.1 a p_2 e p_3 troviamo q_2 e p_4 con $\deg p_4 < \deg p_3$ tali che

$$p_2 = q_2 p_3 + p_4.$$

Procedendo in questo modo prima o poi arriveremo a ottenere resto zero, in quanto a ogni passo il grado del resto diminuisce di almeno 1. Dunque abbiamo costruito una successione di polinomi p_1, \dots, p_s con

$$p_{i-1} = q_{i-1} p_i + p_{i+1}$$

per $i = 2, \dots, s-1$ e

$$p_{s-1} = q_{s-1} p_s;$$

l'idea è che il massimo comun divisore di p_1 e p_2 è (a meno del coefficiente direttivo, che potrebbe essere diverso da 1) proprio p_s . Infatti, per definizione p_s divide p_{s-1} ; quindi da

$$p_{s-2} = q_{s-2} p_{s-1} + p_s = (q_{s-2} q_{s-1} + 1) p_s$$

deduciamo che divide anche p_{s-2} . Risalendo la catena, vediamo che p_s divide sia p_2 che p_1 . Viceversa, se \bar{p} divide sia p_1 che p_2 , necessariamente divide anche p_3 ; dunque divide anche p_4 e quindi, discendendo la catena, divide anche p_s , come richiesto.

ESEMPIO 11C.3 Consideriamo i polinomi $p_1, p_2 \in \mathbb{R}[t]$ dati da

$$p_1(t) = 2t^5 + 2t^4 + 2t + 2 \quad \text{e} \quad p_2(t) = t^3 - t^2 - t + 1.$$

Allora si ha

$$\begin{aligned} 2t^5 + 2t^4 + 2t + 2 &= (2t^2 + 4t + 6)(t^3 - t^2 - t + 1) + 8t^2 + 4t - 4, \\ t^3 - t^2 - t + 1 &= \left(\frac{1}{8}t - \frac{3}{16}\right)(8t^2 + 4t - 4) + \frac{1}{4}t + \frac{1}{4}, \\ 8t^2 + 4t - 4 &= (32t - 16)\left(\frac{1}{4}t + \frac{1}{4}\right). \end{aligned}$$

Quindi il massimo comun divisore monico di p_1 e p_2 è

$$(p_1, p_2) = t + 1.$$

L'Esercizio 11C.2 spiega come trovare il massimo comun divisore di più polinomi.

11C.3 Fattorizzazione di polinomi

Un altro concetto fondamentale per lo studio dei numeri interi è quello di numero primo. L'analogo per i polinomi è il concetto di polinomio irriducibile:

Definizione 11C.9 Un polinomio $p \in \mathbb{K}[t]$ è detto *irriducibile* (su \mathbb{K}) se ogni polinomio non costante che lo divide è della forma cp per qualche $c \in \mathbb{K}$. In altre parole, il polinomio p è irriducibile se (a meno di fattori costanti non nulli) gli unici divisori di p sono 1 e lui stesso.

ESEMPIO 11C.4 Se p è irriducibile, anche cp lo è, per ogni costante non nulla c (perché?); quindi, ci limiteremo spesso a considerare polinomi irriducibili monici.

ESEMPIO 11C.5 Un polinomio lineare (cioè di primo grado) è sempre irriducibile. Infatti se $p(t) = t - \alpha$ e $q \in \mathbb{K}[t]$ divide p , allora q è costante oppure il quoziente lo è: In quest'ultimo caso $p = cq$ con $c \neq 0$, e quindi $q = c^{-1}p$.

ESEMPIO 11C.6 Se un polinomio $p \in \mathbb{K}[t]$ di grado almeno 2 ammette una radice in \mathbb{K} , allora non è irriducibile. Infatti, se $\alpha \in \mathbb{K}$ è una radice di p abbiamo visto che $t - \alpha$ divide p . In particolare, quindi, il Teorema 11.7 ci dice che gli unici polinomi irriducibili in $\mathbb{C}[t]$ sono i polinomi di primo grado.

ESEMPIO 11C.7 Un polinomio $p \in \mathbb{K}[t]$ di grado esattamente 2 è irriducibile su \mathbb{K} se e solo se non ammette radici in \mathbb{K} . Infatti, abbiamo appena visto che se ammette una radice non è irriducibile. Viceversa, se non è irriducibile si deve scrivere come prodotto di due fattori di primo grado. Ma un polinomio di primo grado ha sempre una radice in \mathbb{K} , e quindi p stesso ammette radici.

ESEMPIO 11C.8 Gli unici polinomi irriducibili in $\mathbb{R}[t]$ sono quelli lineari e quelli quadratici senza radici reali. Per dimostrarlo, grazie agli Esempi precedenti basta far vedere che ogni polinomio $p \in \mathbb{R}[t]$ di grado almeno 3 non è irriducibile. Prendiamo una radice $z_0 \in \mathbb{C}$ di p , che esiste per il Teorema 11.7 (che verrà dimostrato nel prossimo paragrafo). Se $z_0 \in \mathbb{R}$, allora $t - z_0$ divide p , e p non è irriducibile. Se $z_0 \notin \mathbb{R}$, allora anche \bar{z}_0 è una radice di p , e quindi il polinomio $(t - z_0)(t - \bar{z}_0) \in \mathbb{R}[t]$ divide p , e di nuovo p non è irriducibile.

Osservazione 11C.2 Un polinomio p può essere irriducibile su \mathbb{R} e non esserlo su \mathbb{C} , oppure essere irriducibile su \mathbb{Q} ma non esserlo su \mathbb{R} . Per esempio, consideriamo un polinomio monico p di secondo grado. Grazie all'Esempio 11C.7, p è irriducibile su \mathbb{K} se e solo se ha delle radici in \mathbb{K} . Quindi $p(t) = t^2 - 2$ è irriducibile su \mathbb{Q} , mentre su \mathbb{R} si spezza come $p(t) = (t - \sqrt{2})(t + \sqrt{2})$. Analogamente, $p(t) = t^2 + 4$ è irriducibile su \mathbb{R} , mentre su \mathbb{C} si ha $p(t) = (t - 2i)(t + 2i)$.

Il nostro principale obiettivo in questo paragrafo è dimostrare che ogni polinomio si può scrivere in maniera essenzialmente unica come prodotto di polinomi irriducibili. Per questo ci servirà il

Lemma 11C.8 Sia $p \in \mathbb{K}[t]$ irriducibile, e $q_1, q_2 \in \mathbb{K}[t]$ polinomi non nulli. Supponiamo che p divida il prodotto $q_1 q_2$; allora p divide necessariamente anche uno dei due fattori q_1 e q_2 .

Dimostrazione. Supponiamo che p non divida q_1 . Allora il massimo comun divisore di p e q_1 è 1, per cui (Osservazione 11C.1) esistono polinomi $\tilde{p}_1, \tilde{p}_2 \in \mathbb{K}[t]$ tali che

$$1 = \tilde{p}_1 q_1 + \tilde{p}_2 p.$$

Moltiplicando per q_2 otteniamo

$$q_2 = \tilde{p}_1 (q_1 q_2) + (\tilde{p}_2 q_2) p;$$

dunque siccome p divide $q_1 q_2$ vediamo subito che divide anche q_2 . \square

Teorema 11C.9 Ogni polinomio non costante $p \in \mathbb{K}[t]$ si può scrivere come prodotto

$$p = c p_1 \cdots p_r, \quad (11C.5)$$

dove c è il coefficiente direttivo di p e p_1, \dots, p_r sono polinomi monici irriducibili, univocamente determinati a meno dell'ordine.

Dimostrazione. A meno di dividere p per il suo coefficiente direttivo possiamo supporre p monico. Cominciamo a dimostrare per induzione sul grado l'esistenza della fattorizzazione. Se $\deg p = 1$ abbiamo già visto che p è irriducibile, per cui abbiamo finito. Sia allora $\deg p > 1$. Se p è irriducibile è fatta; altrimenti, possiamo scrivere $p = c q_1 q_2$, con sia q_1 che q_2 monici e di grado minore di p . In particolare, c è il coefficiente direttivo di p , per cui $c = 1$. Per l'ipotesi induttiva, sia q_1 che q_2 si possono scrivere come prodotto di polinomi monici irriducibili, e quindi anche per p è fatta.

Vediamo ora l'unicità. Supponiamo di avere

$$p_1 \cdots p_r = q_1 \cdots q_s,$$

dove i p_i e i q_j sono tutti polinomi monici irriducibili; vogliamo dimostrare che $r = s$ e che i q_j sono uguali ai p_i , a meno dell'ordine. Procediamo per induzione su s . Se $s = 1$, essendo q_1 irriducibile e monico, abbiamo per forza $r = 1$ e quindi $p_1 = q_1$. Sia allora vero per $s - 1$. Applicando il Lemma 11C.8 a p_1 e al prodotto $q_1 (q_2 \cdots q_s)$, vediamo che o p_1 divide q_1 oppure p_1 divide $q_2 \cdots q_s$. Nel primo caso, abbiamo necessariamente $p_1 = q_1$, per cui $p_2 \cdots p_r = q_2 \cdots q_s$ e l'ipotesi induttiva ci fornisce la tesi. Se invece p_1 divide $q_2 \cdots q_s$, applicando di nuovo il Lemma 11C.8 troviamo che o p_1 divide q_2 oppure divide il prodotto $q_3 \cdots q_s$. Procedendo in questo modo scopriamo che esiste necessariamente un j_0 tale che $p_1 = q_{j_0}$; possiamo allora applicare l'ipotesi induttiva come nel caso $j_0 = 1$, ed è fatta. \square

Definizione 11C.10 I polinomi p_1, \dots, p_r in (11C.5) sono detti *fattori irriducibili* del polinomio p .

Corollario 11C.10 Siano $p_1, p_2, q \in \mathbb{K}[t]$ tali che $(p_1, p_2) = 1$ e p_1 divide $p_2 q$. Allora p_1 divide q .

Dimostrazione. Basta applicare il Lemma 11C.8 ai fattori irriducibili di p_1 . \square

Ovviamente, i polinomi monici irriducibili p_1, \dots, p_r potrebbero non essere tutti distinti. Ma in ogni caso possiamo scrivere

$$p = c p_1^{m_1} \cdots p_s^{m_s}, \quad (11C.6)$$

dove c è il coefficiente direttivo di p , i p_j sono polinomi monici irriducibili *distinti* univocamente determinati a meno dell'ordine, e gli $m_j \geq 1$ sono numeri interi univocamente determinati, detti *molteplicità* di p_j come divisore di p .

Un caso molto interessante è quello dei polinomi irriducibili della forma $t - \alpha$.

Definizione 11C.11 Sia $\alpha \in \mathbb{K}$ una radice del polinomio non costante $p \in \mathbb{K}[t]$. Diremo *molteplicità* di α come radice di p la molteplicità di $t - \alpha$ come divisore di p .

Per definizione, possiamo scrivere

$$p(t) = (t - \alpha)^m q(t),$$

dove m è la molteplicità di α come radice di p , e $q(\alpha) \neq 0$ (in quanto $q(t)$ è il prodotto degli altri fattori irriducibili della decomposizione di p , che sono relativamente primi con $t - \alpha$). Dunque la molteplicità di una radice è anche il più grande intero m tale che $(t - \alpha)^m$ divide p (perché?).

Utilizzeremo abbondantemente la molteplicità delle radici nel Capitolo 14; ma adesso dobbiamo parlare del minimo comune multiplo.

Definizione 11C.12 Siano $p_1, p_2 \in \mathbb{K}[t]$ due polinomi non nulli. Il *minimo comune multiplo* di p_1 e p_2 è un polinomio monico $[p_1, p_2] \in \mathbb{K}[t]$ tale che p_1 e p_2 dividono $[p_1, p_2]$, e tale che se $\tilde{q} \in \mathbb{K}[t]$ è un polinomio diviso sia da p_1 che da p_2 , allora anche $[p_1, p_2]$ divide \tilde{q} . Più in generale, se $p_1, \dots, p_k \in \mathbb{K}[t]$ sono polinomi non nulli, il loro *minimo comune multiplo* $[p_1, \dots, p_k] \in \mathbb{K}[t]$ è definito per induzione da

$$[p_1, \dots, p_k] = [[p_1, \dots, p_{k-1}], p_k].$$

Proposizione 11C.11 Siano $p_1, \dots, p_k \in \mathbb{K}[t]$ polinomi non nulli. Allora:

(i) si ha

$$[p_1, p_2] = \frac{1}{c} \frac{p_1 p_2}{(p_1, p_2)},$$

dove c è il coefficiente direttivo di $p_1 p_2$;

(ii) il minimo comune multiplo dei polinomi p_1, \dots, p_k è il generatore monico dell'ideale $\langle p_1 \rangle \cap \cdots \cap \langle p_k \rangle$.

Dimostrazione. (i) Prima di tutto è facile verificare (Esercizio 11C.3) che il minimo comune multiplo se esiste è unico. Poi siccome (p_1, p_2) divide sia p_1 che p_2 è chiaro

che p_1 e p_2 dividono $p_1 p_2 / (p_1, p_2)$. Supponiamo ora che sia p_1 che p_2 dividano il polinomio \tilde{q} . Quindi possiamo scrivere

$$\tilde{q} = p_1 q_1 = \frac{p_1}{(p_1, p_2)} q_1(p_1, p_2).$$

Siccome p_2 divide \tilde{q} , per il Corollario 11C.10 il polinomio p_2 deve dividere $q_1(p_1, p_2)$, per cui otteniamo $q_1(p_1, p_2) = p_2 \tilde{q}_1$ e quindi

$$\tilde{q} = \frac{p_1 p_2}{(p_1, p_2)} \tilde{q}_1,$$

cioè $p_1 p_2 / (p_1, p_2)$ divide \tilde{q} . Dunque $c^{-1} p_1 p_2 / (p_1, p_2)$ è un polinomio monico che soddisfa la definizione di minimo comune multiplo di p_1 e p_2 , e ci siamo.

(ii) Procediamo per induzione su k . Sia $k = 2$, e $q \in \mathbb{K}[t]$ il generatore monico dell'ideale (vedi l'Esercizio 11C.1) $\langle p_1 \rangle \cap \langle p_2 \rangle$. Prima di tutto, dalla definizione è evidente che $[p_1, p_2] \in \langle p_1 \rangle \cap \langle p_2 \rangle$, per cui q divide $[p_1, p_2]$. Viceversa, p_1 e p_2 dividono q , per cui $[p_1, p_2]$ divide q . Trattandosi di polinomi monici ne segue che $q = [p_1, p_2]$.

Sia ora vero per $k - 1$. Allora

$$\langle p_1 \rangle \cap \dots \cap \langle p_k \rangle = \langle [p_1, \dots, p_{k-1}] \rangle \cap \langle p_k \rangle = \langle [[p_1, \dots, p_{k-1}], p_k] \rangle = \langle [p_1, \dots, p_k] \rangle,$$

ed è fatta. \square

11C.4 Dimostrazione del Teorema fondamentale dell'algebra

In questo paragrafo descriveremo la dimostrazione del Teorema fondamentale dell'algebra promessa nel Paragrafo 11.3. Ci serve un risultato preliminare di Analisi.

Lemma 11C.12 Fissato $r \geq 0$ sia $D_r = \{z \in \mathbb{C} \mid |z| \leq r\}$ il disco di raggio r , e sia $f: D_r \rightarrow \mathbb{R}$ una funzione continua definita su D_r a valori reali. Allora esiste un $z_0 \in D_r$ in cui la funzione f assume il suo valore minimo.

Dimostrazione. Se l'insieme $f(D_r) \subseteq \mathbb{R}$ non è limitato inferiormente, scegliamo una successione di punti $z_n = x_n + iy_n \in D_r$ tali che $f(z_n) < -n$. Se invece $f(D_r)$ è limitato inferiormente, sia $a = \inf f(D_r)$ e scegliamo una successione $z_n = x_n + iy_n \in D_r$ tali che $a \leq f(z_n) < a + 1/n$. In entrambi i casi le successioni $\{x_n\}$ e $\{y_n\}$ sono contenute nell'intervallo $[-r, r]$, grazie alla Proposizione 11.3.(iii); quindi il Teorema di Weierstrass (Teorema 1.6) ci permette di estrarre due sottosuccessioni convergenti, $x_{n_k} \rightarrow x_0$ e $y_{n_k} \rightarrow y_0$. Dunque $z_{n_k} = x_{n_k} + iy_{n_k}$ tende a $z_0 = x_0 + iy_0$; siccome $|z_{n_k}| \leq r$ per ogni n_k , per definizione di limite si ha $|z_0| \leq r$, cioè $z_0 \in D_r$. Ma f è una funzione continua; quindi $f(z_0) = \lim f(z_{n_k})$. Questo vuol dire che $\inf f(D_r)$ non può essere uguale a $-\infty$, per cui $f(z_0) = \lim f(z_{n_k})$ è necessariamente uguale ad a , che è dunque il valore minimo assunto da f su D_r . \square

Teorema 11C.13 Sia $p \in \mathbb{C}[z]$ un polinomio di grado n . Allora p ha esattamente n radici, contate con la relativa molteplicità.

Dimostrazione. Procediamo per induzione sul grado di p . Se p ha grado 1, cioè si ha $p(z) = az + b$ con $a \neq 0$, allora ha un'unica radice di molteplicità 1, e ci siamo.

Supponiamo ora che p abbia grado $n > 1$; vogliamo prima di tutto dimostrare che p ha almeno una radice. Supponiamo per assurdo che p non abbia radici; a meno di moltiplicare p per una costante non nulla (che non cambia la situazione) possiamo assumere che p sia della forma

$$p(z) = z^n + a_{n-1} z^{n-1} + \dots + a_0.$$

Sia $C > 1$ una costante più grande del modulo di tutti i coefficienti di p , cioè

$$C > \max\{1, |a_{n-1}|, \dots, |a_0|\},$$

e prendiamo $z \in \mathbb{C}$ tale che $|z| \geq 2nC$. In particolare, $|z| > 1$, e quindi $|z| \leq |z|^k$ per ogni $1 \leq k \leq n$. Dunque

$$\left| \frac{a_{n-1}}{z} + \dots + \frac{a_0}{z^n} \right| \leq \frac{|a_{n-1}|}{|z|} + \dots + \frac{|a_0|}{|z|^n} \leq \frac{nC}{|z|} \leq \frac{1}{2},$$

per cui

$$\begin{aligned} |p(z)| &= \left| z^n \left(1 + \frac{a_{n-1}}{z} + \dots + \frac{a_0}{z^n} \right) \right| \geq |z|^n \left| 1 - \left| \frac{a_{n-1}}{z} + \dots + \frac{a_0}{z^n} \right| \right| \\ &\geq \frac{1}{2} |z|^n \geq 2^{n-1} (nC)^n > |a_0| \end{aligned}$$

(in questi conti abbiamo ampiamente utilizzato la Proposizione 11.3). In altri termini, abbiamo fatto vedere che $|p(z)| > |a_0| = |p(0)|$ non appena $|z| \geq 2nC$.

Ora, per il Lemma 11C.12 la funzione $|p(z)|$ (che è una funzione continua, in quanto modulo di un polinomio) assume un valore minimo nel disco $|z| \leq 2nC$. Il punto z_0 in cui $|p(z)|$ assume il minimo è necessariamente dentro al disco (non può essere sulla circonferenza), in quanto $|p(z_0)|$ dev'essere minore o uguale di $|p(0)|$, mentre per ipotesi $|p(z)| > |p(0)| = |a_0|$ sulla circonferenza.

A meno di sostituire a p il polinomio $p_1(z) = p(z + z_0)$, che ovviamente ha esattamente tante radici quante ne ha p , possiamo supporre che il punto di minimo sia $z_0 = 0$. Siccome abbiamo supposto per assurdo che p_1 non abbia radici, si ha $p_1(0) = b_0 \neq 0$, per cui possiamo scrivere

$$p_1(z) = b_0 + b_1 z^k + z^{k+1} q(z)$$

per qualche $k \geq 1$ e $b_1 \neq 0$, dove $q(z) \in \mathbb{C}[z]$ è un polinomio opportuno.

Sia $z_1 \neq 0$ tale che $z_1^k = -b_0/b_1$ (un tale z_1 esiste per la Proposizione 11.5), e prendiamo $t \in [0, 1]$ reale. Allora

$$p_1(tz_1) = b_0 - t^k b_0 + t^{k+1} z_1^{k+1} q(tz_1) = b_0 [1 - t^k - t^{k+1} z_1 q(tz_1)/b_1].$$

Ora, la funzione $t \mapsto |z_1 q(tz_1)/b_1|$ è continua, per cui (Teorema 1.7) ammette massimo nell'intervallo chiuso e limitato $[0, 1]$. Questo vuol dire che esiste $C_1 > 1$ tale che

$$\forall t \in [0, 1] \quad |z_1 q(tz_1)/b_1| \leq C_1,$$

e quindi

$$\forall t \in [0, 1] \quad |p_1(tz_1)| \leq |b_0|(1 - t^k + C_1 t^{k+1}).$$

Prendiamo $0 < t_0 < 1/C_1$; si ha (esercizio)

$$0 < 1 - t_0^k + C_1 t_0^{k+1} < 1,$$

e quindi $|p_1(t_0 z_1)| < |b_0| = |p_1(0)|$, contraddicendo così l'ipotesi che 0 fosse il punto di minimo del modulo di p_1 .

La contraddizione deriva dall'aver supposto $b_0 \neq 0$, cioè dall'aver supposto che il nostro polinomio p non abbia radici. Questo si è rivelato impossibile, e quindi p deve avere almeno una radice z_0 . Ma allora possiamo scrivere $p(z) = (z - z_0)q(z)$, dove q è un polinomio di grado $n - 1$. Per ipotesi induttiva, q ha esattamente $n - 1$ radici, contate con la relativa molteplicità. Siccome ognuna di queste è anche radice di p , abbiamo fatto vedere che p ha esattamente n radici, contate con la relativa molteplicità, come richiesto. \square

Esercizi

11C.1 Siano $\mathcal{I}, \mathcal{J} \subseteq \mathbb{K}[t]$ due ideali. Dimostra che $\mathcal{I} \cap \mathcal{J}$ è ancora un ideale.

11C.2 Siano $p_1, \dots, p_k \in \mathbb{K}[t]$ polinomi non nulli. Dimostra che

$$(p_1, \dots, p_k) = ((p_1, \dots, p_{k-1}), p_k).$$

11C.3 Siano $p_1, p_2 \in \mathbb{K}[t]$ due polinomi non nulli, e $\tilde{p}_1, \tilde{p}_2 \in \mathbb{K}[t]$ due polinomi monici che soddisfano la definizione di minimo comune multiplo di p_1 e p_2 . Dimostra che $\tilde{p}_1 = \tilde{p}_2$. (Suggerimento: usa il Lemma 11C.5).

Prodotti scalari ed hermitiani

È finalmente giunto il momento di occuparci di distanze e angoli, ma senza dimenticarci quanto visto finora: vogliamo trattare questioni metriche con tecniche lineari, in modo da poterle inserire facilmente nel contesto degli spazi vettoriali. Questo ci porterà a introdurre il concetto di prodotto scalare (e il suo equivalente complesso, il prodotto hermitiano), che è l'argomento principale di questo capitolo. Dopo una lunga serie di definizioni ed esempi, ci concentreremo sui prodotti scalari (o hermitiani) definiti positivi, dimostrando la disuguaglianza di Cauchy-Schwarz, introducendo il concetto di base ortonormale e descrivendo il procedimento di ortogonalizzazione di Gram-Schmidt. Dopo aver definito l'ortogonale di un sottospazio e le proiezioni ortogonali, mostreremo come rappresentare un prodotto scalare rispetto a una base tramite una matrice. Infine definiremo l'aggiunta di un'applicazione lineare fra spazi vettoriali dotati di un prodotto scalare, ed esamineremo in particolare le isometrie lineari e gli endomorfismi autoaggiunti.

12.1 Definizioni ed esempi, caso reale

Cominciamo a vedere come si può procedere in \mathcal{A}^2 . In questo caso, dovresti già avere un'idea di cos'è la distanza fra due punti, e l'angolo fra due direzioni; il nostro obiettivo è trovare un modo per calcolarli. Prima di tutto, fissiamo un sistema di riferimento affine $RA(O, \vec{i}, \vec{j})$, in modo da poter lavorare in \mathbb{R}^2 ; inoltre supponiamo che i vettori \vec{i} e \vec{j} siano di lunghezza 1 e formino un angolo di $\pi/2$ radianti¹.

Prendiamo un punto P del piano, di coordinate $v_1 = (x_1, y_1) \in \mathbb{R}^2$. Il Teorema di Pitagora (Teorema 1.2) ci dice subito che la distanza $d(O, P)$ di P dall'origine O , cioè la lunghezza del vettore \vec{OP} , è data da

$$d(O, P) = \sqrt{x_1^2 + y_1^2}. \quad (12.1)$$

¹ In altri termini, abbiamo fissato un sistema di riferimento ortonormale, o cartesiano. Ne ripareremo nel prossimo capitolo.

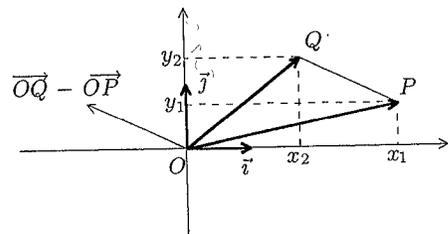


Figura 12.1 La distanza fra due punti.

Spesso scriveremo $\|\overrightarrow{OP}\|$ o anche $\|v_1\|$ per indicare la distanza $d(O, P)$. Il numero $\|v_1\|$ viene detto *norma* del vettore v_1 ; la doppia sbarra è stata scelta per ricordare che si tratta di una generalizzazione del modulo di un numero reale (o di un numero complesso).

Il Teorema di Pitagora ci dice anche (vedi la Figura 12.1) come calcolare la distanza $d(P, Q)$ fra P e un altro punto Q del piano, di coordinate $v_2 = (x_2, y_2) \in \mathbb{R}^2$:

$$d(P, Q) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} = \|v_2 - v_1\| = \|\overrightarrow{OQ} - \overrightarrow{OP}\|. \quad (12.2)$$

Osservazione 12.1 Attenzione a un particolare apparentemente secondario ma importante: le formule (12.1) e (12.2) ci forniscono l'usuale distanza euclidea del piano *soltanto perché i vettori di base del sistema di riferimento affine scelto sono di lunghezza unitaria e perpendicolari fra loro*. Se non fosse stato così, potevamo ancora identificare \mathcal{A}^2 con \mathbb{R}^2 come sempre, e introdurre una "distanza" fra punti con la formula (12.2); ma questa distanza non sarebbe stata quella euclidea. Per esempio, come indicato nell'Esercizio 12.1, l'insieme dei punti di distanza uno dall'origine invece di una circonferenza poteva essere un'ellisse.

Torniamo al nostro problema. La formula (12.2) non sembra (per ora) dare indicazioni su come esprimere le distanze tramite oggetti lineari; vediamo se con gli angoli ci andrà meglio.

Osservazione 12.2 Per noi, l'angolo fra due direzioni (o rette, o segmenti, o vettori) sarà sempre *convesso*, e mai concavo, per cui varierà fra 0 e π (inclusi). In particolare, siccome il coseno nell'intervallo $[0, \pi]$ è una funzione iniettiva, per trovare l'angolo fra due vettori basterà calcolarne il coseno.

Consideriamo due vettori in \mathbb{R}^2 , di coordinate $v_1 = (x_1, y_1)$ e $v_2 = (x_2, y_2)$; vogliamo l'angolo θ che determinano. Per quanto detto sopra, sarà sufficiente trovare $\cos \theta$ (vedi la Figura 12.2).

Calcolare $\cos \theta$ direttamente non è immediato; la trigonometria invece ci dice come

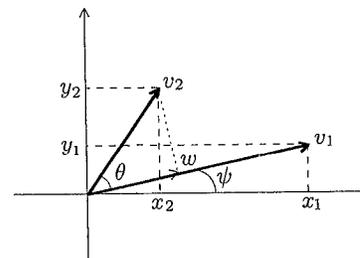


Figura 12.2 L'angolo fra due vettori.

trovare l'angolo ψ fra v_1 e l'asse x , e l'angolo $\psi + \theta$ fra v_2 e l'asse x . Infatti si ha

$$\begin{cases} \cos \psi = \frac{x_1}{\|v_1\|}, \\ \sin \psi = \frac{y_1}{\|v_1\|}, \end{cases} \quad (12.3)$$

e

$$\begin{cases} \cos(\psi + \theta) = \frac{x_2}{\|v_2\|}, \\ \sin(\psi + \theta) = \frac{y_2}{\|v_2\|}. \end{cases} \quad (12.4)$$

Le formule d'addizione per le funzioni trigonometriche trasformano (12.4) in

$$\begin{cases} \cos \psi \cos \theta - \sin \psi \sin \theta = \frac{x_2}{\|v_2\|}, \\ \sin \psi \cos \theta + \cos \psi \sin \theta = \frac{y_2}{\|v_2\|}, \end{cases}$$

che è un sistema lineare quadrato nelle incognite $\cos \theta$ e $\sin \theta$. La matrice dei coefficienti di questo sistema ha determinante

$$\det \begin{vmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{vmatrix} = \cos^2 \psi + \sin^2 \psi = 1,$$

per cui il sistema ha un'unica soluzione, che possiamo calcolare per esempio col Teorema di Cramer (Corollario 9.11):

$$\cos \theta = \det \begin{vmatrix} x_2/\|v_2\| & -\sin \psi \\ y_2/\|v_2\| & \cos \psi \end{vmatrix} = \det \begin{vmatrix} x_2/\|v_2\| & -y_1/\|v_1\| \\ y_2/\|v_2\| & x_1/\|v_1\| \end{vmatrix} = \frac{x_1 x_2 + y_1 y_2}{\|v_1\| \|v_2\|}.$$

Viene quindi naturale associare ai vettori v_1 e v_2 il numero

$$\langle v_1, v_2 \rangle = x_1 x_2 + y_1 y_2, \quad (12.5)$$

per poter poi scrivere

$$\cos \theta = \frac{\langle v_1, v_2 \rangle}{\|v_1\| \|v_2\|}$$

Ora, $\langle \cdot, \cdot \rangle$ è lineare in entrambe le variabili, esattamente come lo era il determinante². Inoltre, anche la norma può venire definita tramite questo nuovo oggetto: infatti

$$\|v\| = \sqrt{\langle v, v \rangle}$$

per qualunque vettore $v \in \mathbb{R}^2$, come avrai cura di verificare. Dunque sembra proprio che abbiamo trovato l'oggetto lineare che stavamo cercando.

Osservazione 12.3 Ci sono altri concetti di geometria metrica del piano che si possono esprimere facilmente usando $\langle \cdot, \cdot \rangle$. Per esempio, due vettori v_1 e v_2 sono ortogonali se e solo se formano un angolo θ di $\pi/2$ radianti, e quindi se e solo se $\cos \theta = 0$, ovvero se e solo se

$$\langle v_1, v_2 \rangle = 0.$$

Analogamente, sono linearmente dipendenti se e solo se formano un angolo θ di 0 o π radianti, e quindi se e solo se $\cos \theta = \pm 1$, ovvero se e solo se

$$\langle v_1, v_2 \rangle = \pm \|v_1\| \|v_2\|.$$

Un altro concetto che possiamo esprimere tramite $\langle \cdot, \cdot \rangle$ è la *proiezione ortogonale* di v_2 su v_1 . Si tratta del vettore w ottenuto tracciando da v_2 la perpendicolare a v_1 (vedi la Figura 12.2), per cui è un multiplo di $v_1/\|v_1\|$, e ha lunghezza $\|v_2\| \cos \theta$. Dunque

$$w = (\|v_2\| \cos \theta) \frac{v_1}{\|v_1\|} = \frac{\langle v_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1.$$

A questo punto potremmo procedere a considerare \mathcal{A}^3 (o meglio \mathbb{R}^3), facendo considerazioni analoghe a quelle già viste. Ma tanto vale fare il passo tutto in una volta e introdurre la seguente definizione:

Definizione 12.1 Il *prodotto scalare canonico* $\langle \cdot, \cdot \rangle: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ su \mathbb{R}^n è la funzione data da

$$\langle v, w \rangle = v_1 w_1 + \dots + v_n w_n = w^T \cdot v \quad (12.6)$$

per ogni $v = (v_1, \dots, v_n)$, $w = (w_1, \dots, w_n) \in \mathbb{R}^n$, dove il prodotto all'ultimo membro è il solito prodotto righe per colonne. La *norma* associata al prodotto scalare canonico è la funzione $\|\cdot\|: \mathbb{R}^n \rightarrow \mathbb{R}^+$ data da

$$\forall v \in \mathbb{R}^n \quad \|v\| = \sqrt{\langle v, v \rangle} = (v_1^2 + \dots + v_n^2)^{1/2}. \quad (12.7)$$

² Ma attenzione: scambiando v_1 e v_2 in $\langle v_1, v_2 \rangle$ otteniamo lo stesso numero, mentre scambiando v_1 e v_2 il determinante $\det(v_1, v_2)$ cambia di segno. La funzione $\langle \cdot, \cdot \rangle$ e il determinante sono oggetti ben diversi, accomunati solo dal fatto di essere entrambi lineari in ciascun argomento.

Nota che la norma è ben definita in quanto

$$\langle v, v \rangle = v_1^2 + \dots + v_n^2 \geq 0$$

per tutti i vettori $v \in \mathbb{R}^n$.

Chiaramente il prodotto scalare canonico generalizza (12.5) a dimensione qualunque, e gode di tutta una serie di proprietà importanti di verifica elementare (e quindi lasciata per esercizio):

(a₁) per ogni v_1, v_2 e $w \in \mathbb{R}^n$ si ha

$$\langle v_1 + v_2, w \rangle = \langle v_1, w \rangle + \langle v_2, w \rangle$$

(additività rispetto alla prima variabile);

(a₂) per ogni $v, w \in \mathbb{R}^n$ e $\lambda \in \mathbb{R}$ si ha

$$\langle \lambda v, w \rangle = \lambda \langle v, w \rangle$$

(omogeneità rispetto alla prima variabile);

(b₁) per ogni v, w_1 e $w_2 \in \mathbb{R}^n$ si ha

$$\langle v, w_1 + w_2 \rangle = \langle v, w_1 \rangle + \langle v, w_2 \rangle$$

(additività rispetto alla seconda variabile);

(b₂) per ogni $v, w \in \mathbb{R}^n$ e $\lambda \in \mathbb{R}$ si ha

$$\langle v, \lambda w \rangle = \lambda \langle v, w \rangle$$

(omogeneità rispetto alla seconda variabile).

Le proprietà (a₁)–(a₂) si riassumono dicendo che il prodotto scalare canonico è *lineare rispetto alla prima variabile*; le (b₁)–(b₂) dicendo che è *lineare rispetto alla seconda variabile*; tutte e quattro dicendo che è *bilineare*.

(c) per ogni $v, w \in \mathbb{R}^n$ si ha $\langle v, w \rangle = \langle w, v \rangle$ (*simmetria*);

(d) si ha $\langle v, w \rangle = 0$ per ogni $w \in V$ se e solo se $v = O$ (il prodotto scalare canonico è *non degenerato*);

(e) per tutti i $v \in \mathbb{R}^n$ con $v \neq O$ si ha $\langle v, v \rangle > 0$ (il prodotto scalare canonico è *definito positivo*).

Usando solo queste proprietà è possibile dimostrare che la distanza fra due punti definita da $\|v - w\|$ e l'angolo fra due vettori definito da $\cos \theta = \langle v, w \rangle / (\|v\| \|w\|)$ godono di tutte le proprietà usuali³, per cui è possibile ricostruire la geometria euclidea

³ Nota che bisogna anche dimostrare che $\langle v, w \rangle / (\|v\| \|w\|)$ è sempre in modulo minore di 1, in modo da poter essere il coseno di qualcosa. La differenza rispetto al caso di \mathcal{A}^2 è che mentre prima *sapevamo a priori* cos'era l'angolo fra due vettori e dovevamo solo calcolarlo, stavolta vogliamo *definire* l'angolo fra due vettori partendo dal prodotto scalare canonico.

di \mathbb{R}^n usando il prodotto scalare canonico. Ma, come oramai potrai immaginare, le parole chiave qui sono "usando solo queste proprietà". In altri termini, non appena abbiamo un'applicazione $\langle \cdot, \cdot \rangle$ che soddisfa le proprietà (a₁)–(e) possiamo ricostruire la geometria euclidea. Ma allora, perché limitarci al solo prodotto scalare canonico? E, soprattutto, perché limitarci al solo \mathbb{R}^n e non invece farci prendere dall'ebbrezza di calcolare la distanza fra due polinomi, l'angolo fra due applicazioni lineari, la proiezione ortogonale di una matrice? Presentiamo quindi subito una serie di definizioni.

Definizione 12.2 Sia V uno spazio vettoriale su \mathbb{R} . Una *forma bilineare* su V è un'applicazione $g: V \times V \rightarrow \mathbb{R}$ tale che

$$g(v_1 + v_2, w) = g(v_1, w) + g(v_2, w), \quad g(v, w_1 + w_2) = g(v, w_1) + g(v, w_2), \\ g(\lambda v, w) = \lambda g(v, w) = g(v, \lambda w),$$

per tutti i $v, v_1, v_2, w, w_1, w_2 \in V$ e $\lambda \in \mathbb{R}$.

Definizione 12.3 Sia V uno spazio vettoriale su \mathbb{R} . Un *prodotto scalare* su V è una forma bilineare g *simmetrica*, cioè tale che

$$\forall v, w \in V \quad g(v, w) = g(w, v).$$

In questo caso scriveremo sempre $\langle v, w \rangle$ invece di $g(v, w)$.

Osservazione 12.4 La notazione $\langle v, w \rangle$ non è l'unica in commercio per i prodotti scalari; si trovano anche $\langle v|w \rangle$, (v, w) , $v \cdot w$, $v \times w$ e altre. Noi useremo esclusivamente $\langle v, w \rangle$, tranne nel Capitolo 16 dove, dovendo considerare contemporaneamente due prodotti scalari, sarà comodo usare anche la notazione (v, w) .

Definizione 12.4 Sia $\langle \cdot, \cdot \rangle$ un prodotto scalare su uno spazio vettoriale V . Il *nucleo* di $\langle \cdot, \cdot \rangle$ è l'insieme⁴

$$V^\perp = \{v \in V \mid \langle v, w \rangle = 0 \quad \forall w \in V\}.$$

Il nucleo di un prodotto scalare è un sottospazio vettoriale: infatti se $v_1, v_2, v \in V^\perp$ e $\lambda \in \mathbb{R}$ si ha

$$\langle v_1 + v_2, w \rangle = \langle v_1, w \rangle + \langle v_2, w \rangle = 0, \quad \langle \lambda v, w \rangle = \lambda \langle v, w \rangle = 0,$$

per ogni $w \in V$, per cui $v_1 + v_2, \lambda v \in V^\perp$.

Definizione 12.5 Sia V uno spazio vettoriale su \mathbb{R} . Un prodotto scalare $\langle \cdot, \cdot \rangle$ su V si dice *non degenerare* se $V^\perp = \{O\}$; *degenerare* altrimenti.

⁴ La scelta del simbolo V^\perp per indicare il nucleo di un prodotto scalare sarà giustificata nella Definizione 12.20.

Nota che $\langle O, O \rangle = 0$ per qualunque prodotto scalare (esercizio). Questo ci porta all'ultima serie di definizioni. \spadesuit

Definizione 12.6 Sia $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$ un prodotto scalare su uno spazio vettoriale V su \mathbb{R} . Questo prodotto scalare è *definito positivo* se $\langle v, v \rangle > 0$ per qualunque $v \neq O$; *definito negativo* se $\langle v, v \rangle < 0$ per qualunque $v \neq O$; in entrambi i casi è necessariamente non degenerare (perché?). Diremo invece che è *semidefinito positivo* se $\langle v, v \rangle \geq 0$ per ogni $v \in V$ ed esiste un $v_0 \neq O$ con $\langle v_0, v_0 \rangle = 0$; che è *semidefinito negativo* se $\langle v, v \rangle \leq 0$ per ogni $v \in V$, ed esiste un $v_0 \neq O$ tale che $\langle v_0, v_0 \rangle = 0$. Infine, diremo che il prodotto scalare è *indefinito* se non è né (semi)definito positivo né (semi)definito negativo, cioè se esistono $v, w \in V$ tali che $\langle v, v \rangle > 0$ e $\langle w, w \rangle < 0$.

Dimostreremo fra poco (Proposizione 12.1) che ogni prodotto scalare semidefinito è necessariamente degenerare, mentre i prodotti scalari indefiniti possono essere sia degeneri che non degeneri (ma se $\langle \cdot, \cdot \rangle$ è indefinito esiste sempre un vettore $v_0 \neq O$ tale che $\langle v_0, v_0 \rangle = 0$). Prima di procedere, però, vediamo qualche esempio.

ESEMPIO 12.1 Il prodotto scalare canonico è chiaramente un prodotto scalare definito positivo su \mathbb{R}^n . Non utilizzeremo un simbolo specifico per indicare il prodotto scalare canonico; quando v e w sono vettori di \mathbb{R}^n il contesto indicherà chiaramente se con $\langle v, w \rangle$ intenderemo il prodotto scalare canonico o un prodotto scalare diverso.

ESEMPIO 12.2 Consideriamo l'applicazione $\langle \cdot, \cdot \rangle: \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}$ data da

$$\langle v, w \rangle = 2v_1w_1 + v_1w_2 + v_2w_1 + v_2w_2 + v_3w_3.$$

È facile verificare che si tratta di un prodotto scalare su \mathbb{R}^3 : infatti, se fissiamo v (ovvero, consideriamo le coordinate di v come parametri), quello che resta è un polinomio di primo grado senza termini noti nelle coordinate di w , per cui è lineare in w ; analogamente, fissando w , quello che resta risulta essere lineare in v . Anche la simmetria è immediata: infatti, ogni volta che è presente un termine del tipo v_iw_j è presente anche un termine del tipo v_jw_i , ed esattamente con lo stesso coefficiente. Nota che queste osservazioni si possono riassumere dicendo che questo prodotto scalare è dato da una formula del tipo

$$\langle v, w \rangle = w^T S v,$$

dove S è la matrice simmetrica (cioè tale che $S^T = S$)

$$S = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Infine, questo prodotto scalare è definito positivo: infatti

$$\langle v, v \rangle = v_1^2 + (v_1 + v_2)^2 + v_3^2 \geq 0,$$

e inoltre $\langle v, v \rangle = 0$ se e solo se $v_1 = v_1 + v_2 = v_3 = 0$, cioè se e solo se $v = O$.

Osservazione 12.5 Vedremo più oltre (nel Capitolo 16) che la norma al quadrato (definita da $\|v\|^2 = \langle v, v \rangle$) di un prodotto scalare definito positivo può sempre essere scritta come somma di quadrati di opportune combinazioni lineari delle coordinate.

ESEMPIO 12.3 Consideriamo invece l'applicazione $\langle \cdot, \cdot \rangle: \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}$ data da

$$\langle v, w \rangle = v_1 w_1 + v_1 w_2 + v_2 w_1 + v_2 w_2 + v_3 w_3 = w^T S v,$$

dove stavolta S è la matrice

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

Anche in questo caso è chiaro che si tratta di un prodotto scalare; ma, contrariamente a prima, è solo semidefinito positivo. Infatti,

$$\langle v, v \rangle = (v_1 + v_2)^2 + v_3^2 \geq 0,$$

ma $\langle v, v \rangle = 0$ se e solo se $v_1 + v_2 = v_3 = 0$, che non implica $v = 0$, in quanto esistono dei vettori non nulli tali che $\langle v, v \rangle = 0$; per esempio,

$$v = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}.$$

ESEMPIO 12.4 L'applicazione $\langle \cdot, \cdot \rangle: \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}$ data da

$$\langle v, w \rangle = -2v_1 w_1 + v_1 w_2 + v_2 w_1 - v_2 w_2 - v_3 w_3 = w^T S v,$$

dove stavolta S è la matrice

$$S = \begin{pmatrix} -2 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

è chiaramente un prodotto scalare, ma definito negativo. Infatti,

$$\langle v, v \rangle = -v_1^2 - (v_1 - v_2)^2 - v_3^2.$$

ESEMPIO 12.5 L'applicazione $\langle \cdot, \cdot \rangle: \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}$ data da

$$\langle v, w \rangle = v_1 w_2 + v_2 w_1 + v_3 w_3 = w^T S v,$$

dove stavolta S è la matrice

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

è ancora un prodotto scalare, ma indefinito. Infatti,

$$\langle v, v \rangle = 2v_1 v_2 + v_3^2,$$

e per esempio abbiamo

$$\left\langle \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right\rangle = 2, \quad \left\langle \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\rangle = 0, \quad \left\langle \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \right\rangle = -2.$$

Pur essendo indefinito, questo prodotto scalare è non degenere: infatti, $\langle v, w \rangle = 0$ per tutti i $w \in \mathbb{R}^3$ se e solo se $v_1 = v_2 = v_3 = 0$, cioè se e solo se $v = 0$.

Osservazione 12.6 Il considerare prodotti scalari non necessariamente definiti positivi non è masochismo da matematici. Tanto per fare un esempio, nella relatività ristretta di Einstein è fondamentale considerare nello spazio-tempo \mathbb{R}^4 il prodotto scalare

$$\left\langle \begin{pmatrix} x_1 \\ y_1 \\ z_1 \\ t_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \\ z_2 \\ t_2 \end{pmatrix} \right\rangle = x_1 x_2 + y_1 y_2 + z_1 z_2 - t_1 t_2,$$

che è non degenere ma indefinito.

Come annunciato prima, ci sono esempi di prodotti scalari anche su spazi vettoriali che non sono \mathbb{R}^n :

ESEMPIO 12.6 Il prodotto scalare probabilmente più importante in assoluto, oltre al prodotto scalare canonico, è il seguente:

$$\langle p, q \rangle = \int_0^1 p(t)q(t) dt,$$

dove p e q sono due polinomi a coefficienti reali (o anche semplicemente due funzioni continue definite sull'intervallo $[0, 1]$). Le proprietà dell'integrale permettono di dimostrare che si tratta di un prodotto scalare definito positivo, uno dei più utili sia per la matematica pura che per quella applicata (te ne renderai conto quando studierai serie e trasformate di Fourier, o l'Analisi Funzionale).

ESEMPIO 12.7 Un prodotto scalare più alla nostra portata è invece il seguente, definito sullo spazio $\mathbb{R}_2[t]$:

$$\langle p, q \rangle = p(0)q(0) + p(1)q(1) + p(-1)q(-1).$$

Dimostriamo che è un prodotto scalare. Siano $p_1, p_2, q \in \mathbb{R}_2[t]$; allora

$$\begin{aligned}\langle p_1 + p_2, q \rangle &= (p_1 + p_2)(0)q(0) + (p_1 + p_2)(1)q(1) + (p_1 + p_2)(-1)q(-1) \\ &= (p_1(0)q(0) + p_1(1)q(1) + p_1(-1)q(-1)) \\ &\quad + (p_2(0)q(0) + p_2(1)q(1) + p_2(-1)q(-1)) \\ &= \langle p_1, q \rangle + \langle p_2, q \rangle.\end{aligned}$$

Analogamente si dimostra l'omogeneità rispetto alla prima variabile. Siccome $\langle \cdot, \cdot \rangle$ è evidentemente simmetrico, la linearità rispetto alla seconda variabile è conseguenza di quanto già visto (perché?), e quindi abbiamo verificato che è un prodotto scalare. Dimostriamo ora che è definito positivo. Infatti,

$$\forall p \in \mathbb{R}_2[t] \quad \langle p, p \rangle = p(0)^2 + p(1)^2 + p(-1)^2 \geq 0,$$

e inoltre $\langle p, p \rangle = 0$ se e solo se il polinomio p ha almeno le tre radici 0, 1 e -1 . Ma questo può accadere se e solo se $p \equiv 0$, in quanto p è di secondo grado.

ESEMPIO 12.8 Definiamo $\langle \cdot, \cdot \rangle: M_{m,n}(\mathbb{R}) \times M_{m,n}(\mathbb{R}) \rightarrow \mathbb{R}$ ponendo

$$\langle A, B \rangle = \text{tr}(B^T A),$$

dove $\text{tr}: M_{n,n}(\mathbb{R}) \rightarrow \mathbb{R}$ è la *traccia* definita nell'Esercizio 5.6; nota che $B^T A$ è una matrice quadrata di ordine n , per cui possiamo calcolarne la traccia. Vogliamo dimostrare che $\langle \cdot, \cdot \rangle$ è un prodotto scalare definito positivo. Che sia bilineare segue subito dalla linearità della traccia. Poi, è chiaro che $\text{tr}(C^T) = \text{tr}(C)$ per qualunque matrice quadrata C ; quindi

$$\langle B, A \rangle = \text{tr}(A^T B) = \text{tr}((A^T B)^T) = \text{tr}(B^T A) = \langle A, B \rangle.$$

Dunque è un prodotto scalare. Rimane da far vedere che è definito positivo; ma infatti:

$$\begin{aligned}\langle A, A \rangle &= \text{tr}(A^T A) = (a_{11}a_{11} + \cdots + a_{m1}a_{m1}) + \cdots + (a_{1n}a_{1n} + \cdots + a_{mn}a_{mn}) \\ &= \sum_{i=1}^m \sum_{j=1}^n a_{ij}^2 \geq 0,\end{aligned}$$

che inoltre è uguale a zero se e solo se A è la matrice nulla.

12.2 Definizioni ed esempi, caso complesso

Dopo tutta questa serie di definizioni ed esempi sarai ormai ansioso di vedere i prodotti scalari all'opera. Sfortunatamente dobbiamo aspettare ancora un po', in quanto è necessario esaminare anche la situazione sui complessi.

Nel seguito sarà utile (è una promessa) considerare anche prodotti "scalari" definiti su \mathbb{C}^n , o più in generale su spazi vettoriali sui numeri complessi. Il primo tentativo naturale sarebbe introdurre un "prodotto scalare canonico" su \mathbb{C}^n ponendo

$$\forall v, w \in \mathbb{C}^n \quad \langle v, w \rangle = v_1 w_1 + \cdots + v_n w_n.$$

Questa definizione ha però un problema fondamentale: $\langle v, v \rangle$ si guarda bene dall'essere sempre positivo. Anzi, potrebbe benissimo essere negativo, o addirittura immaginario puro (verificalo su degli esempi). Quindi se volessimo definire una lunghezza come radice quadrata di $\langle v, v \rangle$ ci troveremmo in grossi guai. Se ci pensi bene, è lo stesso tipo di problema che avevamo avuto nel tentare di generalizzare al caso complesso la Proposizione 5.11. Allora, il problema era stato risolto (nella Proposizione 11.10) ponendo dei coniugati in posti strategici; qui possiamo operare nello stesso modo.

Definizione 12.7 Il *prodotto hermitiano canonico* $\langle \cdot, \cdot \rangle: \mathbb{C}^n \times \mathbb{C}^n \rightarrow \mathbb{C}$ su \mathbb{C}^n è la funzione data da

$$\langle v, w \rangle = v_1 \bar{w}_1 + \cdots + v_n \bar{w}_n = w^H \cdot v \quad (12.8)$$

per ogni $v = (v_1, \dots, v_n)$, $w = (w_1, \dots, w_n) \in \mathbb{C}^n$, dove il prodotto all'ultimo membro è il solito prodotto righe per colonne, e $w^H = \bar{w}^T$. La *norma* associata al prodotto hermitiano canonico è la funzione $\|\cdot\|: \mathbb{C}^n \rightarrow \mathbb{R}^+$ data da

$$\forall v \in \mathbb{C}^n \quad \|v\| = \sqrt{\langle v, v \rangle} = (|v_1|^2 + \cdots + |v_n|^2)^{1/2}. \quad (12.9)$$

Nota che la norma adesso è ben definita in quanto

$$\langle v, v \rangle = |v_1|^2 + \cdots + |v_n|^2 \geq 0$$

per tutti i vettori $v \in \mathbb{C}^n$.

Il prodotto hermitiano canonico gode di proprietà simili a quelle del prodotto scalare canonico, ma non proprio delle stesse: ogni tanto appare qualche coniugato. Per essere precisi, valgono (a_1) , (a_2) e (b_1) , ma (b_2) è sostituita da

(b_2') per ogni $v, w \in \mathbb{C}^n$ e $\lambda \in \mathbb{C}$ si ha

$$\langle v, \lambda w \rangle = \bar{\lambda} \langle v, w \rangle$$

(anti-omogeneità nella seconda variabile).

Analogamente, la proprietà (c) è sostituita dalla

(c') per ogni $v, w \in \mathbb{C}^n$ si ha $\langle w, v \rangle = \overline{\langle v, w \rangle}$ (hermitianità).

A questo punto abbiamo ovviamente una serie di definizioni simili a quelle viste nel caso reale.

Definizione 12.8 Sia V uno spazio vettoriale su \mathbb{C} . Una *forma sesquilineare* su V è un'applicazione $g: V \times V \rightarrow \mathbb{C}$ tale che

$$\begin{aligned} g(v_1 + v_2, w) &= g(v_1, w) + g(v_2, w), & g(v, w_1 + w_2) &= g(v, w_1) + g(v, w_2), \\ g(\lambda v, w) &= \lambda g(v, w), & g(v, \lambda w) &= \bar{\lambda} g(v, w), \end{aligned}$$

per tutti $v, v_1, v_2, w, w_1, w_2 \in V$ e $\lambda \in \mathbb{C}$.

Definizione 12.9 Sia V uno spazio vettoriale su \mathbb{C} . Un *prodotto hermitiano* su V è una forma sesquilineare g *hermitiana*, cioè tale che

$$\forall v, w \in V \quad g(w, v) = \overline{g(v, w)}.$$

Di nuovo, in tal caso scriveremo sempre $\langle v, w \rangle$ invece di $g(v, w)$.

Definizione 12.10 Sia $\langle \cdot, \cdot \rangle$ un prodotto hermitiano su uno spazio vettoriale V . Il *nucleo* di $\langle \cdot, \cdot \rangle$ è l'insieme

$$V^\perp = \{v \in V \mid \langle v, w \rangle = 0 \quad \forall w \in V\}.$$

Come nel caso reale si dimostra che V^\perp è un sottospazio di V .

Definizione 12.11 Sia V uno spazio vettoriale su \mathbb{C} . Un prodotto hermitiano $\langle \cdot, \cdot \rangle$ su V è detto *non degenerare* se $V^\perp = \{0\}$; *degenerare* altrimenti.

Nota che se $\langle \cdot, \cdot \rangle$ è un prodotto hermitiano su V , allora $\langle v, v \rangle \in \mathbb{R}$ per qualunque $v \in V$ (esercizio); quindi le definizioni successive sono del tutto analoghe a quelle nel caso reale.

Definizione 12.12 Sia $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{C}$ un prodotto hermitiano su uno spazio vettoriale V su \mathbb{C} . Questo prodotto hermitiano è *definito positivo* se $\langle v, v \rangle > 0$ per qualunque $v \neq 0$; *definito negativo* se $\langle v, v \rangle < 0$ per qualunque $v \neq 0$; in entrambi i casi è necessariamente non degenerare. Diremo invece che è *semidefinito positivo* se $\langle v, v \rangle \geq 0$ per ogni $v \in V$ ma esiste $v_0 \neq 0$ tale che $\langle v_0, v_0 \rangle = 0$; che è *semidefinito negativo* se $\langle v, v \rangle \leq 0$ per ogni $v \in V$ ma esiste $v_0 \neq 0$ tale che $\langle v_0, v_0 \rangle = 0$. Infine, diremo che il prodotto hermitiano è *indefinito* se non è né (semi)definito positivo né (semi)definito negativo, cioè se esistono $v, w \in V$ tali che $\langle v, v \rangle > 0$ e $\langle w, w \rangle < 0$.

ESEMPIO 12.9 Sia $\langle \cdot, \cdot \rangle: \mathbb{C}^2 \times \mathbb{C}^2 \rightarrow \mathbb{C}$ data da

$$\langle v, w \rangle = 3v_1\bar{w}_1 + iv_1\bar{w}_2 - iv_2\bar{w}_1 + v_2\bar{w}_2 = w^H S v,$$

dove S è la matrice hermitiana (cioè tale che $S^H = S$)

$$S = \begin{vmatrix} 3 & -i \\ i & 1 \end{vmatrix}.$$

Questo è un buon esempio di prodotto hermitiano su \mathbb{C}^2 . Che sia una forma sesquilineare si vede esattamente come nel caso reale (con l'avvertenza di notare che i coniugati sono tutti sulle coordinate di w , e nessuno sulle coordinate di v); che sia hermitiano segue dall'osservazione che a ogni termine del tipo $v_h \bar{w}_k$ corrisponde un termine del tipo $v_k \bar{w}_h$ con coefficiente *coniugato* (in particolare, i coefficienti dei termini $v_h \bar{w}_h$ sono tutti reali). Infine, è definito positivo: infatti,

$$\langle v, v \rangle = 3|v_1|^2 + 2\operatorname{Re}(iv_1\bar{v}_2) + |v_2|^2 = 2|v_1|^2 + |iv_1 + v_2|^2 \geq 0,$$

e $\langle v, v \rangle = 0$ se e solo se $v_1 = iv_1 + v_2 = 0$, cioè se e solo se $v = 0$.

Nel seguito, molti dei risultati che dimostreremo varranno sia per prodotti scalari sui reali che per prodotti hermitiani sui complessi, con enunciati e dimostrazioni praticamente identici. In tal caso, talvolta verrà riportata solo la dimostrazione per il caso complesso; la dimostrazione per il caso reale la si ottiene semplicemente dimenticandosi di scrivere (o leggere) gli eventuali segni di coniugato e di parte reale presenti. Ci saranno ovviamente anche delle situazioni in cui gli enunciati o le dimostrazioni, o entrambi saranno diversi; in tal caso, le differenze saranno ben indicate.

Infine, per evitare di utilizzare in continuazione frasi complicate del tipo "consideriamo un prodotto scalare (hermitiano) su uno spazio vettoriale reale (complesso)", introduciamo la seguente definizione:

Definizione 12.13 Uno *spazio vettoriale metrico* reale o complesso è uno spazio vettoriale V su \mathbb{R} o su \mathbb{C} provvisto di un prodotto scalare (o hermitiano nel caso complesso) $\langle \cdot, \cdot \rangle$ definito positivo. La *norma* $\|\cdot\|: V \rightarrow \mathbb{R}^+$ in questo spazio vettoriale metrico è definita da

$$\|v\| = \sqrt{\langle v, v \rangle}.$$

La norma $\|v\|$ di un vettore v si dice anche *lunghezza* di v .

12.3 Disuguaglianza di Cauchy-Schwarz

Possiamo finalmente cominciare a fare qualcosa di concreto con prodotti scalari e hermitiani. Cominciamo con un piccolo risultato introduttivo promesso nel Paragrafo 12.1:

Proposizione 12.1 Sia $\langle \cdot, \cdot \rangle$ un prodotto scalare (hermitiano) su uno spazio vettoriale V su \mathbb{R} (o \mathbb{C}). Allora:

- (i) Se $\langle v, v \rangle \neq 0$ per tutti $v \in V$ diversi dal vettore nullo, allora $\langle \cdot, \cdot \rangle$ è non degenerare;
- (ii) Supponiamo che $\langle \cdot, \cdot \rangle$ sia non degenerare. Allora $\langle v_1, w \rangle = \langle v_2, w \rangle$ per tutti $w \in V$ se e solo se $v_1 = v_2$;

- (iii) $\langle \cdot, \cdot \rangle$ è (semi)definito negativo se e solo se $-\langle \cdot, \cdot \rangle$ è (semi)definito positivo;
 (iv) Se $\langle \cdot, \cdot \rangle$ è semidefinito positivo o negativo, allora è degenere;
 (v) Se $\langle \cdot, \cdot \rangle$ è indefinito, allora esiste $v \neq O$ tale che $\langle v, v \rangle = 0$.

Dimostrazione. (i) Se $\langle \cdot, \cdot \rangle$ fosse degenere, esisterebbe $v_0 \in V^\perp \setminus \{O\}$, e quindi in particolare avremmo $\langle v_0, v_0 \rangle = 0$ con $v_0 \neq O$.

(ii) Se $\langle v_1, w \rangle = \langle v_2, w \rangle$, allora $\langle v_2 - v_1, w \rangle = 0$; siccome questo deve valere per tutti i w e $V^\perp = \{O\}$ per ipotesi, ne segue che $v_2 - v_1 = O$, cioè $v_1 = v_2$.

(iii) Ovvio.

(iv) Supponiamo $\langle \cdot, \cdot \rangle$ semidefinito positivo; grazie a (iii), il caso semidefinito negativo è identico. Siccome il nostro prodotto scalare (o hermitiano) è soltanto semidefinito, esiste un $v_0 \neq O$ tale che $\langle v_0, v_0 \rangle = 0$; vogliamo dimostrare che $\langle v_0, w \rangle = 0$ per tutti i $w \in V$.

Supponiamo per assurdo che esista $w_0 \in V$ tale che $\langle v_0, w_0 \rangle \neq 0$. L'idea è che in tal caso dobbiamo trovare una contraddizione esaminando la retta passante per w_0 in direzione v_0 .

Assumiamo per il momento $a = \operatorname{Re}\langle v_0, w_0 \rangle \neq 0$. Il nostro prodotto scalare (o hermitiano) è semidefinito positivo; quindi per ogni $t \in \mathbb{R}$ abbiamo

$$\begin{aligned} 0 \leq \langle w_0 + tv_0, w_0 + tv_0 \rangle &= \langle w_0, w_0 \rangle + t[\langle w_0, v_0 \rangle + \langle v_0, w_0 \rangle] + t^2 \langle v_0, v_0 \rangle \\ &= \langle w_0, w_0 \rangle + 2t \operatorname{Re}\langle v_0, w_0 \rangle, \end{aligned}$$

(in quanto $\langle v_0, v_0 \rangle = 0$ per ipotesi), cioè $-2at \leq \langle w_0, w_0 \rangle$ qualunque sia $t \in \mathbb{R}$, che chiaramente è impossibile. Se siamo su \mathbb{R} abbiamo finito; se siamo su \mathbb{C} potrebbe ancora aversi $\operatorname{Re}\langle v_0, w_0 \rangle = 0$ ma $\langle v_0, w_0 \rangle \neq 0$. In tal caso $\operatorname{Im}\langle v_0, w_0 \rangle \neq 0$, e quindi

$$\operatorname{Re}\langle v_0, iw_0 \rangle = \operatorname{Re}[-i\langle v_0, w_0 \rangle] = \operatorname{Im}\langle v_0, w_0 \rangle \neq 0.$$

Dunque possiamo ripetere il ragionamento precedente con iw_0 al posto di w_0 , giungendo di nuovo a una contraddizione. Quindi un tale w_0 non può esistere, e il nostro prodotto scalare (o hermitiano) è degenere.

(v) Siano $v_0, v_1 \in V$ tali che $\langle v_0, v_0 \rangle < 0$ e $\langle v_1, v_1 \rangle > 0$; l'idea è che il vettore che cerchiamo deve trovarsi da qualche parte nella retta che congiunge v_0 e v_1 .

Poniamo $v_t = v_0 + t(v_1 - v_0)$, e definiamo $f: [0, 1] \rightarrow \mathbb{R}$ con

$$f(t) = \langle v_t, v_t \rangle.$$

Chiaramente, $f(0) = \langle v_0, v_0 \rangle < 0$ e $f(1) = \langle v_1, v_1 \rangle > 0$. Inoltre, f è una funzione continua, in quanto è un polinomio di secondo grado in t :

$$\begin{aligned} f(t) &= \langle v_0 + t(v_1 - v_0), v_0 + t(v_1 - v_0) \rangle \\ &= \langle v_0, v_0 + t(v_1 - v_0) \rangle + t\langle v_1 - v_0, v_0 + t(v_1 - v_0) \rangle \\ &= \langle v_0, v_0 \rangle + t[\langle v_0, v_1 - v_0 \rangle + \langle v_1 - v_0, v_0 \rangle] + t^2 \langle v_1 - v_0, v_1 - v_0 \rangle. \end{aligned}$$

Per il Teorema dei valori intermedi (Teorema 1.5), esiste $t_0 \in (0, 1)$ tale che $f(t_0) = 0$, cioè tale che $\langle v_{t_0}, v_{t_0} \rangle = 0$. Quindi se dimostriamo che $v_{t_0} \neq O$ abbiamo finito. Se fosse $v_{t_0} = O$, avremmo $v_1 = \lambda v_0$ per $\lambda = (t_0 - 1)/t_0 \in \mathbb{R}^*$; ma allora

$$\langle v_1, v_1 \rangle = \langle \lambda v_0, \lambda v_0 \rangle = |\lambda|^2 \langle v_0, v_0 \rangle < 0,$$

mentre sappiamo che $\langle v_1, v_1 \rangle > 0$, contraddizione. \square

La nostra idea è che su uno spazio vettoriale metrico è possibile, usando il prodotto scalare (o hermitiano) e la norma associata, studiare la geometria metrica in modo simile a quanto avviene nel piano o, nello spazio euclideo. Per far ciò abbiamo però bisogno di definire la distanza fra due punti, e l'angolo fra due vettori (o direzioni). Per la distanza è facile:

Definizione 12.14 Sia V uno spazio vettoriale metrico (su \mathbb{R} o su \mathbb{C} , è indifferente). Allora la distanza $d(v, w)$ fra due punti $v, w \in V$ è data da

$$d(v, w) = \|w - v\|.$$

ESEMPIO 12.10 Se $\langle \cdot, \cdot \rangle$ è il prodotto scalare canonico su \mathbb{R}^2 , la Definizione 12.14 dà esattamente la distanza euclidea, come visto in (12.2). Considerando invece degli altri prodotti scalari definiti positivi otteniamo distanze diverse. Per esempio, col prodotto scalare definito positivo

$$\langle v, w \rangle = 2v_1w_1 + v_1w_2 + v_2w_1 + 3v_2w_2$$

otteniamo la distanza

$$\begin{aligned} d(v, w) &= \sqrt{2(w_1 - v_1)^2 + 2(w_1 - v_1)(w_2 - v_2) + 3(w_2 - v_2)^2} \\ &= \sqrt{(w_1 - v_1)^2 + (w_1 + w_2 - v_1 - v_2)^2 + 2(w_2 - v_2)^2}. \end{aligned}$$

Osservazione 12.7 Il voler considerare distanze diverse dalla solita distanza euclidea non è un capriccio da matematici; al contrario, è essenziale per descrivere correttamente il mondo reale. Supponi di voler andare a trovare un amico che abita a un chilometro di distanza (euclidea, ovvero in linea d'aria) da casa tua. Se ci pensi un attimo, ti renderai conto che quel chilometro non indica minimamente la fatica che devi fare per raggiungere casa sua: la situazione cambia sensibilmente a seconda che abitate in pianura oppure separati da un chilometro di salita. Se poi il tuo amico si trovasse esattamente un chilometro sopra (o sotto) di te, non avresti molte speranze di andarlo a trovare: per quel che ti riguarda, sarebbe a una distanza enorme. Il fatto è che se vuoi misurare la distanza fra due punti tenendo presente l'effettiva fatica che devi fare per andare dal primo al secondo, spostamenti nelle direzioni destra-sinistra e avanti-dietro hanno un peso ben minore di spostamenti nella direzione alto-basso. Diventa quindi naturale introdurre una distanza diversa da quella euclidea, una distanza del tipo

$$d(v, w) = \sqrt{(w_1 - v_1)^2 + (w_2 - v_2)^2 + 10000(w_3 - v_3)^2},$$

che proviene da un prodotto scalare diverso da quello canonico.

