Algebra lineare, per matematici

Marco Manetti

versione preliminare, 26 agosto 2024

Marco Manetti
Dipartimento di Matematica Guido Castelnuovo
Sapienza Università di Roma

In copertina: "Trasformazione del caffè in teoremi" (2017), creata usando il software libero disponibile all'indirizzo www.wordclouds.com.

Note scritte dall'autore con il programma di composizione tipografica TEX (leggasi tek) mediante marcatura LATEX (leggasi latek), vedi it.wikipedia.org/wiki/LaTeX.

Avvertenza. La presente versione è ancora in forma preliminare e su certe parti ancora in forma sperimentale, ossia non sottoposta alla prova delle lezioni ed al feedback con gli studenti. Come conseguenza di ciò il numero di errori, di incongruenze e di argomenti trattati in maniera incompleta è ancora elevato: *i lettori sono avvisati!*

Sarò naturalmente lieto di ricevere da chiunque commenti, suggerimenti e segnalazioni di errori.

Esercizi. Alla fine di ogni sezione saranno proposti alcuni esercizi di diversa natura e difficoltà. Il simbolo ☕ indica gli esercizi ritenuti più difficili, il simbolo ♡ quelli per cui è riportata la soluzione, o una sua traccia, nell'ultimo capitolo[1] ed il simbolo Ⓐ quelli che richiedono nozioni impartite usualmente in altri insegnamenti universitari, tipicamente in quelli di Analisi Matematica. Questa versione contiene 869 esercizi.

Complementi. Ogni capitolo si conclude con una sezione di complementi, ossia di argomenti che di norma non fanno parte del programma di algebra lineare e che possono essere omessi in prima lettura senza compromettere la comprensione generale del testo.

Test preliminare di autovalutazione. Per ciascuno dei seguenti 15 concetti matematici assegnate un punteggio da 0 (mai sentito) a 5 (conosco perfettamente l'argomento). Rispondete onestamente senza consultare né amici né internet, sommate i punteggi e andate a verificare il vostro livello a pagina 391.

Concetti: Funzione esponenziale, divisore, funzione quadratica, numero proprio, equazione lineare, vettore, numero complesso, numero razionale, determinante, cambio di scala congiuntivo, poligono, frazione dichiarativa, figure congruenti, coseno, media aritmetica.

Questo lavoro è rilasciato sotto la licenza Creative Commons Attribuzione - Non commerciale - Condividi allo stesso modo 4.0 Internazionale (CC BY-NC-SA 4.0).

Ognuno è libero:

- Condividere, riprodurre, distribuire, comunicare al pubblico, esporre in pubblico, rappresentare, eseguire e recitare questo materiale con qualsiasi mezzo e formato.
- di creare opere derivate.

Alle seguenti condizioni:

Attribuzione: Di riconoscere il contributo dell'autore originario. In occasione di ogni atto di riutilizzazione o distribuzione, bisogna chiarire agli altri i termini della licenza di quest'opera.

Non commerciale: Di non usare quest'opera per scopi commerciali.

Condividi allo stesso modo: Le opere derivate devono essere distribuite con la stessa licenza del materiale originario.

Se si ottiene il permesso dal titolare del diritto d'autore, è possibile rinunciare ad ognuna di queste condizioni.

[1] Nella versione pdf, ad ogni simbolo ♡ è associato un collegamento ipertestuale alla soluzione

Indice

Capitolo 1. I sistemi lineari 1
1.1. Sistemi lineari 1
1.2. Sistemi ridondanti e rango 4
1.3. Il linguaggio degli insiemi 7
1.4. Brevi cenni sul metodo di Gauss 10
1.5. Alcune cose che si trovano nei libri di matematica 13
1.6. Prima esercitazione 15
1.7. Complementi: tavole della verità e fallacie logiche 19

Capitolo 2. Numeri interi e razionali 23
2.1. Numeri naturali, interi e razionali 23
2.2. Applicazioni tra insiemi 26
2.3. Il principio di induzione matematica 29
2.4. Il teorema fondamentale dell'aritmetica 35
2.5. Attenti all'errore! 37
2.6. Fattoriali e coefficienti binomiali 38
2.7. Il prodotto di composizione 44
2.8. Complementi: i numeri di Fibonacci e di Bernoulli 48

Capitolo 3. Numeri reali e complessi 53
3.1. I numeri reali 53
3.2. Estensioni quadratiche 58
3.3. I numeri complessi 60
3.4. Rappresentazione geometrica dei numeri complessi 63
3.5. Potenze e radici di numeri complessi 67
3.6. Campi di numeri 70
3.7. Campi, polinomi e funzioni razionali 71
3.8. Complementi: la formula di Cardano 78

Capitolo 4. Spazi e sottospazi vettoriali 81
4.1. Vettori numerici 81
4.2. Spazi vettoriali 84
4.3. Sottospazi vettoriali 86
4.4. Combinazioni lineari e generatori 89
4.5. Indipendenza lineare e teorema di scambio 92
4.6. Basi e dimensione 96
4.7. Semisemplicità e formula di Grassmann 100
4.8. Complementi: i numeri algebrici 101

Capitolo 5. Applicazioni lineari 105
5.1. Applicazioni lineari 105
5.2. Nucleo, iperpiani e sistemi di coordinate 109
5.3. Immagine e teorema del rango 112
5.4. Matrici ed applicazioni lineari 115
5.5. Spazi di applicazioni lineari 118
5.6. Complementi: successioni esatte e caccia al diagramma 122

Capitolo 6. Operazioni con le matrici 127
6.1. Traccia e trasposta 127

6.2. L'algebra delle matrici 129
6.3. Matrici invertibili 135
6.4. Rango di una matrice 139
6.5. Matrici speciali 142
6.6. Complementi: attenti a chi si incontra in rete 146

Capitolo 7. Riduzione a scala ed applicazioni 149
7.1. L'algoritmo di divisione 149
7.2. Matrici a scala 153
7.3. Operazioni sulle righe e riduzione a scala 156
7.4. Il teorema di Rouché–Capelli 160
7.5. Riduzione di Gauss–Jordan e calcolo della matrice inversa 165
7.6. Prodotto scalare e proiezioni ortogonali 168
7.7. Complementi: matrici a coefficienti interi e riduzione di Smith 174

Capitolo 8. Il determinante 177
8.1. Una formula per il determinante 177
8.2. Segnatura delle permutazioni ed unicità del determinante 181
8.3. Incroci e segnatura 186
8.4. Sviluppi di Laplace 190
8.5. Aggiunta classica e regola di Cramer 195
8.6. Complementi: lo Pfaffiano 198

Capitolo 9. Endomorfismi e polinomio caratteristico 203
9.1. Matrici simili 203
9.2. Spettro e polinomio caratteristico 206
9.3. Matrici compagne 209
9.4. Endomorfismi ed autovalori 211
9.5. Autovettori e sottospazi invarianti 214
9.6. Endomorfismi triangolabili 219
9.7. Endomorfismi diagonalizzabili 221
9.8. Il teorema fondamentale dell'algebra 225
9.9. Complementi: similitudine complessa di matrici reali 227

Capitolo 10. Polinomio minimo 229
10.1. Il polinomio minimo 229
10.2. Il teorema di Cayley–Hamilton 233
10.3. Polinomio minimo e diagonalizzazione 237
10.4. Matrici ed endomorfismi nilpotenti 240
10.5. Matrici simmetriche ed antisimmetriche reali 242
10.6. Criterio di Sylvester e regola dei segni di Cartesio 246
10.7. Complementi: il teorema di Cayley–Hamilton–Frobenius 249

Capitolo 11. Autospazi generalizzati e forma canonica di Jordan 253
11.1. La decomposizione di Fitting 253
11.2. Diagrammi di Young 256
11.3. Autospazi generalizzati 260
11.4. La forma canonica di Jordan 263
11.5. Moduli di persistenza 267
11.6. Complementi: la decomposizione di Dunford 268

Capitolo 12. Spazi duali 271
12.1. Spazi duali 271
12.2. Basi duali e sistemi di coordinate 275
12.3. Biduale e trasposta 277
12.4. Dualità vettoriale 279
12.5. Il principio del massimo 281
12.6. Complementi: forme alternanti 286

Capitolo 13. Spazi quoziente 295
13.1. Relazioni di equivalenza 295
13.2. Spazi vettoriali quoziente 297
13.3. La costruzione dei numeri reali 301
13.4. Complementi: insiemi ordinati e lemma di Zorn 307

Capitolo 14. Fattorizzazione di polinomi e forma canonica razionale 311
14.1. Il massimo comune divisore di polinomi 311
14.2. Polinomi irriducibili e fattorizzazione unica 313
14.3. Decomposizione primaria ed endomorfismi semisemplici 315
14.4. Spazi ciclici, irriducibili e indecomponibili 317
14.5. La forma canonica razionale 320
14.6. Complementi: il risultante di due polinomi 322

Capitolo 15. Forme bilineari e quadratiche 325
15.1. Nozioni base 325
15.2. Rango e congruenza di forme bilineari 328
15.3. Forme bilineari simmetriche 332
15.4. Applicazioni ortogonali e riflessioni 337
15.5. Forme quadratiche reali e complesse 339
15.6. Eliminazione di Gauss simmetrica 347
15.7. Il teorema spettrale 349
15.8. I gruppi ortogonali 353
15.9. Complementi: Radici quadrate e scomposizione polare 357

Capitolo 16. Spazi e trasformazioni affini 359
16.1. Combinazioni baricentriche, spazi e sottospazi affini 359
16.2. Il rapporto semplice 363
16.3. Inviluppo affine, dimensione e formula di Grassmann 366
16.4. Polinomi di Bernstein e curve di Bézier 368
16.5. Complementi: spazi affini astratti e modellati 370

Capitolo 17. Complementi: le trascendenze famose 377
17.1. Irrazionalità di e ed l 378
17.2. L'operatore di derivazione 379
17.3. Irrazionalità di π 381
17.4. La trascendenza di l 382
17.5. La trascendenza di e 384
17.6. Polinomi simmetrici 386
17.7. La trascendenza di π 389

Capitolo 18. Note, commenti, curiosità e riferimenti bibliografici 391

Capitolo 19. Soluzioni e suggerimenti di alcuni esercizi 395

CAPITOLO 1

I sistemi lineari

Per gli addetti ai lavori, l'algebra lineare è lo studio degli spazi vettoriali e delle applicazioni lineari. Per i novizi, e probabilmente per molti studenti appena iscritti all'università, la precedente definizione può suonare un po' troppo autoreferenziale. Obiettivo di questo capitolo è di dare, in maniera molto informale e privilegiando gli aspetti intuitivi, un primo assaggio di algebra lineare rivisitando i ben noti sistemi di equazioni lineari, mentre spazi vettoriali ed applicazioni lineari verranno introdotti nei prossimi capitoli. È bene precisare subito che l'algebra lineare non serve solo a studiare i sistemi lineari ma possiede innumerevoli legami e relazioni con quasi tutti gli ambiti e le aree matematiche.

1.1. Sistemi lineari

Problema: un ragazzo vede conigli e polli in un cortile. Conta 18 teste e 56 zampe, quanti polli e conigli ci sono nel cortile?

In tale problema abbiamo due quantità incognite, ossia il numero di polli ed il numero di conigli. Essendo questo un libro che vuole insegnare il mestiere di matematico, non perdiamo tempo in chiacchiere fuori contesto e indichiamo con le lettere x, y le nostre due incognite, più precisamente chiamiamo x il numero di conigli ed y il numero di polli nel cortile. Quello che dobbiamo fare è trovare i valori di x, y che soddisfano *entrambe* le equazioni:

(1) $x + y = 18$ (equazione delle teste),
(2) $4x + 2y = 56$ (equazione delle zampe).

Più in generale, quando abbiamo alcune equazioni e cerchiamo i valori che le risolvono tutte, parliamo di *sistema di equazioni*; solitamente i sistemi di equazioni si rappresentano con una parentesi graffa alla sinistra delle equazioni incolonnate in verticale. Nel nostro caso:

$$\begin{cases} x + y = 18 \\ 4x + 2y = 56 \end{cases}$$

Tale sistema si può risolvere usando il *metodo di sostituzione*: in tale metodo si suppone che il sistema abbia soluzioni e si utilizza un'equazione per calcolare il valore di un'incognita in funzione delle altre, poi si sostituisce tale valore nelle rimanenti equazioni ottenendo un sistema con un'equazione ed un'incognita in meno. Nel nostro caso:

$$\begin{cases} x = 18 - y \\ 4x + 2y = 56 \end{cases}, \quad \begin{cases} x = 18 - y \\ 4(18 - y) + 2y = 56 \end{cases}, \quad \begin{cases} x = 18 - y \\ 72 - 2y = 56 \end{cases},$$

$$\begin{cases} x = 18 - y \\ -2y = -16 \end{cases}, \quad \begin{cases} x = 18 - y \\ y = 8 \end{cases}, \quad \begin{cases} x = 10 \\ y = 8 \end{cases}.$$

Abbiamo quindi dimostrato che il precedente problema dei polli e conigli ammette una *unica soluzione*, ossia $x = 10$, $y = 8$.

Le cose però possono andare diversamente. Consideriamo per esempio il seguente problema: *In un cortile ci sono conigli e polli, tutti integri e senza amputazioni. Sapendo che ci sono 10 teste e 20 orecchie, dire quanti sono i polli e quanti sono i conigli.*

In questo caso il sistema diventa

$$\begin{cases} x + y = 10 \\ 2x + 2y = 20 \end{cases}$$

Se proviamo a risolverlo con il metodo di sostituzione troviamo

$$\begin{cases} x = 10 - y \\ 2x + 2y = 20 \end{cases}, \qquad \begin{cases} x = 10 - y \\ 2(10 - y) + 2y = 20 \end{cases}, \qquad \begin{cases} x = 10 - y \\ 20 = 20 \end{cases}.$$

Ma l'equazione $20 = 20$ è sempre soddisfatta, non ci dà alcuna informazione su polli e conigli e quindi la possiamo omettere. Dunque il nostro sistema si riduce alla sola equazione $x = 10 - y$ che non ha una unica soluzione.

Consideriamo adesso un altro problema: *in un cortile ci sono conigli e polli, tutti integri ed in ottima salute. Sapendo che ci sono 10 teste e 21 orecchie, dire quanti sono i polli e quanti sono i conigli.*

In questo caso il sistema diventa

$$\begin{cases} x + y = 10 \\ 2x + 2y = 21 \end{cases}$$

e se proviamo a risolverlo con il metodo di sostituzione troviamo

$$\begin{cases} x = 10 - y \\ 2x + 2y = 20 \end{cases}, \qquad \begin{cases} x = 10 - y \\ 2(10 - y) + 2y = 21 \end{cases}, \qquad \begin{cases} x = 10 - y \\ 20 = 21 \end{cases}.$$

In questo caso l'equazione $20 = 21$ non è mai soddisfatta (è impossibile) e quindi *l'ipotesi che il sistema abbia soluzioni porta ad una contraddizione*. In tale caso non rimane quindi che dedurre che il sistema *non ammette soluzioni*. Un sistema di equazioni lineari che non ammette soluzioni viene anche detto **inconsistente**.

Riepilogando, dato un sistema di equazioni lineari, la procedura di soluzione per sostituzione porta ad una delle seguenti tre conclusioni:

(1) unica soluzione,
(2) nessuna soluzione,
(3) soluzioni multiple.

Adesso però ci sorge un dubbio: nell'applicazione del metodo di sostituzione abbiamo scelto sia l'incognita da esplicitare sia l'equazione da utilizzare allo scopo. Diverse scelte portano a diversi procedimenti; ma chi ci assicura che *diverse scelte portano alla stessa conclusione?*

Forse in questo caso la preoccupazione è eccessiva, in fin dei conti il metodo di sostituzione, qualunque strada percorra, porta sempre all'insieme delle soluzioni. Ci sono però altre importanti informazioni ottenibili dal metodo di sostituzione la cui indipendenza dalle scelte non è affatto chiara.

L'algebra lineare nasce dall'esigenza di fornire un quadro teorico alla teoria dei sistemi di equazioni lineari, in grado di fornire la risposta al precedente dubbio (e non solo).

Esempio 1.1.1. Cip e Ciop calcolano le soluzioni del sistema di due equazioni in tre incognite

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

Applicando il metodo di sostituzione Cip trova:

$$\begin{cases} x = 1 - y - z \\ (1 - y - z) - y + z = 0 \end{cases}, \qquad \begin{cases} x = 1 - y - z \\ -2y = -1 \end{cases}, \qquad \begin{cases} y = \frac{1}{2} \\ x = \frac{1}{2} - z \end{cases}.$$

Invece Ciop trova:

$$\begin{cases} y = 1 - x - z \\ x - (1 - x - z) + z = 0 \end{cases}, \qquad \begin{cases} y = 1 - x - z \\ 2x + 2z = 1 \end{cases},$$

$$\begin{cases} y = 1 - x - z \\ z = \frac{1}{2} - x \end{cases} \qquad \begin{cases} y = 1 - x - (\frac{1}{2} - x) \\ z = \frac{1}{2} - x \end{cases} \qquad \begin{cases} y = \frac{1}{2} \\ z = \frac{1}{2} - x \end{cases}.$$

Le due soluzioni sono entrambe corrette e solo apparentemente diverse. Infatti Cip trova che le soluzioni sono l'insieme delle terne (x, y, z) tali che $y = 1/2$ e $x = 1/2 - z$, mentre Ciop trova che le soluzioni sono l'insieme delle terne (x, y, z) tali che $y = 1/2$ e $z = 1/2 - x$. Tali insiemi chiaramente coincidono.

ESEMPIO 1.1.2. Vogliamo trovare due numeri a, b tali che

$$\frac{1}{(t-1)(t-2)} = \frac{a}{t-1} + \frac{b}{t-2}.$$

Eseguendo la somma del secondo membro mediante l'usuale regola di messa a denominatore comune si ha

$$\frac{a}{t-1} + \frac{b}{t-2} = \frac{a(t-2)}{(t-1)(t-2)} + \frac{b(t-1)}{(t-1)(t-2)} = \frac{a(t-2)+b(t-1)}{(t-1)(t-2)}$$

e quindi i numeri a, b devono soddisfare l'uguaglianza

$$1 = a(t-2) + b(t-1) = (a+b)t + (-b-2a).$$

Equiparando i coefficienti delle potenze di t troviamo il sistema

$$\begin{cases} a+b=0 \\ -b-2a=1 \end{cases}$$

che ha come soluzione $a=-1$ e $b=1$.

Esercizi.

1 (Tratto da un concorso a dirigente comunale). Un contadino alleva mucche e galline. Se possiede 60 capi che hanno complessivamente 172 zampe, quante sono rispettivamente le mucche e le galline?

2. Risolvere i seguenti sistemi di equazioni lineari:

$$\begin{cases} 3x+y=5 \\ 3x+2y=4 \end{cases}, \qquad \begin{cases} 2x+7y=3 \\ 6x+21y=4 \end{cases}, \qquad \begin{cases} 3x-2y+z=1 \\ 12x-8y-z=4 \\ x+y+z=1 \end{cases}.$$

3. Un contadino, avendo incontrato dei politici, voleva tirare 5 pomodori a ciascuno, ma per fare questo gli mancavano 2 pomodori. Allora egli tirò 4 pomodori a ciascuno e così gli rimasero 5 pomodori. Quanti erano i politici?

4. Batman, Robin e Catwoman corrono per le strade di Gotham City con le loro potenti moto. La moto di Batman viaggia a una velocità doppia di quella di Catwoman e il tempo che impiega Robin per attraversare il Robert Kane Memorial Bridge è uguale alla somma dei tempi che impiegano Batman e Catwoman per percorrere il medesimo ponte. Ad uno stesso istante Batman e Robin imboccano i due estremi dell'Old Steam Tunnel, lungo 736 metri. Quanti metri percorre Batman prima di scontrarsi con Robin? (Si suppone che tutti viaggiano a velocità costante.)

5. Determinare quattro numeri sapendo che le loro somme a tre a tre sono $9, 10, 11$ e 12.

6 (♡). Francesca ha 26 anni, ama disegnare paesaggi e leggere poesie. A 18 anni si è iscritta al WWF ed ha partecipato attivamente ad iniziative contro la vivisezione. Possiede una forte personalità ed ama gli abbigliamenti etnici. Quale tra le seguenti affermazioni ritenete sia la meno probabile?

(1) Francesca lavora in banca;
(2) Francesca è vegetariana;
(3) Francesca è vegetariana e lavora in banca.

7 (Eureka!). Una moneta del peso 16 grammi è fatta di oro e piombo ed il suo peso in acqua è di 15 grammi. Sapendo che il peso specifico dell'oro è $19,3$ volte quello dell'acqua e quello del piombo $11,3$ volte, calcolare quanti grammi di oro contiene la moneta.

8. Trovare tre numeri a, b, c tali che

$$\frac{2}{t^3-t} = \frac{a}{t-1} + \frac{b}{t+1} + \frac{c}{t}.$$

9. Determinare per quali valori del parametro k il sistema lineare

$$\begin{cases} (k-1)x + (3k+1)y = 2k \\ (k-1)x + (4k-2)y = (k+3) \\ 2x + (3k+1)y = k-3 \end{cases},$$

di tre equazioni nelle variabili x, y, possiede soluzioni.

1.2. Sistemi ridondanti e rango

Domanda: che cosa hanno in comune i seguenti sistemi di equazioni lineari?

$$(A) \quad \begin{cases} x+y=1 \\ 2x-y=3 \\ 0=0 \end{cases}, \qquad (B) \quad \begin{cases} x+y=1 \\ 2x-y=3 \\ x+y=1 \end{cases}, \qquad (C) \quad \begin{cases} x+y=1 \\ 2x-y=3 \\ 3x=4 \end{cases}.$$

Risposta: hanno tutti più equazioni del necessario.

Spieghiamo caso per caso la risposta. Il sistema (A) contiene come terza equazione $0=0$ che è sempre verificata, non influisce sul sistema e può essere tolta. Nel sistema (B) la terza equazione è uguale alla prima: in particolare se una coppia di numeri x, y soddisfa le prime due equazioni di (B) allora soddisfa anche la terza. Dunque anche in questo caso la terza equazione è ridondante e può essere tolta.

Nel sistema (C) le tre equazioni sono diverse tra loro, tuttavia è facile osservare che la terza è la somma delle prime due: infatti $(x+y)+(2x-y)=3x$ e $1+3=4$. Ne segue che se x, y soddisfano le prime due equazioni, allora

$$3x = (x+y)+(2x-y) = 1+3 = 4$$

e quindi soddisfano anche la terza: anche in questo caso la terza equazione non aggiunge alcuna ulteriore informazione e può essere tolta. Vediamo adesso un caso leggermente più complicato:

$$(D) \quad \begin{cases} x+y=1 \\ 2x-y=3 \\ 3y=-1 \end{cases}.$$

Siccome $2(x+y)-(2x-y)=3y$ e $2\cdot 1-3=-1$, la terza equazione è uguale al doppio della prima meno la seconda; dunque se x, y soddisfano le prime due equazioni allora

$$3y = 2(x+y)-(2x-y) = 2(1)-(3) = -1$$

e soddisfano anche la terza. Dunque la terza equazione si può omettere dal sistema senza alterare l'insieme delle soluzioni.

Definizione 1.2.1. Diremo che un'equazione di un sistema è **combinazione lineare** delle altre se è la somma delle rimanenti equazioni moltiplicate per opportuni numeri.

Esempio 1.2.2. Nel sistema

$$\begin{cases} x+y+z=1 \\ 2x-y-z=3 \\ x-y+2z=0 \\ 4y+7z=-3 \end{cases}$$

la quarta equazione è combinazione lineare delle altre e più precisamente è la somma di 3 volte la prima, di -2 volte la seconda e della terza. Infatti si ha:

$$3(x+y+z)-2(2x-y-z)+(x-y+2z) = 4y+7z, \qquad 3(1)-2(3)+(0) = -3.$$

Osserviamo inoltre che ognuna delle quattro equazioni è combinazione lineare delle altre tre: ad esempio

$$x+y+z = \frac{2}{3}(2x-y-z)-\frac{1}{3}(x-y+2z)+\frac{1}{3}(4y+7z); \qquad 1 = \frac{2}{3}(3)-\frac{1}{3}(0)+\frac{1}{3}(-3).$$

Se i valori $x, y, z, \dots$ soddisfano un insieme finito di equazioni lineari, allora soddisfano anche ogni loro combinazione lineare. Ciò implica che ogni equazione che è combinazione lineare delle rimanenti può essere tolta dal sistema senza modificare le soluzioni.

DEFINIZIONE 1.2.3. Diremo che un sistema di equazioni lineari è **ridondante** se qualche sua equazione è combinazione lineare delle altre.

Dunque ogni sistema ridondante può essere "semplificato" togliendo una equazione. Se dopo aver tolto un'equazione che è combinazione lineare delle altre il sistema è ancora ridondante possiamo ripetere la procedura fino a quando il sistema non è più ridondante ed il numero di equazioni di tale sistema viene detto **rango**. Uno degli obiettivi dell'algebra lineare è quello di mostrare che *il rango non dipende dalla scelta delle equazioni che sono tolte* perché combinazione lineare delle rimanenti.

ESEMPIO 1.2.4. Il rango del sistema dell'Esempio 1.2.2 è uguale a 3. Infatti togliendo una qualunque delle 4 equazioni otteniamo un sistema che non è ridondante. Per esempio, togliendo la quarta equazione si ottiene

$$\begin{cases} x+y+z=1 \\ 2x-y-z=3 \\ x-y+2z=0 \end{cases}$$

Il valore $x = 4/3$, $y = -1/3$, $z = 0$ soddisfa le prime due equazioni ma non la terza e questo implica che la terza equazione non è combinazione lineare delle prime due. Similmente si mostra che nessuna delle tre è combinazione lineare delle rimanenti. Nel seguito vedremo metodi più semplici e pratici per determinare se un sistema è o meno ridondante.

ESEMPIO 1.2.5. Consideriamo il sistema

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

e cerchiamo di capire se la terza equazione è combinazione lineare delle prime due. Per definizione la terza equazione è combinazione lineare delle prime due se e solo se esistono due numeri a, b tali che

$$a(x+y+z)+b(2x+y-z)=x-y+3z, \qquad a(0)+b(0)=0.$$

Uguagliando membro a membro i coefficienti di x, y, z otteniamo il sistema di tre equazioni

$$\begin{cases} a+2b=1 \\ a+b=-1 \\ a-b=3 \end{cases}$$

che si dimostra facilmente essere senza soluzioni. Quindi $x-y+3z=0$ **non è** combinazione lineare di $x+y+z=0$ e $2x+y-z=0$.

Esercizi.

10. Per ciascuno dei tre sistemi lineari:

$$\begin{cases} x+y+z=0 \\ 2x+y-z=0 \\ 3x+2y=0 \end{cases}, \qquad \begin{cases} x+2y+z=1 \\ 2x+y-z=0 \\ x-y-2z=-1 \end{cases}, \qquad \begin{cases} x+2y+z+w=1 \\ 2x+y-z=0 \\ x-y-2z-w=-1 \\ 4x+2y-2z=0 \end{cases},$$

scrivere l'ultima equazione come combinazione lineare delle precedenti.

11. Dei seguenti tre sistemi lineari, solamente uno è inconsistente. Senza fare i conti, ma solamente guardandoli, dire quale:

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

12. Determinare il rango dei seguenti sistemi lineari:

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

13. Discutere il sistema di equazioni lineari nelle incognite x, y, z

$$\begin{cases} x + ky + z = 0 \\ kx - y + 2z = 3 \\ x + y - 2z = k - 2 \end{cases}$$

al variare del parametro k. Più precisamente, si chiede di determinare per quali valori di k il sistema è inconsistente, e per quali valori ammette soluzioni multiple; inoltre, per ciascun valore di k per cui esistono soluzioni multiple si chiede di calcolare il rango del sistema.[1]

14. I seguenti problemi sono tratti dal "Compendio d'Analisi" di Girolamo Saladini, pubblicato in Bologna MDCCLXXV.

(1) Un cane si dà ad inseguire una lepre in distanza di passi numero $= a$, e la velocità del cane è alla velocità della lepre come m/n. Si cerca dopo quanti passi il cane giugnerà la lepre.
(2) Caio interrogato, che ora fosse, rispose, che le ore scorse dalla mezza notte, alle ore, che rimanevano fino al meriggio[2] erano come 2/3. Si vuol sapere qual ora fosse accennata da Caio.
(3) Sempronio volendo distribuire certi denari a certi poveri osserva, che se ne dà tre a ciascuno, ne mancano otto, se ne dà due, ne avanzano tre. Si vuol sapere il numero de' poveri, e de' denari.
(4) Rispose Tizio ad un, che domandavagli quanti anni avesse: se il numero de' miei anni si moltiplichi per 4, ed al prodotto si aggiunga 15, si ha un numero, che di tanto eccede il 150, quanto il numero 100 eccede il numero de' miei anni. Si cerca il numero degli anni di Tizio.
(5) Caio per mantenimento della sua famiglia spende il primo anno scudi 380, il rimanente dell'entrata lo mette a traffico, ed il frutto, che ne trae è un quarto della somma messa a traffico; il secondo anno spesi i soliti 380 scudi pone il rimanente a guadagno e ne ricava pure un quarto; lo stesso in tutto e per tutto gli succede nel terzo anno, passato il quale si accorge che la sua entrata è cresciuta di un sesto. Si vuol sapere quanto fosse nel primo anno l'entrata di Caio.
(6) Si fa una cena, che importa 12 scudi. Due de' commensali non pagano, gli altri sborsano uno scudo in più di quello che avrebbero dato se la spesa si fosse egualmente ripartita in tutti i commensali. Ciò supposto si vuol sapere quanti essi sono.

15. Per ciascuno dei seguenti sistemi lineari, dire senza fare calcoli (semplicemente osservandoli) se possiedono soluzioni oppure se sono inconsistenti:

$$\begin{cases} x_1 - 2x_2 + x_3 + 3x_4 = 1 \\ 2x_1 - x_3 = 0 \\ 2x_1 - 4x_2 + 2x_3 + 6x_4 = 3 \\ x_1 + x_2 + x_3 + x_4 = 318 \end{cases}, \qquad \begin{cases} x_1 - x_2 + x_3 + x_4 = 0 \\ 2x_1 - x_2 - x_3 = 0 \\ x_1 - 4x_2 + 2x_3 + 3x_4 = 0 \\ x_1 + x_2 - x_3 + x_4 = 0 \end{cases}.$$

16. Dato il sistema lineare

$$\begin{cases} kx + y + z + w = 1 \\ x + ky + z + w = 1 \\ x + y + kz + w = 1 \\ x + y + z + kw = 1 \end{cases}$$

nelle quattro incognite x, y, z, w e dipendente dal parametro k:

[1] Vedremo in seguito che è questo l'unico caso dove è interessante calcolare il rango. Dimostreremo inoltre che se il sistema ammette una unica soluzione, allora il suo rango è uguale al numero di incognite.

[2] Mezzogiorno [nda].

(1) aggiungere come quinta equazione la somma delle quattro equazioni date;
(2) dire per quali valori di k il sistema possiede una soluzione con $x = y = z = w$;
(3) dire per quali valori di k il sistema è incompatibile.

17. Il problema dei polli e conigli nel cortile è tratto da una poesia di Elio Pagliarani, intitolata "La merce esclusa". È divertente leggere la soluzione proposta nel medesimo testo e qui di seguito riportata con alcune (lievi) variazioni rispetto all'originale.

> "Si consideri una specie di animale a sei zampe e due teste: il conigliopollo. Ci sono nel cortile 56 zampe diviso 6 zampe = 9 coniglipolli, nove coniglipolli che necessitano di $9 \times 2 = 18$ teste. Restano dunque $18 - 18 = 0$ teste nel cortile. Ma questi animali hanno $9 \times 6 = 54$ zampe allora $56 - 54 = 2$. Restano due zampe nel cortile. Si consideri quindi un'altra specie di animale, che potrebbe essere il coniglio spollato, che ha 1 testa -1 testa $= 0$ teste, 4 zampe -2 zampe $= 2$ zampe: le due zampe che stanno nel cortile. C'è dunque nel cortile 9 coniglipolli + 1 coniglio spollato. Detto in altri termini 9 conigli +9 polli +1 coniglio -1 pollo. Ed ora *i conigli coi conigli e i polli coi polli*, si avrà $9 + 1 = 10$ conigli, $9 - 1 = 8$ polli."

Provate a riscrivere la stessa soluzione in linguaggio un po' più algebrico? Cosa intende dire l'autore con *i conigli coi conigli e i polli coi polli*?

18 (♡). Vedremo ben presto che l'algebra lineare sviluppa tutta la sua potenza quando gli è consentito, tra le altre cose, di fare combinazioni lineari di oggetti non necessariamente omogenei, come ad esempio polli e conigli. È talvolta utile rappresentare ogni possibile combinazione lineare

$$a\,\text{Coniglio} + b\,\text{Pollo}$$

con il punto nel piano cartesiano di coordinate (a, b); abbiamo quindi che $(1, 0)$ è il coniglio, $(0, 1)$ il pollo, $(1, 1)$ il conigliopollo, $(1, -1)$ il coniglio spollato e così via. Sapreste dare un nome ai fantasiosi animali corrispondenti ai punti $(-1, 1), (-1, 0), (0, 2), (1, -2), (3, 0), (5, -6)$?

1.3. Il linguaggio degli insiemi

Nelle precedenti sezioni abbiamo incontrato il termine "insieme" nel senso usualmente inteso in matematica, ossia con il significato di *collezione, aggregato, cumulo, famiglia, raccolta, conglomerato, classe, coacervo, accozzaglia ecc..*

Il concetto di insieme, che rappresenta una pluralità di elementi considerati come un tutt'uno, sta alla base della matematica ed è talmente intuitivo che non richiede (per il momento) ulteriori spiegazioni; ogni insieme è caratterizzato dagli elementi appartenenti ad esso. Ad esempio l'insieme dei numeri pari coincide con l'insieme dei numeri non dispari: pur avendo diverso nome i due insiemi possiedono gli stessi elementi.

Talvolta, se possibile, indicheremo un insieme elencandone gli elementi racchiusi tra parentesi graffe, per cui $\{1, 2, 6\}$ rappresenta l'insieme i cui elementi sono i numeri $1, 2$ e 6. Quando gli elementi dell'insieme sono troppi per poter essere elencati tutti, ma sono tuttavia determinati da alcuni di essi in maniera induttiva e non ambigua, si può utilizzare la notazione dei puntini di sospensione: ad esempio, l'insieme dei numeri interi compresi tra 1 e 100 può essere efficacemente indicato $\{1, 2, \dots, 100\}$, mentre la tabellina del 7 può essere indicata $\{7, 14, 21, \dots, 70\}$.

Un insieme si dice **finito** se possiede un numero finito (limitato) di elementi; un insieme si dice **infinito** se non è finito, ossia se che contiene un numero infinito (illimitato) di elementi. Il mondo che ci circonda offre moltissimi esempi di insiemi finiti. Il concetto di insieme infinito è decisamente più astratto e concettuale, e tuttavia comprensibile, almeno a livello intuitivo, a tutte le persone funzionalmente alfabetizzate.

La formula $a \in A$ sta ad indicare che a è un elemento dell'insieme A. La stessa formula si legge anche a **appartiene** ad A, od anche A **contiene** a. Scriveremo invece $a \notin A$ se l'elemento a non appartiene all'insieme A. Ad esempio:

$$2 \in \{1, 2, 6\}, \quad 3 \notin \{1, 2, 6\}, \quad 6 \in \{1, 2, 6\}, \quad 5 \notin \{1, 2, 6\},$$

$$\text{mamma dei bischeri} \in \{\text{donne perennemente in stato interessante}\}.$$

Quando tutti gli elementi di un insieme A sono anche elementi dell'insieme B scriveremo $A \subseteq B$ e diremo che A è un **sottoinsieme di** B (espressioni equivalenti: A è incluso in B, A è contenuto in B, A è parte di B ecc.): ad esempio

$$\{1,2\} \subseteq \{1,2,3\} \subseteq \{1,2,3\}, \qquad \{\text{uomini}\} \subseteq \{\text{mammiferi}\} \subseteq \{\text{animali}\}.$$

Due insiemi sono uguali se hanno gli stessi elementi: dati due insiemi A, B, si ha $A = B$ se e solo se valgono contemporaneamente entrambe le inclusioni $A \subseteq B$ e $B \subseteq A$. Scriveremo $A \not\subseteq B$ per indicare che A non è un sottoinsieme di B, ossia che esiste almeno un elemento $a \in A$ che non appartiene a B. Scriveremo $A \neq B$ per indicare che gli insiemi A, B non sono uguali.

Un sottoinsieme $A \subseteq B$ si dice **proprio** se $A \neq B$, ossia se esiste almeno un elemento $b \in B$ che non appartiene ad A; scriveremo $A \subset B$ per indicare che A è un sottoinsieme proprio di B. L'esistenza del numero 3, del gatto Silvestro e del canarino Titti folnisce una plova convincente che si hanno le inclusioni proprie

$$\{1,2\} \subset \{1,2,3\}, \qquad \{\text{uomini}\} \subset \{\text{mammiferi}\} \subset \{\text{animali}\}.$$

Se A e B sono due insiemi indichiamo con $A \cap B$ la loro **intersezione**. Per definizione, l'intersezione $A \cap B$ è l'insieme formato dagli elementi che appartengono sia ad A che a B:

$$\{2,3\} \cap \{3,4\} = \{3\}, \quad \{\text{numeri pari}\} \cap \{\text{numeri compresi tra 1 e 5}\} = \{2,4\}.$$

È chiaro per definizione che $A \cap B$ è sottoinsieme sia di A che di B:

$$A \cap B = \{x \in A \text{ tali che } x \in B\} = \{x \in B \text{ tali che } x \in A\}.$$

L'operazione di intersezione è commutativa ed associativa, dove con ciò intendiamo che valgono le uguaglianze

$$A \cap B = B \cap A, \quad A \cap (B \cap C) = (A \cap B) \cap C,$$

per ogni terna di insiemi A, B, C. Il significato delle parentesi è quello abituale: quando scriviamo $A \cap (B \cap C)$ si intende che si effettua prima l'intersezione tra B e C e l'insieme risultante $B \cap C$ viene successivamente intersecato con A.

Se A e B sono due insiemi indichiamo con $A \cup B$ la loro **unione**. Per definizione, l'unione $A \cup B$ è l'insieme formato dagli elementi che appartengono ad A oppure a B, intendendo con questo che possono appartenere ad entrambi: ad esempio

$$\{0,1\} \cup \{3,4\} = \{0,1,3,4\}, \qquad \{1,2,3\} \cup \{3,4\} = \{1,2,3,4\}.$$

Pure l'operazione di unione è commutativa ed associativa, intendendo con questo che per ogni terna di insiemi A, B, C si ha

$$A \cup B = B \cup A, \quad A \cup (B \cup C) = (A \cup B) \cup C.$$

È inoltre del tutto evidente che per ogni insieme A vale $A \cup A = A \cap A = A$.

Se A e B non hanno elementi in comune, la loro intersezione è l'**insieme vuoto**, indicato con il simbolo $\emptyset$. Possiamo quindi scrivere

$$\{1,2\} \cap \{3,4\} = \emptyset, \quad \{\text{numeri pari}\} \cap \{\text{numeri dispari}\} = \emptyset.$$

Osservazione 1.3.1. Per ogni insieme B vale $\emptyset \subseteq B$. Infatti, la condizione da soddisfare affinché $A \subseteq B$ è che ogni elemento di A appartenga a B. Se A non ha elementi, allora non ci sono condizioni da verificare e quindi $\emptyset \subseteq B$ è sempre vera. Viceversa, la relazione $B \subseteq \emptyset$ è vera se e solo se B è l'insieme vuoto.

Basta un po' di buonsenso e di logica comune per convincersi della validità dei seguenti quattro sillogismi, in ciascuno dei quali le tre lettere S, M, P denotano altrettanti insiemi:

(1) se $M \subseteq P$ e $S \subseteq M$, allora $S \subseteq P$;
(2) se $M \cap P = \emptyset$ e $S \subseteq M$, allora $S \cap P = \emptyset$;
(3) se $M \subseteq P$ e $S \cap M \neq \emptyset$, allora $S \cap P \neq \emptyset$;
(4) se $M \cap P = \emptyset$ e $S \cap M \neq \emptyset$, allora $S \not\subseteq P$.

Osservazione 1.3.2. Nella trattazione medievale della logica Aristotelica, i quattro tipi di relazioni tra insiemi che compaiono nei suddetti punti venivano indicati con le 4 vocali a,e,i,o, secondo la seguente tabella:

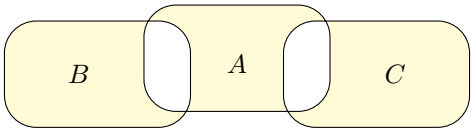


FIGURA 1.1. Visualizzazione grafica di $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

AaB	$A \subseteq B$	ogni A è B (universale affermativa)
AeB	$A \cap B = \emptyset$	ogni A non è B (universale negativa)
AiB	$A \cap B \neq \emptyset$	qualche A è B (particolare affermativa)
AoB	$A \not\subseteq B$	qualche A non è B (particolare negativa)

Sono detti sillogismi della prima figura quelli del tipo

$$\text{“ } \textit{se } M \diamond P \textit{ e } S \dagger M\textit{, allora } S \star P \text{ ”},$$

dove i tre simboli $\diamond, \dagger, \star$ assumono valori nelle quattro vocali a,e,i,o. Tenendo presente che il concetto del vuoto come insieme è una conquista recente dell'intelletto, tra le $4^3 = 64$ combinazioni possibili, solo le 6 elencate nella seguente proposizione risultavano corrette dal punto di vista logico e le vocali del corrispondente nome latino rappresentano la successione $\diamond, \dagger, \star$.

PROPOSIZIONE 1.3.3 (Prima figura dei sillogismi). *Siano S, M, P insiemi non vuoti:*

(Barbara) *se $M \subseteq P$ e $S \subseteq M$, allora $S \subseteq P$;*
(Barbari) *se $M \subseteq P$ e $S \subseteq M$, allora $S \cap P \neq \emptyset$;*
(Celarent) *se $M \cap P = \emptyset$ e $S \subseteq M$, allora $S \cap P = \emptyset$;*
(Celaront) *se $M \cap P = \emptyset$ e $S \subseteq M$, allora $S \not\subseteq P$;*
(Darii) *se $M \subseteq P$ e $S \cap M \neq \emptyset$, allora $S \cap P \neq \emptyset$;*
(Ferio) *se $M \cap P = \emptyset$ e $S \cap M \neq \emptyset$, allora $S \not\subseteq P$.*

In totale si hanno quattro figure e le rimanenti tre si ottengono dalla prima scambiando nelle premesse la posizione di M rispetto a S e P.

Se A è un insieme e $B \subseteq A$ è un sottoinsieme, il **complementare** di B in A è definito come il sottoinsieme degli elementi di A che non appartengono a B e si indica $A - B$, e cioè:

$$A - B = \{a \in A \text{ tali che } a \notin B\}.$$

Equivalentemente, $A - B$ è l'unico sottoinsieme di A tale che

$$B \cup (A - B) = A, \qquad B \cap (A - B) = \emptyset.$$

Dati comunque tre insiemi A, B, C, sono verificate le uguaglianze

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \tag{1.1}$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C). \tag{1.2}$$

Dimostrare la (1.1) equivale a dimostrare entrambe le inclusioni

$$A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C), \qquad A \cap (B \cup C) \supseteq (A \cap B) \cup (A \cap C).$$

Dire che $x \in A \cap (B \cup C)$ significa dire che $x \in A$ e $x \in B \cup C$; equivalentemente, dire che $x \in A \cap (B \cup C)$ significa dire che $x \in A$ e $x \in B$ oppure che $x \in A$ e $x \in C$. Nel primo caso $x \in A \cap B$, mentre nel secondo $x \in A \cap C$ ed in entrambi i casi $x \in (A \cap B) \cup (A \cap C)$. Abbiamo quindi dimostrato $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$. Lasciamo per esercizio al lettore la dimostrazione di $A \cap (B \cup C) \supseteq (A \cap B) \cup (A \cap C)$ e di (1.2).

ESEMPIO 1.3.4. L'insieme $\{1, 2\}$ contiene esattamente quattro sottoinsiemi: $\emptyset$, $\{1\}$, $\{2\}$ e $\{1, 2\}$. L'insieme $\{1, 2, 3\}$ contiene esattamente otto sottoinsiemi, e cioè i quattro precedenti più $\{3\}$, $\{1, 3\}$, $\{2, 3\}$, $\{1, 2, 3\}$.

Dato un qualunque insieme X, indichiamo con $\mathcal{P}(X)$ quello che viene chiamato **l'insieme delle parti di** X. Per definizione gli elementi di $\mathcal{P}(X)$ sono tutti e soli i sottoinsiemi di X. Ad esempio

$$\mathcal{P}(\{a\}) = \{\emptyset, \{a\}\}, \qquad \mathcal{P}(\{a,b\}) = \{\emptyset, \{a\}, \{b\}, \{a,b\}\}.$$

Si noti il curioso fatto che l'insieme $\mathcal{P}(\emptyset) = \{\emptyset\}$ non è vuoto in quanto contiene, come unico elemento, il sottoinsieme vuoto; talvolta l'insieme $\mathcal{P}(\emptyset)$ viene indicato con un asterisco $*$ e chiamato singoletto o singoletta. In alcuni contesti, gli insiemi $\emptyset$ e $\mathcal{P}(\emptyset)$ vengono chiamati rispettivamente insieme iniziale ed insieme finale.

Le nozioni di unione ed intersezione si estendono nel modo più ovvio possibile a terne, quadruple e, più in generale a successioni finite di insiemi $A_1, \dots, A_n$: si pone infatti

$$A_1 \cap \cdots \cap A_n = \text{insieme degli elementi che appartengono a tutti gli } A_i,$$

$$A_1 \cup \cdots \cup A_n = \text{insieme degli elementi che appartengono ad almeno un } A_i.$$

Si ha quindi, a titolo di esempio,

$$\{1,2,3\} \cap \{1,3,4\} \cap \{3,4\} = \{3\}, \quad \{1,2\} \cup \{1,3\} \cup \{3,4\} = \{1,2,3,4\}.$$

In maniera ancora più generale, per ogni insieme X ed ogni sottoinsieme $\Gamma \subseteq \mathcal{P}(X)$ ha senso considerare

$$\bigcap_{A \in \Gamma} A = \text{insieme degli elementi che appartengono a tutti gli } A \in \Gamma,$$

$$\bigcup_{A \in \Gamma} A = \text{insieme degli elementi che appartengono ad almeno un } A \in \Gamma.$$

Esercizi.

19. Descrivere tutti i sottoinsiemi di $\{a,b,c,d\}$ formati da un numero dispari di elementi.

20. Convincetevi che per ogni coppia di insiemi A, B:
(1) valgono le inclusioni $A \cap B \subseteq A$ e $A \subseteq A \cup B$;
(2) valgono le uguaglianze $A \cup (A \cap B) = A$ e $A \cap (A \cup B) = A$;
(3) vale $A \subseteq B$ se e solo se $A \cap B = A$;
(4) vale $A \subseteq B$ se e solo se $A \cup B = B$.

21. Quanti sono i sottoinsiemi di $\{1,2,3,4,5\}$ che contengono esattamente due numeri dispari?

22 (♡, Vero o Falso?). Se qualche criminale è milionario e tutti i magnati sono milionari, allora alcuni magnati sono criminali. Usare il linguaggio degli insiemi, magari accompagnato con un disegnino, per spiegare se la conclusione è vera o falsa.

23 (♡). Per ogni insieme finito X, indichiamo con $|X|$ il numero di elementi di X. Ad esempio, $|\emptyset| = 0$, $|\{\emptyset\}| = 1$, $|\{\emptyset, \{\emptyset\}\}| = 2$, $|\{a,b,c\}| = 3$, $|\mathcal{P}(\{a,b\})| = 4$ eccetera. Nelle seguenti espressioni mettere il giusto segno ($+$ oppure $-$) al posto di $\pm$ in modo che le uguaglianze siano verificate per ogni terna di insiemi finiti A, B, C:

$$|A \cap B| = |A| \pm |A - B|,$$

$$|A \cup B| = |A| \pm |B| \pm |A \cap B|,$$

$$|A \cup B \cup C| = |A| \pm |B| \pm |C| \pm |A \cap B| \pm |A \cap C| \pm |B \cap C| \pm |A \cap B \cap C|.$$

1.4. Brevi cenni sul metodo di Gauss

Non c'è bisogno di spiegazioni per dire che se in un sistema lineare scambiamo l'ordine delle equazioni, allora le soluzioni non cambiano; la stessa conclusione vale se ad un sistema togliamo od aggiungiamo una equazione che è combinazione lineare delle altre.

Da ciò possiamo ricavare un metodo di soluzione dei sistemi lineari, noto ai matematici cinesi da oltre 2000 anni con il nome di *Fangcheng*, riscoperto indipendentemente da Gauss agli inizi del XIX secolo e basato essenzialmente sul seguente teorema.

TEOREMA 1.4.1. *Se un sistema di equazioni lineari viene trasformato in un altro effettuando una delle seguenti operazioni:*

(1) *scambiare l'ordine delle equazioni;*
(2) *moltiplicare un'equazione per un numero diverso da* 0*;*
(3) *aggiungere ad una equazione un multiplo di un'altra equazione del sistema.*

Allora i due sistemi hanno le stesse soluzioni.

La validità del precedente teorema è abbastanza intuitiva e non richiede particolari commenti, se non osservare che le tre operazioni descritte sono usate comunemente nelle trattazioni dei sistemi lineari fatte alle scuole superiori.

Al fine di introdurre un tipo di ragionamento utile in matematica, diamo una dimostrazione formale del fatto che l'operazione (3) non cambia l'insieme delle soluzioni. Per semplicità espositiva consideriamo il caso dei sistemi a due equazioni e due incognite, ma le stesse argomentazioni valgono in qualsiasi generalità.

Consideriamo quindi un sistema

$$(A) \quad \begin{cases} ax + by = \alpha \\ cx + dy = \beta \end{cases}$$

dove a, b, c, d, α e β sono numeri e x, y le incognite. Sia k un altro numero e consideriamo il sistema

$$(B) \quad \begin{cases} ax + by = \alpha \\ (c + ka)x + (d + kb)y = \beta + k\alpha \end{cases}$$

ottenuto aggiungendo alla seconda equazione la prima moltiplicata per k. Indichiamo con S_A l'insieme delle soluzioni del sistema A, ossia l'insieme delle coppie di numeri (x, y) tali che $ax + by = \alpha$ e $cx + dy = \beta$. Similmente indichiamo con S_B l'insieme delle soluzioni del sistema B. Vogliamo dimostrare che $S_A = S_B$, e per fare ciò dimostriamo che valgono entrambe le relazioni

$$S_A \subseteq S_B, \qquad S_B \subseteq S_A.$$

Sia dunque (x, y) una soluzione di A, vogliamo dimostrare che risolve anche (B): basta chiaramente mostrare che $(c + ka)x + (d + kb)y = \beta + k\alpha$; siccome $cx + dy = \beta$ e $ax + by = \alpha$ una semplice sostituzione ci dà

$$(c + ka)x + (d + kb)y = cx + dy + k(ax + by) = \beta + k\alpha.$$

Abbiamo quindi dimostrato che $S_A \subseteq S_B$. Per dimostrare che $S_B \subseteq S_A$ si procede alla stessa maniera: se (x, y) è una soluzione di S_B allora $ax + by = \alpha$, $(c + ka)x + (d + kb)y = \beta + k\alpha$ e quindi

$$cx + dy = (c + ka)x + (d + kb)y - k(ax + by) = \beta + k\alpha - k\alpha = \beta.$$

Osservazione 1.4.2. È possibile provare che se due sistemi nelle stesse incognite con lo stesso numero di equazioni sono risolubili e hanno le stesse soluzioni, allora si può passare dall'uno all'altro con una successione finita di operazioni descritte nel Teorema 1.4.1 (Esercizio 401).

Con il termini **eliminazione di Gauss** e **Fangcheng** intenderemo l'applicazione di una successione finita delle operazioni descritte nel Teorema 1.4.1 per trasformare un sistema lineare in un altro, che ha le stesse soluzioni, ma con alcune variabili eliminate da alcune equazioni: saremo più precisi e metodici in proposito nel Capitolo 7. In altre parole, si usa il metodo di Gauss per far comparire quanti più zeri possibile tra i coefficienti delle variabili del sistema.

Esempio 1.4.3. Utilizziamo l'eliminazione di Gauss per risolvere il sistema

$$\begin{cases} x + y + z = 3 \\ x + 2y + 3z = 6 \\ x + 4y + 9z = 14 \end{cases}$$

Vogliamo eliminare la variabile x dalle equazioni successive alla prima: per annullare il coefficiente di x dalla terza equazione possiamo sottrarre la seconda equazione alla terza:

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

Adesso annulliamo il coefficiente di x dalla seconda equazione sottraendo la prima:

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

Adesso dobbiamo eliminare la y dalle equazioni successive alla seconda: a tal fine annulliamo il coefficiente di y dalla terza equazione sottraendo 2 volte la seconda:

$$\begin{cases} x+y+z=3 \\ \quad y+2z=3 \\ \qquad 2z=2 \end{cases}$$

Adesso il sistema si è di molto semplificato e si può proseguire con il metodo di sostituzione:

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

Osservazione 1.4.4. È molto intuitivo, e non difficile da dimostrare, che se applicando l'eliminazione di Gauss troviamo l'equazione $0=0$ allora il sistema è ridondante, mentre se troviamo l'equazione $0=\alpha$, con α un qualsiasi numero diverso da 0, allora il sistema è inconsistente, ossia non ha soluzioni. Rimandiamo la dimostrazione rigorosa di questo fatto al Capitolo 7, quando la teoria ci permetterà di farlo in maniera elegante e indolore.