Ovviamente, perché la funzione introdotta nella Definizione 12.14 meriti il nome di “distanza” deve verificare le proprietà principali che noi intuitivamente associamo al concetto di distanza fra punti. Per esempio, un punto deve essere a distanza zero da se stesso: e infatti, $d(v_0, v_0) = \|v_0 - v_0\| = \|O\| = 0$. Poi, la distanza fra v_0 e v_1 dev'essere uguale alla distanza fra v_1 e v_0 : e infatti

$$d(v_0, v_1) = \|v_1 - v_0\| = \|-(v_0 - v_1)\| = \|v_0 - v_1\| = d(v_1, v_0),$$

(dove abbiamo usato il fatto che $\| -v \| = \|v\|$ per ogni $v \in V$; vedi la Proposizione 12.2.(iii) qui sotto). Infine, se andiamo da v_0 direttamente a v_2 dobbiamo percorrere al più una distanza uguale a quella percorsa andando da v_0 a v_1 e poi da v_1 a v_2 , qualunque sia v_1 ; in simboli,

$$d(v_0, v_2) \leq d(v_0, v_1) + d(v_1, v_2). \quad (12.10)$$

Anche questa proprietà è verificata dalla nostra distanza, ma non è ovvia come le altre due. La dimostrazione è contenuta nella seguente Proposizione, che raccoglie tutte le proprietà fondamentali della norma in uno spazio vettoriale metrico:

Proposizione 12.2 Sia V uno spazio vettoriale metrico con norma $\|\cdot\|$. Allora:

- (i) $\|v\| = 0$ se e solo se $v = O$;
- (ii) $\|v\| > 0$ per ogni $v \neq O$;
- (iii) $\|\lambda v\| = |\lambda| \|v\|$ per ogni scalare λ e $v \in V$, e

$$\|v + w\|^2 = \|v\|^2 + 2 \operatorname{Re}\langle v, w \rangle + \|w\|^2$$

per ogni $v, w \in V$;

- (iv) per ogni $v, w \in V$ si ha

$$|\langle v, w \rangle| \leq \|v\| \|w\|,$$

(disuguaglianza di Cauchy-Schwarz), e l'uguaglianza vale se e solo se v e w sono linearmente dipendenti;

- (v) per ogni $v, w \in V$ si ha

$$\left| \|v\| - \|w\| \right| \leq \|v + w\| \leq \|v\| + \|w\|,$$

(disuguaglianza triangolare);

- (vi) se V è uno spazio vettoriale su \mathbb{R} si ha

$$\langle v, w \rangle = \frac{1}{4} [\|v + w\|^2 - \|v - w\|^2]$$

per ogni $v, w \in V$. Se invece V è uno spazio vettoriale su \mathbb{C} per ogni $v, w \in V$ si ha

$$\operatorname{Re}\langle v, w \rangle = \frac{1}{4} [\|v + w\|^2 - \|v - w\|^2], \quad \operatorname{Im}\langle v, w \rangle = \frac{1}{4} [\|v + iw\|^2 - \|v - iw\|^2].$$

Osservazione 12.8 La disuguaglianza triangolare implica che per ogni v_0, v_1 e $v_2 \in V$ si ha

$$\|v_2 - v_0\| \leq \|v_1 - v_0\| + \|v_2 - v_1\|$$

(basta prendere $v = v_1 - v_0$ e $w = v_2 - v_1$), per cui (12.10) è verificata. Il nome “disuguaglianza triangolare” è dovuto al fatto che può essere interpretata come l'affermazione che un lato del triangolo di vertici O, v e $v + w$ è più corto della somma e più lungo della differenza degli altri due lati (vedi la Figura 12.3), un risultato classico della geometria euclidea dei triangoli.

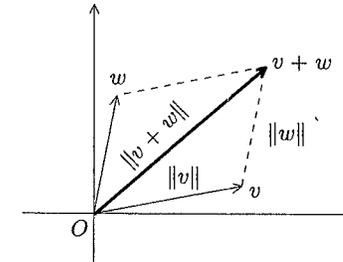


Figura 12.3 La disuguaglianza triangolare.

Osservazione 12.9 La Proposizione 12.2.(vi) ci dice una cosa importante: in uno spazio vettoriale metrico, la norma determina completamente il prodotto scalare. In altri termini, se due prodotti scalari (o hermitiani) hanno la stessa norma, allora coincidono. Tra l'altro, come vedrai dalla dimostrazione, la Proposizione 12.2.(vi) vale per prodotti scalari (hermitiani) qualunque, anche non definiti positivi.

Dimostrazione. La (i) e la (ii) seguono dalla definizione di prodotto scalare (hermitiano) definito positivo.

- (iii) Infatti

$$\|\lambda v\| = \sqrt{\langle \lambda v, \lambda v \rangle} = \sqrt{|\lambda|^2 \langle v, v \rangle} = |\lambda| \|v\|.$$

Inoltre,

$$\begin{aligned} \|v + w\|^2 &= \langle v + w, v + w \rangle = \langle v, v \rangle + \langle v, w \rangle + \langle w, v \rangle + \langle w, w \rangle \\ &= \|v\|^2 + 2 \operatorname{Re}\langle v, w \rangle + \|w\|^2. \end{aligned}$$

(iv) Se $v = O$ oppure $w = O$ è ovvio; quindi possiamo supporli entrambi non nulli. Siccome $\langle \cdot, \cdot \rangle$ è definito positivo, si ha

$$\begin{aligned} 0 \leq \|av + bw\|^2 &= \langle av + bw, av + bw \rangle = a\langle v, av + bw \rangle + b\langle w, av + bw \rangle \\ &= |a|^2 \langle v, v \rangle + a\bar{b}\langle v, w \rangle + \bar{a}b\langle w, v \rangle + |b|^2 \langle w, w \rangle \\ &= |a|^2 \|v\|^2 + 2 \operatorname{Re}(a\bar{b}\langle v, w \rangle) + |b|^2 \|w\|^2 \end{aligned}$$

per ogni coppia di scalari a e b . In particolare, prendendo $a = \|w\|^2$ e $b = -\langle v, w \rangle$ otteniamo

$$\begin{aligned} 0 &\leq \|w\|^4 \|v\|^2 - 2\|w\|^2 |\langle v, w \rangle|^2 + \|w\|^2 |\langle v, w \rangle|^2 \\ &= \|w\|^2 (\|v\|^2 \|w\|^2 - |\langle v, w \rangle|^2), \end{aligned}$$

per cui dividendo per $\|w\|^2$ otteniamo la disuguaglianza di Cauchy-Schwarz. L'uguaglianza vale se e solo se $av + bw = 0$, cioè se e solo se v e w sono linearmente dipendenti.

(v) Prima di tutto, notiamo che grazie al punto (iv) e alla Proposizione 11.3.(iii) si ha

$$|\operatorname{Re}\langle v, w \rangle| \leq |\langle v, w \rangle| \leq \|v\| \|w\|.$$

Quindi

$$\begin{aligned} \|\|v\| - \|w\|\|^2 &= \|v\|^2 + \|w\|^2 - 2\|v\| \|w\| \leq \|v\|^2 + \|w\|^2 + 2\operatorname{Re}\langle v, w \rangle = \|v + w\|^2 \\ &\leq \|v\|^2 + \|w\|^2 + 2\|v\| \|w\| = (\|v\| + \|w\|)^2, \end{aligned}$$

e quindi abbiamo dimostrato la disuguaglianza triangolare.

(vi) Si tratta di un facile conto; svolgiamo qui quello nel caso complesso, lasciando il caso reale come esercizio. Abbiamo

$$\begin{aligned} \|v + w\|^2 &= \|v\|^2 + \|w\|^2 + 2\operatorname{Re}\langle v, w \rangle, & \|v + iw\|^2 &= \|v\|^2 + \|w\|^2 + 2\operatorname{Im}\langle v, w \rangle, \\ \|v - w\|^2 &= \|v\|^2 + \|w\|^2 - 2\operatorname{Re}\langle v, w \rangle, & \|v - iw\|^2 &= \|v\|^2 + \|w\|^2 - 2\operatorname{Im}\langle v, w \rangle, \end{aligned}$$

e l'asserto segue sommando e sottraendo opportunamente. \square

Dunque la nostra distanza su uno spazio vettoriale metrico soddisfa effettivamente le proprietà basilari che si richiedono a una distanza. Inoltre, sempre grazie alla disuguaglianza di Cauchy-Schwarz, almeno nel caso reale possiamo definire l'angolo fra due vettori:

Definizione 12.15 Sia V uno spazio vettoriale metrico sui reali, e $v, w \in V$ due vettori non nulli. Diremo *angolo* fra v e w quel numero reale $\widehat{vw} \in [0, \pi]$ tale che

$$\cos \widehat{vw} = \frac{\langle v, w \rangle}{\|v\| \|w\|}.$$

Grazie alla disuguaglianza di Cauchy-Schwarz, questo rapporto è in modulo minore o uguale a 1, per cui è il coseno di qualcosa; e siccome il coseno è una funzione iniettiva nell'intervallo $[0, \pi]$, questo definisce in modo univoco l'angolo fra v e w .

Dunque abbiamo recuperato anche la definizione di angolo in uno spazio vettoriale metrico. Nota che il caso di uguaglianza nella Proposizione 12.2.(iv) ci dice che due vettori v e w sono linearmente dipendenti se e solo se il coseno del loro angolo è uguale a ± 1 , cioè se e solo se formano un angolo di 0 o π radianti, com'è giusto che

sia. D'altro canto, formano un angolo di $\pi/2$ se e solo se il coseno è zero, cioè se e solo se $\langle v, w \rangle = 0$. Questo ci porta alla seguente

Definizione 12.16 Sia $\langle \cdot, \cdot \rangle$ un prodotto scalare (hermitiano) su uno spazio vettoriale V . Due vettori $v, w \in V$ si dicono *ortogonali* (o *perpendicolari*) se $\langle v, w \rangle = 0$, e scriveremo $v \perp w$. Se poi $S_1, S_2 \subset V$ sono due sottoinsiemi di V tali che tutti gli elementi di S_1 sono ortogonali a tutti gli elementi di S_2 diremo che S_1 è *ortogonale* a S_2 , e scriveremo $S_1 \perp S_2$. Se $S_1 = \{v_1\}$, scriveremo $v_1 \perp S_2$ invece di $\{v_1\} \perp S_2$.

12.4 Basi ortogonali

Uno dei concetti più utili per lavorare con gli spazi vettoriali è stato senza dubbio quello di base; vediamo se riusciamo ad adattarlo in modo da renderlo utile anche nel caso degli spazi vettoriali metrici. Cominciamo col seguente

Lemma 12.3 Sia V uno spazio vettoriale metrico, e $v_1, \dots, v_k \in V$ vettori non nulli ortogonali a due a due, cioè tali che $v_i \perp v_j$ non appena $i \neq j$. Allora v_1, \dots, v_k sono linearmente indipendenti. In particolare, se $\dim V = n$ allora n vettori non nulli a due a due ortogonali sono automaticamente una base di V .

Dimostrazione. Sia

$$\alpha_1 v_1 + \dots + \alpha_k v_k = 0$$

una relazione di dipendenza lineare fra v_1, \dots, v_k . Moltiplicando scalarmente per v_j otteniamo

$$0 = \langle 0, v_j \rangle = \alpha_1 \langle v_1, v_j \rangle + \dots + \alpha_j \langle v_j, v_j \rangle + \dots + \alpha_k \langle v_k, v_j \rangle = \alpha_j \|v_j\|^2,$$

perché $\langle v_i, v_j \rangle = 0$ se $i \neq j$; quindi $\alpha_j = 0$. Siccome questo valeva per qualunque j , tutti gli α_j sono zero, e dunque v_1, \dots, v_k sono linearmente indipendenti. \square

Questo risultato ci porta a identificare il tipo di base che ci interessa:

Definizione 12.17 Una *base ortogonale* di uno spazio vettoriale metrico V è una base $\{v_1, \dots, v_n\}$ di V composta da vettori a due a due ortogonali, ovvero tali che $\langle v_i, v_j \rangle = 0$ per $1 \leq i \neq j \leq n$.

ESEMPIO 12.11 La base canonica $\{e_1, e_2\}$ è una base ortogonale di \mathbb{R}^2 col prodotto scalare canonico. Un'altra base ortogonale è

$$\left\{ \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} -3 \\ 1 \end{pmatrix} \right\},$$

e se ne possono trovare facilmente infinite altre.

Vi è anche un'altro tipo di base ancora più speciale:

Definizione 12.18 Una base ortonormale di uno spazio vettoriale metrico V è una base ortogonale composta da vettori di lunghezza unitaria (ovvero, con una terminologia che useremo raramente, da *versori*).

In altri termini, una base $\{v_1, \dots, v_n\}$ di uno spazio vettoriale metrico è una base ortonormale se e solo se

$$\langle v_i, v_j \rangle = \delta_{ij} = \begin{cases} 1 & \text{se } i = j, \\ 0 & \text{se } i \neq j. \end{cases} \quad (12.11)$$

ESEMPIO 12.12 La base canonica è sempre una base ortonormale di \mathbb{R}^n rispetto al prodotto scalare canonico.

ESEMPIO 12.13 Un'altra base ortonormale di \mathbb{R}^2 rispetto al prodotto scalare canonico è data da

$$\left\{ \left| \begin{array}{c} 1/\sqrt{10} \\ 3/\sqrt{10} \end{array} \right|, \left| \begin{array}{c} -3/\sqrt{10} \\ 1/\sqrt{10} \end{array} \right| \right\}.$$

ESEMPIO 12.14 La base canonica non è una base ortonormale per il prodotto scalare definito positivo dell'Esempio 12.2 su \mathbb{R}^3 (non è neppure ortogonale: $\langle e_1, e_2 \rangle = 1$). Per trovare una base ortonormale $\{v_1, v_2, v_3\}$ per questo prodotto scalare cominciamo prendendo e_1 . Siccome $\langle e_1, e_1 \rangle = 2$, il vettore $v_1 = e_1/\sqrt{2}$ ha norma 1. Il secondo vettore v_2 dev'essere ortogonale a v_1 ; quindi se scriviamo $v_2 = (x, y, z)$ si deve avere

$$\langle v_1, v_2 \rangle = \frac{2}{\sqrt{2}}x + \frac{1}{\sqrt{2}}y = 0.$$

Ricordandosi che $\|v_2\| = 1$, una possibile scelta è $v_2 = (1/\sqrt{2}, -2/\sqrt{2}, 0)$. Le coordinate del terzo vettore v_3 devono soddisfare il sistema

$$\begin{cases} \langle v_1, v_3 \rangle = \frac{2}{\sqrt{2}}x + \frac{1}{\sqrt{2}}y = 0, \\ \langle v_2, v_3 \rangle = -\frac{1}{\sqrt{2}}y = 0, \end{cases}$$

in modo che $\|v_3\| = 1$; una scelta possibile è $v_3 = e_3$. Dunque una base ortonormale per questo prodotto scalare è data da

$$\left\{ \left| \begin{array}{c} 1/\sqrt{2} \\ 0 \\ 0 \end{array} \right|, \left| \begin{array}{c} 1/\sqrt{2} \\ -2/\sqrt{2} \\ 0 \end{array} \right|, \left| \begin{array}{c} 0 \\ 0 \\ 1 \end{array} \right| \right\}.$$

Ogni vettore di uno spazio vettoriale si scrive in maniera unica come combinazione lineare degli elementi di una base, e i coefficienti sono le coordinate del vettore rispetto a quella base. Nel caso delle basi ortogonali, le coordinate si trovano immediatamente usando il prodotto scalare:

Proposizione 12.4 Sia $\{v_1, \dots, v_n\}$ una base ortogonale di uno spazio vettoriale metrico V . Allora

$$v = \frac{\langle v, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 + \dots + \frac{\langle v, v_n \rangle}{\langle v_n, v_n \rangle} v_n \quad (12.12)$$

per ogni $v \in V$. In particolare, se $\{v_1, \dots, v_n\}$ è una base ortonormale si ha

$$\forall v \in V \quad v = \langle v, v_1 \rangle v_1 + \dots + \langle v, v_n \rangle v_n. \quad (12.13)$$

Dimostrazione. Siccome $\{v_1, \dots, v_n\}$ è una base, esistono unici gli scalari $\alpha_1, \dots, \alpha_n$ tali che

$$v = \alpha_1 v_1 + \dots + \alpha_n v_n.$$

Moltiplicando scalarmente per v_j , otteniamo $\langle v, v_j \rangle = \alpha_j \langle v_j, v_j \rangle$, grazie all'ortogonalità della base, ed è fatta. \square

Dunque il vettore $v \in V$ è completamente individuato dai numeri $\langle v, v_j \rangle / \langle v_j, v_j \rangle$, che talvolta sono detti *coefficienti di Fourier* di v rispetto alla base $\{v_1, \dots, v_n\}$. In particolare, quindi, la sua norma deve essere esprimibile in termini di questi coefficienti. E infatti:

Corollario 12.5 Sia $\{v_1, \dots, v_n\}$ una base ortogonale di uno spazio vettoriale metrico V . Allora

$$\forall v, w \in V \quad \langle v, w \rangle = \frac{\langle v, v_1 \rangle \langle v_1, w \rangle}{\langle v_1, v_1 \rangle} + \dots + \frac{\langle v, v_n \rangle \langle v_n, w \rangle}{\langle v_n, v_n \rangle} \quad (12.14)$$

$$\forall v \in V \quad \|v\|^2 = \frac{|\langle v, v_1 \rangle|^2}{\|v_1\|^2} + \dots + \frac{|\langle v, v_n \rangle|^2}{\|v_n\|^2}. \quad (12.15)$$

In particolare, se $\{v_1, \dots, v_n\}$ è una base ortonormale di V si ha

$$\forall v, w \in V \quad \langle v, w \rangle = \langle v, v_1 \rangle \langle v_1, w \rangle + \dots + \langle v, v_n \rangle \langle v_n, w \rangle \quad (12.16)$$

(formula di Parseval) e

$$\forall v \in V \quad \|v\|^2 = |\langle v, v_1 \rangle|^2 + \dots + |\langle v, v_n \rangle|^2 \quad (12.17)$$

(Teorema di Pitagora).

Dimostrazione. Moltiplicando scalarmente la (12.12) e la (12.13) per w si ottengono le (12.14) e (12.16); le (12.15) e (12.17) seguono prendendo $w = v$. \square

A questo punto si pone il fondamentale problema dell'esistenza di basi ortonormali. La risposta è che non solo ci sono sempre, ma si possono costruire a partire da una base qualunque con una procedura effettiva che conserva molte proprietà (per esempio gli span) della base di partenza. Questa procedura, che è descritta nella dimostrazione del prossimo Teorema, è detta *procedimento di ortogonalizzazione di Gram-Schmidt*.

Teorema 12.6 Sia V uno spazio vettoriale metrico, e $v_1, \dots, v_r \in V$ linearmente indipendenti. Allora esistono $w_1, \dots, w_r \in V$ che soddisfano le seguenti proprietà:

- (i) $\text{Span}(w_1, \dots, w_j) = \text{Span}(v_1, \dots, v_j)$ per ogni $j = 1, \dots, r$;
- (ii) w_j è ortogonale a ogni elemento di $\text{Span}(w_1, \dots, w_{j-1})$ qualunque sia $j = 2, \dots, r$;
- (iii) $\langle w_j, v_j \rangle > 0$ per $j = 1, \dots, r$.

Gli elementi w_1, \dots, w_r sono univocamente determinati a meno di fattori scalari positivi; in altri termini, se w'_1, \dots, w'_r sono altri elementi di V che soddisfano le proprietà (i)–(iii) allora $w'_j = \alpha_j w_j$ con $\alpha_j > 0$ per $j = 1, \dots, r$. Infine, gli elementi w_1, \dots, w_r sono ottenuti ponendo $w_1 = v_1$ e

$$w_j = v_j - \sum_{h=1}^{j-1} \frac{\langle v_j, w_h \rangle}{\langle w_h, w_h \rangle} w_h, \quad (12.18)$$

per $j = 2, \dots, r$.

Dimostrazione. Procediamo per induzione su r . Per $r = 1$, la (i) ci dice che dev'essere $w_1 = cv_1$ per qualche $c \neq 0$. La (ii) non fornisce alcuna condizione, e la (iii) diventa $c\|v_1\|^2 > 0$, cioè $c > 0$, e quindi in questo caso ci siamo.

Supponiamo allora il Teorema vero per $r - 1$ vettori, e vediamo di dimostrarlo per r vettori. Abbiamo quindi costruito w_1, \dots, w_{r-1} usando la (12.18); dobbiamo trovare w_r . Prima di tutto notiamo che, essendo i vettori v_1, \dots, v_r linearmente indipendenti, la dimensione di $\text{Span}(v_1, \dots, v_r)$ è esattamente r . Siccome, per ipotesi induttiva, $\{w_1, \dots, w_{r-1}, v_r\}$ è un sistema di generatori di $\text{Span}(v_1, \dots, v_r)$, questo implica che w_1, \dots, w_{r-1} e v_r sono linearmente indipendenti. In particolare, il nostro w_r dev'essere della forma

$$w_r = \alpha v_r - \beta_1 w_1 - \dots - \beta_{r-1} w_{r-1},$$

dove $\alpha, \beta_1, \dots, \beta_{r-1}$ sono degli scalari che dobbiamo determinare in modo che (i)–(iii) valgano. Prima di tutto, la (i) implica $\alpha \neq 0$. La condizione (ii) è equivalente (Esercizio 12.14) a dire che w_r dev'essere ortogonale a w_1, \dots, w_{r-1} (i quali a loro volta sono ortogonali a due a due, per ipotesi induttiva); quindi per $j = 1, \dots, r - 1$ si deve avere

$$\begin{aligned} 0 &= \langle w_r, w_j \rangle = \alpha \langle v_r, w_j \rangle - \beta_1 \langle w_1, w_j \rangle - \dots - \beta_{r-1} \langle w_{r-1}, w_j \rangle \\ &= \alpha \langle v_r, w_j \rangle - \beta_j \langle w_j, w_j \rangle, \end{aligned}$$

cioè $\beta_j = \alpha \langle v_r, w_j \rangle / \langle w_j, w_j \rangle$. Quindi

$$w_r = \alpha \left[v_r - \sum_{h=1}^{r-1} \frac{\langle v_r, w_h \rangle}{\langle w_h, w_h \rangle} w_h \right].$$

Dobbiamo scoprire per quali valori di α la (iii) è verificata. Notiamo innanzitutto che

$$\langle w_r, v_r \rangle = \alpha \left[\|v_r\|^2 - \sum_{h=1}^{r-1} \frac{|\langle v_r, w_h \rangle|^2}{\|w_h\|^2} \right]. \quad (12.19)$$

Vogliamo dimostrare che $\langle w_r, v_r \rangle > 0$ se e solo se $\alpha > 0$; per far ciò dobbiamo trovare il segno della quantità fra parentesi quadre. Per costruzione, $v_r \in \text{Span}(w_1, \dots, w_r)$, e $\{w_1, \dots, w_r\}$ è una base ortogonale di $\text{Span}(w_1, \dots, w_r)$. Quindi la (12.15) ci dice che

$$\begin{aligned} \|v_r\|^2 &= \sum_{h=1}^r \frac{|\langle v_r, w_h \rangle|^2}{\|w_h\|^2} \\ &= \sum_{h=1}^{r-1} \frac{|\langle v_r, w_h \rangle|^2}{\|w_h\|^2} + \frac{|\langle v_r, w_r \rangle|^2}{\|w_r\|^2} > \sum_{h=1}^{r-1} \frac{|\langle v_r, w_h \rangle|^2}{\|w_h\|^2}, \end{aligned}$$

in quanto $v_r \notin \text{Span}(w_1, \dots, w_{r-1}) = \text{Span}(w_1, \dots, w_{r-1})$ per cui la (12.12) implica che $\langle v_r, w_r \rangle \neq 0$. Dunque la quantità fra parentesi quadre in (12.19) è positiva, e vediamo che (iii) è verificata se e solo se $\alpha > 0$. Quindi abbiamo dimostrato l'esistenza e l'unicità a meno di fattori scalari positivi dei vettori $w_1, \dots, w_r \in V$ soddisfacenti (i)–(iii), come voluto. \square

Come conseguenza otteniamo l'esistenza delle basi ortonormali:

Corollario 12.7 Sia $\mathcal{B} = \{v_1, \dots, v_n\}$ una base di uno spazio vettoriale metrico V . Allora esiste un'unica base ortonormale $\mathcal{B}' = \{u_1, \dots, u_n\}$ di V tale che $\langle u_j, v_j \rangle > 0$ e $\text{Span}(v_1, \dots, v_j) = \text{Span}(u_1, \dots, u_j)$ per $j = 1, \dots, n$.

Dimostrazione. Il Teorema 12.6 fornisce una base ortogonale $\{w_1, \dots, w_n\}$ soddisfacente le condizioni indicate; ponendo $u_j = w_j / \|w_j\|$ per $j = 1, \dots, n$ troviamo la base ortonormale cercata. Se $\{u'_1, \dots, u'_n\}$ è un'altra base ortonormale che soddisfa le stesse condizioni, per il Teorema 12.6 si deve avere $u'_j = \alpha_j u_j$ con opportuni $\alpha_j > 0$, per $j = 1, \dots, n$. Ma allora

$$1 = \|u'_j\| = \|\alpha_j u_j\| = \alpha_j \|u_j\| = \alpha_j,$$

e quindi $u'_j = u_j$ per $j = 1, \dots, n$. \square

Otteniamo anche una versione per spazi vettoriali metrici del Teorema del completamento (Teorema 4.10):

Corollario 12.8 Siano v_1, \dots, v_r vettori a due a due ortogonali di uno spazio vettoriale metrico V di dimensione $n \geq r$. Allora esistono $w_{r+1}, \dots, w_n \in V$ tali che $\mathcal{B} = \{v_1, \dots, v_r, w_{r+1}, \dots, w_n\}$ sia una base ortogonale di V . Inoltre, se $\|v_j\| = 1$ per $j = 1, \dots, r$ allora w_{r+1}, \dots, w_n possono essere scelti in modo che \mathcal{B} sia una base ortonormale di V .

Dimostrazione. Completiamo v_1, \dots, v_r a una base $\{v_1, \dots, v_r, v'_{r+1}, \dots, v'_n\}$ di V ; il Teorema 12.6 fornisce una base ortogonale $\{w_1, \dots, w_r, w_{r+1}, \dots, w_n\}$ di V , in cui, grazie a (12.18), possiamo supporre $w_j = v_j$ per $j = 1, \dots, r$. Il resto dell'enunciato è ora ovvio. \square

Quindi non solo le basi ortonormali esistono sempre, ma possiamo anche fare tutte le operazioni di completamento che facevamo con le basi qualunque.

ESEMPIO 12.15 Consideriamo lo spazio vettoriale

$$V = \text{Span} \left(v_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, v_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} \right) \subset \mathbb{R}^4$$

equipaggiato col prodotto scalare canonico; ne vogliamo una base ortonormale. Appliciamo il procedimento di Gram-Schmidt, come descritto in (12.18), e il Corollario 12.7. Siccome $\|v_1\|^2 = 2$, il primo passo è

$$w_1 = v_1, \quad u_1 = \frac{w_1}{\|w_1\|} = \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix}.$$

Per il secondo passo calcoliamo $\|w_1\|^2 = 2$ e $\langle v_2, w_1 \rangle = 2$. Quindi

$$w_2 = v_2 - \frac{\langle v_2, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{2} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad u_2 = \frac{w_2}{\|w_2\|} = \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \\ 0 \end{pmatrix}.$$

Infine per il terzo passo calcoliamo anche $\|w_2\|^2 = 2$, $\langle v_3, w_1 \rangle = 1$ e $\langle v_3, w_2 \rangle = 2$. Quindi

$$w_3 = v_3 - \frac{\langle v_3, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 - \frac{\langle v_3, w_2 \rangle}{\langle w_2, w_2 \rangle} w_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} - \frac{2}{2} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1/2 \\ 0 \\ -1/2 \end{pmatrix},$$

$$u_3 = \frac{w_3}{\|w_3\|} = \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{pmatrix},$$

e la base ortonormale cercata è

$$\left\{ \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix}, \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{pmatrix} \right\}.$$

12.5 Proiezioni ortogonali

Vogliamo esaminare meglio dal punto di vista geometrico il procedimento di Gram-Schmidt. Se $r = 1$ non c'è nulla da dire, ma già se $r = 2$ notiamo qualcosa di interessante. Per trovare w_2 abbiamo sottratto da v_2 un vettore w in modo da ottenere qualcosa ortogonale a v_1 ; inoltre questo w era un multiplo di v_1 . È un procedimento che ricorda la proiezione ortogonale di v_2 su v_1 (vedi la Figura 12.2), e la somiglianza diventa ancora più spiccata osservando che

$$w = \frac{\langle v_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 = \|v_2\| \cos \theta \frac{v_1}{\|v_1\|},$$

dove θ è l'angolo fra v_1 e v_2 ; in altre parole abbiamo usato proprio la stessa formula della proiezione ortogonale nel piano (vedi l'Osservazione 12.3).

Per $r > 2$ succede essenzialmente la stessa cosa: per trovare w_r prendiamo v_r e vi sottraiamo la sua proiezione ortogonale sul sottospazio generato da w_1, \dots, w_{r-1} . In altre parole, a ogni passo applichiamo il seguente risultato:

Proposizione 12.9 Sia V uno spazio vettoriale metrico, e U un sottospazio di V . Allora per ogni $v_0 \in V$ esiste un'unico $u_0 \in U$ tale che $v_0 - u_0$ sia ortogonale a tutti gli elementi di U .

Dimostrazione. Fissiamo una base ortonormale $\{u_1, \dots, u_r\}$ di U (costruita per esempio applicando il Corollario 12.7 a una qualunque base di U), e supponiamo che $u_0 = \alpha_1 u_1 + \dots + \alpha_r u_r \in U$ sia come richiesto. Ma allora

$$0 = \langle v_0 - u_0, u_j \rangle = \langle v_0, u_j \rangle - \alpha_1 \langle u_1, u_j \rangle - \dots - \alpha_r \langle u_r, u_j \rangle = \langle v_0, u_j \rangle - \alpha_j$$

(in quanto $\{u_1, \dots, u_r\}$ è una base ortonormale di U), e quindi si ha $\alpha_j = \langle v_0, u_j \rangle$ per $j = 1, \dots, r$. Dunque il vettore u_0 , se esiste, è unico ed è dato da

$$u_0 = \langle v_0, u_1 \rangle u_1 + \dots + \langle v_0, u_r \rangle u_r. \quad (12.20)$$

Definiamo allora $u_0 \in U$ con la (12.20). Il conto già fatto mostra che $v_0 - u_0$ è ortogonale a u_1, \dots, u_r , per cui (Esercizio 12.14) è ortogonale a tutto U , come voluto. \square

Questo risultato suggerisce di definire un'applicazione $P_U: V \rightarrow V$ ponendo

$$\forall v \in V \quad P_U(v) = \langle v, u_1 \rangle u_1 + \dots + \langle v, u_r \rangle u_r,$$

dove $\{u_1, \dots, u_r\}$ è una base ortonormale del sottospazio U . L'applicazione P_U è chiaramente lineare, e non dipende dalla base scelta: infatti la dimostrazione della Proposizione 12.9 mostra che $P_U(v)$ è l'unico elemento di U tale che $v - P_U(v)$ sia ortogonale a tutti gli elementi di U .

Definizione 12.19 Sia U un sottospazio vettoriale di uno spazio vettoriale metrico V . L'applicazione lineare $P_U: V \rightarrow V$ appena definita si dice *proiezione ortogonale* di V su U . Nota che $P_U(u) = u$ per ogni $u \in U$ (grazie alla Proposizione 12.4), e che $\text{Im } P_U = U$ (ovvio).

Visto che abbiamo parlato di proiezione ortogonale, introduciamo anche i supplementi ortogonali (in modo da poter generalizzare al caso degli spazi vettoriali metrici concetti come quello di retta ortogonale a un piano in \mathbb{R}^3).

Definizione 12.20 Sia V uno spazio vettoriale equipaggiato con un prodotto scalare (hermitiano) $\langle \cdot, \cdot \rangle$. L'ortogonale di un sottoinsieme $S \subseteq V$ è l'insieme S^\perp di tutti gli elementi di V ortogonali a S :

$$S^\perp = \{v \in V \mid \langle v, w \rangle = 0 \quad \forall w \in S\}.$$

Nel caso in cui S sia un sottospazio vettoriale U di V (e $\langle \cdot, \cdot \rangle$ sia definito positivo), allora U^\perp è detto *supplemento ortogonale* di U .

Quest'ultima definizione richiede un paio di chiarimenti: bisogna prima di tutto far vedere che U^\perp è un sottospazio, e poi che è un supplementare di U (nel senso della Definizione 4.11). Cominciamo col

Lemma 12.10 Sia V uno spazio vettoriale metrico. Allora:

- (i) Se S è un sottoinsieme di V , allora S^\perp è un sottospazio vettoriale;
- (ii) Se U è un sottospazio di V , allora $U^\perp = \text{Ker } P_U$.

Dimostrazione. (i) Infatti se $v_1, v_2 \in S^\perp$, $w \in S$ e λ è uno scalare allora abbiamo

$$\langle v_1 + v_2, w \rangle = \langle v_1, w \rangle + \langle v_2, w \rangle = 0 + 0 = 0, \quad \langle \lambda v_1, w \rangle = \lambda \langle v_1, w \rangle = \lambda 0 = 0,$$

per cui $v_1 + v_2, \lambda v_1 \in S^\perp$.

(ii) Infatti $v \in \text{Ker } P_U$ se e solo se (perché?) v è ortogonale a tutti gli elementi di U se e solo se $v \in U^\perp$. \square

Allora possiamo dimostrare la

Proposizione 12.11 Sia U un sottospazio vettoriale di uno spazio vettoriale metrico V . Allora $V = U \oplus U^\perp$; in particolare, $\dim U^\perp = \dim V - \dim U$. Infine, se scriviamo $v = u + u^\perp$ con $v \in V$, $u \in U$ e $u^\perp \in U^\perp$, si ha

$$\|v\|^2 = \|u\|^2 + \|u^\perp\|^2. \quad (12.21)$$

Dimostrazione. Prima di tutto, $U \cap U^\perp = \{0\}$; infatti, se $u \in U \cap U^\perp$ allora il vettore u (essendo in U^\perp) è ortogonale a se stesso (essendo in U), per cui ha norma zero e quindi è il vettore nullo.

Per dimostrare che $V = U + U^\perp$ basta notare che ogni $v \in V$ può essere decomposto come

$$v = P_U(v) + (v - P_U(v))$$

con, per definizione, $P_U(v) \in U$ e $v - P_U(v) \in U^\perp$.

Infine, se scriviamo $v = u + u^\perp$ si ha

$$\|v\|^2 = \langle u + u^\perp, u + u^\perp \rangle = \|u\|^2 + \|u^\perp\|^2,$$

perché $\langle u, u^\perp \rangle = 0$. \square

Come corollario abbiamo la seguente caratterizzazione della proiezione ortogonale:

Corollario 12.12 Sia U un sottospazio di uno spazio vettoriale metrico V , e prendiamo $v \in V$. Allora $P_U(v)$ è l'elemento di U più vicino a v . In altri termini,

$$\forall u \in U \quad \|v - u\|^2 \geq \|v - P_U(v)\|^2,$$

e l'uguaglianza vale se e solo se $u = P_U(v)$.

Dimostrazione. Scriviamo $v = u_0 + u_0^\perp$, dove $u_0 = P_U(v)$ e $u_0^\perp = v - P_U(v) \in U^\perp$. Allora se $u \in U$ la (12.21) ci dice che

$$\|v - u\|^2 = \|u_0 - u\|^2 + \|u_0^\perp\|^2 = \|u_0 - u\|^2 + \|v - P_U(v)\|^2 \geq \|v - P_U(v)\|^2,$$

e l'uguaglianza vale se e solo se $\|u_0 - u\|^2 = 0$, cioè se e solo se $u = u_0 = P_U(v)$. \square

Questo risultato giustificherà la definizione di distanza fra un punto e una retta o fra un punto e un piano che vedremo nel prossimo capitolo.

12.6 Prodotti scalari e matrici

Abbiamo visto che per fare i conti con le applicazioni lineari conviene fissare una base; in questo modo si associa all'applicazione lineare una matrice e poi si lavora con quest'ultima. Nel caso dei prodotti scalari, o più in generale delle forme bilineari, possiamo fare esattamente la stessa cosa.

Affrontiamo prima di tutto il caso reale. Sia V uno spazio vettoriale su \mathbb{R} , e g una forma bilineare qualunque su V . Scelta una base $\mathcal{B} = \{v_1, \dots, v_n\}$ di V , per $v, w \in V$ indichiamo con $x = F_{\mathcal{B}}(v)$, rispettivamente $y = F_{\mathcal{B}}(w) \in \mathbb{R}^n$, le coordinate di v e w rispetto a \mathcal{B} . Allora si ha

$$\begin{aligned} g(v, w) &= g(x_1 v_1 + \dots + x_n v_n, y_1 v_1 + \dots + y_n v_n) = g\left(\sum_{k=1}^n x_k v_k, \sum_{h=1}^n y_h v_h\right) \\ &= \sum_{k=1}^n \sum_{h=1}^n x_k y_h g(v_k, v_h). \end{aligned}$$

(Quindi se introduciamo la matrice $S = (s_{hk})$ data da (attenzione all'ordine degli indici!)

$$s_{hk} = g(v_k, v_h),$$

abbiamo

$$g(v, w) = \sum_{h,k=1}^n y_h s_{hk} x_k = y^T S x. \quad (12.22)$$

Definizione 12.21 Sia $g: V \times V \rightarrow \mathbb{R}$ una forma bilineare su uno spazio vettoriale V , e $\mathcal{B} = \{v_1, \dots, v_n\}$ una base di V . La matrice $S \in M_{n,n}(\mathbb{R})$ che al posto (h, k) contiene il numero $g(v_k, v_h)$ si dice *matrice associata a* (o che *rappresenta*) la forma bilineare g rispetto alla base \mathcal{B} .

ESEMPIO 12.16 Se $S \in M_{n,n}(\mathbb{R})$ rappresenta una forma bilineare g su \mathbb{R}^n rispetto alla base canonica, allora s_{hk} è il coefficiente di $x_k y_h$ (perché? Vedi anche gli Esempi 12.2–12.5).

Dunque fissata una base a ogni forma bilineare possiamo associare una matrice in modo che (12.22) sia valida. Viceversa, se $S \in M_{n,n}(\mathbb{R})$ è una matrice quadrata (e abbiamo scelto una base \mathcal{B}) allora possiamo usare (12.22) per definire una forma bilineare su V . Potremmo formalizzare questo discorso dicendo che lo spazio delle forme bilineari su V è isomorfo a $M_{n,n}(\mathbb{R})$, dove $n = \dim V$; ma siccome non useremo questo fatto lo lasciamo per esercizio (vedi gli Esercizi 12.17–12.18).

Vediamo ora cosa succede nel caso complesso; si tratterà solo di aggiungere opportunamente qualche coniugato. Sia V uno spazio vettoriale su \mathbb{C} , e $g: V \times V \rightarrow \mathbb{C}$ una forma sesquilineare qualunque su V . Scelta una base $\mathcal{B} = \{v_1, \dots, v_n\}$ di V , per $v, w \in V$ indichiamo con $x = F_{\mathcal{B}}(v)$, rispettivamente $y = F_{\mathcal{B}}(w) \in \mathbb{C}^n$, le coordinate di v e w rispetto a \mathcal{B} . Allora si ha

$$\begin{aligned} g(v, w) &= g(x_1 v_1 + \dots + x_n v_n, y_1 v_1 + \dots + y_n v_n) = g\left(\sum_{k=1}^n x_k v_k, \sum_{h=1}^n y_h v_h\right) \\ &= \sum_{k=1}^n \sum_{h=1}^n x_k \bar{y}_h g(v_k, v_h). \end{aligned}$$

Quindi se introduciamo la matrice $S = (s_{hk})$ data da

$$s_{hk} = g(v_k, v_h)$$

(di nuovo, attenzione all'ordine degli indici!) abbiamo

$$g(v, w) = \sum_{h,k=1}^n \bar{y}_h s_{hk} x_k = y^H S x. \quad (12.23)$$

In particolare, *le forme bilineari e hermitiane sono univocamente determinate dai valori che assumono su una base.*

Definizione 12.22 Sia $g: V \times V \rightarrow \mathbb{C}$ una forma sesquilineare su uno spazio vettoriale V , e $\mathcal{B} = \{v_1, \dots, v_n\}$ una base di V . La matrice $S \in M_{n,n}(\mathbb{C})$ che al posto (h, k) contiene il numero $g(v_k, v_h)$ si dice *matrice associata a* (o che *rappresenta*) la forma sesquilineare g rispetto alla base \mathcal{B} .

A questo punto si pongono due problemi fondamentali. Come cambia la matrice associata cambiando base? Come possiamo trovare le proprietà della forma bilineare (o sesquilineare) semplicemente guardando la matrice associata?

Il primo problema ha una risposta immediata:

Proposizione 12.13 Sia $g: V \times V \rightarrow \mathbb{R}$ una forma bilineare (sesquilineare) su uno spazio vettoriale V su \mathbb{R} (o \mathbb{C}). Siano $\mathcal{B} = \{v_1, \dots, v_n\}$ e $\mathcal{B}' = \{v'_1, \dots, v'_n\}$ due basi di V , e B la matrice di cambiamento di base da \mathcal{B} a \mathcal{B}' . Indichiamo infine con S e S' le matrici associate a g rispetto a \mathcal{B} e \mathcal{B}' . Allora si ha

$$S' = B^T S B \quad (S' = B^H S B).$$

Dimostrazione. La dimostrazione è identica in entrambi i casi. Infatti, per definizione la i -esima colonna di B contiene le coordinate di v'_i rispetto alla base \mathcal{B} , per $i = 1, \dots, n$. Quindi

$$s'_{hk} = g(v'_k, v'_h) = (B^h)^T S B^k \quad \implies \quad S' = B^T S B,$$

oppure

$$s'_{hk} = g(v'_k, v'_h) = (B^h)^H S B^k \quad \implies \quad S' = B^H S B. \quad \square$$

Osservazione 12.10 Qui bisogna fare attenzione. Nonostante le apparenze, le matrici associate alle applicazioni lineari (o meglio, agli endomorfismi) e quelle associate alle forme bilineari *non* sono la stessa cosa. Certo, in entrambi i casi le matrici sembrano essere semplicemente tabelle quadrate di numeri, ma in realtà si comportano in modo profondamente diverso. Per esempio, cambiando base le matrici associate a forme bilineari cambiano secondo la legge $S \mapsto B^T S B$, mentre abbiamo visto che le matrici associate a endomorfismi cambiano secondo la legge $A \mapsto B^{-1} A B$. Il punto è che in entrambi i casi le matrici sono soltanto dei mezzi per fare calcoli con oggetti che sono intrinsecamente diversi; quello che studiamo, a cui siamo interessati, sono questi oggetti (gli endomorfismi, le forme bilineari), e non le matrici in sé. Quindi quando si fanno conti con le matrici bisogna avere ben chiaro se si tratta di endomorfismi o di forme bilineari.

Definizione 12.23 Due matrici $A, A' \in M_{n,n}(\mathbb{R})$ sono *congruenti su \mathbb{R}* se esiste una matrice invertibile $B \in GL_n(\mathbb{R})$ tale che $A' = B^T A B$. La *classe di congruenza C_A* della matrice A è l'insieme delle matrici congruenti ad A .

Definizione 12.24 Due matrici $A, A' \in M_{n,n}(\mathbb{C})$ sono *congruenti su \mathbb{C}* se esiste una matrice invertibile $B \in GL_n(\mathbb{C})$ tale che $A' = B^H A B$. La *classe di congruenza C_A* della matrice A è di nuovo l'insieme delle matrici congruenti ad A .

Dunque se due matrici rappresentano la stessa forma bilineare (sesquilineare) rispetto a basi diverse allora sono congruenti. Come nel caso degli endomorfismi (vedi la Proposizione 8.2) vale anche il viceversa (Esercizio 12.20).

Osservazione 12.11 Nota che la matrice associata a un prodotto scalare (o hermitiano) rispetto a una base ortogonale è necessariamente diagonale (perché?). Inoltre, se il prodotto scalare (o hermitiano) è definito positivo, allora la base \mathcal{B} è ortonormale se e solo se la matrice associata rispetto a questa base è la matrice identica I_n . Dunque se rispetto alla base ortonormale \mathcal{B} i vettori $v, w \in V$ hanno come coordinate rispettivamente i vettori x e y , il loro prodotto scalare (hermitiano) sarà dato da

$$\langle v, w \rangle = y^T x$$

nel caso reale, o da

$$\langle v, w \rangle = y^H x$$

nel caso complesso. In altri termini, ogni prodotto scalare (hermitiano) definito positivo si esprime in coordinate rispetto a una base *ortonormale* esattamente come il prodotto scalare (hermitiano) canonico.

Questa osservazione ha due conseguenze rilevanti.

Lemma 12.14 Sia $\langle \cdot, \cdot \rangle$ un prodotto scalare (hermitiano) definito positivo su uno spazio vettoriale V , ed S la matrice associata a $\langle \cdot, \cdot \rangle$ rispetto a una base \mathcal{B} . Allora $\det S > 0$.

Dimostrazione. Sia \mathcal{B}_0 una base ortonormale di V , e B la matrice di cambiamento di base da \mathcal{B}_0 a \mathcal{B} . Allora $\langle \cdot, \cdot \rangle$ è rappresentato da I_n rispetto a \mathcal{B}_0 , per cui $S = B^T B$ (o $S = B^H B$ nel caso complesso). Dunque

$$\det S = \det(B^T B) = (\det B)^2 > 0,$$

o rispettivamente $\det S = \det(B^H B) = |\det B|^2 > 0$, grazie all'Esercizio 11.14. \square

Proposizione 12.15 Sia V uno spazio vettoriale su \mathbb{R} (o su \mathbb{C}), e $\mathcal{B} = \{v_1, \dots, v_n\}$ una base di V . Allora esiste un unico prodotto scalare (hermitiano) definito positivo su V rispetto a cui \mathcal{B} è una base ortonormale.

Dimostrazione. Per l'Osservazione 12.11, questo prodotto scalare (hermitiano) se esiste dev'essere definito ponendo

$$\forall v, w \in V \quad \langle v, w \rangle = y^T x \quad (\text{rispettivamente, } \langle v, w \rangle = y^H x),$$

dove $x = F_{\mathcal{B}}(v)$ e $y = F_{\mathcal{B}}(w)$ sono le coordinate di v e w rispetto a \mathcal{B} . Dunque basta verificare che $\langle \cdot, \cdot \rangle$ è davvero un prodotto scalare (hermitiano) definito positivo su V . Per esempio, è omogeneo nella prima variabile: se $x = F_{\mathcal{B}}(v)$ allora $\lambda x = F_{\mathcal{B}}(\lambda v)$ e quindi

$$\langle \lambda v, w \rangle = y^T (\lambda x) = \lambda (y^T x) = \lambda \langle v, w \rangle.$$

In modo analogo (vedi l'Esercizio 12.21) si dimostra la bilinearità e la simmetria (o la sesquilinearità e l'hermitianità nel caso complesso). È anche definito positivo: infatti

$$\langle v, v \rangle = x^T x = x_1^2 + \dots + x_n^2 \geq 0,$$

(nel caso complesso si procede in modo analogo) e l'uguaglianza vale se e solo se $x = 0$, cioè se e solo se $v = 0$. Infine, \mathcal{B} è una base ortonormale, in quanto $F_{\mathcal{B}}(v_j) = e_j$ per $j = 1, \dots, n$, per cui

$$\langle v_i, v_j \rangle = e_j^T e_i = \delta_{ij}.$$

\square

Osservazione 12.12 Detta in modo diverso, la dimostrazione della Proposizione 12.15 suona così: V è isomorfo tramite $F_{\mathcal{B}}$ allo spazio \mathbb{R}^n (o \mathbb{C}^n), su cui è definito il prodotto scalare (hermitiano) canonico. Allora per calcolare il prodotto scalare (o hermitiano) di due elementi di V ci basta trasferirli in \mathbb{R}^n (o \mathbb{C}^n) con $F_{\mathcal{B}}$ e fare i conti lì con i prodotti canonici. Siccome \mathcal{B} finisce nella base canonica, che è ortonormale, abbiamo quanto richiesto.

Torniamo al secondo problema, ancora aperto: come leggere dalle matrici associate le proprietà delle forme bilineari (o sesquilineari)? Ci serve una definizione.

Definizione 12.25 Una matrice quadrata $A \in M_{n,n}(\mathbb{R})$ a coefficienti reali si dice *simmetrica* se $A^T = A$; *antisimmetrica* se $A^T = -A$. Una matrice $B \in M_{n,n}(\mathbb{C})$ a coefficienti complessi si dice *hermitiana* se $B^H = B$; *anti-hermitiana* se $B^H = -B$.

Ciò detto, abbiamo un primo risultato:

Lemma 12.16 Sia V uno spazio vettoriale su \mathbb{R} (o su \mathbb{C}) di dimensione n , e g una forma bilineare (sesquilineare) su V . Fissiamo una base $\mathcal{B} = \{v_1, \dots, v_n\}$ di V , e sia S la matrice rappresentativa di g rispetto a \mathcal{B} . Allora:

- (i) La forma g è simmetrica (hermitiana) se e solo se S è simmetrica (hermitiana).
- (ii) Supponiamo g simmetrica (hermitiana). Allora $\text{Ker } S = F_{\mathcal{B}}(V^{\perp})$. In particolare, g è non degenere se e solo se S è non singolare, cioè se e solo se $\det S \neq 0$.

Dimostrazione. (i) Se g è simmetrica, allora

$$s_{hk} = g(v_k, v_h) = g(v_h, v_k) = s_{kh},$$

o S è simmetrica; analogamente, se g è hermitiana si ha

$$s_{hk} = g(v_k, v_h) = \overline{g(v_h, v_k)} = \overline{s_{kh}},$$

o S è hermitiana. Viceversa, prendiamo $v, w \in V$ con coordinate rispettivamente x e y rispetto a \mathcal{B} . Se S è simmetrica si ha

$$g(v, w) = y^T S x = y^T S^T x = (x^T S y)^T = x^T S y = g(w, v),$$

(dove $(x^T S y)^T = x^T S y$ per il semplice motivo che $x^T S y$ è un numero reale), e quindi g è simmetrica. Analogamente, se S è hermitiana si ha

$$g(v, w) = y^H S x = y^H S^H x = (x^H S y)^H = \overline{x^H S y} = \overline{g(w, v)},$$

(dove $(x^H S y)^H = \overline{x^H S y}$ semplicemente perché $x^H S y$ è un numero complesso), e dunque g è hermitiana.

(ii) Un vettore $v_0 \in V$ appartiene a V^\perp se e solo se $g(v_0, w) = 0$ per ogni $w \in V$. Ponendo $x_0 = F_B(v_0) \in \mathbb{R}^n$, questo accade se e solo se $y^T S x_0 = 0$ per ogni $y \in \mathbb{R}^n$, che a sua volta può succedere se e solo se $S x_0 = O$ (perché?), cioè se e solo se $x_0 \in \text{Ker } S$. Dunque $v_0 \in V^\perp$ se e solo se $F_B(v_0) \in \text{Ker } S$, e quindi $\text{Ker } S = F_B(V^\perp)$. In particolare, $V^\perp = \{O\}$ se e solo se $\text{Ker } S = \{O\}$, per cui g è non degenera se e solo se S è non singolare. La stessa dimostrazione funziona anche nel caso complesso (esercizio). \square

Dunque simmetria (o hermitianità) si leggono facilmente dalle matrici rappresentative. Ben più interessanti invece sono i criteri che dicono come stabilire se un certo prodotto scalare (hermitiano) è definito positivo, semplicemente guardandone una matrice associata. Dobbiamo però aspettare il Capitolo 16 per parlarne, in quanto dipendono da strumenti teorici che non abbiamo ancora introdotto.

12.7 L'aggiunta

Vogliamo ora esplorare che relazioni ci sono fra applicazioni lineari e prodotti scalari. Se $T: V \rightarrow W$ è un'applicazione lineare fra spazi vettoriali metrici, presi $v \in V$ e $w \in W$ possiamo considerare il prodotto $\langle T(v), w \rangle_W$, dove $\langle \cdot, \cdot \rangle_W$ è il prodotto scalare in W . Domanda: c'è qualche relazione fra questo prodotto e uno della forma $\langle v, \tilde{v} \rangle_V$ per qualche $\tilde{v} \in V$? La risposta è positiva, come indicato nel seguente Teorema:

Teorema 12.17 Sia $T: V \rightarrow W$ un'applicazione lineare fra due spazi vettoriali metrici V e W . Allora esiste un'unica applicazione lineare $T^*: W \rightarrow V$ tale che

$$\forall v \in V \forall w \in W \quad \langle T(v), w \rangle_W = \langle v, T^*(w) \rangle_V, \quad (12.24)$$

dove $\langle \cdot, \cdot \rangle_V$ (rispettivamente, $\langle \cdot, \cdot \rangle_W$) è il prodotto scalare — o hermitiano — su V (rispettivamente, su W).

Dimostrazione. Fissiamo una base ortonormale $\mathcal{B} = \{v_1, \dots, v_n\}$ di V , e cominciamo col far vedere che esiste al più un'unica applicazione lineare $T^*: W \rightarrow V$ che soddisfa (12.24). Infatti, se T^* esiste, la (12.13) ci dice che si deve avere

$$\begin{aligned} T^*(w) &= \langle T^*(w), v_1 \rangle_V v_1 + \dots + \langle T^*(w), v_n \rangle_V v_n \\ &= \langle w, T(v_1) \rangle_W v_1 + \dots + \langle w, T(v_n) \rangle_W v_n, \end{aligned}$$

dove abbiamo usato (12.24). Dunque T^* se esiste deve essere data da

$$T^*(w) = \langle w, T(v_1) \rangle_W v_1 + \dots + \langle w, T(v_n) \rangle_W v_n = \sum_{j=1}^n \langle w, T(v_j) \rangle_W v_j, \quad (12.25)$$

per cui è univocamente determinata. Per l'esistenza, definiamo $T^*: W \rightarrow V$ tramite la (12.25). Chiaramente T^* è lineare; si tratta solo di dimostrare che soddisfa (12.24).

Ma infatti ogni $v \in V$ si scrive come $v = \sum_{j=1}^n \langle v, v_j \rangle_V v_j$, e quindi

$$\begin{aligned} \langle v, T^*(w) \rangle_V &= \sum_{j=1}^n \langle v, \langle w, T(v_j) \rangle_W v_j \rangle_V = \sum_{j=1}^n \langle T(v_j), w \rangle_W \langle v, v_j \rangle_V \\ &= \sum_{j=1}^n \langle \langle v, v_j \rangle_V T(v_j), w \rangle_W = \left\langle T \left(\sum_{j=1}^n \langle v, v_j \rangle_V v_j \right), w \right\rangle_W = \langle T(v), w \rangle_W, \end{aligned}$$

come volevamo. \square

Definizione 12.26 Sia $T: V \rightarrow W$ un'applicazione lineare fra spazi vettoriali metrici. L'applicazione $T^*: W \rightarrow V$ definita nel Teorema 12.17 si dice *aggiunta* di T . La formula (12.24) si chiama *formula di aggiunta*.

In pratica, l'aggiunta è per le applicazioni lineari quello che è la trasposta (o la trasposta coniugata, nel caso complesso) per le matrici:

ESEMPIO 12.17 Sia $A \in M_{m,n}(\mathbb{R})$, e consideriamo \mathbb{R}^n ed \mathbb{R}^m con il prodotto scalare canonico. Allora

$$(L_A)^* = L_{A^T}.$$

Infatti, $(L_A)^*$ è l'unica applicazione lineare $T: \mathbb{R}^m \rightarrow \mathbb{R}^n$ tale che

$$\langle L_A(x), y \rangle = \langle x, T(y) \rangle$$

per ogni $x \in \mathbb{R}^n$ e $y \in \mathbb{R}^m$; ma si ha

$$\langle L_A(x), y \rangle = y^T A x = (A^T y)^T x = \langle x, L_{A^T}(y) \rangle,$$

e quindi $(L_A)^* = L_{A^T}$.

ESEMPIO 12.18 Sia $B \in M_{m,n}(\mathbb{C})$, e consideriamo \mathbb{C}^n e \mathbb{C}^m con il prodotto hermitiano canonico. Allora

$$(L_B)^* = L_{B^H}.$$

Infatti, $(L_B)^*$ è l'unica applicazione lineare $T: \mathbb{C}^m \rightarrow \mathbb{C}^n$ tale che

$$\langle L_B(x), y \rangle = \langle x, T(y) \rangle$$

per ogni $x \in \mathbb{C}^n$ e $y \in \mathbb{C}^m$; ma si ha

$$\langle L_B(x), y \rangle = y^H B x = (B^H y)^H x = \langle x, L_{B^H}(y) \rangle,$$

e quindi $(L_B)^* = L_{B^H}$.

ESEMPIO 12.19 Cambiando prodotto scalare, l'aggiunta ovviamente cambia. Consideriamo per esempio su \mathbb{R}^2 il prodotto scalare definito positivo dato da

$$\langle x, y \rangle = 2x_1y_1 + x_1y_2 + x_2y_1 + x_2y_2, \quad (12.26)$$

e sia $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ l'applicazione lineare

$$T \begin{vmatrix} x_1 \\ x_2 \end{vmatrix} = \begin{vmatrix} x_1 + x_2 \\ -x_1 \end{vmatrix} = A \begin{vmatrix} x_1 \\ x_2 \end{vmatrix},$$

dove $A \in M_{2,2}(\mathbb{R})$ è la matrice

$$A = \begin{vmatrix} 1 & 1 \\ -1 & 0 \end{vmatrix}.$$

Vogliamo calcolare $T^*: \mathbb{R}^2 \rightarrow \mathbb{R}^2$. Se $y \in \mathbb{R}^2$, il vettore $\tilde{y} = T^*(y)$ è l'unico vettore di \mathbb{R}^2 tale che $\langle T(x), y \rangle = \langle x, \tilde{y} \rangle$ per ogni $x \in \mathbb{R}^2$; siccome si ha

$$\begin{aligned} \langle T(x), y \rangle &= 2(x_1 + x_2)y_1 + (x_1 + x_2)y_2 - x_1y_1 - x_1y_2 = x_1y_1 + x_2(2y_1 + y_2) \\ \langle x, \tilde{y} \rangle &= 2x_1\tilde{y}_1 + x_1\tilde{y}_2 + x_2\tilde{y}_1 + x_2\tilde{y}_2 = x_1(2\tilde{y}_1 + \tilde{y}_2) + x_2(\tilde{y}_1 + \tilde{y}_2), \end{aligned}$$

le coordinate $(\tilde{y}_1, \tilde{y}_2)$ di \tilde{y} devono soddisfare il sistema

$$\begin{cases} 2\tilde{y}_1 + \tilde{y}_2 = y_1, \\ \tilde{y}_1 + \tilde{y}_2 = 2y_1 + y_2. \end{cases}$$

Risolviendo troviamo

$$\begin{cases} \tilde{y}_1 = -y_1 - y_2, \\ \tilde{y}_2 = 3y_1 + 2y_2, \end{cases}$$

per cui l'aggiunta T^* è data da

$$T^* \begin{vmatrix} y_1 \\ y_2 \end{vmatrix} = \begin{vmatrix} -y_1 - y_2 \\ 3y_1 + 2y_2 \end{vmatrix} = \begin{vmatrix} -1 & -1 \\ 3 & 2 \end{vmatrix} \begin{vmatrix} y_1 \\ y_2 \end{vmatrix},$$

che è ben diversa da L_{A^T} .

Come conseguenza della definizione abbiamo subito le seguenti proprietà dell'aggiunta, analoghe alle proprietà della trasposta:

Corollario 12.18 Siano U, V e W tre spazi vettoriali metrici, e consideriamo tre applicazioni lineari $T, T_1: V \rightarrow W$ e $T_2: U \rightarrow V$. Allora:

$$(T^*)^* = T \quad \text{e} \quad (T_1 \circ T_2)^* = T_2^* \circ T_1^*.$$

Dimostrazione. $(T^*)^*: V \rightarrow W$ è l'unica applicazione lineare $L: V \rightarrow W$ che soddisfa

$$\langle T^*(w), v \rangle_V = \langle w, L(v) \rangle_W$$

per ogni $v \in V$ e $w \in W$; siccome anche T soddisfa questa uguaglianza, si ha necessariamente $(T^*)^* = T$.

Poi abbiamo

$$\langle (T_1 \circ T_2)(u), w \rangle_W = \langle T_2(u), T_1^*(w) \rangle_V = \langle u, (T_2^* \circ T_1^*)(w) \rangle_U,$$

per ogni $u \in U$ e $w \in W$, e quindi $(T_1 \circ T_2)^* = T_2^* \circ T_1^*$. \square

Ovviamente, una volta fissate delle basi di V e W vorremmo conoscere che relazione c'è fra la matrice che rappresenta T e quella che rappresenta T^* . La risposta è contenuta nella seguente

Proposizione 12.19 Siano V, W spazi vettoriali metrici, \mathcal{B}_V una base di V , \mathcal{B}_W una base di W , e $T: V \rightarrow W$ un'applicazione lineare. Siano A e A^* le matrici che rappresentano rispettivamente T e T^* rispetto a \mathcal{B}_V e \mathcal{B}_W , e S_V, S_W le matrici che rappresentano il prodotto scalare (hermitiano) rispettivamente di V e W rispetto a queste basi. Allora

$$A^* = S_V^{-1} A^T S_W \quad (A^* = S_V^{-1} A^H S_W \text{ se siamo su } \mathbb{C}).$$

In particolare, se \mathcal{B}_V e \mathcal{B}_W sono basi ortonormali, allora

$$A^* = A^T \quad (A^* = A^H \text{ se siamo su } \mathbb{C}).$$

Dimostrazione. Prendiamo $v \in V$ e $w \in W$, e indichiamo con $x = F_{\mathcal{B}_V}(v)$ le coordinate di v , e con $y = F_{\mathcal{B}_W}(w)$ le coordinate di w . Allora

$$y^T S_W A x = \langle T(v), w \rangle_W = \langle v, T^*(w) \rangle_V = (A^* y)^T S_V x = y^T (A^*)^T S_V x,$$

per ogni x e y . Quindi $S_W A = (A^*)^T S_V$, per cui trasponendo si ha $S_V A^* = A^T S_W$, e nel caso reale è fatta. Nel caso complesso, basta sostituire dovunque la trasposta coniugata alla trasposta. \square

12.8 Endomorfismi autoaggiunti e isometrie

Usando l'aggiunta possiamo identificare alcune classi particolarmente importanti di endomorfismi.

Definizione 12.27 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale metrico V . Diremo che T è *autoaggiunto* se $T^* = T$, cioè se

$$\forall v_1, v_2 \in V \quad \langle T(v_1), v_2 \rangle = \langle v_1, T(v_2) \rangle.$$

Se lo spazio V è sui reali, diremo anche *simmetrico* invece di autoaggiunto; se è sui complessi, diremo anche *hermitiano* invece di autoaggiunto.

Osservazione 12.13 Grazie alla Proposizione 12.19, vediamo che se \mathcal{B} è una base ortonormale di V allora l'endomorfismo $T: V \rightarrow V$ è simmetrico (hermitiano) se e solo se la matrice associata rispetto a \mathcal{B} è simmetrica (hermitiana). Ma se la base \mathcal{B} non è ortonormale, questo non è più vero, come mostrato nel seguente esempio.

ESEMPIO 12.20 Consideriamo \mathbb{R}^2 col prodotto scalare canonico, e sia $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dato da

$$T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ x_1 - x_2 \end{pmatrix}.$$

È facile vedere che T è simmetrico (per esempio perché rispetto alla base canonica è rappresentato da una matrice simmetrica). Scegliamo invece come base di \mathbb{R}^2 la base

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}.$$

Rispetto a questa base il prodotto scalare canonico è rappresentato dalla matrice

$$S = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix} \quad \Longrightarrow \quad S^{-1} = \begin{pmatrix} 5 & -3 \\ -3 & 2 \end{pmatrix},$$

mentre T è rappresentato dalla matrice

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 7 \\ -2 & -4 \end{pmatrix},$$

che si guarda bene dall'essere simmetrica. In questa base il fatto che T è un endomorfismo simmetrico è indicato dall'eguaglianza $S^{-1}A^T S = A$, eguaglianza che avrai cura di verificare.

Mentre l'importanza degli endomorfismi autoaggiunti diventerà evidente solo nei Capitoli 15 e 16, c'è un'altra classe di endomorfismi il cui interesse è subito chiaro.

Definizione 12.28 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale metrico V . Diremo che T è un'isometria (lineare) se

$$\forall v_1, v_2 \in V \quad \langle T(v_1), T(v_2) \rangle = \langle v_1, v_2 \rangle.$$

Se lo spazio V è sui reali, diremo anche *endomorfismo ortogonale* invece di isometria: se è sui complessi, diremo anche *endomorfismo unitario* invece di isometria.

Dunque un'isometria conserva il prodotto scalare (hermitiano); in particolare, conserva distanze e angoli. Prima di discutere degli esempi, vediamo alcune proprietà:

Proposizione 12.20 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale metrico V . Allora le seguenti affermazioni sono equivalenti:

- (i) T è un'isometria;
- (ii) T è invertibile e $T^{-1} = T^*$;
- (iii) se $\{v_1, \dots, v_n\}$ è una base ortonormale di V , anche $\{T(v_1), \dots, T(v_n)\}$ lo è;
- (iv) $\|T(v)\| = \|v\|$ per ogni $v \in V$.

Dimostrazione. (i) \implies (ii). Prima di tutto, T è invertibile: infatti se $T(v) = 0$ abbiamo

$$0 = \langle T(v), T(v) \rangle = \langle v, v \rangle = \|v\|^2,$$

e quindi $v = 0$. Poi, per ogni $v, w \in V$ si ha

$$\langle v, T^*(T(w)) \rangle = \langle T(v), T(w) \rangle = \langle v, w \rangle,$$

per cui, grazie alla Proposizione 12.1.(ii), $T^* \circ T = \text{id}_V$, cioè $T^* = T^{-1}$.

(ii) \implies (iii). Infatti

$$\langle T(v_i), T(v_j) \rangle = \langle v_i, (T^* \circ T)(v_j) \rangle = \langle v_i, v_j \rangle = \delta_{ij} = \begin{cases} 0 & \text{se } i \neq j, \\ 1 & \text{se } i = j. \end{cases}$$

(iii) \implies (iv). Utilizzando (12.13) e (12.17) si ha

$$\begin{aligned} \|T(v)\|^2 &= \left\langle \sum_{i=1}^n \langle v, v_i \rangle T(v_i), \sum_{j=1}^n \langle v, v_j \rangle T(v_j) \right\rangle \\ &= \sum_{i=1}^n \sum_{j=1}^n \langle v, v_i \rangle \langle v_j, v \rangle \langle T(v_i), T(v_j) \rangle \\ &= \sum_{i=1}^n \sum_{j=1}^n \langle v, v_i \rangle \langle v_j, v \rangle \delta_{ij} = \sum_{h=1}^n |\langle v, v_h \rangle|^2 = \|v\|^2. \end{aligned}$$

(iv) \implies (i). Utilizzando la Proposizione 12.2.(vi) nel caso reale troviamo

$$\begin{aligned} \langle T(v), T(w) \rangle &= \frac{1}{4} [\|T(v) + T(w)\|^2 - \|T(v) - T(w)\|^2] \\ &= \frac{1}{4} [\|T(v+w)\|^2 - \|T(v-w)\|^2] = \frac{1}{4} [\|v+w\|^2 - \|v-w\|^2] \\ &= \langle v, w \rangle. \end{aligned}$$

Il caso complesso è analogo. □

Corollario 12.21 Sia $T:V \rightarrow V$ un endomorfismo di uno spazio vettoriale metrico V . Fissiamo una base \mathcal{B} di V ; sia A la matrice che rappresenta T rispetto a \mathcal{B} , ed S la matrice che rappresenta il prodotto scalare (hermitiano) rispetto a questa base. Allora T è ortogonale (unitario) se e solo se A è invertibile e

$$A^{-1} = S^{-1}A^T S \quad (A^{-1} = S^{-1}A^H S).$$

Dimostrazione. Segue dalle Proposizioni 12.20.(ii) e 12.19. □

Nel caso delle matrici, le isometrie si riconoscono facilmente.

Definizione 12.29 Una matrice $A \in M_{n,n}(\mathbb{R})$ si dice *ortogonale* se $A^T A = I_n$. L'insieme delle matrici ortogonali di ordine n si indica con $O(n)$.

Definizione 12.30 Una matrice $A \in M_{n,n}(\mathbb{C})$ si dice *unitaria* se $A^H A = I_n$. L'insieme delle matrici unitarie di ordine n si indica con $U(n)$.

Osservazione 12.14 Nota che il Corollario 7.8 ci dice che ogni matrice ortogonale (unitaria) è automaticamente invertibile, e $A^{-1} = A^T$ (rispettivamente, $A^{-1} = A^H$).

Osservazione 12.15 Un endomorfismo $T:V \rightarrow V$ è ortogonale (unitario) se e solo se è rappresentato da una matrice ortogonale (unitaria) rispetto a una base *ortonormale*; confronta con l'Osservazione 12.13.

Corollario 12.22 Sia A una matrice quadrata di ordine n a coefficienti reali (complessi). Allora le seguenti affermazioni sono equivalenti:

- (i) A è ortogonale (unitaria);
- (ii) L_A è un'isometria rispetto al prodotto scalare (hermitiano) canonico;
- (iii) le colonne di A formano una base ortonormale di \mathbb{R}^n (\mathbb{C}^n) rispetto al prodotto scalare (hermitiano) canonico.

Dimostrazione. (i) \implies (ii). Infatti se A è ortogonale si ha $A^{-1} = A^T$, e quindi

$$(L_A)^{-1} = L_{A^{-1}} = L_{A^T} = (L_A)^*,$$

grazie all'Esempio 12.17. Il caso complesso è analogo.

(ii) \implies (iii). Le colonne di A sono l'immagine tramite L_A della base canonica, e quindi sono una base ortonormale per la Proposizione 12.20.

(iii) \implies (i). Infatti l'elemento di posto (i, j) del prodotto $A^T A$ è dato da

$$(A^T A)_{ij} = (A^T)_i A^j = (A^i)^T A^j = \langle A^j, A^i \rangle = \delta_{ij},$$

e quindi $A^T A = I_n$. La dimostrazione nel caso complesso è identica. □

ESEMPIO 12.21 Vediamo di trovare tutte le applicazioni ortogonali di \mathbb{R}^2 in sé rispetto al prodotto scalare canonico — ovvero tutte le matrici ortogonali $A \in O(2)$. La relazione $A^T A = I_2$ in coordinate diventa

$$\begin{vmatrix} a_{11}^2 + a_{21}^2 & a_{11}a_{12} + a_{21}a_{22} \\ a_{11}a_{12} + a_{21}a_{22} & a_{12}^2 + a_{22}^2 \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}.$$

Quindi devono esistere angoli $\theta, \phi \in \mathbb{R}$ tali che

$$a_{11} = \cos \theta, \quad a_{21} = \sin \theta, \quad a_{12} = \sin \phi, \quad a_{22} = \cos \phi,$$

e inoltre

$$0 = \cos \theta \sin \phi + \sin \theta \cos \phi = \sin(\theta + \phi).$$

Dunque abbiamo due possibilità: o $\phi = -\theta$ oppure $\phi = \pi - \theta$. Corrispondentemente, la matrice A è necessariamente di uno dei due tipi seguenti:

$$R_\theta = \begin{vmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{vmatrix} \quad \text{oppure} \quad S_\theta = \begin{vmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{vmatrix},$$

ed è facile verificare che $R_\theta, S_\theta \in O(2)$ per ogni $\theta \in \mathbb{R}$.

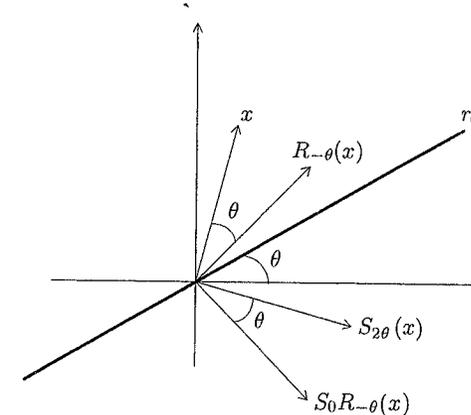


Figura 12.4 Matrici ortogonali e simmetrie.

Possiamo interpretare geometricamente le isometrie R_θ e S_θ (vedi la Figura 12.4). Prendiamo $x \in \mathbb{R}^2$ e scriviamolo in coordinate polari: si ha $x = r(\cos \psi, \sin \psi)$, per qualche $r \geq 0$ e $\psi \in \mathbb{R}$. Allora

$$R_\theta(x) = \begin{vmatrix} r(\cos \theta \cos \psi - \sin \theta \sin \psi) \\ r(\sin \theta \cos \psi + \cos \theta \sin \psi) \end{vmatrix} = \begin{vmatrix} r \cos(\theta + \psi) \\ r \sin(\theta + \psi) \end{vmatrix},$$

cioè R_θ è la rotazione di angolo θ in verso antiorario. Riguardo S_θ , notiamo prima di tutto che

$$S_{2\theta} = R_\theta \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} R_{-\theta};$$

dunque $S_{2\theta}$ si ottiene effettuando prima una rotazione di angolo $-\theta$, poi una simmetria S_0 rispetto all'asse x e infine una rotazione di angolo θ . Il risultato netto è una simmetria rispetto alla retta r_θ che forma un angolo θ con l'asse x .

In conclusione, gli endomorfismi ortogonali di \mathbb{R}^2 rispetto al prodotto scalare canonico sono tutte e sole le simmetrie rispetto a rette passanti per l'origine e le rotazioni.

Esercizi

12.1 Supponiamo che $\mathcal{R}_0 = RA(O, \vec{i}, \vec{j})$ sia un sistema di riferimento affine in \mathcal{A}^2 tale che i vettori \vec{i} e \vec{j} siano perpendicolari e di lunghezza unitaria. Considera il sistema di riferimento affine $\mathcal{R} = RA(O, \vec{i}, \vec{j})$, dove $\vec{j} = 2\vec{j}$, e definisci la "distanza" fra due punti di \mathcal{A}^2 con la formula (12.2), ma usando le coordinate rispetto a \mathcal{R} invece di quelle rispetto a \mathcal{R}_0 . L'esercizio consiste nel disegnare l'insieme dei punti di \mathcal{A}^2 che distano 1 dall'origine, rispetto a questa particolare "distanza".

12.2 Trova prima di tutto quali delle seguenti applicazioni $\langle \cdot, \cdot \rangle_j: \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}$ sono dei prodotti scalari. Scrivi ciascuno di essi nella forma $\langle v, w \rangle = v^T S w$ e stabilisci quali sono (semi)definiti:

$$\begin{aligned} \langle v, w \rangle_1 &= v_1^2 + w_2^2 + v_3 w_3; \\ \langle v, w \rangle_2 &= v_1 w_1 - v_1 w_2 - v_2 w_1 + 3v_2 w_2 + \pi v_3 w_3; \\ \langle v, w \rangle_3 &= v_1 w_2 + v_2 w_1 + v_2 w_3 + v_3 w_2 + v_3 w_1 + v_1 w_3; \\ \langle v, w \rangle_4 &= 3v_1 w_1 - v_1 w_2 + 3v_2 w_2 + v_3 w_3; \\ \langle v, w \rangle_5 &= -v_1 w_1 + v_1 w_2 + v_2 w_1 - v_2 w_2 - v_3 w_3; \\ \langle v, w \rangle_6 &= v_1 w_1 - v_2 w_2; \\ \langle v, w \rangle_7 &= 400v_1 w_1 + 3\sqrt{\pi}v_1 w_3 + 3\sqrt{\pi}v_3 w_1 + 227v_3 w_3. \end{aligned}$$

12.3 Dimostra che

$$\langle p, q \rangle = p(0)q(0) + p(1)q(1) + \dots + p(n)q(n)$$

è un prodotto scalare definito positivo su $\mathbb{R}_n[t]$.

12.4 Dimostra che

$$\langle v, w \rangle = 2v_1 \bar{w}_1 + 2iv_1 \bar{w}_2 - 2iv_2 \bar{w}_1 + 4v_2 \bar{w}_2$$

è un prodotto hermitiano definito positivo su \mathbb{C}^2 . Che si può dire invece su

$$\langle v, w \rangle = 2v_1 \bar{w}_1 + 2iv_1 \bar{w}_2 - 2iv_2 \bar{w}_1 + 2v_2 \bar{w}_2?$$

12.5 Dimostra che

$$\langle p, q \rangle = p(0)\overline{q(0)} + p(1)\overline{q(1)} + p(i)\overline{q(i)}$$

è un prodotto hermitiano definito positivo sullo spazio $\mathbb{C}_2[t]$ dei polinomi a coefficienti complessi di grado minore o uguale a 2.

12.6 Dimostra che l'applicazione $\langle \cdot, \cdot \rangle: M_{2,2}(\mathbb{R}) \times M_{2,2}(\mathbb{R}) \rightarrow \mathbb{R}$ data da

$$\langle A, B \rangle = \text{tr}(B^T P A),$$

dove P è la matrice

$$P = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix},$$

è un prodotto scalare definito positivo.

12.7 Sia $P \in M_{n,n}(\mathbb{R})$ una matrice simmetrica. Siano $\langle \cdot, \cdot \rangle: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ e $\langle\langle \cdot, \cdot \rangle\rangle: M_{n,n}(\mathbb{R}) \times M_{n,n}(\mathbb{R}) \rightarrow \mathbb{R}$ le applicazioni definite da

$$\langle x, y \rangle = y^T P x \quad \text{e} \quad \langle\langle A, B \rangle\rangle = \text{tr}(B^T P A)$$

per ogni $x, y \in \mathbb{R}^n$ e $A, B \in M_{n,n}(\mathbb{R})$. Dimostra che sono entrambi prodotti scalari, e che $\langle \cdot, \cdot \rangle$ è definito positivo se e solo se $\langle\langle \cdot, \cdot \rangle\rangle$ lo è.

12.8 Sia V uno spazio vettoriale su \mathbb{R} , e $\langle \cdot, \cdot \rangle$ un prodotto scalare su V (che può anche essere degenere). Dimostra che

$$\langle v, w \rangle = \frac{1}{2} [\langle v+w, v+w \rangle - \langle v, v \rangle - \langle w, w \rangle]$$

per ogni $v, w \in V$.

12.9 Sia V uno spazio vettoriale metrico su \mathbb{R} . Dimostra che due vettori v e w sono linearmente dipendenti se e solo se vale una delle uguaglianze nella disuguaglianza triangolare Proposizione 12.2.(v). Cosa succede nel caso di uno spazio vettoriale metrico su \mathbb{C} ?

12.10 Sia V uno spazio vettoriale metrico con norma $\| \cdot \|$. Dimostra che

$$\forall v, w \in V \quad \|v+w\|^2 + \|v-w\|^2 = 2(\|v\|^2 + \|w\|^2)$$

(uguaglianza del parallelogramma). Interpreta geometricamente questa formula considerando il parallelogramma di vertici O, v, w e $v+w$.

12.11 Trova delle basi ortonormali di \mathbb{R}^3 rispetto a quei prodotti scalari dell'Esercizio 12.2 che erano definiti positivi.

12.12 Dimostra che se $\{v_1, \dots, v_n\}$ è una base ortogonale di uno spazio vettoriale metrico V , allora $\{v_1/\|v_1\|, \dots, v_n/\|v_n\|\}$ è una base ortonormale di V .

12.13 Trova una base ortonormale, rispetto al prodotto scalare canonico, del sottospazio di \mathbb{R}^4 dato da

$$V = \text{Span} \left(\begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 5 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -3 \\ -3 \\ 1 \\ -3 \end{pmatrix} \right),$$

e completala a una base ortonormale di tutto \mathbb{R}^4 .

12.14 Sia U un sottospazio vettoriale di uno spazio vettoriale V , $\{u_1, \dots, u_r\}$ una base di U , e sia $\langle \cdot, \cdot \rangle$ un prodotto scalare (o hermitiano) su V . Dimostra che un vettore $v \in V$ è ortogonale a U se e solo se $v \perp u_j$ per $j = 1, \dots, r$.

12.15 Dimostra che una proiezione ortogonale è una proiezione nel senso dell'Esercizio 7.7.

12.16 Siano U, U_1, U_2 sottospazi di uno spazio vettoriale metrico V . Dimostra che $U_1 \subseteq U_2$ implica $U_1^\perp \supseteq U_2^\perp$; che $(U^\perp)^\perp = U$; che $(U_1 + U_2)^\perp = U_1^\perp \cap U_2^\perp$; e che $(U_1 \cap U_2)^\perp = U_1^\perp + U_2^\perp$.

12.17 Dimostra che tutte le forme bilineari su \mathbb{R}^n sono del tipo $g(x, y) = y^T S x$ per qualche matrice $S \in M_{n,n}(\mathbb{R})$.

12.18 Indichiamo con $\text{Bil}(V)$ l'insieme delle forme bilineari su uno spazio vettoriale reale V . Definisci una somma e un prodotto per scalari su $\text{Bil}(V)$ in modo da renderlo uno spazio vettoriale. Fissata una base di V , dimostra che l'applicazione che associa a ogni elemento di $\text{Bil}(V)$ la matrice che lo rappresenta rispetto a questa base è un isomorfismo di $\text{Bil}(V)$ con $M_{n,n}(\mathbb{R})$, dove $n = \dim V$, e deducine che $\dim \text{Bil}(V) = n^2$.

12.19 Trova la matrice associata al prodotto scalare dell'Esempio 12.7 rispetto alla solita base $\mathcal{B} = \{1, t, t^2\}$.

12.20 Sia $g: V \times V \rightarrow \mathbb{R}$ (o $g: V \times V \rightarrow \mathbb{C}$) una forma bilineare (sesquilineare) su uno spazio vettoriale V . Sia \mathcal{B} una base di V , e S la matrice associata a g rispetto a \mathcal{B} . Sia S' una matrice congruente a S ; dimostra che esiste una base \mathcal{B}' di V tale che S' rappresenta g rispetto a \mathcal{B}' .

12.21 Completa i dettagli della dimostrazione della Proposizione 12.15.

12.22 Siano $T, T_1, T_2: V \rightarrow W$ applicazioni lineari, e sia λ uno scalare. Dimostra che $(T_1 + T_2)^* = T_1^* + T_2^*$ e che $(\lambda T)^* = \bar{\lambda} T^*$.

12.23 Sia $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ l'applicazione lineare data da

$$T \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x + y \\ y - z \\ x + y - z \end{pmatrix}.$$

Trova l'aggiunta di T rispetto al prodotto scalare definito positivo dell'Esempio 12.2.

12.24 Sia $T: \mathbb{R}_2[t] \rightarrow \mathbb{R}_2[t]$ l'applicazione lineare data da $[T(p)](t) = p(t-1)$. Trova l'aggiunta di T rispetto al prodotto scalare definito positivo dell'Esempio 12.7.

12.25 Sia $T: V \rightarrow W$ un'applicazione lineare fra spazi vettoriali metrici. Dimostra che $\text{rg } T = \text{rg } T^*$.

12.26 Dimostra che ogni proiezione ortogonale è un endomorfismo autoaggiunto.

12.27 Sia U un sottospazio di uno spazio vettoriale metrico V , \mathcal{B}_1 una base di U e \mathcal{B}_2 una base di U^\perp . Trova la matrice che rappresenta la proiezione ortogonale $P_U: V \rightarrow V$ rispetto alla base $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ di V . Se \mathcal{B}_0 è una base ortonormale di V , dimostra che la matrice P associata a P_U rispetto a \mathcal{B}_0 è $P = B(B^T B)^{-1} B^T$ (su \mathbb{R} ; $P = B(B^H B)^{-1} B^H$ su \mathbb{C}), dove B è la matrice che contiene per colonne le coordinate dei vettori di \mathcal{B}_1 rispetto a \mathcal{B}_0 .

12.28 Sia $T: V \rightarrow V$ un endomorfismo autoaggiunto di uno spazio vettoriale metrico V . Dimostra che la forma $\langle \cdot, \cdot \rangle_T$ definita da $\langle v, w \rangle_T = \langle T(v), w \rangle$ è un prodotto scalare (o hermitiano su \mathbb{C}), non degenerare se e solo se T è invertibile.

12.29 Se $T_1, T_2: V \rightarrow V$ sono isometrie di uno spazio vettoriale metrico V , dimostra che anche $T_1 \circ T_2$ e $T_1^* = T_1^{-1}$ lo sono.

12.30 Dimostra che l'insieme $O(n)$ delle matrici ortogonali di ordine n è un gruppo non commutativo rispetto al prodotto di matrici.

12.31 Dimostra che l'insieme $U(n)$ delle matrici unitarie di ordine n è un gruppo non commutativo rispetto al prodotto di matrici.

12.32 Dimostra che se A è una matrice ortogonale (unitaria), allora $|\det A| = 1$.

12.33 Siano $\mathcal{B}_1 = \{v_1, \dots, v_n\}$ e $\mathcal{B}_2 = \{w_1, \dots, w_n\}$ due basi ortonormali di \mathbb{R}^n . Dimostra che esiste un'unica matrice ortogonale $B \in O(n)$ tale che $Bv_j = w_j$ per $j = 1, \dots, n$. (Suggerimento: utilizza le Proposizioni 5.2, 12.20 e il Corollario 12.22).

COMPLEMENTI

12C.1 Il Teorema di rappresentazione di Riesz

In questo paragrafo dimostreremo un risultato che è l'evoluzione finale degli Esercizi 5.8 e 7.12: il *Teorema di rappresentazione di Riesz*.

Teorema 12C.1 (Riesz) Sia V uno spazio vettoriale metrico, e $\varphi: V \rightarrow \mathbb{R}$ (oppure $\varphi: V \rightarrow \mathbb{C}$) lineare. Allora esiste un unico elemento $v_\varphi \in V$ tale che

$$\forall v \in V \quad \varphi(v) = \langle v, v_\varphi \rangle.$$

Inoltre su \mathbb{R} l'applicazione $\Phi: V' \rightarrow V$ data da $\Phi(\varphi) = v_\varphi$ è un isomorfismo.

Dimostrazione. L'unicità segue subito dalla Proposizione 12.1.(ii). Per l'esistenza, poniamo su \mathbb{R} (o \mathbb{C}) il prodotto scalare (hermitiano) canonico dato da $\langle \lambda, \mu \rangle = \lambda \bar{\mu}$

per $\lambda, \mu \in \mathbb{R}$ (ovvero da $\langle \lambda, \mu \rangle = \lambda \bar{\mu}$ per $\lambda, \mu \in \mathbb{C}$). Allora possiamo considerare l'aggiunta φ^* di φ , e il vettore $v_\varphi = \varphi^*(1)$. Allora

$$\langle v, v_\varphi \rangle = \langle v, \varphi^*(1) \rangle = \langle \varphi(v), 1 \rangle = \varphi(v) \cdot 1 = \varphi(v),$$

e anche l'esistenza è fatta.

Usando l'Esercizio 12.22 è facile verificare che sui reali Φ è lineare⁵ (esercizio); vediamo che è iniettiva. Infatti $\Phi(\varphi) = 0$ vuol dire che $v_\varphi = O$ e quindi

$$\forall v \in V \quad \varphi(v) = \langle v, v_\varphi \rangle = 0,$$

per cui $\varphi = O$. Siccome $\dim V = \dim V'$, l'applicazione Φ è un isomorfismo. \square

Quindi Φ è un isomorfismo canonico fra uno spazio vettoriale metrico V e il suo duale. È facile verificare (Esercizi 12C.1–12C.2) che Φ trasforma annullatori in ortogonali, e applicazioni trasposte in applicazioni aggiunte; quindi parte di questo capitolo non è altro che un caso particolare di quanto studiato nei Complementi al Capitolo 8.

Esercizi

12C.1 Sia S un sottoinsieme di uno spazio vettoriale metrico V su \mathbb{R} , e $\Phi: V' \rightarrow V$ l'isomorfismo del Teorema 12C.1. Dimostra che $\Phi(S^\circ) = S^\perp$, dove $S^\circ \subseteq V'$ è l'annullatore di S definito nei Complementi al Capitolo 8.

12C.2 Sia $T: V \rightarrow W$ un'applicazione lineare fra spazi vettoriali metrici su \mathbb{R} , e $\Phi_V: V' \rightarrow V$, $\Phi_W: W' \rightarrow W$ gli isomorfismi del Teorema 12C.1. Dimostra che

$$T^* = \Phi_V \circ T' \circ \Phi_W^{-1},$$

dove $T': W' \rightarrow V'$ è l'applicazione trasposta definita nei Complementi al Capitolo 8.

⁵ Sui complessi, Φ risulta essere *antilineare*, cioè $\Phi(\lambda\varphi) = \bar{\lambda}\Phi(\varphi)$.

Geometria euclidea

In questo capitolo parleremo della geometria metrica del piano e dello spazio, utilizzando gli strumenti introdotti nei capitoli precedenti. Definiremo l'angolo fra rette e/o piani, la distanza fra punti, rette e/o piani, e vedremo come calcolarli. Parleremo anche del prodotto vettore, un'operazione definita esclusivamente su \mathbb{R}^3 che permette di associare a una coppia di vettori un terzo vettore ortogonale a entrambi. Nei Complementi a questo capitolo discuteremo le isometrie del piano e dello spazio, studieremo la geometria delle coniche e presenteremo qualche esempio di superficie nello spazio.

13.1 Angoli e ortogonalità

Cominciamo definendo l'ambiente in cui tratteremo la geometria metrica.

Definizione 13.1 Il piano euclideo \mathcal{E}^2 (rispettivamente, lo spazio euclideo \mathcal{E}^3) è il piano affine \mathcal{A}^2 (lo spazio affine \mathcal{A}^3) in cui sia stata fissata un'unità di misura per le lunghezze.

Ricordando che l'angolo di 1 radiante è quello che in una circonferenza di raggio 1 sottende un arco lungo 1, in \mathcal{E}^2 ed \mathcal{E}^3 è determinata univocamente anche un'unità di misura per gli angoli.

Definizione 13.2 Un sistema di riferimento cartesiano $RC(O, A_1, A_2, A_3)$ di \mathcal{E}^3 è un sistema di riferimento affine $RA(O, A_1, A_2, A_3)$ tale che i vettori $\vec{i} = \overrightarrow{OA_1}$, $\vec{j} = \overrightarrow{OA_2}$ e $\vec{k} = \overrightarrow{OA_3}$ siano di lunghezza unitaria e ortogonali a due a due.

In altri termini, $\{\vec{i}, \vec{j}, \vec{k}\}$ è una base ortonormale di \mathcal{V}_O^3 , e quindi l'usuale identificazione fra \mathcal{A}^3 (e quindi \mathcal{E}^3) ed \mathbb{R}^3 data dalle coordinate rispetto a questa base ci permette di misurare lunghezze e angoli in \mathcal{E}^3 usando il prodotto scalare canonico di \mathbb{R}^3 . Inutile dire che analoghe definizioni e osservazioni si applicano a \mathcal{E}^2 ed \mathbb{R}^2 .

Prima di proseguire vediamo velocemente cosa succede quando cambiamo sistema di riferimento cartesiano. Siano $\mathcal{R} = RC(O, A_1, A_2, A_3)$ e $\mathcal{R}' = RC(O', A'_1, A'_2, A'_3)$

due sistemi di riferimento cartesiani, e $P \in \mathcal{E}^3$. Indichiamo con $x, x' \in \mathbb{R}^3$ le coordinate di P rispetto a \mathcal{R} , rispettivamente a \mathcal{R}' . Trattandosi di particolari sistemi di riferimento affini, quanto visto nel Paragrafo 10.7 ci dice che x e x' sono legati da

$$x = Bx' + c,$$

dove $c \in \mathbb{R}^3$ mentre $B \in GL_3(\mathbb{R})$ è la matrice di cambiamento di base. In particolare, B contiene per colonne le coordinate della base di \mathcal{V}_O^3 , relativa a \mathcal{R}' , che per ipotesi è una base ortonormale; quindi (Corollario 12.22) la matrice B è ortogonale. Viceversa, se B è ortogonale allora la base di \mathcal{V}_O^3 , relativa a \mathcal{R}' è ortonormale, per cui \mathcal{R}' è un sistema di riferimento cartesiano. In altre parole abbiamo dimostrato la

Proposizione 13.1 Sia \mathcal{R} un sistema di riferimento cartesiano su \mathcal{E}^3 (o \mathcal{E}^2), e \mathcal{R}' un altro sistema di riferimento affine. Allora anche \mathcal{R}' è un sistema di riferimento cartesiano se e solo se la formula di cambiamento di coordinate da \mathcal{R} a \mathcal{R}' è data da $x = Bx' + c$, dove B è una matrice ortogonale.

Definizione 13.3 Un cambiamento di coordinate di \mathcal{E}^2 o \mathcal{E}^3 che trasforma un sistema di riferimento cartesiano in un altro sistema di riferimento cartesiano verrà detto *cambiamento di coordinate metrico*.

Fissiamo allora una volta per tutte un riferimento cartesiano $RC(O, A_1, A_2, A_3)$, in modo da identificare \mathcal{E}^3 con \mathbb{R}^3 (ed \mathcal{E}^2 con \mathbb{R}^2); poniamo anche $\vec{i} = \overrightarrow{OA_1}$, $\vec{j} = \overrightarrow{OA_2}$ e $\vec{k} = \overrightarrow{OA_3}$.

Il nostro primo obiettivo è definire l'angolo fra rette e/o piani. Cominciamo con le rette. Ci troviamo subito di fronte a un piccolo ostacolo: due rette che si intersecano determinano *due* angoli convessi, e non uno (vedi la Figura 13.1); quale prendiamo?

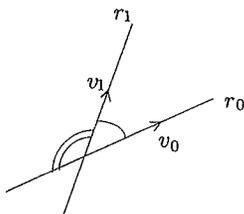


Figura 13.1 L'angolo fra due rette.

La risposta è che la domanda è mal posta: entrambi vanno bene, per poter decidere occorrono altre informazioni. Per esempio, scegliendo un verso di percorrenza delle due rette (equivalentemente, fissando un vettore direttore per le due rette), allora quale angolo bisogna prendere risulta chiarissimo: l'unico determinato dalle due semirette positive. Questo suggerisce una definizione.

Definizione 13.4 Una *retta orientata* è una retta in cui si è fissato un vettore direttore (e quindi una direzione di percorrenza). Due diversi vettori direttori determinano la stessa *orientazione* (la stessa direzione di percorrenza) se e solo se sono un multiplo positivo l'uno dell'altro (vedi anche il Paragrafo 10.8). In particolare, esiste un unico vettore di lunghezza unitaria che determina l'orientazione scelta: il *versore direttore* della retta orientata. Se $v = \alpha\vec{i} + \beta\vec{j} + \gamma\vec{k}$ è l'espressione del versore direttore v come combinazione lineare della base $\{\vec{i}, \vec{j}, \vec{k}\}$ di \mathcal{V}_O^3 , allora $\alpha, \beta, \gamma \in \mathbb{R}$ sono detti *coseni direttori* della retta.

I coseni direttori sono effettivamente dei coseni: infatti la Proposizione 12.4 implica che $\alpha = \langle v, \vec{i} \rangle = \cos \widehat{v\vec{i}}$, in quanto $\|v\| = 1 = \|\vec{i}\|$, e analogamente per β e γ .

Definizione 13.5 L'*angolo* $\widehat{r_0 r_1}$ fra due rette orientate r_0 ed r_1 è per definizione l'angolo fra i loro versori direttori.

In particolare, quindi, l'angolo fra due rette dipende soltanto dai loro sottospazi di giacitura¹.

Dunque se le due rette r_0 ed r_1 hanno equazioni parametriche $P = P_0 + tv_0$ e $P = P_1 + sv_1$ rispettivamente, dove $v_0 = (l_0, m_0, n_0)$ e $v_1 = (l_1, m_1, n_1)$ sono vettori direttori che determinano le orientazioni scelte, allora l'angolo fra le rette r_0 ed r_1 è dato da

$$\cos \widehat{r_0 r_1} = \frac{\langle v_0, v_1 \rangle}{\|v_0\| \|v_1\|} = \frac{l_0 l_1 + m_0 m_1 + n_0 n_1}{\sqrt{l_0^2 + m_0^2 + n_0^2} \sqrt{l_1^2 + m_1^2 + n_1^2}}. \quad (13.1)$$

In particolare, le due rette sono ortogonali se e solo se $\cos \widehat{r_0 r_1} = 0$, cioè se e solo se

$$l_0 l_1 + m_0 m_1 + n_0 n_1 = 0 \quad (13.2)$$

(condizione di ortogonalità).

Osservazione 13.1 In \mathcal{E}^2 l'angolo fra due rette orientate si definisce e si calcola nello stesso modo; semplicemente nelle formule abbiamo una coordinata in meno.

Consideriamo ora una retta orientata r_0 passante per l'origine, di vettore direttore $v_0 = (l_0, m_0, n_0)$. Un altro vettore $v = (x, y, z)$ è ortogonale a questa retta se e solo se $l_0 x + m_0 y + n_0 z = 0$; quindi l'ortogonale r_0^\perp di r_0 è un piano (come si poteva anche dedurre dalla Proposizione 12.11). In particolare, l'equazione del piano ortogonale a r_0 e passante per il punto $P_0 = (x_0, y_0, z_0)$ è

$$l_0(x - x_0) + m_0(y - y_0) + n_0(z - z_0) = 0. \quad (13.3)$$

Nota che questa è l'equazione del piano ortogonale a r_0 (perché?) anche se r_0 non passa per l'origine.

¹ E quindi due rette possono essere ortogonali anche se non si intersecano.

Definizione 13.6 La proiezione ortogonale di un punto P_0 su una retta r è il punto d'intersezione di r con l'unico piano passante per P_0 ortogonale a r .

Per calcolare l'angolo fra due piani, ci serve la seguente:

Definizione 13.7 Un piano orientato è un piano π in cui è stata fissata una base nel sottospazio di giacitura. Due basi definiscono la stessa orientazione se la matrice di cambiamento di base ha determinante positivo (vedi il Paragrafo 10.8).

Sia π un piano orientato, con base $\{v_1, v_2\}$ del sottospazio di giacitura. Se v_3 e $-v_3$ sono i due versori direttori della retta ortogonale π^\perp , allora solo una delle due basi $\{v_1, v_2, v_3\}$ e $\{v_1, v_2, -v_3\}$ ha la stessa orientazione della base $\{\vec{i}, \vec{j}, \vec{k}\}$ (vedi l'Esercizio 10.29). Dunque:

Definizione 13.8 Sia π un piano orientato, e $\{v_1, v_2\}$ una base del sottospazio di giacitura che determina l'orientazione scelta. Il versore ortogonale di π è l'unico versore direttore v_3 della retta ortogonale al piano tale che $\{v_1, v_2, v_3\}$ sia una base di \mathcal{V}_O^3 con la stessa orientazione della base $\{\vec{i}, \vec{j}, \vec{k}\}$.

Osservazione 13.2 Prendiamo un piano π di equazione cartesiana $ax + by + cz = d$. Per quanto detto prima dell'equazione (13.3), il vettore $v = (a, b, c)$ è un vettore direttore della retta π^\perp (passante per l'origine) ortogonale al piano. Dunque il versore ortogonale di π è $v/\|v\|$ oppure $-v/\|v\|$ a seconda dell'orientazione scelta per π .

Definizione 13.9 L'angolo $\widehat{\pi_0\pi_1}$ fra due piani orientati π_0 e π_1 è per definizione l'angolo fra i loro versori ortogonali.

Osservazione 13.3 In geometria elementare l'"angolo diedro" formato da due piani non paralleli è l'angolo formato dalle due rette ottenute intersecando i due piani con un terzo piano ortogonale a entrambi. La nostra definizione sostituisce a queste due rette altre due, ruotate di $\pi/2$; l'angolo quindi non è cambiato (vedi la Figura 13.2). Infine, di nuovo l'angolo fra due piani dipende solo dai sottospazi di giacitura.

Per calcolare l'angolo fra due piani π_0 e π_1 , supponiamo che abbiano equazione cartesiana $a_0x + b_0y + c_0z = d_0$ e $a_1x + b_1y + c_1z = d_1$ rispettivamente. Allora i vettori $v_0 = (a_0, b_0, c_0)$ e $v_1 = (a_1, b_1, c_1)$ sono vettori direttori delle rette ortogonali; possiamo supporre che i piani siano orientati in modo che $v_0/\|v_0\|$ e $v_1/\|v_1\|$ siano i loro versori ortogonali. Allora

$$\cos \widehat{\pi_0\pi_1} = \cos \widehat{v_0v_1} = \frac{a_0a_1 + b_0b_1 + c_0c_1}{\sqrt{a_0^2 + b_0^2 + c_0^2} \sqrt{a_1^2 + b_1^2 + c_1^2}} \quad (13.4)$$

In particolare, i due piani sono ortogonali se e solo se

$$a_0a_1 + b_0b_1 + c_0c_1 = 0 \quad (13.5)$$

(condizione di ortogonalità).

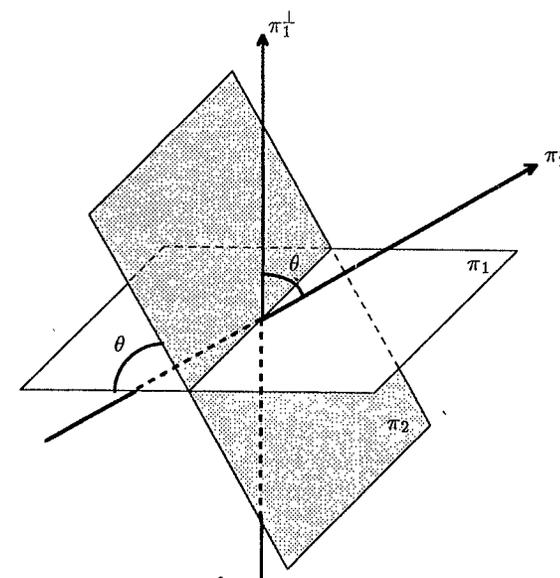


Figura 13.2 L'angolo fra due piani.

Di nuovo, dato un piano π di equazione cartesiana $ax + by + cz = d$ e un punto P_0 , esiste un'unica retta ortogonale a π passante per P_0 , che ha equazione parametrica $P = P_0 + tv$, dove $v = (a, b, c)$.

Definizione 13.10 La proiezione ortogonale di un punto P_0 su un piano π è il punto d'intersezione di π con l'unica retta passante per P_0 ortogonale al piano.

Rimane da definire l'angolo fra una retta e un piano. Prendiamo una retta r orientata con vettore direttore $v = (l, m, n)$, e un piano π di equazione $ax + by + cz = d$. Allora possiamo considerare la retta r' proiezione ortogonale di r su π ; la retta r' ha come vettore direttore la proiezione ortogonale di v sul sottospazio di giacitura di π .

Definizione 13.11 L'angolo $\widehat{r\pi}$ fra una retta r e un piano π è l'angolo fra r e la sua proiezione ortogonale su π . Cambiando l'orientazione di r cambia anche quella di r' , per cui l'angolo non dipende dall'orientazione di r scelta.

Per definizione, l'angolo fra una retta e un piano è sempre compreso fra 0 e $\pi/2$; quindi per trovarlo esplicitamente ci basta calcolarne il seno. Se r e π sono paralleli, allora il loro angolo è zero; altrimenti, sia P_0 il loro punto d'intersezione, ed s la retta ortogonale a π passante per P_0 .

Se indichiamo con $\theta \in [0, \pi/2]$ l'angolo fra r e π , a meno di moltiplicare per -1 l'equazione del piano possiamo supporre che la retta orientata s di vettore diret-

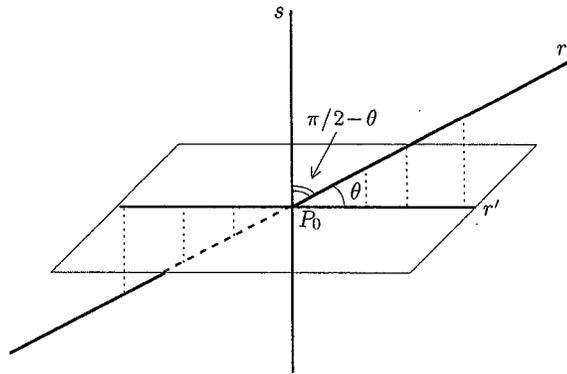


Figura 13.3 L'angolo fra una retta e un piano.

tore (a, b, c) abbia un angolo di $\pi/2 - \theta$ radianti con r (confronta l'Esercizio 13.3). Allora (vedi la Figura 13.3)

$$\sin \widehat{r\pi} = \sin \widehat{rr'} = \cos \widehat{rs} = \frac{al + bm + cn}{\sqrt{a^2 + b^2 + c^2} \sqrt{l^2 + m^2 + n^2}}. \quad (13.6)$$

In particolare, retta e piano sono ortogonali se e solo se $\sin \widehat{r\pi} = 1$ se e solo se (per la disuguaglianza di Cauchy-Schwarz) i vettori (a, b, c) e (l, m, n) sono linearmente dipendenti (*condizione di ortogonalità*).

Vediamo ora un esempio di un tipico problema di geometria metrica.

ESEMPIO 13.1 Vogliamo trovare la retta r passante per $P_0 = (1, 0, -1)$ ortogonale e incidente alla retta s di equazioni cartesiane

$$\begin{cases} x - 2z = 1, \\ y + 2z = 0. \end{cases}$$

Prima di tutto, le equazioni parametriche di s sono (esercizio)

$$P = P_1 + tv_1 = \begin{vmatrix} 1 \\ 0 \\ 0 \end{vmatrix} + t \begin{vmatrix} 2 \\ -2 \\ 1 \end{vmatrix};$$

vogliamo trovare le equazioni parametriche di r , che devono essere del tipo

$$P = P_0 + sv_0 = \begin{vmatrix} 1 \\ 0 \\ -1 \end{vmatrix} + s \begin{vmatrix} l \\ m \\ n \end{vmatrix}.$$

Ora, la condizione di ortogonalità ci dice che

$$2l - 2m + n = 0. \quad (13.7)$$

In particolare, $\text{rg} \{v_0, v_1\} = 2$. La condizione d'incidenza (vedi il Paragrafo 10.4) ci dice invece che

$$0 = \det \{v_0, v_1, P_0 - P_1\} = \det \begin{vmatrix} l & 2 & 0 \\ m & -2 & 0 \\ m & 1 & -1 \end{vmatrix} = 2l + 2m. \quad (13.8)$$

Risolvendo il sistema composto da (13.7) e (13.8) otteniamo quindi la soluzione: la retta r ha equazioni parametriche

$$P = P_0 + sv_0 = \begin{vmatrix} 1 \\ 0 \\ -1 \end{vmatrix} + s \begin{vmatrix} -1 \\ 1 \\ 4 \end{vmatrix}.$$

Ma vi è anche un altro modo per risolvere questo problema. Prima di tutto, la retta r deve appartenere al piano π_1 passante per P_0 e ortogonale alla retta s , che ha equazione cartesiana

$$2(x - 1) - 2y + (z + 1) = 0. \quad (13.9)$$

Poi la retta r deve anche appartenere al piano π_2 passante per P_0 e contenente s (infatti questo piano contiene sia P_0 che il punto d'intersezione fra r ed s). Per quanto visto nel Paragrafo 10.5.(I) il piano π_2 ha equazione

$$-2(x - 2z - 1) - 2(y + 2z) = 0. \quad (13.10)$$

Mettendo insieme (13.9) e (13.10) abbiamo trovato che r ha equazioni cartesiane

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

13.2 Distanze

Passiamo ora a calcolare le distanze fra punti, rette e piani. La distanza fra due punti $P_0 = (x_0, y_0, z_0)$ e $P_1 = (x_1, y_1, z_1)$ è ovviamente data da

$$d(P_0, P_1) = \|\overrightarrow{OP_1} - \overrightarrow{OP_0}\| = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2}.$$

Più interessante è invece il calcolo della distanza fra un punto e una retta.

Definizione 13.12 Sia r una retta e P_0 un punto. La *distanza* $d(P_0, r)$ fra il punto e la retta è per definizione la distanza fra P_0 e la sua proiezione ortogonale su r .

Osservazione 13.4 La scelta di questa definizione per la distanza è dovuta al Corollario 12.12.

Supponiamo che la retta r abbia equazione parametrica $P = P_1 + tv_1$, e indichiamo con θ l'angolo fra v_0 e il vettore $v_1 = \overrightarrow{OP_1} - \overrightarrow{OP_0}$ (vedi la Figura 13.4, dove abbiamo supposto r orientata in modo che $\theta \in [0, \pi/2]$).

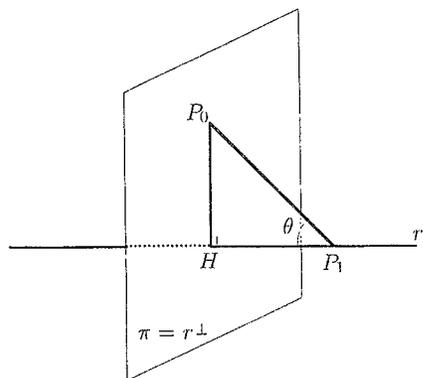


Figura 13.4 La distanza fra un punto e una retta.

Sia H la proiezione ortogonale di P_0 su r . Allora considerando il triangolo rettangolo di vertici P_0, P_1 e H vediamo che

$$\begin{aligned} d(P_0, r) &= d(P_0, H) = d(P_0, P_1) |\sin \theta| = d(P_0, P_1) \sqrt{1 - (\cos \theta)^2} \\ &= \|v_1\| \sqrt{1 - \frac{|\langle v_1, v_0 \rangle|^2}{\|v_0\|^2 \|v_1\|^2}} = \sqrt{\|v_1\|^2 - \frac{|\langle v_1, v_0 \rangle|^2}{\|v_0\|^2}} \\ &= \sqrt{d(P_0, P_1)^2 - \frac{|\langle \overrightarrow{OP_1} - \overrightarrow{OP_0}, v_0 \rangle|^2}{\|v_0\|^2}}. \end{aligned} \tag{13.11}$$

Osservazione 13.5 Ovviamente, questa stessa formula letta in \mathcal{E}^2 dà la distanza fra una retta del piano espressa in equazioni parametriche e un punto. Un'altra formula utilizzabile in \mathcal{E}^3 è invece descritta nel prossimo paragrafo.

ESEMPIO 13.2 Vogliamo la distanza fra il punto $P_0 = (1, 0, 1)$ e la retta r di equazione parametrica

$$P = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + t \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}.$$

Si ha

$$\overrightarrow{OP_1} - \overrightarrow{OP_0} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix},$$

per cui

$$d(P_0, r) = \sqrt{1 - \frac{1}{2}} = \frac{1}{\sqrt{2}}.$$

Passiamo ora alla distanza fra un punto e un piano.

Definizione 13.13 Sia π un piano e P_0 un punto. La distanza $d(P_0, \pi)$ fra il punto e il piano è per definizione la distanza fra P_0 e la sua proiezione ortogonale su π .

Sia $P_0 = (x_0, y_0, z_0)$ un punto, e π un piano di equazione cartesiana $ax + by + cz = d$. Sia H la proiezione ortogonale di P_0 su π ; indichiamo con r la retta ortogonale a π passante per P_0 (e H ; vedi la Figura 13.5).

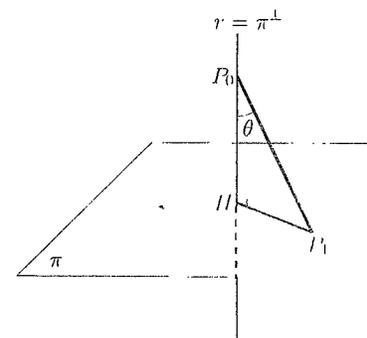


Figura 13.5 La distanza fra un punto e un piano.

Sia $P_1 = (x_1, y_1, z_1) \in \pi$ un punto del piano, e indichiamo con θ l'angolo fra il vettore $v_1 = \overrightarrow{OP_1} - \overrightarrow{OP_0}$ e il vettore $v_0 = (a, b, c)$, che è un vettore direttore della retta r : a meno di moltiplicare per -1 l'equazione del piano possiamo supporre che $\theta \in [0, \pi/2]$. Allora considerando il triangolo rettangolo di vertici P_0, P_1 e H vediamo che

$$\begin{aligned} d(P_0, \pi) &= d(P_0, H) = d(P_0, P_1) |\cos \theta| = \frac{|\langle v_0, v_1 \rangle|}{\|v_0\|} \\ &= \frac{|a(x_0 - x_1) + b(y_0 - y_1) + c(z_0 - z_1)|}{\sqrt{a^2 + b^2 + c^2}}. \end{aligned}$$

Ma ora il fatto che $P_1 \in \pi$ ci dice che $ax_1 + by_1 + cz_1 = d$: quindi

$$d(P_0, \pi) = \frac{|ax_0 + by_0 + cz_0 - d|}{\sqrt{a^2 + b^2 + c^2}}. \tag{13.12}$$

Osservazione 13.6 Una formula assolutamente analoga (in cui non compare c) fornisce la distanza in \mathcal{E}^2 fra una retta del piano in equazioni cartesiane e un punto.

La distanza fra una retta e un piano, o fra due piani, è ottenuta molto facilmente: se la retta e il piano (o i due piani) sono incidenti, la distanza è zero; se invece sono paralleli, è uguale alla distanza di un qualunque punto della retta (o del primo piano) dal (secondo) piano, per cui si calcola come abbiamo appena visto. Dunque l'unica distanza rimasta da determinare è quella fra due rette.

Siano r_0 ed r_1 due rette, di equazione parametrica rispettivamente $P = P_0 + tv_0$ e $P = P_1 + sv_1$. Se le due rette non sono parallele, allora esiste un unico piano che contiene r_1 ed è parallelo a r_0 : è il piano π di equazione parametrica $P = P_1 + sv_1 + tv_0$.

Definizione 13.14 Siano r_0 ed r_1 due rette. Se sono parallele, definiamo loro distanza $d(r_0, r_1)$ il numero $d(P_0, r_1)$, dove P_0 è un qualunque punto di r_0 ; se non sono parallele, la loro distanza $d(r_0, r_1)$ è invece data da $d(P_0, \pi)$, dove P_0 è ancora un punto qualunque di r_0 e π è l'unico piano contenente r_1 e parallelo a r_0 .

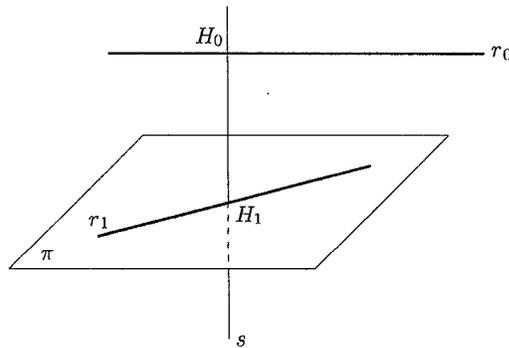


Figura 13.6 La distanza fra due rette.

Vediamo allora di trovare la formula della distanza di r_0 ed r_1 quando non sono parallele (vedi la Figura 13.6). Scriviamo $v_0 = (l_0, m_0, n_0)$ e $v_1 = (l_1, m_1, n_1)$. Per l'Esercizio 10.7, il piano π ha equazione cartesiana

$$ax + by + cz = ax_1 + by_1 + cz_1,$$

dove

$$a = \det \begin{vmatrix} m_0 & m_1 \\ n_0 & n_1 \end{vmatrix} = m_0 n_1 - m_1 n_0, \quad b = -\det \begin{vmatrix} l_0 & l_1 \\ n_0 & n_1 \end{vmatrix} = n_0 l_1 - n_1 l_0,$$

$$c = \det \begin{vmatrix} l_0 & l_1 \\ m_0 & m_1 \end{vmatrix} = l_0 m_1 - l_1 m_0.$$

In particolare si verifica facilmente (esercizio) che

$$ax_0 + by_0 + cz_0 - (ax_1 + by_1 + cz_1) = \det |v_0 \ v_1 \ P_0 - P_1|,$$

e che

$$\begin{aligned} a^2 + b^2 + c^2 &= (m_0 n_1 - m_1 n_0)^2 + (n_0 l_1 - n_1 l_0)^2 + (l_0 m_1 - l_1 m_0)^2 \\ &= (l_0^2 + m_0^2 + n_0^2)(l_1^2 + m_1^2 + n_1^2) - (l_0 l_1 + m_0 m_1 + n_0 n_1)^2 \\ &= \|v_0\|^2 \|v_1\|^2 - (\langle v_0, v_1 \rangle)^2 = \|v_0\|^2 \|v_1\|^2 (1 - (\cos \widehat{v_0 v_1})^2) \\ &= \|v_0\|^2 \|v_1\|^2 (\sin \widehat{v_0 v_1})^2. \end{aligned}$$

Quindi applicando la (13.12) troviamo che la distanza fra r_0 ed r_1 è data da

$$d(r_0, r_1) = d(P_0, \pi) = \frac{|\det |v_0 \ v_1 \ P_1 - P_0||}{\|v_0\| \|v_1\| \sin \widehat{v_0 v_1}}. \quad (13.13)$$

Nota che le due rette (non parallele) sono incidenti se e solo se $\det |v_0 \ v_1 \ P_1 - P_0| = 0$, cioè se e solo se $d(r_0, r_1) = 0$, come dev'essere. Nel prossimo paragrafo vedremo un'ulteriore interpretazione di questa formula.

ESEMPIO 13.3 Vogliamo calcolare la distanza fra le rette r_0 ed r_1 di equazione parametrica rispettivamente

$$r_0: P = \begin{vmatrix} 1 \\ 0 \\ 1 \end{vmatrix} + s \begin{vmatrix} 1 \\ 1 \\ 1 \end{vmatrix} \quad \text{e} \quad r_1: P = \begin{vmatrix} -1 \\ 1 \\ 0 \end{vmatrix} + t \begin{vmatrix} -2 \\ 1 \\ 2 \end{vmatrix}.$$

Utilizzando le notazioni precedenti troviamo $a = 1$, $b = -4$ e $c = 3$; dunque

$$\|v_0\| \|v_1\| |\sin \widehat{v_0 v_1}| = (a^2 + b^2 + c^2)^{1/2} = \sqrt{26}, \quad \det |v_0 \ v_1 \ P_1 - P_0| = -5,$$

e quindi $d(r_0, r_1) = 5/\sqrt{26}$.

13.3 Il prodotto vettore

Questo paragrafo è dedicato alla descrizione di una particolare operazione definita unicamente su \mathbb{R}^3 . Due vettori $v, w \in \mathbb{R}^3$ linearmente indipendenti determinano un unico piano π che li contiene, di equazioni parametriche $P = sv + tw$. L'equazione cartesiana di π è $ax + by + cz = 0$, dove (vedi l'Esercizio 10.7)

$$a = \det \begin{vmatrix} v_2 & w_2 \\ v_3 & w_3 \end{vmatrix}, \quad b = -\det \begin{vmatrix} v_1 & w_1 \\ v_3 & w_3 \end{vmatrix}, \quad c = \det \begin{vmatrix} v_1 & w_1 \\ v_2 & w_2 \end{vmatrix}.$$

Inoltre, come abbiamo visto nel Paragrafo 13.1, il vettore (a, b, c) è ortogonale al piano π ; in particolare, è ortogonale sia a v che a w . Dunque a ogni coppia di vettori (linearmente indipendenti) di \mathbb{R}^3 possiamo associare un terzo vettore ortogonale a entrambi.

Definizione 13.15 Siano $v, w \in \mathbb{R}^3$. Il prodotto vettore di v e w è il vettore

$$v \wedge w = \det \begin{vmatrix} v_2 & w_2 \\ v_3 & w_3 \end{vmatrix} e_1 - \det \begin{vmatrix} v_1 & w_1 \\ v_3 & w_3 \end{vmatrix} e_2 + \det \begin{vmatrix} v_1 & w_1 \\ v_2 & w_2 \end{vmatrix} e_3 \in \mathbb{R}^3. \quad (13.14)$$

A volte si scrive $v \times w$ invece di $v \wedge w$.

Osservazione 13.7 La formula (13.14) viene spesso scritta nel seguente modo:

$$v \wedge w = \det \begin{pmatrix} e_1 & v_1 & w_1 \\ e_2 & v_2 & w_2 \\ e_3 & v_3 & w_3 \end{pmatrix},$$

dove si intende che questo “determinante” viene sviluppato solo lungo la prima colonna.

Vediamo che proprietà possiede il prodotto vettore.

Proposizione 13.2 Sia $v \wedge w$ il prodotto vettore di due vettori $v, w \in \mathbb{R}^3$. Allora:

- (i) $v \wedge w$ è ortogonale sia a v che a w ;
- (ii) $\|v \wedge w\|$ è l'area del parallelogramma di vertici O, v, w e $v + w$, cioè

$$\|v \wedge w\| = \|v\| \|w\| |\sin \theta|,$$

dove θ è l'angolo fra v e w ;

- (iii) $v \wedge w = O$ se e solo se v e w sono linearmente dipendenti;
- (iv) se v e w sono linearmente indipendenti, allora la base $\{v, w, v \wedge w\}$ determina la stessa orientazione della base canonica;
- (v) $v \wedge w$ è l'unico vettore di \mathbb{R}^3 che soddisfa (i), (ii) e (iv);
- (vi) $w \wedge v = -v \wedge w$;
- (vii) l'applicazione $\wedge: \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^3$ è bilineare.

Dimostrazione. (i) Infatti

$$\langle v, v \wedge w \rangle = \det \begin{vmatrix} v_1 & v_1 & w_1 \\ v_2 & v_2 & w_2 \\ v_3 & v_3 & w_3 \end{vmatrix} = 0 = \det \begin{vmatrix} w_1 & v_1 & w_1 \\ w_2 & v_2 & w_2 \\ w_3 & v_3 & w_3 \end{vmatrix} = \langle w, v \wedge w \rangle.$$

(ii) Prima di tutto, la Figura 13.7 mostra che l'area del parallelogramma di vertici O, v, w e $v + w$ è effettivamente data da $\|v\| \|w\| |\sin \theta|$. Poi,

$$\begin{aligned} \|v \wedge w\|^2 &= (v_2 w_3 - w_2 v_3)^2 + (v_1 w_3 - w_1 v_3)^2 + (v_1 w_2 - w_1 v_2)^2 \\ &= (v_1^2 + v_2^2 + v_3^2)(w_1^2 + w_2^2 + w_3^2) - (v_1 w_1 + v_2 w_2 + v_3 w_3)^2 \\ &= \|v\|^2 \|w\|^2 (1 - \cos^2 \theta), \end{aligned}$$

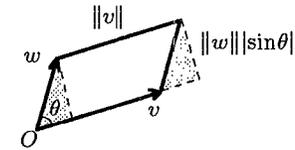


Figura 13.7 La norma del prodotto vettore.

e quindi (ii) è soddisfatta.

(iii) I vettori v e w sono linearmente dipendenti se e solo se $\sin \theta = 0$, per cui l'asserto segue da (ii).

(iv) Infatti

$$\begin{aligned} \det \begin{vmatrix} v & w & v \wedge w \end{vmatrix} &= \det \begin{vmatrix} v_1 & w_1 & \det \begin{vmatrix} v_2 & w_2 \\ v_3 & w_3 \end{vmatrix} \\ v_2 & w_2 & -\det \begin{vmatrix} v_1 & w_1 \\ v_3 & w_3 \end{vmatrix} \\ v_3 & w_3 & \det \begin{vmatrix} v_1 & w_1 \\ v_2 & w_2 \end{vmatrix} \end{vmatrix} \\ &= \left(\det \begin{vmatrix} v_2 & w_2 \\ v_3 & w_3 \end{vmatrix} \right)^2 + \left(\det \begin{vmatrix} v_1 & w_1 \\ v_3 & w_3 \end{vmatrix} \right)^2 + \left(\det \begin{vmatrix} v_1 & w_1 \\ v_2 & w_2 \end{vmatrix} \right)^2 \geq 0, \end{aligned}$$

cd è uguale a zero se e solo se $v \wedge w = O$, cioè se e solo se v e w sono linearmente dipendenti, e anche (iv) vale.

(v) Supponiamo che $u \in \mathbb{R}^3$ soddisfi (i), (ii) e (iv). Se v e w sono linearmente dipendenti, allora (ii) forza $u = O = v \wedge w$. Se sono linearmente indipendenti, l'ortogonale di $\text{Span}(v, w)$ ha dimensione 1, per cui necessariamente $u = \lambda(v \wedge w)$ per qualche $\lambda \in \mathbb{R}$. La (ii) allora forza $\lambda = \pm 1$, e la (iv) infine implica $\lambda = 1$.

(vi) e (vii) sono conseguenze immediate della definizione. \square

Osservazione 13.8 Il prodotto vettore non è un “prodotto” come lo intendiamo di solito: non è associativo. Per esempio,

$$e_1 \wedge (e_2 \wedge e_2) = e_1 \wedge O = O \neq -e_1 = e_3 \wedge e_2 = (e_1 \wedge e_2) \wedge e_2.$$

Il punto è che il prodotto vettore in realtà non è un'operazione su \mathbb{R}^3 ma sul suo duale, che per una strana combinazione di dimensioni si può rileggere di nuovo su \mathbb{R}^3 . Per spiegare di cosa si tratta, consideriamo uno spazio vettoriale V , e prendiamo due elementi $\varphi, \psi \in V'$ nel duale. Allora possiamo definire una forma bilineare $\varphi \wedge \psi: V \times V \rightarrow \mathbb{R}$ ponendo

$$\forall v, w \in V \quad (\varphi \wedge \psi)(v, w) = \varphi(v)\psi(w) - \varphi(w)\psi(v);$$

nota che $(\varphi \wedge \psi)(w, v) = -(\varphi \wedge \psi)(v, w)$, cioè $\varphi \wedge \psi$ è antisimmetrica. Inoltre, se V è uno spazio vettoriale metrico a ogni $v \in V$ possiamo associare l'applicazione lineare $\varphi_v \in V'$ definita da $\varphi_v(w) = \langle w, v \rangle$ per ogni $w \in V$ (vedi il Teorema 12C.1). Di

conseguenza a ogni coppia v, w di vettori di uno spazio vettoriale metrico V possiamo associare la forma bilineare antisimmetrica $\varphi_v \wedge \varphi_w$.

Su \mathbb{R}^3 , lo spazio delle forme bilineari antisimmetriche è isomorfo allo spazio delle matrici quadrate di ordine 3 antisimmetriche, che ha esattamente dimensione 3 (Esercizio 5.9) — e quindi a sua volta è isomorfo a \mathbb{R}^3 . Dunque in questo caso particolare la forma bilineare antisimmetrica $\varphi_v \wedge \varphi_w$ può essere interpretata come un vettore di \mathbb{R}^3 , che risulta essere proprio il prodotto vettore $v \wedge w$. Questa combinazione di dimensioni è il motivo per cui il “prodotto vettore” (che in realtà ha ben poco sia del prodotto che del vettore) può venire introdotto solo su \mathbb{R}^3 .

Osservazione 13.9 Il prodotto vettore può venire usato per esprimere la distanza di un punto da una retta, e la distanza fra due rette. Consideriamo il punto $P_0 \in \mathbb{R}^3$ e la retta r di equazione parametrica $P = P_1 + tv_1$. Ponendo $v_0 = \overrightarrow{OP_1} - \overrightarrow{OP_0}$ la formula (13.11) ci dice che

$$d(P_0, r) = d(P_0, P_1) |\sin \theta| = \|v_0\| |\sin \theta| = \frac{\|v_0 \wedge v_1\|}{\|v_1\|}. \quad (13.15)$$

Siano poi r_0 ed r_1 due rette di equazione parametrica $P = P_j + tv_j$ con $v_j = (x_j, y_j, z_j)$ per $j = 0, 1$. Poniamo $v_2 = \overrightarrow{OP_1} - \overrightarrow{OP_0}$; allora si verifica subito che

$$\langle v_0 \wedge v_1, v_2 \rangle = \det |v_0 \ v_1 \ v_2|. \quad (13.16)$$

Di conseguenza la (13.13) diventa

$$d(r_0, r_1) = \left| \frac{\langle v_0 \wedge v_1, \overrightarrow{OP_1} - \overrightarrow{OP_0} \rangle}{\|v_0 \wedge v_1\|} \right|. \quad (13.17)$$

Quest'ultima Osservazione suggerisce la seguente

Definizione 13.16 Se $v_1, v_2, v_3 \in \mathbb{R}^3$ sono tre vettori, il loro *prodotto misto* è

$$[v_1, v_2, v_3] = \langle v_1 \wedge v_2, v_3 \rangle \in \mathbb{R}.$$

Se θ è l'angolo fra $v_1 \wedge v_2$ e v_3 , si ha

$$[v_1, v_2, v_3] = \|v_1 \wedge v_2\| \|v_3\| \cos \theta.$$

Ora, il parallelepipedo generato da v_1, v_2 e v_3 (vedi la Figura 13.8) ha area di base esattamente $\|v_1 \wedge v_2\|$ e altezza $\|v_3\| \cos \theta$; quindi il prodotto misto dà il volume del parallelepipedo. Ricordando la (13.16) abbiamo quindi fatto vedere come il determinante dia il volume (orientato) del parallelepipedo generato dai tre vettori; confronta con le Osservazioni 9.1 e 9.11.

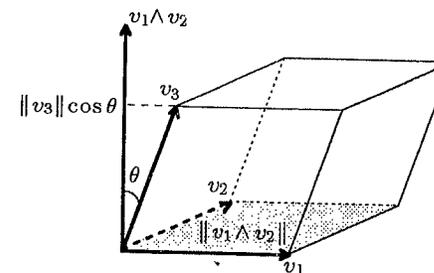


Figura 13.8 Prodotto misto e volume del parallelepipedo.

Esercizi

13.1 Presa una retta $r \subset \mathcal{E}^3$ passante per l'origine, dimostra che la Definizione 13.6 di proiezione ortogonale su r coincide con quella data nel Paragrafo 12.5. Se r non passa per l'origine, prendi un punto $Q \in r$ e dimostra la stessa cosa lavorando in \mathcal{V}_Q^3 .

13.2 Preso un piano π passante per l'origine, dimostra che la Definizione 13.10 di proiezione ortogonale su π coincide con quella data nel Paragrafo 12.5.

13.3 Nel ricavare la formula (13.6) dell'angolo fra una retta e un piano, abbiamo implicitamente assunto che le tre rette r, r' ed s siano complanari (altrimenti l'angolo fra s ed r non sarebbe stato $\pi/2 - \theta$). Dimostra che sono davvero complanari.

13.4 Calcola il coseno dell'angolo fra le rette r_0 ed r_1 (orientate a tuo piacimento) di equazioni cartesiane

$$r_0: \begin{cases} 2x + y - z = 0, \\ x - y + z = 1; \end{cases} \quad r_1: \begin{cases} x - y = 1, \\ y - z = 0. \end{cases}$$

13.5 Trova equazioni parametriche e cartesiane delle rette e del piano passanti per $P_0 = (0, 1, 0)$ e ortogonali alla retta r_1 dell'Esercizio precedente.

13.6 Calcola il coseno dell'angolo fra i piani π_0 e π_1 (orientati a tuo piacimento) di equazioni cartesiane rispettivamente $x + y + z = 1$ e $x - y - z = 2$.

13.7 Calcola il seno dell'angolo fra la retta r_0 dell'Esercizio 13.4 e il piano π_1 dell'Esercizio 13.6.

13.8 Date due rette sghembe, esiste sempre un piano parallelo a entrambe? E un piano ortogonale a entrambe?

13.9 Trova equazioni cartesiane e parametriche della retta r passante per il punto $P_0 = (1, 1, 1)$ e ortogonale e incidente alla retta s di equazioni cartesiane $2x - y + z = 1$ e $x - y - z = 1$.

13.10 Sia $P_0 = (1, 0, 1)$. Calcola la distanza di P_0 dalle rette dell'Esercizio 13.4 e dai piani dell'Esercizio 13.6.

13.11 Calcola la distanza fra le due rette dell'Esercizio 13.4.

13.12 Trova equazioni cartesiane e parametriche dei piani che distano $2/\sqrt{6}$ dal punto $P_0 = (1, 1, 1)$ e contenenti la retta r di equazioni cartesiane $x - y = 1$ e $y + z = 1$.

13.13 Siano r_0 e r_1 due rette non parallele. Dimostra che esiste un'unica retta s ortogonale e incidente sia a r_0 che a r_1 . Dimostra che $d(r_0, r_1)$ è uguale a $d(H_0, H_1)$, dove H_j è il punto d'intersezione fra s ed r_j , per $j = 0, 1$ (vedi la Figura 13.6).

13.14 Date le rette r_1 ed r_2 di equazioni rispettivamente

$$r_1 : \begin{cases} x = 2 + s, \\ y = -2s, \\ z = -1 + 3s, \end{cases}, \quad r_2 : \begin{cases} x + y + 2z = 0, \\ x + y + z = 0, \end{cases}$$

calcola la distanza fra r_1 e la retta s passante per $P_0 = (2, 1, 3)$, ortogonale a r_1 e incidente a r_2 .

13.15 Trova il prodotto vettore di $v = (1, 2, 1)$ e $w = (a, 2a, 1)$ al variare di $a \in \mathbb{R}$.

13.16 Fissato un vettore $v_0 \in \mathbb{R}^3$, siano $S, T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ le applicazioni definite da $S(v) = v \wedge v_0$ e $T(v) = v \wedge v_0 + \langle v, v_0 \rangle v_0$, dove $\langle \cdot, \cdot \rangle$ è il prodotto scalare canonico. Dimostra che S e T sono lineari, e determina le loro immagini e i loro nuclei.

COMPLEMENTI

13C.1 Isometrie

L'obiettivo di questo paragrafo è scoprire quali affinità (vedi i Complementi al Capitolo 10) conservano le distanze. Cominciamo con due definizioni.

Definizione 13C.1 Uno spazio (piano, retta) euclideo è uno spazio (piano, retta) affine su cui sia stata fissata un'unità di misura.

Definizione 13C.2 Sia φ un'affinità fra due spazi (piani, rette) euclidei. Diremo che φ è un'isometria (non necessariamente lineare) se conserva le distanze, cioè se

$$d(\varphi(P_1), \varphi(P_2)) = d(P_1, P_2)$$

per ogni coppia di punti P_1, P_2 dello spazio (piano, retta) affine di partenza.

Consideriamo un'affinità $\varphi: \Sigma_1 \rightarrow \Sigma_2$ fra due spazi euclidei; il caso di affinità fra piani o rette sarà analogo. Fissato un sistema di riferimento cartesiano \mathcal{R}_1 in Σ_1 e uno \mathcal{R}_2 in Σ_2 sappiamo che φ si esprime in coordinate come $y = Bx + c$ per qualche $B \in GL_3(\mathbb{R})$ e $c \in \mathbb{R}^3$, dove $x \in \mathbb{R}^3$ sono le coordinate di un punto $P \in \Sigma_1$ rispetto a \mathcal{R}_1 e $y \in \mathbb{R}^3$ sono le coordinate di $\varphi(P) \in \Sigma_2$ rispetto a \mathcal{R}_2 . Dunque φ è un'isometria se e solo se per ogni coppia di punti $P, P' \in \Sigma_1$ si ha

$$\|x' - x\| = d(P, P') = d(\varphi(P), \varphi(P')) = \|(Bx' + c) - (Bx + c)\| = \|B(x' - x)\|.$$

Dunque la matrice B conserva la norma; la Proposizione 12.20 e il Corollario 12.22 implicano che B è ortogonale.

Viceversa, se B è una matrice ortogonale, è chiaro che un'affinità data in coordinate da $y = Bx + c$ conserva le distanze (per non parlare degli angoli); quindi

Proposizione 13C.1 Sia φ un'affinità fra due spazi (piani, rette) euclidei. Allora φ è un'isometria se e solo se rispetto a qualunque sistema di riferimento cartesiano scelto nei due spazi (piani, rette) la φ è rappresentata in coordinate da $y = Bx + c$, dove B è una matrice ortogonale. In particolare, un'isometria conserva anche gli angoli.

In realtà possiamo dimostrare un risultato molto più forte: una qualunque applicazione fra spazi (rette, piani) euclidei che conserva le distanze è necessariamente un'affinità — e quindi un'isometria nel senso della Definizione 13C.2.

Considereremo solo il caso degli spazi; per rette e piani la dimostrazione è analoga.

Teorema 13C.2 Sia $\varphi: \Sigma_1 \rightarrow \Sigma_2$ un'applicazione fra spazi euclidei tale che

$$\forall P, Q \in \Sigma_1 \quad d_2(\varphi(P), \varphi(Q)) = d_1(P, Q),$$

dove d_j è la distanza in Σ_j per $j = 1, 2$. Allora φ è necessariamente un'affinità — e quindi un'isometria.