Osservazione 1.4.5. Un risultato importante nella teoria dei sistemi lineari è il teorema di Rouché–Capelli[3], secondo il quale un sistema lineare

$$\begin{cases} a_{11}x_1+a_{12}x_2+\cdots+a_{1m}x_m=b_1 \\ \quad\vdots \qquad\quad \vdots \qquad\qquad \vdots \\ a_{n1}x_1+a_{n2}x_2+\cdots+a_{nm}x_m=b_n \end{cases}$$

possiede soluzioni se e soltanto se il suo rango è uguale al rango del sistema ottenuto ponendo i "termini noti" $b_1,\dots,b_n$ tutti uguali a 0. Dimostreremo il teorema di Rouché–Capelli in più modi nei prossimi capitoli.

Esercizi.

24. Utilizzando l'eliminazione di Gauss, risolvere i seguenti sistemi lineari

$$\begin{cases} x+y+z=1 \\ x-y+2z=1 \\ x+y+4z=1 \end{cases} \qquad \begin{cases} x+y+2z=4 \\ x+2y+4z=7 \\ x+4y+10z=15 \end{cases} \qquad \begin{cases} x+y-2z=3 \\ x+2y+2z=6 \end{cases}$$

25. Usare il metodo di Gauss e l'Osservazione 1.4.4 per stabilire quali tra i seguenti sistemi sono ridondanti e quali sono inconsistenti:

$$\begin{cases} x+y+z=1 \\ x-y+2z=1 \\ 3x+y+4z=3 \end{cases} \qquad \begin{cases} x+y+z=1 \\ x-y+2z=1 \\ 3x+y+4z=1 \end{cases} \qquad \begin{cases} x+2y+2z=0 \\ x+3y+4z=0 \\ x+5y+10z=0 \end{cases}$$

26 (♡). Risolvere il sistema lineare

$$\begin{cases} x+2y+3z+4w=11 \\ 2x+y+2z+3w=9 \\ 2x+2y+z+2w=6 \\ 2x+2y+2z+w=7 \end{cases}$$

27. Nella tradizione occidentale, uno dei primi sistemi lineari di cui si ha notizia storica è l'*epantema di Timarida*, oggi più simile ad un gioco enigmistico che ad un problema matematico: *Paperino, Qui, Quo e Qua pesano assieme 150 kg. Paperino e Qui pesano assieme*

[3]Leggasi ruscé–capelli.

91 kg, Paperino e Quo pesano assieme 90 kg, Paperino e Qua pesano assieme 89 kg. Quanto pesa ciascun papero?

In linguaggio algebrico, il problema si riconduce ad un sistema lineare del tipo

$$\begin{cases} x + x_1 + \cdots + x_n = S \\ x + x_1 = a_1 \\ \cdots \\ x + x_n = a_n \end{cases}, \qquad n > 1.$$

che ha come soluzione (provate a spiegare perché?)

$$x = \frac{(a_1 + \cdots + a_n) - S}{n-1}, \quad x_1 = a_1 - x, \quad \cdots, \quad x_n = a_n - x.$$

1.5. Alcune cose che si trovano nei libri di matematica

Per motivi di natura storica noti a tutti, la Grecia antica ha avuto un ruolo fondamentale nella storia della matematica. Come conseguenza di questo fatto molti termini matematici hanno una chiara origine greca, come ad esempio *ipotesi* (premessa) e *tesi* (conclusione). Come se non bastasse, le formule matematiche sono piene zeppe di lettere dell'alfabeto greco:

α (alpha)	β (beta)	γ (gamma)	δ (delta)
ϵ, ε (epsilon)	ϕ, φ (phi)	η (eta)	θ, ϑ (theta)
ζ (zeta)	μ (mu)	ν (nu)	λ (lambda)
ξ (xi)	ρ, ϱ (rho)	π (pi)	σ, ς (sigma)
τ (tau)	χ (chi)	ψ (psi)	ω (omega)
ι (iota)	κ (kappa)		

Inoltre, sono di uso frequente in matematica:

- le lettere greche maiuscole Γ (Gamma), Δ (Delta), Θ (Theta), Λ (Lambda), Σ (Sigma), Π (Pi), Φ (Phi), Ψ (Psi), Ξ (Xi), Ω (Omega).
- i simboli $+, \times, \vee, \oplus, \otimes, \odot, \boxplus, \boxtimes$ derivati dagli alfabeti delle lingue arcaiche del Mediterraneo (Fenicio, Euboico, Etrusco ecc.): con il passare dei secoli tali simboli si sono evoluti nelle usuali lettere greche e latine ($+$ diventa τ, $\oplus$ diventa Θ ecc.) ed hanno perso il loro valore fonetico.
- i simboli ∇ (nabla) e $\amalg$ che sono Delta e Pi rovesciati, ∂ (de) derivato dall'alfabeto Cirillico, $\aleph$ (aleph) dell'alfabeto ebraico, ∞ (infinito)[4], $\overline{\partial}$ (debar), $\hbar$ (accatagliato), $\wedge$ (wedge), $\in$ (appartiene), $\forall$ (per ogni), $\exists$ (esiste) e $\iff$ (se e solo se).

Ogni scritto matematico, come il presente, richiede qualcosa in più della semplice punteggiatura per essere comprensibile a chi legge. In particolare è necessario dichiarare il senso e la funzione di alcune parti, usando i nomi di *enunciato*, *definizione*, *teorema*, *lemma*, *corollario*, *proposizione*, *congettura*, *speculazione*, *dimostrazione*, *confutazione*, *notazione*, *osservazione*, *esempio*, *esercizio* eccetera. Per ragioni puramente linguistiche, accade che nella lingua italiana, i vocaboli enunciato, teorema, lemma e corollario sono di genere maschile, mentre definizione, proposizione, congettura, speculazione, dimostrazione, notazione ed osservazione sono di genere femminile.

Senza alcuna pretesa di completezza e rigore, diamo una breve spiegazione, grossolana ed informale, del significato di alcuni dei suddetti termini. Per enunciato, o affermazione, o asserzione, o proposizione (logica), si intende un insieme di frasi che esprimono un messaggio di senso compiuto, che è o vero o falso, ma non vero e falso contemporaneamente: ciascuna frase dell'enunciato è formata da parole e simboli matematici, è organizzata intorno ad un verbo ed è delimitata da segni di punteggiatura.[5]

---

[4] Il simbolo ∞ deriva da DD, o per meglio dire dal doppio antenato della lettera D, e si usava in alcune versioni primitive della numerazione romana per indicare il numero $1000 = 500 + 500$; pure gli Etruschi utilizzavano il doppio antenato della D per indicare 1000, usando però l'allineamento verticale ed ottenendo quello che oggi ci appare come un 8.

[5] Il gruppo verbale di una frase può essere benissimo rappresentato da simboli matematici: ad esempio, nella formula $x = 1$ il gruppo verbale è il simbolo $=$.

La dimostrazione è la prova che un certo enunciato è vero, mentre la confutazione è la prova che un certo enunciato è falso.

Dato un enunciato A, il suo **opposto** è un enunciato che è vero se e soltanto se A è falso. Ad esempio l'opposto di $1+1=2$ è $1+1\neq 2$, mentre l'opposto di *ogni uomo è mortale* è rappresentato da *esiste un uomo immortale*.

I termini teorema, lemma, corollario e proposizione (asserita) hanno tutti il significato di enunciato dimostrato (e quindi vero). La scelta di quale vocabolo usare tra i quattro precedenti è materia alquanto opinabile e dipende in larga misura dal gusto di chi scrive. In linea di massima per teorema si intende un enunciato di primaria importanza, per proposizione un enunciato di secondaria importanza, mentre per corollario si intende una conseguenza, più o meno ovvia, di un precedente teorema.

Il lemma è un risultato che serve alla dimostrazione di un teorema o una proposizione; c'è da dire che l'importanza di un enunciato dimostrato può cambiare nel tempo e che la matematica è piena di lemmi che hanno assunto successivamente un'importanza maggiore dei teoremi per i quali erano inizialmente usati nella dimostrazione. I termini congettura, speculazione e problema aperto indicano un enunciato non dimostrato né confutato: mentre una congettura è un enunciato che si ritiene vero e sul quale si hanno evidenze più o meno forti a supporto, il termine speculazione viene usato per portare all'attenzione un fatto sul quale si hanno evidenze molto deboli sulla sua possibile validità.

La definizione è il nome che viene dato ad un insieme di cose collegate tra loro da determinate relazioni. La definizione di un ente non implica l'esistenza dell'ente medesimo: ad esempio ciascun matematico è libero di definire un numero sarcastico come un numero intero che è contemporaneamente pari e dispari, sebbene tali numeri non esistono. Similmente ogni teologo è libero di definire l'inferno, ogni filosofo il mondo perfetto, ogni uomo/donna il partner ideale eccetera.

Nei testi matematici non è difficile trovare tracce di latino, come ad esempio nelle abbreviazioni *i.e.* (id est), *e.g.* (exempli gratia), *N.B.* (nota bene), *Q.E.D.* (quod erat demonstrandum), *viz.* (videlicet), *et al.* (et alia), *cf.* (confer) e nelle locuzioni *mutatis mutandis* (cambiando quello che bisogna cambiare), *cum grano salis* (da interpretare con discernimento), *a fortiori* (a maggior ragione), *ergo* (di conseguenza). I termini i.e. e viz. sono entrambi sinonimi dei vocaboli *ossia* e *cioè*; la sottile differenza consiste nel fatto che mentre i.e. indica una semplice riscrittura di un concetto con altre parole, il termine viz. implica un maggior contenuto esplicativo. Il termine e.g. introduce una lista di esempi, mentre Q.E.D. indica la conclusione di una dimostrazione. Per evitare un errore comune ribadiamo che cf. ha il significato di "confronta (quello che abbiamo appena scritto) con" e non è sinonimo di "vedi". Tali termini sono usati anche nei testi di lingua inglese; da notare il curioso fatto che nell'American English i termini i.e. ed e.g. sono sempre seguiti dalla virgola, a differenza del British English dove la virgola non è richiesta. Mentre viz. e Q.E.D. sono considerati un po' obsoleti e stanno lentamente scomparendo, sostituiti rispettivamente da i.e. e dal quadratino $\square$, i termini i.e. ed e.g. risultano tuttora ampiamente usati. Va detto che, nella maggioranza dei casi, i termini latini non sono apposti per esibire cialtronescamente la cultura classica dell'autore, ma per rendere meno legnosa la scrittura in un ambito, quello della matematica, dove per forza di cose determinati termini si ripetono in continuazione.

A proposito di latino, un principio fondamentale della matematica[6] è il "Tertium non datur", secondo il quale ogni affermazione matematica o è vera oppure è falsa (ma non vera e falsa contemporaneamente). Questo implica che ogni enunciato matematico richiede una dose di chiarezza molto superiore a quella usata nelle discussioni politiche. Ad esempio una frase del tipo "i marziani rubano" contiene un livello di ambiguità tale da poter essere smentita ed interpretata in una miriade di modi: in un ipotetico paese dove ogni discussione è valutata alla luce del pensiero logico e del ragionamento, tale frase andrebbe sostituita con una priva di ambiguità, del tipo "tutti i marziani rubano" oppure "esistono dei marziani che rubano" oppure "in ogni regione esiste almeno un marziano che ruba" eccetera.

Per aiutare la comprensione, non è raro accompagnare le risposte ad un quesito matematico con alcune sfumature lessicali: ad esempio un'affermazione falsa riguardante una molteplicità di situazioni può essere completamente falsa oppure generalmente falsa. Ad esempio

[6] Almeno della sua corrente stradominante basata sulla logica binaria: il suo dominio non è ideologico ma è dovuto al fatto che riesce a descrivere il mondo in cui viviamo con un'efficacia incredibile.

dire che tutti gli uomini viventi sono alti più di 7 metri è completamente falso, mentre dire che tutti gli uomini viventi sono alti più di 2 metri è generalmente falso (esistono alcuni uomini più alti di 2 metri). Notare che per essere falsa un'affermazione è sufficiente che sia falsa in un singolo caso: siccome un mio rivale è alto un metro ed un tappo, anche l'affermazione che tutti gli uomini sono più alti di un metro e due tappi è generalmente falsa, sebbene sia vera per il 99,9% del genere umano.

Esercizi.

28. Prendere carta e penna e allenarsi a scrivere le lettere minuscole dell'alfabeto greco, con particolare cura alle lettere ξ, ζ e θ (notoriamente le più ostiche).

29. Per ciascuna delle seguenti affermazioni, dire se sono vere, generalmente false o completamente false:

(1) ogni uomo è mortale;
(2) ogni studente che si iscrive all'università riesce a laurearsi;
(3) ogni tortellino ripieno possiede una coscienza.

30 (♡). Il comune canovaccio di molte barzellette matematiche è la commistione di argomenti tipicamente matematici con altri relativi alla vita comune. Un esempio è dato dal seguente arcinoto problema: tra madre e figlio ci sono 21 anni di differenza, tra 6 anni la madre avrà 5 volte l'età del figlio. Dove si trova il padre?

31 (♡). Sul tavolo sono disposte quattro carte come in Figura 1.2; ciascuna carta ha disegnato un numero su di una faccia e una lettera sulla faccia opposta.

Quali sono le due carte da rivoltare se vogliamo dimostrare o confutare l'affermazione che se su di una faccia c'è la lettera A, allora sulla faccia opposta c'è il numero 2?

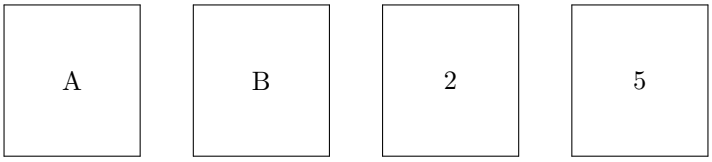


FIGURA 1.2. Il test delle quattro carte

32. Durante il suo spettacolo teatrale, il mago Flip estrae dalla tasca il portafoglio e, rivolgendosi ad uno spettatore dice: "sento con i miei poteri che in questo portafoglio c'è lo stesso numero di banconote che nel tuo, anzi mi correggo, qui ce ne sono 3 di più. No, fermi! Il mio spirito guida dice che qui ci sono le stesse che nel tuo, più altre 3, più quelle che mancano alle tue per arrivare a 20 banconote". Dopo aver verificato che lo spirito guida del mago aveva ragione, dal teatro parte una standing ovation. Quante banconote aveva il mago nel portafoglio?

1.6. Prima esercitazione

Una delle caratteristiche che deve avere ogni laureato in matematica rispettabile è quella di saper produrre dimostrazioni rigorose di risultati matematici non identici a quelli già conosciuti e di difficoltà non eccessiva.

Nelle precedenti sezioni sono stati proposti degli esercizi di calcolo e risoluzione di equazioni. Da adesso in poi saranno proposti pure esercizi in cui si chiede di *produrre dimostrazioni*; per facilitare il lettore in questo (inizialmente arduo) compito, in questa sezione saranno illustrati, ed in parte svolti, alcuni esercizi dimostrativi, corredati da commenti informali sul tipo di procedura adottata.

A. Una **dimostrazione** non è altro che un'argomentazione in grado di convincere un lettore intelligente e sufficientemente informato della veridicità di una asserzione. Ogni teorema matematico richiede almeno una dimostrazione, che deve essere chiara e convincente.

ESERCIZIO 1.6.1. Dimostrare che per ogni intero dispari n, il numero n^2-1 è divisibile per 8.

Soluzione. Dalla ben nota formula di scomposizione della differenza di due quadrati si ottiene $n^2-1=(n+1)(n-1)$. Per ipotesi n è dispari, quindi sia $n+1$ che $n-1$ sono pari, ossia esistono due numeri interi a,b tali che $n+1=2a$, $n-1=2b$. Inoltre

$$a-b=\frac{n+1}{2}-\frac{n-1}{2}=\frac{n+1-n+1}{2}=1,$$

a,b sono due interi consecutivi e quindi esattamente uno dei due è pari. Se a è pari si ha $a=2c$ e quindi $n^2-1=(2a)(2b)=4ab=8cb$ risulta divisibile per 8. Se invece b è pari si ha $b=2d$ e quindi $n^2-1=(2a)(2b)=4ab=8ad$ risulta ugualmente divisibile per 8.

ESERCIZIO 1.6.2. Dimostrare che per ogni terna di insiemi A,B,C si ha:

$$A\cap(B\cup C)=(A\cap B)\cup(A\cap C).$$

Soluzione. Sia $x\in A\cap(B\cup C)$ un qualsiasi elemento, allora in particolare $x\in B\cup C$ e quindi $x\in B$ oppure $x\in C$. Nel primo caso $x\in A$, $x\in B$ e quindi $x\in A\cap B$; nel secondo caso $x\in A$, $x\in C$ e quindi $x\in A\cap C$; in entrambi i casi si ha dunque $x\in(A\cap B)\cup(A\cap C)$. Abbiamo quindi dimostrato che se $x\in A\cap(B\cup C)$, allora $x\in(A\cap B)\cup(A\cap C)$, ossia che $A\cap(B\cup C)\subseteq(A\cap B)\cup(A\cap C)$.

Supponiamo viceversa che $y\in(A\cap B)\cup(A\cap C)$, allora $x\in A\cap B$ oppure $x\in A\cap C$. Nel primo caso $x\in A$, $x\in B$ ed a maggior ragione $x\in B\cup C$; nel secondo caso $x\in A$, $x\in C$ ed a maggior ragione $x\in B\cup C$. In entrambi i casi $x\in A$, $x\in B\cup C$ e quindi $x\in A\cap(B\cup C)$. Abbiamo quindi dimostrato che se $x\in(A\cap B)\cup(A\cap C)$, allora $x\in A\cap(B\cup C)$, ossia che $(A\cap B)\cup(A\cap C)\subseteq A\cap(B\cup C)$.

Mettendo assieme le due parti della precedente argomentazione abbiamo dimostrato che $x\in A\cap(B\cup C)$ se e solamente se $x\in(A\cap B)\cup(A\cap C)$, ossia che $A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$.

ESERCIZIO 1.6.3. Siano A,B,C,D sottoinsiemi di un insieme X tali che $X=A\cup B=C\cup D$. Dimostrare che $X=(A\cap C)\cup(B\cup D)$.

Soluzione. Bisogna dimostrare che per un qualsiasi elemento $x\in X$ si ha $x\in B\cup D$ oppure $x\in A\cap C$. Se x non appartiene all'unione $B\cup D$ a maggior ragione non appartiene a B e quindi, siccome $A\cup B=X$ deve necessariamente essere $x\in A$. Similmente x non appartiene a D e quindi $x\in C$. In conclusione abbiamo provato che se $x\notin B\cup D$ allora $x\in A\cap C$.

B. A volte è utile fare dimostrazioni **per assurdo**: se si vuole dimostrare che da un'ipotesi segue una determinata tesi, si suppone che l'ipotesi sia vera, che la tesi sia falsa e da ciò si riesce a dedurre una contraddizione.

ESERCIZIO 1.6.4. Dimostrare che il sistema lineare

$$\begin{cases} x+y+z=1\\ x-y+2z=1\\ 2x+3z=1 \end{cases} \tag{1.3}$$

non possiede soluzioni.

Soluzione. Supponiamo per assurdo che le tre equazioni (1.3) siano vere (ipotesi vera) e che il sistema possieda soluzioni (tesi falsa). Sia (x_0,y_0,z_0) una di tali soluzioni. Allora si hanno le uguaglianze $x_0+y_0+z_0=1$, $x_0-y_0+2z_0=1$ e sommando membro a membro si ottiene $2x_0+3z_0=2$ in contraddizione con l'uguaglianza $2x_0+3z_0=1$ ottenuta direttamente dalla terza equazione del sistema. Riassumendo, abbiamo dimostrato che supponendo che il sistema abbia soluzioni produce una contraddizione, e ciò equivale a dire che il sistema non ha soluzioni.

ESERCIZIO 1.6.5. Ricordiamo che un primo positivo è un numero intero $p\geq 2$ che, all'interno dei numeri interi positivi, è divisibile solamente per 1 e per p. Dal fatto che ogni numero maggiore di 1 è divisibile per almeno un numero primo, dedurre che esistono infiniti numeri primi positivi.

Soluzione. Riportiamo la classica dimostrazione di Euclide, probabilmente già nota a molti lettori, che a distanza di millenni rimane uno dei migliori modelli di ragionamento matematico. Supponiamo per assurdo che esista solamente un numero finito n di primi positivi, che indicheremo $p_1, \dots, p_n$, e consideriamo il numero

$$q = p_1 p_2 \cdots p_n + 1,$$

ossia il prodotto di tutti i numeri primi più 1. Dato che $q > 1$ si ha che q è divisibile per almeno un primo positivo. D'altra parte, il numero q non è divisibile per p_1 perché la divisione ha resto 1; per lo stesso motivo il numero q non è divisibile per nessuno dei numeri $p_2, \dots, p_n$, e questo rappresenta una contraddizione.

C. Una **confutazione**, o refutazione, è un'argomentazione in grado di convincere un lettore intelligente della falsità di una asserzione. Talvolta per dimostrare la verità di un'asserzione si confuta l'asserzione opposta.

Esercizio 1.6.6. Scrivere gli opposti dei seguenti enunciati:

(1) il numero $2^{17} - 1$ è primo;
(2) l'equazione di secondo grado $x^2 + x + 1$ possiede soluzioni intere;
(3) per ogni numero k l'equazione $x + k = 0$ possiede soluzioni;
(4) esiste un numero intero positivo n che non divide $2^n - 2$;
(5) per ogni insieme A esiste un sottoinsieme $B \subseteq A$ tale che $A - B$ è un insieme finito e non vuoto;
(6) il caffè della Peppina è corretto e zuccherato;
(7) se mio nonno aveva le ruote, allora era un carretto.[7]

Soluzione. Ricordiamo che se P è un enunciato, che può essere vero o falso, il suo opposto è l'enunciato definito dalla proprietà di essere vero se e solo se P è falso. Gli opposti degli enunciati precedenti sono nell'ordine:

(1) il numero $2^{17} - 1$ non è primo;
(2) l'equazione di secondo grado $x^2 + x + 1$ non possiede soluzioni intere;
(3) esiste un numero k tale che l'equazione $x + k = 0$ non possiede soluzioni;
(4) ogni intero positivo n divide $2^n - 2$;
(5) esiste un insieme A tale che per ogni suo sottoinsieme $B \subseteq A$ la differenza $A - B$ o è vuota oppure infinita;
(6) il caffè della Peppina non è corretto oppure non è zuccherato;
(7) mio nonno aveva le ruote ma non era un carretto.

Si noti che ogni enunciato del tipo "se A allora B" è de tutto equivalente a dire che "A è falso oppure B è vero": il suo opposto diventa quindi "A è vero e B è falso".

Esercizio 1.6.7. Dire quali dei seguenti enunciati che coinvolgono il connettivo logico "se ... allora ..." sono veri e quali falsi:

(1) se $3 < 5$, allora $5 < 3$;
(2) se $3 > 5$, allora $3 < 5$;

Soluzione. Il primo enunciato è falso, in quanto $3 < 5$ è vero, mentre $5 < 3$ è falso. Il secondo enunciato è vero poiché $3 > 5$ è falso: per lo stesso motivo, dato un qualunque enunciato P, l'implicazione "se $3 > 5$, allora P" è sempre vera.

D. Per confutare un enunciato che coinvolge una pluralità di casi è sufficiente provare che è falso in almeno uno di tali casi. Un tale caso viene detto un **controesempio**, o esempio in contrario, dell'enunciato.

Esercizio 1.6.8. Dimostrare o confutare che per ogni valore del parametro k il sistema linaere

$$\begin{cases} 3x + ky = 0 \\ kx + 12y = 1 \end{cases}$$

possiede soluzioni.

[7] Trattasi di detto popolare: non venitemi a dire che ho sbagliato i congiuntivi!.

Soluzione. L'enunciato riguarda la totalità dei possibili valori di k ed è falso in quanto per $k = 6$ e $k = -6$ i sistemi corrispondenti

$$\begin{cases} 3x + 6y = 0 \\ 6x + 12y = 1 \end{cases}, \qquad \begin{cases} 3x - 6y = 0 \\ -6x + 12y = 1 \end{cases},$$

non possiedono soluzioni, e quindi $k = 6$ e $k = -6$ sono due possibili controesempi. Il lettore può inoltre facilmente verificare che per ogni valore di k diverso da ± 6 il sistema è risolubile.

Esercizio 1.6.9. Dimostrare o confutare che per ogni intero positivo n il numero $F_n = 2^{2^n} + 1$ è primo.

Soluzione. L'enunciato riguarda la totalità degli interi positivi e la sua ipotetica validità è suggerita dal fatto che $F_1 = 5$, $F_2 = 17$, $F_3 = 257$ e $F_4 = 65537$ sono numeri primi. Fu osservato da Eulero nel 1732 che l'enunciato è falso nella sua totalità, e che il numero $n = 5$ rappresenta un controesempio: infatti si ha

$$F_5 = 2^{2^5} + 1 = 4294967297 = 641 \cdot 6700417.$$

Esercizi.

33. Provare che per ogni successione finita di insiemi $A_1, \dots, A_n$ si ha:

$$A_1 \cap \cdots \cap A_n = A_1 \cap (A_2 \cap \cdots \cap A_n), \quad A_1 \cup \cdots \cup A_n = A_1 \cup (A_2 \cup \cdots \cup A_n).$$

34. Dimostrare l'uguaglianza (1.2).

35. Siano date due successioni finite di sottoinsiemi $A_1, \dots, A_n$, $B_1, \dots, B_n$, di un insieme X tali che $X = A_i \cup B_i$ per ogni indice i. Dimostrare che

$$X = (A_1 \cap \cdots \cap A_n) \cup (B_1 \cup \cdots \cup B_n).$$

36. Siano A, B, C insiemi. Mostrare la validità delle seguenti affermazioni:
(1) Se $A \subseteq C$ e $B \subseteq C$, allora $A \cup B \subseteq C$.
(2) Se $A \subseteq B$ e $A \subseteq C$, allora $A \subseteq B \cap C$.