Dimostrazione. Fissiamo dei sistemi di riferimento cartesiani in Σ_1 e Σ_2 in modo da poter identificare sia Σ_1 che Σ_2 con \mathbb{R}^3 . Allora φ diventa un'applicazione $\tilde{\varphi}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ che conserva la distanza indotta dal prodotto scalare canonico; dobbiamo dimostrare che $\tilde{\varphi}(x) = Bx + c$ per qualche $c \in \mathbb{R}^3$ e $B \in GL_3(\mathbb{R})$.

Prima di tutto, definiamo $\tilde{\varphi}_1: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ ponendo

$$\tilde{\varphi}_1(x) = \tilde{\varphi}(x) - \tilde{\varphi}(O);$$

chiaramente $\tilde{\varphi}_1$ conserva ancora le distanze, e inoltre $\tilde{\varphi}_1(O) = O$. Vogliamo dimostrare che $\tilde{\varphi}_1$ è lineare.

Siccome $\tilde{\varphi}_1$ conserva le distanze abbiamo

$$\forall v, w \in \mathbb{R}^3 \quad \|\tilde{\varphi}_1(v) - \tilde{\varphi}_1(w)\|^2 = \|v - w\|^2;$$

quindi

$$\|\tilde{\varphi}_1(v)\|^2 - 2\langle \tilde{\varphi}_1(v), \tilde{\varphi}_1(w) \rangle + \|\tilde{\varphi}_1(w)\|^2 = \|v\|^2 - 2\langle v, w \rangle + \|w\|^2.$$

In particolare prendendo $w = O$ otteniamo $\|\tilde{\varphi}_1(v)\| = \|v\|$. Ne segue che

$$\forall v, w \in \mathbb{R}^3 \quad \langle \tilde{\varphi}_1(v), \tilde{\varphi}_1(w) \rangle = \langle v, w \rangle. \quad (13C.1)$$

Sia $\{e_1, e_2, e_3\}$ la base canonica di \mathbb{R}^3 . La (13C.1) implica che $\{\tilde{\varphi}_1(e_1), \tilde{\varphi}_1(e_2), \tilde{\varphi}_1(e_3)\}$ è una base ortonormale di \mathbb{R}^3 . Allora se $v = (x_1, x_2, x_3) \in \mathbb{R}^3$ la Proposizione 12.4 dà

$$\tilde{\varphi}_1(v) = \sum_{j=1}^3 \langle \tilde{\varphi}_1(v), \tilde{\varphi}_1(e_j) \rangle \tilde{\varphi}_1(e_j) = \sum_{j=1}^3 \langle v, e_j \rangle \tilde{\varphi}_1(e_j) = \sum_{j=1}^3 x_j \tilde{\varphi}_1(e_j).$$

per cui $\tilde{\varphi}_1$ è lineare (vedi la dimostrazione della Proposizione 5.2). Essendo necessariamente iniettiva (perché?) esiste una matrice invertibile $B \in GL_3(\mathbb{R})$ tale che $\tilde{\varphi}_1(x) = Bx$, e quindi $\tilde{\varphi}(x) = Bx + c$ dove $c = \tilde{\varphi}(O)$. \square

Osservazione 13C.1 La stessa dimostrazione può venire adattata per far vedere che un'applicazione $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}^n$ che conserva la distanza indotta dal prodotto scalare canonico è necessariamente della forma $\varphi(x) = Bx + c$ per qualche $B \in O(n)$ e $c \in \mathbb{R}^n$.

13C.2 Coniche

Rette e piani non sono i soli oggetti studiati dalla geometria metrica; ve ne sono moltissimi altri, fra cui i più importanti sono senza dubbio le coniche, a cui è dedicata questo paragrafo.

Cominciamo col caso più semplice e noto.

Definizione 13C.3 Una circonferenza X di centro $C \in \mathcal{E}^2$ e raggio $r > 0$ è il luogo² dei punti $P \in \mathcal{E}^2$ che distano esattamente r da C . In simboli,

$$X = \{P \in \mathcal{E}^2 \mid d(P, C) = r\}.$$

È facile trovare equazioni cartesiane e parametriche di una circonferenza. Scelto un sistema di riferimento cartesiano $\mathcal{R} = RC(O, A_1, A_2)$ di \mathcal{E}^2 , supponiamo che C abbia coordinate $(c_1, c_2) \in \mathcal{R}^2$. Allora il punto P di coordinate (x, y) appartiene a X se e solo se

$$(x - c_1)^2 + (y - c_2)^2 = r^2. \tag{13C.2}$$

Dividendo (13C.2) per r^2 troviamo che il punto di coordinate $((x - c_1)/r, (y - c_2)/r)$ dista esattamente 1 dall'origine, per cui è della forma $(\cos t, \sin t)$ per qualche $t \in \mathbb{R}$. Questo vuol dire che equazioni parametriche della circonferenza sono (vedi la Figura 13.9)

$$\begin{cases} x = c_1 + r \cos t, \\ y = c_2 + r \sin t. \end{cases} \tag{13C.3}$$

La circonferenza è un tipo particolare di ellisse:

Definizione 13C.4 Un'ellisse E di fuochi $F, F' \in \mathcal{E}^2$ e semiasse $a > 0$ è il luogo dei punti $P \in \mathcal{E}^2$ tali che la somma delle distanze di P da F ed F' sia la costante $2a$, cioè

$$E = \{P \in \mathcal{E}^2 \mid d(P, F) + d(P, F') = 2a\}.$$

Se i due fuochi coincidono otteniamo una circonferenza di centro $F = F'$ e raggio a .

In modo analogo possiamo definire l'iperbole:

² "Luogo" è il sinonimo di "insieme" tradizionalmente usato in questo contesto.

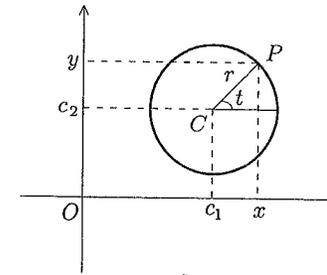


Figura 13.9 La circonferenza.

Definizione 13C.5 Un'iperbole I di fuochi $F, F' \in \mathcal{E}^2$ e semiasse $a > 0$ è il luogo dei punti $P \in \mathcal{E}^2$ tali che il valore assoluto della differenza delle distanze di P da F ed F' sia la costante $2a$, cioè

$$I = \{P \in \mathcal{E}^2 \mid |d(P, F) - d(P, F')| = 2a\}.$$

Osservazione 13C.2 La disuguaglianza triangolare ci dice che

$$|d(P, F) - d(P, F')| \leq d(F, F') \leq d(P, F) + d(P, F');$$

quindi se $2c = d(F, F') \geq 0$ è la distanza fra i due fuochi troviamo che per un'ellisse dev'essere $a \geq c$, mentre per un'iperbole dev'essere $a \leq c$. Se $a = c$, i punti F, P ed F' sono allineati; l'ellisse si riduce allora al segmento $\overline{F'F}$, mentre l'iperbole si riduce al complementare di questo segmento sulla retta passante per F' ed F .

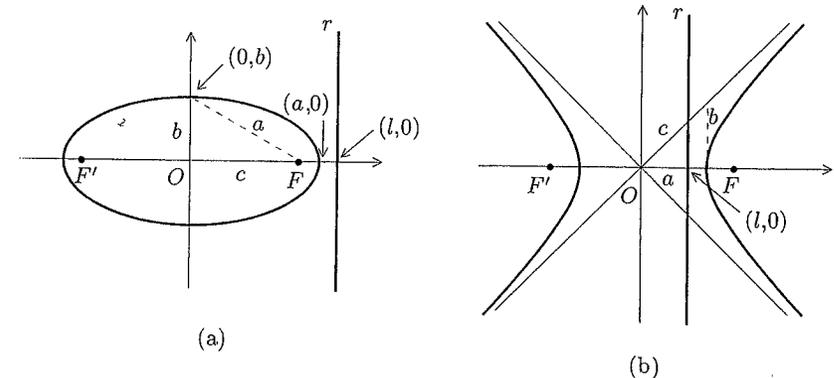


Figura 13.10 (a) L'ellisse. (b) L'iperbole.

Vogliamo trovare equazioni cartesiane per ellissi e iperboli. Per semplificare i conti, scegliamo un sistema di riferimento cartesiano adattato alla situazione (Figura 13.10).

Prendiamo il sistema $\mathcal{R} = RC(O, A_1, A_2)$ dove O è il punto medio del segmento $\overline{FF'}$, e $\overrightarrow{OA_1}$ è parallelo allo stesso segmento, con verso che punta da F' a F . Rispetto a \mathcal{R} , F ha coordinate $(c, 0)$ e F' coordinate $(-c, 0)$.

Indichiamo con (x, y) le coordinate del punto $P \in \mathcal{E}^2$ rispetto al sistema di riferimento cartesiano \mathcal{R} scelto. Allora P appartiene all'ellisse E se e solo se

$$\sqrt{(x-c)^2 + y^2} + \sqrt{(x+c)^2 + y^2} = 2a.$$

Elevando al quadrato entrambi i membri e semplificando otteniamo che $P \in E$ se e solo se

$$x^2 + y^2 + c^2 - 2a^2 = -\sqrt{(x-c)^2 + y^2} \sqrt{(x+c)^2 + y^2};$$

elevando di nuovo al quadrato e risemplificando troviamo che $P \in E$ se e solo se

$$(a^2 - c^2)x^2 + a^2y^2 = a^2(a^2 - c^2),$$

ovvero se e solo se (ricordando che $a > c$).

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad (13C.4)$$

dove $b^2 = |a^2 - c^2|$.

Analogamente P appartiene all'iperbole I se e solo se

$$|\sqrt{(x-c)^2 + y^2} - \sqrt{(x+c)^2 + y^2}| = 2a.$$

Procedendo come prima troviamo che $P \in I$ se e solo se

$$x^2 + y^2 + c^2 - 2a^2 = \sqrt{(x-c)^2 + y^2} \sqrt{(x+c)^2 + y^2};$$

se e solo se

$$(a^2 - c^2)x^2 + a^2y^2 = a^2(a^2 - c^2),$$

ovvero se e solo se (ricordando che stavolta $a < c$)

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \quad (13C.5)$$

dove di nuovo $b^2 = |a^2 - c^2|$.

Da queste equazioni è semplice ricavare equazioni parametriche. Nel caso dell'ellisse, (13C.4) implica che

$$\begin{cases} x = a \cos t, \\ y = b \sin t, \end{cases} \quad (13C.6)$$

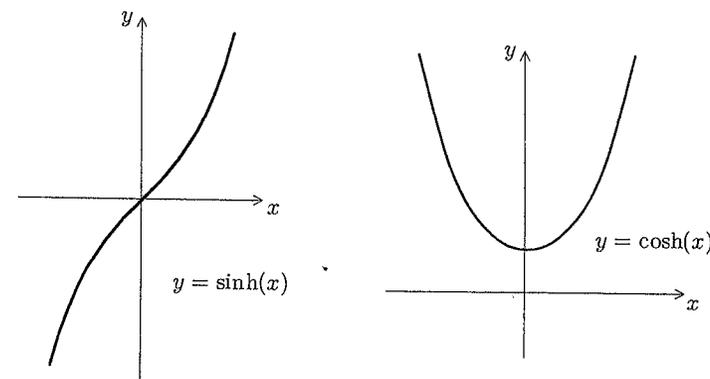


Figura 13.11 Seno e coseno iperbolici.

sono equazioni parametriche di E . Per l'iperbole, ci serve una definizione.

Definizione 13C.6 Le funzioni coseno iperbolico $\cosh: \mathbb{R} \rightarrow \mathbb{R}$ e seno iperbolico $\sinh: \mathbb{R} \rightarrow \mathbb{R}$ sono definite da

$$\cosh t = \frac{e^t + e^{-t}}{2}, \quad \sinh t = \frac{e^t - e^{-t}}{2}.$$

La Figura 13.11 ne contiene i grafici.

È facile verificare (esercizio) che

$$(\cosh t)^2 - (\sinh t)^2 = 1.$$

Viceversa, se $(x, y) \in \mathbb{R}^2$ è tale che $x^2 - y^2 = 1$, poniamo

$$t = \log(y + \sqrt{1 + y^2});$$

allora $\sinh t = y$, e $1 + y^2 = (\cosh t)^2$, per cui $x = \pm \cosh t$. Dunque le coppie $(x, y) \in \mathbb{R}^2$ che soddisfano l'equazione $x^2 - y^2 = 1$ sono tutte e sole quelle della forma $(\pm \cosh t, \sinh t)$, e quindi (13C.5) ci dice che equazioni parametriche dell'iperbole sono

$$\begin{cases} x = \pm a \cosh t, \\ y = b \sinh t. \end{cases} \quad (13C.7)$$

Osservazione 13C.3 L'iperbole è fatta da due rami separati; conseguentemente abbiamo due equazioni parametriche distinte (una col segno più, l'altra col segno meno), una per ciascun ramo.

Osservazione 13C.4 Se $(x, y) \in I$, la (13C.7) implica che

$$\frac{y}{x} = \pm \frac{b \sinh t}{a \cosh t} = \pm \frac{b}{a} \frac{e^t - e^{-t}}{e^t + e^{-t}} = \pm \frac{b}{a} \frac{e^{2t} - 1}{e^{2t} + 1}.$$

Questa quantità tende a $\pm b/a$ per $t \rightarrow \infty$. In altre parole, andando verso l'infinito l'iperbole tende ad avvicinarsi alle rette $y = \pm(b/a)x$, dette *asintoti* dell'iperbole (che sono indicati anche nella Figura 13.10).

Un'altra curva parente stretta di iperboli ed ellissi è la parabola.

Definizione 13C.7 Una parabola C di fuoco $F \in \mathcal{E}^2$ e direttrice la retta r è il luogo dei punti $P \in \mathcal{E}^2$ la cui distanza da r è uguale alla distanza da F , cioè

$$C = \{P \in \mathcal{E}^2 \mid d(P, r) = d(P, F)\}.$$

Il numero $p = d(F, r)/2$ è detto *parametro* della parabola.

Per trovare un'equazione cartesiana della parabola, scegliamo di nuovo opportunamente il sistema di riferimento cartesiano $\mathcal{R} = RC(O, A_1, A_2)$. Sia H la proiezione ortogonale di F su r ; prendiamo come origine il punto medio O del segmento \overline{FH} , e poi scegliamo $\overrightarrow{OA_1}$ parallelo a r e $\overrightarrow{OA_2}$ ortogonale a r diretto verso F (vedi la Figura 13.12).

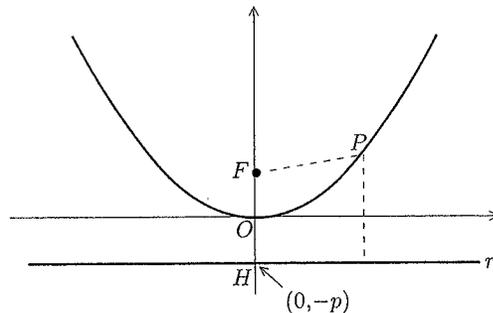


Figura 13.12 La parabola.

In questo riferimento il fuoco F ha coordinate $(0, p)$, e la direttrice r ha equazione cartesiana $y = -p$. Dunque il punto $P \in \mathcal{E}^2$ di coordinate (x, y) appartiene alla parabola C se e solo se

$$\sqrt{x^2 + (y - p)^2} = |y + p|,$$

ovvero $P \in C$ se e solo se

$$y = \frac{1}{4p} x^2. \quad (13C.8)$$

Le equazioni parametriche sono ovvie:

$$\begin{cases} x = t, \\ y = t^2/4p. \end{cases} \quad (13C.9)$$

Osservazione 13C.5 Le equazioni (13C.4), (13C.5) e (13C.8) sono equazioni cartesiane di ellissi, iperboli e parabole in sistemi di riferimento cartesiani particolari. Per scoprire che aspetto hanno le equazioni cartesiane di ellissi, iperboli e parabole in un sistema di riferimento cartesiano qualsiasi basta cambiare coordinate, passando dal sistema iniziale qualunque al sistema finale adattato alla figura. Il cambiamento di coordinate è della forma

$$\begin{pmatrix} x \\ y \end{pmatrix} = B \begin{pmatrix} x' \\ y' \end{pmatrix} + \begin{pmatrix} c_1 \\ c_2 \end{pmatrix},$$

come visto all'inizio del Paragrafo 13.1; quindi le equazioni cartesiane diventano del tipo

$$ax^2 + 2bxy + cy^2 + 2dx + 2ey + f = 0, \quad (13C.10)$$

per opportuni $a, b, c, d, e, f \in \mathbb{R}$, indipendentemente dal fatto che si tratti di ellissi, iperboli o parabole (o circonferenze). Nel Capitolo 16 vedremo dei metodi per stabilire che tipo di curva rappresenta un'equazione del tipo (13C.10).

Osservazione 13C.6 Non è un caso che ellissi, iperboli e parabole siano rappresentabili da equazioni dello stesso tipo; infatti, si ottengono tutte intersecando un cono circolare retto³ con un piano. Per questo motivo ellissi, iperboli e parabole sono collettivamente note come *sezioni coniche* — o più semplicemente *coniche*.

Un'ulteriore dimostrazione del fatto che ellissi, iperboli e parabole sono strettamente imparentate segue dal fatto che si possono definire contemporaneamente usando il concetto di eccentricità.

Teorema 13C.3 Siano dati una retta r , un punto $F \notin r$ e un numero $e > 0$. Sia $X \subset \mathcal{E}^2$ il luogo dei punti $P \in \mathcal{E}^2$ tali che

$$d(P, F) = e d(P, r). \quad (13C.11)$$

Allora X è un'ellisse se $e < 1$, una parabola se $e = 1$, un'iperbole se $e > 1$. Viceversa, ogni conica (che non sia una circonferenza) si ottiene in questo modo.

Definizione 13C.8 Il numero e è detto *eccentricità* della conica, la retta r *direttrice* e il punto F *fuoco*.

Dimostrazione. Se $e = 1$ non c'è nulla da dimostrare; supporremo quindi $e \neq 1$.

³ Vedi il prossimo paragrafo per una definizione precisa di cono circolare retto.

Scegliamo un sistema di riferimento cartesiano $RC(O, A_1, A_2)$ in modo che il vettore $\overrightarrow{OA_1}$ sia ortogonale a r , il vettore $\overrightarrow{OA_2}$ sia parallelo a r , e il fuoco F abbia coordinate $(c, 0)$, con

$$c = \frac{e^2}{|1 - e^2|} d > 0,$$

dove $d = d(F, r)$; inoltre, scegliamo l'origine e il verso di $\overrightarrow{OA_1}$ in modo che F si trovi fra O e la direttrice se $e < 1$, mentre sia r a trovarsi fra O ed F se $e > 1$ (è la situazione rappresentata nella Figura 13.10); nota che $d < c$ quando $e > 1$. In particolare la direttrice ha equazione cartesiana $x = l$, con

$$l = \begin{cases} c + d & \text{se } e < 1, \\ c - d & \text{se } e > 1. \end{cases}$$

Rispetto a questo sistema di riferimento cartesiano, un punto $P \in \mathcal{E}^2$ di coordinate (x, y) appartiene a X se e solo se

$$\sqrt{(x - c)^2 + y^2} = e|x - l|.$$

Elevando al quadrato e notando che c è stato scelto in modo che si abbia $c = e^2 l$ sempre, troviamo che $P \in X$ se e solo se

$$(1 - e^2)x^2 + y^2 = e^2 l^2 - c^2 = \frac{e^2 d^2}{1 - e^2},$$

cioè se e solo se

$$\frac{(1 - e^2)^2}{e^2 d^2} x^2 + \frac{1 - e^2}{e^2 d^2} y^2 = 1,$$

che è proprio l'equazione di un'ellisse se $e < 1$, e di un'iperbole se $e > 1$.

Viceversa, consideriamo un'ellisse (o un'iperbole) di semiasse $a > 0$ e distanza fra i fuochi $2c > 0$. Abbiamo visto che in tal caso l'equazione è data da

$$\frac{x^2}{a^2} + \frac{y^2}{a^2 - c^2} = 1;$$

allora ci basta trovare $e, d > 0$ tali che

$$a^2 = \frac{e^2 d^2}{(1 - e^2)^2} \quad \text{e} \quad a^2 - c^2 = \frac{e^2 d^2}{1 - e^2} \quad (13C.12)$$

per dimostrare che ogni conica (tranne le circonferenze) si scrive come in (13C.11). Ma infatti (13C.12) implica che

$$1 - e^2 = \frac{a^2 - c^2}{a^2} = 1 - \frac{c^2}{a^2} \quad \text{per cui} \quad e = \frac{c}{a},$$

e quindi

$$d^2 = \frac{1 - e^2}{e^2} (a^2 - c^2) = \frac{(a^2 - c^2)^2}{c^2} \quad \text{per cui} \quad d = \frac{b^2}{c},$$

dove come al solito $b^2 = |a^2 - c^2|$. \square

Osservazione 13C.7 In un certo senso, le circonferenze possono essere considerate come il caso limite di coniche con eccentricità zero e direttrice la retta all'infinito (vedi il Paragrafo 10C.2).

Concludiamo questo paragrafo studiando le possibili posizioni reciproche di una retta e una conica. Sia r una retta di equazione parametrica

$$\begin{cases} x \\ y \end{cases} = \begin{cases} x_0 \\ y_0 \end{cases} + t \begin{cases} l \\ m \end{cases}, \quad (13C.13)$$

e prendiamo una conica X di equazione cartesiana (13C.10). Sostituendo (13C.13) in (13C.10) troviamo che il punto $(x_0 + tl, y_0 + tm) \in r$ appartiene alla conica se e solo se $t \in \mathbb{R}$ è soluzione dell'equazione

$$(al^2 + 2blm + cm^2)t^2 + 2(alx_0 + bly_0 + bmx_0 + cmy_0 + dl + cm)t + (ax_0^2 + 2bx_0y_0 + cy_0^2 + 2dx_0 + 2cy_0 + f) = 0. \quad (13C.14)$$

Questa è un'equazione di secondo grado nell'incognita t , per cui può avere zero, una o due (o infinite) soluzioni reali, che corrispondono a zero, uno o due (o infiniti) punti d'intersezione fra la retta e la conica.

Definizione 13C.9 Un vettore $v_0 = (l, m) \neq (0, 0)$ tale che $al^2 + 2blm + cm^2 = 0$ è detto *vettore asintotico* della conica di equazione (13C.10). Una retta di vettore direttore un vettore asintotico è detta *retta asintotica* della conica. Se r è una retta asintotica tale che anche il coefficiente di t in (13C.14) si annulla, diremo che r è un *asintoto* della conica se il termine noto in (13C.14) è diverso da zero; che è *contenuta* nella conica se il termine noto in (13C.14) è nullo.

Definizione 13C.10 Diremo che una retta non asintotica è *secante* (rispettivamente, *tangente a*, *esterna a*) una conica X se interseca la conica in due punti (rispettivamente, un punto, nessun punto).

Supponiamo che il punto $P_0 = (x_0, y_0)$ appartenga alla conica X , cioè si abbia

$$ax_0^2 + 2bx_0y_0 + cy_0^2 + 2dx_0 + 2ey_0 + f = 0. \quad (13C.15)$$

Allora una retta di equazione parametrica (13C.13) è tangente alla conica se e solo se $al^2 + 2blm + cm^2 \neq 0$ e (13C.14) ha come unica soluzione $t = 0$, che accade se e solo se

$$(ax_0 + by_0 + d)l + (bx_0 + cy_0 + e)m = 0. \quad (13C.16)$$

Questa equazione ha sempre almeno uno dei coefficienti non nulli (e quindi lo spazio delle soluzioni ha sempre dimensione 1) tranne nel caso in cui si abbia

$$\begin{cases} ax_0 + by_0 + d = 0, \\ bx_0 + cy_0 + e = 0, \\ dx_0 + ey_0 + f = 0, \end{cases}$$

dove la terza equazione è conseguenza delle prime due e di (13C.15). Questo succede se e solo se $(x_0, y_0, 1)$ appartiene al nucleo della matrice

$$A = \begin{vmatrix} a & b & d \\ b & c & e \\ d & e & f \end{vmatrix}. \quad (13C.17)$$

Definizione 13C.11 Una conica per cui si abbia $\text{Ker } A = \{O\}$, dove A è la matrice data in (13C.17), è detta *non degenera*.

Dunque abbiamo dimostrato che *per ogni punto di una conica non degenera passa una e una sola retta tangente*, il cui vettore direttore è soluzione di (13C.16). Come vedremo nel Capitolo 16, le coniche non degeneri sono esattamente le ellissi, le iperboli e le parabole.

ESEMPIO 13C.1 Vediamo cosa diventano questi concetti nel caso di ellissi, iperboli e parabole. Si verifica subito che un'ellisse non possiede rette asintotiche di alcun genere; inoltre se $P_0 = (x_0, y_0)$ appartiene all'ellisse E di equazione (13C.4), allora il vettore direttore della retta tangente a E passante per P_0 soddisfa l'equazione

$$\frac{x_0}{a}l + \frac{y_0}{b}m = 0.$$

In particolare, se la retta è una circonferenza (cioè $a = b$) allora la retta tangente è sempre perpendicolare al raggio.

Per una parabola di equazione (13C.8) le rette ortogonali alla direttrice sono rette asintotiche; in particolare, la retta asintotica passante per il fuoco è detta *asse della parabola*, e il punto in cui interseca la parabola si chiama *vertice della parabola*. Il vettore direttore della retta tangente alla parabola nel punto (x_0, y_0) soddisfa l'equazione

$$\frac{x_0}{2p}l - m = 0.$$

Infine, un'iperbole I di equazione (13C.5) ammette (a meno di multipli) due vettori asintotici: $(1, b/a)$ e $(1, -b/a)$. In particolare, si verifica facilmente che l'iperbole ammette esattamente due asintoti: le rette di equazione cartesiana $y = \pm(b/a)x$, come già anticipato nell'Osservazione 13C.4. Infine, il vettore direttore della retta tangente all'iperbole nel punto (x_0, y_0) soddisfa l'equazione

$$\frac{x_0}{a}l - \frac{y_0}{b}m = 0.$$

13C.3 Esempi di superfici

In questo paragrafo daremo alcuni esempi di superfici in \mathcal{E}^3 , cominciando col descrivere due modi per costruire superfici a partire da una curva in un piano.

Definizione 13C.12 Sia $\pi \subset \mathcal{E}^3$ un piano, $\gamma \subset \pi$ una curva e $r \subset \mathcal{E}^3$ una retta non parallela a π . Il *cilindro di base γ e asse r* è l'insieme delle rette parallele a r passanti per i punti di γ .

Definizione 13C.13 Sia $\pi \subset \mathcal{E}^3$ un piano, $\gamma \subset \pi$ una curva e $P_0 \notin \pi$ un punto esterno a π . Il *cono di base γ e vertice P_0* è l'insieme delle rette congiungenti P_0 ai punti di γ .

Il cilindro e il cono più noti sono quelli *circolari retti*, in cui la base è una circonferenza, e l'asse è ortogonale al piano π (rispettivamente, la proiezione ortogonale del vertice P_0 sul piano π è il centro della circonferenza). Ne esistono però molti altri; la Figura 13.13 contiene alcuni esempi.

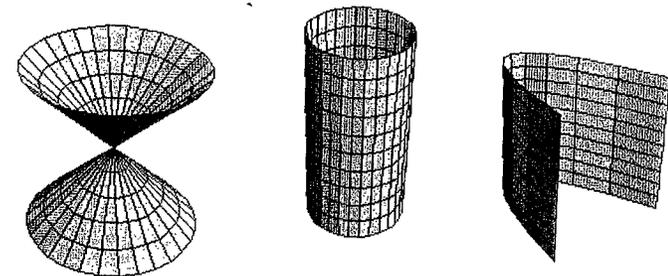


Figura 13.13 Esempi di coni e cilindri.

È facile trovare le equazioni parametriche di cilindri e coni; cominciamo coi cilindri. Scegliamo il sistema di riferimento cartesiano in modo che il piano π abbia equazione $z = 0$, e l'asse r passi per l'origine. In questo sistema la curva γ avrà equazioni parametriche del tipo

$$(x, y, z) = (f(t), g(t), 0)$$

per opportune funzioni f, g , e quindi il cilindro avrà equazioni parametriche

$$\begin{cases} x = f(t) + sl, \\ y = g(t) + sm, \\ z = sn, \end{cases} \quad (13C.18)$$

dove (l, m, n) con $n \neq 0$ sono le coordinate di un vettore direttore di r .

Per i coni, poniamo l'origine nel vertice, e facciamo in modo che il piano π abbia equazione $z = 1$. Stavolta la curva γ avrà equazioni parametriche del tipo

$$(x, y, z) = (f(t), g(t), 1),$$

per cui il cono avrà equazioni parametriche

$$\begin{cases} x = sf(t), \\ y = sg(t), \\ z = s. \end{cases} \quad (13C.19)$$

Vediamo ora le equazioni cartesiane, cominciando nuovamente dai cilindri. Il sistema di riferimento cartesiano scelto è sempre quello di prima. Supponiamo che la curva γ abbia equazione cartesiana $F(x, y) = 0$; in particolare $F: \mathbb{R}^2 \rightarrow \mathbb{R}$ è una funzione tale che

$$F(f(t), g(t)) \equiv 0.$$

Da (13C.18) ricaviamo $s = z/n$, $f(t) = x - ls$ e $g(t) = y - ms$, per cui le equazioni cartesiane del cilindro sono

$$F(x - lz/n, y - mz/n) = 0. \quad (13C.20)$$

In maniera assolutamente analoga troviamo che equazioni cartesiane per il cono dato da (13C.19) sono

$$F(x/z, y/z) = 0. \quad (13C.21)$$

ESEMPIO 13C.2 Vogliamo equazioni parametriche e cartesiane del cilindro circolare retto. Prendiamo come base γ la circonferenza di raggio $\rho > 0$ che ha equazione parametrica $(x, y) = (\rho \cos t, \rho \sin t)$ ed equazione cartesiana $x^2 + y^2 = \rho^2$. Un vettore direttore dell'asse del cilindro retto è e_3 ; quindi otteniamo

$$\begin{cases} x = \rho \cos t, \\ y = \rho \sin t, \\ z = s, \end{cases} \quad \text{e} \quad x^2 + y^2 = \rho^2$$

come equazioni parametriche e cartesiane del cilindro circolare retto. Più in generale, se il vettore direttore del cilindro ha coordinate (l, m, n) — cioè se il cilindro non è necessariamente retto — troviamo

$$\begin{cases} x = \rho \cos t + sl, \\ y = \rho \sin t + sm, \\ z = sn, \end{cases} \quad \text{e} \quad \left(x - \frac{l}{n}z\right)^2 + \left(y - \frac{m}{n}z\right)^2 = \rho^2.$$

Osservazione 13C.8 Ogni punto di \mathbb{R}^3 è contenuto in un unico cilindro circolare retto di base una circonferenza di equazione $x^2 + y^2 = \rho^2$. In altre parole, per

ogni $P = (x, y, z) \in \mathbb{R}^3$ esistono (e sono unici se $x^2 + y^2 > 0$) numeri $\rho \geq 0$, $\theta \in [0, 2\pi)$ e $z \in \mathbb{R}$ tali che

$$\begin{cases} x = \rho \cos \theta, \\ y = \rho \sin \theta, \\ z = z. \end{cases}$$

I tre numeri (ρ, θ, z) sono le coordinate cilindriche del punto $P \in \mathbb{R}^3$.

ESEMPIO 13C.3 Vogliamo equazioni parametriche e cartesiane del cono circolare retto. Le equazioni di γ sono come nell'Esempio precedente; quindi otteniamo

$$\begin{cases} x = s\rho \cos t, \\ y = s\rho \sin t, \\ z = s, \end{cases}$$

come equazioni parametriche, e $(x/z)^2 + (y/z)^2 = 1$, o meglio

$$x^2 + y^2 - z^2 = 0,$$

come equazioni cartesiane. Più in generale, se la circonferenza γ ha centro in $(x_0, y_0, 1)$ — cioè se il cono non è necessariamente retto — troviamo

$$\begin{cases} x = s(\rho \cos t + x_0), \\ y = s(\rho \sin t + y_0), \\ z = s, \end{cases}$$

come equazioni parametriche, e $((x/z) - x_0)^2 + ((y/z) - y_0)^2 = 1$, cioè

$$(x - x_0z)^2 + (y - y_0z)^2 = z^2,$$

come equazioni cartesiane.

Un'altra superficie di fondamentale importanza è, ovviamente, la sfera.

Definizione 13C.14 La sfera S di centro $C \in \mathcal{E}^3$ e raggio $r > 0$ è il luogo dei punti di \mathcal{E}^3 che sono a distanza r da C , cioè

$$S = \{P \in \mathcal{E}^3 \mid d(P, C) = r\}.$$

Fissiamo un sistema di riferimento cartesiano qualunque. L'equazione cartesiana della sfera di centro $(x_0, y_0, z_0) \in \mathbb{R}^3$ e raggio $r > 0$ è ovviamente

$$(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r^2.$$

Per trovare le equazioni parametriche, consideriamo prima il caso $C = O$ ed $r = 1$. In tal caso si ha

$$x^2 + y^2 = 1 - z^2$$

con $z \in [-1, 1]$. Ora, esiste un unico $\theta \in [-\pi/2, \pi/2]$ tale che $z = \sin \theta$; ma allora $1 - z^2 = \cos^2 \theta$ e quindi (x, y) appartiene alla circonferenza di centro O e raggio $\cos \theta$. Questo implica che

$$(x, y) = (\cos \theta \cos \varphi, \cos \theta \sin \varphi)$$

per qualche $\varphi \in [0, 2\pi)$. Di conseguenza (perché?) le equazioni parametriche della sfera generica S sono

$$\begin{cases} x = r \cos \theta \cos \varphi + x_0, \\ y = r \cos \theta \sin \varphi + y_0, \\ z = r \sin \theta + z_0. \end{cases}$$

Osservazione 13C.9 Ciascun punto $P = (x, y, z) \in \mathbb{R}^3$ appartiene a un'unica sfera di centro l'origine; quindi esistono (unici se $x^2 + y^2 + z^2 > 0$) numeri $r \geq 0$, $\theta \in [-\pi/2, \pi/2]$ e $\varphi \in [0, 2\pi)$ tali che

$$\begin{cases} x = r \cos \theta \cos \varphi, \\ y = r \cos \theta \sin \varphi, \\ z = r \sin \theta. \end{cases}$$

I tre numeri (r, θ, φ) sono le *coordinate sferiche* del punto $P \in \mathbb{R}^3$; θ è la *colatitudine* (la *latitudine* è $\pi/2 - \theta$), e φ è la *longitudine* di P .

Vediamo infine in che posizioni reciproche si possono trovare un piano π e una sfera S . Fissiamo un sistema di riferimento cartesiano in cui l'origine sia il centro di S , in modo che quest'ultima abbia equazione cartesiana $x^2 + y^2 + z^2 = r^2$. Sia $P = P_0 + sv_1 + tv_2$ un'equazione parametrica del piano π , dove possiamo supporre che $\{v_1, v_2\}$ sia una base ortonormale del sottospazio di giacitura di π . Sostituendo l'equazione parametrica del piano nell'equazione cartesiana della sfera troviamo

$$(s + \langle v_1, P_0 \rangle)^2 + (t + \langle v_2, P_0 \rangle)^2 = r^2 - \|P_0\|^2 + \langle v_1, P_0 \rangle^2 + \langle v_2, P_0 \rangle^2,$$

che è l'equazione di una circonferenza in π se il secondo membro è positivo (e diremo che il piano è *secante* alla sfera); di un punto di π se il secondo membro è nullo (il piano è *tangente* alla sfera); e non è mai soddisfatta se il secondo membro è negativo (il piano è *esterno* alla sfera). In particolare, se $P_0 \in S$ (cioè se $\|P_0\|^2 = r^2$) l'intersezione si riduce a P_0 se e solo se il piano π è ortogonale a P_0 (cioè al raggio della sfera). In altre parole, il piano tangente a una sfera S qualunque in un punto $P_0 \in S$ è il piano π ortogonale al raggio della sfera terminante in P_0 : se la sfera ha centro $C = (a, b, c)$ e $P_0 = (x_0, y_0, z_0)$, allora il piano π ha equazione

$$(x_0 - a)(x - x_0) + (y_0 - b)(y - y_0) + (z_0 - c)(z - z_0) = 0,$$

grazie a (13.3).

Cilindri e coni di base una conica, come pure le sfere, sono esempi di *quadriche*: superfici di \mathbb{R}^3 descritte da un'equazione di secondo grado. Torneremo a studiare le quadriche nel Capitolo 16.

Autovalori e autovettori

Questo capitolo è dedicato a uno degli argomenti più importanti, sia dal punto di vista teorico che per le applicazioni, dell'Algebra Lineare: autovettori e autovalori. Un autovettore di un endomorfismo T di uno spazio vettoriale V è un vettore $v \in V$ non nullo tale che $T(v)$ sia un multiplo di v ; il coefficiente di proporzionalità si chiama autovalore. Nelle applicazioni, autovettori e autovalori compaiono come assi preferenziali di rotazione, frequenze di risonanza, direzioni di maggior sforzo, e così via; nella matematica pura sono uno dei mezzi più utili per lo studio degli endomorfismi degli spazi vettoriali.

Dopo aver esaminato molteplici esempi, definiremo il polinomio caratteristico di un endomorfismo, che è lo strumento principale per la determinazione di autovalori e autovettori. Poi studieremo il problema (che ritroveremo anche nel prossimo capitolo) dell'esistenza di una base di autovettori per un dato endomorfismo, dimostrando un criterio necessario e sufficiente in termini della molteplicità algebrica e geometrica degli autovalori. Nei Complementi introdurremo il polinomio minimo di un endomorfismo e dimostreremo un algoritmo esplicito per scoprire se un endomorfismo ammette una base di autovettori senza bisogno di determinare gli autovalori, algoritmo basato sul Teorema di Sturm per l'individuazione delle radici reali di un polinomio.

14.1 Definizioni ed esempi

Per giustificare l'importanza dell'argomento a cui è dedicato questo capitolo vediamo un modello di situazione fisica (e ingegneristica) in cui compare naturalmente.

ESEMPIO 14.1 Consideriamo un corpo statico Σ sottoposto a delle forze; può essere un ponte, o una macchina, o una colonna, o qualsiasi altra cosa. Vogliamo capire se c'è il rischio che il corpo, a causa dello stress a cui è sottoposto, si rompa.

Prendiamo un punto $P_0 \in \Sigma$; cominciamo col rappresentare le forze a cui è soggetto il corpo in P_0 tenendo presente che, siccome abbiamo a che fare con un corpo solido, ciò che accade in P_0 è influenzato da quanto succede nelle immediate vicinanze. Un possibile modo di procedere consiste nel considerare le forze che agiscono sulle facce di un cubetto infinitesimo centrato in P_0 . Chiamiamo $T(e_j)$ la somma di

tutte le forze applicate sulla faccia ortogonale alla direzione e_j , per $j = 1, 2, 3$. Estendendo per linearità (cioè applicando la Proposizione 5.2) otteniamo un'applicazione lineare $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ che descrive che tipo di forze agiscono sul nostro materiale in P_0 : il vettore $T(v)$ è la somma delle forze che agiscono su una superficie piana infinitesima passante per P_0 e ortogonale a v . A questo punto possiamo definire lo sforzo F a cui il corpo è soggetto in P_0 nella direzione $v \neq O$ ponendo

$$F(v) = \frac{\langle T(v), v \rangle}{\langle v, v \rangle}.$$

Infatti, la componente di $T(v)$ parallela alla superficie ortogonale alla direzione v non crea alcuno stress su questa superficie (si limita a traslarla su se stessa, non la tira verso l'esterno); nello sforzo dunque conta solo la componente di $T(v)$ ortogonale a questa superficie, cioè la proiezione ortogonale di $T(v)$ lungo v , proiezione che ha proprio modulo $F(v)$.

La direzione in cui il nostro materiale si potrebbe rompere è ovviamente quella di massimo sforzo; vogliamo vedere che relazione ha col nostro endomorfismo T . Prima di tutto, notiamo che $F(\lambda v) = F(v)$ per ogni $\lambda \in \mathbb{R}^*$ (esercizio). Come conseguenza, questa direzione di massimo sforzo esiste: infatti, per trovarla possiamo limitarci a considerare i vettori v di lunghezza uno, che è un insieme chiuso e limitato (un insieme compatto). Ma F è una funzione continua (è un rapporto di polinomi), e quindi ammette massimo su un insieme chiuso e limitato (vedi il Teorema 1.7 e il Lemma 11C.12).

Sia allora v_0 questa direzione di massimo sforzo, con $\|v_0\| = 1$. Prendiamo un'altra direzione $v \neq O$, e consideriamo la funzione $f(t) = F(v_0 + tv)$. Chiaramente, la f ha un massimo in 0; quindi $f'(0) = 0$. Calcoliamo questa derivata; prima di tutto

$$f(t) = \frac{\langle T(v_0 + tv), v_0 + tv \rangle}{\langle v_0 + tv, v_0 + tv \rangle} = \frac{\langle T(v_0), v_0 \rangle + t[\langle T(v), v_0 \rangle + \langle T(v_0), v \rangle] + t^2 \langle T(v), v \rangle}{\langle v_0, v_0 \rangle + 2t \langle v_0, v \rangle + t^2 \langle v, v \rangle}.$$

Ora si può dimostrare che, siccome il nostro corpo è fermo, l'endomorfismo T è simmetrico¹; dunque abbiamo

$$f(t) = \frac{\langle T(v_0), v_0 \rangle + 2t \langle T(v_0), v \rangle + t^2 \langle T(v), v \rangle}{\langle v_0, v_0 \rangle + 2t \langle v_0, v \rangle + t^2 \langle v, v \rangle}.$$

Derivando e ponendo $t = 0$ otteniamo

$$0 = f'(0) = 2 \frac{\langle T(v_0), v \rangle \langle v_0, v_0 \rangle - \langle v_0, v \rangle \langle T(v_0), v_0 \rangle}{\langle v_0, v_0 \rangle^2}.$$

Dunque, ricordando che $\langle v_0, v_0 \rangle = 1$ e indicando con $\lambda = \langle T(v_0), v_0 \rangle = F(v_0)$ lo sforzo in v_0 , troviamo che

$$\langle T(v_0), v \rangle = \lambda \langle v_0, v \rangle.$$

¹ L'idea è che se non lo fosse allora il corpo dovrebbe cominciare a ruotare intorno a P_0

Siccome questo dev'essere vero per ogni $v \in \mathbb{R}^3$ ne segue che

$$T(v_0) = \lambda v_0. \quad (14.1)$$

Riassumendo, abbiamo trovato che la direzione di massimo sforzo in P_0 deve soddisfare la (14.1), dove λ è proprio lo sforzo che il nostro corpo subisce nella direzione v_0 . Dunque per scoprire se il nostro materiale sottoposto alle sollecitazioni riassunte dall'endomorfismo T si romperà dobbiamo trovare tutti i vettori che soddisfano (14.1), e vedere se la resistenza del materiale in P_0 è maggiore del valore dello sforzo lungo quei vettori. Se lo è, non ci sono problemi; se non lo è, meglio cambiare materiale.

A questo punto le seguenti definizioni divengono piuttosto naturali:

Definizione 14.1 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale V . Un vettore $v_0 \neq O$ di V è un *autovettore* di T relativo all'*autovalore* λ se si ha

$$T(v_0) = \lambda v_0. \quad (14.2)$$

L'insieme degli autovalori di T — cioè degli scalari λ per cui esista un vettore non nullo v_0 soddisfacente (14.2) — è lo *spettro* di T ; ed è indicato con $\text{sp}(T)$. Se $\lambda \in \text{sp}(T)$, l'insieme

$$V_\lambda = \{v \in V \mid T(v) = \lambda v\}$$

è detto *autospazio* di T relativo all'autovalore λ , ed è un sottospazio di V . (Esercizio 14.1). Ovviamente, λ è un autovalore se e solo se $V_\lambda \neq \{O\}$.

Osservazione 14.1 Dalla definizione è chiaro che $V_0 = \text{Ker } T$, cioè lo zero è un autovalore di T se e solo se T è singolare. Più in generale, $V_\lambda = \text{Ker}(T - \lambda \text{id})$, per cui λ è un autovalore se e solo se l'endomorfismo $T - \lambda \text{id}$ è singolare.

Con questa terminologia l'Esempio 14.1 può venire riassunto dicendo che la direzione di massimo sforzo è esattamente un autovettore dell'endomorfismo T , e lo sforzo è il relativo autovalore. Dunque dovrebbe essere chiara l'importanza del seguente problema: *dato un endomorfismo $T: V \rightarrow V$, trovare i suoi autovettori e relativi autovalori.*

Prima di affrontare il problema in generale vediamo alcuni esempi.

ESEMPIO 14.2 Sia $V = \mathbb{R}^3$ e $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ la rotazione di angolo π attorno all'asse z . Allora $T(e_1) = -e_1$, $T(e_2) = -e_2$ e $T(e_3) = e_3$; quindi T è rappresentato dalla matrice

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

In particolare, T ha almeno due autovalori, 1 e -1 , e i relativi autospazi contengono per lo meno i vettori della base canonica; per l'esattezza, $V_1 \supseteq \text{Span}(e_3)$ e $V_{-1} \supseteq \text{Span}(e_1, e_2)$. Per scoprire se ce ne sono altri, studiamo il sistema $T(x) = \lambda x$, dipendente dal parametro λ ; vogliamo vedere se esistono dei valori del parametro

per cui questo sistema ammette una soluzione non nulla. In coordinate, dobbiamo studiare il sistema

$$\begin{cases} -x_1 = \lambda x_1, \\ -x_2 = \lambda x_2, \\ x_3 = \lambda x_3, \end{cases} \quad \text{cioè} \quad \begin{cases} (\lambda + 1)x_1 = 0, \\ (\lambda + 1)x_2 = 0, \\ (\lambda - 1)x_3 = 0. \end{cases} \quad (14.3)$$

Quindi se $\lambda \neq \pm 1$ l'unica soluzione possibile è $x = O$, per cui un tale λ non è un autovalore. Di conseguenza, $\text{sp}(T) = \{1, -1\}$; inoltre risolvendo (14.3) per $\lambda = 1$ o $\lambda = -1$ troviamo esattamente che $V_1 = \text{Span}(e_3)$ e $V_{-1} = \text{Span}(e_1, e_2)$, in quanto V_λ è proprio lo spazio delle soluzioni del sistema $T(x) = \lambda x$. Abbiamo dunque recuperato tutti gli autovettori e autovalori di T ; notiamo in particolare che la base canonica è una base di autovettori per il nostro endomorfismo T .

ESEMPIO 14.3 Sia ancora $V = \mathbb{R}^3$, ma stavolta $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ è la rotazione di angolo $\pi/2$ attorno all'asse z . Questa volta $T(e_1) = e_2$, $T(e_2) = -e_1$ e $T(e_3) = e_3$, per cui T è rappresentato dalla matrice

$$\begin{vmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix}. \quad (14.4)$$

Stavolta vi è un solo autovalore evidente, $\lambda = 1$ con autospazio $V_1 \supseteq \text{Span}(e_3)$. Vediamo se ce ne sono altri, studiando il sistema $T(x) = \lambda x$. In coordinate abbiamo

$$\begin{cases} -x_2 = \lambda x_1, \\ x_1 = \lambda x_2, \\ x_3 = \lambda x_3, \end{cases} \quad \text{cioè} \quad \begin{cases} x_1 = \lambda x_2, \\ (1 + \lambda^2)x_2 = 0, \\ (1 - \lambda)x_3 = 0. \end{cases} \quad (14.5)$$

Siccome siamo su \mathbb{R} , il numero $1 + \lambda^2$ è sempre positivo; quindi (14.5) forza $x_1 = x_2 = 0$ per qualunque valore di λ . Se poi $\lambda \neq 1$, la terza equazione implica anche $x_3 = 0$, e $\lambda \notin \text{sp}(T)$; se invece $\lambda = 1$ la terza equazione è soddisfatta per qualunque valore di x_3 . Quindi $\text{sp}(T) = \{1\}$ e $V_1 = \text{Span}(e_3)$. Dunque T ha un solo autovalore, e l'autospazio relativo ha dimensione 1; in particolare, non esiste una base di \mathbb{R}^3 composta da autovettori di T .

ESEMPIO 14.4 Ora consideriamo lo stesso endomorfismo T , ma sui complessi. In altri termini, prendiamo $V = \mathbb{C}^3$, e sia $T: \mathbb{C}^3 \rightarrow \mathbb{C}^3$ l'endomorfismo rappresentato ancora dalla matrice (14.4). Ovviamente, $1 \in \text{sp}(T)$ e $V_1 \supseteq \text{Span}(e_3)$ esattamente come prima; ma nel momento in cui andiamo a risolvere il sistema (14.5) le cose cambiano. Infatti, se λ è un numero complesso non è più detto che $1 + \lambda^2$ sia sempre diverso da zero; per $\lambda = \pm i$ abbiamo $1 + \lambda^2 = 0$. Dunque, se $\lambda \neq 1, \pm i$ il sistema (14.5) continua ad ammettere soltanto la soluzione $x = O$, e se $\lambda = 1$ otteniamo ancora semplicemente $V_1 = \text{Span}(e_3)$. Ma se $\lambda = \pm i$ la seconda equazione è automaticamente soddisfatta, la terza fornisce $x_3 = 0$ e la prima diventa $x_1 = \pm i x_2$. Dunque esistono

soluzioni non nulle: per l'esattezza,

$$V_i = \text{Span} \left(\begin{vmatrix} i \\ 1 \\ 0 \end{vmatrix} \right), \quad V_{-i} = \text{Span} \left(\begin{vmatrix} -i \\ 1 \\ 0 \end{vmatrix} \right).$$

Di conseguenza, $\text{sp}(T) = \{1, i, -i\}$, e \mathbb{C}^3 ammette una base composta da autovettori di T : la base

$$\mathcal{B} = \left\{ \begin{vmatrix} i \\ 1 \\ 0 \end{vmatrix}, \begin{vmatrix} -i \\ 1 \\ 0 \end{vmatrix}, \begin{vmatrix} 0 \\ 0 \\ 1 \end{vmatrix} \right\}.$$

Rispetto a \mathcal{B} , l'endomorfismo T è rappresentato dalla matrice diagonale

$$\begin{vmatrix} i & 0 & 0 \\ 0 & -i & 0 \\ 0 & 0 & 1 \end{vmatrix}.$$

Osservazione 14.2 La differenza fra questi due esempi è uno dei motivi che ci costringono a lavorare anche con spazi vettoriali complessi. Mentre un endomorfismo su uno spazio vettoriale complesso di dimensione n ammette sempre esattamente n autovalori, contati con le relative molteplicità², in spazi vettoriali reali questo non è sempre vero: possono mancare degli autovalori. Il problema è che, come vedremo fra poco, gli autovalori si ottengono come radici di un polinomio di grado n ; sui complessi, grazie al Teorema fondamentale dell'algebra, tutti i polinomi di grado n hanno, contando le molteplicità, esattamente n radici, mentre sui reali può non essere così (pensa al polinomio $p(t) = t^2 + 1$). Questo rende la teoria nel caso reale sensibilmente più complicata: bisogna iniziare a distinguere fra endomorfismi con tutti gli autovalori reali ed endomorfismi con anche autovalori complessi. Allora tanto vale lavorare direttamente sui complessi, dove tale distinzione svanisce, e poi applicare i risultati al caso particolare in cui tutti i numeri complessi coinvolti sono reali — e questo è quanto faremo.

Osservazione 14.3 Un'altra differenza evidente negli ultimi due esempi è che nel secondo caso esisteva una base di autovettori, mentre nel primo no. Si potrebbe pensare che sui complessi basi di autovettori esistano sempre: questo sarebbe particolarmente piacevole, in quanto rispetto a una base di autovettori un endomorfismo è rappresentato da una matrice molto semplice, addirittura diagonale. Sfortunatamente, questo non è vero: come mostrato nel prossimo Esempio, anche sui complessi ci sono endomorfismi che non ammettono basi di autovettori. Rimane comunque vero che sui complessi è sempre possibile trovare basi rispetto a cui l'endomorfismo si esprime con matrici semplici (nella fattispecie, triangolari superiori), mentre sui reali questo non accade sempre. Ne ripareremo.

² Nell'Esempio 14.2 l'autovalore -1 appariva con molteplicità 2, come segue dal fatto che l'autospazio relativo aveva dimensione 2. Invece, nell'Esempio 14.3 l'autovalore 1 aveva molteplicità 1. Daremo fra breve la definizione precisa di molteplicità di un autovalore.

ESEMPIO 14.5 Prendiamo $V = \mathbb{R}^2$ e sia $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ rappresentato dalla matrice

$$\begin{vmatrix} -5 & -9 \\ 4 & 7 \end{vmatrix}.$$

Questa volta non ci sono autovettori o autovalori evidenti, per cui consideriamo subito il sistema $T(x) = \lambda x$. Questo sistema è equivalente al sistema $(T - \lambda I_2)x = O$, che a sua volta ammette soluzioni diverse dal vettore nullo se e solo se la matrice $T - \lambda I_2$ è singolare, cioè se e solo se $\det(T - \lambda I_2) = 0$. Ora,

$$\det(T - \lambda I_2) = \det \begin{vmatrix} -5 - \lambda & -9 \\ 4 & 7 - \lambda \end{vmatrix} = (-5 - \lambda)(7 - \lambda) - 4(-9) = \lambda^2 - 2\lambda + 1 = (\lambda - 1)^2.$$

Dunque $\det(T - \lambda I_2) = 0$ se e solo se $\lambda = 1$, per cui $\text{sp}(T) = \{1\}$. Per trovare l'autospazio dobbiamo risolvere il sistema $T(x) = x$, cioè

$$\begin{cases} -5x_1 - 9x_2 = x_1, \\ 4x_1 + 7x_2 = x_2, \end{cases} \quad \text{ovvero} \quad \begin{cases} 6x_1 = -9x_2, \\ 4x_1 = -6x_2, \end{cases}$$

che ha come soluzione

$$V_1 = \text{Span} \left(\begin{pmatrix} -3 \\ 2 \end{pmatrix} \right).$$

Dunque non esiste una base di autovettori. E questo accade anche considerando T come endomorfismo su \mathbb{C}^2 . Infatti, in nessun momento abbiamo utilizzato il fatto che λ potesse essere reale: la stessa dimostrazione funziona anche sui complessi. Infine notiamo che rispetto alla base $\left\{ \begin{pmatrix} -3 \\ 2 \end{pmatrix}, e_2 \right\}$ l'endomorfismo T è rappresentato dalla matrice

$$\begin{vmatrix} 1 & 3 \\ 0 & 1 \end{vmatrix},$$

che è triangolare superiore.

Dunque un'altra questione che questi esempi invitano a considerare è: *quando esiste una base di autovettori per un dato endomorfismo T ?* Questo problema è legato a quello, già accennato alla fine del Paragrafo 8.2, di trovare una base in cui l'endomorfismo sia rappresentato da una matrice semplice, come mostra la seguente

Proposizione 14.1 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale V , e \mathcal{B} una base di V . Allora \mathcal{B} è composta da autovettori di T se e solo se la matrice che rappresenta T rispetto a \mathcal{B} è diagonale.

Dimostrazione. Sia $\mathcal{B} = \{v_1, \dots, v_n\}$ una base composta da autovettori di T . Allora $T(v_j) = \lambda_j v_j$ per $j = 1, \dots, n$, e quindi la matrice che rappresenta T rispetto a questa base è la matrice diagonale di elementi $\lambda_1, \dots, \lambda_n$.

Viceversa, T è rappresentato da una matrice diagonale di elementi $\lambda_1, \dots, \lambda_n$ rispetto alla base $\mathcal{B} = \{v_1, \dots, v_n\}$ esattamente quando $T(v_j) = \lambda_j v_j$ per $j = 1, \dots, n$, e quindi \mathcal{B} è composta da autovettori di T . \square

Corollario 14.2 Sia A una matrice quadrata di ordine n . Allora esiste una base di autovettori per L_A se e solo se A è simile a una matrice diagonale se e solo se la classe di similitudine \mathcal{O}_A contiene una matrice diagonale.

Dimostrazione. Segue dalle Proposizioni 14.1 e 8.2. \square

Definizione 14.2 Un endomorfismo $T: V \rightarrow V$ è *diagonalizzabile* se esiste una base di V composta da autovettori di T .

Dunque un endomorfismo è diagonalizzabile se e solo se esiste una base rispetto a cui è rappresentato da una matrice diagonale. Buona parte di questo e del prossimo capitolo sarà dedicato a identificare classi di endomorfismi diagonalizzabili (gli endomorfismi autoaggiunti, per esempio, o quelli unitari — ma non quelli ortogonali), anche se non daremo una soluzione completa del problema di quali endomorfismi sono diagonalizzabili (vedi però l'Osservazione 14.8 e i Complementi a questo capitolo, e anche il Teorema 15.3 e relativi corollari). In compenso, faremo vedere che almeno sui complessi tutti gli endomorfismi possono venire rappresentati da matrici triangolari superiori (Teorema 15.10).

Definizione 14.3 Un endomorfismo $T: V \rightarrow V$ è *triangolabile* se esiste una base \mathcal{B} di V rispetto a cui T è rappresentato da una matrice triangolare superiore; diremo che la base \mathcal{B} *triangolarizza* T .

A questo punto è chiaro che abbiamo bisogno di maggiori informazioni su autovalori e autovettori, e soprattutto su come trovarli. Quindi concludiamo questa introduzione e mettiamoci al lavoro.

14.2 Il polinomio caratteristico

Nell'Esempio 14.5 abbiamo trovato gli autovalori di un endomorfismo come radici dell'equazione $\det(T - \lambda I_2) = 0$. Questo è esattamente il modo con cui si recuperano in generale:

Teorema 14.3 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale V sul campo \mathbb{K} . Fissiamo una base $\mathcal{B} = \{v_1, \dots, v_n\}$ di V , e sia $A \in M_{n,n}(\mathbb{K})$ la matrice che rappresenta T rispetto a \mathcal{B} . Allora si ha:

(i) la funzione $p_T: \mathbb{K} \rightarrow \mathbb{K}$ data da

$$p_T(\lambda) = \det(A - \lambda I_n).$$

non dipende dalla base \mathcal{B} scelta;

(ii) p_T è un polinomio di grado n in cui il coefficiente direttivo è $(-1)^n$, il termine noto è $\det A$ e il coefficiente di λ^{n-1} è $(-1)^{n-1} \text{tr } A$;

(iii) $\lambda_0 \in \mathbb{K}$ è un autovalore di T se e solo se $p_T(\lambda_0) = 0$.

Osservazione 14.4 Come già accadeva nella prima parte di questo testo, buona parte di quanto faremo vale non solo su \mathbb{R} o su \mathbb{C} ma anche su altri campi³, quali per esempio \mathbb{Q} . In pratica, però, in questo e nel prossimo capitolo frasi del tipo “sul campo \mathbb{K} ” saranno da intendersi come abbreviazioni di “su \mathbb{R} o su \mathbb{C} ”.

Dimostrazione. (i) Supponiamo che la matrice A' rappresenti T in un'altra base. Per la Proposizione 8.2 esiste una matrice quadrata invertibile B tale che $A' = B^{-1}AB$. Dunque

$$\begin{aligned} \det(A' - \lambda I_n) &= \det(B^{-1}AB - \lambda I_n) = \det(B^{-1}AB - \lambda B^{-1}B) \\ &= \det(B^{-1}(A - \lambda I_n)B) = \det(B^{-1}) \det(A - \lambda I_n) \det(B) \\ &= (\det B)^{-1} (\det B) \det(A - \lambda I_n) = \det(A - \lambda I_n), \end{aligned}$$

e p_T non dipende dalla base scelta.

(ii) Procediamo per induzione su n . Per $n = 2$ si ha

$$\begin{aligned} p_T(\lambda) &= \det \begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} = (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} \\ &= \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21}, \end{aligned}$$

e ci siamo. Supponiamolo vero per $n - 1$, e dimostriamolo per n . Prima di tutto, il determinante è ottenuto facendo somme e prodotti degli elementi della matrice; quindi $p_T(\lambda)$ contiene solo somme e prodotti di costanti e di λ , per cui è un polinomio in λ . Inoltre, siccome nel calcolo del determinante ogni elemento della matrice interviene al più una volta sola in ogni singolo addendo di uno sviluppo di Laplace, e nella matrice $A - \lambda I_n$ il parametro λ compare solo in n elementi, ne segue che $p_T(\lambda)$ ha al più grado n . Ora, possiamo scrivere

$$\begin{aligned} p_T(\lambda) &= \det \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} \\ &= (a_{11} - \lambda) \det \begin{vmatrix} a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} + \text{termini di grado al più } n - 2, \end{aligned}$$

dove abbiamo sviluppato il determinante lungo la prima colonna, e il resto sono determinanti di matrici in cui λ appare soltanto $n - 2$ volte, per cui forniscono polinomi

³ Ma non proprio su tutti: ce ne sono alcuni, detti di “caratteristica 2” (in cui accade che “1+1=0”) dove possono esserci dei problemi.

di grado al più $n - 2$. Per ipotesi induttiva abbiamo allora

$$\begin{aligned} p_T(\lambda) &= (a_{11} - \lambda)[(-1)^{n-1}\lambda^{n-1} + (-1)^{n-2}(a_{22} + \cdots + a_{nn})\lambda^{n-2} + \cdots] + \cdots \\ &= (-1)^n \lambda^n + (-1)^{n-1}(\text{tr } A)\lambda^{n-1} + \cdots, \end{aligned}$$

dove i puntini rappresentano termini di grado minore o uguale a $n - 2$. Dunque abbiamo dimostrato che p_T è un polinomio di grado esattamente n con coefficiente direttivo $(-1)^n$, e con $(-1)^{n-1} \text{tr } A$ come coefficiente di λ^{n-1} . Infine il termine noto di $p_T(\lambda)$ è dato da

$$p_T(0) = \det(A - 0I_n) = \det A.$$

(iii) Lo scalare $\lambda_0 \in \mathbb{K}$ è un autovalore di T se e solo se il sistema $Ax = \lambda_0 x$ ammette una soluzione x diversa dal vettore nullo, se e solo se il sistema $(A - \lambda_0 I_n)x = 0$ ammette una soluzione diversa dal vettore nullo, se e solo se

$$0 = \det(A - \lambda_0 I_n) = p_T(\lambda_0),$$

(Teorema 7.7 e Corollario 9.10). □

Definizione 14.4 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale V sul campo \mathbb{K} . Il polinomio $p_T \in \mathbb{K}[\lambda]$ definito nel Teorema 14.3 si chiama *polinomio caratteristico* dell'endomorfismo T . Ricordando la Definizione 9.3, si vede subito che $p_T(\lambda) = \det(T - \lambda \text{id}_V)$.

Definizione 14.5 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale V . La *traccia* $\text{tr } T$ di T è la traccia della matrice associata a T rispetto a una qualunque base di V . Grazie al Teorema 14.3, la traccia di un endomorfismo è ben definita, in quanto a meno del segno è uguale al coefficiente di λ^{n-1} del polinomio caratteristico di T (vedi anche l'Esercizio 7.25).

Dunque gli autovalori di un endomorfismo T sono esattamente le radici del suo polinomio caratteristico. Questo è il motivo per cui su \mathbb{R} l'endomorfismo T potrebbe non avere tutti gli autovalori necessari: le radici del polinomio caratteristico potrebbero non essere tutte reali. Sui complessi questo non può succedere; infatti, il Teorema fondamentale dell'algebra (Teorema 11.7) ci assicura che un polinomio a coefficienti complessi di grado n ha esattamente n radici, contate con la relativa molteplicità.

Definizione 14.6 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale di dimensione n sul campo \mathbb{K} . Allora diremo che T ha *tutti gli autovalori* in \mathbb{K} se il polinomio caratteristico p_T ha esattamente n radici in \mathbb{K} , contate con la relativa molteplicità (vedi la Definizione 11.8), cioè se esistono $\lambda_1, \dots, \lambda_n \in \mathbb{K}$, non necessariamente distinti, tali che

$$p_T(\lambda) = (\lambda_1 - \lambda) \cdots (\lambda_n - \lambda).$$

Chiaramente, ogni endomorfismo di uno spazio vettoriale complesso ha tutti gli autovalori in \mathbb{C} , mentre sui reali questo non capita sempre (vedi per esempio l'Esercizio 14.2). Tra parentesi, l'aver tutti gli autovalori nel campo è una condizione necessaria per la triangolabilità.

Lemma 14.4 Sia $T: V \rightarrow V$ un endomorfismo triangolabile di uno spazio vettoriale sul campo \mathbb{K} . Allora T ha tutti gli autovalori in \mathbb{K} .

Dimostrazione. Sia \mathcal{B} una base di V che triangolarizza T , e sia

$$A = \begin{pmatrix} \lambda_1 & a_{12} & \cdots & a_{1n} \\ 0 & \lambda_2 & \cdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \lambda_n \end{pmatrix}$$

la matrice triangolare superiore che rappresenta T rispetto a \mathcal{B} . Allora

$$p_T(\lambda) = \det(A - \lambda I_n) = (\lambda_1 - \lambda) \cdots (\lambda_n - \lambda),$$

e T ha tutti gli autovalori in \mathbb{K} . \square

Una conseguenza immediata del Teorema 14.3 è il seguente

Corollario 14.5 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale sul campo \mathbb{K} . Supponiamo che T abbia tutti gli autovalori in \mathbb{K} , e siano $\lambda_1, \dots, \lambda_n$ le radici del polinomio caratteristico di T . Allora

$$\operatorname{tr} T = \lambda_1 + \cdots + \lambda_n \quad \text{e} \quad \det T = \lambda_1 \cdots \lambda_n.$$

Dimostrazione. Per ipotesi possiamo scrivere

$$p_T(\lambda) = (\lambda_1 - \lambda) \cdots (\lambda_n - \lambda); \quad (14.6)$$

in particolare, $\det T = p_T(0) = \lambda_1 \cdots \lambda_n$. Inoltre, è facile dimostrare per induzione su n (esercizio) che (14.6) implica

$$p_T(\lambda) = (-1)^n \lambda^n + (-1)^{n-1} (\lambda_1 + \cdots + \lambda_n) \lambda^{n-1} + \cdots,$$

e il Teorema 14.3.(ii) ci fornisce $\operatorname{tr} T = \lambda_1 + \cdots + \lambda_n$. \square

Osservazione 14.5 Dunque per trovare gli autovalori di un endomorfismo basta calcolare le radici di un polinomio. Questo risultato è molto gradevole dal punto di vista teorico, ma può creare dei problemi sul piano pratico. Infatti, mentre si sa perfettamente come trovare le radici di un polinomio di secondo grado, per gradi più alti la situazione non è così semplice⁴. Comunque, ci sono metodi anche molto efficienti per trovare delle approssimazioni delle radici di un polinomio qualunque, e in molte applicazioni queste approssimazioni sono altrettanto utili del valore esatto.

⁴ Esistono delle formule che esprimono le radici di polinomi di terzo e quarto grado in funzione dei coefficienti, ma si può dimostrare che non esistono formule del genere per i polinomi di grado dal quinto in su.

14.3 Molteplicità

Ora che sappiamo come trovare gli autovalori (basta calcolare le radici del polinomio caratteristico), e quindi i relativi autovettori (basta risolvere il sistema $T(v) = \lambda v$ con λ autovalore), possiamo dedicarci a trovare un criterio di diagonalizzabilità. Cominciamo con la

Proposizione 14.6 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale V sul campo \mathbb{K} , e siano $v_1, \dots, v_k \neq O$ autovettori di T corrispondenti ad autovalori distinti $\lambda_1, \dots, \lambda_k \in \mathbb{K}$. Allora v_1, \dots, v_k sono linearmente indipendenti.