37. Sia p un numero primo maggiore od uguale a 5. Dimostrare che $p^2 - 1$ è divisibile per 24.

38. Siano B, C due sottoinsiemi di A. Provare che valgono le formule

$$(A - B) \cup (A - C) = A - (B \cap C), \qquad (A - B) \cap (A - C) = A - (B \cup C),$$

ossia che il passaggio al complementare scambia le operazioni di unione ed intersezione.

39 (Contrario e subcontrario). Abbiamo già incontrato la nozione di enunciati opposti. Due enunciati si dicono **contrari** se non possono essere entrambi veri, si dicono **subcontrari** se non possono essere entrambi falsi. Ad esempio gli enunciati "Maria è nata a Roma" e "Maria è nata a Napoli" sono contrari, mentre gli enunciati "Maria ha meno di 30 anni" e "Maria ha più di 20 anni" sono subcontrari. Chiaramente due enunciati sono opposti se e solo se sono al tempo stesso contrari e subcontrari.[8]

Dati due insiemi non vuoti A, B, dire per ciascuna delle seguenti 4 coppie di enunciati se sono opposti, contrari o subcontrari:
(1) $A \cap B = \emptyset$, $A \subseteq B$;
(2) $A \cap B = \emptyset$, $A \cap B \neq \emptyset$;
(3) $A \subseteq B$, $A \not\subseteq B$;
(4) $A \cap B \neq \emptyset$, $A \not\subseteq B$.

40. Sul tavolo di fronte a voi ci sono tre scatolette di cibo, etichettate A, B e C. All'interno di ciascuna di esse si trova un diverso tipo di pesce: tonno, sgombro e sardine. Non sapete come sono distribuiti i cibi nelle scatole e vi viene detto che soltanto una delle seguenti affermazioni è vera:
(1) la scatoletta A contiene il tonno;
(2) la scatoletta B non contiene il tonno;

[8] Mentre l'enunciato opposto è unico, di enunciati contrari ne esistono tantissimi, ragion per cui è buona regola diffidare da chi cerca il consenso contro qualcuno/qualcosa in maniera generica.

(3) la scatoletta C non contiene lo sgombro.

Senza sapere quale delle tre affermazioni sia vera, dovete determinare il contenuto di ciascuna scatoletta.

41. Nel lontano stato del Funtoristan vivono personaggi di due tipi: i funtori semplici, che dicono sempre la verità, ed i funtori derivati, che mentono sempre:

(1) Al nostro arrivo incontriamo due indigeni, uno dei due dice: "Siamo entrambi derivati." Che cosa possiamo dedurre sui due funtori?
(2) Poco dopo incontriamo tre funtori, i cui nomi sono Hom, Tor ed Ext. Hom dice: "tra noi c'è almeno un derivato", Tor dice: "tra noi c'è non più di un derivato", Ext dice: "tra noi c'è esattamente un derivato". Chi di loro è semplice e chi derivato?

1.7. Complementi: tavole della verità e fallacie logiche

Precisiamo subito che il termine verità è qui inteso nel senso della logica binaria, quella per cui un enunciato è sempre vero o falso ma non vero e falso contemporaneamente.

A partire da alcuni enunciati, ne possiamo costruire altri usando le operazioni logiche che illustreremo nelle prossime righe: la caratteristica comune di tali operazioni è che la verità o falsità dell'enunciato costruito dipende in maniera univoca dalla verità o falsità degli enunciati di partenza.

Negazione. Dato un enunciato P indichiamo con $\text{NOT}\, P$ il suo opposto. Per definizione di opposto, si ha che $\text{NOT}\, P$ è vero (V) se e solo se P è falso (F). Possiamo rappresentare questo fatto tramite la seguente tabella:

(1.4)

P	NOT P
V	F
F	V

Congiunzione. la congiunzione di due enunciati P, Q viene indicata con P AND Q; per definizione P AND Q è vero se e solo se P e Q sono entrambi veri, ossia se la sua verità o falsità è determinata dalla seguente *tavola di verità*, ossia dalla tabella

(1.5)

P	Q	P AND Q
V	V	V
V	F	F
F	V	F
F	F	F

Nella lingua italiana la congiunzione di due enunciati si esprime mediante la congiunzione (grammaticale) *e*. Ad esempio, l'enunciato "il numero 7 è dispari *e* maggiore di 6" è la congiunzione dei due enunciati "il numero 7 è dispari" e "il numero 7 è maggiore di 6".

Bicondizionale. Il bicondizionale "P *se e solo se* Q", scritto anche $P \Leftrightarrow Q$, stabilisce l'equivalenza logica tra P e Q; due enunciati sono logicamente equivalenti se sono entrambi veri oppure se sono entrambi falsi. Pertanto, se interpretiamo $P \Leftrightarrow Q$ come enunciato composto, la corrispondente tavola di verità è:

(1.6)

P	Q	$P \Leftrightarrow Q$
V	V	V
V	F	F
F	V	F
F	F	V

Condizionale. Nelle sezioni precedenti abbiamo già incontrato il cosiddetto **condizionale**, ossia una regola del tipo "*se P, allora Q*", dove P e Q sono a loro volta enunciati, chiamati rispettivamente **antecedente** e **conseguente**. La regola "se P, allora Q", che si può anche scrivere $P \Rightarrow Q$, è un modo sintetico di dire che se l'antecedente P è vero, allora anche il conseguente Q è vero, ossia che P *implica* Q.

I termini condizionale e bicondizionale sono tipici del gergo della logica ed al momento non hanno ampia diffusione in altri ambiti. Nei testi rivolti ad un pubblico più vasto si preferisce

di norma usare il termine **implicazione** al posto di condizionale e **doppia implicazione** al posto di bicondizionale.

Ogni implicazione può essere vista come un enunciato la cui verità o falsità è descritta dalla seguente *tavola di verità*, ossia dalla tabella

(1.7)

P	Q	$P \Rightarrow Q$
V	V	V
V	F	F
F	V	V
F	F	V

Come si vede, e come già discusso nell'Esercizio 1.6.7, una implicazione è falsa se e solo se l'antecedente è vero ed il conseguente è falso. Come ulteriore motivazione a questo fatto, osserviamo che in questo modo $P \Leftrightarrow Q$ è logicamente equivalente a $(P \Rightarrow Q)$ AND $(Q \Rightarrow P)$, ma non è logicamente equivalente a $P \Rightarrow Q$; dunque la doppia implicazione è data dalla congiunzione di esattamente due implicazioni.

Disgiunzione e disgiunzione esclusiva. Siano P, Q due enunciati: la disgiunzione P OR Q è vera se e solo se almeno uno tra P e Q è vero; la disgiunzione esclusiva P XOR Q è vera se e solo se esattamente uno tra P e Q è vero. Si hanno quindi le tavole di verità:

(1.8)

P	Q	P OR Q
V	V	V
V	F	V
F	V	V
F	F	F

P	Q	P XOR Q
V	V	F
V	F	V
F	V	V
F	F	F

Nei testi matematici in lingua italiana la disgiunzione si esprime mediante le congiunzioni *o*, *oppure*, mentre la disgiunzione esclusiva richiede frasi più articolate.

Le precedenti operazioni logiche si possono comporre per ottenere operazioni più complesse. Ad esempio, possiamo esprimere le implicazioni usando solamente le operazioni OR e NOT:

$$P \Rightarrow Q \text{ è logicamente equivalente a } (\text{NOT } P) \text{ OR } Q.$$

In determinati contesti è utile introdurre le operazioni NAND (composizione di NOT e AND) e NOR (composizione di NOT e OR); più precisamente:

$$P \text{ NAND } Q = \text{NOT } (P \text{ AND } Q); \qquad P \text{ NOR } Q = \text{NOT } (P \text{ OR } Q).$$

Anche le nozioni di enunciati opposti, contrari o subcontrari (Esercizio 39) possono essere viste come enunciati:

(1.9)

P	Q	P, Q opposti
V	V	F
V	F	V
F	V	V
F	F	F

P	Q	P, Q contrari
V	V	F
V	F	V
F	V	V
F	F	V

P	Q	P, Q subcontrari
V	V	V
V	F	V
F	V	V
F	F	F

Si hanno dunque le equivalenze logiche ($\cong$):

- P, Q opposti $\cong$ P XOR Q;
- P, Q contrari $\cong$ P NAND Q $\cong$ NOT(P AND Q);
- P, Q subcontrari $\cong$ P OR Q.

In ambito informatico ed elettronico, si preferisce usare le cifre binarie 1 e 0 per indicare i valori di verità e falsità rispettivamente. In tal caso, ad esempio, le tavole di verità degli operatori e XOR e AND diventano:

(1.10)

P	Q	P XOR Q
0	0	0
0	1	1
1	0	1
1	1	0

P	Q	P AND Q
0	0	0
0	1	0
1	0	0
1	1	1

Le due componenti della implicazione danno i nomi a due celebri fallacie logiche, ossia a errori di ragionamento.[9]

L'*affermazione del conseguente* è una fallacia che si sviluppa secondo lo schema: A implica B, B è vero, quindi A è vero. Ad esempio:

- Premessa: ogni matematico ha studiato algebra lineare;
- Premessa: Ettore ha studiato algebra lineare;
- Conclusione: Ettore è un matematico.

La *negazione dell'antecedente* è una fallacia che si sviluppa secondo lo schema: A implica B, A è falso, quindi B è falso. Ad esempio

- Premessa: ogni matematico ha studiato algebra lineare;
- Premessa: Ettore non è un matematico;
- Conclusione: Ettore non ha studiato algebra lineare.

Entrambi gli esempi sopra riportati sono falsi poiché anche i fisici hanno studiato algebra lineare nel loro corso di studi. Nei ragionamenti errati più complessi, dove intervengono anche enunciati opposti, risulta a volte difficile distinguere l'affermazione del conseguente dalla negazione dell'antecedente. Infatti, dal punto di vista della logica, l'implicazione *se A è vero, allora B è vero* è del tutto equivalente all'implcazione *se B è falso, allora A è falso*; quindi l'affermazione del conseguente nel primo condizionale è del tutto equivalente alla negazione dell'antecedente sul secondo, e viceversa.

Altre fallacie logiche che spesso compaiono in certe prese di posizione sono l'inversione, il non sequitur e l'analogia debole (più tante altre la cui descrizione non rientra tra gli obiettivi di queste note).

L'*inversione* si sviluppa secondo lo schema: A implica B, quindi B implica A. Ad esempio:

- Premessa: Se ho la polmonite allora ho la febbre.
- Conclusione: Se ho la febbre allora ho la polmonite.

Nei casi concreti l'inversione viene presentata in forme più subdole e meno spudorate, spesso come modo per trasformare enunciati contrari in subcontrari e viceversa. Infatti, dire che A e B sono contrari equivale al condizionale *se A è vero, allora B è falso*, mentre dire che A e B sono subcontrari equivale al condizionale *se B è falso, allora A è vero.*

Per *non sequitur* si intende una argomentazione la cui conclusione deriva da premesse che non sono affatto collegate ad essa da un punto di vista logico. Ad esempio:

- Premessa: Ogni multiplo di 2 è pari.
- Premessa: 6 è un multiplo di 2.
- Conclusione: La trota è un pesce.

L'esempio precedente è ovviamente surreale; nell'uso comune il non sequitur si verifica di solito quando la prova presentata sembra essere collegata alla conclusione in modo logico ma in realtà non lo è.

L'*analogia debole* si sviluppa secondo lo schema in cui se due oggetti/soggetti hanno in comune alcune caratteristiche allora devono averne altre in comune. Ad esempio:

- Premessa: i mafiosi votano per il partito X;
- Premessa: Tizio vota X;
- Conclusione: Tizio è mafioso.

Pur non avendo valore dimostrativo, a volte le analogie deboli sono utili in matematica per suggerire congetture e dimostrazioni corrette. In alcuni casi, per facilitare la comprensione di un determinato argomento vengono presentate, avvisando il lettore, anche false dimostrazioni basate su analogie.

Esercizi.

42. Scrivere le tavole della verità dei seguenti enunciati composti:
    - $(P \text{ OR } Q) \Rightarrow (P \text{ AND } Q)$;

[9] Errare humanum est: chiunque può commettere errori di ragionamento. Tuttavia il mondo reale offre numerosi esempi di ragionamenti apparentemente logici, ma volutamente e maliziosamente erronei, fatti per convincere l'interlocutore e condizionarne le scelte: in questi casi alla fallacia logica viene dato il nome di *sofisma*.

- $(P \Rightarrow Q)$ AND (NOT P);
- $(P \Leftrightarrow Q) \Rightarrow Q$.

43. In un lontano paese, ogni abitante ha i capelli biondi oppure mori e può iscriversi all'Università soltanto chi ha o il medesimo sesso o il medesimo colore di capelli del sovrano, ma non entrambe le qualità. Non è noto se il sovrano sia maschio o femmina. Sapendo che un maschio biondo si può iscrivere, quale delle seguenti affermazioni è sicuramente vera?

(1) i maschi mori possono iscriversi,
(2) le femmine bionde possono iscriversi,
(3) i maschi mori o le femmine bionde possono iscriversi, ma non entrambi,
(4) le femmine more possono iscriversi.

44. In una classe ci sono 10 tifosi tra Roma e Lazio e sappiamo che esattamente una delle seguenti frasi è falsa. Dire quale?

(1) ci sono almeno 2 romanisti,
(2) i laziali sono non più di 5,
(3) non è vero che sono tutti romanisti,
(4) i romanisti sono almeno quanti i laziali,
(5) ci sono più di 3 laziali.

45. Spiegate i ragionamenti dei protagonisti del seguente racconto, ambientato in un lontano paese, molto tempo fa.

La Regina deve scegliere tra tre candidati, Kim, Kom e Kam, a chi affidare il governo del paese e decide di farlo valutando le loro capacità logiche e deduttive. Indi li convoca, mostra loro cinque cappelli di identica forma e dimensione, due bianchi e tre neri e dice: adesso sarete bendati e metterò un cappello in testa a ciascuno di voi in modo tale che, quando vi sarà tolta la benda, ognuno potrà vedere i cappelli degli altri due ma non il proprio. Il primo che mi dice il colore del proprio cappello diventa primo ministro, chi sbaglia sarà giustiziato immediatamente.

Kom vede che Kim ha in testa un cappello nero e Kam ha in testa un cappello bianco; dopo aver atteso inutilmente la risposta di Kim per un tempo ragionevole, Kom prende coraggio e afferma di avere un cappello nero.

Subito dopo, pure Kim dice di avere in testa un cappello nero, riceve le congratulazioni della Regina e va a portare doni e condoglianze alla giovane e bella vedova di Kom.

CAPITOLO 2

Numeri interi e razionali

Dopo aver parlato in maniera semplice e poco rigorosa dei sistemi lineari, prima di affrontare nel Capitolo 4 la parte vera e propria di algebra lineare, dedicheremo questo ed il prossimo capitolo alla cosiddetta 'Algebretta', ossia alla trattazione con linguaggio elementare di certi principi e strutture algebriche fondamentali. Nello specifico studieremo in questo capitolo i numeri interi e razionali, il principio di induzione matematica, l'analisi combinatoria ed il teorema fondamentale dell'aritmetica, mentre dedicheremo il prossimo capitolo ai numeri reali e complessi, ai polinomi ed alle funzioni razionali.

2.1. Numeri naturali, interi e razionali

> Whenever you are about to utter something astonishingly false, always begin with, "It is an acknowledged fact" etc. (Quando state per profferire qualche cosa di straordinariamente falso, cominciate sempre con la frase: "È un fatto accertato" ecc.)
> *Edward Bulwer-Lytton, Tomlinsoniana.*[1]

In questo capitolo inizieremo molte frasi con riferimenti a fatti accertati, o espressioni simili, ed in effetti le affermazioni che seguiranno, se non proprio straordinariamente false, sono spesso meno ovvie ed acclarate di come il tono perentorio e categorico usato potrebbe far pensare. Tuttavia, è necessario stabilire un punto di partenza condiviso senza (per il momento) preoccuparsi se è validato o meno alla luce delle leggi della logica e della correttezza formale. Per dirla breve, non ci interessa spendere tempo e fatica per dimostrare in maniera rigorosa che $3+2=2+3$, oppure che $7\times 8=8\times 7$ e $(4+7)\times 5=4\times 5+7\times 5$: di ciò siamo tutti convinti, andiamo avanti!

È un fatto accertato che alla base dell'aritmetica e della matematica ci sono i **numeri naturali**:

$$0,1,2,3,4,5,\dots;$$

il simbolo usato per indicare l'insieme dei numeri naturali è la enne maiuscola a doppio strato:

$$\mathbb{N}=\{0,1,2,3,4,5,6,\dots\}.$$

I numeri naturali possono essere sommati e moltiplicati nel modo che tutti conosciamo, e questo ci autorizza a dire che $\mathbb{N}$ è un **insieme numerico**.

Un altro insieme numerico che merita attenzione è quello degli **interi**, indicato con il simbolo

$$\mathbb{Z}=\{\dots,-4,-3,-2,-1,0,1,2,3,4,\dots\}.$$

Non daremo definizioni assiomatiche né degli interi né dei numeri naturali e assumeremo che il lettore ne conosca le principali proprietà, alcune delle quali saranno tuttavia ridimostrate in seguito per esigenze di tipo didattico-espositivo.

Siccome nella lingua italiana il termine intero può assumere svariati significati, per evitare ambiguità useremo parimenti il termine *numeri interi* per indicare gli elementi di $\mathbb{Z}$.

Il fatto che

$$n+0=0+n=n,\qquad n\cdot 1=1\cdot n=n,$$

per ogni intero n, si esprime a parole dicendo che 0 è *neutro* per la somma e 1 è *neutro* per il prodotto.

[1] Edward George Earle Bulwer-Lytton (1803–1873), meglio noto per la citazione "La penna è più potente della spada" e per l'incipit "Era una notte buia e tempestosa".

Come voi probabilmente sapete, dati due numeri interi a, b si può sempre dire se sono uguali e, in caso contrario, qual è il minore tra i due: scriveremo $a < b$ se a è minore di b e $a \leq b$ se a è minore o uguale a b. A scanso di equivoci diciamo subito che la dicitura *a minore o uguale a b* significa che a è minore di b oppure che a è uguale a b.[2] I numeri positivi positivi sono quelli strettamente maggiori di 0 e quelli negativi sono quelli strettamente minori di 0. Abbiamo quindi:

- interi positivi $= \{1, 2, 3, \ldots\}$;
- interi non negativi $= \{0, 1, 2, 3, \ldots\} = \mathbb{N}$;
- interi non positivi $= \{0, -1, -2, -3, \ldots\}$;
- interi negativi $= \{-1, -2, -3, \ldots\}$.

Mentre sugli interi i matematici sono tutti d'accordo, ci sono diverse opinioni se lo 0 debba essere considerato o meno un numero naturale. Per tale motivo, allo scopo di evitare ambiguità e malintesi, si preferisce spesso dire e scrivere *interi non negativi* in luogo di numeri naturali.

Notiamo che gli elementi dell'insieme $\mathbb{N}$ stanno anche nell'insieme $\mathbb{Z}$, e possiamo quindi scrivere $\mathbb{N} \subseteq \mathbb{Z}$. Un modo equivalente di esprimere la stessa cosa è

$$n \in \mathbb{N} \implies n \in \mathbb{Z},$$

dove $\implies$ è il simbolo che denota **l'implicazione logica**. Se $\mathcal{P}$ e $\mathcal{Q}$ sono due enunciati, la formula $\mathcal{P} \implies \mathcal{Q}$ (che si legge "$\mathcal{P}$ implica $\mathcal{Q}$") è un modo abbreviato per dire che se $\mathcal{P}$ è vero, allora è vero anche $\mathcal{Q}$.[3] Per esigenze grafiche scriveremo talvolta $\mathcal{Q} \impliedby \mathcal{P}$ con lo stesso significato di $\mathcal{P} \implies \mathcal{Q}$. Similmente scriveremo $\mathcal{P}_1, \mathcal{P}_2 \implies \mathcal{Q}$ per indicare che se gli enunciati $\mathcal{P}_1$ e $\mathcal{P}_2$ sono entrambi veri, allora è vero anche $\mathcal{Q}$. Ad esempio si hanno le implicazioni:

$$a < b \implies a \leq b, \qquad a \leq b,\ a \neq b \implies a < b, \qquad a = b,\ b = c \implies a = c,$$

$$a = b \implies a \leq b, \qquad a \leq b,\ b \leq a \implies a = b, \qquad a \leq b,\ b \leq c \implies a \leq c.$$

Quando scriviamo $\mathcal{P} \iff \mathcal{Q}$ intendiamo che valgono entrambe le implicazioni $\mathcal{P} \implies \mathcal{Q}$ e $\mathcal{Q} \implies \mathcal{P}$; in altri termini $\mathcal{P} \iff \mathcal{Q}$ significa che $\mathcal{P}$ è vero se e solo se $\mathcal{Q}$ è vero. Ad esempio, se a, b sono numeri interi si ha $a \leq b \iff b - a \in \mathbb{N}$.

Se $A \subseteq \mathbb{Z}$ è un sottoinsieme, diremo che un intero $m \in \mathbb{Z}$ è il massimo di A, ed in tal caso si scrive $m = \max(A)$ se $m \in A$ e se $m \geq a$ per ogni $a \in A$. Similmente diremo che $m \in \mathbb{Z}$ è il minimo di A, ed in tal caso si scrive $m = \min(A)$ se $m \in A$ e se $m \leq a$ per ogni $a \in A$. Ad esempio:

$$\max\{-1, 3, 5, 17\} = 17, \qquad \min\{-1, 3, 5, 17\} = -1.$$

È del tutto evidente che ogni insieme finito di interi possiede massimo e minimo, mentre un insieme infinito di interi non ha necessariamente né massimo né minimo. Se invece ci restringiamo agli interi non negativi si ha il seguente principio:

Principio del minimo intero. *Ogni sottoinsieme non vuoto di $\mathbb{N}$ possiede un elemento minimo.*

Infatti, se $A \subseteq \mathbb{N}$ non è vuoto deve contenere un numero $n \in A$; dato che l'intersezione di A con $\{0, 1, \ldots, n\}$ è un insieme finito, possiamo trovare un elemento $m \in A \cap \{0, 1, \ldots, n\}$ tale che $m \leq a$ per ogni $a \in A \cap \{0, 1, \ldots, n\}$, ed in particolare $m \leq n$. Se a è un qualunque elemento di A che non appartiene a $\{0, 1, \ldots, n\}$ si ha $n < a$ ed a maggior ragione $m \leq n < a$. Quindi $m = \min(A)$.

---

[2] Su questo punto la lingua inglese è più precisa: infatti la relazione $a \leq b$ viene letta *a is less than or equal to b.*

[3] Il linguaggio naturale offre vari modi equivalenti per esprimere una implicazione $\mathcal{P} \implies \mathcal{Q}$. Tra i più comuni abbiamo:

(1) $\mathcal{P}$ implica $\mathcal{Q}$,
(2) $\mathcal{Q}$ è implicato da $\mathcal{P}$,
(3) se $\mathcal{P}$ è vero, allora $\mathcal{Q}$ è vero,
(4) $\mathcal{Q}$ è vero se $\mathcal{P}$ è vero,
(5) $\mathcal{P}$ è vero solo se $\mathcal{Q}$ è vero,
(6) la verità di $\mathcal{Q}$ è condizione necessaria alla verità di $\mathcal{P}$,
(7) la verità di $\mathcal{P}$ è condizione sufficiente alla verità di $\mathcal{Q}$.

Se A è un insieme, quando si vuole indicare il sottoinsieme formato dagli elementi che godono di una determinata proprietà si usa talvolta la notazione

$$\{a \in A \mid a \text{ soddisfa la determinata proprietà}\}.$$

Ad esempio, se A, B sono sottoinsiemi di C, si ha:

$$A \cap B = \{x \in C \mid x \in A \text{ e } x \in B\},$$

$$A \cup B = \{x \in C \mid x \in A \text{ oppure } x \in B\}.$$

Si può scrivere $\mathbb{N} = \{n \in \mathbb{Z} \mid n \geq 0\}$ per affermare che i naturali altro non sono che gli interi non negativi, e similmente:

$$\begin{aligned}
\mathbb{N} &= \{n \in \mathbb{Z} \mid n+1 > 0\},\\
\{\text{numeri pari}\} &= \{n \in \mathbb{Z} \mid n \text{ è divisibile per } 2\},\\
\{\text{quadrati perfetti}\} &= \{n \in \mathbb{Z} \mid \text{esiste } a \in \mathbb{Z} \text{ tale che } n = a^2\},\\
\{\text{numeri sarchiaponici}\} &= \{n \in \mathbb{Z} \mid \text{esiste un sarchiapone ad } n \text{ zampe}\},\\
\mathbb{N} &= \{n \in \mathbb{Z} \mid n \notin \{m \in \mathbb{Z} \mid m < 0\}\},\\
\mathbb{N} &= \{n \in \mathbb{Z} \mid n + a^2 \geq 0 \text{ per ogni } a \in \mathbb{Z}\}.
\end{aligned}$$

Se a, b sono due numeri interi, diremo che a **divide** b, ed in tal caso scriveremo $a|b$, se esiste un intero c tale che $b = ac$; equivalentemente diremo che a è un **divisore** di b se $a|b$. Ogni intero non nullo, ossia diverso da 0, possiede un numero finito di divisori: infatti se $n > 0$ allora ogni suo divisore è compreso tra $-n$ ed n; se $n < 0$ ogni suo divisore divide pure $-n$ e viceversa. Diremo che due interi hanno un **fattore comune** se esiste un intero $q \geq 2$ che li divide entrambi.

La necessità di risolvere equazioni del tipo $nx = m$ con $n, m \in \mathbb{Z}$ e $n \neq 0$ comporta l'introduzione dell'insieme dei numeri razionali, indicato con il simbolo $\mathbb{Q}$. Tali numeri di solito vengono rappresentati sotto forma di frazione $x = \dfrac{m}{n}$ di due numeri interi. La rappresentazione come frazione di un numero razionale non è unica. Si noti, ad esempio che le equazioni

$$2x = 1, \quad 4x = 2, \quad 100x = 50, \quad -100x = -50,$$

sono soddisfatte tutte dallo stesso valore di $x = \dfrac{1}{2} = \dfrac{2}{4} = \dfrac{50}{100} = \dfrac{-50}{-100}$: terremo presente questo fatto dicendo che due frazioni $\dfrac{a}{b}$ e $\dfrac{c}{d}$ rappresentano lo stesso numero razionale se e solo se $ad = bc$. Dunque esistono infinite frazioni che rappresentano lo stesso numero razionale, e però ne esiste una sola in cui il denominatore è positivo e non ha fattori comuni con il numeratore: se r è un numero razionale, tra tutte le coppie (a, b) di numeri interi tali che $b \neq 0$ e $br = a$, ve ne sono certamente alcune in cui $b > 0$ e per il principio del minimo intero ne esiste una tra queste, chiamiamola (n, m), in cui m assume valore minimo. Se n ed m hanno un fattore comune, ossia $n = pa$, $m = pb$, con $p > 1$, allora $br = a$ e $0 < b < m$ in contraddizione con la minimalità di m.

I numeri razionali si sommano e si moltiplicano secondo le ben note regole, ad esempio:

$$\frac{1}{3} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}, \qquad \frac{5}{2} \cdot \frac{4}{7} = \frac{20}{14} = \frac{10}{7}.$$

Esercizi.

46. Dimostrare che se quattro numeri interi a_1, a_2, a_3, a_4 soddisfano almeno due delle seguenti quattro condizioni, allora sono tutti uguali tra loro.

(1) $a_1 \leq a_2 \leq a_3 \leq a_4$,
(2) $a_3 \leq a_1 \leq a_4 \leq a_2$,
(3) $a_2 \leq a_4 \leq a_1 \leq a_3$,
(4) $a_4 \leq a_3 \leq a_2 \leq a_1$.

47. Si trovino tre interi positivi a, b, c tali che: $a > 2b$, $bc > 3a$ ed a non divide né b né c ma divide il prodotto bc.

48. Mostrare che:

$$\{n \in \mathbb{Z} \mid 6 \text{ divide } n\} = \{n \in \mathbb{Z} \mid 2 \text{ divide } n\} \cap \{n \in \mathbb{Z} \mid 3 \text{ divide } n\},$$

$$\{n \in \mathbb{Z} \mid 6 \text{ non divide } n\} = \{n \in \mathbb{Z} \mid n \text{ è dispari}\} \cup \{n \in \mathbb{Z} \mid 3 \text{ non divide } n\}.$$

49. Determinare le relazione tra un sottoinsieme $A \subseteq \mathbb{Z}$ ed il sottoinsieme

$$B = \{x \in \mathbb{Z} \mid x \notin \{y \in \mathbb{Z} \mid y \notin A\}\}.$$

50. Per ogni successione finita $(a_1, \dots, a_n)$ di interi positivi tali che $a_1 \geq a_2 \geq \cdots \geq a_n$, $n > 0$, definiamo

$$T(a_1, \dots, a_n) = (b_1, \dots, b_k),$$

dove $k = a_1$ e per ogni $1 \leq h \leq k$, si definisce b_h uguale al numero di indici j tali che $a_j \geq h$. Ad esempio:

$$T(3,2) = (2,2,1), \quad T(1,1,1,1) = (4), \quad T(3,3,1) = (3,2,2).$$

Dimostrare che se $T(a_1, \dots, a_n) = (b_1, \dots, b_k)$, allora:

(1) $b_1 = n$,
(2) $a_1 + \cdots + a_n = b_1 + \cdots + b_k$,
(3) $T(b_1, \dots, b_k) = (a_1, \dots, a_n)$.

51 (♡). Siano a, c due numeri razionali tali che $a > c \geq 0$. Provare che esiste un numero razionale $b > 0$ tale che $c < b^2 \leq a$.

2.2. Applicazioni tra insiemi

Definizione 2.2.1. Una **applicazione** da un insieme A ad un insieme B è una legge, di qualunque natura, che ad ogni elemento di A associa uno ed un solo elemento di B. Indicheremo un'applicazione da A in B con il simbolo

$$f \colon A \to B, \qquad a \mapsto f(a),$$

dove, per ogni $a \in A$, l'elemento $f(a) \in B$ è quello associato ad a tramite l'applicazione medesima.

Esempio 2.2.2. Ecco alcuni esempi di applicazioni:

(1)
$$f \colon \mathbb{N} \to \mathbb{N}, \qquad n \mapsto f(n) = 2n$$
è l'applicazione che ad ogni numero naturale associa il suo doppio.

(2)
$$f \colon \mathbb{Z} \to \mathbb{Z}, \qquad n \mapsto f(n) = n^2$$
è l'applicazione che ad ogni numero intero associa il suo quadrato.

(3)
$$f \colon \{\text{Uomini}\} \to \{\text{Date}\} \qquad f(x) = \text{data di nascita di } x,$$
è un'applicazione.

Esempio 2.2.3. Dato un qualunque insieme A, l'applicazione **identità:**

$$\mathrm{Id} \colon A \to A, \qquad \mathrm{Id}(a) = a,$$

è l'applicazione che associa ad ogni elemento se stesso. Più in generale se $B \subseteq A$, l'applicazione di **inclusione** è definita come

$$i \colon B \to A, \qquad i(b) = b.$$

In altri termini se $b \in B$, allora $i(b)$ è lo stesso elemento pensato però come appartenente all'insieme A.

Due applicazioni f, g da un insieme A ad un insieme B sono uguali se $f(a) = g(a)$ per ogni $a \in A$. Conseguentemente sono diverse se esiste almeno un elemento $a \in A$ tale che $f(a) \neq g(a)$.

Esempio 2.2.4. Le due applicazioni

$$f, g\colon \mathbb{Z} \to \mathbb{Z}, \qquad f(n) = n^2, \quad g(n) = (n-1)^2 + 2n - 1$$

sono uguali. Infatti per ogni numero intero n vale

$$g(n) = (n-1)^2 + 2n - 1 = n^2 - 2n + 1 + 2n - 1 = n^2 = f(n).$$

Esempio 2.2.5. Le due applicazioni

$$f, g\colon \mathbb{N} \to \mathbb{N}, \qquad f(n) = n, \quad g(n) = \max(2, n) = \text{massimo tra 2 e } n,$$

sono diverse poiché $f(1) = 1$ e $g(1) = 2$; si noti che $f(n) = g(n)$ per ogni $n > 1$.

Definizione 2.2.6. Chiameremo **immagine** di un'applicazione $f\colon A \to B$, e la denoteremo con $f(A)$, l'insieme degli elementi di B che sono del tipo $f(a)$ per qualche $a \in A$, ossia

$$f(A) = \{f(a) \mid a \in A\}.$$

Equivalentemente

$$f(A) = \{b \in B \mid \text{ esiste } a \in A \text{ tale che } b = f(a)\}.$$

Chiaramente $f(A)$ è un sottoinsieme di B.

Esempio 2.2.7. L'immagine dell'applicazione $f\colon \mathbb{N} \to \mathbb{Z}$, $f(n) = n + 1$, è l'insieme degli interi positivi, ossia $f(\mathbb{N}) = \{x \in \mathbb{Z} \mid x > 0\}$.

Osservazione 2.2.8. Nella lingua italiana, il termine funzione può essere usato come sinonimo di applicazione, anche se la tendenza prevalente è quella di chiamare funzioni le applicazioni che assumono valori numerici. Recentemente si sta imponendo l'uso del sostantivo femminile *mappa* come sinonimo di applicazione. A mio modesto parere tale uso andrebbe evitato: nell'italiano antico il termine mappa indicava una tovaglia (da cui il termine mappina=straccio) ed è passato poi ad indicare le carte geografiche, i cui primi esemplari venivano disegnati per l'appunto su tovaglie. Personalmente ritengo invece accettabile, anche se talvolta poco gradevole, l'uso del verbo *mappare* (dall'inglese "to map") in ambito matematico. Esempio: il numero 7 viene mappato nel numero 45 dell'applicazione $\mathbb{Z} \to \mathbb{Z}$, $n \mapsto n^2 - 4$.

Il seguente principio è nella sostanza del tutto equivalente al principio del minimo intero.

Siano A un insieme non vuoto e $f\colon A \to \mathbb{N}$ un'applicazione. Esiste allora un elemento $x \in A$ tale che $f(x) \leq f(a)$ per ogni $a \in A$.

Partiamo da un qualsiasi $x \in A$, se $f(x) \leq f(a)$ per ogni $a \in A$ abbiamo finito, altrimenti esiste $x_1 \in A$ tale che $f(x_1) < f(x)$. Se $f(x_1)$ è il minimo abbiamo finito, altrimenti esiste $x_2 \in A$ tale che $f(x_2) < f(x_1)$ eccetera. Dato che $f(x_i) \geq 0$ per ogni i, questa procedura non può proseguire all'infinito e ad un certo punto $f(x_n) \leq f(a)$ per ogni a.

Definizione 2.2.9. Un'applicazione $f\colon A \to B$ si dice **iniettiva** se manda elementi distinti di A in elementi distinti di B. In altri termini f è iniettiva se vale l'implicazione

$$a \neq b \implies f(a) \neq f(b),$$

o equivalentemente se vale

$$f(a) = f(b) \implies a = b.$$

Conseguentemente, f non è iniettiva se esistono $a, b \in A$ tali che $a \neq b$ e $f(a) = f(b)$.

Esempio 2.2.10. L'applicazione

$$f\colon \mathbb{Z} \to \mathbb{Z}, \qquad f(n) = n + 1,$$

è iniettiva. Per provarlo bisogna dimostrare che se $f(n) = f(m)$, allora $n = m$. Questo è facile: se $f(n) = f(m)$, allora $n + 1 = m + 1$ e quindi $n = m$.

Esempio 2.2.11. L'applicazione

$$f\colon \mathbb{Z} \to \mathbb{Z}, \qquad f(n) = n^2,$$

non è iniettiva. Per provarlo è sufficiente trovare due numeri interi n, m tali che $n \neq m$ e $f(n) = f(m)$. Anche questo è facile: infatti $f(1) = f(-1)$.

DEFINIZIONE 2.2.12. Un'applicazione $f\colon A \to B$ si dice **surgettiva** se ogni elemento di B è l'immagine di almeno un elemento di A. Equivalentemente $f\colon A \to B$ è surgettiva se $f(A) = B$.

ESEMPIO 2.2.13. L'applicazione

$$f\colon \mathbb{Z} \to \mathbb{Z}, \qquad f(n) = n + 1,$$

è surgettiva. Infatti per ogni $n \in \mathbb{Z}$ si ha $n = f(n-1)$ e quindi $n \in f(\mathbb{Z})$.

ESEMPIO 2.2.14. L'applicazione

$$f\colon \mathbb{Z} \to \mathbb{Z}, \qquad f(n) = n^2,$$

non è surgettiva. Per provarlo è sufficiente trovare un intero n che non appartiene all'immagine di f. Questo non potrebbe essere più facile: infatti $-1 \notin f(\mathbb{Z})$.

DEFINIZIONE 2.2.15. Un'applicazione si dice **bigettiva** se è contemporaneamente iniettiva e surgettiva.

OSSERVAZIONE 2.2.16. A volte si scrive $A \xrightarrow{f} B$ per indicare un'applicazione $f\colon A \to B$.

Dati due insiemi A e B si definisce il **prodotto cartesiano** $A \times B$ come l'insieme di tutte le coppie ordinate (a, b) con $a \in A$ e $b \in B$, e cioè

$$A \times B = \{(a, b) \mid a \in A,\ b \in B\}.$$

Le due applicazioni

$$p_1\colon A \times B \to A, \qquad p_1(a, b) = a,$$

$$p_2\colon A \times B \to B, \qquad p_2(a, b) = b,$$

si dicono **proiezioni**, sul primo e secondo fattore rispettivamente.

In maniera simile si definisce il prodotto cartesiano di tre insiemi $A \times B \times C$ come l'insieme di tutte le terne ordinate (a, b, c), con $a \in A, b \in B, c \in C$, e più in generale il prodotto cartesiano di una qualunque famiglia finita di insiemi

$$A_1 \times \cdots \times A_n = \{(a_1, \ldots, a_n) \mid a_1 \in A_1,\ a_2 \in A_2,\ \ldots, a_n \in A_n\}.$$

Le potenze cartesiane di un insieme A sono date dai prodotti cartesiani di A con se stesso: ad esempio la potenza cartesiana tripla $A \times A \times A$ di A è l'insieme di tutte le terne ordinate (a_1, a_2, a_3) con $a_i \in A$ per ogni i. Poiché le potenze cartesiane ricorrono spesso in matematica è utile introdurre la seguente notazione semplificata:

$$A^{(1)} = A, \quad A^{(2)} = A \times A, \quad A^{(3)} = A \times A \times A, \quad \ldots, \quad A^{(n)} = \underbrace{A \times \cdots \times A}_{n \text{ fattori}}.$$

Per convenzione si pone $A^{(0)} = * = \mathcal{P}(\emptyset)$, ossia il prodotto vuoto è uguale al singoletto.

Alcune semplici ma utili osservazioni sono:

(1) l'insieme $A \times B$ è vuoto se e solo se almeno uno tra A e B è vuoto;
(2) se A e B sono insiemi finiti, allora anche $A \times B$ è un insieme finito ed il numero di elementi di $A \times B$ è uguale al prodotto del numero di elementi di A per il numero di elementi di B;
(3) se $f\colon A \to C$ e $g\colon B \to D$ sono due applicazioni iniettive (risp.: surgettive, bigettive) allora l'applicazione

$$f \times g\colon A \times B \to C \times D, \qquad f \times g(a, b) = (f(a), g(b)),$$

è iniettiva (risp.: surgettiva, bigettiva).

Per prevenire un possibile errore logico, notiamo che se A e B sono due insiemi distinti, allora anche i due insiemi $A \times B$ e $B \times A$ *sono distinti*, pur esistendo una ovvia e naturale bigezione

$$A \times B \to B \times A, \qquad (a, b) \mapsto (b, a).$$

Esercizi.

52 (Somme telescopiche). Mostrare che per ogni intero positivo n si ha:

$$-1+4-9-16+\cdots+(-1)^n n^2 = (-1)^n \frac{n(n+1)}{2},$$

$$\frac{1}{2}+\frac{1}{6}+\frac{1}{12}+\cdots+\frac{1}{n(n+1)} = \frac{n}{n+1}.$$

Suggerimento:

$$n^2 = \frac{n(n+1)}{2} + \frac{(n-1)n}{2}, \qquad \frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}.$$

53. Per ciascuna delle seguenti applicazioni, dire se è iniettiva, se è surgettiva e se è bigettiva:

(1) $f\colon \mathbb{Z} \to \mathbb{Z}$, $f(n) = n^2$;
(2) $f\colon \mathbb{Z} \to \mathbb{Z}$, $f(n) = 2n - 1$;
(3) $f\colon \mathbb{Z} \to \mathbb{Z}$, $f(n) = n^2 + n + 1$.

54. Siano A_1, A_2 sottoinsiemi di un insieme A e B_1, B_2 sottoinsiemi di un insieme B. Per ogni $i = 1, 2$ indichiamo con $A_i \times B_i \subseteq A \times B$ il sottoinsieme formato dalle coppie (a, b), con $a \in A_i$ e $b \in B_i$. Dimostrare che:

(1) $(A_1 \times B_1) \cap (A_2 \times B_2) = (A_1 \cap A_2) \times (B_1 \cap B_2)$;
(2) $(A_1 \cap A_2) \times (B_1 \cup B_2) \subseteq (A_1 \times B_1) \cup (A_2 \times B_2)$;
(3) $(A_1 \times B_1) \cup (A_2 \times B_2) \subseteq (A_1 \cup A_2) \times (B_1 \cup B_2)$ e mostrare con un esempio che in generale non vale l'inclusione inversa $\supseteq$.

55. Dimostrare che l'applicazione

$$f\colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}, \qquad f(x, y) = \frac{(x+y)(x+y+1)}{2} + x,$$

è bigettiva.

2.3. Il principio di induzione matematica

L'induzione matematica è uno strumento utile per stabilire una verità matematica in una successione infinita di casi, il primo, il secondo, il terzo e così via, senza eccezione alcuna. Supponiamo di avere, per ogni intero positivo n una proposizione $\mathcal{A}_n$ relativa ad n. Ad esempio $\mathcal{A}_n$ potrebbe essere "Un segmento può essere diviso in n parti uguali" oppure "Il numero n si può scrivere come somma di quattro quadrati".

Un altro esempio di proposizione $\mathcal{A}_n$ potrebbe essere "Il numero $3n$ è pari": chiaramente $\mathcal{A}_n$ è vera se n è pari ed è falsa per n dispari.

Principio di induzione matematica (prima formulazione). Sia data per ogni intero positivo n una proposizione $\mathcal{A}_n$. Se:

(1) la proposizione $\mathcal{A}_1$ è vera,
(2) in base a qualche ragionamento matematico dimostriamo che, se $\mathcal{A}_n$ è vera, per n intero positivo qualsiasi, allora segue la validità di $\mathcal{A}_{n+1}$.

Allora $\mathcal{A}_n$ è vera per ogni $n \geq 1$.

Trattandosi di un principio alla base dell'aritmetica e delle proprietà dei numeri naturali, la cosa migliore per comprenderlo appieno è vederlo all'opera in una serie di esempi interessanti.

Esempio 2.3.1. Se un insieme finito X contiene n elementi, allora l'insieme $\mathcal{P}(X)$ delle parti di X contiene esattamente 2^n elementi. Si verifica immediatamente che il risultato è vero per piccoli valori di n:

$$\mathcal{P}(\emptyset) = \{\emptyset\}, \qquad \mathcal{P}(\{a\}) = \{\emptyset, \{a\}\}, \quad \mathcal{P}(\{a, b\}) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}, \dots .$$

Sia dunque X un insieme con $n > 0$ elementi, scegliamone uno $x \in X$ e indichiamo con $Y = X - \{x\}$ il sottoinsieme complementare. I sottoinsiemi di X si dividono in due categorie: quelli che non contengono x e quelli che lo contengono. I primi sono esattamente i sottoinsiemi

di Y, mentre i secondi sono tutti e soli quelli della forma $\{x\} \cup A$, con $A \subseteq Y$. Ne deduciamo che $\mathcal{P}(X)$ ha il doppio degli elementi di $\mathcal{P}(Y)$; siccome Y possiede $n-1$ elementi, per l'ipotesi induttiva $\mathcal{P}(Y)$ contiene 2^{n-1} elementi e quindi $\mathcal{P}(X)$ contiene 2^n elementi.

ESEMPIO 2.3.2. Siano n un intero positivo ed S un insieme di $2n$ elementi che è unione disgiunta di suoi sottoinsiemi $S_1, \ldots, S_h$, alcuni dei quali possibilmente vuoti: unione disgiunta significa che $S = S_1 \cup \cdots \cup S_h$ e $S_i \cap S_j = \emptyset$ per ogni $i \neq j$. Supponiamo che ciascun sottoinsieme S_i contenga al massimo n elementi. Dimostriamo per induzione su n che esiste una indicizzazione

$$S = S_1 \cup \cdots \cup S_h = \{x_1, \ldots, x_{2n}\}$$

tale che per ogni indice $i = 1, \ldots, n$ i due elementi x_{2i-1} e x_{2i} non appartengono al medesimo insieme S_j.

Indichiamo con n_i il numero di elementi di S_i; meno di scambiare (permutare) gli indici non è restrittivo supporre

$$n_1 + \cdots + n_h = 2n, \qquad n \geq n_1 \geq n_2 \geq \cdots \geq n_h \geq 0.$$

Si noti che le precedenti condizioni implicano $n_1 \geq n_2 > 0$ e $n_i < n$ per ogni $i > 2$.

Prendiamo due elementi $x_{2n} \in S_1$, $x_{2n-1} \in S_2$ e consideriamo la nuova famiglia di insiemi disgiunti

$$T_1 = S_1 - \{x_{2n}\}, \quad T_2 = S_2 - \{x_{2n-1}\}, \quad T_3 = S_3, \ \ldots, \ T_h = S_h.$$

Ciascun T_i contiene al più $n-1$ punti e la loro unione $T_1 \cup \cdots \cup T_h$ contiene $2n-2$ punti. Per l'ipotesi induttiva possiamo scrivere

$$S - \{x_{2n-1}, x_{2n}\} = T_1 \cup \cdots \cup T_h = \{x_1, \ldots, x_{2n-2}\}$$

con x_{2i-1} e x_{2i} non appartenenti al medesimo insieme T_j. Per finire basta osservare che l'indicizzazione $S = \{x_1, \ldots, x_{2n}\}$ ha le proprietà richieste.

ESEMPIO 2.3.3. Sia $\mathcal{A}_n$ la proposizione: *esistono due interi positivi a, b tali che* $5a+6b = 35+n$. La proposizione $\mathcal{A}_1$ è vera, in quanto $6 \cdot 5 + 6 = 36$. Supponiamo adesso $n > 1$ e che $\mathcal{A}_{n-1}$ sia vera; esistono quindi $a, b > 0$ tali che $5a + 6b = 35 + n - 1$. Se $a > 1$ allora

$$5(a-1) + 6(b+1) = 5a + 6b + 1 = 35 + n.$$

Se invece $a = 1$, allora $6b = 30 + n - 1$ e quindi $b \geq 5$. Possiamo allora scrivere $5a + 6b = 5(a+6) + 6(b-5) = 35 + n - 1$ e ragionando come sopra $5(a+5) + 6(b-4) = 35 + n$. Abbiamo quindi dimostrato $\mathcal{A}_n$ e per il principio di induzione $\mathcal{A}_n$ è vera per ogni n.

ESEMPIO 2.3.4. Dimostriamo che per ogni intero $n > 0$ vale $2^n \geq n + 1$. In questo caso la proposizione $\mathcal{A}_1$ è la disugualianza $2^1 \geq 1 + 1$, la proposizione $\mathcal{A}_2$ è la disugualianza $2^2 \geq 2 + 1$ eccetera. Siccome $2^1 = 2 \geq 1 + 1$, la proposizione $\mathcal{A}_1$ è vera.

Supponiamo adesso che, per un intero qualsiasi n la proposizione $\mathcal{A}_n$ sia vera, ossia che $2^n \geq n + 1$: come detto si tratta per il momento di un'ipotesi, in quanto la verità o la falsità di $\mathcal{A}_n$ sarà stabilita al termine del procedimento. Supponiamo quindi $2^n \geq n + 1$, allora si ha:

$$2^{n+1} = 2^n + 2^n \geq 2^n + n + 1 \geq (n+1) + 1$$

dove nella disuguaglianza a destra abbiamo unato il fatto che $2^n \geq 1$. Dunque supponendo vera $\mathcal{A}_n$ (ossia $2^n \geq n + 1$) abbiamo dimostrato che anche $\mathcal{A}_{n+1}$ è vera (ossia $2^{n+1} \geq n + 2$) e per il principio di induzione $\mathcal{A}_n$ è vera per ogni n.

Alla stessa maniera si dimostra che, per ogni numero razionale positivo a e per ogni intero positivo n vale la disuguaglianza $(a+1)^n \geq a^n + na^{n-1}$. Per $n = 1$ la disuguaglianza diventa $a + 1 \geq a + 1$ che è chiaramente vera. Supponendo quindi vero che $(a+1)^n \geq a^n + na^{n-1}$ si ottiene

$$\begin{aligned}(a+1)^{n+1} &= (a+1)^n(a+1) \geq (a^n + na^{n-1})(a+1) = a^{n+1} + (n+1)a^n + na^{n-1} \\ &\geq a^{n+1} + (n+1)a^n\,.\end{aligned}$$

ESEMPIO 2.3.5 (Disuguaglianza di Bernoulli). Dimostriamo che per ogni numero razionale $t \geq -1$ e per ogni intero positivo n vale la disuguaglianza

$$(1+t)^n \geq 1 + nt.$$

In questo caso la proposizione $\mathcal{A}_n$ è $(1+t)^n \geq 1+nt$: la proposizione $\mathcal{A}_1$ diventa $1+t \geq 1+t$ che è ovviamente vera. Supponiamo adesso vero che $(1+t)^n \geq 1+nt$, allora

$$(1+t)^{n+1} = (1+t)(1+t)^n = (1+t)^n + t(1+t)^n \geq 1 + nt + t(1+t)^n.$$

Se riusciamo a dimostrare che $t(1+t)^n \geq t$ allora dalla disuguaglianza precedente segue che $(1+t)^{n+1} \geq 1+(n+1)t$ ed abbiamo provato la validità di $\mathcal{A}_{n+1}$. Per dimostrare che $t(1+t)^n \geq t$ per ogni $t \geq -1$ ed ogni $n > 0$ trattiamo separatamente i casi $t \geq 0$ e $-1 \leq t < 0$. Se $t \geq 0$ allora $(1+t)^n \geq 1$ e quindi $t(1+t)^n \geq t \cdot 1 = t$. Se invece $-1 \leq t < 0$ allora $0 \leq (1+t)^n < 1$ e, siccome t è negativo si ha $t(1+t)^n > t$.

Notiamo che da tale disuguaglianza segue in particolare che per ogni coppia di numeri razionali b, c, con $b > 1$, esiste un intero positivo n tale che $b^n \geq c$. Infatti, scrivendo $b = 1+t$ si ha $b^n \geq 1+nt$ e basta prendere n sufficientemente grande e tale che $n \geq (c-1)/t$.

Esempio 2.3.6. Dimostriamo che per ogni n la somma $1+2+\cdots+n$ dei primi n interi positivi è uguale a $\dfrac{n(n+1)}{2}$. Tale affermazione è vera per $n = 1$, mentre se la supponiamo vera per n si ha

$$1+2+\cdots+n+(n+1) = \frac{n(n+1)}{2} + (n+1) = \frac{n^2+n+2(n+1)}{2} = \frac{(n+1)(n+2)}{2}.$$

Esempio 2.3.7. Dimostriamo che per ogni intero $n > 0$ vale la disuguaglianza

$$5^n \geq 3^{n-1}(2n+3).$$

Per $n = 1$ tale disuguaglianza diventa $5 \geq 5$ che è dunque vera. Supponiamola vera per un qualsiasi intero positivo n e scriviamo

$$\begin{aligned} 5^{n+1} = 3\cdot 5^n + 2\cdot 5^n &\geq 3\cdot 3^{n-1}(2n+3) + 2\cdot 3^{n-1}(2n+3) \\ &\geq 3^n(2n+3) + 3^{n-1}4n + 2\cdot 3^n \\ &\geq 3^n(2n+5) + 3^{n-1}4n \\ &\geq 3^n(2(n+1)+3). \end{aligned}$$

Esempio 2.3.8. Dimostriamo che per ogni n la somma $1^2+2^2+\cdots+n^2$ dei quadrati dei primi n interi positivi è uguale a $\dfrac{n(n+1)(2n+1)}{6}$. Tale affermazione è vera per $n = 1$, mentre se la supponiamo vera per n si ha

$$1+2^2+\cdots+n^2+(n+1)^2 = \frac{n(n+1)(2n+1)}{6} + (n+1)^2.$$

Lasciamo al lettore il compito di verificare l'uguaglianza

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

Ricordiamo il significato del simbolo di sommatoria Σ: date le quantità $a_1, \ldots, a_n$ si pone

$$\sum_{i=1}^{n} a_i = a_1 + \cdots + a_n,$$

ogniqualvolta la somma a destra del segno di uguaglianza è ben definita.