Dimostrazione. Procediamo per induzione su k . Per $k = 1$ non c'è nulla da dimostrare; sia allora vero per $k - 1$, cioè supponiamo che v_1, \dots, v_{k-1} siano linearmente indipendenti, e siano $\alpha_1, \dots, \alpha_k \in \mathbb{K}$ tali che

$$\alpha_1 v_1 + \cdots + \alpha_k v_k = O. \quad (14.7)$$

Applicando T a questa relazione troviamo

$$O = \alpha_1 T(v_1) + \cdots + \alpha_k T(v_k) = \alpha_1 \lambda_1 v_1 + \cdots + \alpha_k \lambda_k v_k. \quad (14.8)$$

La (14.7) ci dice che $\alpha_k v_k = -\alpha_1 v_1 - \cdots - \alpha_{k-1} v_{k-1}$; inserendolo nella (14.8) otteniamo

$$\alpha_1 (\lambda_1 - \lambda_k) v_1 + \cdots + \alpha_{k-1} (\lambda_{k-1} - \lambda_k) v_{k-1} = O.$$

Essendo v_1, \dots, v_{k-1} linearmente indipendenti, questo implica

$$\alpha_1 (\lambda_1 - \lambda_k) = \cdots = \alpha_{k-1} (\lambda_{k-1} - \lambda_k) = 0.$$

Ma tutti i λ_j sono distinti per ipotesi; quindi $\alpha_1 = \cdots = \alpha_{k-1} = 0$. Ritornando in (14.7) otteniamo anche $\alpha_k = 0$, e dunque v_1, \dots, v_k sono linearmente indipendenti. \square

Osservazione 14.6 In particolare, se $v_1, \dots, v_k \in V$ sono autovettori relativi ad autovalori distinti $\lambda_1, \dots, \lambda_k$ di un endomorfismo $T: V \rightarrow V$, e si ha $v_1 + \cdots + v_k = O$, allora $v_1 = \cdots = v_k = O$. Infatti, se anche uno solo dei v_j fosse diverso dal vettore nullo avremmo trovato una relazione di dipendenza lineare fra autovettori relativi ad autovalori distinti, contraddicendo la Proposizione 14.6.

Una prima conseguenza è la seguente:

Corollario 14.7 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale di dimensione n sul campo \mathbb{K} , e supponiamo che T abbia esattamente n autovalori distinti in \mathbb{K} . Allora T è diagonalizzabile.

Dimostrazione. Siano $v_1, \dots, v_n \in V$ degli autovettori relativi agli n autovalori distinti $\lambda_1, \dots, \lambda_n$. Per la Proposizione 14.6, v_1, \dots, v_n sono linearmente indipendenti, e quindi formano una base di V . \square

ESEMPIO 14.6 Consideriamo l'endomorfismo T rappresentato dalla matrice

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

Allora $p_T(\lambda) = \lambda^2 + 1$, per cui T non ha autovalori reali e non è neppure triangolabile su \mathbb{R} . Sui complessi, invece, ha esattamente due autovalori distinti, i e $-i$, per cui è diagonalizzabile su \mathbb{C} .

Ma cosa accade quando le radici del polinomio caratteristico non sono tutte distinte? Cominciamo con una definizione:

Definizione 14.7 Sia $\lambda_0 \in \text{sp}(T)$ un autovalore di un endomorfismo $T: V \rightarrow V$. Diremo *molteplicità algebrica* di λ_0 la sua molteplicità come radice del polinomio caratteristico $p_T(\lambda)$; *molteplicità geometrica* la dimensione del relativo autospazio V_{λ_0} .

Osservazione 14.7 Per calcolare la molteplicità algebrica dell'autovalore λ_0 di un endomorfismo T bisogna scoprire quante volte λ_0 è radice del polinomio caratteristico. Per calcolare invece la molteplicità geometrica, basta trovare il rango di $T - \lambda_0 \text{id}$; infatti l'autospazio V_{λ_0} è il nucleo di $T - \lambda_0 \text{id}$, per cui

$$\dim V_{\lambda_0} = \dim V - \text{rg}(T - \lambda_0 \text{id}).$$

ESEMPIO 14.7 Sia $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ l'endomorfismo rappresentato dalla matrice

$$\begin{pmatrix} 3 & 1 & 1 \\ 2 & 4 & 2 \\ 3 & 3 & 5 \end{pmatrix};$$

vogliamo calcolare molteplicità algebrica e geometrica dei suoi autovalori. Prima di tutto,

$$p_T(\lambda) = \det(T - \lambda I_3) = -\lambda^3 + 12\lambda^2 - 36\lambda + 32.$$

Si verifica subito che $p_T(2) = 0$; quindi

$$p_T(\lambda) = (2 - \lambda)(\lambda^2 - 10\lambda + 16) = (2 - \lambda)^2(8 - \lambda),$$

e $\text{sp}(T) = \{2, 8\}$, dove 2 ha molteplicità algebrica 2 e 8 ha molteplicità algebrica 1. Per calcolare le molteplicità geometriche, troviamo

$$\text{rg}(T - 2I_3) = \text{rg} \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{pmatrix} = 1,$$

e

$$\text{rg}(T - 8I_3) = \text{rg} \begin{pmatrix} -5 & 1 & 1 \\ 2 & -4 & 2 \\ 3 & 3 & -3 \end{pmatrix} = 2,$$

per cui 2 ha molteplicità geometrica 2, e 8 ha molteplicità geometrica 1.

Le molteplicità geometriche e algebriche di un autovalore non sono completamente scollegate:

Proposizione 14.8 La molteplicità algebrica di un autovalore è sempre maggiore o uguale alla sua molteplicità geometrica.

Dimostrazione. Sia $\lambda_0 \in \text{sp}(T)$ un autovalore di un endomorfismo T di uno spazio vettoriale V sul campo \mathbb{K} , e sia h la molteplicità geometrica di λ_0 . Sia $\{v_1, \dots, v_h\}$ una base dell'autospazio V_{λ_0} , e completiamola a una base $\{v_1, \dots, v_h, v'_{h+1}, \dots, v'_n\}$ di V . Rispetto a questa base, l'endomorfismo T è rappresentato da una matrice della forma

$$A = \left(\begin{array}{c|c} \lambda_0 I_h & B \\ \hline O & C \end{array} \right),$$

dove $B \in M_{h, n-h}(\mathbb{K})$ e $C \in M_{n-h, n-h}(\mathbb{K})$. Dunque

$$A - \lambda I_n = \left(\begin{array}{c|c} (\lambda_0 - \lambda) I_h & B \\ \hline O & C - \lambda I_{n-h} \end{array} \right),$$

per cui sviluppando il determinante rispetto alle prime h colonne troviamo

$$p_T(\lambda) = \det(A - \lambda I_n) = (\lambda_0 - \lambda)^h p_C(\lambda).$$

Siccome la molteplicità algebrica è il più grande m per cui $(\lambda_0 - \lambda)^m$ può dividere $p_T(\lambda)$ ne segue che la molteplicità geometrica è sempre minore o uguale alla molteplicità algebrica. \square

E ora possiamo enunciare (e dimostrare) un criterio di diagonalizzabilità:

Teorema 14.9 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale di dimensione finita sul campo \mathbb{K} . Supponiamo che $\text{sp}(T) = \{\lambda_1, \dots, \lambda_h\}$, e indichiamo con μ_j (rispettivamente, ν_j) la molteplicità algebrica (rispettivamente, geometrica) di λ_j , per $j = 1, \dots, h$. Allora le seguenti affermazioni sono equivalenti:

- (i) T è diagonalizzabile;
- (ii) ogni elemento di V si scrive come somma di autovettori di T relativi ad autovalori distinti;
- (iii) $V = V_{\lambda_1} \oplus \dots \oplus V_{\lambda_h}$;
- (iv) T ha tutti gli autovalori in \mathbb{K} e la molteplicità geometrica di ciascun autovalore coincide con la sua molteplicità algebrica, cioè $\mu_j = \nu_j$ per $j = 1, \dots, h$;
- (v) $\nu_1 + \dots + \nu_h = \dim V$.

Dimostrazione. (i) \implies (ii) Per ipotesi, ogni vettore $v \in V$ si scrive come combinazione lineare di una base di autovettori. Raggruppando i termini corrispondenti ad autovettori relativi a uno stesso autovalore otteniamo v come somma di autovettori relativi ad autovalori distinti.

(ii) \implies (iii) Essendo $\lambda_1, \dots, \lambda_h$ gli autovalori distinti di T , l'ipotesi è che ogni elemento di V si scrive come somma di elementi di $V_{\lambda_1}, \dots, V_{\lambda_h}$; dobbiamo solo dimostrare che si scrive in modo unico. Infatti, se $v \in V$ è tale che

$$v_1 + \dots + v_h = v = v'_1 + \dots + v'_h$$

con $v_j, v'_j \in V_{\lambda_j}$ per $j = 1, \dots, h$, abbiamo

$$(v_1 - v'_1) + \dots + (v_h - v'_h) = O$$

e l'Osservazione 14.6 implica $v_1 - v'_1 = \dots = v_h - v'_h = O$, come voluto.

(iii) \implies (iv) Grazie alla Proposizione 4.19.(ii), dall'ipotesi possiamo dedurre che $\dim V = \nu_1 + \dots + \nu_h$. Inoltre $\dim V \geq \mu_1 + \dots + \mu_h$ in quanto quest'ultima somma è il numero di radici in \mathbb{K} del polinomio caratteristico di T , numero che al più è uguale al grado di p_T , cioè a $\dim V$. Dunque

$$\nu_1 + \dots + \nu_h \geq \mu_1 + \dots + \mu_h. \quad (14.9)$$

Ma la Proposizione 14.8 ci dice che $\nu_j \leq \mu_j$ per $j = 1, \dots, h$; quindi (14.9) può essere valida solo se $\nu_j = \mu_j$ per ogni $j = 1, \dots, h$. Di conseguenza abbiamo anche che $\mu_1 + \dots + \mu_h = \dim V$, per cui T ha tutti gli autovalori in \mathbb{K} .

(iv) \implies (v) Siccome T ha tutti gli autovalori in \mathbb{K} segue che $\mu_1 + \dots + \mu_h = \dim V$, e ci siamo.

(v) \implies (i) Sia $\mathcal{B}^j = \{v_1^j, \dots, v_{\nu_j}^j\}$ una base di V_{λ_j} per $j = 1, \dots, h$. Per ipotesi, l'insieme $\mathcal{B} = \mathcal{B}^1 \cup \dots \cup \mathcal{B}^h$ contiene esattamente $\dim V$ elementi; quindi se dimostriamo che sono linearmente indipendenti abbiamo trovato una base di V composta da autovettori di T . Scriviamo una relazione di dipendenza lineare fra gli elementi di \mathcal{B} , cioè scriviamo

$$\alpha_1^1 v_1^1 + \dots + \alpha_{\nu_1}^1 v_{\nu_1}^1 + \dots + \alpha_1^h v_1^h + \dots + \alpha_{\nu_h}^h v_{\nu_h}^h = O. \quad (14.10)$$

Ponendo

$$w_j = \alpha_1^j v_1^j + \dots + \alpha_{\nu_j}^j v_{\nu_j}^j \in V_{\lambda_j},$$

possiamo riscrivere la (14.10) come

$$w_1 + \dots + w_h = O.$$

Ma allora l'Osservazione 14.6 ci dice che $w_1 = \dots = w_h = O$, e quindi tutti gli α_i^j sono zero. \square

ESEMPIO 14.8 Dunque l'endomorfismo dell'Esempio 14.7 è diagonalizzabile. Una base di autovettori, trovata risolvendo i sistemi $T(v) = 2v$ e $T(v) = 8v$, è

$$\left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right\}.$$

Osservazione 14.8 Il criterio dato nel Teorema 14.9 per decidere quando un endomorfismo è diagonalizzabile è abbastanza efficace ma non è perfetto, in quanto richiede il calcolo esplicito degli autovalori⁵. Vi è un altro criterio che non ha questo difetto, ma è di dimostrazione alquanto più complicata. Sia $T: V \rightarrow V$ un endomorfismo. Indichiamo con T^k la composizione di T con se stesso k volte; per esempio, $T^3 = T \circ T \circ T$ e così via. Per ogni polinomio $p \in \mathbb{K}[t]$ possiamo definire un endomorfismo $p(T): V \rightarrow V$ in questo modo: se $p(t) = a_d t^d + a_{d-1} t^{d-1} + \dots + a_0$, allora

$$p(T) = a_d T^d + a_{d-1} T^{d-1} + \dots + a_1 T + a_0 \text{id}_V.$$

Chiaramente, può succedere che $p(T)$ sia l'applicazione nulla (vedi l'Esercizio 14.11). Si può dimostrare che, a meno di costanti moltiplicative, esiste un unico polinomio p_0 (chiamato *polinomio minimo* di T) di grado minimo tale che $p_0(T) = O$; è esplicitamente calcolabile, e risulta essere un divisore del polinomio caratteristico di T . Il criterio summenzionato è il seguente: T è diagonalizzabile su \mathbb{K} se e solo se tutte le radici del polinomio minimo sono in \mathbb{K} e hanno molteplicità 1. Siccome un polinomio ha tutte le radici in \mathbb{K} e di molteplicità 1 se e solo se lui e la sua derivata non hanno fattori comuni, ed esiste un metodo (basato sul calcolo con l'algoritmo di Euclide del massimo comun divisore) per stabilire se due polinomi hanno fattori comuni senza bisogno di calcolarne le radici, questo criterio ci permette di stabilire quando un endomorfismo è diagonalizzabile senza bisogno di calcolarne gli autovalori. Lo discuteremo soltanto nei Complementi a questo capitolo sia perché è di dimostrazione non semplice sia perché nel prossimo capitolo daremo un criterio necessario e sufficiente molto efficace per l'esistenza di una base ortonormale di autovettori, criterio che basterà abbondantemente per i nostri scopi.

Esercizi

14.1 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale V , e $\lambda \in \text{sp}(T)$. Dimostra che V_λ è un sottospazio vettoriale di V . Se $V_\lambda = V$, che cosa si può dire su T ?

14.2 Sia $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ l'endomorfismo rappresentato dalla matrice

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

Dimostra che non esiste nessuna base di \mathbb{R}^2 rispetto a cui T è rappresentato da una matrice triangolare superiore. Dimostra anche che se consideriamo T come un endomorfismo di \mathbb{C}^2 allora esiste una base di \mathbb{C}^2 composta da autovettori di T .

⁵ Nota che conoscere semplicemente il valore approssimato $\tilde{\lambda}$ di un autovalore λ_0 non basta: anche se $\tilde{\lambda}$ è vicinissimo a λ_0 , l'endomorfismo $T - \tilde{\lambda} \text{id}$ rimane invertibile, e quindi non ci dice nulla sulla molteplicità geometrica di λ_0 .

14.3 Trova autovalori e autovettori degli endomorfismi rappresentati, sui reali e sui complessi, dalle matrici

$$\begin{array}{ccccccc} \begin{vmatrix} 2 & 7 \\ 7 & 24 \end{vmatrix}, & \begin{vmatrix} -11 & -24 & -18 \\ 8 & 17 & 12 \\ -2 & -4 & -2 \end{vmatrix}, & \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix}, & \begin{vmatrix} 2 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{vmatrix}, & \begin{vmatrix} 1 & -2 & -4 \\ 0 & 0 & -2 \\ 0 & 1 & 3 \end{vmatrix}, \\ \begin{vmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{vmatrix}, & \begin{vmatrix} 1 & -2 & -5 \\ 0 & 0 & -4 \\ 0 & 1 & 3 \end{vmatrix}, & \begin{vmatrix} 1 & 0 & 0 & 1 \\ 0 & 4 & -2 & 0 \\ 0 & 3 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}, & \begin{vmatrix} 4 & 0 & -2 & 0 \\ 0 & 2 & 0 & 5 \\ 3 & 0 & -1 & 0 \\ 0 & -1 & 0 & -2 \end{vmatrix}, & \begin{vmatrix} 0 & -2 & 1 \\ 2 & -5 & 2 \\ -1 & 2 & -2 \end{vmatrix}, \end{array}$$

e determina se sono diagonalizzabili su \mathbb{R} o su \mathbb{C} .

14.4 Sia $A \in M_{n,n}(\mathbb{R})$. Dimostra che A e A^T hanno gli stessi autovalori, ma non necessariamente gli stessi autovettori.

14.5 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale di dimensione n . Una bandiera per T è una bandiera V_0, \dots, V_n di V (vedi l'Esercizio 4.37) tale che $T(V_i) \subseteq V_i$ per $i = 0, \dots, n$. Una base a bandiera per T è una base $\{v_1, \dots, v_n\}$ di V tale che, posto $V_i = \text{Span}(v_1, \dots, v_i)$ per $i = 1, \dots, n$, allora V_1, \dots, V_n è una bandiera per T . Dimostra che T è triangolabile se e solo se esiste una bandiera per T . (Suggerimento: rispetto a una base a bandiera T è rappresentato da una matrice di forma particolare.)

14.6 Sia $T: \mathbb{R}_3[t] \rightarrow \mathbb{R}_3[t]$ l'endomorfismo dato da $T(p) = tp'$, dove p' è la derivata di p . Trova gli autovettori e gli autovalori di T .

14.7 Per ogni $a \in \mathbb{R}$ definiamo $T_a: \mathbb{R}_3[t] \rightarrow \mathbb{R}_3[t]$ ponendo $[T_a(p)](t) = tp'(t-a)$. Determina per quali $a \in \mathbb{R}$ (se ne esistono) l'endomorfismo T_a è diagonalizzabile.

14.8 Sia $T: V \rightarrow V$ un endomorfismo tale che $\text{sp}(T) = \{\lambda_0\}$. Dimostra che T è diagonalizzabile se e solo se $T = \lambda_0 \text{id}_V$.

14.9 Siano $S, T: V \rightarrow V$ due endomorfismi. Dimostra che $\text{sp}(S \circ T) = \text{sp}(T \circ S)$. (Suggerimento: se $(S \circ T)(v) = \lambda v$ allora $(T \circ S)(T(v)) = \lambda T(v)$.)

14.10 Trova due matrici $A, A' \in M_{2,2}(\mathbb{R})$ che non siano simili ma che abbiano comunque polinomio caratteristico, autovalori, determinante e traccia uguali.

14.11 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale V di dimensione n . Dimostra che esiste un polinomio p di grado n^2 tale che $p(T) = O$. (Suggerimento: gli endomorfismi $\text{id}, T, T^2, \dots, T^{n^2}$ sono $n^2 + 1$ elementi dello spazio vettoriale $\mathcal{L}(V, V)$ che ha dimensione n^2 .)

14.12 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale V sul campo \mathbb{K} . e siano $\lambda_1, \dots, \lambda_k \in \mathbb{K}$ gli autovalori di T . Preso un polinomio $p \in \mathbb{K}[t]$ calcola gli autovalori di $p(T)$, e dimostra che se T è diagonalizzabile allora anche $p(T)$ lo è. È vero il viceversa?

14.13 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale di dimensione finita sul campo \mathbb{K} , e $U \subseteq V$ un sottospazio di V tale che $T(U) \subseteq U$. Dimostra che il

polinomio caratteristico della restrizione $T|_U$ di T a U divide p_T . In particolare, se T ha tutti gli autovalori in \mathbb{K} anche $T|_U$ li ha. (Suggerimento: usa l'Esercizio 9.10.)

14.14 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale di dimensione finita sul campo \mathbb{K} , e $U \subseteq V$ un sottospazio di V tale che $T(U) \subseteq U$. Supponiamo che T sia diagonalizzabile. Dimostra che esiste un supplementare $W \subseteq V$ di U in V tale che $T(W) \subseteq W$.

14.15 Trova un endomorfismo $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ e un sottospazio $U \subset \mathbb{R}^2$ tali che si abbia $T(U) \subseteq U$ ma $T(W) \not\subseteq W$ per ogni supplementare W di U in \mathbb{R}^2 .

14.16 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale di dimensione finita sul campo \mathbb{K} , e $U \subseteq V$ un sottospazio di V tale che $T(U) \subseteq U$. Supponiamo che T sia diagonalizzabile. Dimostra che anche l'endomorfismo $T|_U: U \rightarrow U$ restrizione di T a U è diagonalizzabile. (Suggerimento: se $V = V_{\lambda_1} \oplus \dots \oplus V_{\lambda_k}$, dimostra che $U = U \cap V_{\lambda_1} \oplus \dots \oplus U \cap V_{\lambda_k}$.)

14.17 Siano $S, T: V \rightarrow V$ due endomorfismi di uno spazio vettoriale di dimensione finita sul campo \mathbb{K} tali che $S \circ T = T \circ S$. Sia $\lambda \in \text{sp}(S)$ un autovalore di S , e V_λ il relativo autospazio. Dimostra che $T(V_\lambda) \subseteq V_\lambda$.

14.18 Siano $S, T: V \rightarrow V$ due endomorfismi diagonalizzabili di uno spazio vettoriale di dimensione finita sul campo \mathbb{K} . Dimostra che esiste una base di V composta da vettori che sono autovettori sia per T che per S se e solo se $S \circ T = T \circ S$. (Suggerimento: usa i due Esercizi precedenti.)

14.19 Sia $A = (a_{ij}) \in M_{n,n}(\mathbb{R})$ tale che $a_{i1} + \dots + a_{in} = 1$ per $i = 1, \dots, n$. Dimostra che $1 \in \text{sp}(LA)$.

14.20 Siano $A, A' \in M_{n,n}(\mathbb{R})$ due matrici simili su \mathbb{C} , cioè supponiamo che esista $\tilde{B} \in GL_n(\mathbb{C})$ con $A' = \tilde{B}^{-1}A\tilde{B}$. Dimostra che A e A' sono simili su \mathbb{R} , cioè che si può trovare una matrice $B \in GL_n(\mathbb{R})$ tale che $A' = B^{-1}AB$. (Suggerimento: scrivi $\tilde{B} = P + iQ$, dove $P, Q \in M_{n,n}(\mathbb{R})$, e dimostra che esiste $r \in \mathbb{R}$ tale che $B = P + rQ$ sia come richiesto.)

14.21 Sia $T: V \rightarrow V$ un endomorfismo tale che $T^2 = T$. Dimostra che T è diagonalizzabile. (Suggerimento: vedi l'Esercizio 7.7.)

14.22 Sia $T: V \rightarrow V$ un endomorfismo tale che $T^2 = \text{id}_V$. Dimostra che T è diagonalizzabile. (Suggerimento: rifletti sull'Esercizio 5.10.)

14.23 Sia $T: V \rightarrow V$ un endomorfismo tale che $T^r = O$ per qualche $r \in \mathbb{N}$; si dice che T è nilpotente. Dimostra che T è diagonalizzabile se e solo se $T = O$.

14.24 Sia $v_0 \neq O$ un autovettore di un endomorfismo $T: V \rightarrow V$ relativo a un autovalore non nullo. Dimostra che $v_0 \in \text{Im } T$. Deduci che se T è diagonalizzabile allora $V = \text{Im } T \oplus \text{Ker } T$. È vero il viceversa?

COMPLEMENTI

14C.1 Il polinomio minimo

L'obiettivo di questo e del prossimo paragrafo è dimostrare il criterio di diagonalizzabilità descritto nell'Osservazione 14.8. Prima di proseguire è necessario che tu legga (o abbia già letto) i Complementi al Capitolo 11.

Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale V sul campo \mathbb{K} . Preso un polinomio

$$p(t) = a_d t^d + a_{d-1} t^{d-1} + \dots + a_0 \in \mathbb{K}[t],$$

l'endomorfismo $p(T) \in \mathcal{L}(V, V)$ è dato per definizione (vedi l'Osservazione 14.8) da

$$p(T) = a_d T^d + a_{d-1} T^{d-1} + \dots + a_0 \text{id}_V.$$

È facile verificare (Esercizio 14C.1) che se $p, q \in \mathbb{K}[t]$ sono due polinomi allora

$$p(T) \circ q(T) = (pq)(T) = q(T) \circ p(T); \quad (14C.1)$$

in particolare, $q(T)(\text{Ker } p(T)) \subseteq \text{Ker } p(T)$ e $q(T)(\text{Im } p(T)) \subseteq \text{Im } p(T)$.

Come indicato nell'Esercizio 14.11, può capitare che $p(T)$ sia l'applicazione nulla; indichiamo con \mathcal{I}_T l'insieme dei polinomi per cui questo accade. In simboli,

$$\mathcal{I}_T = \{p \in \mathbb{K}[t] \mid p(T) = O\}.$$

Lemma 14C.1 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale di dimensione finita sul campo \mathbb{K} . Allora \mathcal{I}_T è un ideale proprio di $\mathbb{K}[t]$.

Dimostrazione. Prima di tutto, se $p_1, p_2 \in \mathcal{I}_T$ allora

$$(p_1 + p_2)(T) = p_1(T) + p_2(T) = O + O = O,$$

e $p_1 + p_2 \in \mathcal{I}_T$. Inoltre, se $p \in \mathbb{K}[t]$ e $q \in \mathcal{I}_T$, la (14C.1) implica

$$(pq)(T) = p(T) \circ q(T) = p(T) \circ O = O,$$

per cui $pq \in \mathcal{I}_T$, e \mathcal{I}_T è un ideale. Infine, $1 \notin \mathcal{I}_T$ e $\mathcal{I}_T \neq \{0\}$ (Esercizio 14.11), per cui \mathcal{I}_T è un ideale proprio. \square

Definizione 14C.1 Il polinomio minimo dell'endomorfismo $T: V \rightarrow V$ è il generatore monico μ_T dell'ideale \mathcal{I}_T (vedi il Teorema 11C.6). Nota che μ_T è il polinomio monico di grado minimo contenuto in \mathcal{I}_T .

Osservazione 14C.1 Si può dimostrare (Esercizio 15.20) che il polinomio caratteristico p_T di T appartiene a \mathcal{I}_T (Teorema di Hamilton-Cayley), per cui μ_T divide p_T .

Il polinomio minimo è esplicitamente calcolabile; per vedere come, ci servono un altro paio di definizioni e risultati.

Definizione 14C.2 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale di dimensione finita sul campo \mathbb{K} , e $v \in V$ un vettore non nullo. Poniamo

$$\mathcal{I}_{T,v} = \{p \in \mathbb{K}[t] \mid v \in \text{Ker } p(T)\} = \{p \in \mathbb{K}[t] \mid p(T)(v) = O\}.$$

È facile verificare (esercizio) che $\mathcal{I}_{T,v}$ è un ideale proprio di $\mathbb{K}[t]$; indichiamo con $\mu_{T,v}$ il suo generatore monico.

Lemma 14C.2 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale di dimensione finita sul campo \mathbb{K} , e $v \in V$ un vettore non nullo. Allora $\mu_{T,v}$ divide μ_T .

Dimostrazione. Per costruzione, $\mathcal{I}_T \subseteq \mathcal{I}_{T,v}$; dunque $\mu_T \in \mathcal{I}_{T,v}$, per cui è un multiplo di $\mu_{T,v}$. \square

Lemma 14C.3 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale di dimensione finita sul campo \mathbb{K} , e $v \in V$ un vettore non nullo. Sia $d > 0$ il più piccolo intero tale che $v, T(v), \dots, T^{d-1}(v)$ siano linearmente indipendenti mentre esistono $\alpha_0, \dots, \alpha_{d-1} \in \mathbb{K}$ tali che

$$\alpha_0 v + \alpha_1 T(v) + \dots + \alpha_{d-1} T^{d-1}(v) + T^d(v) = O.$$

Allora $\mu_{T,v}(t) = t^d + \alpha_{d-1} t^{d-1} + \dots + \alpha_0$.

Dimostrazione. Sia $p(t) = t^k + a_{k-1} t^{k-1} + \dots + a_0 \in \mathcal{I}_{T,v}$ un qualunque polinomio monico nell'ideale $\mathcal{I}_{T,v}$. Dunque

$$a_0 v + a_1 T(v) + \dots + a_{k-1} T^{k-1}(v) + T^k(v) = O,$$

per cui $v, T(v), \dots, T^k(v)$ sono linearmente dipendenti e quindi $k \geq d$. Questo vuol dire che qualunque polinomio monico in $\mathcal{I}_{T,v}$ ha grado almeno d ; siccome il polinomio $q(t) = t^d + \alpha_{d-1} t^{d-1} + \dots + \alpha_0$ per ipotesi appartiene a $\mathcal{I}_{T,v}$ e ha grado esattamente d , dalla minimalità del grado di $\mu_{T,v}$ segue che $\mu_{T,v} = q$. \square

Proposizione 14C.4 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale di dimensione finita su \mathbb{K} , e $\{v_1, \dots, v_n\}$ una base di V . Allora μ_T è il minimo comune multiplo dei polinomi $\mu_{T,v_1}, \dots, \mu_{T,v_n}$.

Dimostrazione. Prima di tutto, dimostriamo che $\mathcal{I}_T = \mathcal{I}_{T,v_1} \cap \dots \cap \mathcal{I}_{T,v_n}$. Infatti, se $p(T) = O$ si ha $v_j \in \text{Ker } p(T) = V$ per $j = 1, \dots, n$, per cui $\mathcal{I}_T \subseteq \mathcal{I}_{T,v_1} \cap \dots \cap \mathcal{I}_{T,v_n}$. Viceversa, se $p \in \mathcal{I}_{T,v_1} \cap \dots \cap \mathcal{I}_{T,v_n}$ allora $p(T)(v_j) = O$ per $j = 1, \dots, n$; essendo $\{v_1, \dots, v_n\}$ una base di V questo implica che $p(T)(v) = O$ per ogni $v \in V$, cioè $p(T) = O$ e $p \in \mathcal{I}_T$.

Dunque $\mathcal{I}_T = \mathcal{I}_{T,v_1} \cap \dots \cap \mathcal{I}_{T,v_n}$, e la tesi segue dalla Proposizione 11C.11.(ii). \square

Usando il Lemma 14C.3 e la Proposizione 14C.4 è possibile determinare esplicitamente il polinomio minimo di un endomorfismo. Scelta una base $\{v_1, \dots, v_n\}$ di V , per ogni $j = 1, \dots, n$ si calcolano $v_j, T(v_j), \dots, T^{d_j}(v_j)$ fino al minimo $d_j \geq 1$ per cui diventano linearmente dipendenti. Allora il Lemma 14C.3 fornisce $\mu_{T.v_1}, \dots, \mu_{T.v_n}$, e applicando le Proposizioni 14C.4 e 11C.11 troviamo μ_T .

Osservazione 14C.2 Vedremo nell'Esercizio 14C.4 che il grado del polinomio minimo è minore o uguale a $n = \dim V$. In particolare, se qualcuno dei $\mu_{T.v_j}$ ha grado n , allora $\mu_T = \mu_{T.v_j}$ senza bisogno di fare altri conti. Infatti in tal caso $\mu_{T.v_j}$ divide μ_T , e inoltre hanno lo stesso grado; basta quindi applicare il Lemma 11C.5.

ESEMPIO 14C.1 Sia $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ l'endomorfismo dato dalla matrice

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

Vogliamo calcolare il polinomio minimo di T . Cominciamo col trovare $\mu_{T.e_1}$. Abbiamo

$$T(e_1) = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad T^2(e_1) = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \quad T^3(e_1) = \begin{pmatrix} 0 \\ 3 \\ -4 \end{pmatrix} = -4T^2(e_1) - T(e_1) + e_1,$$

per cui il Lemma 14C.3 ci assicura che $\mu_{T.e_1}(t) = t^3 + 4t^2 + t - 1$. L'Osservazione 14C.2 implica quindi che $\mu_T = \mu_{T.e_1}$ — e infatti un conto veloce conferma che $\mu_{T.e_2} = \mu_{T.e_3} = \mu_{T.e_1}$.

Abbiamo quindi visto come il polinomio minimo sia effettivamente calcolabile. Il nostro prossimo obiettivo è mostrare come il polinomio minimo ci dica esattamente quando un endomorfismo è diagonalizzabile. Il risultato cruciale è il

Teorema 14C.5 Sia V uno spazio vettoriale su un campo \mathbb{K} , e $\mu_T \in \mathbb{K}[t]$ il polinomio minimo di un endomorfismo $T \in \mathcal{L}(V, V)$. Sia $\mu_T = p_1^{m_1} \cdots p_s^{m_s}$ la fattorizzazione (11C.6) di μ_T , dove $p_1, \dots, p_s \in \mathbb{K}[t]$ sono polinomi monici irriducibili distinti, e poniamo $V_j = \text{Ker } p_j^{m_j}(T)$ per $j = 1, \dots, s$. Allora

$$V = V_1 \oplus \cdots \oplus V_s.$$

Dimostrazione. Se $s = 1$ l'asserto è ovvio; assumiamo quindi $s \geq 2$, e procediamo per induzione sulla dimensione di V . Se $\dim V = 1$ non c'è nulla da dimostrare; sia allora $\dim V > 1$ e supponiamo l'asserto vero per tutti gli spazi vettoriali di dimensione minore di V . Poniamo $\tilde{p}_1 = p_1^{m_1}$ e $\tilde{p}_2 = p_2^{m_2} \cdots p_s^{m_s}$, in modo che $\mu_T = \tilde{p}_1 \tilde{p}_2$ con \tilde{p}_1 e \tilde{p}_2 relativamente primi. Vogliamo dimostrare che $V = V_1 \oplus \tilde{V}_2$, dove $\tilde{V}_2 = \text{Ker } \tilde{p}_2(T)$. Nota che $V_1, \tilde{V}_2 \neq V$ in quanto \tilde{p}_1 e \tilde{p}_2 hanno grado strettamente minore di μ_T e quindi non appartengono a \mathcal{I}_T .

Essendo \tilde{p}_1 e \tilde{p}_2 relativamente primi, l'Osservazione 11C.1 fornisce $q_1, q_2 \in \mathbb{K}[t]$ tali che $q_1 \tilde{p}_1 + q_2 \tilde{p}_2 = 1$; quindi

$$q_1(T) \circ \tilde{p}_1(T) + q_2(T) \circ \tilde{p}_2(T) = \text{id}_V. \tag{14C.2}$$

Prendiamo $v \in V_1 \cap \tilde{V}_2$; allora

$$v = \text{id}_V(v) = q_1(T)(\tilde{p}_1(T)(v)) + q_2(T)(\tilde{p}_2(T)(v)) = q_1(T)(0) + q_2(T)(0) = 0,$$

per cui $V_1 \cap \tilde{V}_2 = \{0\}$. Infine, per $v \in V$ qualunque poniamo $v_j = q_j \tilde{p}_j(T)(v)$, con $j = 1, 2$. La (14C.2) implica che $v = v_1 + v_2$; inoltre

$$\begin{aligned} \tilde{p}_2(T)(v_1) &= (\tilde{p}_2(T) \circ q_1(T) \circ \tilde{p}_1(T))(v) = q_1(T)((\tilde{p}_1(T) \circ \tilde{p}_2(T))(v)) \\ &= q_1(T)(\mu_T(T)(v)) = q_1(T)(0) = 0, \end{aligned}$$

per cui $v_1 \in \tilde{V}_2$. Analogamente si dimostra che $v_2 \in V_1$, e quindi $V = V_1 \oplus \tilde{V}_2$ come volevamo.

Ora, l'Esercizio 14C.1 implica che $T(\tilde{V}_2) \subseteq \tilde{V}_2$; inoltre \tilde{p}_2 è proprio il polinomio minimo di $T|_{\tilde{V}_2}$. Infatti, se così non fosse esisterebbe un polinomio \tilde{q} di grado strettamente minore del grado di \tilde{p}_2 tale che $\tilde{q}(T)(\tilde{v}) = 0$ per ogni $\tilde{v} \in \tilde{V}_2$. Ma allora avremmo $\tilde{p}_1 \tilde{q}(T)(v) = 0$ per ogni $v \in V$ (perché?), col grado di $\tilde{p}_1 \tilde{q}$ strettamente minore del grado di $\tilde{p}_1 \tilde{p}_2 = \mu_T$, impossibile.

Dunque possiamo applicare l'ipotesi induttiva a \tilde{V}_2 , ottenendo $\tilde{V}_2 = V_2 \oplus \cdots \oplus V_k$, ed è fatta. \square

Definizione 14C.3 Diremo che un polinomio $p \in \mathbb{K}[t]$ non ha radici multiple se tutte le radici di p hanno molteplicità 1. Diremo che p ha tutte le radici in \mathbb{K} se si può scrivere

$$p(t) = c(t - \lambda_1) \cdots (t - \lambda_d)$$

con $c \neq 0$ e $\lambda_1, \dots, \lambda_d \in \mathbb{K}$ non necessariamente distinti.

Ed ecco l'annunciato criterio:

Corollario 14C.6 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale di dimensione finita sul campo \mathbb{K} . Allora T è diagonalizzabile se e solo se μ_T ha tutte le radici in \mathbb{K} e non ha radici multiple.

Dimostrazione. Supponiamo che T sia diagonalizzabile, e sia $\mathcal{B} = \{v_1, \dots, v_n\}$ una base di autovettori. Chiaramente, $\mu_{T.v_j}(t) = t - \lambda_j$ per $j = 1, \dots, n$, dove λ_j è l'autovalore di v_j . Quindi il Lemma 14C.4 implica $\mu_T(t) = (t - \lambda_1) \cdots (t - \lambda_n)$, dove $\lambda_1, \dots, \lambda_n \in \mathbb{K}$ sono gli autovalori distinti di T , per cui μ_T ha tutte le radici in \mathbb{K} e non ha radici multiple.

Viceversa, supponiamo che μ_T soddisfi le ipotesi; quindi esistono $\lambda_1, \dots, \lambda_k \in \mathbb{K}$ distinti tali che $\mu_T(t) = (t - \lambda_1) \cdots (t - \lambda_k)$. Per il Teorema 14C.5, possiamo allora scrivere $V = V_1 \oplus \cdots \oplus V_k$, dove $V_j = \text{Ker}(T - \lambda_j \text{id}_V)$ è proprio l'autospazio relativo a λ_j , e abbiamo dimostrato che T è diagonalizzabile.

Osservazione 14C.3 Può capitare che μ_T abbia tutti gli autovalori nel campo ma T non sia diagonalizzabile. Per esempio, sia $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dato dalla matrice

$$\begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix}.$$

Un conto veloce mostra che $\mu_T(t) = (t-1)^2$, per cui T non è diagonalizzabile.

Dunque il polinomio minimo (che, come abbiamo visto, è esplicitamente calcolabile) ci dice se un endomorfismo è diagonalizzabile. O meglio, ce lo dirà quando saremo in grado di stabilire se un polinomio ha tutte le radici nel campo e non ha radici multiple. Anche questo si può fare, ed è l'obiettivo del prossimo paragrafo.

14C.2 Il Teorema di Sturm

In questo paragrafo descriveremo delle tecniche che permettono di stabilire se un polinomio ha radici multiple, e se ha tutte le radici in \mathbb{R} ; come conseguenza troveremo un algoritmo per stabilire se un endomorfismo è diagonalizzabile senza bisogno di calcolarne gli autovalori.

Cominciamo con

Proposizione 14C.7 Sia $p \in \mathbb{K}[t]$ un polinomio non costante, e $p' \in \mathbb{K}[t]$ la sua derivata. Allora le seguenti affermazioni sono equivalenti:

- (i) p ha una radice di molteplicità maggiore di 1;
- (ii) p e p' hanno una radice in comune;
- (iii) il massimo comun divisore di p e p' ha una radice in \mathbb{K} .

Dimostrazione. (i) \implies (ii). Supponiamo che $\lambda_0 \in \mathbb{K}$ sia una radice di molteplicità almeno due. Allora possiamo scrivere $p(t) = (t - \lambda_0)^2 q(t)$ per qualche $q \in \mathbb{K}[t]$. Dunque

$$p'(t) = 2(t - \lambda_0)q(t) + (t - \lambda_0)^2 q'(t),$$

per cui λ_0 è radice anche di p' .

(ii) \implies (i). Supponiamo che $\lambda_0 \in \mathbb{K}$ sia una radice comune di p e p' . Possiamo certamente scrivere $p(t) = (t - \lambda_0)q(t)$ per qualche $q \in \mathbb{K}[t]$. In particolare,

$$p'(t) = q(t) + (t - \lambda_0)q'(t),$$

per cui $p'(\lambda_0) = q(\lambda_0)$. Essendo $p'(\lambda_0) = 0$ per ipotesi, ne segue che $t - \lambda_0$ divide anche q , e quindi λ_0 è radice di p di molteplicità almeno due.

(ii) \iff (iii). Lo scalare $\lambda_0 \in \mathbb{K}$ è radice sia di p che di p' se e solo se $t - \lambda_0$ divide sia p che p' se e solo se $t - \lambda_0$ divide (p, p') se e solo se λ_0 è radice di (p, p') . \square

Dunque per vedere se p ha radici multiple ci basta controllare se il massimo comun divisore di p e p' ha radici in \mathbb{K} . Se $\mathbb{K} = \mathbb{C}$, il massimo comun divisore non ha radici se e solo se è una costante, per cui abbiamo dimostrato il

Corollario 14C.8 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale di dimensione finita su \mathbb{C} . Allora T è diagonalizzabile se e solo se $(\mu_T, \mu'_T) = 1$, dove μ'_T è la derivata del polinomio minimo di T .

Quindi abbiamo trovato un algoritmo completamente soddisfacente per scoprire se un endomorfismo su uno spazio vettoriale complesso è diagonalizzabile: si calcola il polinomio minimo μ_T , se ne fa la derivata μ'_T e si calcola il massimo comun divisore (μ_T, μ'_T) . Se $(\mu_T, \mu'_T) = 1$ l'endomorfismo è diagonalizzabile, altrimenti non lo è.

Sui reali invece per ora abbiamo il meno soddisfacente

Corollario 14C.9 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale di dimensione finita su \mathbb{R} . Allora T è diagonalizzabile se e solo se μ_T ha tutte le radici in \mathbb{R} e $(\mu_T, \mu'_T) = 1$.

Dimostrazione. Basta osservare che se μ_T ha tutte le radici in \mathbb{R} anche (μ_T, μ'_T) lo deve avere, per cui l'asserto segue dal Corollario 14C.6 e dalla Proposizione 14C.7. \square

Dobbiamo quindi trovare un criterio che dica quando un polinomio $p \in \mathbb{R}[t]$ ha tutte le radici in \mathbb{R} , almeno nel caso in cui p e p' siano relativamente primi.

Ci servono due definizioni.

Definizione 14C.4 Sia $\mathbf{c} = (c_0, \dots, c_s) \in \mathbb{R}^s$ una successione finita di numeri reali. Se tutti gli elementi della successione sono diversi da zero, diremo *numero di variazioni* $V_{\mathbf{c}}$ in \mathbf{c} il numero di indici j tali che $c_j c_{j+1} < 0$, cioè tali che c_j e c_{j+1} abbiano segni opposti. Se invece qualcuno degli elementi della successione \mathbf{c} è nullo, diremo *numero di variazioni* $V_{\mathbf{c}}$ in \mathbf{c} il numero di variazioni della successione \mathbf{c}' composta dagli elementi non nulli di \mathbf{c} .

Definizione 14C.5 Sia $p \in \mathbb{R}[t]$ un polinomio non costante. Diremo *sequenza standard* associata a p la successione $p_0, p_1, \dots, p_s \in \mathbb{R}[t]$ di polinomi così definita:

$$\begin{array}{ll} p_0 = p, & p_1 = p', \\ p_0 = q_1 p_1 - p_2, & \text{con } \deg p_2 < \deg p_1, \\ \vdots & \vdots \\ p_{j-1} = q_j p_j - p_{j+1}, & \text{con } \deg p_{j+1} < \deg p_j, \\ \vdots & \vdots \\ p_{s-1} = q_s p_s, & (\text{cioè } p_{s+1} \equiv 0). \end{array}$$

In altri termini, la sequenza standard è ottenuta modificando il segno del resto nell'algoritmo di Euclide per il calcolo del massimo comun divisore di p e p' . Dunque p_s , a meno di una costante moltiplicativa non nulla, è il massimo comun divisore di p e p' ; in particolare, $(p, p') = 1$ se e solo se p_s è una costante non nulla.

Il risultato che cerchiamo sarà una conseguenza del seguente *Teorema di Sturm*:

Teorema 14C.10 (Sturm) Sia $p \in \mathbb{R}[t]$ un polinomio tale che $(p, p') = 1$, e prendiamo due numeri reali $a < b$ tali che $p(a)p(b) \neq 0$; sia inoltre $p_0, \dots, p_s \in \mathbb{R}[t]$ la sequenza standard associata a p . Allora il numero di radici di p nell'intervallo (a, b) è dato da $V_a - V_b$, dove $\mathbf{a} = (p_0(a), \dots, p_s(a))$ e $\mathbf{b} = (p_0(b), \dots, p_s(b))$.

Dimostrazione. Cominciamo elencando alcune proprietà della sequenza standard.

- (a) p_s non ha radici in $[a, b]$, e $p_0(a)p_0(b) \neq 0$: sono le ipotesi.
- (b) Se $\lambda_0 \in [a, b]$ è una radice di p_j , con $0 < j < s$, allora $p_{j-1}(\lambda_0)p_{j+1}(\lambda_0) < 0$. Infatti,

$$p_{j-1}(\lambda_0) = q_j(\lambda_0)p_j(\lambda_0) - p_{j+1}(\lambda_0) = -p_{j+1}(\lambda_0);$$

inoltre, se fosse $p_{j-1}(\lambda_0) = 0$ allora $t - \lambda_0$ dividerebbe sia p_j che p_{j-1} . E quindi dividerebbe p_{j-2} e così via, per cui $t - \lambda_0$ dividerebbe sia p che p' , eventualità che abbiamo escluso.

- (c) Sia $\lambda_0 \in [a, b]$ una radice di p , e $\alpha < \lambda_0 < \beta$ tali che $p_0(t)p_1(t) \neq 0$ per ogni $t \in (\alpha, \lambda_0) \cup (\lambda_0, \beta)$. Allora $p_0(t)p_1(t) < 0$ per ogni $t \in (\alpha, \lambda_0)$ e $p_0(t)p_1(t) > 0$ per ogni $t \in (\lambda_0, \beta)$. Infatti, per la Proposizione 14C.7 il polinomio p non ha radici multiple, per cui $p_0(t) = (t - \lambda_0)q(t)$ con $q(\lambda_0) = p_1(\lambda_0) \neq 0$. Siccome per ipotesi $p_1(t)q(t) \neq 0$ in (α, β) , il Teorema dei valori intermedi (Teorema 1.5) ci assicura che $p_1(t)q(t) > 0$ per ogni $t \in (\alpha, \beta)$. Allora $p_0(t)p_1(t) = (t - \lambda_0)p_1(t)q(t)$ ha lo stesso segno di $t - \lambda_0$ in (α, β) , come voluto.

Ora decomponiamo l'intervallo $[a, b]$ utilizzando le radici dei p_i . In altri termini, prendiamo dei punti

$$a = a_0 < a_1 < \dots < a_m = b$$

tali che nessun p_i abbia radici in (a_j, a_{j+1}) per $i = 0, \dots, s$ e $j = 0, \dots, m - 1$. Scegliamo inoltre $\lambda_j \in (a_j, a_{j+1})$ per $j = 0, \dots, m - 1$.

Per $\lambda \in \mathbb{R}$ poniamo $\lambda = (p_0(\lambda), \dots, p_s(\lambda))$. Vogliamo dimostrare che $V_a = V_\lambda$. Siccome ciascun p_i non ha radici in (a_0, a_1) , il Teorema dei valori intermedi ci assicura che

$$p_i(a_0)p_i(\lambda_0) \geq 0 \tag{14C.3}$$

per $i = 0, \dots, s$; in particolare, se a_0 non è radice di nessuno dei p_i , abbiamo subito $V_a = V_{\lambda_0}$. Supponiamo invece che $p_{i_0}(a_0) = 0$ per qualche i_0 ; chiaramente, dev'essere $0 < i_0 < s$. La proprietà (b) allora ci dice che $p_{i_0-1}(a_0)p_{i_0+1}(a_0) < 0$, per cui (14C.3) implica che $p_{i_0\pm 1}(a_0)$ e $p_{i_0\pm 1}(\lambda_0)$ hanno lo stesso segno, da cui segue che $p_{i_0-1}(\lambda_0)p_{i_0+1}(\lambda_0) < 0$. In conclusione, il numero di variazioni è uno sia in $(p_{i_0-1}(a_0), 0, p_{i_0+1}(a_0))$ che in $(p_{i_0-1}(\lambda_0), p_{i_0}(\lambda_0), p_{i_0+1}(\lambda_0))$. Ripetendo il ragionamento per tutti i possibili i_0 otteniamo quindi che $V_a = V_{\lambda_0}$ sempre, come voluto.

In modo analogo si dimostra che $V_{\lambda_{m-1}} = V_b$. Prendiamo ora $0 \leq j < m - 1$. Il ragionamento di prima mostra che $V_{\lambda_j} = V_{\lambda_{j+1}}$ se $p(a_{j+1}) \neq 0$. Supponiamo invece che $p(a_{j+1}) = 0$. Allora la (c) dice che $p_0(\lambda_j)p_1(\lambda_j) < 0$ e $p_0(\lambda_{j+1})p_1(\lambda_{j+1}) > 0$, per cui $(p_0(\lambda_j), p_1(\lambda_j))$ ha una variazione mentre $(p_0(\lambda_{j+1}), p_1(\lambda_{j+1}))$ non ne ha. Infine, il solito ragionamento mostra che $(p_{i-1}(\lambda_j), p_i(\lambda_j), p_{i+1}(\lambda_j))$ ha lo stesso numero di variazioni di $(p_{i-1}(\lambda_{j+1}), p_i(\lambda_{j+1}), p_{i+1}(\lambda_{j+1}))$ quando $i > 1$; quindi se $p(a_{j+1}) = 0$ si ha $V_{\lambda_j} - V_{\lambda_{j+1}} = 1$.

Ora possiamo scrivere

$$V_a - V_b = (V_a - V_{\lambda_0}) + \sum_{j=0}^{m-2} (V_{\lambda_j} - V_{\lambda_{j+1}}) + (V_{\lambda_{m-1}} - V_b);$$

ma abbiamo dimostrato che $V_a = V_{\lambda_0}$, $V_b = V_{\lambda_{m-1}}$ e che $(V_{\lambda_j} - V_{\lambda_{j+1}}) = 0$ o 1 a seconda che $p(a_{j+1}) \neq 0$ o $p(a_{j+1}) = 0$; quindi $V_a - V_b$ è proprio il numero delle radici di p in $[a, b]$. \square

Osservazione 14C.4 Il Teorema di Sturm vale anche senza l'ipotesi $(p, p') = 1$; vedi l'Esercizio 14C.5.

Dunque il Teorema di Sturm ci dice come localizzare le radici reali di un polinomio; con una piccola variazione ci può anche dire esattamente quante sono. Per dimostrarlo ci serve il

Lemma 14C.11 Sia $p(t) = a_d t^d + \dots + a_0 \in \mathbb{R}[t]$ un polinomio non costante, e prendiamo

$$C_0 = \max\{|a_0|, \dots, |a_d|\}.$$

Allora $p(t_0)$ ha lo stesso segno di a_d e $p(-t_0)$ ha lo stesso segno di $(-1)^d a_d$ non appena $t_0 > dC_0/|a_d| \geq 1$. In particolare, $p(t_0) \neq 0$ per $|t_0| > dC_0/|a_d|$.

Dimostrazione. Infatti abbiamo

$$\begin{aligned} |a_{d-1}t_0^{d-1} + \dots + a_0| &\leq |a_{d-1}t_0^{d-1}| + \dots + |a_0| \leq C_0[|t_0|^{d-1} + \dots + 1] \\ &\leq dC_0|t_0|^{d-1} < |a_d t_0^d|, \end{aligned}$$

per cui il segno di $p(t_0)$ è totalmente determinato dal segno del coefficiente direttivo. \square

E quindi

Corollario 14C.12 Sia $p \in \mathbb{R}[t]$ un polinomio non costante con $(p, p') = 1$. Indichiamo con $p_0, \dots, p_s \in \mathbb{R}[t]$ la sequenza standard associata a p , e siano d_j il grado e $c_j \in \mathbb{R}$ il coefficiente direttivo di p_j , per $j = 0, \dots, s$. Allora il numero di radici reali di p è $V_- - V_+$, dove V_+ è il numero di variazioni nella successione (c_0, \dots, c_s) , e V_- è il numero di variazioni nella successione $((-1)^{d_0}c_0, \dots, (-1)^{d_s}c_s)$.

Dimostrazione. Per il Lemma 14C.11, esiste un $M > 1$ tale che $p_j(M)$ ha lo stesso segno di c_j e $p_j(-M)$ ha lo stesso segno di $(-1)^{d_j}c_j$ per $j = 0, \dots, s$; inoltre possiamo supporre che tutte le radici reali di p siano contenute in $[-M, M]$. Allora ci basta applicare il Teorema 14C.10 con $a = -M$ e $b = M$. \square

Finalmente abbiamo un metodo effettivo per stabilire anche su \mathbb{R} se un endomorfismo è diagonalizzabile. Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale V su \mathbb{R} o su \mathbb{C} . Per stabilire se T è diagonalizzabile bisogna procedere come segue:

- (1) Calcolare il polinomio minimo μ_T di T .
- (2) Calcolare la sequenza standard p_0, \dots, p_s associata a μ_T . Se p_s non è costante, l'endomorfismo T non è diagonalizzabile. Se p_s è costante e siamo sui complessi, T è diagonalizzabile su \mathbb{C} . Se p_s è costante e siamo sui reali, proseguire.
- (3) Calcolare, usando il Corollario 14C.12, il numero di radici reali (che sono necessariamente tutte distinte, in quanto p_s costante implica che μ_T e la sua derivata sono relativamente primi) di μ_T . Se μ_T ha esattamente $d = \deg \mu_T$ radici reali, l'endomorfismo T è diagonalizzabile su \mathbb{R} ; altrimenti non lo è.

ESEMPIO 14C.2 Sia $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ l'endomorfismo dell'Esempio 14C.1. Sappiamo che $\mu_T(t) = t^3 + 4t^2 + t - 1$. La sequenza standard è

$$\begin{aligned} p_0(t) &= t^3 + 4t^2 + t - 1, & p_1(t) &= 3t^2 + 8t + 1, \\ p_2(t) &= \frac{26}{9}t + \frac{13}{9}, & p_3(t) &= \frac{7}{4}. \end{aligned}$$

Dunque sicuramente T è diagonalizzabile su \mathbb{C} . Per vedere cosa accade su \mathbb{R} , il Corollario 14C.12 ci dice di considerare le successioni $(-1, 3, -\frac{26}{9}, \frac{7}{4})$ e $(1, 3, \frac{26}{9}, \frac{7}{4})$, da cui troviamo $V_- - V_+ = 3 - 0 = 3$. Dunque μ_T ha tre radici reali distinte (anche se non sappiamo quali sono), e T è diagonalizzabile su \mathbb{R} .

Esercizi

14C.1 Siano $p, q \in \mathbb{K}[t]$ polinomi, e $T: V \rightarrow V$ un endomorfismo. Dimostra che $(pq)(T) = p(T) \circ q(T)$ e che $p(T) \circ q(T) = q(T) \circ p(T)$. Deduci che

$$q(T)(\text{Ker } p(T)) \subseteq \text{Ker } p(T) \quad \text{e} \quad q(T)(\text{Im } p(T)) \subseteq \text{Im } p(T).$$

14C.2 Sia $T: V \rightarrow V$ un endomorfismo. Dimostra che $v \in V$ è autovettore di T relativo a $\lambda \in \mathbb{K}$ se e solo se $\mu_{T,v}(t) = t - \lambda$. Deduci che gli autovalori di T sono esattamente le radici di μ_T . (Suggerimento: se $\lambda \notin \text{sp}(T)$ allora $T - \lambda \text{id}_V$ è invertibile.)

14C.3 Sia $T: V \rightarrow V$ un endomorfismo tale che $\mu_T = p^m$, dove $p \in \mathbb{K}[t]$ è un polinomio irriducibile. Sia $\mathcal{B} = \{v_1, \dots, v_n\}$ una base di V ; dimostra che esiste un $v_j \in \mathcal{B}$ tale che $\mu_{T,v_j} = \mu_T$. In particolare, $\deg \mu_T \leq \dim V$. (Suggerimento: ciascun μ_{T,v_i} divide μ_T , per cui è della forma p^{k_j} con $k_j \leq m$.)

14C.4 Sia $T: V \rightarrow V$ un endomorfismo. Dimostra che $\deg \mu_T \leq \dim V$. (Suggerimento: sia $\mu_T = p_1^{m_1} \cdots p_s^{m_s}$ la fattorizzazione di μ_T in polinomi irriducibili. Posto $V_j = \text{Ker } p_j^{m_j}(T)$, applica l'Esercizio precedente a ciascun V_j e poi usa il Teorema 14C.5.)

14C.5 Dimostra che il Teorema di Sturm vale anche per polinomi $p \in \mathbb{R}[t]$ tali che $(p, p') \neq 1$. Deduci che il criterio del Corollario 14C.12 fornisce il numero delle radici reali distinte di qualunque polinomio $p \in \mathbb{R}[t]$. (Suggerimento: posto $\tilde{p}_j = p_j / (p, p')$, dove p_0, \dots, p_s è la sequenza standard associata a p , dimostra che la sequenza $\tilde{p}_0, \dots, \tilde{p}_s$ gode delle proprietà (a)–(c) viste nella dimostrazione del Teorema 14C.10.)

Il Teorema spettrale

In questo capitolo continuiamo lo studio di autovettori e autovalori, trattando il caso degli spazi vettoriali metrici. Il problema principale stavolta è determinare quando esiste una base *ortonormale* che diagonalizzi o triangolarizzi un dato endomorfismo. Vedremo che una base ortonormale di autovettori esiste nel caso complesso se e solo se l'endomorfismo commuta con la propria aggiunta; nel caso reale se e solo se l'endomorfismo è simmetrico. Dimostreremo poi che esiste una base ortonormale che triangolarizza un dato endomorfismo se e solo se l'endomorfismo ha tutti gli autovalori nel campo; in particolare, tutti gli endomorfismi di spazi vettoriali complessi sono triangolabili. Infine, i Complementi a questo capitolo sono dedicati a dimostrare l'esistenza e unicità della forma canonica di Jordan di un endomorfismo triangolabile.

15.1 Basi ortonormali di autovettori

Nel capitolo precedente abbiamo dimostrato un criterio necessario e sufficiente perché un endomorfismo $T: V \rightarrow V$ di uno spazio vettoriale qualunque fosse diagonalizzabile. Supponiamo invece che V sia uno spazio vettoriale metrico (su \mathbb{R} o su \mathbb{C}); è naturale chiedersi quando esiste una base *ortonormale* di autovettori di T . Come vedremo, la risposta consisterà in due criteri (uno sui complessi e l'altro sui reali) molto semplici da verificare.

Per capire da dove partire, supponiamo che $\mathcal{B} = \{v_1, \dots, v_n\}$ sia una base ortonormale di autovettori per un endomorfismo $T: V \rightarrow V$ di uno spazio vettoriale metrico V . Indichiamo con $\lambda_j \in \mathbb{K}$ l'autovalore relativo a v_j , in modo che $T(v_j) = \lambda_j v_j$ per $j = 1, \dots, n$. Prendiamo $v \in V$; sappiamo che rispetto alla base \mathcal{B} il vettore v si scrive come

$$v = \sum_{j=1}^n \langle v, v_j \rangle v_j.$$

Applicando T troviamo

$$T(v) = \sum_{j=1}^n \lambda_j \langle v, v_j \rangle v_j.$$

Ma allora

$$\begin{aligned}\|T(v)\|^2 &= \sum_{j=1}^n |\lambda_j|^2 |\langle v, v_j \rangle|^2 = \sum_{j=1}^n |\langle v, \lambda_j v_j \rangle|^2 \\ &= \sum_{j=1}^n |\langle v, T(v_j) \rangle|^2 = \sum_{j=1}^n |\langle T^*(v), v_j \rangle|^2 = \|T^*(v)\|^2,\end{aligned}$$

dove $T^*: V \rightarrow V$ è l'aggiunta di T (vedi il Paragrafo 12.7), e abbiamo usato la Proposizione 12.4 e il Corollario 12.5.

Dunque l'esistenza di una base ortonormale di autovettori implica che necessariamente $\|T(v)\| = \|T^*(v)\|$ per ogni $v \in V$. Questo suggerisce la seguente

Definizione 15.1 Un endomorfismo $T: V \rightarrow V$ di uno spazio vettoriale metrico V è detto *normale* se $\|T(v)\| = \|T^*(v)\|$ per ogni $v \in V$.

ESEMPIO 15.1 Gli endomorfismi autoaggiunti (Paragrafo 12.8) sono chiaramente normali, come pure le isometrie (perché? Vedi l'Esercizio 12.29).

Vi è un modo semplice per verificare se un endomorfismo è normale:

Proposizione 15.1 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale metrico V . Allora le seguenti affermazioni sono equivalenti:

- (i) T è normale;
- (ii) $\langle T(v), T(w) \rangle = \langle T^*(v), T^*(w) \rangle$ per ogni $v, w \in V$;
- (iii) $T \circ T^* = T^* \circ T$.

Dimostrazione. (i) \implies (ii). Supponiamo di essere sui reali; la dimostrazione nel caso complesso è analoga. Ricordando la Proposizione 12.2.(vi) per ogni $v, w \in V$ otteniamo

$$\begin{aligned}\langle T(v), T(w) \rangle &= \frac{1}{4} [\|T(v) + T(w)\|^2 - \|T(v) - T(w)\|^2] \\ &= \frac{1}{4} [\|T(v+w)\|^2 - \|T(v-w)\|^2] = \frac{1}{4} [\|T^*(v+w)\|^2 - \|T^*(v-w)\|^2] \\ &= \frac{1}{4} [\|T^*(v) + T^*(w)\|^2 - \|T^*(v) - T^*(w)\|^2] \\ &= \langle T^*(v), T^*(w) \rangle,\end{aligned}$$

come voluto.

(ii) \implies (iii). Infatti la (ii) implica che

$$\langle v, (T^* \circ T)(w) \rangle = \langle T(v), T(w) \rangle = \langle T^*(v), T^*(w) \rangle = \langle v, (T \circ T^*)(w) \rangle$$

per ogni $v, w \in V$, e quindi $T^* \circ T = T \circ T^*$ (Proposizione 12.1).

(iii) \implies (i). Infatti

$$\begin{aligned}\|T(v)\|^2 &= \langle T(v), T(v) \rangle = \langle v, (T^* \circ T)(v) \rangle = \langle v, (T \circ T^*)(v) \rangle \\ &= \langle T^*(v), T^*(v) \rangle = \|T^*(v)\|^2\end{aligned}$$

per ogni $v \in V$. □

Dunque per vedere se un endomorfismo è normale basta controllare se commuta con la sua aggiunta.

ESEMPIO 15.2 Consideriamo l'endomorfismo $T: \mathbb{C}^2 \rightarrow \mathbb{C}^2$ rappresentato rispetto alla base canonica di \mathbb{C}^2 (che è una base ortonormale per il prodotto hermitiano canonico) dalla matrice

$$A = \begin{bmatrix} 1 & i \\ 1 & 1 \end{bmatrix}.$$

Si verifica facilmente che

$$AA^H = \begin{bmatrix} 2 & 1+i \\ 1-i & 2 \end{bmatrix} = A^H A,$$

per cui $T \circ T^* = T^* \circ T$ (vedi l'Esempio 12.18), e T è normale pur non essendo unitario o hermitiano.

La normalità è esattamente la condizione che ci assicura l'esistenza di una base ortonormale di autovettori, non appena l'endomorfismo ha tutti gli autovalori nel campo. Per dimostrarlo ci serve un risultato di collegamento fra endomorfismi normali e autovettori.

Proposizione 15.2 Sia $T: V \rightarrow V$ un endomorfismo normale di uno spazio vettoriale metrico. Allora:

- (i) il vettore $v_0 \in V$ è autovettore di T relativo all'autovalore λ_0 se e solo se v_0 è autovettore di T^* relativo all'autovalore $\bar{\lambda}_0$;
- (ii) autovettori di T relativi ad autovalori distinti sono automaticamente ortogonali.

Dimostrazione. (i) Sia $v \in V$ un vettore qualunque. Allora si ha

$$\begin{aligned}\|T(v) - \lambda_0 v\|^2 &= \langle T(v) - \lambda_0 v, T(v) - \lambda_0 v \rangle \\ &= \langle T(v), T(v) \rangle - \lambda_0 \langle v, T(v) \rangle - \bar{\lambda}_0 \langle T(v), v \rangle + \lambda_0 \bar{\lambda}_0 \langle v, v \rangle \\ &= \langle T^*(v), T^*(v) \rangle - \lambda_0 \langle T^*(v), v \rangle - \bar{\lambda}_0 \langle v, T^*(v) \rangle + \lambda_0 \bar{\lambda}_0 \langle v, v \rangle \\ &= \|T^*(v) - \bar{\lambda}_0 v\|^2.\end{aligned}$$

Quindi $v_0 \in V$ è un autovettore di T relativo a λ_0 se e solo se $\|T(v_0) - \lambda_0 v_0\| = 0$ e solo se $\|T^*(v_0) - \bar{\lambda}_0 v_0\| = 0$ se e solo se v_0 è un autovettore di T^* relativo a $\bar{\lambda}_0$.

(ii) Infatti prendiamo $v \in V_\lambda$ e $w \in V_\mu$. Grazie a (i) troviamo che

$$\lambda \langle v, w \rangle = \langle T(v), w \rangle = \langle v, T^*(w) \rangle = \langle v, \bar{\mu}w \rangle = \bar{\mu} \langle v, w \rangle,$$

ed essendo $\lambda \neq \mu$ segue che $\langle v, w \rangle = 0$. \square

E allora abbiamo il

Teorema 15.3 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale metrico V sul campo \mathbb{K} . Allora esiste una base ortonormale \mathcal{B} di V composta da autovettori di T se e solo se T è normale e ha tutti gli autovalori in \mathbb{K} .

Dimostrazione. Supponiamo che esista una base ortonormale di autovettori di T . Allora T ha necessariamente tutti gli autovalori in \mathbb{K} (Lemma 14.4) e il ragionamento che ci ha portato alla Definizione 15.1 mostra che T è normale.

Viceversa, supponiamo T normale e con tutti gli autovalori in \mathbb{K} , e procediamo per induzione sulla dimensione di V . Se $\dim V = 1$ non c'è nulla da dimostrare; sia allora $\dim V = n > 1$, e supponiamo l'asserto vero per tutti gli spazi vettoriali metrici di dimensione $n - 1$.

Sia $\lambda_1 \in \text{sp}(T)$ un autovalore di T (per ipotesi almeno uno esiste), e prendiamo un autovettore $v_1 \neq 0$ relativo a λ_1 ; possiamo supporre che $\|v_1\| = 1$. Sia $W = \text{Span}(v_1)^\perp$; vogliamo dimostrare che $T(W) \subseteq W$. Ma infatti sia $w \in W$: allora la Proposizione 15.2 implica che

$$\langle v_1, T(w) \rangle = \langle T^*(v_1), w \rangle = \langle \bar{\lambda}_1 v_1, w \rangle = \bar{\lambda}_1 \langle v_1, w \rangle = 0,$$

per cui $T(W) \subseteq W$. Dunque $T|_W: W \rightarrow W$ è un endomorfismo normale di uno spazio vettoriale metrico di dimensione $n - 1$. Inoltre, l'Esercizio 14.13 ci assicura che anche $T|_W$ ha tutti gli autovalori in \mathbb{K} . Per l'ipotesi induttiva, quindi, esiste una base ortonormale $\{v_2, \dots, v_n\}$ di W composta da autovettori di T ; e dunque $\{v_1, v_2, \dots, v_n\}$ è una base ortonormale di V composta da autovettori di T . \square

Nel caso complesso, il Teorema fondamentale dell'algebra ci permette di semplificare ulteriormente questo criterio:

Corollario 15.4 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale metrico su \mathbb{C} . Allora esiste una base ortonormale di V composta da autovettori di T se e solo se T è normale. In particolare, gli endomorfismi hermitiani e gli endomorfismi unitari ammettono basi ortonormali di autovettori.

Dimostrazione. Infatti ogni endomorfismo di uno spazio vettoriale complesso ha tutti gli autovalori nel campo, grazie al Teorema fondamentale dell'algebra, e gli endomorfismi hermitiani e unitari sono automaticamente normali (Esempio 15.1). \square

Corollario 15.5 Sia $A \in M_{n,n}(\mathbb{C})$ una matrice quadrata a coefficienti complessi. Allora le seguenti affermazioni sono equivalenti:

- (i) esiste una base ortonormale di \mathbb{C}^n composta da autovettori di L_A ;
- (ii) esiste una matrice unitaria $U \in U(n)$ tale che $U^{-1}AU = U^H AU$ sia diagonale;
- (iii) si ha $A^H A = AA^H$.