Più in generale se $m \leq n$ si pone

$$\sum_{i=m}^{n} a_i = a_m + \cdots + a_n.$$

Possiamo quindi riformulare i risultati degli Esempi 2.3.6 e 2.3.8 dicendo che per ogni n valgono le uguaglianze

$$\sum_{i=1}^{n} i = \frac{n(n-1)}{2}, \qquad \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}.$$

Esempio 2.3.9. Dimostriamo per induzione su n che

$$\sum_{i=0}^{n} \frac{1}{10^i} \leq \frac{10}{9}.$$

La disuguaglianza è vera per $n = 1$ in quanto $1 + 1/10 < 10/9$. Supponiamo adesso $n > 1$ e che $\sum_{i=0}^{n-1} \frac{1}{10^i} \leq \frac{10}{9}$; allora

$$\sum_{i=0}^{n} \frac{1}{10^i} = 1 + \sum_{i=1}^{n} \frac{1}{10^i} = 1 + \frac{1}{10} \sum_{i=0}^{n-1} \frac{1}{10^i} \leq 1 + \frac{1}{10}\frac{10}{9} = \frac{10}{9}.$$

Giova ricordare che la variabile i usata nei precedenti simboli di sommatoria è muta (o apparente, o fittizia) e nulla cambia nella sostanza se viene sostituita con una variabile di nome diverso: la nuova variabile può essere uguale alla precedente oppure diversa ma dipendente dalla vecchia in maniera biunivoca. Ad esempio si hanno le uguaglianze:

$$\sum_{i=1}^{n} a_i = \sum_{j=1}^{n} a_j = \sum_{k=0}^{n-1} a_{k+1} = \sum_{h=0}^{n-1} a_{n-h} = \sum_{l=1}^{n} a_{n+1-l}\,.$$

Esistono diverse variazioni del principio di induzione matematica che possono risultare utili in determinate situazioni:

Principio di induzione matematica (seconda formulazione). Sia data per ogni intero positivo n una proposizione $\mathcal{A}_n$. Se:

(1) la proposizione $\mathcal{A}_1$ è vera,
(2) in base a qualche ragionamento matematico dimostriamo che $\mathcal{A}_n$ è vera ogniqualvolta le proposizioni $\mathcal{A}_k$ sono vere per ogni $1 \leq k < n$.

Allora $\mathcal{A}_n$ è vera per ogni $n \geq 1$.

È facile mostrare l'equivalenza tra le due formulazioni del principio di induzione: infatti se $\mathcal{A}_n$, $n > 0$, è una successione infinita di proposizioni definiamo una nuova successione $\mathcal{B}_n$ di proposizioni mediante la regola:

$$\mathcal{B}_n \text{ è vera se e solo se } \mathcal{A}_k \text{ è vera per ogni } 1 \leq k \leq n\,.$$

È chiaro che $\mathcal{A}_n$ è vera per ogni n se e solo se $\mathcal{B}_n$ è vera per ogni n. Basta adesso osservare che $\mathcal{A}_n$ soddisfa la seconda formulazione del principio di induzione se e solo se $\mathcal{B}_n$ soddisfa la prima formulazione.

Un utile esercizio teorico è mostrare che il principio di induzione segue dal principio del minimo intero. Sia data una successione di proposizioni $\mathcal{A}_n$ che soddisfa le ipotesi del principio di induzione e supponiamo per assurdo che l'insieme S degli interi positivi s per cui A_s è falsa sia non vuoto. Per il principio del minimo l'insieme S possiede un valore minimo $n = \min(S)$ che non può essere $= 1$ poiché A_1 è vera per ipotesi. Dunque $n > 1$ e questo implica $n-1 > 0$, $n-1 \notin S$ e quindi A_k è vero per ogni $k < n$. Adesso, in entrambe le formulazioni del principio di induzione, le ipotesi implicano che A_n è vera, in contraddizione con l'appartenenza di n ad S.

Esempio 2.3.10. Usiamo la seconda formulazione del principio di induzione per dimostrare che per ogni intero positivo n esiste un numero naturale a tale che $n = 2^a m$ con m dispari. Per $n = 1$ il risultato è vero (con $a = 0$). Sia $n > 1$ e supponiamo il risultato vero per ogni $1 \leq k < n$: se n è dispari si ha $n = 2^0 n$, mentre se n è pari si ha $n = 2k$ con $1 \leq k < n$. Per l'ipotesi induttiva $k = 2^b m$ con $b \in \mathbb{N}$ ed m dispari. Questo implica $n = 2^{b+1} m$ come volevasi dimostrare.

Un'altra ovvia variazione del principio di induzione si ha quando gli indici risultano spostati di un qualsiasi numero intero: più precisamente, supponiamo di avere un numero intero N e sia data per ogni intero $n \geq N$ una proposizione $\mathcal{A}_n$. Se:

(1) la proposizione $\mathcal{A}_N$ è vera,
(2) in base a qualche ragionamento matematico dimostriamo che, se $\mathcal{A}_n$ è vera, per $n \geq N$ intero qualsiasi, allora segue la validità di $\mathcal{A}_{n+1}$.

Allora $\mathcal{A}_n$ è vera per ogni $n \geq N$.

Per rendersi conto della validità delle precedente affermazione basta applicare il principio di induzione alle proposizioni $\mathcal{B}_n = \mathcal{A}_{N+n-1}$, $n > 0$.

Parenti stretti del principio di induzione sono le definizioni **ricorsive**, frequentemente usate per definire applicazioni $f\colon \mathbb{N} \to X$, con X insieme qualunque. Detto in parole semplici

un'applicazione $f\colon \mathbb{N} \to X$ è definita in modo ricorsivo se per ogni $n > 0$ il valore $f(n)$ dipende, secondo una determinata regola, dai valori $f(0), f(2), \dots, f(n-1)$. Ad esempio un modo alternativo di definire l'applicazione $f(n) = n + 3$ è:

$$f(0) = 3, \qquad f(n) = f(n-1) + 1, \quad n > 1.$$

Similmente, modi alternativi per definire le funzioni $f(n) = n^2$ e $g(n) = 2^n$ sono:

$$f(0) = 0, \qquad f(n) = f(n-1) + 2n - 1, \quad n > 0;$$

$$g(0) = 1, \qquad g(n) = 2g(n-1), \quad n > 0;$$

$$g(0) = 1, \quad g(1) = 2, \qquad g(n) = g(n-1) + 2g(n-2), \quad n > 1.$$

Senza voler entrare in dettagli eccessivamente pedanti, possiamo dire che nel metodo ricorsivo il principio di induzione viene usato non per dimostrare una successione di enunciati ma per definire un'applicazione.

Esercizi.

56. Consideriamo l'applicazione $f\colon \mathbb{N} \to \mathbb{Z}$ definita ricorsivamente dalle formule

$$f(0) = 1, \qquad f(n) = n - f(n-1), \quad n > 0.$$

Calcolare $f(5)$.

57. Consideriamo l'applicazione $f\colon \mathbb{N} \to \mathbb{Z}$ definita ricorsivamente dalle formule

$$f(0) = 1, \quad f(1) = 1 \qquad f(n) = f(n-1) + f(n-2), \quad n > 1.$$

Calcolare $f(5)$.

58. Siano $a_1, \dots, a_n$ quantità numeriche. Provare che per ogni $1 \le p \le q \le n$ vale

$$\sum_{i=1}^{n} a_i + \sum_{i=p}^{q} a_i = \sum_{i=1}^{q} a_i + \sum_{i=p}^{n} a_i.$$

59. Sia $f\colon \{1, \dots, n\} \to \{1, \dots, n\}$ un'applicazione bigettiva. Dimostrare che

$$\sum_{i=1}^{n} i(i - f(i)) = \frac{1}{2} \sum_{i=1}^{n} (i - f(i))^2 \ge 0.$$

60. Dire se le seguenti formule sono giuste o sbagliate:

$$\sum_{i=1}^{n} \sum_{j=1}^{i} (i^2 + j^3) = \sum_{j=1}^{n} \sum_{i=j}^{n} (i^2 + j^3),$$

$$\sum_{i=0}^{n} \sum_{j=0}^{n-i} (i^2 + j^3) = \sum_{k=0}^{n} \sum_{h=0}^{k} ((k-h)^2 + h^3).$$

61. Dimostrare, usando il principio di induzione, che la somma

$$1 + 3 + \cdots + (2n-1) = \sum_{i=1}^{n} (2i - 1)$$

dei primi n numeri dispari positivi è uguale a n^2. Successivamente, formulare in maniera matematicamente precisa e dimostrare la seguente osservazione attribuita a Galileo (1615):

$$\frac{1}{3} = \frac{1+3}{5+7} = \frac{1+3+5}{7+9+11} = \frac{1+3+5+7}{9+11+13+15} = \cdots .$$

62. Scrivere la successione dei numeri dispari positivi in infinite righe secondo lo schema

$$\begin{array}{ccccccc} & & & & 1 & & \\ & & & 3 & 5 & & \\ & & 7 & 9 & 11 & , & \\ & 13 & 15 & 17 & 19 & & \\ 21 & 23 & \cdots & & & & \end{array}$$

formulare una congettura su quanto vale la somma degli n numeri sulla riga n e darne una dimostrazione rigorosa.

63. Sia q un numero diverso da 1. Usare il principio di induzione per mostrare che la somma delle prime n potenze di q è uguale a

$$q + q^2 + \cdots + q^n = q\frac{q^n - 1}{q - 1}.$$

64. Usare il principio di induzione per mostrare che per ogni intero $n > 0$ vale

$$\sum_{i=1}^{n} \frac{1}{i+n} = 1 - \frac{1}{2} + \cdots + \frac{1}{2n-1} - \frac{1}{2n} = \sum_{i=1}^{2n} \frac{(-1)^{i-1}}{i}.$$

65. Usando la diseguaglianza $\left(\frac{4}{3}\right)^3 < 3$, dimostrare che $n^3 \leq 3^n$ per ogni numero intero n.

(Suggerimento: trattare prima i casi $n < 0$, $n = 0, 1, 2$. Successivamente trattare il caso $n \geq 3$ usando il principio di induzione.)

66. Dimostrare per induzione che per ogni intero $n > 0$ si ha

$$1 + \frac{1}{4} + \frac{1}{9} + \cdots + \frac{1}{n^2} \leq 2 - \frac{1}{n}.$$

67. Siano date per ogni intero $n > 0$ due proposizioni P_n e Q_n. Si supponga inoltre che:
(1) P_1 è vera.
(2) Se P_n è vera, allora anche Q_n è vera.
(3) Se Q_s è vera per ogni $s < n$, allora anche P_n è vera.

Dimostrare che P_n, Q_n sono vere per ogni n.

68 (☕). Congetturate una formula generale per il prodotto

$$\left(1 - \frac{1}{4}\right)\left(1 - \frac{1}{9}\right)\left(1 - \frac{1}{16}\right)\cdots\left(1 - \frac{1}{n^2}\right), \qquad n \geq 2,$$

e dimostratela usando il principio di induzione.

69 (☕). Dopo aver osservato che

$$1 + 2 = 3$$
$$4 + 5 + 6 = 7 + 8$$
$$9 + 10 + 11 + 12 = 13 + 14 + 15$$
$$16 + 17 + 18 + 19 + 20 = 21 + 22 + 23 + 24,$$

ee pensate che si può andare avanti illimitatamente, descrivete questo fatto con una formula e dimostratela.

70 (☕). Il numero di coppie (a, b) di interi non negativi tali che $a + b = n$ è chiaramente uguale a $n + 1$ (le coppie sono $(n, 0), (n-1, 1), \ldots, (0, n)$).

Usare questo fatto ed il principio di induzione per mostrare che il numero di terne (a, b, c) di interi non negativi tali che $a + b + c = n$ è uguale a $\dfrac{(n+1)(n+2)}{2}$.

71. Dopo aver calcolato per $n = 1, 2, 3, 4, 5, 6$ il valore dell'espressione

$$P(n) = (n^2 - 5n + 5)^{(n^2 - 11n + 30)},$$

siete in grado di congetturare (senza fare i conti) quanto vale $P(10)$?

72 (Il problema del fidanzamento, ☕, ♡). Sia M un insieme di n ragazzi, ciascun $m \in M$ conosce un insieme F_m di ragazze. Si chiede se è possibile far fidanzare tutti i ragazzi in maniera univoca, ossia se esiste un'applicazione iniettiva

$$f\colon M \to \bigcup_{m \in M} F_m, \qquad \text{tale che} \quad f(m) \in F_m \quad \text{per ogni } m.$$

Una condizione necessaria affinché ciò sia possibile è che ogni sottogruppo di k ragazzi conosca cumulativamente almeno k ragazze, ossia se per ogni $B \subseteq M$ esistono applicazioni iniettive $B \to \bigcup_{m \in B} F_m$. Dimostrare che tale condizione è anche sufficiente.

2.4. Il teorema fondamentale dell'aritmetica

In questa sezione applicheremo i principi del minimo intero e di induzione matematica per dimostrare in maniera rigorosa alcuni fatti ben noti sulla fattorizzazione dei numeri interi e naturali.

LEMMA 2.4.1 (divisione Euclidea). *Siano m, n numeri interi con $n > 0$. Allora esistono $q, r \in \mathbb{Z}$ tali che*

$$m = qn + r, \qquad 0 \leq r < n.$$

DIMOSTRAZIONE. Siccome $n > 0$ esiste almeno un intero t tale che $m + tn \geq 0$. Tra tutti gli interi non negativi del tipo $m + tn$ indichiamo con r il più piccolo e sia $q \in \mathbb{Z}$ l'unico intero tale che $r = m - qn$. Per costruzione $r \geq 0$ e per concludere basta dimostrare che $r < n$. Se fosse $r \geq n$ allora $r - n = m - (q + 1)n \geq 0$ in contraddizione con la scelta di r come minimo tra i numeri $m + tn$ non negativi. □

LEMMA 2.4.2. *Siano a, b, c numeri interi tali che $a|bc$. Allora esistono $a_1, a_2 \in \mathbb{Z}$ tali che $a = a_1 a_2$, $a_1|b$ e $a_2|c$.*

DIMOSTRAZIONE. A meno di cambiamenti di segno, che non influiscono sulla divisibilità, non è restrittivo supporre a, b, c non negativi. Se $b = 0$ basta considerare $a_1 = a$ e $a_2 = 1$, similmente se $c = 0$. Dunque non è restrittivo supporre $bc > 0$. Dimostriamo per induzione su n la seguente proposizione $\mathcal{A}_n$: *se $a, b, c, d > 0$ sono interi positivi tali che $a + b + c + d \leq n$ e $ad = bc$, allora esistono $a_1, a_2 \in \mathbb{N}$ tali che $a = a_1 a_2$, $a_1|b$ e $a_2|c$.* La proposizione $\mathcal{A}_1$ è vera in quanto vuota.

Supponiamo adesso $\mathcal{A}_n$ vera e dimostriamo $\mathcal{A}_{n+1}$: siano $a, b, c, d > 0$ interi positivi tali che $a + b + c + d \leq n + 1$ e $ad = bc$. Se $a = b$ basta porre $a_1 = a$ e $a_2 = 1$; altrimenti si ha una delle seguenti possibilità:

(1) $a < b$. In questo caso si ha $a(d - c) = (b - a)c$ e poiché $a + (b - a) + c + (d - c) \leq n$ per l'ipotesi induttiva esiste una fattorizzazione $a = a_1 a_2$ tale che $a_1|b - a$ e $a_2|c$; se $b - a = \delta a_1$, allora $b = (\delta + a_2)a_1$ e quindi $a_1|b$.
(2) $a > b$. Scambiando a con b e c con d, il ragionamento precedente mostra che si ha una fattorizzazione $b = b_1 b_2$ con $b_1|a$ e $b_2|d$. Ponendo $a_1 = b_1$ e $a_2 = a/b_1$ si ha

$$\frac{c}{a_2} = \frac{ad}{ba_2} = \frac{a_1 d}{b} = \frac{d}{b_2} \in \mathbb{N}.$$

□

Terminiamo la sezione dimostrando un'importante e ben nota proprietà dei numeri primi che è alla base del principio di fattorizzazione unica. Ricordiamo che per **primo positivo** si intende un numero naturale $p \geq 2$ che è divisibile, su $\mathbb{Z}$, solo per 1 e per se stesso.

LEMMA 2.4.3. *Se p è un primo positivo che divide un prodotto finito $b_1 b_2 \cdots b_s$ di interi positivi, allora p divide almeno uno dei fattori $b_1, \ldots, b_s$.*

DIMOSTRAZIONE. Induzione su s, con il caso $s = 1$ vero in maniera tautologica. Se $s > 1$, per il Lemma 2.4.2 possiamo scrivere $p = a_1 a_2$ con $a_1|b_1$ e $a_2|b_2 \cdots b_s$. Per ipotesi p è primo e quindi $a_1 = p$ oppure $a_2 = p$: nel primo caso $p|b_1$, nel secondo $p|b_2 \cdots b_s$ e per l'ipotesi induttiva p divide almeno uno dei fattori $b_2, \ldots, b_s$. □

TEOREMA 2.4.4 (Teorema fondamentale dell'aritmetica). *Ogni intero $c > 1$ si fattorizza in maniera unica come prodotto di una successione finita non decrescente di primi positivi.*

DIMOSTRAZIONE. Sia X l'insieme dei numeri interi $c > 1$ che non si possono scrivere come prodotto di numeri primi e supponiamo per assurdo $X \neq \emptyset$; indichiamo con $n \in X$ il più piccolo di loro. Se n è primo abbiamo finito, altrimenti si ha $n = ab$ con $0 < a < n$, $0 < b < n$. Dunque a, b non appartengono ad X e possono essere scritti ciascuno come prodotto di primi

$$a = p_1 \cdots p_i, \qquad b = q_1 \cdots q_j.$$

Ma allora anche $n = ab = p_1 \cdots p_i q_1 \cdots q_j$ è un prodotto di numeri primi. Ovviamente se $c = p_1 \cdots p_s$ è una fattorizzazione come prodotto di numeri primi, a meno di scambiare l'ordine dei vari fattori possiamo sempre supporre $p_1 \leq p_2 \leq p_3 \leq \cdots \leq p_s$.

Per quanto riguarda l'unicità, dimostriamo per induzione su s che se $c = p_1 \cdots p_s = q_1 \cdots q_t$ con $2 \le p_1 \le \cdots \le p_s$ e $2 \le q_1 \le \ldots \le q_t$ numeri primi, allora $s = t$ e $p_i = q_i$ per ogni indice i. Se $s = 1$ allora $c = p_1$ è primo e quindi vale $c = p_1 = q_1 \cdots q_t$ se e solo se $t = 1$ e $q_1 = p_1$; Se $s > 0$, per il Lemma 2.4.2 il primo p_1 divide uno dei fattori $q_1, \ldots, q_t$, e siccome ogni q_i è primo si ha $p_1 = q_j$ per qualche j. Per lo stesso motivo $q_1 = p_h$ per qualche h e quindi $p_h = q_1 \le q_j = p_1$ da cui segue $q_1 = p_h = p_1$. Dividendo c per $p_1 = q_1$ si ottiene l'uguaglianza $p_2 \cdots p_s = q_2 \cdots q_t$ e per l'ipotesi induttiva $s = t$ e $q_i = p_i$ per ogni i. □

Raggruppando i fattori primi uguali tra loro si ha quindi che ogni intero $c > 1$ si può scrivere nella forma

$$c = p_1^{a_1} \cdots p_s^{a_s}, \qquad a_1 > 0, \ldots, a_s > 0, \quad p_i \neq p_j \text{ per } i \neq j.$$

Dati due interi a, b non entrambi nulli, l'insieme $CD(a, b)$ degli interi positivi che dividono sia a che b è non vuoto (contiene 1) ed è limitato; il massimo di $CD(a, b)$ viene detto **massimo comune divisore** di a e b e viene indicato con $\mathrm{MCD}(a, b)$: dunque, due numeri hanno un fattore comune se e solo se il loro massimo comune divisore è maggiore di 1.

Esercizi.

73. Si trovino quoziente q e resto r della divisione Euclidea di -630 per 36, ossia si trovino gli interi q, r tali che

$$-630 = 36q + r, \qquad 0 \le r < 36.$$

74. Sia $d \ge 1$ il massimo comune divisore di due interi positivi m e n. Mostrare che esiste una fattorizzazione $d = ab$ tale che i due interi m/a e n/b non hanno fattori comuni.

75. Sia $s \colon \mathbb{N} \to \mathbb{N}$ l'applicazione che ad ogni intero positivo associa la somma delle cifre della sua rappresentazione decimale, ossia $s(1) = 1$, $s(13) = 4$, $s(308) = 11$ eccetera. Dimostrare per induzione su n che $n - s(n)$ è divisibile per 9, o equivalentemente che la divisione per 9 di n e $s(n)$ produce lo stesso resto.

76. Trovare tutte le terne (x, y, z) di numeri naturali tali che

$$x + y + z = \frac{x}{20} + y + 5z = 100\,.$$

77. Siano a, b, c interi positivi. Dimostrare che

(1) $\mathrm{MCD}(a, b) \cdot \mathrm{MCD}(a, c) \ge \mathrm{MCD}(a, bc)$,
(2) $\mathrm{MCD}(ab, ac) = a\,\mathrm{MCD}(b, c)$,
(3) se a divide b ed a divide c, allora a divide anche $\mathrm{MCD}(b, c)$.

78. Diremo che un sottoinsieme $H \subseteq \mathbb{N}$ è un semigruppo se per ogni $a, b \in H$ vale $a + b \in H$. Ad esempio l'insieme di tutti i numeri maggiori di 17 e l'insieme di tutti i numeri pari positivi sono semigruppi, mentre l'insieme di tutti gli interi dispari positivi non è un semigruppo. Dato un semigruppo $H \subseteq \mathbb{N}$, dimostrare che:

(1) il complementare $\mathbb{N} - H$ è finito se e solo se H contiene due interi consecutivi;
(2) (☕) supponiamo che $0 \in H$ e che $\mathbb{N} - H = \{n_1, \ldots, n_g\}$ sia formato da un numero finito g di elementi, ordinati in maniera strettamente crescente, ossia $0 < n_1 < n_2 < \cdots < n_g$. Provare che $n_i \le 2i - 1$ per ogni i e vale la formula

$$\frac{g(g+1)}{2} \le n_1 + n_2 + \cdots + n_g \le g^2.$$

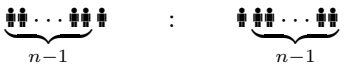
(suggerimento: se $n_i \ge 2i$ considerare le coppie $(j, n_i - j)$, $j = 1, \ldots, i$.)

2.5. Attenti all'errore!

Il ragionamento matematico si presta all'errore ed un ragionamento errato può portare a conseguenze false, anche se si parte da premesse vere. Vedremo adesso alcuni esempi di ragionamenti errati che portano a conclusioni paradossali, e nei quali si capisce subito che qualcosa non va. Di solito però le conseguenze di un ragionamento errato sono molto verosimili e a prima vista non hanno nulla di sbagliato. Per questo è utile imparare a riconoscere la correttezza di una dimostrazione a prescindere dal risultato finale.[4] Nel linguaggio dotto ed accademico una falsa dimostrazione viene chiamata *pseudodimostrazione*.

Esempio 2.5.1. Diamo una pseudodimostrazione del fatto che tutti i bambini hanno gli occhi dello stesso colore. I bambini sono in numero finito che però non conosciamo; indichiamo tale numero con n e se dimostriamo che per ogni n vale l'affermazione $\mathcal{A}_n$: *dato un insieme di n bambini, hanno tutti gli occhi dello stesso colore*, allora abbiamo dimostrato quello che vogliamo.

La proposizione $\mathcal{A}_1$ è certamente vera, un solo bambino ha gli occhi di un solo colore. Prendiamo adesso n bambini, mettiamoli in riga e prendiamo i primi $n-1$; per l'ipotesi induttiva hanno gli occhi dello stesso colore.



Lo stesso si applica se prendiamo gli ultimi $n-1$ e di conseguenza tutti hanno lo stesso colore degli occhi.

L'errore consiste chiaramente nel fatto che la dimostrazione di $\mathcal{A}_{n-1} \implies \mathcal{A}_n$ che abbiamo dato funziona solo per $n > 2$, rimane quindi non dimostrato che $\mathcal{A}_1 \implies \mathcal{A}_2$.

Esempio 2.5.2. Utilizziamo il principio di induzione per pseudodimostrare che $5 = 8$; dati due interi positivi n, m indichiamo con $\max(n, m)$ il più grande dei due: ad esempio $\max(2, 3) = 3$. Per ogni $n > 0$ indichiamo con $\mathcal{A}_n$ l'affermazione: *se vale* $\max(a, b) = n$ *allora* $a = b$.

La $\mathcal{A}_1$ è certamente vera, infatti vale $\max(a, b) = 1$ se e solo se $a = b = 1$. Supponiamo adesso che $\mathcal{A}_{n-1}$ sia vera e dimostriamo che vale anche $\mathcal{A}_n$: supponiamo che si abbia $\max(a, b) = n$, allora $\max(a-1, b-1) = n-1$ e, siccome abbiamo assunto $\mathcal{A}_{n-1}$ vera si ha $a - 1 = b - 1$ e quindi $a = b$.

Per il principo di induzione $\mathcal{A}_n$ è vero per ogni n, anche $\mathcal{A}_8$ è vero e quindi siccome $\max(5, 8) = 8$ si ha $5 = 8$.

L'errore fatto nel ragionamento è un esempio di *dicto simpliciter*.[5] Per dimostrare $\mathcal{A}_1$ abbiamo implicitamente assunto che a, b fossero entrambi maggiori di 0 e tale restrizione può impedire le sottrazioni fatte nella dimostrazione di $\mathcal{A}_{n-1} \implies \mathcal{A}_n$.

Esempio 2.5.3 (☕, Gli isolani dagli occhi blu). Questo celebre rompicapo logico possiede due soluzioni apparentemente corrette ma palesemente contraddittorie tra di loro. È talmente bello che sarebbe uno spoileraggio criminale svelare quale delle due è invece errata: lo riporto nella versione esposta nel blog[6] di Terence Tao, rispetto al quale non potrei fare meglio.

There is an island upon which a tribe resides. The tribe consists of 1000 people, with various eye colours. Yet, their religion forbids them to know their own eye color, or even to discuss the topic; thus, each resident can (and does) see the eye colors of all other residents, but has no way of discovering his or her own (there are no reflective surfaces). If a tribesperson does discover his or her own eye color, then their religion compels them to commit ritual suicide at noon the following day in the village square for all to witness. All the tribespeople are highly logical and devout, and they all know that each other is also highly logical and

[4] Le dimostrazioni, in quanto opere dell'intelletto umano, si dividono in tre categorie: quelle corrette, quelle sbagliate e quelle che non sono nemmeno sbagliate. Una dimostrazione non è nemmeno sbagliata quando è vuota o talmente illeggibile e lacunosa da non consentire una sua verifica da parte di chi la legge. Se una dimostrazione sbagliata può contenere comunque idee interessanti ed originali, quella che non è nemmeno sbagliata è solo da buttare nel cassonetto!

[5] Applicazione di una regola generale ad una situazione particolare in condizioni che rendono quella regola inapplicabile.

[6] terrytao.wordpress.com

devout (and they all know that they all know that each other is highly logical and devout, and so forth; "highly logical" means that any conclusion that can logically deduced from the information and observations available to an islander, will automatically be known to that islander.

Of the 1000 islanders, it turns out that 100 of them have blue eyes and 900 of them have brown eyes, although the islanders are not initially aware of these statistics (each of them can of course only see 999 of the 1000 tribespeople). One day, a blue-eyed foreigner visits to the island and wins the complete trust of the tribe. One evening, he addresses the entire tribe to thank them for their hospitality. However, not knowing the customs, the foreigner makes the mistake of mentioning eye color in his address, remarking "how unusual it is to see another blue-eyed person like myself in this region of the world".

What effect, if anything, does this faux pas have on the tribe?

Argument 1. The foreigner has no effect, because his comments do not tell the tribe anything that they do not already know (everyone in the tribe can already see that there are several blue-eyed people in their tribe).

Argument 2. 100 days after the address, all the blue eyed people commit suicide. This is proven as a special case of the following proposition.

Suppose that the tribe had n blue-eyed people for some positive integer n. Then n days after the traveller's address, all n blue-eyed people commit suicide.

Proof: We induct on n. When $n = 1$, the single blue-eyed person realizes that the traveler is referring to him or her, and thus commits suicide on the next day. Now suppose inductively that n is larger than 1. Each blue-eyed person will reason as follows: "If I am not blue-eyed, then there will only be $n-1$ blue-eyed people on this island, and so they will all commit suicide $n-1$ days after the traveler's address". But when $n-1$ days pass, none of the blue-eyed people do so (because at that stage they have no evidence that they themselves are blue-eyed). After nobody commits suicide on the $(n-1)$-th day, each of the blue eyed people then realizes that they themselves must have blue eyes, and will then commit suicide on the n-th day.

Which argument is valid?

Esercizi.

79. È vero o falso che se $A \subseteq \mathbb{Z}$ è un sottoinsieme con la proprietà che $a+1 \in A$ e $a-1 \in A$ per ogni $a \in A$, allora $A = \mathbb{Z}$?

80 (♡). Nel famoso detto *Morto un Papa se ne fa un altro* si nasconde un chiaro esempio di dicto simpliciter. Sapete scovarlo?

81. Anche se del tutto controintuitivo, talvolta per dimostrare un teorema T risulta più semplice dimostrare un risultato più forte del quale T è un corollario molto particolare. Questo è particolarmente vero nelle dimostrazioni per induzione, in quanto l'enunciato del teorema ha un ruolo fondamentale nella dimostrazione. Ad esempio, se $S_n = 1 + 4 + \cdots + n^2$ indica la somma dei quadrati dei primi n interi positivi, allora il teorema: *Il numero $n+1$ divide $6S_n$ per ogni n,* è al tempo stesso più debole e più difficile da dimostrare del teorema *Per ogni intero positivo n si ha $6S_n = n(n+1)(2n+1)$.*

2.6. Fattoriali e coefficienti binomiali

Dati due insiemi A, B si denota con B^A l'insieme di tutte le applicazioni $f\colon A \to B$. La notazione è motivata dal fatto che se A e B sono insiemi finiti, con k e n elementi rispettivamente, allora B^A possiede n^k elementi: infatti, per definire un'applicazione $f\colon A \to B$ possiamo prendere per ciascuno dei k elementi di A un qualsiasi elemento di B, per il quale abbiamo n scelte possibili.

Dati due interi non negativi k, n indichiamo con:

- $D_{k,n}$ il numero di applicazioni iniettive $f\colon \{1,\dots,k\} \to \{1,\dots,n\}$;
- $W_{k,n}$ il numero di applicazioni surgettiva $f\colon \{1,\dots,k\} \to \{1,\dots,n\}$.

I numeri $D_{k,n}$ si calcolano molto facilmente: infatti per definire un'applicazione $f\colon \{1,\dots,k\} \to \{1,\dots,n\}$ iniettiva possiamo prendere:

☞ $f(1)$ un qualsiasi numero compreso tra 1 e n: (n scelte),

- ☞ $f(2)$ un qualsiasi numero compreso tra 1 e n diverso da $f(1)$: $(n-1$ scelte),

⋮

- ☞ $f(k)$ un qualsiasi numero compreso tra 1 e n diverso da $f(1)$, $f(2)$, …, $f(k-1)$: $(n-k+1$ scelte).

Otteniamo quindi

$$D_{k,n} = n(n-1)(n-2)\cdots(n-k+1).$$

In particolare per $k = n$ otteniamo che $D_{n,n}$ è uguale al **fattoriale** di n:

$$D_{n,n} = n! = n(n-1)\cdots 2\cdot 1.$$

Si ha inoltre $D_{0,n} = 1$ per ogni $n \geq 0$ (esiste un'unica applicazione dall'insieme vuoto in un qualsiasi altro insieme) e $D_{k,n} = 0$ per ogni $k > n$. In particolare $0! = D_{0,0} = 1$ e quindi, se $0 \leq k \leq n$ allora vale la formula $D_{k,n} = \dfrac{n!}{(n-k)!}$.

Il calcolo dei numeri $W_{k,n}$ è decisamente più complicato. Ovviamente si ha $W_{0,0} = 1$, $W_{k,n} = 0$ se $k < n$, $W_{k,0} = 0$ se $k > 0$. In generale i numeri $W_{k,n}$ si possono calcolare in maniera ricorsiva mediante le formule:

$$W_{k,n} = n(W_{k-1,n} + W_{k-1,n-1}), \qquad k, n > 0. \tag{2.1}$$

Per dimostrare le formule (2.1) scriviamo

$$W_{k,n} = W'_{k,n} + W''_{k,n},$$

dove con $W'_{k,n}$ indichiamo il numero di applicazioni la cui restrizione $\{1,\dots,k-1\} \to \{1,\dots,n\}$ è surgettiva, e con $W''_{k,1}$ il numero di applicazioni surgettive la cui restrizione $\{1,\dots,k-1\} \to \{1,\dots,n\}$ non è surgettiva. Nel primo caso $f(k)$ può assumere qualunque valore e quindi $W'_{k,n} = nW_{k-1,n}$. Nel secondo caso, la restrizione di f a $\{1,\dots,k-1\}$ è un'applicazione surgettiva a valori in $\{1,\dots,n\} - \{f(k)\}$ e quindi $W''_{k,n} = nW_{k-1,n-1}$. Un'altra formula per il calcolo dei numeri $W_{k,n}$ sarà proposta nell'Esercizio 90.

Notiamo che se A è un insieme finito, allora un'applicazione $f\colon A \to A$ è iniettiva se e solo se è surgettiva e quindi $W_{n,n} = D_{n,n} = n!$ per ogni $n \geq 0$.

Dati due interi k, n, con $0 \leq k \leq n$, indichiamo con $\binom{n}{k}$ il numero di sottoinsiemi distinti di $\{1,\dots,n\}$ formati da k elementi. Siccome ogni applicazione iniettiva $f\colon \{1,\dots,k\} \to \{1,\dots,n\}$ è univocamente determinata da un sottoinsieme $A \subseteq \{1,\dots,n\}$ formato da k elementi (l'immagine di f) e da un'applicazione bigettiva $\{1,\dots,k\} \to A$ abbiamo la formula

$$D_{k,n} = \binom{n}{k} k!$$

da cui segue

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{1}{k!} n(n-1)\cdots(n-k+1)\,. \tag{2.2}$$

Lemma 2.6.1. *I numeri definiti in (2.2) soddisfano le seguenti uguaglianze:*

1. *per ogni $n \geq 0$ vale $\binom{n}{0} = \binom{n}{n} = 1$;*
2. *per ogni $0 \leq k \leq n$ si ha $\binom{n}{k} = \binom{n}{n-k}$;*
3. *per ogni $0 < k < n$ si ha*

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

Dimostrazione. I primi due punti seguono immediatamente dalla formula (2.2). Dalla stessa formula si ricava che

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n}{n-k}\,\frac{(n-1)!}{k!(n-1-k)!} = \frac{n}{n-k}\binom{n-1}{k},$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n}{k}\,\frac{(n-1)!}{(k-1)!(n-k)!} = \frac{n}{k}\binom{n-1}{k-1},$$

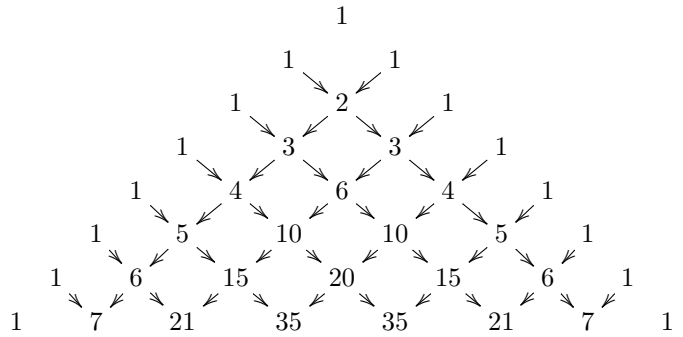


FIGURA 2.1. Nel triangolo di Tartaglia (di Pascal per i francesi) ogni numero è la somma dei due che lo sovrastano ed il coefficiente $\binom{n}{k}$ occupa la $k+1$-esima posizione nella $n+1$-esima riga.

e quindi

$$\binom{n-1}{k-1}+\binom{n-1}{k}=\frac{k}{n}\binom{n}{k}+\frac{n-k}{n}\binom{n}{k}=\left(\frac{k}{n}+\frac{n-k}{n}\right)\binom{n}{k}=\binom{n}{k}.$$

□

OSSERVAZIONE 2.6.2. L'ultima formula del Lemma 2.6.1 ha una chiara interpretazione combinatoria. Infatti i sottoinsiemi di $\{1,\dots,n\}$ di k elementi si dividono in due classi tra loro disgiunte: la classe dei sottoinsiemi che contengono n e la classe dei sottoinsiemi che non contengono n. La seconda classe coincide con la classe dei sottoinsiemi di $\{1,\dots,n-1\}$ di k elementi, e quindi è formata da $\binom{n-1}{k}$ sottoinsiemi. Ogni sottoinsieme nella prima classe è ottenuto in maniera unica aggiungendo n ad un sottoinsieme di $\{1,\dots,n-1\}$ di $k-1$ elementi: dunque la prima classe contiene $\binom{n-1}{k-1}$ sottoinsiemi e di conseguenza

$$\binom{n}{k}=\binom{n-1}{k-1}+\binom{n-1}{k}.$$

TEOREMA 2.6.3 (Binomio di Newton). *Se scambiando l'ordine dei fattori il prodotto non cambia, ossia se $ab=ba$, allora per ogni intero positivo n vale la formula:*

$$(a+b)^n=a^n+b^n+\sum_{i=1}^{n-1}\binom{n}{i}a^{n-i}b^i.$$

Ponendo per convenzione $a^0=b^0=1$ possiamo riscrivere la formula precedente nella forma

$$(a+b)^n=\sum_{i=0}^{n}\binom{n}{i}a^{n-i}b^i.$$

Ad esempio, per $n=2,3$ il Teorema 2.6.3 si riduce alle ben note formule

$$(a+b)^2=\binom{2}{0}a^2+\binom{2}{1}ab+\binom{2}{2}b^2=a^2+2ab+b^2,$$

$$(a+b)^3=\binom{3}{0}a^3+\binom{3}{1}a^2b+\binom{3}{2}ab^2+\binom{3}{3}b^3=a^3+3a^2b+3ab^2+b^3.$$

Nell'enunciato siamo stati vaghi su cosa siano le quantità a,b; per il momento possiamo supporre, per fissare le idee, che a,b siano numeri razionali (dove l'ipotesi $ab=ba$ è automaticamente soddisfatta) tenendo presente che la stessa formula vale più in generale per numeri reali, numeri complessi, polinomi, matrici quadrate, endomorfismi lineari ed altri oggetti matematici che incontreremo nel corso di queste note.

DIMOSTRAZIONE. La formula è senz'altro vera per $n = 1$ (ed anche per $n = 2, 3$). Se $n > 1$, per induzione possiamo supporre vero lo sviluppo di Newton di $(a+b)^{n-1}$ e quindi

$$\begin{aligned}(a+b)^n &= a(a+b)^{n-1} + b(a+b)^{n-1} \\ &= \sum_{i=0}^{n-1} \binom{n-1}{i} a^{n-i} b^i + \sum_{i=0}^{n-1} \binom{n-1}{j} a^{n-1-i} b^{i+1}.\end{aligned}$$

Il coefficiente di $a^{n-i}b^i$ in $(a+b)^n$ è pertanto uguale a $\binom{n-1}{i} + \binom{n-1}{i-1}$ che, per il Lemma 2.6.1, è uguale a $\binom{n}{i}$. □

I numeri del tipo $\binom{n}{k}$ vengono detti **coefficienti binomiali**. Come prima applicazione del binomio di Newton dimostriamo il seguente risultato.

COROLLARIO 2.6.4 (Piccolo teorema di Fermat). *Sia p un primo positivo. Per ogni intero n, il numero $n^p - n$ è divisibile per p.*

DIMOSTRAZIONE. Il risultato è banalmente vero per $p = 2$ in quanto $n^2 - n = n(n-1)$ ed uno tra n e $n-1$ è un numero pari. Non è restrittivo quindi supporre p dispari. Se il risultato vale per n, allora vale anche per $-n$ in quando $(-n)^p - (-n) = -(n^p - n)$ e di conseguenza basta dimostrare il teorema per $n \geq 0$. Poiché il risultato è banalmente vero per $n = 0$ e $n = 1$, per il principio di induzione è sufficiente dimostrare che se p divide $n^p - n$, allora p divide anche $(n+1)^p - (n+1)$. Per il binomio di Newton si ha

$$(n+1)^p - (n+1) = (n^p - n) + \sum_{i=1}^{p-1} \binom{p}{i} n^i$$

e per l'ipotesi induttiva basta dimostrare che p divide ogni coefficiente binomiale $\binom{p}{i}$, con $0 < i < p$. Siccome p divide $p!$, dalla formula

$$\binom{p}{i} i!(p-i)! = p!$$

ne consegue che p divide almeno uno dei tre fattori a sinistra. Siccome $i, p - i < p$ il numero primo p non può dividere né $i!$ né $(p-i)!$ e giocoforza deve dividere il coefficiente binomiale. □

ESEMPIO 2.6.5. Tra le prime applicazioni del principio di induzione abbiamo visto la dimostrazione delle uguaglianze

$$0 + 1 + 2 + \cdots + (n-1) = \sum_{i=0}^{n-1} i = \frac{1}{2}n^2 - \frac{1}{2}n,$$

$$0^2 + 1^2 + 2^2 + \cdots + (n-1)^2 = \sum_{i=0}^{n-1} i^2 = \frac{n^3}{3} - \frac{n^2}{2} + \frac{n}{6},$$

ed in maniera simile si può dimostrare che

$$0^3 + 1^3 + 2^3 + \cdots + (n-1)^3 = \sum_{i=0}^{n-1} i^3 = \frac{n^4}{4} - \frac{n^3}{2} + \frac{n^2}{4}.$$

Non è sorprendente scoprire che formule simili valgano per esponenti maggiori di 3; nel 1636 il matematico tedesco Faulhaber, pubblicò le formule chiuse per la somma delle potenze d-esime dei primi n numeri naturali, per ogni $d \leq 17$. Come possiamo ritrovare le formule di Faulhaber, e come possiamo proseguire la sua opera per trovare, se lo desideriamo, la formula per la somma delle potenze 124-esime dei primi numeri naturali?

Anche qui ci viene in aiuto il principio di induzione. Per ogni $d, n \geq 0$ definiamo

$$g_d(n) = 0^d + 1^d + 2^d + \cdots + (n-1)^d = \sum_{i=0}^{n-1} i^d.$$

Abbiamo $g_0(n) = n$ (si pone per convenzione $0^0 = 1$), $g_1(n) = n(n-1)/2$ eccetera. Dimostriamo per induzione su n che, per ogni intero $d \geq 0$ si ha

$$\sum_{r=0}^{d} \binom{d+1}{r} g_r(n) = n^{d+1}.$$

Supponendo vera la formula precedente si può scrivere

$$(n+1)^{d+1} = n^{d+1} + \sum_{r=0}^{d} \binom{d+1}{r} n^r = \sum_{r=0}^{d} \binom{d+1}{r} (g_r(n) + n^r) = \sum_{r=0}^{d} \binom{d+1}{r} g_r(n+1).$$

Siccome $\binom{d+1}{d} = d+1$, ricaviamo la formula

$$(d+1)g_d(n) + \sum_{r=0}^{d-1} \binom{d+1}{r} g_r(n) = n^{d+1}$$

che equivale a

$$g_d(n) = \frac{n^{d+1}}{d+1} - \frac{1}{d+1} \sum_{r=0}^{d-1} \binom{d+1}{r} g_r(n). \tag{2.3}$$

Da ciò segue, sempre per induzione, che $g_d(n)$ è un polinomio di grado $d+1$ in n a coefficienti razionali. Possiamo quindi, almeno in teoria, calcolare in maniera ricorsiva tutti i polinomi $g_d(n)$ con d grande a piacere.

Esercizi.

82. Dimostrare che ogni numero razionale a si scrive in modo unico come

$$a = a_1 + \frac{a_2}{2!} + \frac{a_3}{3!} + \cdots + \frac{a_k}{k!}$$

con $a_1, \ldots, a_k$ numeri interi tali che $0 \leq a_2 < 2$, $0 \leq a_3 < 3$, $\ldots$, $0 \leq a_k < k$.

83. Sia p un intero positivo. Provare che per ogni intero $n > 0$ valgono le diseguaglianze

$$(n+p)^n \geq (p+1)n^n, \qquad (n+p)! \geq p^{n+1}.$$

84. Usare il principio di induzione per provare che $\left(\frac{n+1}{2}\right)^{n+1} \geq (n+1)\left(\frac{n}{2}\right)^n$ per ogni $n \geq 1$ e $\left(\frac{m}{2}\right)^m \geq m!$ per ogni $m \geq 6$.

85. Usare il binomio di Newton per dimostrare che per ogni n valgono le formule

$$\sum_{i=0}^{n} (-1)^i \binom{n}{i} = 0, \qquad \sum_{i=0}^{n} (-2)^i \binom{n}{i} = (-1)^n.$$

86. Dato un intero $n > 0$, quanto vale

$$n \cdot n! + (n-1) \cdot (n-1)! + \cdots + 2 \cdot 2! + 2\,?$$

Fare il conto per piccoli valori di n, congetturare la risposta e poi dimostrarla usando il principio di induzione.

87. È utile estendere la definizione del coefficiente binomiale $\binom{n}{k}$ ad ogni coppia di numeri n, k, con k intero, ponendo $\binom{n}{k} = 0$ se $k < 0$, $\binom{n}{0} = 1$ e

$$\binom{n}{k} = \frac{1}{k!} n(n-1)\cdots(n-k+1), \qquad \text{se } k > 0.$$

Ad esempio si ha

$$\binom{-1}{3} = -1 \qquad \binom{\frac{1}{2}}{2} = -\frac{1}{8}.$$

Provare che continua a valere la formula

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$$

e dedurre che $\binom{n}{k} \in \mathbb{Z}$ per ogni $n, k \in \mathbb{Z}$.

88 (Identità della mazza da hockey). Dimostrare che per ogni $0 \leq k \leq n$ vale

$$\binom{n+1}{k+1} = \sum_{i=k}^{n} \binom{i}{k}.$$

89. Verificare che valgono le uguaglianze

$$\binom{n}{k}\binom{k}{i} = \binom{n}{i}\binom{c-i}{k-i}, \qquad 0 \leq i \leq k \leq n,$$

$$\binom{n}{k}\binom{n-k+1}{i} = \binom{n}{i+k-1}\binom{i+k-1}{k} + \binom{n}{i}\binom{n-i}{k}.$$

90. Dimostrare per induzione su k che vale la formula:

$$W_{k,n} = \sum_{r=0}^{n} (-1)^{n-r} \binom{n}{r} r^k\,, \qquad k, n \geq 0.$$

Provare inoltre per induzione su $n+k$ che $W_{k,n}$ è divisibile per $n!$. I relativi quozienti sono chiamati **numeri di Stirling di seconda specie**, sono indicati con i simboli

$$\left\{ {k \atop n} \right\} = \frac{W_{k,n}}{n!},$$

e soddisfano le relazioni

$$\left\{ {k \atop 1} \right\} = \left\{ {k \atop k} \right\} = 1, \qquad \left\{ {k \atop n} \right\} = n \left\{ {k-1 \atop n} \right\} + \left\{ {k-1 \atop n-1} \right\}.$$

91. In questo esercizio su principio di induzione e coefficienti binomiali arriveremo, tra le altre cose, a dimostrare le disuguaglianze

$$\sum_{r=0}^{n} \frac{1}{r!} \geq \left(1 + \frac{1}{n}\right)^n \geq \left(1 - \frac{1}{2n}\right) \sum_{r=0}^{n} \frac{1}{r!}\,, \qquad n \geq 1, \tag{2.4}$$

che, passando al limite, ci forniscono la ben nota identità

$$e = \sum_{r=0}^{\infty} \frac{1}{r!} = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n.$$

(1) Sia $a_1, a_2 \ldots$ una successione di numeri razionali compresi tra 0 e 1. Dimostrare per induzione su r che

$$1 \geq (1-a_1)(1-a_2)\cdots(1-a_r) \geq 1 - \sum_{i=1}^{r} a_i$$

e dedurre che per ogni $2 \leq r \leq n$ si ha

$$1 \geq \frac{r!}{n^r}\binom{n}{r} \geq 1 - \frac{0+1+\cdots+(r-1)}{n} = 1 - \frac{r(r-1)}{2n}.$$

(2) Sia $n \geq 2$ un intero fissato. Usando il punto precedente e lo sviluppo di Newton

$$\left(1 + \frac{1}{n}\right)^n = 2 + \sum_{r=2}^{n} \binom{n}{r} \frac{1}{n^r}$$

dedurre che

$$\sum_{r=0}^{n} \frac{1}{r!} \geq \left(1 + \frac{1}{n}\right)^n \geq 2 + \sum_{r=2}^{n} \frac{1}{r!}\left(1 - \frac{r(r-1)}{2n}\right) = \sum_{r=0}^{n} \frac{1}{r!} - \frac{1}{2n} \sum_{r=0}^{n-2} \frac{1}{r!}\,.$$

(3) Dimostrare (2.4) per ogni $n \geq 1$.

92 (☕☕). Sia n un intero positivo fissato. Per ogni sottoinsieme $M \subset \{1, 2, \dots, 2n\}$ tale che $|M| = n$, ossia che contiene esattamente n elementi, sia dato un numero razionale a_M. Si supponga che per ogni $N \subset \{1, 2, \dots, 2n\}$ tale che $|N| = n - 1$ si abbia

$$\sum_{N \subset M,\, |M|=n} a_M = \sum_{i \notin N} a_{N \cup \{i\}} = 0\,.$$

(Si ha quindi un sistema lineare omogeneo di $\binom{2n}{n-1}$ equazioni in $\binom{2n}{n}$ incognite.) Dimostrare che $a_{n+1,n+2,\dots,2n} = (-1)^n a_{1,2,\dots,n}$.

2.7. Il prodotto di composizione

Dati tre insiemi A, B, C e due applicazioni $f\colon A \to B$, $g\colon B \to C$ si indica con $g \circ f\colon A \to C$ la **composizione** di f e g definita da:

$$g \circ f(a) = g(f(a)), \qquad \text{per ogni } a \in A.$$

Ad esempio se $f, g\colon \mathbb{N} \to \mathbb{N}$ sono le applicazioni

$$f(n) = 2n, \qquad g(m) = m^2,$$

si ha

$$g \circ f(n) = g(2n) = (2n)^2 = 4n^2.$$

La composizione di applicazioni gode della **proprietà associativa**, ossia se $f\colon A \to B$, $g\colon B \to C$ e $h\colon C \to D$ sono applicazioni si ha

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

Infatti per ogni $a \in A$ vale:

$$(h \circ (g \circ f))(a) = h(g \circ f(a)) = h(g(f(a)) = (h \circ g)(f(a)) = ((h \circ g) \circ f)(a).$$

Non presenta quindi alcuna ambiguità scrivere $h \circ g \circ f$ intendendo con essa una qualsiasi delle due composizioni $h \circ (g \circ f)$, $(h \circ g) \circ f$, ossia

$$h \circ g \circ f(a) = h(g(f(a)).$$

Se $f\colon A \to A$ è un'applicazione di un insieme in sé, si denota con f^n la composizione di f con sé stessa n volte:

$$f^1 = f, \quad f^2 = f \circ f, \quad f^3 = f \circ f \circ f, \quad \dots$$

Per convenzione si pone f^0 uguale all'identità; notiamo che vale la formula $f^n \circ f^m = f^{n+m}$ per ogni $n, m \geq 0$.

Spesso, per alleggerire la notazione, quando ciò non crea problemi o ambiguità si scrive semplicemente gf in luogo di $g \circ f$, per cui vale $gf(a) = g(f(a))$ e l'associatività della composizione diventa $(hg)f = h(gf)$.

Esempio 2.7.1. In generale la composizione non gode della proprietà commutativa, ossia in generale $fg \neq gf$, anche quando entrambe le composizioni sono definite. Si considerino ad esempio le tre applicazioni $f, g, h\colon \mathbb{N} \to \mathbb{N}$ definite da:

$$f(n) = 2n, \qquad g(m) = m^2, \qquad h(n) = n + 1.$$

Si ha:

(1) $fg(n) = 2n^2$, $gf(n) = 4n^2$,
(2) $fh(n) = 2n + 2$, $hf(n) = 2n + 1$,
(3) $gh(n) = n^2 + 2n + 1$, $hg(n) = n^2 + 1$.