Dimostrazione. L'equivalenza di (i) e (iii) segue dal Corollario 15.4 e dall'Esempio 12.18. La matrice di cambiamento di base dalla base canonica a una base ortonormale \mathcal{B} contiene per colonne i vettori di \mathcal{B} , per cui è necessariamente unitaria (Corollario 12.22). L'equivalenza di (i) e (ii) segue allora dal Corollario 14.2. \square

ESEMPIO 15.3 L'endomorfismo $T: \mathbb{C}^2 \rightarrow \mathbb{C}^2$ dell'Esempio 15.2 è normale, per cui è diagonalizzabile. Cerchiamo una base ortonormale di autovettori. Cominciamo col calcolare il polinomio caratteristico; si trova

$$p_T(\lambda) = \lambda^2 - 2\lambda + 1 - i,$$

le cui radici sono $\lambda_\pm = 1 \pm (1+i)/\sqrt{2}$, per cui $\text{sp}(T) = \{1 + (1+i)/\sqrt{2}, 1 - (1+i)/\sqrt{2}\}$. L'autospazio V_{λ_+} è il nucleo della matrice

$$\begin{vmatrix} -\frac{1}{\sqrt{2}}(1+i) & i \\ 1 & -\frac{1}{\sqrt{2}}(1+i) \end{vmatrix};$$

risolvendo si trova

$$V_{\lambda_+} = \text{Span} \left(\begin{vmatrix} 1 \\ \frac{1}{\sqrt{2}}(1-i) \end{vmatrix} \right) = \text{Span} \left(\begin{vmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{2}(1-i) \end{vmatrix} \right),$$

dove nell'ultimo passaggio abbiamo semplicemente reso di lunghezza unitaria il vettore della base. Il secondo autospazio è invece il nucleo della matrice

$$\begin{vmatrix} \frac{1}{\sqrt{2}}(1+i) & i \\ 1 & \frac{1}{\sqrt{2}}(1+i) \end{vmatrix},$$

per cui

$$V_{\lambda_-} = \text{Span} \left(\begin{vmatrix} 1 \\ -\frac{1}{\sqrt{2}}(1-i) \end{vmatrix} \right) = \text{Span} \left(\begin{vmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{2}(1-i) \end{vmatrix} \right);$$

nota che, coerentemente con la Proposizione 15.2, V_+ e V_- sono automaticamente ortogonali. La base ortonormale cercata quindi è

$$\left\{ \begin{vmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{2}(1-i) \end{vmatrix}, \begin{vmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{2}(1-i) \end{vmatrix} \right\}.$$

Anche nel caso reale c'è un enunciato molto preciso, che caratterizza tutti gli endomorfismi diagonalizzabili da una base ortonormale. Per dimostrarlo ci serve il

Lemma 15.6 Sia $T: V \rightarrow V$ un endomorfismo hermitiano di uno spazio vettoriale metrico V complesso. Allora tutti gli autovalori di T sono numeri reali.

Dimostrazione. Sia $\lambda_0 \in \mathbb{C}$ un autovalore di T , e $v_0 \neq 0$ un suo autovettore. Allora

$$\lambda_0 \|v_0\|^2 = \langle \lambda_0 v_0, v_0 \rangle = \langle T(v_0), v_0 \rangle = \langle v_0, T(v_0) \rangle = \langle v_0, \lambda_0 v_0 \rangle = \overline{\lambda_0} \|v_0\|^2.$$

Quindi $\lambda_0 = \overline{\lambda_0}$, cioè λ_0 è reale. \square

Lemma 15.7 Sia $T: V \rightarrow V$ un endomorfismo simmetrico di uno spazio vettoriale metrico V reale. Allora T ha tutti gli autovalori in \mathbb{R} .

Dimostrazione. Fissiamo una base ortonormale \mathcal{B} di V , e sia $A \in M_{n,n}(\mathbb{R})$ la matrice simmetrica (vedi l'Osservazione 12.13) che rappresenta T rispetto a questa base. Praticamente per definizione, gli autovalori di T sono gli stessi di $L_A: \mathbb{R}^n \rightarrow \mathbb{R}^n$. Ora possiamo considerare L_A anche come un endomorfismo di \mathbb{C}^n , rappresentato esattamente dalla stessa matrice. Essendo A simmetrica, è in particolare hermitiana; quindi $L_A: \mathbb{C}^n \rightarrow \mathbb{C}^n$ è un endomorfismo hermitiano rispetto al prodotto hermitiano canonico di \mathbb{C}^n , e dunque ha tutti gli autovalori in \mathbb{R} (Lemma 15.6). \square

Di conseguenza gli endomorfismi simmetrici sono sempre diagonalizzabili su \mathbb{R} tramite una base ortonormale. Il risultato notevole, noto come *Teorema spettrale*, è che non ce ne sono altri:

Teorema 15.8 (spettrale) Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale metrico V su \mathbb{R} . Allora esiste una base ortonormale di V composta da autovettori di T se e solo se T è simmetrico.

Dimostrazione. Se T è simmetrico, è ovviamente normale e per il Lemma 15.7 ha tutti gli autovalori in \mathbb{R} ; quindi il Teorema 15.3 ci fornisce una base ortonormale di V composta da autovettori di T .

Viceversa, supponiamo che una tale base \mathcal{B} esista. Allora rispetto a questa base T è rappresentato da una matrice simmetrica (addirittura diagonale); essendo la base ortonormale, per l'Osservazione 12.13 questo vuol dire che T è simmetrico. \square

Corollario 15.9 Sia $A \in M_{n,n}(\mathbb{R})$ una matrice quadrata a coefficienti reali. Allora le seguenti affermazioni sono equivalenti:

- (i) esiste una base ortonormale di \mathbb{R}^n composta da autovettori di L_A ;
- (ii) esiste una matrice ortogonale $U \in O(n)$ tale che $U^{-1}AU = U^T AU$ è diagonale;
- (iii) A è simmetrica.

Dimostrazione. Segue dal Teorema 15.8 procedendo come per il Corollario 15.5. \square

Il Teorema 15.8 è la principale ragione dell'importanza degli endomorfismi simmetrici: sono gli unici endomorfismi diagonalizzabili su \mathbb{R} tramite basi ortonormali. Questo risultato è anche ciò che ci ha spinto a studiare i prodotti hermitiani: infatti la dimostrazione dipende in maniera essenziale dai numeri complessi.

ESEMPIO 15.4 Prendiamo la matrice simmetrica

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

che ora sappiamo essere diagonalizzabile; vogliamo una base ortonormale di autovettori. Prima di tutto troviamo gli autovalori:

$$p_A(\lambda) = \det(A - \lambda I_3) = (1 - \lambda)[(1 - \lambda)^2 - 5] = (1 - \lambda)(1 + \sqrt{5} - \lambda)(1 - \sqrt{5} - \lambda),$$

per cui $\text{sp}(A) = \{1, 1 + \sqrt{5}, 1 - \sqrt{5}\}$. Risolvendo i relativi sistemi troviamo

$$V_1 = \text{Span} \left(\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} \right), \quad V_{1+\sqrt{5}} = \text{Span} \left(\begin{pmatrix} 1 \\ 2 \\ \sqrt{5} \end{pmatrix} \right), \quad V_{1-\sqrt{5}} = \text{Span} \left(\begin{pmatrix} 1 \\ 2 \\ -\sqrt{5} \end{pmatrix} \right).$$

Quindi una base ortonormale di autovettori di A è

$$\left\{ \begin{pmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{pmatrix}, \begin{pmatrix} 1/\sqrt{10} \\ 2/\sqrt{10} \\ 1/\sqrt{2} \end{pmatrix}, \begin{pmatrix} 1/\sqrt{10} \\ 2/\sqrt{10} \\ -1/\sqrt{2} \end{pmatrix} \right\}.$$

Osservazione 15.1 Ovviamente esistono anche endomorfismi diagonalizzabili non simmetrici; semplicemente, la base di autovettori non è mai ortonormale.

ESEMPIO 15.5 L'endomorfismo $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ rappresentato dalla matrice

$$\begin{pmatrix} 3 & 0 \\ 1 & 2 \end{pmatrix}$$

è diagonalizzabile, in quanto ha esattamente due autovalori distinti ($\text{sp}(T) = \{3, 2\}$), ma non è simmetrico. Una base di autovettori è data per esempio da

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

15.2 Endomorfismi triangolabili

Se l'endomorfismo $T: V \rightarrow V$ non è normale, una base ortonormale di autovalori non esiste. Una variante della tecnica utilizzata nella dimostrazione del Teorema 15.3 ci permette però di trovare una base ortonormale che triangolarizza T non appena tutti gli autovalori sono nel campo:

Teorema 15.10 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale V sul campo \mathbb{K} . Allora T è triangolabile se e solo se ha tutti gli autovalori in \mathbb{K} . Inoltre, se V è uno spazio vettoriale metrico la base che triangolarizza T può essere scelta ortonormale.

Dimostrazione. Se T è triangolabile sappiamo già che deve avere tutti gli autovalori in \mathbb{K} (Lemma 14.4); quindi dobbiamo dimostrare solo il viceversa.

Siccome (Proposizione 12.15) su V è sempre possibile mettere un prodotto scalare (o hermitiano) definito positivo, possiamo direttamente supporre V metrico e cercare una base ortonormale che triangolarizzi T .

Posto $n = \dim V$, siano $\lambda_1, \dots, \lambda_n \in \mathbb{K}$ gli autovalori di T , ripetuti secondo la rispettiva molteplicità algebrica. Vogliamo una base ortonormale $\mathcal{B} = \{v_1, \dots, v_n\}$ di V tale che $T(v_1) = \lambda_1 v_1$ e

$$T(v_{j+1}) - \lambda_{j+1} v_{j+1} \in V_j \quad (15.1)$$

per $j = 1, \dots, n-1$, dove $V_j = \text{Span}(v_1, \dots, v_j)$. Infatti (15.1) vale se e solo se esistono $a_{ij} \in \mathbb{K}$ tali che

$$T(v_{j+1}) = \lambda_{j+1} v_{j+1} + a_{1,j+1} v_1 + \dots + a_{j,j+1} v_j,$$

e quindi la matrice che rappresenta T rispetto a \mathcal{B} è triangolare superiore.

Come v_1 prendiamo un autovettore di T relativo a λ_1 di norma unitaria. Supponiamo ora di aver determinato v_1, \dots, v_j come richiesto; vogliamo trovare v_{j+1} . Sia $W_j = V_j^\perp$: indichiamo con $P_j: V \rightarrow W_j$ la proiezione ortogonale su W_j , e poniamo $T_j = P_j \circ T|_{W_j}: W_j \rightarrow W_j$. Se $\{w_{j+1}, \dots, w_n\}$ è una base qualunque di W_j rispetto alla base $\{v_1, \dots, v_j, w_{j+1}, \dots, w_n\}$ di V l'endomorfismo T è rappresentato da una matrice della forma

$$\left| \begin{array}{c|c} A & B \\ \hline O & C \end{array} \right|,$$

dove $A \in M_{j,j}(\mathbb{K})$ è triangolare superiore con $\lambda_1, \dots, \lambda_j$ sulla diagonale principale. Siccome $P_j(v_h) = 0$ per $1 \leq h \leq j$ e $P_j(w_k) = w_k$ per $j+1 \leq k \leq n$, si verifica facilmente (esercizio) che C è la matrice che rappresenta T_j rispetto alla base $\{w_{j+1}, \dots, w_n\}$ di W_j . Grazie all'Esercizio 9.10 vediamo che

$$(\lambda_1 - \lambda) \cdots (\lambda_n - \lambda) = p_T(\lambda) = \det(A - \lambda I) \det(C - \lambda I) = (\lambda_1 - \lambda) \cdots (\lambda_j - \lambda) p_{T_j}(\lambda)$$

quindi gli autovalori di T_j sono $\lambda_{j+1}, \dots, \lambda_n$. Prendiamo allora $v_{j+1} \in W_j$ un autovettore di T_j relativo a λ_{j+1} di norma unitaria. In particolare $P_j(T(v_{j+1})) = \lambda_{j+1} v_{j+1}$ cioè (per definizione di proiezione ortogonale) $T(v_{j+1}) - \lambda_{j+1} v_{j+1} \in W_j^\perp = V_j$, come voluto. Procedendo in questo modo costruiamo la base cercata.

Corollario 15.11 Sia A una matrice quadrata di ordine n a coefficienti complessi. Allora esiste sempre una matrice unitaria $U \in U(n)$ tale che $U^{-1}AU = U^H A U$ sia triangolare superiore.

Dimostrazione. Segue dal Teorema 15.10 procedendo come per il Corollario 15.5. \square

Corollario 15.12 Sia A una matrice quadrata di ordine n a coefficienti reali. Allora le seguenti affermazioni sono equivalenti:

- (i) L_A è triangolabile;
- (ii) L_A ha tutti gli autovalori in \mathbb{R} ;
- (iii) esiste una matrice invertibile $B \in GL_n(\mathbb{R})$ tale che $B^{-1}AB$ sia triangolare superiore;
- (iv) esiste una matrice ortogonale $U \in O(n)$ tale che $U^{-1}AU = U^T A U$ sia triangolare superiore.

Dimostrazione. Di nuovo segue subito dal Teorema 15.10. \square

La dimostrazione del Teorema 15.10 suggerisce anche una tecnica per trovare una base che triangolarizza un dato endomorfismo.

ESEMPIO 15.6 Vogliamo trovare una base $\{v_1, v_2, v_3\}$ ortonormale rispetto al prodotto scalare canonico che triangolarizzi l'endomorfismo $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ rappresentato dalla matrice

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

Cominciamo col cercare gli autovalori di T . Si ha

$$p_T(\lambda) = \det(A - \lambda I_3) = -\lambda^3 + 2\lambda^2 - \lambda = (-\lambda)(1 - \lambda)^2,$$

per cui $\text{sp}(T) = \{0, 1\}$. L'autovalore 0 ha molteplicità algebrica (e quindi geometrica) 1; l'autovalore 1 invece ha molteplicità algebrica 2 ma molteplicità geometrica 1, in quanto

$$\text{rg}(A - I_3) = \text{rg} \begin{pmatrix} 2 & 1 & 2 \\ -2 & -1 & -2 \\ -1 & 0 & -2 \end{pmatrix} = 2.$$

In particolare, T non è diagonalizzabile. Per determinare una base ortonormale che triangolarizzi T cominciamo col trovare un autovettore relativo all'autovalore 1. Risolvendo il sistema $Ax = x$ otteniamo

$$V_1 = \text{Span} \left(\begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix} \right),$$

per cui possiamo prendere come primo vettore della nostra base

$$v_1 = \begin{pmatrix} -2/3 \\ 2/3 \\ 1/3 \end{pmatrix}.$$

Ora, l'Esercizio 12.27 ci dice che la matrice che rappresenta rispetto alla base canonica la proiezione ortogonale su V_1 è $v_1 v_1^T$ (nota che $v_1^T v_1 = \|v_1\|^2 = 1$); quindi la matrice che rappresenta la proiezione ortogonale P_1 su $W_1 = V_1^\perp$ è $I_3 - v_1 v_1^T$ (in quanto la somma delle due proiezioni ortogonali deve dare l'identità). Quindi la matrice che rappresenta $T_1 = P_1 \circ T$ è

$$A_1 = (I_3 - v_1 v_1^T)A = \begin{pmatrix} 5/9 & 4/9 & 2/9 \\ 4/9 & 5/9 & -2/9 \\ 2/9 & -2/9 & 8/9 \end{pmatrix} \begin{vmatrix} 3 & 1 & 2 \\ -2 & 0 & -2 \\ -1 & 0 & -1 \end{vmatrix} = \begin{pmatrix} 5/9 & 5/9 & 0 \\ 4/9 & 4/9 & 0 \\ 2/9 & 2/9 & 0 \end{pmatrix}.$$

Vogliamo ora un autovettore v_2 di T_1 relativo a $\lambda_2 = 1$ di norma unitaria; nota che siccome

$$v_2 = \lambda_2^{-1} P_1(T(v_2)) \in \text{Im } P_1 = V_1^\perp,$$

il vettore v_2 è automaticamente ortogonale a v_1 . Calcolando l'autospazio relativo a 1 di A_1 troviamo

$$v_2 = \begin{pmatrix} 5/3\sqrt{5} \\ 4/3\sqrt{5} \\ 2/3\sqrt{5} \end{pmatrix}.$$

Posto $V_2 = \text{Span}(v_1, v_2)$, l'Esercizio 12.27 dice che la matrice associata alla proiezione ortogonale su V_2 è BB^T , dove $B = [v_1 \ v_2]$ (nota che $B^T B = I$ in quanto $\{v_1, v_2\}$ è una base ortonormale di V_2). Quindi la matrice associata a $T_2 = P_2 \circ T$ è

$$A_2 = (I_3 - BB^T)A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1/5 & -2/5 \\ 0 & -2/5 & 4/5 \end{pmatrix} \begin{vmatrix} 3 & 1 & 2 \\ -2 & 0 & -2 \\ -1 & 0 & -1 \end{vmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Questo vuol dire (perché?) che $\text{Im } T \subseteq V_2$; quindi il terzo vettore della base cercata può essere un qualunque vettore di lunghezza unitaria ortogonale a V_2 . Per trovarlo, prendiamo $w_3 \notin V_2$ qualunque; allora $v_3 = P_2(w_3)/\|P_2(w_3)\|$ è come voluto. Per esempio, se $w_3 = e_3$ troviamo

$$v_3 = \begin{pmatrix} 0 \\ -1/\sqrt{5} \\ 2/\sqrt{5} \end{pmatrix}.$$

Rispetto alla base $\{v_1, v_2, v_3\}$ l'endomorfismo T è rappresentato dalla matrice

$$A' = C^{-1}AC = C^T AC = \begin{pmatrix} 1 & -9/\sqrt{5} & -16/3\sqrt{5} \\ 0 & 1 & -1/3 \\ 0 & 0 & 0 \end{pmatrix},$$

dove

$$C = \begin{pmatrix} -2/3 & 5/3\sqrt{5} & 0 \\ 2/3 & 4/3\sqrt{5} & -1/\sqrt{5} \\ 1/3 & 2/3\sqrt{5} & 2/\sqrt{5} \end{pmatrix}.$$

Osservazione 15.2 In generale non c'è alcuna relazione fra la matrice triangolare ottenuta nel Teorema 15.10 (matrice strettamente legata a T come endomorfismo) e quella che si ottiene effettuando un'eliminazione di Gauss su una matrice che rappresenta T rispetto a qualche base. L'eliminazione di Gauss cambia drasticamente l'immagine della matrice, e quindi non c'entra nulla con la triangolabilità di un endomorfismo. In particolare, i pivot e gli autovalori di una matrice non hanno nulla a che spartire.

Osservazione 15.3 In realtà, si può fare più di quanto indicato nel Teorema 15.10. Non solo esiste una base che triangolarizza l'endomorfismo T , ma possiamo scegliere questa base in modo che la matrice associata a T abbia come elementi non nulli fuori dalla diagonale principale soltanto degli 1 posti nella diagonale subito soprastante, come succede per esempio nella matrice

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

Inoltre questa matrice è univocamente determinata (a meno dell'ordine dei blocchi), ed è detta *forma canonica di Jordan* dell'endomorfismo T . Se sei interessato alla dimostrazione dell'esistenza e unicità della forma canonica di Jordan leggi i Complementi a questo capitolo.

Esercizi

15.1 Trova una base ortonormale di \mathbb{R}^3 che diagonalizzi la matrice

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

15.2 Trova, se esistono, basi ortonormali di \mathbb{C}^3 rispetto al prodotto hermitiano canonico che diagonalizzino le seguenti matrici:

$$\begin{pmatrix} 2+17i & 2-i & -4(2-i) \\ 2-i & 2+17i & -4(2-i) \\ -4(2-i) & -4(2-i) & 32+2i \end{pmatrix}, \quad \begin{pmatrix} 0 & 2+i & 0 \\ 2-i & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

15.3 Trova un esempio di endomorfismo *non* normale $T: V \rightarrow V$ di uno spazio vettoriale metrico V tale che se $v_0 \neq 0$ è un autovettore di T e $W_0 = \text{Span}(v_0)$ allora $T(W_0)$ *non* è contenuto in W_0 .

15.4 Dimostra che tutti gli autovalori di un endomorfismo unitario sono numeri complessi di modulo uno.

15.5 Dimostra che una matrice ortogonale $A \in O(n)$ è diagonalizzabile se e solo se $\text{sp}_{\mathbb{C}}(A) \subseteq \{1, -1\}$, dove $\text{sp}_{\mathbb{C}}(A)$ indica lo spettro di $L_A: \mathbb{C}^n \rightarrow \mathbb{C}^n$.

15.6 Sia $T: V \rightarrow V$ un endomorfismo normale su uno spazio vettoriale metrico complesso V . Dimostra che T è hermitiano se e solo se $\text{sp}(T) \subset \mathbb{R}$, e che T è unitario se e solo se ha tutti gli autovalori di modulo uno.

15.7 Sia $T: V \rightarrow V$ un endomorfismo diagonalizzabile di uno spazio vettoriale su \mathbb{R} . Dimostra che esiste un prodotto scalare definito positivo su V rispetto a cui T è simmetrico.

15.8 Sia $A \in M_{n,n}(\mathbb{R})$ una matrice antisimmetrica. Dimostra che

$$p_A(\lambda) = (-1)^n p_A(-\lambda),$$

e deduci che se n è dispari allora $\det A = 0$.

15.9 Dimostra che una matrice antisimmetrica $A \in M_{n,n}(\mathbb{R})$ è sempre diagonalizzabile su \mathbb{C} , mentre l'unica matrice antisimmetrica diagonalizzabile su \mathbb{R} è la matrice nulla. (*Suggerimento*: se A fosse diagonalizzabile su \mathbb{R} , necessariamente $\text{sp}(A) \subset \mathbb{R}$. Ma allora sarebbe possibile trovare una matrice ortogonale $U \in O(n)$ in modo che la matrice $U^T A U$ risulti triangolare superiore.)

15.10 Sia $A \in M_{n,n}(\mathbb{R})$ una matrice antisimmetrica. Dimostra che se $\lambda_0 \in \mathbb{R}$ è autovalore di A allora $\lambda_0 = 0$. (*Suggerimento*: l'endomorfismo L_A dell'autospazio relativo a λ_0 è ancora antisimmetrico; applica l'Esercizio precedente.)

15.11 Sia $A \in M_{n,n}(\mathbb{R})$ una matrice antisimmetrica. Dimostra che se $\lambda_0 \in \mathbb{C}^*$ è un autovalore non nullo di $L_A: \mathbb{C}^n \rightarrow \mathbb{C}^n$ allora λ_0 è immaginario puro. (*Suggerimento*: A^2 è simmetrica, per cui $\lambda_0^2 \in \text{sp}(A^2) \subset \mathbb{R}$.)

15.12 Trova, se esistono, basi ortonormali di \mathbb{R}^n rispetto al prodotto scalare canonico che triangolarizzino le matrici

$$\begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 2 & 4 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

15.13 Trova una base ortonormale di \mathbb{C}^3 rispetto al prodotto hermitiano canonico che triangolarizzi la matrice

$$\begin{pmatrix} 1+i & i & 1 \\ -1+i & i & -1 \\ 1-i & 2-i & 3 \end{pmatrix}.$$

15.14 Trova per quali valori dei parametri $a, b \in \mathbb{R}$ le seguenti matrici sono triangolari (diagonalizzabili, diagonalizzabili da una base ortonormale) su \mathbb{R} , e ove possibile

determina una matrice ortogonale che le triangolarizzi (o diagonalizzi):

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & a \\ 0 & b & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 0 \\ 1 & a & b \\ 0 & -1 & 1 \end{pmatrix}, \quad \begin{pmatrix} a & 0 & 0 \\ b & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 1 \\ 0 & b & 0 \\ a & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} a+1 & 2 & a-1 \\ 0 & a & 0 \\ a & a & -1 \end{pmatrix}.$$

15.15 Per ogni $p \in \mathbb{R}^*$ trova una matrice quadrata $A \in M_{2,2}(\mathbb{R})$ su cui si possa effettuare l'eliminazione di Gauss senza scambi di righe con pivot $p_1 = p$ e $p_2 = 2/p$ e tale che $\text{sp}(A) = \{1, 2\}$.

15.16 Sia $A \in M_{2,2}(\mathbb{R})$. Dimostra che se $\det A < 0$ allora A è necessariamente diagonalizzabile, e che se $\det A = 0$ allora A è necessariamente triangolabile. Trova due matrici $B, C \in M_{2,2}(\mathbb{R})$ tali che B sia non triangolabile e abbia $\det B > 0$, e C sia non diagonalizzabile e abbia $\det C = 0$.

15.17 Sia V uno spazio vettoriale metrico, e $u \in V$ con $\|u\| = 1$. La riflessione rispetto all'iperpiano $V_1 = \text{Span}(u)^\perp$ è l'endomorfismo $S_u: V \rightarrow V$ dato da

$$S_u(v) = v - 2\langle v, u \rangle u.$$

Dimostra che $S_u^* = S_u = S_u^{-1}$ (per cui S_u è un'isometria autoaggiunta) e trova una base ortonormale di autovettori per S_u .

15.18 Scrivi la matrice $A \in M_{3,3}(\mathbb{R})$ associata (rispetto alla base canonica) alla riflessione $S: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ rispetto al piano π_1 di equazione cartesiana $x + 2y - z = 0$, e trova una matrice ortogonale $U \in O(3)$ tale che $U^T A U$ sia diagonale.

15.19 Dopo aver riletto l'Osservazione 14.8, dimostra che se $T: V \rightarrow V$ è un endomorfismo triangolabile di uno spazio vettoriale V sul campo \mathbb{K} allora $p_T(T) = 0$. (*Suggerimento*: se $A \in M_{n,n}(\mathbb{K})$ è una matrice triangolare superiore che rappresenta T rispetto a un'opportuna base $\{v_1, \dots, v_n\}$ di V , dimostra per induzione su j che $(A - \lambda_1 I_n) \cdots (A - \lambda_j I_n)v = 0$ per ogni $v \in V_j = \text{Span}(v_1, \dots, v_j)$, dove $\lambda_1, \dots, \lambda_n$ sono gli elementi sulla diagonale di A .)

15.20 Dimostra il *Teorema di Hamilton-Cayley*: se $T: V \rightarrow V$ è un endomorfismo di uno spazio vettoriale su \mathbb{R} o su \mathbb{C} allora $p_T(T) = 0$.

15.21 Sia V lo spazio vettoriale delle matrici antisimmetriche di ordine 3, e indichiamo con $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$ il prodotto scalare definito positivo dell'Esempio 12.8. Sia $T: V \rightarrow V$ l'endomorfismo dato da $T(X) = AX + XA$, dove $A \in M_{3,3}(\mathbb{R})$ è la matrice

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

Dimostra che T è simmetrico rispetto a $\langle \cdot, \cdot \rangle$, e determina una base ortonormale di autovettori per T .

15.22 Sia $A \in M_{n,n}(\mathbb{R})$ simmetrica. Dimostra che esiste una matrice simmetrica $B \in M_{n,n}(\mathbb{R})$ tale che $B^3 = A$. Trova una matrice $A \in M_{2,2}(\mathbb{R})$ simmetrica per cui non esista una matrice $C \in M_{2,2}(\mathbb{R})$ tale che $C^2 = A$.

COMPLEMENTI

15C.1 Esistenza della forma canonica di Jordan

Abbiamo visto che se un endomorfismo $T: V \rightarrow V$ di uno spazio vettoriale V ha tutti gli autovalori nel campo allora è triangolabile. In generale, però, cambiando base la matrice triangolare superiore che rappresenta T muta drasticamente. In questi Complementi vogliamo far vedere come sia possibile associare a ogni endomorfismo triangolabile una base rispetto a cui la matrice triangolare superiore che lo rappresenta è particolarmente semplice — ed essenzialmente unica.

Ma cominciamo con due definizioni.

Definizione 15C.1 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale V sul campo \mathbb{K} . Una catena di Jordan per T di lunghezza $m \geq 1$ associata a $\lambda \in \mathbb{K}$ è un insieme $v_1, \dots, v_m \in V$ di vettori linearmente indipendenti tali che

$$T(v_i) = \lambda v_i + v_{i-1}$$

per $i = 1, \dots, m$, dove abbiamo posto $v_0 = O$. Nota che $T(v_1) = \lambda v_1$, per cui λ è necessariamente un autovalore di T .

Definizione 15C.2 Una base di Jordan per un endomorfismo $T: V \rightarrow V$ è una base di V composta dall'unione di un certo numero di catene di Jordan per T a due a due disgiunte.

Osservazione 15C.1 Una base di autovettori è una base di Jordan composta tutta da catene di lunghezza uno; quindi il concetto di base di Jordan in un certo senso generalizza quello di base di autovettori.

Supponiamo che esista una base $\mathcal{B} = \{v_1, \dots, v_m\}$ di V composta da un'unica catena di Jordan per T . Rispetto a \mathcal{B} l'endomorfismo T è rappresentato dalla matrice

$$J_m(\lambda) = \begin{vmatrix} \lambda & 1 & & & \\ & \lambda & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda \end{vmatrix} \in M_{m,m}(\mathbb{K}),$$

dove gli spazi bianchi sono riempiti di zeri. Più in generale, se \mathcal{B} è una base di Jordan per T , allora la matrice J che rappresenta T rispetto a \mathcal{B} è composta da una serie di blocchi di Jordan disposti lungo la diagonale principale, uno per ogni catena di Jordan in \mathcal{B} .

Definizione 15C.3 La matrice $J_m(\lambda)$ si chiama blocco di Jordan relativo a $\lambda \in \mathbb{K}$ di ordine m . Una matrice J composta da blocchi di Jordan disposti lungo la diagonale principale si chiama matrice di Jordan.

In questo paragrafo dimostreremo che ogni endomorfismo triangolabile ammette una base di Jordan, e nel prossimo paragrafo che i blocchi di Jordan in J sono univocamente determinati da T a meno dell'ordine. Ci serve un lemma:

Lemma 15C.1 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale V sul campo \mathbb{K} . Sia $W \subseteq V$ un sottospazio di V tale che $\text{Im } T = T(W) \subseteq W$ e $\text{Ker } T \subseteq W$. Allora $W = V$.

Dimostrazione. Infatti se indichiamo con $S: W \rightarrow W$ la restrizione di T a W abbiamo che $\text{Ker } S = \text{Ker } T$, $\text{Im } S = \text{Im } T$ e quindi

$$\dim V = \dim \text{Ker } T + \dim \text{Im } T = \dim \text{Ker } S + \dim \text{Im } S = \dim W,$$

per cui $V = W$. \square

Teorema 15C.2 Sia $T: V \rightarrow V$ un endomorfismo di uno spazio vettoriale V di dimensione finita sul campo \mathbb{K} . Supponiamo che T abbia tutti gli autovalori in \mathbb{K} . Allora esiste una base di Jordan per T .

Dimostrazione. Per induzione su $n = \dim V$. Se $n = 1$ non c'è nulla da dimostrare; supponiamo allora il Teorema vero per tutti gli spazi di dimensione minore di n .

Sia $\lambda_0 \in \text{sp}(T)$ un autovalore, e prendiamo un autovettore $v_0 \neq O$ relativo a λ_0 . Sia poi $S = T - \lambda_0 \text{id}_V$. Se $S = O$ abbiamo $T = \lambda_0 \text{id}_V$, per cui qualunque base di V è una base di Jordan, ed è fatta. Assumiamo allora $S \neq O$. Poniamo $U = \text{Im } S$, e sia $R = S|_U: U \rightarrow U$ la restrizione di S a U . Nota che $v_0 \in \text{Ker } S$, per cui $m = \dim U < \dim V$.

Ora, siccome $p_S(\lambda) = p_T(\lambda + \lambda_0)$, anche S ha tutti gli autovalori in \mathbb{K} ; grazie all'Esercizio 14.13 questo è vero anche per R . L'ipotesi induttiva, quindi, ci fornisce una base di Jordan \mathcal{B} di U per R . Scriviamo $\mathcal{B} = \mathcal{B}_1 \cup \dots \cup \mathcal{B}_k$, dove $\mathcal{B}_i = \{v_{i,1}, \dots, v_{i,m_i}\}$ è una catena di Jordan in U per R (e quindi per S) di lunghezza m_i relativa a λ_i , e $\mathcal{B}_i \cap \mathcal{B}_j = \emptyset$ se $i \neq j$. Ordiniamo i \mathcal{B}_i in modo che $\lambda_1 = \dots = \lambda_d = 0$ e $\lambda_i \neq 0$ quando $i > d$. I vettori v_{i,m_i} appartengono a $U = \text{Im } S$ per $i = 1, \dots, d$; quindi troviamo $v_{i,m_i+1} \in V$ tali che $S(v_{i,m_i+1}) = v_{i,m_i}$. Per costruzione $v_{1,1}, \dots, v_{d,1} \in \text{Ker } S$; scegliamo $w_1, \dots, w_q \in \text{Ker } S$ in modo che $\{w_1, \dots, w_q, v_{1,1}, \dots, v_{d,1}\}$ sia una base di $\text{Ker } S$, dove $q = \dim \text{Ker } S - d = n - m - d$. Poniamo

$$\mathcal{C} = \mathcal{B} \cup \{v_{1,m_1+1}, \dots, v_{d,m_d+1}, w_1, \dots, w_q\}.$$

Questo insieme contiene $m + d + q = n$ elementi. Inoltre è unione disgiunta di catene di Jordan per S : le catene $\mathcal{B}_i \cup \{v_{i,m_i+1}\}$ per $1 \leq i \leq d$, le catene \mathcal{B}_i per $i > d$, e le catene $\{w_j\}$ di lunghezza 1 associate a 0. Nota che una catena di Jordan per S associata a λ è una catena di Jordan per T associata a $\lambda + \lambda_0$; quindi per concludere ci basta dimostrare che \mathcal{C} è una base di V .

Sia $W = \text{Span}(\mathcal{C})$. Siccome \mathcal{C} contiene una base di $\text{Ker } S$, chiaramente $W \supseteq \text{Ker } S$. Inoltre \mathcal{C} contiene anche una base di $\text{Im } S$; quindi $S(W) \subseteq \text{Im } S \subseteq W$. Di più, si ha anche $\mathcal{B} \subset S(W)$. Infatti per $i = 1, \dots, d$ e $1 \leq j \leq m_i$ abbiamo $v_{i,j} = S(v_{i,j+1})$, per

cui $v_{i,j} \in S(W)$. Poi se $i > d$ si ha $\lambda_i v_{i,1} = S(v_{i,1})$, per cui $S(\lambda_i^{-1} v_{i,1}) = v_{i,1} \in S(W)$. Supponiamo infine che per $i > d$ fissato e $j-1 < m_i$ si abbia $v_{i,j-1} \in S(W)$; allora

$$v_{i,j} = S(\lambda_i^{-1} v_{i,j}) - \lambda_i^{-1} v_{i,j-1} \in S(W),$$

e procedendo per induzione su j abbiamo fatto vedere che $\mathcal{B} \subset S(W)$. Essendo \mathcal{B} una base di $\text{Im } S$ otteniamo quindi $S(V) = S(W)$, e il Lemma 15C.1 ci dà $W = V$. Ma allora \mathcal{C} è un sistema di generatori di V costituito da $n = \dim V$ elementi, per cui \mathcal{C} è una base di Jordan per T . \square

15C.2 Unicità della forma canonica di Jordan

Abbiamo dunque dimostrato l'esistenza di una base di Jordan per ogni endomorfismo con tutti gli autovalori nel campo.

Definizione 15C.4 Sia \mathcal{B} una base di Jordan per l'endomorfismo $T: V \rightarrow V$. La matrice che rappresenta T rispetto a \mathcal{B} è detta *forma canonica di Jordan* per T .

Per definizione, la forma canonica di Jordan è una matrice di Jordan; in particolare è triangolare superiore di un tipo particolarmente semplice. Ma il vero motivo della sua importanza è che, a meno dell'ordine dei blocchi, la forma canonica di Jordan è univocamente associata all'endomorfismo T . Per dimostrarlo ci serve ancora un lemma.

Lemma 15C.3 Sia $J_m(\lambda)$ un blocco di Jordan. Allora

$$\forall k \in \mathbb{N} \quad \text{rg } J_m(\lambda)^k = \begin{cases} m & \text{se } \lambda \neq 0; \\ \max\{0, m-k\} & \text{se } \lambda = 0. \end{cases}$$

Dimostrazione. Se $\lambda \neq 0$ la matrice $J_m(\lambda)$ è invertibile, per cui lei e le sue potenze hanno tutte rango m . D'altra parte si verifica facilmente per induzione su k che

$$J_m(0)^k = \begin{vmatrix} O & I_{m-k} \\ O & O \end{vmatrix}$$

se $k < m$, e che $J_m(0)^k = O$ se $k \geq m$, per cui l'asserto è dimostrato. \square

Teorema 15C.4 Sia $T: V \rightarrow V$ un endomorfismo triangolabile di uno spazio vettoriale V sul campo \mathbb{K} di dimensione n . Sia $\lambda_0 \in \mathbb{K}$ un autovalore di T , e per $k \in \mathbb{N}$ poniamo $s_k = \text{rg}(T - \lambda_0 \text{id}_V)^k$. Allora ogni base di Jordan \mathcal{B} per T contiene esattamente

$$r_k = s_{k+1} - 2s_k + s_{k-1} \quad (15C.1)$$

catene disgiunte associate a λ_0 di lunghezza k . In particolare, la matrice che rappresenta T rispetto a una base di Jordan è univocamente determinata a meno dell'ordine dei blocchi.

Dimostrazione. Sia $J \in M_{n,n}(\mathbb{K})$ la matrice di Jordan che rappresenta T rispetto a \mathcal{B} , e sia $B = J - \lambda_0 I_n$ la matrice che rappresenta $T - \lambda_0 \text{id}_V$ rispetto a \mathcal{B} . Allora B è della forma

$$B = \begin{vmatrix} B_0 & O \\ O & B_1 \end{vmatrix},$$

dove B_0 è formata da tanti blocchi di Jordan relativi a 0 quanti erano i blocchi di Jordan di J relativi a λ_0 (e quindi r_1 blocchi di lunghezza 1, r_2 di lunghezza 2, e così via), mentre $B_1 \in M_{n-\mu, n-\mu}(\mathbb{K})$ è una matrice invertibile, dove $\mu \in \mathbb{N}$ è la molteplicità algebrica di λ_0 come autovalore di T . Il Lemma 15C.3 allora ci dice che

$$s_k = \text{rg}(B^k) = \sum_{m=k+1}^n (m-k)r_m + (n-\mu)$$

per $k = 0, \dots, n$. Quindi

$$\text{rg}(B^k) - \text{rg}(B^{k-1}) = - \sum_{m=k}^n r_m$$

e

$$s_{k+1} - 2s_k + s_{k-1} = (s_{k+1} - s_k) - (s_k - s_{k-1}) = - \sum_{m=k+1}^n r_m + \sum_{m=k}^n r_m = r_k.$$

Dunque il numero di blocchi di Jordan relativi a λ_0 di lunghezza k in J è indipendente dalla base di Jordan scelta, per cui la matrice J è univocamente determinata a meno dell'ordine dei blocchi. \square

Questo Teorema suggerisce anche un metodo per calcolare la forma canonica di Jordan di un endomorfismo $T: V \rightarrow V$: basta trovare gli autovalori $\lambda_1, \dots, \lambda_h \in \mathbb{K}$ di T , e per ognuno dei $\lambda_j \in \text{sp}(T)$ posto $S_j = T - \lambda_j \text{id}_V$ calcolare $\text{rg } S_j$, $\text{rg } S_j^2$, $\text{rg } S_j^3$ e così via fino al minimo $m_j \geq 1$ tale che $\text{rg } S_j^{m_j} = \text{rg } S_j^{m_j+1}$. Allora il numero di blocchi di Jordan relativi a λ_j di ordine k con $1 \leq k \leq m_j$ è

$$\text{rg } S_j^{k+1} - 2 \text{rg } S_j^k + \text{rg } S_j^{k-1}.$$

ESEMPIO 15C.1 Vogliamo trovare la forma canonica di Jordan J dell'endomorfismo $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ rappresentato dalla matrice

$$\begin{vmatrix} 14 & -1 & -1 \\ -24 & 12 & 4 \\ 44 & 6 & 6 \end{vmatrix}.$$

Il polinomio caratteristico di T è

$$p_T(\lambda) = -\lambda^3 + 32\lambda^2 - 320\lambda + 1024 = (16 - \lambda)(8 - \lambda)^2,$$

per cui $\text{sp}(T) = \{8, 16\}$. Poniamo $S_1 = T - 8I_3$ e $S_2 = T - 16I_3$. Allora si trova facilmente che $\text{rg } S_1 = 2$ e $\text{rg } S_1^2 = 1$; siccome $S_1^3 \neq O$, dev'essere per forza $\text{rg } S_1^3 = 1$ e possiamo fermarci. Quindi J contiene

$$(\text{rg } S_1^0 - \text{rg } S_1) - (\text{rg } S_1 - \text{rg } S_1^2) = (3 - 2) - (2 - 1) = 0$$

blocchi di lunghezza 1 relativi a 8, e

$$(\text{rg } S_1 - \text{rg } S_1^2) - (\text{rg } S_1^2 - \text{rg } S_1^3) = (2 - 1) - (1 - 1) = 1$$

blocco di lunghezza 2 relativo a 8. Poi $\text{rg } S_2 = 2 = \text{rg } S_2^2$; quindi J contiene solo un blocco di lunghezza 1 relativo a 16 (e del resto non poteva essere altrimenti, visto che J è di ordine 3). In conclusione,

$$J = \left| \begin{array}{cc|c} 8 & 1 & 0 \\ 0 & 8 & 0 \\ \hline 0 & 0 & 16 \end{array} \right|.$$

La forma canonica di Jordan ci permette anche di stabilire quando due matrici sono simili:

Corollario 15C.5 Siano $A, A' \in M_{n,n}(\mathbb{K})$ due matrici quadrate, dove $\mathbb{K} = \mathbb{R}$ o $\mathbb{K} = \mathbb{C}$. Allora A e A' sono simili se e solo se succede che $\text{sp}(A) = \text{sp}(A')$ e $\text{rg}(A - \lambda_0 I_n)^k = \text{rg}(A' - \lambda_0 I_n)^k$ per ogni autovalore λ_0 e ogni $n \geq k \geq 1$.

Dimostrazione. Se $A' = B^{-1}AB$ per qualche $B \in GL_n(\mathbb{K})$, allora $p_{A'}(\lambda) = p_A(\lambda)$ e

$$\begin{aligned} (A' - \lambda_0 I_n)^k &= (B^{-1}AB - \lambda_0 B^{-1}B)^k = (B^{-1}(A - \lambda_0 I_n)B)^k \\ &= (B^{-1}(A - \lambda_0 I_n)B)(B^{-1}(A - \lambda_0 I_n)B) \cdots (B^{-1}(A - \lambda_0 I_n)B) \\ &= B^{-1}(A - \lambda_0 I_n)^k B, \end{aligned}$$

per cui $\text{sp}(A) = \text{sp}(A')$ e $\text{rg}(A - \lambda_0 I_n)^k = \text{rg}(A' - \lambda_0 I_n)^k$ per ogni autovalore λ_0 e ogni $k \geq 1$.

Viceversa, se le suddette condizioni sono soddisfatte e le matrici sono triangolabili (se lo è una lo è anche l'altra grazie al Teorema 15.10) il Lemma 15C.3 ci dice che L_A ed $L_{A'}$ hanno la stessa forma canonica di Jordan, per cui A e A' sono simili. Infine, se le matrici non sono triangolabili (per cui necessariamente $\mathbb{K} = \mathbb{R}$) la tesi segue dal caso complesso e dall'Esercizio 14.20. \square

Adesso che sappiamo come trovare la forma canonica di Jordan, vediamo come si trova una base di Jordan. L'idea è la seguente: preso $\lambda_0 \in \text{sp}(T)$, le catene di Jordan di T associate a λ_0 sono le catene di Jordan di $S = T - \lambda_0 \text{id}_V$ associate a 0. Ora, il Teorema della dimensione applicato a $S|_{\text{Im } S^{k-1}}: \text{Im } S^{k-1} \rightarrow \text{Im } S^k$ ci dice che

$$\text{rg } S^{k-1} - \text{rg } S^k = \dim(\text{Ker } S \cap \text{Im } S^{k-1}). \quad (15C.2)$$

Sia allora m la lunghezza massima di una catena di Jordan di S associata a 0. Abbiamo $\text{rg } S^m = \text{rg } S^{m+1}$, per cui (15C.1) e (15C.2) ci dicono che il numero $h_1 = r_m$ di catene di lunghezza m è esattamente $\dim(\text{Ker } S \cap \text{Im } S^{m-1})$, e i primi elementi di queste catene formano una base di $\text{Ker } S \cap \text{Im } S^{k-1}$ (in quanto vi appartengono e sono linearmente indipendenti). Prendiamo una base $\{v_{1,1}, \dots, v_{h_1,1}\}$ di $\text{Ker } S \cap \text{Im } S^{m-1}$. Per ipotesi esistono $v_{i,m} \in V$ tali che $S^{m-1}(v_{i,m}) = v_{i,1}$ per $1 \leq i \leq h_1$; poniamo

$$\forall 1 \leq j \leq m \quad \forall 1 \leq i \leq h_1 \quad v_{i,j} = S^{m-j}(v_{i,m}),$$

e abbiamo trovato le catene di lunghezza m . Poi (15C.1) e (15C.2) ci dicono che il numero r_{m-1} di catene di lunghezza $m-1$ è

$$r_{m-1} = \dim(\text{Ker } S \cap \text{Im } S^{m-2}) - \dim(\text{Ker } S \cap \text{Im } S^{m-1}).$$

Dunque completiamo $\{v_{1,1}, \dots, v_{h_1,1}\}$ a una base $\{v_{1,1}, \dots, v_{h_1,1}, v_{h_1+1,1}, \dots, v_{h_2,1}\}$ di $\text{Ker } S \cap \text{Im } S^{m-2}$ (dove $h_2 = r_m + r_{m-1}$), e per i che varia da $h_1 + 1$ a h_2 prendiamo $v_{i,m-1} \in V$ tale che $S^{m-2}(v_{i,m-1}) = v_{i,1}$. Ponendo

$$\forall 1 \leq j \leq m-1 \quad \forall h_1 + 1 \leq i \leq h_2 \quad v_{i,j} = S^{m-1-j}(v_{i,m-1}),$$

abbiamo le catene di lunghezza $m-1$. Continuiamo così fino a completare una base di $\text{Ker } S \cap \text{Im } S$ a una base di $\text{Ker } S$; in questo modo otteniamo tutte le catene di T associate a λ_0 che ci servono. Ripetendo questo procedimento per tutti gli autovalori di T troviamo la base di Jordan cercata.

Osservazione 15C.2 Questo metodo richiede di determinare basi di sottospazi del tipo $\text{Ker } S \cap \text{Im } S^i$ con $i \geq 0$. Fissando una base di V in modo da poter rappresentare S con una matrice, noi vogliamo trovare gli $x \in \mathbb{K}^n$ tali che $Sx = O$ ed esista $y \in \mathbb{K}^n$ con $S^i y = x$. Applicando le tecniche viste nel Paragrafo 6.4 possiamo trovare equazioni cartesiane del sottospazio $\text{Im } S^i$, che saranno del tipo $R_i x = O$ per qualche matrice R_i . Allora i vettori di $\text{Ker } S \cap \text{Im } S^i$ sono le soluzioni del sistema

$$\begin{cases} Sx = O, \\ R_i x = O. \end{cases}$$

ESEMPIO 15C.2 Vogliamo una base di Jordan per l'endomorfismo T dell'Esempio 15C.1. Abbiamo già trovato che $\text{sp}(T) = \{8, 16\}$ e calcolato $S = T - 8I_3$ e le sue potenze. Essendoci solo un blocco di ordine 2 relativo a 8, dobbiamo trovare una base di $\text{Ker } S \cap \text{Im } S$. Calcoliamo le equazioni cartesiane di $\text{Im } S$ operando come indicato nel Paragrafo 6.4:

$$\begin{aligned} \left| \begin{array}{ccc|c} 6 & -1 & -1 & x_1 \\ -24 & 4 & 4 & x_2 \\ 44 & 6 & -2 & x_3 \end{array} \right| & \longrightarrow & \left| \begin{array}{ccc|c} 6 & -1 & -1 & x_1 \\ 0 & 0 & 0 & x_2 + 4x_1 \\ 0 & 40/3 & 16/3 & x_3 - 22x_1/3 \end{array} \right| \\ & & \longrightarrow & \left| \begin{array}{ccc|c} 6 & -1 & -1 & x_1 \\ 0 & 40/3 & 16/3 & x_3 - 22x_1/3 \\ 0 & 0 & 0 & x_2 + 4x_1 \end{array} \right| \end{aligned}$$

per cui $\text{Im } S = \{x \in \mathbb{R}^3 \mid 4x_1 + x_2 = 0\}$. Dunque per trovare $\text{Ker } S \cap \text{Im } S$ dobbiamo risolvere il sistema

$$\begin{cases} 6x_1 - x_2 - x_3 = 0, \\ -24x_1 + 4x_2 + 4x_3 = 0, \\ 44x_1 + 6x_2 - 2x_3 = 0, \\ 4x_1 + x_2 = 0. \end{cases}$$

Si ottiene

$$\text{Ker } S \cap \text{Im } S = \text{Span} \left(v_{1,1} = \begin{vmatrix} 1 \\ -4 \\ 10 \end{vmatrix} \right).$$

Risolvendo poi il sistema $Sx = v_{1,1}$ vediamo che possiamo per esempio prendere

$$v_{1,2} = \begin{vmatrix} 1/4 \\ 0 \\ 1/2 \end{vmatrix},$$

per cui $\{v_{1,1}, v_{1,2}\}$ è una catena di Jordan per T associata a 8.

Infine, per $\lambda_0 = 16$ abbiamo solo un blocco di ordine 1, per cui ci basta trovare un autovettore di T relativo a 16. Per esempio,

$$v_{2,1} = \begin{vmatrix} 1 \\ -4 \\ 2 \end{vmatrix}.$$

Quindi una base di Jordan per T è

$$\left\{ \begin{vmatrix} 1 \\ -4 \\ 10 \end{vmatrix}, \begin{vmatrix} 1/4 \\ 0 \\ 1/2 \end{vmatrix}, \begin{vmatrix} 1 \\ -4 \\ 2 \end{vmatrix} \right\}.$$

Forme quadratiche reali

Questo capitolo è essenzialmente il risultato della combinazione dei Capitoli 12 e 15. Useremo il Teorema spettrale per classificare completamente i prodotti scalari su uno spazio vettoriale; dopodiché utilizzeremo quanto ottenuto per descrivere coniche e quadriche in \mathbb{R}^n . I Complementi infine contengono la dimostrazione di un risultato di notevole utilità, il Criterio di Cartesio.

16.1 Criteri di positività

Il primo passo verso una classificazione dei prodotti scalari¹ ce lo suggerisce l'Osservazione 12.9: invece di lavorare con tutti i possibili prodotti della forma $\langle v, w \rangle$ possiamo concentrarci semplicemente su quelli del tipo $\langle v, v \rangle$, in quanto la Proposizione 12.2.(vi) ci dice come passare dagli uni agli altri.

Definizione 16.1 Una forma quadratica su uno spazio vettoriale reale V è una funzione $\varphi: V \rightarrow \mathbb{R}$ della forma

$$\varphi(v) = \langle v, v \rangle$$

dove $\langle \cdot, \cdot \rangle$ è un prodotto scalare su V , che è detto associato a φ .

Osservazione 16.1 La funzione φ viene detta forma quadratica perché $\varphi(\lambda v) = \lambda^2 \varphi(v)$ per ogni $\lambda \in \mathbb{R}$ e $v \in V$.

Osservazione 16.2 Grazie all'Osservazione 12.9, esiste un unico prodotto scalare associato a una data forma quadratica φ : si ottiene con la formula

$$\langle v, w \rangle = \frac{1}{4}[\varphi(v+w) - \varphi(v-w)].$$

¹ In questo capitolo finale considereremo soltanto il caso reale: i complessi hanno concluso la loro corsa col capitolo precedente.

Sia $\varphi: V \rightarrow \mathbb{R}$ una forma quadratica. Fissando una base \mathcal{B} di V , e indicando con $x = F_{\mathcal{B}}(v)$ le coordinate di un vettore $v \in V$, si ha

$$\varphi(v) = x^T S x = \sum_{h,k=1}^n s_{hk} x_h x_k, \quad (16.1)$$

dove S è la matrice simmetrica che rappresenta rispetto alla base \mathcal{B} il prodotto scalare associato a φ ; diremo che S rappresenta (o che è associata a) φ rispetto a \mathcal{B} . Dunque φ è un polinomio omogeneo di secondo grado nelle coordinate di v ; viceversa, ogni polinomio omogeneo di secondo grado del tipo (16.1) definisce una forma quadratica su V (esercizio).

Vediamo ora un paio di esempi.

ESEMPIO 16.1 Su \mathbb{R} tutte le forme quadratiche sono della forma $\varphi(x) = ax^2$, con $a \in \mathbb{R}$. Il prodotto scalare associato è $\langle x, y \rangle = axy$, e la matrice rappresentativa è $S = |a|$. Notiamo inoltre che $\varphi(x) > 0$ (o $\varphi(x) < 0$) per tutti gli $x \neq 0$ se e solo se $a > 0$ (o $a < 0$), e ovviamente $\varphi(x) \equiv 0$ se $a = 0$. Dunque φ ha sempre un segno ben definito.

ESEMPIO 16.2 Su \mathbb{R}^2 le forme quadratiche sono polinomi del tipo

$$\varphi(x) = ax_1^2 + 2bx_1x_2 + cx_2^2,$$

con $a, b, c \in \mathbb{R}$. Il prodotto scalare associato è

$$\langle x, y \rangle = ax_1y_1 + bx_1y_2 + bx_2y_1 + cx_2y_2,$$

rappresentato rispetto alla base canonica dalla matrice simmetrica

$$S = \begin{vmatrix} a & b \\ b & c \end{vmatrix}.$$

Vogliamo studiare anche stavolta il segno di φ . Cominciamo supponendo $a \neq 0$; allora si ha

$$\varphi(x) = \frac{1}{a} (ax_1 + bx_2)^2 - \frac{b^2}{a} x_2^2 + cx_2^2 = \frac{1}{a} [(ax_1 + bx_2)^2 + \det(S)x_2^2]. \quad (16.2)$$

Abbiamo quindi la seguente situazione:

- (i) se $a > 0$ e $\det(S) > 0$, allora $\varphi(x) > 0$ per ogni $x \neq 0$;
- (ii) se $a < 0$ e $\det(S) > 0$, allora $\varphi(x) < 0$ per ogni $x \neq 0$;
- (iii) se $a > 0$ e $\det(S) = 0$, allora $\varphi(x) \geq 0$ per ogni $x \in \mathbb{R}^2$, ma esistono degli $x_0 \neq 0$ con $\varphi(x_0) = 0$;
- (iv) se $a < 0$ e $\det(S) = 0$, allora $\varphi(x) \leq 0$ per ogni $x \in \mathbb{R}^2$, ma esistono degli $x_0 \neq 0$ con $\varphi(x_0) = 0$;
- (v) se $\det(S) < 0$, allora $\varphi(x)$ può avere un segno qualsiasi.

Se $a = 0$ ma $c \neq 0$, possiamo ripetere lo stesso discorso scambiando x_1 e x_2 ; infine $a = c = 0$ allora $\varphi(x) = 2bx_1x_2$ può chiaramente assumere qualunque segno.

Una definizione a questo punto s'impone:

Definizione 16.2 Una forma quadratica φ è (semi)definita positiva (o (semi)definita negativa, o indefinita, o degenera) se il prodotto scalare associato lo è. Diremo poi che una matrice simmetrica $S \in M_{n,n}(\mathbb{R})$ è (semi)definita positiva (negativa, indefinita) se la forma quadratica $\varphi(x) = x^T S x$ su \mathbb{R}^n lo è.

Il nostro primo obiettivo è trovare dei criteri efficaci per stabilire il segno di una forma quadratica. In particolare, saremo in grado di risolvere il problema rimasto aperto nel Capitolo 12 di stabilire se un prodotto scalare è definito positivo o meno semplicemente guardando la matrice associata. L'idea è che questa matrice rappresenta anche un endomorfismo simmetrico (rispetto a un diverso prodotto scalare definito positivo) il cui comportamento, essendo diagonalizzabile, è completamente determinato dagli autovalori. Quindi è ragionevole pensare che il segno della forma quadratica sia collegato al segno degli autovalori della matrice associata. Cominciamo allora col dimostrare il seguente

Teorema 16.1 Sia V uno spazio vettoriale metrico su \mathbb{R} , e indichiamo con (\cdot, \cdot) il suo prodotto scalare definito positivo. Allora:

- (i) se $T: V \rightarrow V$ è un endomorfismo simmetrico, la funzione $\varphi: V \rightarrow \mathbb{R}$ definita da $\varphi(v) = (T(v), v)$ è una forma quadratica;
- (ii) se $\varphi: V \rightarrow \mathbb{R}$ è una forma quadratica, allora esiste un unico endomorfismo simmetrico $T: V \rightarrow V$ tale che $\varphi(v) = (T(v), v)$ per ogni $v \in V$. In tal caso, il prodotto scalare (\cdot, \cdot) associato a φ è dato da

$$\langle v, w \rangle = (T(v), w). \quad (16.3)$$

Dimostrazione. (i) Infatti (Esercizio 12.28) la forma bilineare $\langle v, w \rangle = (T(v), w)$ è il prodotto scalare su V associato alla forma quadratica $\varphi(v) = (T(v), v)$.

(ii) Sia $\mathcal{B} = \{v_1, \dots, v_n\}$ una base ortonormale di V rispetto al prodotto scalare definito positivo (\cdot, \cdot) . Per la Proposizione 12.4, ogni vettore $w \in V$ si scrive come

$$w = \sum_{j=1}^n (w, v_j) v_j;$$

dunque

$$\langle v, w \rangle = \left\langle v, \sum_{j=1}^n (w, v_j) v_j \right\rangle = \sum_{j=1}^n \langle v, (w, v_j) v_j \rangle = \sum_{j=1}^n (v_j, w) \langle v, v_j \rangle = \left(\sum_{j=1}^n \langle v, v_j \rangle v_j, w \right)$$

Definiamo $T: V \rightarrow V$ ponendo

$$\forall v \in V \quad T(v) = \sum_{j=1}^n \langle v, v_j \rangle v_j.$$

L'applicazione T è chiaramente lineare e, per costruzione, (16.3) è soddisfatta. Inoltre,

$$(T(v), w) = \langle v, w \rangle = \langle w, v \rangle = (T(w), v) = (v, T(w)),$$

per cui T è simmetrico come desiderato.

Rimane da dimostrare che l'endomorfismo simmetrico T è unico. Se ce ne fossero due, diciamo T_1 e T_2 , posto $S = T_1 - T_2$ per ogni $v, w \in V$ avremmo

$$(S(v), w) = (T_1(v), w) - (T_2(v), w) = \langle v, w \rangle - \langle v, w \rangle = 0.$$

Essendo (\cdot, \cdot) non degenerare, questo implica che $S(v) = 0$ per ogni $v \in V$, e $T_1 \equiv T_2$. \square

Definizione 16.3 L'endomorfismo $T: V \rightarrow V$ simmetrico la cui esistenza e unicità abbiamo appena dimostrato si dice *associato* alla forma quadratica φ .

Osservazione 16.3 Fissiamo una base \mathcal{B} dello spazio vettoriale metrico V , e indichiamo con R_0 la matrice (simmetrica, invertibile) che rappresenta il prodotto scalare definito positivo (\cdot, \cdot) rispetto a questa base. Se poi indichiamo con S la matrice che rappresenta la forma quadratica φ rispetto a \mathcal{B} , e con A la matrice che rappresenta l'endomorfismo T , la (16.3) ci dice che

$$y^T R_0(Ax) = (T(v), w) = \langle v, w \rangle = y^T Sx$$

per ogni $v, w \in V$ (dove $x, y \in \mathbb{R}^n$ sono le coordinate di v e w rispetto a \mathcal{B}) e quindi

$$A = R_0^{-1}S. \quad (16.4)$$

In particolare, φ e T sono rappresentati dalla stessa matrice se e solo se la base \mathcal{B} è ortonormale rispetto a (\cdot, \cdot) .

Osservazione 16.4 Di primo acchito il Teorema 16.1 può lasciare interdetti: stavamo studiando le forme quadratiche su un qualunque spazio vettoriale; perché utilizzare anche un prodotto scalare definito positivo? La risposta è che questo prodotto scalare definito positivo ci permette di introdurre l'endomorfismo simmetrico T , che è lo strumento cardine per studiare le forme quadratiche. Inoltre, se fissiamo una base \mathcal{B} di uno spazio vettoriale V , la Proposizione 12.15 ci dice che possiamo trovare un prodotto scalare definito positivo su V per cui la base \mathcal{B} è ortonormale. Dunque rispetto a questa base (e a questo prodotto scalare) l'endomorfismo simmetrico associato a una forma quadratica φ è rappresentato dalla stessa matrice che rappresenta φ , e quindi l'introduzione del prodotto scalare definito positivo serve soltanto a mostrare che possiamo fare i conti con questa matrice come fosse quella di un endomorfismo. Infine, \mathbb{R}^n nasce già fornito di un prodotto scalare definito positivo, il prodotto scalare canonico; quindi in questo caso l'ambientazione prevista dal Teorema 16.1 è del tutto naturale.

Per cominciare, dimostriamo che una forma quadratica φ è degenerare se e solo se l'endomorfismo simmetrico associato T non è invertibile.

Proposizione 16.2 Sia V uno spazio vettoriale metrico su \mathbb{R} , con prodotto scalare definito positivo (\cdot, \cdot) . Sia $\varphi: V \rightarrow \mathbb{R}$ una forma quadratica su V , e indichiamo con (\cdot, \cdot) il prodotto scalare e con T l'endomorfismo simmetrico associato a φ . Allora il nucleo di (\cdot, \cdot) coincide con $\text{Ker } T$. In particolare, φ è non degenerare se e solo se T è invertibile.

Dimostrazione. Infatti $v_0 \in V^\perp$ se e solo se $\langle v_0, w \rangle = 0$ per ogni $w \in V$ se e solo se $(T(v_0), w) = 0$ per ogni $w \in V$ se e solo se (essendo (\cdot, \cdot) non degenerare) $T(v_0) = 0$ se e solo se $v_0 \in \text{Ker } T$. \square

Ma soprattutto l'endomorfismo T ci permette di stabilire il segno di una forma quadratica:

Teorema 16.3 Sia V uno spazio vettoriale metrico su \mathbb{R} , con prodotto scalare definito positivo (\cdot, \cdot) . Sia $\varphi: V \rightarrow \mathbb{R}$ una forma quadratica, e $T: V \rightarrow V$ l'endomorfismo simmetrico associato a φ . Allora:

- (i) φ è definita positiva (negativa) se e solo se tutti gli autovalori di T sono positivi (negativi);
- (ii) φ è semidefinita positiva (negativa) se e solo se tutti gli autovalori di T sono non negativi (non positivi) e T non è invertibile;
- (iii) φ è indefinita se e solo se T ha sia autovalori positivi che autovalori negativi.

Dimostrazione. (i) Supponiamo che φ sia definita positiva, e sia $v_0 \neq 0$ un autovettore di T di autovalore λ_0 . Allora, se indichiamo con $\|\cdot\|$ la norma associata al prodotto scalare definito positivo (\cdot, \cdot) , si ha

$$0 < \varphi(v_0) = (T(v_0), v_0) = \lambda_0 \|v_0\|^2,$$

e quindi $\lambda_0 > 0$.

Viceversa, supponiamo che tutti gli autovalori di T siano positivi. Siccome T è simmetrico, per il Teorema spettrale (Teorema 15.8) esiste una base $\mathcal{C} = \{v_1, \dots, v_n\}$ di V ortonormale rispetto al prodotto scalare definito positivo (\cdot, \cdot) composta da autovettori di T ; dunque $T(v_j) = \lambda_j v_j$ con $\lambda_j > 0$ per $j = 1, \dots, n$. Prendiamo $v \in V$ e scriviamo

$$v = (v, v_1)v_1 + \dots + (v, v_n)v_n.$$

Allora

$$T(v) = \lambda_1 (v, v_1)v_1 + \dots + \lambda_n (v, v_n)v_n$$

e la formula di Parseval (12.16) ci dice che se $v \neq 0$ si ha

$$\varphi(v) = (T(v), v) = \sum_{j=1}^n \lambda_j |(v, v_j)|^2 > 0,$$

per cui φ è definita positiva. Il caso negativo è identico.

(ii) Come in (i), tenendo presente le Proposizioni 16.2 e 12.1.(iv).

(iii) Infatti è l'unica possibilità rimasta. \square

Dunque il segno degli autovalori dell'endomorfismo simmetrico T dice tutto sul segno della forma quadratica φ . Il fatto che rende questo risultato davvero efficace è l'esistenza di un modo per trovare il segno delle radici di un polinomio semplicemente guardandone i coefficienti. Si tratta del cosiddetto *Criterio di Cartesio*:

Teorema 16.4 (Cartesio) Sia $p(t) = a_n t^n + a_{n-1} t^{n-1} + \dots + a_d t^d$ un polinomio di grado n a coefficienti reali, con $0 \leq d \leq n$ e $a_d \neq 0$. Supponiamo che tutte le radici di p siano reali. Allora:

- (i) 0 è radice di p se e solo se $d \geq 1$, e in tal caso è una radice di molteplicità esattamente d ;
- (ii) p ha tante radici positive, contate con la relativa molteplicità, quante sono le variazioni di segno nella successione dei coefficienti non nulli di p .

La parte (i) di questo Teorema è ovvia; la dimostrazione della parte (ii) è contenuta nei Complementi a questo capitolo.

Osservazione 16.5 Il numero delle radici negative del polinomio p non è il numero delle non-variazioni (o permanenze) di segno nella successione dei coefficienti non nulli di p . Il numero delle radici negative lo si trova sottraendo al grado del polinomio il numero delle radici positive o nulle.

Se T è un endomorfismo simmetrico sappiamo già che il suo polinomio caratteristico ha tutte le radici reali; quindi il criterio di Cartesio è ideale per studiare efficacemente il segno delle forme quadratiche.

ESEMPIO 16.3 Sia $\varphi_a: \mathbb{R}^3 \rightarrow \mathbb{R}$ la forma quadratica rappresentata dalla matrice

$$S_a = \begin{pmatrix} 4 & 1 & -4 \\ 1 & 1 & 0 \\ -4 & 0 & a \end{pmatrix},$$

con $a \in \mathbb{R}$. Vogliamo trovare il segno di φ_a al variare di $a \in \mathbb{R}$. Prendiamo come prodotto scalare definito positivo il prodotto scalare canonico, e come base la base canonica; quindi l'endomorfismo simmetrico associato a φ_a è rappresentato proprio dalla matrice S_a . Il suo polinomio caratteristico è

$$\det(S_a - \lambda I_3) = -\lambda^3 + (a+5)\lambda^2 + (13-5a)\lambda + (3a-16).$$

Per esempio, per $a = -5$ abbiamo

$$\det(S_{-5} - \lambda I_3) = -\lambda^3 + 38\lambda - 31,$$

per cui la sequenza dei coefficienti non nulli è $(-1, 38, -31)$ che varia due volte di segno. Dunque S_{-5} ha due autovalori positivi e, essendo $\det S_{-5} \neq 0$, uno negativo, per cui φ_{-5} è indefinita non degenera. Invece per $a = 6$ abbiamo

$$\det(S_6 - \lambda I_3) = -\lambda^3 + 11\lambda^2 - 17\lambda + 2,$$

per cui abbiamo tre variazioni e conseguentemente tre autovalori positivi: φ_6 è definita positiva. Ripetendo queste considerazioni per ogni valore di a troviamo che φ_a è indefinita non degenera (2 autovalori positivi e uno negativo) per $a < 16/3$, semidefinita positiva per $a = 16/3$ e definita positiva per $a > 16/3$.

16.2 Classificazione metrica delle forme quadratiche

Il nostro prossimo obiettivo è classificare le forme quadratiche, trovandone delle "forme canoniche". Vediamo di spiegare cosa s'intende. Supponiamo di avere due forme quadratiche $\varphi_1, \varphi_2: V \rightarrow \mathbb{R}$ su uno spazio vettoriale V . Quando possiamo dire che queste due forme quadratiche si comportano essenzialmente allo stesso modo? Certamente è così se esiste un isomorfismo $L: V \rightarrow V$ tale che $\varphi_2(v) = \varphi_1(L(v))$ per ogni $v \in V$; infatti, essendo L invertibile e lineare, fare i conti con v o fare i conti con $L(v)$ è esattamente la stessa cosa.

Nel caso però in cui V sia uno spazio vettoriale metrico, ammettendo isomorfismi L qualunque non si terrebbe conto delle proprietà metriche dello spazio. Quindi in questo caso considereremo φ_1 e φ_2 equivalenti solo se esiste un'isometria $L: V \rightarrow V$ tale che $\varphi_2(v) = \varphi_1(L(v))$ per ogni $v \in V$.

Definizione 16.4 Siano $\varphi_1, \varphi_2: V \rightarrow \mathbb{R}$ due forme quadratiche su uno stesso spazio vettoriale V . Diremo che φ_1 e φ_2 sono *affinemente equivalenti* se esiste un isomorfismo $L: V \rightarrow V$ tale che $\varphi_2 = \varphi_1 \circ L$ (e quindi $\varphi_1 = \varphi_2 \circ L^{-1}$). Se inoltre V è uno spazio vettoriale metrico, diremo che φ_1 e φ_2 sono *metricamente equivalenti* se esiste un'isometria $L: V \rightarrow V$ con $\varphi_2 = \varphi_1 \circ L$.

Osservazione 16.6 Ovviamente, due forme quadratiche metricamente equivalenti sono anche affinemente equivalenti; il viceversa invece non è vero (vedi l'Esercizio 16.7).

"Classificare" le forme quadratiche su uno spazio vettoriale V vuol allora dire:

- (i) trovare una procedura efficace per stabilire se due forme quadratiche sono affinemente (o metricamente) equivalenti; e
- (ii) preparare una lista di forme quadratiche a due a due non equivalenti e tali che ogni altra forma quadratica sia equivalente a una di quelle nella lista.

Il passo (i) di solito viene risolto tramite gli invarianti. Parlando alla buona, un *invariante affine* (o *metrico*) è una funzione che associa a ogni forma quadratica un numero in modo che a due forme quadratiche affinemente (o metricamente) equivalenti venga associato lo stesso numero. Il passo (i) viene allora considerato risolto se si riesce a trovare un insieme di invarianti tali che due forme quadratiche siano equivalenti se e solo se hanno esattamente gli stessi invarianti. La lista del passo (ii) allora deve contenere una forma quadratica per ogni possibile valore degli invarianti; queste forme quadratiche si chiamano di solito *forme canoniche affini* (o *metriche*) delle forme quadratiche.

Per vedere come trovare invarianti e forme canoniche, esaminiamo meglio il concetto di equivalenza. Fissiamo una base $\mathcal{B} = \{v_1, \dots, v_n\}$ di V (che supporremo ortonormale nel caso metrico). Prendiamo due forme quadratiche φ_1 e φ_2 su V ,

rappresentate rispetto alla base B dalle matrici simmetriche S_1 ed S_2 . Supponiamo che φ_1 e φ_2 siano affinemente (o metricamente) equivalenti, tramite l'isomorfismo (o isometria) $L: V \rightarrow V$. Indichiamo con B la matrice invertibile (e ortogonale nel caso metrico) che rappresenta L rispetto a B . Allora l'uguaglianza $\varphi_2(v) = \varphi_1(L(v))$ in coordinate diventa

$$x^T S_2 x = (Bx)^T S_1 Bx = x^T B^T S_1 Bx$$

per ogni $x \in \mathbb{R}^n$, e quindi (perché?) $S_2 = B^T S_1 B$. In altre parole, le matrici simmetriche S_1 e S_2 sono congruenti (tramite una matrice ortogonale nel caso metrico). Viceversa, se abbiamo $S_2 = B^T S_1 B$ è chiaro (esercizio) che se indichiamo con L l'isomorfismo rappresentato rispetto alla base B dalla matrice B si ha $\varphi_2 = \varphi_1 \circ L$.

Riassumendo (e ricordando la Proposizione 12.13), abbiamo dimostrato la

Proposizione 16.5 Due forme quadratiche su uno spazio vettoriale (metrico) sono affinemente (metricamente) equivalenti se e solo se le matrici che le rappresentano rispetto a una base (ortonormale) sono congruenti (tramite una matrice ortogonale). Dunque due forme quadratiche sono affinemente (metricamente) equivalenti se e solo se le matrici che le rappresentano rispetto a una data base (ortonormale) rappresentano anche una stessa forma quadratica ma rispetto a due basi (ortonormali) diverse.

Osservazione 16.7 Una conseguenza di questo risultato è che classificare le forme quadratiche è l'altra faccia del seguente problema: data una forma quadratica, trovare una base rispetto a cui è rappresentata da una matrice di tipo particolare. Infatti in entrambi i casi si tratta di studiare le classi di congruenza delle matrici che rappresentano le forme quadratiche.

Dunque, fissata una base di V , trovare degli invarianti delle forme quadratiche vuol dire trovare dei numeri che non cambiano per congruenza; per esempio, il segno del determinante della matrice associata a una forma quadratica è un'invariante affine (Esercizio 16.8). Per trovare una forma canonica, invece, dovremo identificare una matrice particolarmente semplice fra quelle congruenti alla matrice associata alla forma quadratica.

Ma basta coi discorsi generali. Cominciamo con la classificazione metrica: abbiamo già tutto quello che ci serve.

Teorema 16.6 Sia V uno spazio vettoriale metrico su \mathbb{R} di dimensione n . Allora:

- (i) gli autovalori dell'endomorfismo simmetrico associato a una forma quadratica sono degli invarianti metrici;
- (ii) due forme quadratiche sono metricamente equivalenti se e solo se gli endomorfismi simmetrici associati hanno gli stessi autovalori;
- (iii) se $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ sono gli autovalori dell'endomorfismo simmetrico associato a una forma quadratica φ , allora φ è metricamente equivalente alla forma quadratica φ_0 data, nelle coordinate relative a una base ortonormale B fissata, da

$$\varphi_0(x) = \lambda_1 x_1^2 + \dots + \lambda_n x_n^2.$$

Dimostrazione. Fissiamo una base ortonormale B di V , e siano φ_1, φ_2 due forme quadratiche metricamente equivalenti, rappresentate dalle matrici S_1 e S_2 . Per la Proposizione 16.5 esiste una matrice ortogonale B tale che $S_2 = B^T S_1 B = B^{-1} S_1 B$ (dove la seconda uguaglianza segue dal fatto che B è ortogonale). Essendo B ortonormale, gli endomorfismi simmetrici associati alle due forme quadratiche sono rappresentati ancora dalle matrici S_1 ed S_2 . Ma abbiamo appena visto che S_1 ed S_2 sono simili; quindi hanno lo stesso polinomio caratteristico, e di conseguenza gli stessi autovalori, per cui (i) è dimostrata.

Sia ora φ una forma quadratica su V , e S la matrice simmetrica associata rispetto a una base ortonormale B fissata. Il Corollario 15.9 ci dice che esiste una matrice ortogonale B tale che $B^T S B = B^{-1} S B$ sia la matrice diagonale con gli autovalori $\lambda_1, \dots, \lambda_n$ lungo la diagonale principale — matrice che è proprio quella associata a φ_0 rispetto a B . Quindi la Proposizione 16.5 implica che φ è metricamente equivalente a φ_0 , e la (iii) è dimostrata.

Infine, se gli endomorfismi simmetrici associati alle forme quadratiche φ_1 e φ_2 hanno gli stessi autovalori, la (iii) implica che φ_1 e φ_2 sono metricamente equivalenti alla stessa forma quadratica, e quindi sono metricamente equivalenti fra loro. \square

Definizione 16.5 La forma quadratica φ_0 si dice *forma canonica metrica* della forma quadratica φ . Chiaramente, è univocamente determinata a meno dell'ordine degli autovalori.

Definizione 16.6 Diremo che una base B di V porta la forma quadratica φ in *forma canonica metrica* se φ si esprime come la sua forma canonica metrica φ_0 nelle coordinate relative a B . In pratica, B dev'essere una base ortonormale di autovettori per l'endomorfismo simmetrico associato a φ .

ESEMPIO 16.4 Vogliamo la forma canonica metrica della forma quadratica φ su \mathbb{R}^3 rappresentata dalla matrice

$$S = \begin{vmatrix} 4 & 0 & -4 \\ 0 & 1 & 0 \\ -4 & 0 & 0 \end{vmatrix}.$$

Il polinomio caratteristico di S è

$$p_S(\lambda) = (1 - \lambda)(\lambda^2 - 4\lambda - 16) = (1 - \lambda)(2 + 2\sqrt{5} - \lambda)(2 - 2\sqrt{5} - \lambda).$$

Quindi la forma canonica metrica di φ è

$$\varphi_0(x) = x_1^2 + 2(\sqrt{5} + 1)x_2^2 - 2(\sqrt{5} - 1)x_3^2,$$

e una base che porta φ in forma canonica metrica (ovvero una base ortonormale di autovettori per S) è

$$\left\{ \begin{vmatrix} 0 \\ 1 \\ 0 \end{vmatrix}, \frac{1}{\sqrt{2(5 + \sqrt{5})}} \begin{vmatrix} \sqrt{5} + 1 \\ 0 \\ -2 \end{vmatrix}, \frac{1}{\sqrt{2(5 - \sqrt{5})}} \begin{vmatrix} \sqrt{5} - 1 \\ 0 \\ 2 \end{vmatrix} \right\}.$$

16.3 Classificazione affine delle forme quadratiche

Affrontiamo ora la classificazione affine. Per trovare degli invarianti dobbiamo identificare delle caratteristiche intrinseche delle forme quadratiche. Per esempio, se una forma quadratica è definita positiva ogni forma affinemente equivalente a lei è ancora definita positiva. Viene quindi spontaneo considerare come possibili invarianti le dimensioni dei sottospazi su cui la forma quadratica assume un segno preciso.

Definizione 16.7 Sia $\varphi: V \rightarrow \mathbb{R}$ una forma quadratica su uno spazio vettoriale V , e $\langle \cdot, \cdot \rangle$ il prodotto scalare associato. Diremo *indice di nullità* $i_0(\varphi)$ di φ la dimensione del nucleo di $\langle \cdot, \cdot \rangle$; *rango* il numero $r(\varphi) = \dim V - i_0(\varphi)$; *indice di positività* $i_+(\varphi)$ la massima dimensione di un sottospazio di V su cui φ è definita positiva; *indice di negatività* $i_-(\varphi)$ la massima dimensione di un sottospazio di V su cui φ è definita negativa; e *segnatura* il numero $s(\varphi) = i_+(\varphi) - i_-(\varphi)$. Se $A \in M_{n,n}(\mathbb{R})$ è una matrice simmetrica, diremo *indice di nullità (di positività, di negatività, segnatura)* di A l'indice di nullità (di positività, di negatività, segnatura) della forma quadratica $\varphi(x) = x^T A x$ definita su \mathbb{R}^n .

Lemma 16.7 *Gli indici di nullità, positività, negatività, la segnatura e il rango sono invarianti affini delle forme quadratiche.*

Dimostrazione. Siano φ_1 e $\varphi_2 = \varphi_1 \circ L$ due forme quadratiche affinemente equivalenti. Siccome φ_2 è definita positiva (negativa) su un sottospazio W di V se e solo se φ_1 è definita positiva (negativa) sul sottospazio di uguale dimensione $L(W)$, chiaramente φ_1 e φ_2 hanno gli stessi indici di positività e negatività, e quindi la stessa segnatura. Se poi $\langle \cdot, \cdot \rangle_j$ è il prodotto scalare associato a φ_j per $j = 1, 2$, si ha

$$\forall v, w \in V \quad \langle v, w \rangle_2 = \langle L(v), L(w) \rangle_1,$$

(perché?) e quindi φ_1 e φ_2 hanno anche uguali indice di nullità e rango. \square

Dunque abbiamo trovato degli invarianti affini. Per calcolarli possiamo usare di nuovo l'endomorfismo simmetrico associato (e il segno dei suoi autovalori), come mostrato nel *Teorema di Sylvester*:

Teorema 16.8 (Sylvester) *Sia $\varphi: V \rightarrow \mathbb{R}$ una forma quadratica su uno spazio vettoriale V . Poniamo su V un prodotto scalare definito positivo $\langle \cdot, \cdot \rangle$, e sia $T: V \rightarrow V$ l'endomorfismo simmetrico associato a φ . Allora:*

- (i) $r(\varphi) = \text{rg } T$ e $i_0(\varphi) = \dim \text{Ker } T$, che è anche uguale alla molteplicità algebrica di 0 come autovalore di T ;
- (ii) $i_+(\varphi)$ è il numero degli autovalori positivi (contati con la relativa molteplicità) di T ;
- (iii) $i_-(\varphi)$ è il numero degli autovalori negativi (contati con la relativa molteplicità) di T ;
- (iv) $i_0(\varphi) + i_+(\varphi) + i_-(\varphi) = \dim V$.

Dimostrazione. Sia $\mathcal{B} = \{v_1, \dots, v_n\}$ una base ortonormale di V composta da autovettori di T , dove v_j è relativo a $\lambda_j \in \mathbb{R}$, ordinati in modo tale che

$$\lambda_j \begin{cases} > 0 & \text{se } j = 1, \dots, p, \\ < 0 & \text{se } j = p+1, \dots, p+q, \\ = 0 & \text{se } j = p+q+1, \dots, p+q+h = n, \end{cases} \quad (16.5)$$

dove $n = \dim V$. Siccome $\text{Ker } T = \text{Span}(v_{p+q+1}, \dots, v_n)$, la Proposizione 16.2 implica che $h = i_0(\varphi)$ e che $r(\varphi) = \text{rg } T$; dobbiamo dimostrare che $p = i_+(\varphi)$ e $q = i_-(\varphi)$.

Poniamo $V_+ = \text{Span}(v_1, \dots, v_p)$, $V_- = \text{Span}(v_{p+1}, \dots, v_{p+q})$ e $V_0 = \text{Ker } T$. Se prendiamo $v = c_1 v_1 + \dots + c_p v_p \in V_+$ con $v \neq O$ abbiamo

$$\varphi(v) = (T(v), v) = \sum_{i,j=1}^p c_i c_j \lambda_i (v_i, v_j) = \sum_{i,j=1}^p \lambda_i c_i c_j \delta_{ij} = \sum_{k=1}^p \lambda_k (c_k)^2 > 0,$$

per cui φ è definita positiva su V_+ , e quindi $i_+(\varphi) \geq \dim V_+ = p$. Analogamente si dimostra che φ è definita negativa su V_- .

Viceversa, sia W un sottospazio di V su cui φ è definita positiva; cominciamo col far vedere che

$$W \cap (V_0 \oplus V_-) = \{O\}. \quad (16.6)$$

Infatti, prendiamo $w \in W \cap (V_0 \oplus V_-)$; dunque possiamo scrivere $w = v_0 + v_-$, con $v_0 \in V_0$ e $v_- \in V_-$. Allora

$$\begin{aligned} 0 \leq \varphi(w) &= \varphi(v_0 + v_-) = (T(v_0 + v_-), v_0 + v_-) = (T(v_-), v_0 + v_-) \\ &= (v_-, T(v_0)) + (T(v_-), v_-) = (T(v_-), v_-) = \varphi(v_-) \leq 0, \end{aligned}$$

dove abbiamo usato $V_0 = \text{Ker } T$ e la simmetria di T . Quindi $\varphi(w) = 0$; ma φ è definita positiva su W , e dunque $w = O$.

Ora, ricordando che $V = V_+ \oplus (V_0 \oplus V_-)$, il Teorema di Grassmann e (16.6) implicano

$$\dim V \geq \dim(W + (V_0 \oplus V_-)) = \dim W + \dim(V_0 \oplus V_-) = \dim W + \dim V - \dim V_+,$$

e quindi $\dim W \leq \dim V_+ = p$. Siccome l'indice di positività è il massimo delle dimensioni dei sottospazi W su cui φ è definita positiva, otteniamo $i_+(\varphi) \leq p$ e, per quanto già visto, $i_+(\varphi) = p$.

In maniera assolutamente analoga si dimostra che $i_-(\varphi) = q = n - h - p$, e ci siamo. \square

E ora siamo in grado di completare anche la classificazione affine:

Teorema 16.9 Sia V uno spazio vettoriale su \mathbb{R} di dimensione n . Allora:

- (i) due forme quadratiche sono affinemente equivalenti se e solo se hanno gli stessi indici di positività e negatività;
 (ii) fissiamo una base \mathcal{B} di V . Se i_+ e i_- sono l'indice di positività e l'indice di negatività di una forma quadratica φ , allora φ è affinemente equivalente alla forma quadratica φ_0 data, nelle coordinate relative alla base \mathcal{B} , da

$$\varphi_0(x) = x_1^2 + \cdots + x_{i_+}^2 - x_{i_++1}^2 - \cdots - x_{i_++i_-}^2.$$

Dimostrazione. Siccome sappiamo già che gli indici di positività e negatività sono degli invarianti affini, ci basta dimostrare (ii); infatti in tal caso due forme quadratiche con gli stessi indici di positività e negatività sono affinemente equivalenti alla stessa forma quadratica, e quindi sono affinemente equivalenti fra loro.

Per dimostrare (ii) scegliamo un prodotto scalare definito positivo (\cdot, \cdot) su V rispetto a cui \mathcal{B} sia una base ortonormale (si può sempre fare: vedi la Proposizione 12.15), e sia $T: V \rightarrow V$ l'endomorfismo simmetrico associato a φ . Allora esiste una base ortonormale $\mathcal{C} = \{v_1, \dots, v_n\}$ composta da autovettori di T ; possiamo anche supporre che gli autovalori siano ordinati come in (16.5). Rispetto a questa base, φ in coordinate è data da

$$\varphi(x) = \lambda_1 x_1^2 + \cdots + \lambda_{i_+} x_{i_+}^2 + \lambda_{i_++1} x_{i_++1}^2 + \cdots + \lambda_{i_++i_-} x_{i_++i_-}^2.$$

Poniamo ora $v'_j = v_j / \sqrt{|\lambda_j|}$ per $j = 1, \dots, i_++i_-$, e $v'_j = v_j$ per $j = i_++i_-+1, \dots, n$. Allora $\mathcal{C}' = \{v'_1, \dots, v'_n\}$ è una base di V (non più ortonormale!) rispetto a cui φ in coordinate è data da

$$\varphi(x') = (x'_1)^2 + \cdots + (x'_{i_+})^2 - (x'_{i_++1})^2 - \cdots - (x'_{i_++i_-})^2.$$

Dunque abbiamo trovato una base di V rispetto a cui φ è rappresentata dalla stessa matrice che rappresenta φ_0 rispetto a \mathcal{B} ; per la Proposizione 16.5, questo vuol esattamente dire che φ e φ_0 sono affinemente equivalenti. \square

Definizione 16.8 La forma quadratica φ_0 si dice *forma canonica affine* della forma quadratica φ . Diremo che una base \mathcal{B} di V porta φ in forma canonica affine se φ si esprime come φ_0 nelle coordinate relative a \mathcal{B} .

Osservazione 16.8 I vari indici e la segnatura di una forma quadratica non sono indipendenti fra loro. Infatti, se $\varphi: V \rightarrow \mathbb{R}$ è una forma quadratica su uno spazio vettoriale di dimensione n , il Teorema 16.8.(iv) ci dice che

$$i_0(\varphi) = n - i_+(\varphi) - i_-(\varphi) \quad \text{e} \quad s(\varphi) = i_+(\varphi) - i_-(\varphi);$$

viceversa,

$$i_+(\varphi) = \frac{n - i_0(\varphi) + s(\varphi)}{2} \quad \text{e} \quad i_-(\varphi) = \frac{n - i_0(\varphi) - s(\varphi)}{2}.$$

In particolare, due forme quadratiche sono affinemente equivalenti se e solo se hanno uguale indice di nullità e segnatura.

Osservazione 16.9 Il Teorema 16.9 ci dice che esiste un isomorfismo $L: V \rightarrow V$ tale che $\varphi = \varphi_0 \circ L$, cioè tale che

$$\varphi(v) = \sum_{i=1}^{i_+} (a_{i1}x_1 + \cdots + a_{in}x_n)^2 - \sum_{i=i_++1}^{i_++i_-} (a_{i1}x_1 + \cdots + a_{in}x_n)^2,$$

dove $x = F_{\mathcal{B}}(v)$ sono le coordinate di v rispetto a una base \mathcal{B} , e $A = (a_{ij})$ è la matrice che rappresenta L rispetto a \mathcal{B} . In particolare, φ è definita positiva se e solo se $\varphi(v)$ si scrive come somma di quadrati di combinazioni lineari delle coordinate, come preannunciato nell'Osservazione 12.5.

ESEMPIO 16.5 Sia $\varphi: \mathbb{R}^3 \rightarrow \mathbb{R}$ la forma quadratica rappresentata dalla matrice

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

vogliamo la sua forma canonica affine, e una base di \mathbb{R}^3 che porti φ in forma canonica affine. Il polinomio caratteristico di S è

$$p_S(\lambda) = -\lambda^3 + 7\lambda^2 + 2\lambda - 8.$$

Abbiamo due variazioni di segno, e $\det S = -8 \neq 0$; dunque il Criterio di Cartesio ci dice che l'indice di nullità è zero, l'indice di positività è due. Di conseguenza, la forma canonica affine è

$$\varphi_0(x) = x_1^2 + x_2^2 - x_3^2.$$

Per trovare la base abbiamo bisogno degli autovalori di S . Non è difficile verificare che $\text{sp}(S) = \{1, 3 \pm \sqrt{17}\}$, e che gli autospazi sono generati dai vettori

$$v_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} \frac{2\sqrt{2}}{\sqrt{17+\sqrt{17}}} \\ 0 \\ \frac{1+\sqrt{17}}{\sqrt{2(17+\sqrt{17})}} \end{pmatrix}, \quad v_3 = \begin{pmatrix} \frac{2\sqrt{2}}{\sqrt{17-\sqrt{17}}} \\ 0 \\ \frac{1-\sqrt{17}}{\sqrt{2(17-\sqrt{17})}} \end{pmatrix},$$

dove v_1, v_2 e v_3 hanno tutti norma uno. Per ottenere i vettori della base cercata dobbiamo quindi dividere questi autovettori per la radice quadrata del modulo del relativo autovalore; quindi una base che porta φ in forma canonica affine è

$$\left\{ \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{2}{\sqrt{7\sqrt{17}-17}} \\ 0 \\ \frac{1+\sqrt{17}}{2\sqrt{7\sqrt{17}-17}} \end{pmatrix}, \begin{pmatrix} \frac{2}{\sqrt{7\sqrt{17}+17}} \\ 0 \\ \frac{1-\sqrt{17}}{2\sqrt{7\sqrt{17}+17}} \end{pmatrix} \right\}.$$

16.4 Quadriche in \mathbb{R}^n

Nei Capitoli 10 e 13 abbiamo studiato insiemi (rette e piani) che possono essere descritti come luogo di zeri di equazioni (o sistemi) di primo grado. In questo capitolo abbiamo invece incontrato le forme quadratiche, il primo oggetto davvero non lineare (quadratico appunto) studiato in questo testo; e adesso vedremo come applicare quanto visto per studiare i luoghi di zeri di polinomi di secondo grado.

Definizione 16.9 Indichiamo con $\mathbb{R}_2[x_1, \dots, x_n]$ l'insieme dei polinomi a coefficienti reali in n variabili di grado minore o uguale a 2. Sia $p \in \mathbb{R}_2[x_1, \dots, x_n]$ un polinomio di secondo grado. La *quadrica* \mathcal{Q}_p di equazione p è il luogo dei punti di \mathbb{R}^n le cui coordinate soddisfano l'equazione $p(x_1, \dots, x_n) = 0$. In altri termini,

$$\mathcal{Q}_p = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid p(x_1, \dots, x_n) = 0\}.$$

Se $n = 2$, la quadrica \mathcal{Q}_p sarà detta *conica* di equazione p , e indicata con \mathcal{C}_p .

ESEMPIO 16.6 Se $p(x, y) = x^2 + y^2 - 2x = (x - 1)^2 + y^2 - 1$, la conica \mathcal{C}_p è la circonferenza di centro $(1, 0)$ e raggio 1. Se invece $p(x, y) = y^2 + 2x$, la conica \mathcal{C}_p è la parabola di equazione $x = -y^2/2$.

Osservazione 16.10 Le coniche si chiamano in questo modo perché si possono ottenere intersecando un cono con un piano; vedi l'Osservazione 13C.6.

Osservazione 16.11 Attenzione a non confondere quadriche e forme quadratiche: sono due concetti ben diversi. Le forme quadratiche sono polinomi *omogenei* di secondo grado; le quadriche sono sottoinsiemi di \mathbb{R}^n che soddisfano un'equazione (di solito non omogenea) di secondo grado.

Il nostro obiettivo è classificare le quadriche di \mathbb{R}^n . Per far ciò dobbiamo decidere quando due polinomi di secondo grado p e q determinano la stessa quadrica. Vi è una situazione ovvia in cui questo accade: se esiste un $\sigma \neq 0$ tale che $q = \sigma p$, è evidente che \mathcal{Q}_q e \mathcal{Q}_p sono proprio lo stesso insieme.

Ma questo non è l'unico caso. Una quadrica è un sottoinsieme di \mathbb{R}^n , con una sua forma e un suo aspetto geometrico indipendente da dove si trova in quel momento — ovvero indipendente dal sistema di coordinate scelto. In altri termini, se cambiamo coordinate l'equazione che descrive la quadrica può mutare, ma la quadrica come insieme rimane essenzialmente la stessa. Dunque se due polinomi sono ottenuti l'uno dall'altro tramite un cambiamento di coordinate allora descrivono essenzialmente la stessa quadrica. E questo accade anche se fra i cambiamenti di coordinate ammettiamo le traslazioni (come facevamo su \mathcal{A}^2 e \mathcal{A}^3 ; vedi i Paragrafi 10.7 e 13.1).

Definizione 16.10 Un *cambiamento affine di coordinate* in \mathbb{R}^n è un cambiamento di coordinate del tipo

$$x = Bx' + c, \quad (16.7)$$

con $B \in GL_n(\mathbb{R})$ e $c \in \mathbb{R}^n$. Un *cambiamento metrico di coordinate* in \mathbb{R}^n è un cambiamento affine di coordinate dove B è una matrice ortogonale.

Definizione 16.11 Due polinomi $p, q \in \mathbb{R}_2[x_1, \dots, x_n]$ sono *affinemente equivalenti* se esistono $\sigma \in \mathbb{R}^*$, $B \in GL_n(\mathbb{R})$ e $c \in \mathbb{R}^n$ tali che

$$q(x') = \sigma p(Bx' + c),$$

e in tal caso $p(x) = \sigma^{-1}q(B^{-1}x - B^{-1}c)$. Se inoltre $B \in O(n)$ è ortogonale, diremo che p e q sono *metricamente equivalenti*. Se p e q sono affinemente (metricamente) equivalenti, diremo anche che le quadriche \mathcal{Q}_p e \mathcal{Q}_q sono *affinemente (metricamente) equivalenti*.

In altri termini, due quadriche sono affinemente (metricamente) equivalenti se esiste un cambiamento affine (metrico) di coordinate che porta l'equazione della prima in un multiplo dell'equazione della seconda.

Osservazione 16.12 Due quadriche metricamente equivalenti sono anche affinemente equivalenti; il viceversa non è vero (per esempio, un'ellisse di equazione $x^2 + 2y^2 - 1 = 0$ è affinemente ma non metricamente equivalente alla circonferenza $x^2 + y^2 - 1 = 0$).

Osservazione 16.13 La fondamentale differenza fra l'equivalenza metrica e quella affine è la seguente: due quadriche metricamente equivalenti sono a tutti gli effetti lo stesso insieme disegnato in posti diversi di \mathbb{R}^n , mentre due quadriche affinemente equivalenti hanno soltanto la stessa "forma". Infatti, l'equivalenza metrica conserva tutte le proprietà metriche (distanze fra punti, angoli, eccetera), mentre l'equivalenza affine conserva solo le proprietà affini (appartenenza, limitatezza, numero di pezzi, allineamenti, eccetera). Per esempio, due circonferenze di raggio diverso (o due ellissi, se è per questo) sono affinemente equivalenti (hanno essenzialmente la stessa forma), ma non sono metricamente equivalenti (hanno raggi diversi).

Osservazione 16.14 La nostra definizione privilegia l'equivalenza *algebraica* delle equazioni sull'equivalenza *geometrica* degli insiemi: se esiste (a sistema di coordinate fissato) un'isometria (o un'affinità) che porta la quadrica \mathcal{Q}_p nella quadrica \mathcal{Q}_q allora le due quadriche sono metricamente (o affinemente) equivalenti; il viceversa, in alcuni casi degeneri, non è vero. Il problema è che siccome siamo sui numeri reali l'equazione $p(x_1, \dots, x_n) = 0$ potrebbe non avere soluzioni. Per esempio, vedremo che le coniche di equazione $x^2 + y^2 + 1 = 0$ e $x^2 + 1 = 0$ non sono neppure affinemente equivalenti, anche se entrambe descrivono l'insieme vuoto (cioè nessun punto $(x, y) \in \mathbb{R}^2$ soddisfa una delle due equazioni).

Osservazione 16.15 Se le precedenti Osservazioni ti risultano oscure, conviene rileggerle dopo aver finito di studiare questo paragrafo; a quel punto saranno molto più chiare.

Il nostro obiettivo è classificare le quadriche: dunque vogliamo associare ai polinomi di secondo grado degli invarianti, e poi trovare una breve lista di quadriche in forma canonica, a cui tutte le altre siano equivalenti. Per dare un'idea degli insiemi di cui stiamo parlando, elenchiamo subito quelle che risulteranno essere le forme canoniche affini delle coniche in \mathbb{R}^2 e delle quadriche in \mathbb{R}^3 :

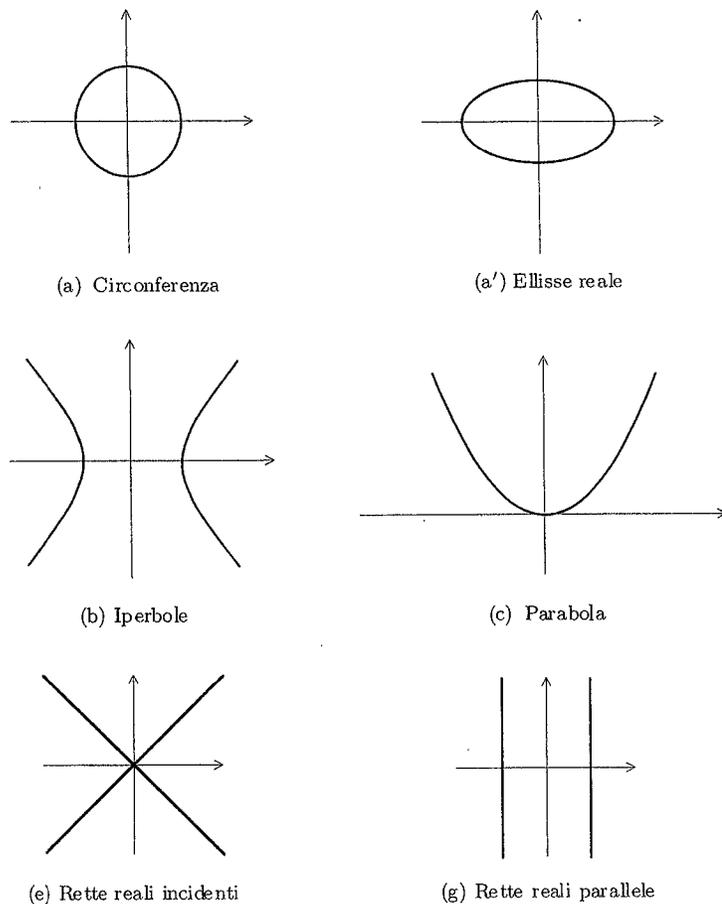


Figura 16.1 Coniche in forma canonica affine.

ESEMPIO 16.7 Le coniche tipo, a cui cercheremo di ricondurre tutte le altre, sono:

- (a) l'ellisse reale, di equazione $x^2 + y^2 - 1 = 0$;
- (b) l'iperbole, di equazione $x^2 - y^2 - 1 = 0$;
- (c) la parabola, di equazione $x^2 - y = 0$;
- (d) l'ellisse immaginaria, di equazione $x^2 + y^2 + 1 = 0$;
- (e) le rette reali incidenti, di equazione $x^2 - y^2 = 0$;
- (f) le rette complesse incidenti, di equazione $x^2 + y^2 = 0$;
- (g) le rette reali parallele, di equazione $x^2 - 1 = 0$;
- (h) le rette complesse parallele, di equazione $x^2 + 1 = 0$;
- (i) le rette coincidenti, di equazione $x^2 = 0$.

Come vedremo, le coniche di tipo (a)–(d) sono *non degeneri*, e le altre *degeneri*. L'ellisse immaginaria e le rette complesse parallele descrivono in realtà entrambe l'insieme vuoto, come detto nell'Osservazione 16.14; le rette complesse incidenti sono un punto, e le rette coincidenti sono l'asse y "contato due volte". Le altre coniche in forma canonica affine sono mostrate nella Figura 16.1. Leggi anche i Complementi al Capitolo 13 per altre informazioni sulle coniche non degeneri.

Osservazione 16.16 Il motivo per cui abbiamo chiamato "ellisse reale" la circonferenza di equazione $x^2 + y^2 - 1 = 0$ è che, come vedremo, tutte le ellissi sono affinemente (ma non metricamente) equivalenti a essa. La Figura 16.1.(a) rappresenta proprio la circonferenza $x^2 + y^2 - 1 = 0$, mentre la Figura 16.1.(a') rappresenta un'ellisse la cui equazione non è in forma canonica affine.

Osservazione 16.17 Le "rette complesse incidenti" (la conica di equazione $x^2 + y^2 = 0$) hanno questo nome perché il polinomio $x^2 + y^2$ si spezza nel prodotto $(x + iy)(x - iy)$ di due fattori lineari a coefficienti complessi, come il polinomio definente le "rette reali incidenti" si spezza nel prodotto di due fattori lineari a coefficienti reali. Il motivo del nome "rette complesse parallele" per la conica di equazione $x^2 + 1 = 0$ è analogo.

ESEMPIO 16.8 La lista delle quadriche tipo in \mathbb{R}^3 è sensibilmente più lunga. Si tratta di:

- (a) l'ellissoide reale, di equazione $x^2 + y^2 + z^2 - 1 = 0$;
- (b) l'ellissoide immaginario, di equazione $x^2 + y^2 + z^2 + 1 = 0$;
- (c) l'iperboloide ellittico, di equazione $x^2 + y^2 - z^2 + 1 = 0$;
- (d) l'iperboloide iperbolico, di equazione $x^2 + y^2 - z^2 - 1 = 0$;
- (e) il paraboloido ellittico, di equazione $x^2 + y^2 - z = 0$;
- (f) il paraboloido iperbolico, di equazione $x^2 - y^2 - z = 0$;
- (g) il cono immaginario, di equazione $x^2 + y^2 + z^2 = 0$;
- (h) il cono reale, di equazione $x^2 + y^2 - z^2 = 0$;
- (i) il cilindro immaginario, di equazione $x^2 + y^2 + 1 = 0$;
- (j) il cilindro ellittico, di equazione $x^2 + y^2 - 1 = 0$;
- (k) il cilindro parabolico, di equazione $x^2 - y = 0$;
- (l) il cilindro iperbolico, di equazione $x^2 - y^2 - 1 = 0$;
- (m) i piani complessi incidenti, di equazione $x^2 + y^2 = 0$;
- (n) i piani reali incidenti, di equazione $x^2 - y^2 = 0$;
- (o) i piani complessi paralleli, di equazione $x^2 + 1 = 0$;
- (p) i piani reali paralleli, di equazione $x^2 - 1 = 0$;
- (q) i piani coincidenti, di equazione $x^2 = 0$.

Alcune di queste quadriche sono rappresentate nella Figura 16.2.

Osservazione 16.18 Come già fatto per ellissi e circonferenze, abbiamo chiamato "ellissoide reale" la sfera di equazione $x^2 + y^2 + z^2 - 1 = 0$. Un esempio di ellissoide non in forma canonica affine è nella Figura 16.2.(a').

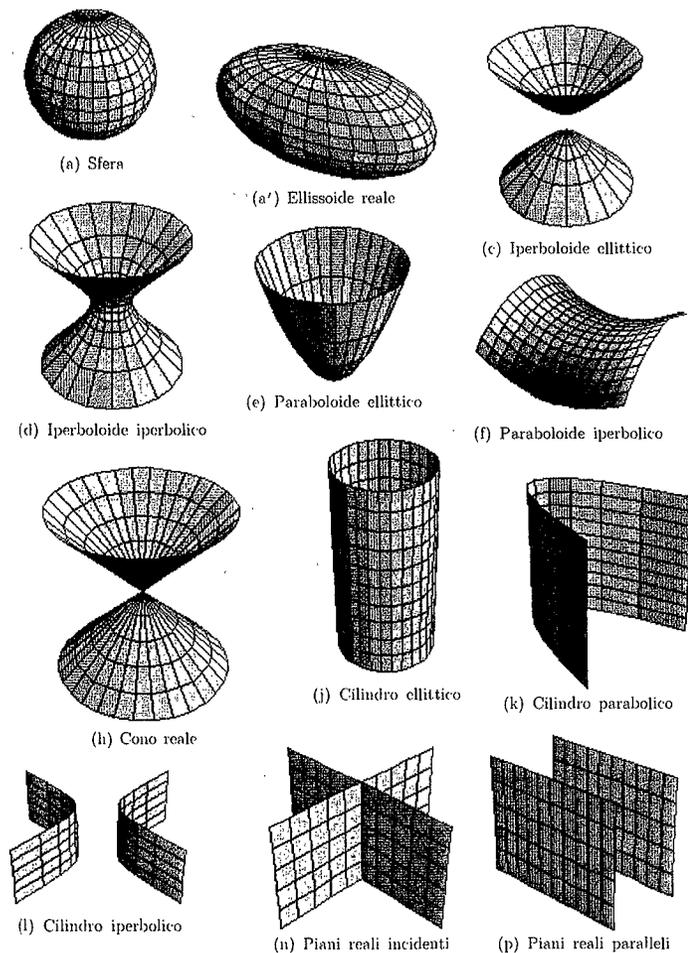


Figura 16.2 Le quadriche in forma canonica affine.

Un generico polinomio $p \in \mathbb{R}_2[x_1, \dots, x_n]$ di secondo grado in n variabili è della forma

$$p(x_1, \dots, x_n) = \sum_{i,j=1}^n a_{ij} x_i x_j + 2 \sum_{j=1}^n a_{j,n+1} x_j + a_{n+1,n+1}. \quad (16.8)$$

Possiamo scrivere p in una maniera più semplice. Al polinomio p (e quindi alla

quadrica \mathcal{Q}_p) possiamo associare la matrice simmetrica

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1,n+1} \\ \vdots & \ddots & \vdots \\ a_{1,n+1} & \cdots & a_{n+1,n+1} \end{pmatrix} \in M_{n+1,n+1}(\mathbb{R}),$$

e la sua sottomatrice (principale; vedi l'Esercizio 16.4)

$$\mathcal{A}_n = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{1n} & \cdots & a_{nn} \end{pmatrix} \in M_{n,n}(\mathbb{R}).$$

Queste matrici hanno una relazione molto stretta con p ; infatti si ha (esercizio):

$$p(x_1, \dots, x_n) = \begin{vmatrix} x_1 & \cdots & x_n & 1 \\ \vdots & & \vdots & \\ x_n & & & \\ 1 & & & \end{vmatrix} A \begin{vmatrix} x_1 \\ \vdots \\ x_n \\ 1 \end{vmatrix},$$

o anche

$$p(x) = \tilde{x}^T A \tilde{x}, \quad (16.9)$$

dove useremo sempre la convenzione che se $x \in \mathbb{R}^n$ allora $\tilde{x} \in \mathbb{R}^{n+1}$ è ottenuto da x aggiungendo 1 come ultima coordinata:

$$x = \begin{vmatrix} x_1 \\ \vdots \\ x_n \end{vmatrix} \implies \tilde{x} = \begin{vmatrix} x_1 \\ \vdots \\ x_n \\ 1 \end{vmatrix}.$$

Inoltre, se indichiamo con

$$p^{(2)}(x_1, \dots, x_n) = \sum_{i,j=1}^n a_{ij} x_i x_j$$

la parte quadratica di p , abbiamo anche che

$$p^{(2)}(x_1, \dots, x_n) = \begin{vmatrix} x_1 & \cdots & x_n \\ \vdots & & \vdots \\ x_n & & \end{vmatrix} \mathcal{A}_n \begin{vmatrix} x_1 \\ \vdots \\ x_n \end{vmatrix}.$$

Definizione 16.12 La matrice A è detta associata alla quadrica \mathcal{Q}_p (o al polinomio p). Diremo che la quadrica è *degenere* se $\det A = 0$, e *non degenere* altrimenti.

Osservazione 16.19 La matrice A è l'unica matrice simmetrica per cui vale (16.9), e quindi grazie alla quale possiamo studiare le quadriche con le tecniche viste per le forme quadratiche. È per questo motivo che associamo al polinomio p proprio la matrice A e non un'altra.

Ora al cambiamento di coordinate affine (o metrico) $x = Bx' + c$ possiamo associare la matrice

$$\tilde{B} = \begin{vmatrix} B & c \\ O & 1 \end{vmatrix} \in GL_{n+1}(\mathbb{R}), \quad (16.10)$$

in modo da riassumere la (16.7) in

$$\tilde{x} = \tilde{B}\tilde{x}'.$$

Allora la condizione di equivalenza fra due polinomi $p(x) = \tilde{x}^T A \tilde{x}$ e $q(x') = (\tilde{x}')^T A' \tilde{x}'$ diventa

$$(\tilde{x}')^T A' \tilde{x}' = q(x') = \sigma p(Bx' + c) = \sigma p(x) = \tilde{x}^T \sigma A \tilde{x} = (\tilde{x}')^T (\sigma \tilde{B}^T A \tilde{B}) \tilde{x}'.$$

Dunque due quadriche di matrici associate A e A' sono equivalenti se e solo se

$$A' = \sigma \tilde{B}^T A \tilde{B}, \quad (16.11)$$

e abbiamo dimostrato la

Proposizione 16.10 Due quadriche di matrici associate A e A' sono affinemente equivalenti se e solo se esiste $\sigma \in \mathbb{R}^*$ tale che σA e A' siano congruenti tramite una matrice della forma (16.10). Le due quadriche sono metricamente equivalenti se e solo se esiste $\sigma \in \mathbb{R}^*$ tale che σA e A' siano congruenti tramite una matrice della forma (16.10) con B ortogonale.

Dunque possiamo richiamare immediatamente tutta la trattazione sugli invarianti affini delle forme quadratiche, vista nei paragrafi precedenti. Nota anche che (16.11) implica (esercizio)

$$A'_n = \sigma B^T A_n B, \quad b' = \sigma B^T (A_n c + b) \quad \text{e} \quad \alpha' = \sigma p(c). \quad (16.12)$$

In particolare abbiamo:

Corollario 16.11 Il valore assoluto della segnatura e il rango della matrice A associata a una quadrica, e il valore assoluto della segnatura e il rango della sottomatrice A_n sono invarianti affini delle quadriche. In particolare, una quadrica degenera e una non degenera non possono mai essere affinemente equivalenti.

Dimostrazione. Dalla Proposizione 16.5 e dal Lemma 16.7 segue che A e $\tilde{B}^T A \tilde{B}$ (come pure A_n e $B^T A_n B$) hanno uguale segnatura e rango. Siccome A e σA hanno uguale rango e segnatura per $\sigma > 0$, e uguale rango ma segnatura opposta per $\sigma < 0$, abbiamo finito. \square

Osservazione 16.20 Contrariamente a quanto si potrebbe immaginare, la matrice \tilde{B} può non essere ortogonale anche se B è ortogonale; quindi gli autovalori di A non sono invarianti metrici delle quadriche.

Il nostro piano di lavoro adesso è il seguente: dare una procedura effettiva per trovare un cambiamento affine (o metrico) di coordinate che porti l'equazione di una quadrica in una forma particolarmente semplice (del tipo di quelle viste negli Esempi 16.7 e 16.8). Fatto ciò, basterà notare che gli invarianti elencati nel Corollario 16.11 permettono di distinguere completamente queste quadriche semplici e avremo completato la classificazione affine delle quadriche. Lungo la strada, troveremo anche la classificazione metrica.

Sia allora \mathcal{Q}_p una quadrica di equazione $p(x_1, \dots, x_n) = 0$. Indichiamo con A la matrice associata, con A_n la solita sottomatrice, e con $p^{(2)}$ la parte quadratica del polinomio p . Poniamo inoltre

$$b = \begin{vmatrix} a_{1,n+1} \\ \vdots \\ a_{n,n+1} \end{vmatrix} \in \mathbb{R}^n \quad \text{e} \quad \alpha = a_{n+1,n+1},$$

in modo da scrivere la matrice A nella forma

$$A = \begin{vmatrix} A_n & b \\ b^T & \alpha \end{vmatrix},$$

e il polinomio p nella forma

$$p(x) = x^T A_n x + 2b^T x + \alpha.$$

Infine, sia $r = \text{rg}(A_n)$ il rango di A_n ed $s = s(A_n)$ la segnatura di A_n .

Prima di tutto, moltiplichiamo se necessario il polinomio p per -1 in modo da essere certi che la matrice A_n abbia segnatura non negativa (e quindi almeno un autovalore positivo; vedi l'Esercizio 16.13). Siccome A_n è una matrice simmetrica, il Corollario 15.9 ci fornisce una matrice ortogonale $B \in O(n)$ tale che $A'_n = B^T A_n B$ sia diagonale; possiamo anche supporre che gli elementi sulla diagonale (che sono gli autovalori di A_n) siano ordinati come in (16.5). Allora effettuando il cambiamento metrico di coordinate

$$x = Bx'$$

troviamo che il polinomio p è metricamente equivalente a un polinomio della forma

$$p_1(x'_1, \dots, x'_n) = \lambda_1 (x'_1)^2 + \dots + \lambda_r (x'_r)^2 + 2(b')^T x' + \alpha, \quad (16.13)$$

dove

$$b' = B^T b = \begin{vmatrix} a'_{1,n+1} \\ \vdots \\ a'_{n,n+1} \end{vmatrix} \in \mathbb{R}^n$$

e $\lambda_1, \dots, \lambda_r \in \mathbb{R}$ sono gli autovalori non nulli di \mathcal{A}_n . Inoltre, a meno dell'ordine possiamo anche supporre che

$$\lambda_1 \geq \dots \geq \lambda_\rho > 0 > \lambda_{\rho+1} \geq \dots \geq \lambda_r,$$

dove $\rho = (r + |s|)/2$ è l'indice di positività di \mathcal{A}_n (vedi l'Osservazione 16.8).

Semplificata la parte quadratica, passiamo ai termini di primo grado. Ricordando (16.12) si vede che una traslazione $x' = X' + c$ elimina i termini di primo grado in p_1 se e solo se $c \in \mathbb{R}^n$ è soluzione del sistema $\mathcal{A}'_n x = -b'$. L'esistenza di una tale soluzione è una caratteristica affine delle quadriche:

Lemma 16.12 *Sia $p \in \mathbb{R}_2[x_1, \dots, x_n]$ un polinomio di secondo grado di matrice associata $A \in M_{n+1, n+1}(\mathbb{R})$. Allora:*

- (i) $\text{rg}(\mathcal{A}_n) \leq \text{rg}(A) \leq \text{rg}(\mathcal{A}_n) + 2$;
- (ii) *il sistema $\mathcal{A}_n x = -b$ ammette soluzione se e solo se $\text{rg}(A) \leq \text{rg}(\mathcal{A}_n) + 1$;*
- (iii) *se $q \in \mathbb{R}_2[x_1, \dots, x_n]$ è un polinomio affinementemente equivalente a p di matrice associata $A' \in M_{n+1, n+1}(\mathbb{R})$, allora il sistema $\mathcal{A}_n x = -b$ ammette soluzione se e solo se il sistema $\mathcal{A}'_n x' = -b'$ ammette soluzione.*

Dimostrazione. Indichiamo con \tilde{A} la matrice completa del sistema $\mathcal{A}_n x = -b$. Siccome \tilde{A} si ottiene da \mathcal{A}_n aggiungendo una colonna, e A si ottiene da \tilde{A} aggiungendo una riga, è chiaro che

$$\text{rg}(\mathcal{A}_n) \leq \text{rg}(\tilde{A}) \leq \text{rg}(\mathcal{A}_n) + 1 \quad \text{e} \quad \text{rg}(\tilde{A}) \leq \text{rg}(A) \leq \text{rg}(\tilde{A}) + 1;$$

in particolare (i) segue immediatamente. Se il sistema $\mathcal{A}_n x = -b$ ha soluzione, il Teorema di Rouché-Capelli ci dice che $\text{rg}(\tilde{A}) = \text{rg}(\mathcal{A}_n)$, per cui $\text{rg}(A) \leq \text{rg}(\mathcal{A}_n) + 1$. Viceversa, se il sistema non ha soluzione necessariamente $\text{rg}(\tilde{A}) = \text{rg}(\mathcal{A}_n) + 1$. Ma allora la matrice composta dalle prime n colonne di A (matrice che è la trasposta di \tilde{A}) ha rango $\text{rg}(\mathcal{A}_n) + 1$; se a questa aggiungiamo l'ultima colonna in modo da ottenere A , il rango aumenta ancora di 1, in quanto l'ultima colonna è linearmente indipendente dalle prime n (essendolo in \tilde{A}). Dunque se il sistema non ha soluzione necessariamente $\text{rg}(A) = \text{rg}(\mathcal{A}_n) + 2$, e anche (ii) è dimostrata.

Infine, se q è affinementemente equivalente a p , si ha $\text{rg}(A') = \text{rg}(A)$ e $\text{rg}(\mathcal{A}'_n) = \text{rg}(\mathcal{A}_n)$, per cui (iii) è conseguenza di (ii). □

Definizione 16.13 Sia \mathcal{Q}_p una quadrica di matrice associata $A \in M_{n+1, n+1}(\mathbb{R})$. Diremo che \mathcal{Q}_p è a centro se il sistema $\mathcal{A}_n x = -b$ ammette una soluzione; altrimenti diremo che \mathcal{Q}_p è un paraboloide. Ogni soluzione del sistema $\mathcal{A}_n x = -b$ è detta centro della quadrica. Nota che il Lemma 16.12 ci dice che \mathcal{Q}_p è a centro se e solo se $\text{rg}(A) \leq \text{rg}(\mathcal{A}_n) + 1$, e che è un paraboloide se e solo se $\text{rg}(A) = \text{rg}(\mathcal{A}_n) + 2$.

Osservazione 16.21 Il motivo del nome è il seguente: se \mathcal{Q}_p è a centro, e $c_0 \in \mathbb{R}^n$ è una qualunque soluzione del sistema $\mathcal{A}_n x = -b$, allora \mathcal{Q}_p è simmetrica rispetto a c_0 . Per dimostrarlo, effettuiamo il cambiamento di coordinate $x = x' + c_0$ che porta c_0 nell'origine, e indichiamo con $q \in \mathbb{R}_2[x_1, \dots, x_n]$ il polinomio dato da $q(x') = p(x' + c_0)$.

chiaramente, \mathcal{Q}_p è simmetrica rispetto a c_0 se e solo se \mathcal{Q}_q è simmetrica rispetto all'origine. Ora, la (16.12) ci dice che

$$q(x') = (x')^T \mathcal{A}_n x' + p(c_0);$$

in particolare, $q(-x') = q(x')$. Dunque $x' \in \mathcal{Q}_q$ se e solo se $-x' \in \mathcal{Q}_q$, cioè \mathcal{Q}_q è simmetrica rispetto all'origine.

Proseguiamo nello studio della quadrica \mathcal{Q}_p considerando prima di tutto il caso in cui sia a centro. Eravamo arrivati a dimostrare che \mathcal{Q}_p è metricamente equivalente a \mathcal{Q}_{p_1} , dove p_1 è dato da (16.13). Allora effettuando un cambiamento metrico di coordinate del tipo

$$x' = X' + c_0,$$

dove $c_0 \in \mathbb{R}^n$ è un centro di \mathcal{Q}_{p_1} , il polinomio p_1 si trasforma in

$$p_2(X'_1, \dots, X'_n) = \lambda_1 (X'_1)^2 + \dots + \lambda_r (X'_r)^2 + \alpha',$$

dove $\alpha' = p_1(c_0)$. A questo punto dobbiamo considerare due sottocasi.

Se $\text{rg}(A) = \text{rg}(\mathcal{A}_n)$, allora necessariamente (perché?) $\alpha' = 0$. Possiamo dividere p_2 per λ_1 (che è positivo per ipotesi); dunque p_2 è equivalente a un polinomio della forma

$$q_0(X'_1, \dots, X'_n) = (X'_1)^2 + h_2^2 (X'_2)^2 + \dots + h_\rho^2 (X'_\rho)^2 - h_{\rho+1}^2 (X'_{\rho+1})^2 - \dots - h_r^2 (X'_r)^2,$$

dove $\rho = i_+(\mathcal{A}_n)$ e $h_j = \sqrt{|\lambda_j|/|\lambda_1|} > 0$ per $j = 2, \dots, r$.

Se invece $\text{rg}(A) = \text{rg}(\mathcal{A}_n) + 1$, allora (perché?) $\alpha' \neq 0$; quindi dividendo p_2 per $|\alpha'| > 0$ vediamo che p è metricamente equivalente a uno dei seguenti due polinomi:

$$q_\pm(X'_1, \dots, X'_n) = h_1^2 (X'_1)^2 + \dots + h_\rho^2 (X'_\rho)^2 - h_{\rho+1}^2 (X'_{\rho+1})^2 - \dots - h_r^2 (X'_r)^2 \pm 1,$$

dove $\rho = i_+(\mathcal{A}_n)$ e $h_j = \sqrt{|\lambda_j|/|\alpha'|} > 0$ per $j = 1, \dots, r$. In particolare,

$$|s(A)| = |\rho - (r - \rho) \pm 1| = ||s| \pm 1|.$$

Se $|s| > 0$, questo implica che p è metricamente equivalente a un polinomio di tipo q_+ (rispettivamente, q_-) se e solo se $|s(A)| > |s|$ (rispettivamente, $|s(A)| < |s|$). Se invece $|s| = 0$ (cioè se ci sono tanti autovalori positivi quanti autovalori negativi) il polinomio p è metricamente equivalente sia a un polinomio di tipo q_+ che a un polinomio di tipo q_- : infatti il cambiamento metrico di coordinate

$$X'_j = \begin{cases} X_{j+\rho} & \text{se } 1 \leq j \leq \rho, \\ X_{j-\rho} & \text{se } \rho + 1 \leq j \leq 2\rho, \\ X_j & \text{se } 2\rho + 1 \leq j \leq n, \end{cases}$$

trasforma il polinomio q_+ nel polinomio

$$q(X_1, \dots, X_n) = -(h_{\rho+1}^2 X_1^2 + \dots + h_r^2 X_\rho^2 - h_1^2 X_{\rho+1}^2 + \dots + h_\rho^2 X_n^2) \pm 1.$$

che è metricamente equivalente a un polinomio di tipo q_- (basta moltiplicare per -1).

ESEMPIO 16.9 Le coniche di equazione $x^2 - y^2 - 1 = 0$ e $x^2 - y^2 + 1 = 0$ sono metricamente equivalenti: basta scambiare x con y e moltiplicare per -1 .

Finora abbiamo effettuato solo cambiamenti metrici di coordinate. Se ora invece effettuiamo il cambiamento affine di coordinate (con $h_1 = 1$ nel caso $\text{rg } A = \text{rg } \mathcal{A}_n$)

$$X'_j = \begin{cases} X_j/h_j & \text{per } j = 1, \dots, r, \\ X_j & \text{per } j = r+1, \dots, n, \end{cases} \quad (16.14)$$

trasformiamo infine q_0 e q_{\pm} rispettivamente nei polinomi

$$\begin{aligned} p_0(X_1, \dots, X_n) &= X_1^2 + \dots + X_{\rho}^2 - X_{\rho+1}^2 - \dots - X_r^2, \\ p_{\pm}(X_1, \dots, X_n) &= X_1^2 + \dots + X_{\rho}^2 - X_{\rho+1}^2 - \dots - X_r^2 \pm 1, \end{aligned}$$

che sono le possibili forme canoniche affini delle quadriche a centro.

Consideriamo ora il caso dei paraboloidi, in cui $\text{rg}(A) = \text{rg}(\mathcal{A}_n) + 2$. Per semplificare il termine di primo grado prima di tutto effettuiamo il cambiamento metrico di coordinate dato da

$$x'_j = \begin{cases} x''_j - a'_{j,n+1}/\lambda_j & \text{per } j = 1, \dots, r, \\ x''_j & \text{per } j = r+1, \dots, n, \end{cases}$$

che porta p_1 nel polinomio

$$p_2(x''_1, \dots, x''_n) = \lambda_1(x''_1)^2 + \dots + \lambda_r(x''_r)^2 + 2(b'')^T x + \alpha',$$

dove

$$b'' = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ a'_{r+1,n+1} \\ \vdots \\ a'_{n,n+1} \end{pmatrix} \quad \text{e} \quad \alpha' = \alpha - \frac{(a'_{1,n+1})^2}{\lambda_1} - \dots - \frac{(a'_{r,n+1})^2}{\lambda_r}.$$

Siccome \mathcal{Q}_p non è a centro, si ha $b'' \neq O$ (perché?). Dunque b'' è un vettore non nullo ortogonale a $\text{Span}(e_1, \dots, e_r)$; quindi possiamo completare $\{e_1, \dots, e_r, -b''/\|b''\|\}$ a una base ortonormale $\mathcal{C} = \{e_1, \dots, e_r, -b''/\|b''\|, v_{r+2}, \dots, v_n\}$ di \mathbb{R}^n . Effettuiamo quindi il cambiamento metrico di coordinate

$$x'' = CX'' ,$$

dove $C \in O(n)$ è la matrice di cambiamento di base dalla base canonica a \mathcal{C} . Questo cambiamento mantiene invariati i vettori e_1, \dots, e_r e manda e_{r+1} in $-b''/\|b''\|$; quindi il polinomio p_2 viene trasformato in un polinomio della forma

$$p_3(X''_1, \dots, X''_n) = \lambda_1(X''_1)^2 + \dots + \lambda_r(X''_r)^2 - 2kX''_{r+1} + \alpha',$$

dove $k = \|b''\| > 0$. A questo punto per liberarci del termine noto basta effettuare il cambiamento metrico di coordinate

$$X''_j = \begin{cases} X'_j & \text{per } j \neq r+1, \\ X'_{r+1} + \alpha'/2k & \text{per } j = r+1, \end{cases}$$

in modo da portare p_3 nel polinomio

$$p_4(X'_1, \dots, X'_n) = \lambda_1(X'_1)^2 + \dots + \lambda_r(X'_r)^2 - 2kX'_{r+1}.$$

Dividendo infine per $2k > 0$ vediamo che p è metricamente equivalente a

$$q_*(X'_1, \dots, X'_n) = h_1^2(X'_1)^2 + \dots + h_{\rho}^2(X'_{\rho})^2 - h_{\rho+1}^2(X'_{\rho+1})^2 - \dots - h_r^2(X'_r)^2 - X'_{r+1},$$

con $\rho = i_+(\mathcal{A}_n)$ e $h_j = \sqrt{|\lambda_j|/2k} > 0$ per $j = 1, \dots, r$. Effettuando poi il cambiamento affine di coordinate

$$X'_j = \begin{cases} X_j/h_j & \text{se } j = 1, \dots, r, \\ X_j & \text{se } j = r+1, \dots, n, \end{cases} \quad (16.15)$$

trasformiamo q in

$$p_*(X_1, \dots, X_n) = X_1^2 + \dots + X_{\rho}^2 - X_{\rho+1}^2 - \dots - X_r^2 - X_{r+1},$$

che è la forma canonica affine di un paraboloide.

Riassumendo, abbiamo dimostrato i seguenti due Teoremi:

Teorema 16.13 Sia $p \in \mathbb{R}_2[x_1, \dots, x_n]$ un polinomio di secondo grado in n variabili. Indichiamo con A la matrice simmetrica associata, e poniamo $r = \text{rg}(\mathcal{A}_n)$, $s = s(\mathcal{A}_n)$ e $\rho = (r + |s|)/2$. Allora esattamente una delle seguenti situazioni si verifica:

(i) $\text{rg}(A) = \text{rg}(\mathcal{A}_n)$, $|s(A)| = |s|$, la quadrica \mathcal{Q}_p è a centro e p è metricamente equivalente a un polinomio della forma

$$q_0(x_1, \dots, x_n) = x_1^2 + h_2^2 x_2^2 + \dots + h_{\rho}^2 x_{\rho}^2 - h_{\rho+1}^2 x_{\rho+1}^2 - \dots - h_r^2 x_r^2,$$

con $h_2 \geq \dots \geq h_{\rho} > 0$ e $0 < h_{\rho+1} \leq \dots \leq h_r$;

(ii) $\text{rg}(A) = \text{rg}(\mathcal{A}_n) + 1$, $|s(A)| < |s|$, la quadrica \mathcal{Q}_p è a centro e p è metricamente equivalente a un polinomio della forma

$$q_-(x_1, \dots, x_n) = h_1^2 x_1^2 + \dots + h_{\rho}^2 x_{\rho}^2 - h_{\rho+1}^2 x_{\rho+1}^2 - \dots - h_r^2 x_r^2 - 1,$$

con $h_1 \geq \dots \geq h_{\rho} > 0$ e $0 < h_{\rho+1} \leq \dots \leq h_r$;

(iii) $\text{rg}(A) = \text{rg}(\mathcal{A}_n) + 1$, $|s(A)| > |s|$, la quadrica \mathcal{Q}_p è a centro e p è metricamente equivalente a un polinomio della forma

$$q_+(x_1, \dots, x_n) = h_1^2 x_1^2 + \dots + h_{\rho}^2 x_{\rho}^2 - h_{\rho+1}^2 x_{\rho+1}^2 - \dots - h_r^2 x_r^2 + 1,$$

con $h_1 \geq \dots \geq h_{\rho} > 0$ e $0 < h_{\rho+1} \leq \dots \leq h_r$;

(iv) $\text{rg}(A) = \text{rg}(\mathcal{A}_n) + 2$, la quadrica \mathcal{Q}_p è un paraboloide e p è metricamente equivalente a un polinomio della forma

$$q_*(x_1, \dots, x_n) = h_1^2 x_1^2 + \dots + h_\rho^2 x_\rho^2 - h_{\rho+1}^2 x_{\rho+1}^2 - \dots - h_r^2 x_r^2 - x_{r+1},$$

con $h_1 \geq \dots \geq h_\rho > 0$ e $0 < h_{\rho+1} \leq \dots \leq h_r$.

Teorema 16.14 Sia $p \in \mathbb{R}_2[x_1, \dots, x_n]$ un polinomio di secondo grado in n variabili. Indichiamo con A la matrice simmetrica associata, e poniamo $r = \text{rg}(\mathcal{A}_n)$, $s = s(\mathcal{A}_n)$ e $\rho = (r + |s|)/2$. Allora esattamente una delle seguenti situazioni si verifica:

(i) $\text{rg}(A) = \text{rg}(\mathcal{A}_n)$, $|s(A)| = |s|$, la quadrica \mathcal{Q}_p è a centro e p è affinemente equivalente a un polinomio della forma

$$p_0(x_1, \dots, x_n) = x_1^2 + x_2^2 + \dots + x_\rho^2 - x_{\rho+1}^2 - \dots - x_r^2;$$

(ii) $\text{rg}(A) = \text{rg}(\mathcal{A}_n) + 1$, $|s(A)| < |s|$, la quadrica \mathcal{Q}_p è a centro e p è affinemente equivalente a un polinomio della forma

$$p_-(x_1, \dots, x_n) = x_1^2 + \dots + x_\rho^2 - x_{\rho+1}^2 - \dots - x_r^2 - 1;$$

(iii) $\text{rg}(A) = \text{rg}(\mathcal{A}_n) + 1$, $|s(A)| > |s|$, la quadrica \mathcal{Q}_p è a centro e p è affinemente equivalente a un polinomio della forma

$$p_+(x_1, \dots, x_n) = x_1^2 + \dots + x_\rho^2 - x_{\rho+1}^2 - \dots - x_r^2 + 1;$$

(iv) $\text{rg}(A) = \text{rg}(\mathcal{A}_n) + 2$, la quadrica \mathcal{Q}_p è un paraboloide e p è affinemente equivalente a un polinomio della forma

$$p_*(x_1, \dots, x_n) = x_1^2 + \dots + x_\rho^2 - x_{\rho+1}^2 - \dots - x_r^2 - x_{r+1}.$$

Questo conclude le classificazioni metrica e affine: ogni quadrica è metricamente (affinemente) equivalente a una di quelle elencate nel Teorema 16.13 (rispettivamente, 16.14), le quali sono a loro volta a due a due non metricamente (affinemente) equivalenti (Esercizio 16.20), con l'eccezione già vista del caso $\text{rg}(A) = \text{rg}(\mathcal{A}_n) + 1$ e $|s(\mathcal{A}_n)| = 0$.

Definizione 16.14 La quadrica di questa lista a cui la quadrica di partenza risulta essere equivalente si dice *forma canonica metrica (affine)* della quadrica, e questo procedimento si chiama *riduzione a forma canonica metrica (affine)*.

Dunque per trovare la forma canonica affine di una quadrica di matrice associata A basta calcolare $\text{rg}(A)$, $\text{rg}(\mathcal{A}_n)$, $|s(A)|$, $|s(\mathcal{A}_n)|$ e andarsi a leggere nel Teorema 16.14 la forma canonica affine corrispondente. Usando il Criterio di Cartesio questo è quindi un problema completamente risolto.

Osservazione 16.22 In pratica, per trovare la forma canonica affine di una quadrica \mathcal{Q}_p , si procede in questo modo. Prima di tutto si scrive la matrice simmetrica A associata al polinomio p . Utilizzando il Criterio di Cartesio si trovano il numero $i_+ = \rho$ di autovalori positivi e il numero $i_- = r - \rho$ di autovalori negativi di \mathcal{A}_n ; possiamo supporre $i_+ \geq i_-$ (se non è così, basta moltiplicare per -1 il polinomio p). Allora la parte quadratica della forma canonica affine di p sarà data dalla somma di i_+ quadrati col segno positivo e i_- quadrati col segno negativo. Poi si confronta $r = i_+ + i_-$ col rango di A : se sono uguali si è nel caso (i) del Teorema 16.14; se differiscono di 2, si è nel caso (iv) del Teorema. Infine, se $\text{rg}(A) = r + 1$, si calcola la segnatura di A e la si confronta con $s = i_+ - i_-$, per stabilire se si è nel caso (ii) o nel caso (iii) del Teorema.

ESEMPIO 16.10 Vogliamo trovare un cambiamento di coordinate che riduca in forma canonica affine la conica di equazione

$$p(x, y) = 5x^2 + 5y^2 - 6xy + 16\sqrt{2}x + 32 = 0.$$

Le matrici associate a questa conica sono

$$A = \begin{pmatrix} 5 & -3 & 8\sqrt{2} \\ -3 & 5 & 0 \\ 8\sqrt{2} & 0 & 32 \end{pmatrix}, \quad \mathcal{A}_2 = \begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix}.$$

Il polinomio caratteristico di \mathcal{A}_2 è $p_{\mathcal{A}_2}(\lambda) = (2 - \lambda)(8 - \lambda)$, per cui $\text{sp}(\mathcal{A}_2) = \{2, 8\}$, $\text{rg}(\mathcal{A}_2) = 2$ e $|s| = 2$. Inoltre

$$p_A(\lambda) = -\lambda^3 + 42\lambda^2 - 208\lambda - 128,$$

per cui $\text{rg}(A) = 3$ e $|s(A)| = 1$. Quindi siamo nel caso (ii) del Teorema 16.14 con $\rho = 2$, e la forma canonica affine è

$$X^2 + Y^2 - 1 = 0,$$

cioè \mathcal{C}_p è un'ellisse reale. Ma era richiesto il cambiamento di coordinate, per cui dobbiamo procedere. Una base ortonormale di \mathbb{R}^2 composta da autovettori di \mathcal{A}_2 è

$$\left\{ \left| \begin{array}{c} 1/\sqrt{2} \\ -1/\sqrt{2} \end{array} \right|, \left| \begin{array}{c} 1/\sqrt{2} \\ 1/\sqrt{2} \end{array} \right| \right\},$$

per cui il primo cambiamento di coordinate $x = Bx'$, effettuato utilizzando la matrice

$$B = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix},$$

ci dà il polinomio

$$p_1(x', y') = 8(x')^2 + 2(y')^2 + 16x' + 16y' + 32.$$

Il sistema $\mathcal{A}'_2 x' = -b'$ ha come unica soluzione

$$c_0 = -(\mathcal{A}'_2)^{-1} b' = \begin{pmatrix} -1 \\ -4 \end{pmatrix},$$

per cui il secondo cambiamento di coordinate $x' = X' + c_0$ ci porta a

$$p_2(X', Y') = 8(X')^2 + 2(Y')^2 - 8.$$

Quindi la forma canonica metrica (dividendo per 8) è

$$q_-(X', Y') = (X')^2 + \left(\frac{1}{2}\right)^2 (Y')^2 - 1.$$

Infine rimane il cambiamento di coordinate finale $X' = X$ e $Y' = 2Y$, che ci conduce alla forma canonica affine

$$p_-(X, Y) = X^2 + Y^2 - 1,$$

come previsto. Riassumendo, il cambiamento affine di coordinate effettuato è

$$\begin{cases} x = (x' + y')/\sqrt{2} = (X' + Y' - 5)/\sqrt{2} = (X + 2Y - 5)/\sqrt{2} \\ y = (-x' + y')/\sqrt{2} = (-X' + Y' - 3)/\sqrt{2} = (-X + 2Y - 3)/\sqrt{2}, \end{cases}$$

ovvero

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 2/\sqrt{2} \\ -1/\sqrt{2} & 2/\sqrt{2} \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} -5/\sqrt{2} \\ -3/\sqrt{2} \end{pmatrix}.$$

La Figura 16.3 mostra come si è modificata la posizione (e forma) della conica con i vari cambiamenti di coordinate.

ESEMPIO 16.11 Consideriamo la quadrica $\mathcal{Q}_p \subset \mathbb{R}^3$ di equazione

$$p(x, y, z) = x^2 + y^2 + 2xy + 2\sqrt{2}x + 2\sqrt{2}y + 2z + 4 = 0.$$

Le matrici associate sono

$$A = \begin{pmatrix} 1 & 1 & 0 & \sqrt{2} \\ 1 & 1 & 0 & \sqrt{2} \\ 0 & 0 & 0 & 1 \\ \sqrt{2} & \sqrt{2} & 1 & 4 \end{pmatrix} \quad \text{e} \quad \mathcal{A}_3 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix};$$

calcolando i polinomi caratteristici e applicando il Criterio di Cartesio vediamo subito che $\text{rg}(A) = 3$, $s(A) = 1$, $\text{rg}(\mathcal{A}_3) = 1$ e $s(\mathcal{A}_3) = 1$. Dunque si tratta di un paraboloide degenero, di forma canonica affine

$$X^2 - Y = 0;$$

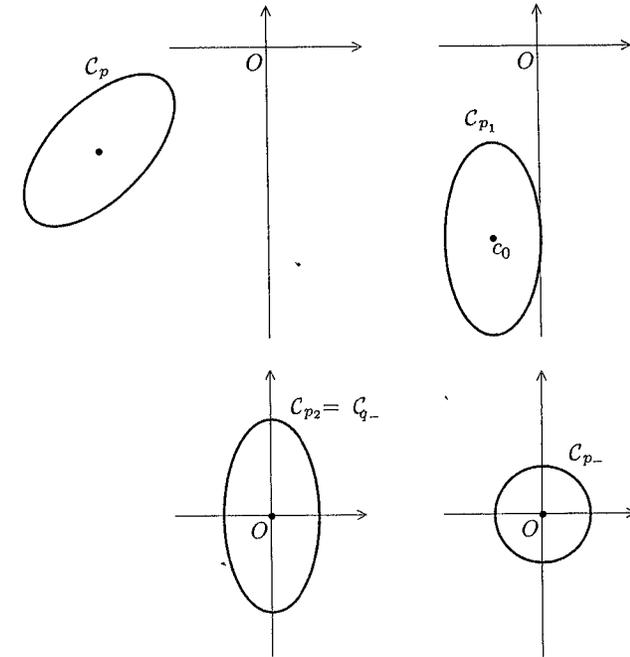


Figura 16.3 Riduzione a forma canonica di una conica.

è un cilindro parabolico. Vogliamo trovare un cambiamento di coordinate che lo riduca a forma canonica. Prima di tutto, $\text{sp}(\mathcal{A}_3) = \{2, 0\}$; il calcolo degli autovettori mostra che prendendo

$$B = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

il cambiamento di coordinate $x = Bx'$ trasforma p nel polinomio

$$p_1(x', y', z') = 2(x')^2 + 4x' + 2z' + 4.$$

Il cambiamento successivo è

$$\begin{cases} x' = x'' - 1, \\ y' = y'', \\ z' = z'', \end{cases}$$

che porta p_1 nel polinomio

$$p_2(x'', y'', z'') = 2(x'')^2 + 2z'' + 2;$$

in particolare,

$$b'' = \begin{vmatrix} 0 \\ 0 \\ 1 \end{vmatrix}.$$

Dunque dobbiamo completare l'insieme $\{e_1, -e_3\}$ a una base ortonormale di \mathbb{R}^3 ; aggiungendo per esempio $v_3 = e_2$ ricaviamo come matrice di cambiamento di base la matrice

$$C = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{vmatrix}.$$

Allora il cambiamento metrico di coordinate di matrice C

$$\begin{cases} x'' = X'', \\ y'' = Z'', \\ z'' = -Y'', \end{cases}$$

trasforma p_2 nel polinomio

$$p_3(X'', Y'', Z'') = 2(X'')^2 - 2Y'' + 2.$$

Procediamo col cambiamento

$$\begin{cases} X'' = X', \\ Y'' = Y' + 1, \\ Z'' = Z', \end{cases}$$

arrivando al polinomio

$$q_*(X', Y', Z') = 2(X')^2 - 2Y',$$

che è la forma canonica metrica. Infine il cambiamento

$$\begin{cases} X' = X/\sqrt{2}, \\ Y' = Y/2, \\ Z' = Z, \end{cases}$$

ci conduce alla forma canonica affine. Riassumendo, il cambiamento di coordinate che porta alla forma canonica affine è

$$\begin{cases} x = (x' + y')/\sqrt{2} = (x'' + y'' - 1)/\sqrt{2} = (X'' + Z'' - 1)/\sqrt{2} = (X' + Z' - 1)/\sqrt{2} \\ \quad = X/2 + Z/\sqrt{2} - 1/\sqrt{2}, \\ y = (x' - y')/\sqrt{2} = (x'' - y'' - 1)/\sqrt{2} = (X'' - Z'' - 1)/\sqrt{2} = (X' - Z' - 1)/\sqrt{2} \\ \quad = X/2 - Z/\sqrt{2} - 1/\sqrt{2}, \\ z = z' = z'' = -Y'' = -Y' - 1 = -Y/2 - 1, \end{cases}$$

cioè

$$\begin{vmatrix} x \\ y \\ z \end{vmatrix} = \begin{vmatrix} 1/2 & 0 & 1/\sqrt{2} \\ 1/2 & 0 & -1/\sqrt{2} \\ 0 & -1/2 & 0 \end{vmatrix} \begin{vmatrix} X \\ Y \\ Z \end{vmatrix} - \begin{vmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 1 \end{vmatrix}.$$

Esercizi

16.1 Studia il segno delle forme quadratiche su \mathbb{R}^2 date da $\varphi_1(x) = 2x_1^2 + 4x_1x_2 + 2x_2^2$ e $\varphi_2(x) = 3x_1^2 - 4x_1x_2 + 2x_2^2$.

16.2 Studia il segno della forma quadratica φ_a su \mathbb{R}^3 rappresentata dalla matrice

$$S_a = \begin{vmatrix} 4 & 0 & 4 \\ 0 & a & 1+a \\ 4 & 1+a & 5+a \end{vmatrix}$$

al variare di $a \in \mathbb{R}$.

16.3 Sia $g: V \times V \rightarrow \mathbb{R}$ una forma bilineare su uno spazio vettoriale V . Definiamo le funzioni $\varphi: V \rightarrow \mathbb{R}$ e $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$ ponendo

$$\varphi(v) = g(v, v) \quad \text{e} \quad \langle v, w \rangle = \frac{1}{2}[g(v, w) + g(w, v)].$$

Dimostra che φ è una forma quadratica su V il cui prodotto scalare associato è $\langle \cdot, \cdot \rangle$.

16.4 Sia $A \in M_{n,n}(\mathbb{R})$ una matrice simmetrica. Per $k = 1, \dots, n$ la *sottomatrice principale* $A_k \in M_{k,k}(\mathbb{R})$ di ordine k di A è la sottomatrice ottenuta considerando le prime k righe e colonne di A . Dimostra che A è definita positiva se e solo se $\det A_k > 0$ per $k = 1, \dots, n$. (Suggerimento: procedi per induzione su n , notando che A_k definisce un prodotto scalare su $\text{Span}\{e_1, \dots, e_k\}$, e che se A è definita positiva su un sottospazio di dimensione k allora ha almeno k autovalori positivi, per il Teorema di Sylvester.)

16.5 Trova la forma canonica affine e una base che porti in forma canonica affine la forma quadratica $\varphi: \mathbb{R}^3 \rightarrow \mathbb{R}$ data da

$$\varphi(x) = 4x_1^2 + x_2^2 + 3x_3^2 - 8x_1x_3.$$

16.6 Trova forme canoniche metrica e affine e basi che portino in forma canonica metrica e affine la forma quadratica $\varphi: \mathbb{R}^3 \rightarrow \mathbb{R}$ data da

$$\varphi(x) = x_1^2 + 2x_2^2 + 6x_3^2 + 8x_1x_3 + 2x_2x_3.$$

16.7 Trova due forme quadratiche su \mathbb{R}^2 affinementemente equivalenti ma non metricamente equivalenti.

16.8 Siano $A, A' \in M_{n,n}(\mathbb{R})$ due matrici congruenti. Dimostra che $\det A$ e $\det A'$ hanno lo stesso segno.

16.9 Dimostra che se $\langle \cdot, \cdot \rangle_0$ e $\langle \cdot, \cdot \rangle_1$ sono due prodotti scalari su V , con $\langle \cdot, \cdot \rangle_0$ definito positivo, allora esiste una base \mathcal{B} di V che è contemporaneamente ortonormale per $\langle \cdot, \cdot \rangle_0$ e ortogonale per $\langle \cdot, \cdot \rangle_1$.

16.10 Se $\ell \subset \mathbb{R}^2$ è una retta, e $v_0 \in \mathbb{R}^2$, indichiamo con $d(v_0, \ell)$ la distanza euclidea fra v_0 ed ℓ . Siano $v_1, \dots, v_k \in \mathbb{R}^2$ punti distinti tali che $v_1 + \dots + v_k = 0$. Dimostrare

che la retta ℓ che meglio approssima i punti v_1, \dots, v_k (cioè per cui la somma

$$f(\ell) = d(v_1, \ell)^2 + \dots + d(v_k, \ell)^2$$

è minima) ha equazione parametrica $x = tv_0$, dove $v_0 \in \mathbb{R}^2$ è un autovettore relativo all'autovalore più grande della matrice simmetrica $A = XX^T$, dove X è la matrice

$$X = [v_1 \ \dots \ v_k] \in M_{2,k}(\mathbb{R}).$$

Questo risultato è un caso particolare del *metodo dei minimi quadrati*. (*Suggerimento*: prima di tutto dimostra che se ℓ è una retta qualunque ed ℓ' è la sua retta di giacitura allora $f(\ell) \geq f(\ell')$, per cui possiamo limitarci a considerare rette per l'origine. Poi dimostra che se ℓ ha equazione parametrica $x = tv$ con $\|v\| = 1$ allora $f(\ell) = \sum_j \|v_j\|^2 - v^T Av$, per cui bisogna massimizzare la forma quadratica $\varphi(v) = v^T Av$ per $\|v\| = 1$. Diagonalizza A e concludi.)

16.11 Trova l'endomorfismo simmetrico associato alla forma quadratica $\varphi: \mathbb{R}^3 \rightarrow \mathbb{R}$ data da

$$\varphi(x) = 2x_1^2 + x_2^2 + 2x_1x_3 - 2x_2x_3$$

rispetto al prodotto scalare definito positivo

$$(x, y) = 2x_1y_1 + x_1y_2 + x_2y_1 + x_2y_2 + x_3y_3.$$

16.12 Sia $\varphi: \mathbb{R}_3[t] \rightarrow \mathbb{R}$ la funzione data da

$$\varphi(p) = p(0)^2 + p(1)^2 - p(-1)^2.$$

Dimostra che φ è una forma quadratica su $\mathbb{R}_3[t]$, e trovasi gli indici di positività, negatività e nullità.

16.13 Sia $S \in M_{n,n}(\mathbb{R})$ una matrice simmetrica tale che $\text{sp}(S) = \{0\}$. Dimostra che S è la matrice nulla. Deduci che se $p \in \mathbb{R}_2[x_1, \dots, x_n]$ è un polinomio di secondo grado di matrice associata A , allora \mathcal{A}_n ha sempre un autovalore non nullo.

16.14 Sia $\varphi: \mathbb{R}^2 \rightarrow \mathbb{R}$ una forma quadratica non identicamente nulla su \mathbb{R}^2 . Dimostra che φ è non degenere se e solo se la segnatura è pari. In particolare, dimostra che in dimensione 2 la segnatura determina completamente il rango.

16.15 Identifica quali coniche nell'Esempio 16.7 e quadriche nell'Esempio 16.8 sono a centro, e quali dei paraboloidi.

16.16 Trova un cambiamento di coordinate che riduca in forma canonica affine la conica di equazione $x^2 + 4xy + 2y^2 - x - y - 1 = 0$.

16.17 Trova usando gli invarianti la forma canonica affine delle coniche di equazione

$$3x^2 + 6xy + 5y^2 - 2x - 4y - 2 = 0, \quad xy + x - 3y + 4 = 0,$$

e solo dopo trova cambiamenti di coordinate che le riducano in forma canonica affine.

16.18 Trova usando gli invarianti la forma canonica affine delle quadriche di equazione

$$5x^2 + 4xy + 2y^2 + 3x - 2z = 0, \quad x^2 + y^2 + z^2 + 2xz + \sqrt{2}x + 2y + 3\sqrt{2}z + 3 = 0,$$

e solo dopo trova cambiamenti di coordinate che le riducano in forma canonica affine.

16.19 Trova la forma canonica affine della quadrica in \mathbb{R}^3 di equazione

$$x^2 + 2yz - 2x + 2z - 1 = 0.$$

16.20 Dimostra che due polinomi di uno stesso tipo fra quelli elencati nel Teorema 16.13 ma con coefficienti diversi non sono mai metricamente equivalenti. Dimostra che due polinomi di tipo diverso fra quelli elencati nel Teorema 16.14 possono essere affinemente equivalenti solo se sono uno di tipo p_+ , l'altro di tipo p_- e si ha $|s| = 0$.

16.21 Sia V uno spazio vettoriale su \mathbb{C} ; una *forma quadratica* su V è una funzione $\psi: V \rightarrow \mathbb{R}$ della forma $\psi(v) = \langle v, v \rangle$, dove $\langle \cdot, \cdot \rangle$ è un prodotto hermitiano su V . Dimostra l'equivalente dei Teoremi 16.1, 16.3, 16.6, 16.8 e 16.9 in questo caso. In particolare, il segno di un prodotto hermitiano si trova applicando il Criterio di Cartesio come per i prodotti scalari. (*Suggerimento*: gli autovalori di un endomorfismo hermitiano sono tutti reali.)

16.22 Dimostra che una matrice simmetrica $S \in M_{n,n}(\mathbb{R})$ è definita positiva se e solo se esiste una matrice simmetrica $C \in GL_n(\mathbb{R})$ invertibile tale che $S = C^2$.

COMPLEMENTI

16C.1 Il Criterio di Cartesio

In questo paragrafo vogliamo dimostrare il Criterio di Cartesio, ovvero il Teorema 16.4.(ii). Cominciamo con un lemma:

Lemma 16C.1 Sia $p \in \mathbb{R}[t]$ un polinomio di grado $n \geq 1$, e per $i = 1, \dots, n$ indichiamo con $p^{(i)} \in \mathbb{R}[t]$ la derivata i -esima di p . Siano $0 \leq i_0 < i_1 \leq n$ e $a_0 \in \mathbb{R}$ tali che

$$p^{(i_0)}(a_0) = \dots = p^{(i_1-1)}(a_0) = 0 \quad \text{e} \quad p^{(i_1)}(a_0) \neq 0.$$

Allora esistono $\lambda_1 < a_0 < \lambda_2$ tali che

$$p^{(i)}(\lambda_1)p^{(i+1)}(\lambda_1) < 0 \quad \text{e} \quad p^{(i)}(\lambda_2)p^{(i+1)}(\lambda_2) > 0$$

per $i = i_0, \dots, i_1 - 1$.

Dimostrazione. Scegliamo $\lambda_1 < a_0 < \lambda_2$ in modo che per $i = i_0, \dots, i_1 - 1$ l'unico radice di $p^{(i)}$ in $[\lambda_1, \lambda_2]$ sia a_0 , e in modo che $p^{(i_1)}$ non abbia radici in $[\lambda_1, \lambda_2]$.

Supponiamo prima di tutto che $p^{(i_1)}(a_0) > 0$. Allora $p^{(i_1-1)}$ è crescente in a_0 , per cui $p^{(i_1-1)}(\lambda_1) < 0$ e $p^{(i_1-1)}(\lambda_2) > 0$. Dunque $p^{(i_1-2)}$ è decrescente a sinistra di a_0 , e crescente alla sua destra; quindi $p^{(i_1-2)}(\lambda_1) > 0$ e $p^{(i_1-2)}(\lambda_2) > 0$. Continuando così otteniamo la tesi. Il caso $p^{(i_1)}(a_0) < 0$ è del tutto analogo. \square

Come conseguenza possiamo dimostrare un risultato preliminare nello spirito del Teorema di Sturm (vedi i Complementi al Capitolo 14), il *Teorema di Budan*:

Teorema 16C.2 (Budan) Sia $p \in \mathbb{R}[t]$ un polinomio di grado $n \geq 1$, e $a < b$ due numeri tali che $p(a)p(b) \neq 0$. Se $c \in \mathbb{R}$ indichiamo con V_c il numero di variazioni nella successione $(p(c), p'(c), \dots, p^{(n)}(c))$, dove $p^{(i)}$ indica la derivata i -esima di p . Sia inoltre $s \in \mathbb{N}$ il numero di radici, contate con la relativa molteplicità, di p nell'intervallo (a, b) . Allora esiste un $k \in \mathbb{N}$ tale che

$$V_a - V_b = s + 2k.$$

Dimostrazione. Scegliamo

$$a = a_0 < a_1 < \dots < a_m = b$$

in modo che nessuna derivata di p si annulli in alcuno degli intervalli aperti (a_{j-1}, a_j) . Prendiamo poi $\lambda_j \in (a_{j-1}, a_j)$ per $j = 1, \dots, m$.

Cominciamo col far vedere che $V_a = V_{\lambda_1}$. Per il Teorema dei valori intermedi, per $i = 0, \dots, n$ si ha $p^{(i)}(a_0)p^{(i)}(\lambda_1) \geq 0$; quindi se $p^{(i)}(a_0) \neq 0$ per ogni i ci siamo. Altrimenti, siano $i_0 < i_1$ tali che

$$p^{(i_0-1)}(a_0) \neq 0, \quad p^{(i_0)}(a_0) = \dots = p^{(i_1-1)}(a_0) = 0, \quad p^{(i_1)}(a_0) \neq 0;$$

siccome $p(a_0) \neq 0$ e $p^{(n)}(a_0) \neq 0$ abbiamo sicuramente $0 < i_0 < i_1 \leq n$. Per il Lemma 16C.1 (e il Teorema dei valori intermedi) $p^{(i)}(\lambda_1)$ ha lo stesso segno di $p^{(i_1)}(\lambda_1)$ per $i = i_0, \dots, i_1$. Dunque la successione $(p^{(i_0-1)}(a_0), 0, \dots, 0, p^{(i_1)}(a_0))$ e la successione $(p^{(i_0-1)}(\lambda_1), p^{(i_0)}(\lambda_1), \dots, p^{(i_1)}(\lambda_1))$ hanno lo stesso numero di variazioni. Ripetendo questo ragionamento tutte le volte che è necessario troviamo che $V_a = V_{\lambda_1}$.

Vediamo ora che relazione c'è fra V_{λ_j} e $V_{\lambda_{j+1}}$. Sicuramente esiste un minimo $i_0 \geq 0$ tale che $p^{(i_0)}(a_j) = 0$; distinguiamo due casi.

(a) $i_0 > 0$. Allora troviamo un $i_1 > i_0$ tale che

$$p^{(i_0-1)}(a_j) \neq 0, \quad p^{(i_0)}(a_j) = \dots = p^{(i_1-1)}(a_j) = 0, \quad p^{(i_1)}(a_j) \neq 0.$$

Il Lemma 16C.1 ci dice che nella successione $\mathbf{c}_j = (p^{(i_0)}(\lambda_j), \dots, p^{(i_1)}(\lambda_j))$ ci sono $i_1 - i_0$ variazioni, mentre in $\mathbf{c}_{j+1} = (p^{(i_0)}(\lambda_{j+1}), \dots, p^{(i_1)}(\lambda_{j+1}))$ non ci sono variazioni. Abbiamo quattro sottocasi:

(a.1) $p^{(i_0-1)}(a_j)p^{(i_1)}(a_j) > 0$ e $i_1 - i_0$ pari. In questo caso nella successione ottenuta aggiungendo $p^{(i_0-1)}(\lambda_j)$ all'inizio della successione \mathbf{c}_j ci sono ancora $i_1 - i_0$ variazioni, mentre nella successione ottenuta aggiungendo $p^{(i_0-1)}(\lambda_{j+1})$ all'inizio della successione \mathbf{c}_{j+1} continuano a non esserci variazioni.

(a.2) $p^{(i_0-1)}(a_j)p^{(i_1)}(a_j) > 0$ e $i_1 - i_0$ dispari. In questo caso la successione ottenuta aggiungendo $p^{(i_0-1)}(\lambda_j)$ all'inizio della successione \mathbf{c}_j ha $i_1 - i_0 + 1$ variazioni, mentre la successione ottenuta aggiungendo $p^{(i_0-1)}(\lambda_{j+1})$ all'inizio della successione \mathbf{c}_{j+1} continua a non avere variazioni.

(a.3) $p^{(i_0-1)}(a_j)p^{(i_1)}(a_j) < 0$ e $i_1 - i_0$ pari. In questo caso nella successione ottenuta aggiungendo $p^{(i_0-1)}(\lambda_j)$ all'inizio della successione \mathbf{c}_j ci sono $i_1 - i_0 + 1$ variazioni, mentre nella successione ottenuta aggiungendo $p^{(i_0-1)}(\lambda_{j+1})$ all'inizio della successione \mathbf{c}_{j+1} si trova una sola variazione.

(a.4) $p^{(i_0-1)}(a_j)p^{(i_1)}(a_j) < 0$ e $i_1 - i_0$ dispari. In questo caso nella successione ottenuta aggiungendo $p^{(i_0-1)}(\lambda_j)$ all'inizio della successione \mathbf{c}_j ci sono ancora $i_1 - i_0$ variazioni, mentre nella successione ottenuta aggiungendo $p^{(i_0-1)}(\lambda_{j+1})$ all'inizio della successione \mathbf{c}_{j+1} si è creata una variazione.

In tutti questi sottocasi, comunque, passando da λ_j a λ_{j+1} si perdono un numero *pari* di variazioni.

(b) $i_0 = 0$. Sia allora $i_1 > 0$ tale che

$$p(a_j) = \dots = p^{(i_1-1)}(a_j) = 0, \quad p^{(i_1)}(a_j) \neq 0;$$

quindi a_j è radice di p di molteplicità i_1 , e in questo caso il Lemma 16C.1 ci dice che passando da λ_j a λ_{j+1} perdiamo esattamente i_1 variazioni.

Infine, il ragionamento in (a) ci dice anche che passando da λ_m ad $a_m = b$ si perdono di nuovo un numero pari di variazioni (in quanto $p(b) \neq 0$). Quindi

$$V_a - V_b = (V_a - V_{\lambda_1}) + \sum_{j=1}^{m-1} (V_{\lambda_j} - V_{\lambda_{j+1}}) + (V_{\lambda_m} - V_b) = s + 2k$$

per qualche $k \in \mathbb{N}$. \square

Corollario 16C.3 Sia $p \in \mathbb{R}[t]$ un polinomio di grado $n \geq 1$. Allora il numero di variazioni di segno nella successione dei coefficienti non nulli di p è maggiore o uguale al numero delle radici positive di p , contate con la relativa molteplicità, e la differenza è un numero pari.

Dimostrazione. A meno di dividere p per un appropriato t^d , possiamo supporre che $p(0) \neq 0$. Scegliamo $M > 0$ in modo che (vedi il Lemma 14C.11) $[0, M]$ contenga tutte le radici positive di p , e $p^{(i)}(M)$ abbia sempre il segno del coefficiente direttivo di p ; in particolare $V_M = 0$. Poi notiamo che la successione $(p(0), p'(0), \dots, p^{(n)}(0))$ differisce dalla successione dei coefficienti di p solo per fattori positivi; quindi $V_0 = v$. Allora ci basta applicare il Teorema 16C.2 con $a = 0$ e $b = M$. \square

Da questo risultato possiamo finalmente dedurre il Criterio di Cartesio:

Teorema 16C.4 (Cartesio) Sia $p \in \mathbb{R}[t]$ un polinomio di grado n con tutte le radici reali. Allora il numero di radici positive, contate con la relativa molteplicità,

di p è uguale al numero di variazioni di segno nella successione dei coefficienti non nulli di p .

Dimostrazione. Chiaramente possiamo supporre che $p(0) \neq 0$. Siano v_+ il numero delle variazioni, e s_+ il numero delle radici positive; indichiamo poi con v_- il numero delle variazioni nella successione dei coefficienti non nulli del polinomio $p(-t)$, e con s_- il numero delle radici negative di p (che sono le radici positive di $p(-t)$, naturalmente). Per il Corollario 16C.3 sappiamo che $v_+ \geq s_+$ e $v_- \geq s_-$; inoltre $s_+ + s_- = n$ per ipotesi. Dunque ci basta dimostrare che $v_+ + v_- \leq n$.

Procediamo per induzione su n . Per $n = 1$, $p(t) = a_1t + a_0$, $p(-t) = -a_1t + a_0$ per cui $v_+ + v_- = 1$ o 0 a seconda che $a_0 \neq 0$ o $a_0 = 0$. Supponiamo allora che sia vero per tutti i polinomi di grado minore di n , e scriviamo

$$p(t) = a_nt^n + a_{n-1}t^{n-1} + \dots + a_0 = a_nt^n + q(t),$$

con $a_l \neq 0$; in particolare,

$$p(-t) = (-1)^n a_nt^n + q(-t).$$

Indichiamo con v'_+ le variazioni di segno nella successione dei coefficienti non nulli di $q(t)$, e con v'_- le variazioni di segno nella successione dei coefficienti non nulli di $q(-t)$. Per ipotesi induttiva, $v'_+ + v'_- \leq l$. Se $l = n - 1$ abbiamo

$$v_+ + v_- = v'_+ + v'_- + 1 \leq l + 1 = n,$$

e ci siamo. Se invece $l \leq n - 2$ allora abbiamo

$$v_+ + v_- \leq v'_+ + v'_- + 2 \leq l + 2 \leq n - 2 + 2 = n,$$

e ci siamo di nuovo. □

Conclusioni

Nel Capitolo 16 abbiamo visto (o meglio, cominciato a vedere) come sia possibile trattare con gli strumenti a nostra disposizione anche oggetti (polinomi, curve, equazioni) di secondo grado. Il passo successivo consisterebbe nel passare al terzo grado, e poi via via a tutti i gradi successivi... Si tratta ovviamente di tutta un'altra vicenda, che però ha le basi in quanto abbiamo visto: si entra nel reame della *Geometria Algebrica*, una delle branche più vaste e affascinanti della matematica.

Ma questo non è l'unico possibile sbocco degli argomenti trattati in questo libro. Un serio studio degli spazi vettoriali di dimensione infinita porta direttamente all'*Analisi Funzionale*, una teoria con ricchissime applicazioni e vaste implicazioni anche in fisica e ingegneria.

Oppure ci si potrebbe concentrare sulle matrici, ed esaminare meglio i gruppi di matrici che abbiamo incontrato (le matrici ortogonali, le matrici unitarie...), entrando nella teoria dei *Gruppi e Algebre di Lie*, teoria esteticamente ineccepibile e di vasta portata.

O anche si potrebbe voler cercare metodi più efficienti per risolvere grossi sistemi lineari, o per trovare le radici di polinomi di grado alto; tutti argomenti di *Analisi Numerica*, disciplina con importanti applicazioni praticamente in qualunque campo della scienza.

Insomma: l'Algebra Lineare descritta in questo libro è solo l'ingresso a quel vasto (molto più vasto di quanto probabilmente immaginavi solo qualche mese fa), importante e affascinante territorio che è la matematica contemporanea. Forse in futuro diventerai uno degli esploratori di questo nuovo continente, o forse ti limiterai a utilizzare i tesori che ne vengono estratti, o addirittura risalirai sulla tua nave per tornare verso lidi per te più accoglienti; ma in ogni caso non rimane che augurarsi che questa prima passeggiata lungo la spiaggia sia stata sufficientemente istruttiva e, perché no, persino interessante.

Appendice

Soluzione di alcuni esercizi

Come detto nella Prefazione, per spingerti a una lettura attiva del testo alcuni dettagli di alcune dimostrazioni sono stati lasciati per esercizio. Per completezza, riportiamo qui le soluzioni di questi esercizi; s'intende che vanno lette soltanto dopo che hai ripetutamente provato (possibilmente con successo) a risolverli da solo.

1.18 Ricorda che il grado di un polinomio è l'esponente della massima potenza di t che compare nel polinomio. Moltiplicando due polinomi, le due potenze di grado massimo si moltiplicano fra loro, e gli esponenti si sommano; quindi

$$\deg(p_1 p_2) = \deg p_1 + \deg p_2.$$

Sommando due polinomi, certo non compaiono potenze di t di grado più alto di quelle che già ci sono; quindi $\deg(p_1 + p_2) \leq \max\{\deg p_1, \deg p_2\}$. Se i due polinomi hanno lo stesso grado e coefficienti direttivi opposti, la somma ha grado strettamente minore; per esempio, se $p_1(t) = 2t$ e $p_2(t) = -2t$, la somma è il polinomio 0 che ha grado $-\infty$, come richiesto.

1.19 Gode delle proprietà (i)–(vii) e della proprietà (ix) del Paragrafo 1.3, per cui $\mathbb{R}[t]$ è un anello commutativo in cui esiste l'elemento neutro per la moltiplicazione (il polinomio costante 1) ma non un campo: non esiste l'inverso dei polinomi non costanti. Infatti se $\deg p > 0$ allora $\deg(pq) \geq \deg p > 0$ per ogni polinomio non nullo q (grazie all'Esercizio 1.18), per cui pq non può mai essere uguale al polinomio costante 1, che ha grado 0.

2.1 Se i tre punti allineati sono A_1 , A_2 e A_3 , si procede esattamente come nella dimostrazione della Proposizione 2.1. Se tutti e quattro i punti sono allineati, la tesi segue semplicemente dall'associatività della somma fra numeri reali. Grazie alla commutatività della somma, rimane soltanto da considerare il caso in cui O , A_1 e A_2 appartengano alla stessa retta r , mentre $A_3 \notin r$. Anche in questo caso la dimostrazione procede come già visto, con l'unica differenza che il parallelismo fra $\overline{A_2 B_1}$ e $\overline{O A_1}$ è ovvio (appartengono alla stessa retta), e la congruenza fra i due segue direttamente dalla definizione di somma di vettori.

2.2 La dimostrazione della (i) della Proposizione 2.2 nel caso $\lambda < 0$ è identica parola per parola a quella del caso $\lambda > 0$; cambia solo il disegno.

2.4 Che $0\overrightarrow{OA} = \overrightarrow{OO}$ per ogni vettore \overrightarrow{OA} è ovvio. Supponiamo invece che si abbia $\lambda\overrightarrow{OA} = \overrightarrow{OO}$ per un qualche vettore \overrightarrow{OA} , con $\lambda \neq 0$. Allora possiamo moltiplicare per λ^{-1} ottenendo

$$\overrightarrow{OO} = \lambda^{-1}\overrightarrow{OO} = \lambda^{-1}(\lambda\overrightarrow{OA}) = (\lambda^{-1}\lambda)\overrightarrow{OA} = 1\overrightarrow{OA} = \overrightarrow{OA},$$

per cui \overrightarrow{OA} è necessariamente il vettore nullo.

4.16 $\left\{ \begin{vmatrix} a \\ c \end{vmatrix}, \begin{vmatrix} b \\ d \end{vmatrix} \right\}$ è una base di \mathbb{R}^2 se e solo se i due vettori sono linearmente indipendenti se e solo se la matrice $A = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$ è non singolare. Supponiamo $a \neq 0$. Effettuando un'eliminazione di Gauss si vede che A è non singolare se e solo se $d - bc/a \neq 0$. Se invece $a = 0$, scambiando le righe vediamo che A è non singolare se e solo se $bc \neq 0$. In entrambi i casi A è non singolare se e solo se $ad - bc \neq 0$.

4.20 Bisogna dimostrare che $\mathcal{B} = \{v_1, \dots, v_n\}$ è un sistema di generatori. Se $v_0 \in V$, per la massimalità i vettori v_0, v_1, \dots, v_n sono linearmente dipendenti; quindi esistono $\alpha_0, \alpha_1, \dots, \alpha_n \in \mathbb{R}$ non tutti nulli tali che

$$\alpha_0 v_0 + \alpha_1 v_1 + \dots + \alpha_n v_n = 0.$$

Essendo v_1, \dots, v_n linearmente indipendenti, sicuramente $\alpha_0 \neq 0$; quindi ricaviamo

$$v_0 = -\frac{\alpha_1}{\alpha_0} v_1 - \dots - \frac{\alpha_n}{\alpha_0} v_n \in \text{Span}(v_1, \dots, v_n),$$

come desiderato.

7.13 Si ha

$$\begin{aligned} L_{A(B+C)} &= L_A \circ L_{B+C} = L_A \circ (L_B + L_C) = (L_A \circ L_B) + (L_A \circ L_C) = L_{AB+AC}, \\ L_{(\lambda A)B} &= L_{\lambda A} \circ L_B = (\lambda L_A) \circ L_B = \lambda(L_A \circ L_B) = \lambda L_{AB} = L_{\lambda(AB)}, \\ L_{A(\lambda B)} &= L_A \circ L_{\lambda B} = L_A \circ (\lambda L_B) = \lambda(L_A \circ L_B) = \lambda L_{AB} = L_{\lambda(AB)}, \end{aligned}$$

per cui $A(B+C) = AB+AC$ e $(\lambda A)B = \lambda(AB) = A(\lambda B)$.

7.19 A è invertibile se e solo se le sue colonne sono linearmente indipendenti se e solo se $ad - bc \neq 0$ (Esercizio 4.16). Siccome

$$\begin{vmatrix} d & -b \\ -c & a \end{vmatrix} \begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} ad - bc & 0 \\ 0 & ad - bc \end{vmatrix} = (ad - bc)I_2,$$

se A è invertibile l'inversa A^{-1} è data dalla formula indicata.

8.1 Si ha $w \in T(U_1) + T(U_2)$ se e solo se esistono $u_1 \in U_1$ e $u_2 \in U_2$ tali che si abbia $w = T(u_1) + T(u_2) = T(u_1 + u_2)$ se e solo se $w \in T(U_1 + U_2)$.

9.10 Se $h = 1$ basta calcolare il determinante sviluppando lungo la prima colonna. Sia allora vero per $h-1$, e dimostriamolo per h . Sviluppando lungo la prima colonna troviamo

$$\det \begin{vmatrix} B & C \\ O & D \end{vmatrix} = b_{11} \det \begin{vmatrix} B_{11} & \hat{C}_1 \\ O & D \end{vmatrix} + \dots + (-1)^{h+1} b_{h1} \det \begin{vmatrix} B_{h1} & \hat{C}_h \\ O & D \end{vmatrix},$$

dove \hat{C}_j è la matrice C a cui è stata tolta la riga j -esima. Per ipotesi induttiva,

$$\det \begin{vmatrix} B_{j1} & \hat{C}_j \\ O & D \end{vmatrix} = (\det B_{j1})(\det D);$$

quindi

$$\det A = [b_{11} \det(B_{11}) + \dots + (-1)^{h+1} b_{h1} \det(B_{h1})](\det D) = (\det B)(\det D).$$

10.1 Se $(a', b', c', d') = \lambda(a, b, c, d)$ con $\lambda \neq 0$ è chiaro che $ax + by + cz = d$ e $a'x + b'y + c'z = d'$ hanno le stesse soluzioni, per cui descrivono lo stesso piano. Viceversa, se le due equazioni descrivono lo stesso piano allora l'insieme delle soluzioni del sistema

$$\begin{cases} ax + by + cz = d, \\ a'x + b'y + c'z = d', \end{cases}$$

è un piano; per i Teoremi della dimensione e di Rouché-Capelli questo implica che

$$1 = \text{rg} \begin{vmatrix} a & b & c \\ a' & b' & c' \end{vmatrix} = \text{rg} \begin{vmatrix} a & b & c & d \\ a' & b' & c' & d' \end{vmatrix},$$

per cui $(a', b', c', d') = \lambda(a, b, c, d)$ per qualche $\lambda \neq 0$.

10.7 Infatti

$$a(ls + l't) + b(ms + m't) + c(ns + n't) = s \det \begin{vmatrix} l & l & l' \\ m & m & m' \\ n & n & n' \end{vmatrix} + t \det \begin{vmatrix} l' & l & l' \\ m' & m & m' \\ n' & n & n' \end{vmatrix} = 0$$

per ogni $s, t \in \mathbb{R}$, per cui il piano π ha equazione cartesiana

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

come annunciato.

10.8 Seguendo il suggerimento si vede che i tre punti sono allineati se e solo se i vettori $P_1 - P_0$ e $P_2 - P_0$ sono linearmente dipendenti, che accade se e solo se l'indicato determinante è zero.

10.13 Infatti sono tutte e sole le rette con vettore direttore proporzionale a quello della retta r .

10.29 Una base ha chiaramente la stessa orientazione di se stessa. Siccome una matrice e la sua inversa hanno determinante dello stesso segno, una base \mathcal{B}_1 ha la stessa orientazione di una base \mathcal{B}_2 se e solo se \mathcal{B}_2 ha la stessa orientazione di \mathcal{B}_1 . Infine il Teorema di Binet ci assicura che se \mathcal{B}_1 ha la stessa orientazione di \mathcal{B}_2 e \mathcal{B}_2 ha la stessa orientazione di \mathcal{B}_3 allora \mathcal{B}_1 ha la stessa orientazione di \mathcal{B}_3 , in quanto la matrice di cambiamento di base da \mathcal{B}_1 a \mathcal{B}_3 è il prodotto delle matrici di cambiamento di base da \mathcal{B}_1 a \mathcal{B}_2 e da \mathcal{B}_2 a \mathcal{B}_3 . Dunque "avere la stessa orientazione" è una relazione d'equivalenza sull'insieme delle basi. Fissiamo poi una base \mathcal{B}_+ di V , e sia \mathcal{B}_- la base ottenuta cambiando di segno l'ultimo vettore di \mathcal{B}_+ . Allora ogni altra base ha la stessa orientazione di \mathcal{B}_+ oppure la stessa orientazione di \mathcal{B}_- (perché il determinante può essere solo positivo o negativo); quindi l'insieme quoziente contiene solo due elementi (si dice che esistono solo due possibili orientazioni di V).

11.13 Se $\alpha_1 A^{j_1} + \dots + \alpha_k A^{j_k} = O$ è una relazione di dipendenza lineare fra le colonne di A , allora $\overline{\alpha_1} \overline{A}^{j_1} + \dots + \overline{\alpha_k} \overline{A}^{j_k} = O$ è una relazione di dipendenza lineare fra le colonne di \overline{A} . Quindi la dimensione di $\text{Im } A$ è uguale alla dimensione di $\text{Im } \overline{A}$, cioè $\text{rg } A = \text{rg } \overline{A}$.

11.14 Il determinante si calcola facendo opportune somme e prodotti degli elementi della matrice; dunque la Proposizione 11.2 ci assicura che coniugando gli elementi e poi calcolando il determinante oppure prima calcolando il determinante e poi coniugando si ottiene lo stesso risultato. Quindi $\det \overline{A} = \overline{\det A}$, da cui $\det A^H = \overline{\det A^T} = \det \overline{A}$. Di conseguenza $A^H = A$ implica $\det \overline{A} = \det A$, cioè $\det A \in \mathbb{R}$.

12.14 Se $v \perp U$ allora $\langle v, u_j \rangle = 0$ per $j = 1, \dots, r$ in quanto $u_1, \dots, u_r \in U$. Viceversa, supponiamo $\langle v, u_j \rangle = 0$ per $j = 1, \dots, r$. Allora

$$\langle v, \alpha_1 u_1 + \dots + \alpha_r u_r \rangle = \alpha_1 \langle v, u_1 \rangle + \dots + \alpha_r \langle v, u_r \rangle = 0$$

per ogni $\alpha_1, \dots, \alpha_r \in \mathbb{K}$, per cui $v \in U^\perp$.

12.21 Prendiamo $v, v_1, v_2, w \in V$ di coordinate $x, x_1, x_2, y \in \mathbb{R}^n$ rispettivamente. Allora

$$\begin{aligned} \langle w, v \rangle &= x^T y = (y^T x)^T = y^T x = \langle v, w \rangle; \\ \langle v_1 + v_2, w \rangle &= y^T (x_1 + x_2) = y^T x_1 + y^T x_2 = \langle v_1, w \rangle + \langle v_2, w \rangle. \end{aligned}$$

Il caso complesso è identico.

12.22 Per ogni $v \in V$ e $w \in W$ si ha

$$\begin{aligned} \langle (\lambda T)(v), w \rangle &= \lambda \langle T(v), w \rangle = \lambda \langle v, T^*(w) \rangle = \langle v, \overline{\lambda} T^*(w) \rangle, \\ \langle (T_1 + T_2)(v), w \rangle &= \langle T_1(v), w \rangle + \langle T_2(v), w \rangle \\ &= \langle v, T_1^*(w) \rangle + \langle v, T_2^*(w) \rangle = \langle v, (T_1^* + T_2^*)(w) \rangle, \end{aligned}$$

e quindi $(T_1 + T_2)^* = T_1^* + T_2^*$ e $(\lambda T)^* = \overline{\lambda} T^*$.

12.27 Abbiamo $P_U(u) = u$ per ogni $u \in U$ e $P_U(u^\perp) = O$ per ogni $u^\perp \in U^\perp$; quindi la matrice $A \in M_{n,n}(\mathbb{K})$ che rappresenta P_U rispetto alla base $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ è

$$A = \begin{vmatrix} I_r & O \\ O & O \end{vmatrix},$$

dove $r = \dim U$. Indichiamo con \tilde{B} la matrice che contiene per colonne le coordinate dei vettori di \mathcal{B}_2 rispetto a \mathcal{B}_0 . Essendo \mathcal{B}_0 ortonormale, le colonne di B sono ortogonali (rispetto al prodotto scalare canonico) alle colonne di \tilde{B} ; in particolare, $B^T \tilde{B} = O$. La matrice di cambiamento di base da \mathcal{B}_0 a \mathcal{B} è per definizione data da

$$C = |B \tilde{B}| \in GL_n(\mathbb{R});$$

siccome

$$C^T C = \begin{vmatrix} B^T B & O \\ O & \tilde{B}^T \tilde{B} \end{vmatrix},$$

l'Esercizio 9.10 ci dà

$$(\det B^T B)(\det \tilde{B}^T \tilde{B}) = \det(C^T C) \neq 0,$$

e le due matrici $B^T B$ e $\tilde{B}^T \tilde{B}$ sono necessariamente invertibili. Quindi

$$C^{-1} = \begin{vmatrix} (B^T B)^{-1} & O \\ O & (\tilde{B}^T \tilde{B})^{-1} \end{vmatrix} C^T.$$

Dunque la matrice che rappresenta P_U rispetto a \mathcal{B}_0 è

$$A' = C \begin{vmatrix} I_r & O \\ O & O \end{vmatrix} C^{-1} = |B \tilde{B}| \begin{vmatrix} (B^T B)^{-1} & O \\ O & O \end{vmatrix} \begin{vmatrix} B^T \\ \tilde{B}^T \end{vmatrix} = B(B^T B)^{-1} B^T.$$

Il caso complesso è analogo.

12.28 Infatti

$$\begin{aligned} \langle v_1 + v_2, w \rangle_T &= \langle T(v_1 + v_2), w \rangle = \langle T(v_1), w \rangle + \langle T(v_2), w \rangle = \langle v_1, w \rangle_T + \langle v_2, w \rangle_T; \\ \langle \lambda v, w \rangle_T &= \langle T(\lambda v), w \rangle = \lambda \langle T(v), w \rangle = \lambda \langle v, w \rangle_T; \\ \langle w, v \rangle_T &= \langle T(w), v \rangle = \langle w, T(v) \rangle = \langle T(v), w \rangle = \langle v, w \rangle_T; \end{aligned}$$

analogamente nel caso complesso. Infine, $v_0 \in \text{Ker } T$ se e solo se $T(v_0) = O$ se e solo se $\langle T(v_0), w \rangle = 0$ per ogni $w \in V$ se e solo se $\langle v_0, w \rangle_T = 0$ per ogni $w \in V$, per cui $\langle \cdot, \cdot \rangle_T$ è degenere se e solo se T è singolare.

12.29 T_1^* e $T_1 \circ T_2$ sono isometrie in quanto $(T_1^*)^* = T_1 = (T_1^{-1})^{-1} = (T_1^*)^{-1}$ e $(T_1 \circ T_2)^* = T_2^* \circ T_1^* = T_2^{-1} \circ T_1^{-1} = (T_1 \circ T_2)^{-1}$.

13.3 Infatti passano per lo stesso punto e hanno vettori direttori linearmente dipendenti.

14.1 Se $v, v_1, v_2 \in V_\lambda$ allora $T(v_1 + v_2) = T(v_1) + T(v_2) = \lambda v_1 + \lambda v_2 = \lambda(v_1 + v_2)$ e $T(\mu v) = \mu T(v) = \lambda(\mu v)$, per cui $v_1 + v_2, \mu v \in V_\lambda$. Se $V_\lambda = V$ allora $T(v) = \lambda v$ per ogni $v \in V$, cioè $T = \lambda \text{id}_V$.

14.13 Sia $\mathcal{C} = \{u_1, \dots, u_r\}$ una base di U e $\mathcal{B} = \{u_1, \dots, u_r, v_{r+1}, \dots, v_n\}$ una base di V ottenuta completando \mathcal{C} . Rispetto a \mathcal{B} l'endomorfismo T è rappresentato da una matrice della forma

$$A = \begin{array}{c|c} B & C \\ \hline O & D \end{array},$$

dove B è la matrice associata a $T|_U$ rispetto alla base $\{u_1, \dots, u_r\}$. Quindi l'Esercizio 9.10 ci dà

$$p_T(\lambda) = \det(A - \lambda I) = \det(B - \lambda I) \det(D - \lambda I) = p_{T|_U}(\lambda) \det(D - \lambda I),$$

per cui $p_{T|_U}$ divide p_T . In particolare, le radici di $p_{T|_U}$ sono anche radici di p_T , e da questo segue l'ultima affermazione.

14.20 Per ipotesi, $\tilde{B}A' = A\tilde{B}$. Scrivendo $\tilde{B} = P + iQ$ e separando parte reale e parte immaginaria troviamo $PA' = AP$ e $QA' = AQ$. Il polinomio $p(r) = \det(P + rQ)$ non è identicamente nullo (in quanto $p(i) = \det \tilde{B} \neq 0$); quindi esiste $r_0 \in \mathbb{R}$ tale che $p(r_0) \neq 0$. Dunque $B = P + r_0Q \in M_{n,n}(\mathbb{R})$ è invertibile e $BA' = AB$, cioè $A' = B^{-1}AB$ come richiesto.

16.13 Essendo S simmetrica, è diagonalizzabile; quindi esiste $B \in GL_n(\mathbb{R})$ tale che $S = BOB^{-1}$ (in quanto $\text{sp}(S) = \{0\}$) e dunque $S = O$. Siccome p è un polinomio di secondo grado, $\mathcal{A}_n \neq O$ e dunque $\text{sp}(\mathcal{A}_n) \neq \{0\}$.

16.20 Prima di tutto, il Corollario 16.11 ci assicura che polinomi di tipo diverso non possono essere affinemente (o metricamente) equivalenti, con l'eccezione già vista. Indichiamo con $A^{(\#)}$ la matrice associata ad un polinomio di tipo $q_\#$ (dove $\# = 0, +, -$ o $*$), e supponiamo che esistano $\sigma \in \mathbb{R}^*$, $B \in O(n)$ e $c \in \mathbb{R}^n$ tali che il polinomio q di matrice associata $A' = \sigma \tilde{B}^T A^{(\#)} \tilde{B}$ sia ancora di tipo $q_\#$. Prima di tutto $c = O$, perché altrimenti q conterrebbe dei termini di primo grado non permessi. Ora, $\mathcal{A}_n^{(\#)}$ è diagonale, e ha tutti gli autovalori disposti lungo la diagonale in modo univocamente determinato; essendo B ortogonale, $\mathcal{A}'_n = B^T \mathcal{A}_n^{(\#)} B$ ha gli stessi autovalori nello stesso ordine, per cui $\mathcal{A}'_n = \mathcal{A}_n^{(\#)}$. Infine in tutti e quattro i casi vi è un coefficiente fissato: quindi anche $\sigma = 1$ (oppure $\sigma = -1$ nel caso eccezionale già discusso), e dunque $q = q_\#$.

Indice analitico

- Affinità, 232.
- Aggiunta, 301.
- Algoritmo di Euclide, 262.
- Anello, 12.
- Angolo, 286, 315-317.
- Annullatore, 171, 173.
- Anti-omogeneità, 279.
- Applicazione, 2.
 - additiva, 103.
 - bilineare, 273.
 - identica, 107.
 - lineare, 103, 107, 141.
 - lineare invertibile, 143.
 - nulla, 108.
 - quoziente, 41.
- Argomento, 243.
- Asintoto, 334, 337.
- Asse, 218, 227, 338, 339.
- Assioma del buon ordinamento, 95.
 - del prodotto non vuoto, 95.
 - di scelta, 94.
- Autospazio, 345.
- Autovalore, 345, 379.
- Autovettore, 345.

- Bandiera, 93, 358.
- Base, 27, 31, 76, 96, 97, 339.
 - a bandiera, 358.
 - canonica, 76, 78, 253.
 - di autovettori, 372, 374.
 - di Jordan, 382, 383.
 - duale, 169.
 - ortogonale, 287.
 - ortonormale, 288, 291.
- Biduale, 170, 171.
- Birapporto, 240.
- Blocco di Jordan, 382.

- Cambiamento di coordinate affini, 224, 402.
 - di coordinate metrico, 314, 402.
- Campo, 12.
- Cardinale, 98.
- Cardinalità, 98, 104.
 - del continuo, 100.
- Catena, 94.
 - di Jordan, 382.
- Centro, 209, 211, 330, 341, 410.
- Cilindro, 339.
 - circolare retto, 339.
 - ellittico, 405.
 - immaginario, 405.
 - iperbolico, 405.
 - parabolico, 405.
- Circonferenza, 130, 330, 405.
- Classe d'equivalenza, 41.
 - di congruenza, 297.
 - di similitudine, 165.
- Codominio, 2.
- Coefficiente, 48.
 - di Fourier, 289.
 - direttivo, 5, 256.
 - di un polinomio, 256.
- Colatitudine, 342.
- Colonna, 49.
- Combinazione lineare, 72.
 - di equazioni, 54.
- Complesso coniugato, 247.
- Completamento a una base, 83.
- Composizione, 7, 17, 142.
- Condizione di ortogonalità, 315, 316, 318.
 - di parallelismo, 213-215, 217, 221.
- Conica, 335, 402.
 - degenerare, 405.
 - non degenerare, 338, 405.
- Coniugio, 247.
- Cono, 339.
 - circolare retto, 339.
 - immaginario, 405.
 - reale, 405.
- Coordinate, 26, 27, 31, 77.

- Coordinate (segue):
 - affini, 222.
 - cilindriche, 341.
 - non omogenee, 239, 242.
 - omogenee, 239, 242.
 - polari, 244.
 - sferiche, 342.
- Coseno direttore, 315.
 - iperbolico, 333.
- Criterio di Cartesio, 394, 423.
- Delta di Kronecker, 146.
- Determinante, 180-183, 186.
 - dell'inversa, 193.
 - della trasposta, 189.
 - di matrici diagonali, 181.
 - di matrici triangolari superiori, 182.
 - di un endomorfismo, 194, 351.
- Diagonale principale, 49.
- Diagramma commutativo, 157.
- Differenza di insiemi, 2.
- Dimensione, 84, 103, 134.
 - infinita, 84.
- Dimostrazione diretta, 51.
 - inversa, 51.
 - per assurdo, 51.
 - per induzione, 52.
- Dipendenza lineare, 75.
- Direttrice, 334, 335.
- Discriminante, 14.
- Distanza, 283, 319-322, 326.
- Disuguaglianza di Cauchy-Schwarz, 284.
 - triangolare, 247, 284.
- Divisione fra polinomi, 258.
- Dominio, 2.
- Eccentricità, 335.
- Elemento, 49.
 di volume, 191.
 - massimale, 95.
 - neutro, 11, 12.
- Eliminazione di Gauss, 59.
- Ellisse, 330, 405.
 - immaginaria, 404.
 - reale, 404.
- Ellissoide, 405.
 - immaginario, 405.
 - reale, 405.
- Endomorfismo, 107, 143, 164.
- Endomorfismo (segue):
 - associato a una forma quadratica, 392.
 - autoaggiunto, 303.
 - con tutti gli autovalori nel campo, 351.
 - diagonalizzabile, 349, 355, 363.
 - hermitiano, 303, 372, 374.
 - nilpotente, 359.
 - normale, 370, 372.
 - ortogonale, 304.
 - simmetrico, 303, 374.
 - triangolabile, 349, 376.
 - unitario, 304, 372.
- Equazione, 48.
 - cartesiana, 130, 135, 206.
 - cartesiana di un piano, 206.
 - cartesiana di una retta, 206, 207.
 - cartesiana lineare, 132.
 - parametrica, 33, 131, 135, 207.
 - parametrica di un piano, 35, 208.
 - parametrica di una retta, 34, 207, 208.
 - vettoriale di un piano, 33.
 - vettoriale di una retta, 32.
- Esponenziale complesso, 255.
- Fascio di piani, 218, 229.
 - di rette, 209, 228.
 - improprio, 227, 229.
- Fattori irriducibili di un polinomio, 264.
- Forma bilineare, 274, 310.
 - bilineare simmetrica, 274.
 - canonica, 395.
 - canonica affine, 400, 414.
 - canonica di Jordan, 379, 383, 384.
 - canonica metrica, 397, 413, 414.
 - quadratica, 389, 421.
 - quadratica affinementemente equivalente, 395.
 - quadratica degenera, 391.
 - quadratica indefinita, 391.
 - quadratica metricamente equivalente, 395.
 - quadratica (semi)definita negativa, 391.
 - quadratica (semi)definita positiva, 391.
 - sesquilineare, 280.
 - sesquilineare hermitiana, 280.

- Forma trigonometrica dei numeri complessi, 249.
- Formula di aggiunzione, 301.
 - di De Moivre, 249.
 - di Parseval, 289.
 - induttiva, 183.
 - ricorsiva, 183.
- Funzione, 2.
 - additiva, 107.
 - bigettiva, 6.
 - iniettiva, 6.
 - inversa, 6.
 - invertibile, 6.
 - omogenea, 107.
 - surgettiva, 5.
- Fuoco, 330, 331, 334, 335.
- Giacitura, 229.
- Grado di un polinomio, 5, 18, 256.
- Gruppo, 12.
 - commutativo, 12.
- Hermitianità, 279.
- Ideale, 259.
 - banale, 260.
 - generato, 260.
 - proprio, 260.
- Identità, 107.
- Immagine, 2, 111, 112.
 - inversa, 2.
- Invariante, 395.
- Incognita, 48.
- Indice di colonna, 49.
 - di negatività, 398, 400.
 - di nullità, 398, 400.
 - di positività, 398, 400.
 - di riga, 49.
 - nullo, 200.
- Indipendenza lineare, 75, 76.
- Insieme, 1.
 - bene ordinato, 94.
 - delle parti, 93.
 - equipotente, 98.
 - finito, 98, 104.
 - induttivo, 95.
 - infinito, 98.
 - libero, 96.
 - massimale di vettori linearmente indipendenti, 79.
- Insieme (segue):
 - numerabile, 100.
 - ordinato, 93.
 - quoziente, 41.
 - totalmente ordinato, 93.
 - vuoto, 1.
- Intersezione, 2, 86.
 - di sottospazi, 126.
- Inversa, 143, 149, 151.
- Inverso, 12.
- Iperbole, 331, 404.
- Iperboloide ellittico, 405.
 - iperbolico, 405.
- Iperpiano, 134.
- Ipotesi del continuo, 101.
 - induttiva, 52.
- Isometria, 328, 329.
 - lineare, 304.
- Isomorfismo, 29, 144.
 - canonico, 144, 171.
- Latitudine, 342.
- Lemma di Zorn, 95.
- Longitudine, 342.
- Lunghezza di un vettore, 281.
- Maggiorante, 95.
- Massimo, 94.
 - comun divisore, 261.
- Matrice, 49.
 - a scala, 120.
 - anti-hermitiana, 299.
 - antisimmetrica, 118, 299.
 - associata a un'applicazione lineare, 160.
 - associata a una forma bilineare, 296.
 - associata a una forma quadratica, 390.
 - associata a una forma sesquilineare, 296.
 - associata a una quadratica, 407.
 - cofattore, 198.
 - completa di un sistema lineare, 48.
 - congruente, 297.
 - coniugata, 253.
 - dei coefficienti di un sistema lineare, 48.
 - di cambiamento di base, 158.
 - di Jordan, 382.

- Matrice (segue):
 - diagonale, 49.
 - hermitiana, 299.
 - identica, 146.
 - indefinita, 391.
 - invertibile, 149, 150.
 - non singolare, 59.
 - nulla, 146.
 - ortogonale, 306, 307.
 - quadrata, 49.
 - (semi)definita negativa, 391.
 - (semi)definita positiva, 391.
 - simile, 164.
 - simmetrica, 118, 299, 374.
 - singolare, 59.
 - sparsa, 189.
 - trasposta coniugata, 253.
 - triangolare inferiore, 49.
 - triangolare superiore, 49.
 - unitaria, 306.
- Metodo dei minimi quadrati, 420.
 - di eliminazione di Gauss, 55.
- Minimo, 94.
 - comune multiplo, 265.
- Minore di una matrice, 183.
- Modulo di un numero complesso, 246.
 - di un vettore del piano, 243.
- Molteplicità algebrica, 354, 355.
 - di una radice, 251, 252, 265.
 - geometrica, 354, 355
- Multiplo, 26.
- Norma, 270, 272, 279, 281, 284.
- Nucleo, 111, 114, 126.
 - di un prodotto hermitiano, 280.
 - di un prodotto scalare, 274.
- Numero complesso, 245.
 - di variazioni, 365.
 - immaginario puro, 246.
 - intero, 10.
 - naturale, 10.
 - razionale, 10.
 - reale, 11.
 - relativo, 10.
- Omogeneità, 273.
- Omotetia, 25.
- Operazione elementare, 54, 55.
- Opposto, 11.
- Ordine di un sistema lineare, 48, 49.
- Orientazione, 224, 230, 315, 316.
- Origine, 20, 222.
- Orlare una sottomatrice, 195.
- Ortogonale, 294.
- Ortogonalità, 287.
- Parabola, 334, 404.
- Paraboloide, 410.
 - ellittico, 405.
 - iperbolico, 405.
- Parametrizzazione, 130.
 - lineare, 130.
- Parametro, 334.
 - di giacitura, 206.
 - direttore, 207.
- Parte immaginaria, 246.
 - quadratica di un polinomio, 407.
 - reale, 246.
- Permutazione degli indici, 201.
- Piano, 134.
 - affine, 232.
 - coincidente, 405.
 - complesso incidente, 405.
 - complesso parallelo, 405.
 - esterno, 342.
 - euclideo, 19, 313, 328.
 - generato, 31.
 - incidente, 220.
 - orientato, 316.
 - parallelo, 220.
 - proiettivo, 242.
 - reale incidente, 405.
 - reale parallelo, 405.
 - secante, 342.
 - tangente, 342.
- Pivot, 55, 59, 120, 122, 125, 379.
- Polinomio, 5, 12.
 - a coefficienti in un campo, 256.
 - affinementemente equivalente, 403.
 - caratteristico, 351.
 - con tutte le radici nel campo, 363.
 - irriducibile, 263.
 - metricamente equivalente, 403.
 - minimo, 357, 360, 361.
 - monico, 260.
 - relativamente primo, 262.
 - senza radici multiple, 363.
- Principio d'induzione, 52, 63.
 - del buon ordinamento, 63.
 - di Hausdorff, 96.

- Procedimento di ortogonalizzazione di Gram-Schmidt, 289.
- Prodotto, 11.
 - cartesiano, 2.
 - di cardinali, 101.
 - di matrici, 147.
 - di polinomi, 257.
 - hermitiano, 280.
 - hermitiano canonico, 279.
 - hermitiano degenerare, 280.
 - hermitiano indefinito, 280.
 - hermitiano non degenerare, 280.
 - hermitiano (semi)definito negativo, 280.
 - hermitiano (semi)definito positivo, 280.
 - misto, 326.
 - per scalari, 23, 29, 68, 69.
 - scalare, 274.
 - scalare associato a una forma quadratica, 389.
 - scalare canonico, 272.
 - scalare degenerare, 274.
 - scalare indefinito, 275.
 - scalare non degenerare, 273, 274.
 - scalare (semi)definito negativo, 275.
 - scalare (semi)definito positivo, 275.
 - vettore, 324.
- Proiettività, 240.
- Proiezione, 109, 153.
 - centrale, 235.
 - ortogonale, 272, 293, 316, 317.
 - parallela, 235.
- Proprietà affine, 230.
 - associativa, 11, 12.
 - commutativa, 11, 12.
 - distributiva, 12, 201.
- Punto all'infinito, 238, 239, 242.
 - improprio, 239, 242.
 - proprio, 239, 242.
- Quadrato del binomio, 13.
- Quadrica, 402.
 - a centro, 410.
 - affinementemente equivalente, 403.
 - degenerare, 407.
 - metricamente equivalente, 403.
 - non degenerare, 407.
- Quoziente, 258.
- Radiante, 15.
- Radice n -esima, 249, 250.
- Radice di un polinomio, 251, 258.
- Raggio, 330, 341.
- Rango, 113, 114, 116, 126.
 - di una forma quadratica, 398.
- Rapporto semplice, 231.
- Rappresentante, 41.
- Regola del parallelogramma, 20, 21.
- Relazione binaria, 40.
 - d'equivalenza, 40.
 - d'ordine, 93.
- Resto, 258.
- Restrizione, 5.
- Retta, 134.
 - affine, 232.
 - all'infinito, 242.
 - asintotica, 337.
 - coincidente, 213, 404.
 - complessa incidente, 404.
 - complessa parallela, 404.
 - contenuta, 337.
 - e piano incidenti, 216.
 - e piano paralleli, 216.
 - esterna, 337.
 - euclidea, 19, 328.
 - incidente, 33, 213.
 - orientata, 315.
 - parallela, 34, 213.
 - per due punti, 210, 211.
 - proiettiva, 239.
 - reale incidente, 404.
 - reale parallela, 404.
 - secante, 337.
 - sghemba, 34, 213.
 - tangente, 337.
- Riduzione a scala, 122, 125.
- Riflessione, 381.
- Riflessività, 40.
- Riga, 49.
- Risoluzione all'indietro, 51, 121.
- Scalare, 25.
- Segnatura, 398.
- Semiassi, 330, 331.
- Seno iperbolico, 333.
- Sequenza standard, 365.
- Sezioni coniche, 335.
- Sfera, 341, 405.
- Simmetria, 40.

- Sistema a scala, 120, 121.
 - compatibile, 48, 74.
 - di generatori, 76, 96.
 - di riferimento affine, 31, 222.
 - di riferimento cartesiano, 313.
 - di riferimento affine contraverso, 224.
 - di riferimento affine equiverso, 224.
 - equivalente, 54.
 - lineare, 48, 126.
 - lineare omogeneo, 71, 76.
 - lineare omogeneo associato, 71.
 - lineare quadrato, 50.
 - triangolare superiore, 50.
- Soluzione, 48.
- Somma, 11, 68, 69, 86, 89, 140.
 - di cardinali, 101.
 - di polinomi, 257.
 - di sottospazi, 126.
 - di vettori, 21, 29.
 - diretta, 88, 89.
- Sommatoria, 199, 200.
- Sottoinsieme, 1.
 - proprio, 1.
- Sottomatrice, 195.
 - principale, 419.
- Sottospazio affine, 105, 134.
 - affine parallelo, 136.
 - di giacitura, 134.
 - generato, 72.
 - vettoriale, 70, 85.
- Span, 31, 72, 74, 126.
- Spazio affine, 232.
 - delle applicazioni lineari, 162.
 - duale, 142, 168.
 - euclideo, 19, 313, 328.
 - vettoriale, 69.
 - vettoriale isomorfo, 144.
 - vettoriale metrico, 281.
 - vettoriale riflessivo, 171.
- Spettro, 345.
- Stella di piani, 211.
 - di rette, 209.
- Supplementare, 88, 104.
- Supplemento ortogonale, 294.
- Sviluppo di Laplace, 183, 187, 188.
- Teorema, 23.
 - degli orlati, 195.
 - del completamento, 81, 97.
- Teorema (*segue*):
 - del prodotto, 101.
 - della dimensione, 114.
 - di Binet, 192.
 - di Budan, 421.
 - di Cantor-Bernstein, 98.
 - di Cramer, 193.
 - di Grassmann, 87, 140.
 - di Hamilton-Cayley, 381.
 - di Pitagora, 14, 289.
 - di rappresentazione di Riesz, 311.
 - di Rouché-Capelli, 115.
 - di struttura, 105.
 - di Sturm, 366.
 - di Sylvester, 398.
 - di Talete, 14.
 - fondamentale dell'algebra, 252, 267.
 - spettrale, 374.
- Termine noto, 5, 48, 256.
- Traccia, 117, 351.
- Transitività, 40.
- Traslazione, 25.
- Trasposizione, 110.
- Trasposta, 110, 173, 301.
 - coniugata, 301.
 - di un'applicazione lineare, 173.
- Uguaglianza del parallelogramma, 309.
- Unione, 2.
- Unità immaginaria, 246.
- Valutazione, 169.
- Variabile dipendente, 122.
 - libera, 47, 122.
- Versore, 288.
 - direttore, 315.
 - ortogonale, 316.
- Vertice, 338, 339.
- Vettore, 67, 69.
 - applicato, 20.
 - asintotico, 337.
 - di giacitura, 33, 208.
 - direttore, 32, 207, 208.
 - libero, 42.
 - nullo, 20, 69.
 - proporzionale, 26.

Questo volume, sprovvisto del talloncino a fronte, è da considerarsi copia saggio-campione gratuito fuori commercio. Fuori campo applicazione IVA ed esente da bolla di accompagnamento (art. 22 L. 67/1987, art. 2, lett. I DPR 633/1972 e art. 4 n. 6 D.P.R. 627/1978).

GRUPPO EDITORIALE
 L'ESPRESSO
 00187 ROMA
 02 4781 2111