Lemma 2.7.2. *Siano $f, g\colon A \to A$ due applicazioni che commutano, ossia tali che $fg = gf$. Allora vale $f^n g^m = g^m f^n$ per ogni coppia di interi positivi n, m.*

DIMOSTRAZIONE. Prima di dimostrare il caso generale è istruttivo studiare alcuni casi particolari. Per $n = 2$ e $m = 1$, per l'associatività del prodotto si ha

$$f^2 g = f(fg) = f(gf) = (fg)f = (gf)f = gf^2.$$

Se $n > 1$ e $m = 1$ si procede im maniera simile, ossia:

$$f^n g = f^{n-1}(fg) = f^{n-1}(gf) = f^{n-2}(fg)f = f^{n-2}gf^2 = \cdots = gf^n.$$

Per $n, m > 0$ si ha

$$f^n g^m = f^n g g^{m-1} = g f^n g^{m-1} = g f^n g g^{m-2} = g^2 f^n g^{m-2} = \cdots = g^m f^n.$$

□

DEFINIZIONE 2.7.3. Siano $f\colon A \to B$ e $g\colon B \to A$ due applicazioni. Diremo che f e g sono una l'**inversa** dell'altra se

$$gf = \text{identità su } A, \qquad fg = \text{identità su } B.$$

Un'applicazione $f\colon A \to B$ si dice **invertibile** se possiede un'inversa.

TEOREMA 2.7.4. *Un'applicazione di insiemi è invertibile se e solo se è bigettiva; in tal caso l'inversa è unica.*

DIMOSTRAZIONE. Lasciata per esercizio. □

DEFINIZIONE 2.7.5. Dato un insieme non vuoto A, si definisce l'insieme delle **permutazioni** di A come

$$\Sigma_A = \{f\colon A \to A \mid f \text{ bigettiva}\} \subseteq A^A.$$

Quando $A = \{1, \dots, n\}$ si scrive $\Sigma_n = \Sigma_A$.

Se A è un insieme finito con n elementi allora Σ_A coincide con l'insieme delle applicazioni iniettive di A in A e quindi contiene esattamente $D_{n,n} = n!$ elementi. Poiché Σ_A coincide anche con l'insieme delle applicazioni surgettive si ha $W_{n,n} = n!$.

Se $f\colon A \to B$ è surgettiva, allora per ogni sottoinsieme finito $C \subseteq B$ esiste un'applicazione $g\colon C \to A$ tale che $f(g(x)) = x$ per ogni $x \in C$. Infatti, se $C = \{x_1, \dots, x_n\}$, allora per ogni $i = 1, \dots, n$ possiamo *scegliere* un qualsiasi elemento $a_i \in A$ tale che $f(a_i) = x_i$ (un tale elemento esiste per la surgettività di f) e definire l'applicazione

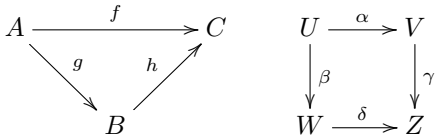
$$g\colon C \to A, \qquad g(x_i) = a_i, \qquad i = 1, \dots, n.$$

A prima vista il precedente argomento sembra valere anche se C è un insieme infinito, ma non esiste alcun modo di dimostrare tale fatto partendo da concetti più elementari. Per tale ragione introduciamo ed accettiamo come principio evidente il cosiddetto:

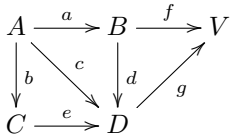
Assioma della scelta: *Per ogni applicazione surgettiva $f\colon A \to B$ e per ogni sottoinsieme $C \subseteq B$ esiste un'applicazione $g\colon C \to A$ tale che $f(g(x)) = x$ per ogni $x \in C$.*

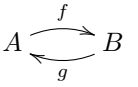
OSSERVAZIONE 2.7.6. Se A è un insieme finito, allora un'applicazione $A \to A$ è iniettiva se e soltanto se è surgettiva: questo fatto è talmente chiaro ed intuitivo che non richiede ulteriori spiegazioni. Tra le applicazioni dell'assioma della scelta vi è la prova che vale anche il viceversa, e cioè che un insieme A è finito se e solo se ogni applicazione iniettiva $A \to A$ è anche surgettiva, o equivalentemente che un insieme B è infinito se e solo se esiste un'applicazione $B \to B$ che è iniettiva ma non surgettiva (Esercizio 104). Ad esempio l'applicazione $g\colon \mathbb{N} \to \mathbb{N}$, $g(n) = n + 1$, è iniettiva ma non surgettiva.

Faremo spesso ricorso alla rappresentazione delle applicazioni tramite diagrammi: per **diagramma** di insiemi si intende una collezione di insiemi e di applicazioni tra essi. Un diagramma si dice **commutativo** quando coincidono tutte le possibili composizioni di applicazioni da un insieme ad un altro. Ad esempio, dire che i diagrammi



sono commutativi equivale a dire che $hg = f$ e $\gamma\alpha = \delta\beta$. Similmente, il diagramma



è commutativo se e solo se $c = da = eb$, $f = gd$, $fa = gda = gc = geb$. Quando nel diagramma sono presenti dei loops si richiede che la composizione delle applicazioni che formano ogni loop sia uguale all'identità; ad esempio dire che il diagramma  è commutativo significa dire che f, g sono invertibili ed una l'inversa dell'altra.

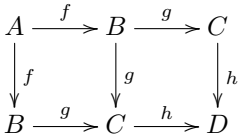
Esercizi.

93. Si consideri una successione di 3 applicazioni di insiemi

$$A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{h} D.$$

Dimostrare:

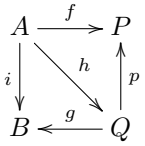
(1) il diagramma



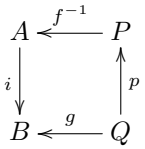
è commutativo;
(2) se gf è iniettiva, allora f è iniettiva;
(3) se gf è surgettiva, allora g è surgettiva;
(4) se gf è bigettiva e hg è iniettiva, allora f è bigettiva.

94. Siano $f\colon A \to B$ e $g\colon B \to A$ due applicazioni tali che $g(f(a)) = a$ per ogni $a \in A$. Dimostrare che f è iniettiva e che g è surgettiva.

95. Si abbia un diagramma commutativo di insiemi



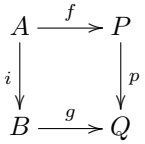
nel quale l'applicazione f è bigettiva. È vero o falso che il diagramma



è commutativo?

96. Diremo che un quadrato commutativo di applicazioni di insiemi

(2.5)



possiede la *proprietà di sollevamento* se esiste un'applicazione $h\colon B \to P$ che rende il diagramma

$$\begin{array}{ccc} A & \xrightarrow{f} & P \\ {\scriptstyle i}\downarrow & \overset{h}{\nearrow} & \downarrow{\scriptstyle p} \\ B & \xrightarrow{g} & Q \end{array}$$

commutativo, ossia tale che $hi = f$ e $ph = g$. Dimostrare che se i è iniettiva e p è surgettiva, allora (2.5) possiede la proprietà di sollevamento.

97. Descrivere un insieme A e due applicazioni $f, g\colon A \to A$ che non commutano, ma i cui quadrati commutano, ossia $f^2g^2 = g^2f^2$ ma $fg \neq gf$.

98. Sia $f\colon A \to A$ una applicazione tra insiemi. Dimostrare che se f^2 è invertibile allora anche f è invertibile.

99. Dimostrare che la classe delle applicazioni bigettive tra insiemi gode delle seguenti proprietà:

(1) (2 su 3) date due applicazioni $A \xrightarrow{f} B \xrightarrow{g} C$. Se due qualsiasi tra le tre applicazioni f, g, gf sono bigettive, allora lo è anche la terza;
(2) (2 su 6) dato un diagramma commutativo

$$\begin{array}{ccc} A & \longrightarrow & B \\ {\scriptstyle f}\downarrow & \searrow\swarrow & \downarrow{\scriptstyle g} \\ C & \longrightarrow & D \end{array}$$

di 4 insiemi e 6 applicazioni, se f e g sono bigettive, allora sono bigettive anche le rimanenti 4 applicazioni.

100. Calcolare il numero di permutazioni $\sigma \in \Sigma_n$ tali che $\sigma^2 = Id$, considerando prima i casi $n \leq 5$ e poi descrivendo una formula per il caso generale.

101 (Il problema delle triadi). Si tratta di un classico problema della prima metà del XIX secolo, all'epoca chiamato, in maniera platealmente sessista, *problema della studentessa*. Si consideri un intero $n \geq 3$ tale che il coefficiente binomiale $\binom{n}{2}$ sia divisibile per 3; si dimostra immediatamente che n deve essere del tipo $3k$ oppure $3k+1$, per un intero positivo k. Ebbene, dato un insieme X di n elementi, si chiede se è possibile trovare $m = \frac{1}{3}\binom{n}{2}$ triadi di X, ossia sottoinsiemi di 3 elementi, tali che ogni coppia di elementi di X è contenuta in una, ed una soltanto, delle m triadi. Dimostrare che:

- il problema delle triadi è risolubile per $n = 3, 7, 9$;
- il problema delle triadi non è risolubile per $n = 4, 6$.

Nel 1850, A. Cayley pubblicò una possibile soluzione al problema delle triadi per $n = 15$ (Figura 2.2).

102 (Un esercizio di astrazione, ☕). I matematici devono spesso lavorare con oggetti di cui ignorano la vera natura e dei quali conoscono solo alcune proprietà. Provate a fare anche voi qualcosa del genere risolvendo il seguente esercizio; se al momento vi rimane ostico potete tornarci in seguito, quando avrete più familiarità con la matematica astratta ed i ragionamenti inferenziali.

Odino ha diviso le applicazioni tra insiemi in buone e cattive. La vera natura di tale distinzione ci è ignota, tuttavia sappiamo che:

(1) per ogni insieme, l'identità è buona;
(2) le applicazioni buone soddisfano la regola del 2 su 6, ossia date tre applicazioni $A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{h} H$, se gf e hg sono buone, allora anche f, g, h e hgf sono buone.

Dimostrare che le applicazioni bigettive sono tutte buone e che le applicazioni buone soddisfano la regola del 2 su 3, ossia date due applicazioni $A \xrightarrow{f} B \xrightarrow{g} C$, se due qualsiasi tra le tre applicazioni f, g, gf sono buone, allora lo è anche la terza.

	a	b	c	d	e	f	g
abc				35	17	82	64
ade		62	84			15	37
afg		13	57	86	42		
bdf	47		16		38		25
bge	58		23	14		67	
cdg	12	78			56	34	
cef	36	45		27			18

FIGURA 2.2. La soluzione di Cayley al problema delle 35 triadi (Esercizio 101) per l'insieme $X = \{a, b, c, d, e, f, g, 1, 2, 3, 4, 5, 6, 7, 8\}$.

103 (Le identità semicosimpliciali, ☕). Per ogni intero non negativo k consideriamo l'applicazione

$$\delta_k \colon \mathbb{Z} \to \mathbb{Z}, \qquad \delta_k(p) = \begin{cases} p & \text{se } p < k, \\ p+1 & \text{se } p \geq k. \end{cases}$$

Provare che per ogni $0 \leq l \leq k$ vale la formula:

$$\delta_l \delta_k = \delta_{k+1} \delta_l.$$

Mostrare inoltre che un'applicazione $f \colon \mathbb{Z} \to \mathbb{Z}$ si può scrivere come composizione di un numero finito n di applicazioni δ_k se e solo se soddisfa le seguenti condizioni:

(1) $f(a) = a$ per ogni $a < 0$;
(2) $f(a+1) > f(a)$ per ogni $a \in \mathbb{Z}$;
(3) esiste $N \in \mathbb{N}$ tale che $f(a) = a + n$ per ogni $a \geq N$.

104 (☕). Sia X un insieme infinito; indichiamo con $\mathcal{F} \subseteq \mathcal{P}(X)$ la famiglia di tutti i sottoinsiemi finiti di X e con $\mathcal{A} \subset \mathcal{F} \times X$ l'insieme delle coppie (A, x) tali che $x \notin A$. Provare che:

(1) l'applicazione $f \colon \mathcal{A} \to \mathcal{F}$, $f(A, x) = A$, è surgettiva;
(2) esiste un'applicazione $g \colon \mathcal{F} \to X$ tale che $g(A) \notin A$ per ogni sottoinsieme finito A di X;
(3) l'applicazione $h \colon \mathbb{N} \to X$, definita in maniera ricorsiva dalla formula

$$h(n) = g(\{h(1), h(2), \ldots, h(n-1)\})$$

è iniettiva;
(4) esiste un'applicazione $k \colon X \to X$ tale che $k(h(n)) = h(n+1)$ per ogni $n \in \mathbb{N}$ e $k(x) = x$ per ogni $x \notin h(\mathbb{N})$. Tale applicazione è iniettiva ma non surgettiva.

2.8. Complementi: i numeri di Fibonacci e di Bernoulli

2.8.1. Numeri di Fibonacci. In risposta ad un problema pratico di coniglicoltura, Leonardo Pisano (1170-1250), figlio di Bonaccio (Fibonacci), scrive nel suo Liber Abaci la successione

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377.$$

In linguaggio moderno si definisce la successione dei **numeri di Fibonacci** tramite la formula ricorsiva

$$F_0 = 0, \quad F_1 = 1, \quad F_{n+1} = F_n + F_{n-1} \quad n \geq 1.$$

I numeri di Fibonacci hanno alcune interessanti proprietà che li rendono interessanti nelle scienze ma anche nelle pseudoscienze: non è infrequente vedere pubblicizzati metodi per vincere alla roulette basati sui numeri di Fibonacci, venduti a caro prezzo da sedicenti esperti al credulone di turno. Ovviamente ci concentreremo qui solo sugli aspetti matematici rimandando ai corsi della Iutubb University per quelli esoterici.

Vogliamo contare quante sono, per ogni intero positivo n, le successioni finite $a_1, \dots, a_k$ di numeri interi tali che

$$2 \le a_1, \quad a_k \le n, \quad a_{i+1} \ge a_i + 2 \text{ per ogni } i.$$

Indichiamo con u_n tale numero. Abbiamo $u_1 = 1$ (la successione vuota), $u_2 = 2$ (la successione vuota e la successione $a_1 = 2$) eccetera.

Una formulazione equivalente del problema è: date $n-1$ forchette in tavola disposte parallelamente tra loro, si vuole contare in quanti modi è possibile sostituire alcune forchette con dei coltelli in modo che non vi siano due coltelli adiacenti.

Possiamo scrivere $u_n = a + b$, dove a è il numero di successioni $a_1, \dots, a_k$ con $a_k < n$ e b è è il numero di successioni $a_1, \dots, a_k$ con $a_k = n$ (e di conseguenza $a_{k-1} \le n-2$. È chiaro che $a = u_{n-1}$ e $b = u_{n-2}$ e quindi, per ogni $n \ge 3$ vale $u_n = u_{n-1} + u_{n-2}$. Siccome $u_1 = F_2$ e $u_2 = F_3$ se ne ricava immediatamente $u_n = F_{n+1}$.

TEOREMA 2.8.1. *Per ogni $a, n \ge 1$ vale la formula*

$$F_n F_a + F_{n-1} F_{a-1} = F_{n+a-1}.$$

DIMOSTRAZIONE. Per $a = 1, 2$ la formula è vera. Si procede per induzione su a (esercizio). □

Nel seguente corollario indicheremo con $\mathrm{MCD}(x, y)$ il massimo comune divisore di due interi positivi x, y.

COROLLARIO 2.8.2. *Per ogni $n, m \ge 1$ vale $\mathrm{MCD}(F_n, F_m) = F_{\mathrm{MCD}(n,m)}$. In particolare, se $n > 4$ e F_n è primo, allora anche n è primo.*

DIMOSTRAZIONE. Mostriamo inizialmente per induzione su n che F_n e F_{n+1} non hanno fattori comuni. Se $p > 0$ divide sia F_n che F_{n+1}, allora divide anche $F_{n+1} - F_n = F_{n-1}$.

Dati due numeri interi positivi a, b, denotiamo con $CD(a, b)$ l'insieme dei divisori comuni di a e b, ossia l'insieme dei numeri $s \ge 1$ che dividono a e b. Per definizione, $\mathrm{MCD}(a, b)$ è il massimo dell'insieme $CD(a, b)$. Notiamo che, se $a > b$, allora

$$CD(a, b) = CD(a - b, b)$$

e di conseguenza $\mathrm{MCD}(a, b) = \mathrm{MCD}(a - b, b)$.

Dimostriamo il corollario per induzione su $n + m$. Se $n = m$ il risultato è banale. Supponiamo dunque $n \ne m$ e, per fissare le idee $n > m$. Ponendo $a = n - m$, siccome $\mathrm{MCD}(n, m) = \mathrm{MCD}(a, m)$, per induzione basta dimostrare che

$$\mathrm{MCD}(F_n, F_m) = \mathrm{MCD}(F_a, F_m).$$

Abbiamo dimostrato che vale

$$F_n = F_{m+1} F_a + F_m F_{a-1}.$$

Ogni divisore di F_n e F_m divide anche $F_{m+1} F_a$ e quindi

$$CD(F_n, F_m) \subseteq CD(F_{m+1} F_a, F_m).$$

Siccome ogni divisore di F_m non ha fattori comuni con F_{m+1}, se segue che

$$CD(F_{m+1} F_a, F_m) \subseteq CD(F_a, F_m).$$

Viceversa, sempre dalla formula $F_n = F_{m+1} F_a + F_m F_{a-1}$ segue che se un numero divide F_a e F_m, allora divide anche F_n e quindi

$$CD(F_a, F_m) \subseteq CD(F_n, F_m).$$

Dunque $CD(F_a, F_m) = CD(F_n, F_m)$.

Supponiamo $n > 4$, F_n primo e supponiamo per assurdo che $n = ab$ con $a > 2$ e $b \ge 2$; allora $1 = \mathrm{MCD}(F_n, F_a) = F_{\mathrm{MCD}(n,a)} = F_a > 1$ che è assurdo.

□

Viceversa, se $n > 2$ è primo, allora F_n può non essere primo; ad esempio $F_{19} = 4181 = 37 \times 113$ e $F_{53} = 53316291173 = 953 \times 55945741$.

COROLLARIO 2.8.3. *Sia $p_1 = 2, p_2 = 3, \dots$ la successione dei numeri primi. Allora per ogni $k \ge 4$ vale $p_{k+1} \le F_{p_k}$.*

DIMOSTRAZIONE. Dalla relazione $\mathrm{MCD}(F_n, 3) = \mathrm{MCD}(F_n, F_4) = F_{\mathrm{MCD}(n,4)}$ segue che se F_n è un multiplo di 3 allora n è un multiplo di 4. Sia $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, $p_4 = 7$, $p_5 = 11$, $p_6 = 13$, $p_7 = 17$, $p_8 = 19$ ecc. la successione dei numeri primi. Notiamo che $p_{i+1} \le F_{p_i}$ per $i = 4, 5, 6, 7$. Se $k \ge 8$, abbiamo visto che i k numeri di Fibonacci

$$F_{p_2}, F_{p_1^2}, F_{p_3}, \ldots, F_{19}, \ldots, F_{p_k}$$

sono tutti maggiori di 1, sono relativamente primi tra loro per il Corollario 2.8.2, e compaiono almeno $k+1$ numeri primi nelle loro scomposizioni dato che F_{19} non è primo. Ne segue che esistono due indici $j > k$ e $a \le k$ tali che p_j divide F_{p_a} e quindi, a maggior ragione $p_{k+1} \le F_{p_k}$. □

Notiamo incidentalmente che la dimostrazione del precedente corollario può essere usata come dimostrazione alternativa che la successione dei numeri primi è infinita (se contiene $k \ge 8$ elementi allora ne contiene almeno $k+1$).

2.8.2. Numeri di Bernoulli. Un'altra successione di numeri che gode di grande popolarità è quella dei numeri di Bernoulli $B_0, B_1, B_2, \ldots$. Si tratta di una successione di numeri razionali definita in maniera ricorsiva dalle equazioni

$$B_0 = 1, \qquad \sum_{i=0}^{n} \binom{n+1}{i} B_i = 0, \quad n > 0, \tag{2.6}$$

oppure, in maniera del tutto equivalente, dalle equazioni

$$B_0 = 1, \qquad B_n = \frac{-1}{n+1} \sum_{i=0}^{n-1} \binom{n+1}{i} B_i, \quad n > 0.$$

Abbiamo quindi:

$$B_1 = \frac{-1}{1+1}\binom{2}{0}B_0 = -\frac{1}{2}, \quad B_2 = \frac{-1}{2+1}\left(\binom{3}{0}B_0 + \binom{3}{1}B_1\right) = \frac{-1}{3}\left(1 - \frac{3}{2}\right) = \frac{1}{6},$$

$$B_3 = \frac{-1}{4}\left(\binom{4}{0}B_0 + \binom{4}{1}B_1 + \binom{4}{2}B_2\right) = 0, \qquad B_4 = -\frac{1}{30},$$

e così via. È facile trovare in rete la successione dei primi 498 numeri di Bernoulli. I primi 13 (da B_0 a B_{12}) sono:

$$1,\ -\frac{1}{2},\ \frac{1}{6},\ 0,\ -\frac{1}{30},\ 0,\ \frac{1}{42},\ 0,\ -\frac{1}{30},\ 0,\ \frac{5}{66},\ 0,\ -\frac{691}{2730}\,.$$

I numeri di Bernoulli trovano vasta applicazione in moltissimi settori della matematica. Storicamente il loro primo uso concerne lo studio dei polinomi $g_d(n)$ introdotti nell'Esempio 2.6.5, ad opera di Jakob Bernoulli nel 1713, riassunto nel seguente teorema.

TEOREMA 2.8.4. *Per ogni coppia di interi positivi d, n si ha:*

$$g_d(n) = \sum_{i=0}^{n-1} i^d = \sum_{s=0}^{d} \binom{d}{s} B_{d-s} \frac{n^{s+1}}{s+1}.$$

DIMOSTRAZIONE. Diamo solamente una traccia della dimostrazione, lasciando come esercizio (☕) per il lettore il compito di completare il ragionamento. Usando la prima formula dell'Esercizio 89 si dimostra che per ogni coppia di interi non negativi n, k si ha

$$\frac{1}{n+1} \sum_{i=0}^{n} \binom{n+1}{i}\binom{i}{k} B_{i-k} = \begin{cases} 1 & \text{se } n = k, \\ 0 & \text{se } n \ne k\,. \end{cases} \tag{2.7}$$

Usando le Formule (2.3) e (2.7) si prova per induzione che per ogni intero positivo d vale

$$\sum_{i=0}^{n-1} i^d = g_d(n) = \frac{1}{d+1} \sum_{k=0}^{d} \binom{d+1}{k} B_k n^{d-k+1} = \sum_{s=0}^{d} \binom{d}{s} B_{d-s} \frac{n^{s+1}}{s+1}\,.$$

□

Esercizi.

105. La definizione dei numeri F_n ha senso anche per n negativo, ad esempio

$$F_{-1} = F_1 - F_0 = 1, \quad F_{-2} = F_0 - F_{-1} = -1, \quad F_{-3} = F_{-1} - F_{-2} = 2, \ \dots$$

Dimostrare che vale $F_{-n} + (-1)^n F_n = 0$ per ogni $n \in \mathbb{Z}$.

106. Usare il principio di induzione per dimostrare che per ogni intero positivo n vale:
(1) $F_1 + F_2 + \cdots + F_{n-1} + F_n = F_{n+2} - 1$,
(2) $F_3 + \cdots + F_{2n-3} + F_{2n-1} = F_{2n} - 1$,
(3) $F_2 + F_4 + \cdots + F_{2n} = F_{2n+1} - 1$,
(4) $F_{2n+1} - 1 = \sum_{i=0}^{n} (n-i) F_{2i+1}$,
(5) $F_{n+1}^2 - F_{n+1}F_n - F_n^2 = (-1)^n$.

Trovare una formula per il valore della sommatoria a segni alterni

$$F_2 - F_3 + F_4 - F_5 + \cdots + (-1)^n F_n.$$

107 (Vedi Figura 2.3). Usare il principio di induzione ed il risultato dell'Esercizio 87 per dimostrare che per ogni intero $n \geq 0$ vale

$$2^n = \sum_{i=0}^{n} \binom{n}{i}, \qquad F_{n+1} = \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n-i}{i},$$

dove $\lfloor \frac{n}{2} \rfloor$ indica la parte intera di $\frac{n}{2}$, ossia il più grande intero minore od uguale ad $\frac{n}{2}$.

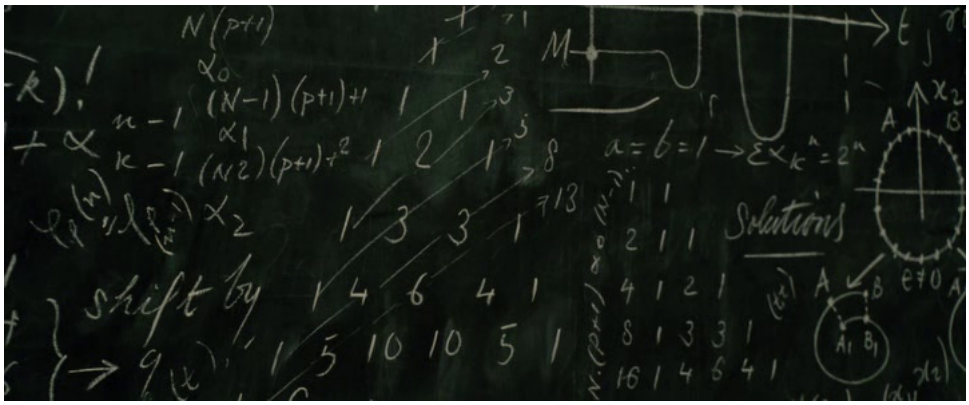


FIGURA 2.3. La lavagna del professor Moriarty (dal film "Sherlock Holmes gioco di ombre", 2011).

108. Dimostrare che per ogni intero positivo n vale:
(1) $\sum_{i=1}^{n} F_i^2 = F_n F_{n+1}$,
(2) $F_{n+1}F_{n-1} = F_n^2 + (-1)^n$,
(3) $F_n F_{n+1} - F_{n-2}F_{n-1} = F_{2n-1}$,
(4) $F_{n+1}F_{n+2} - F_n F_{n+3} = (-1)^n$,
(5) $F_1F_2 + F_2F_3 + \cdots + F_{2n-1}F_{2n} = F_{2n}^2$,
(6) $F_n^3 + F_{n+1}^3 - F_{n-1}^3 = F_{3n}$.

109 (☕, Teorema di Zeckendorf). Dimostrare che per ogni intero positivo N vi è un'unica successione $a_1, a_2, \dots, a_n$ di interi maggiori di 1 tali che:
(1) $a_{i+1} \geq a_i + 2$ per ogni $i = 1, \dots, n-1$;
(2) $N = F_{a_1} + F_{a_2} + \cdots + F_{a_n}$.

110 (☕, Ⓐ). La definizione più comune in letteratura dei numeri di Bernoulli B_n non è la (2.6) ma coinvolge invece lo sviluppo in serie di Taylor della funzione $B(x) = \dfrac{x}{e^x - 1}$, viz.

$$\sum_{n=0}^{\infty} \frac{B_n}{n!} x^n = \frac{x}{e^x - 1} = 1 - \frac{x}{2} + \frac{x^2}{12} - \frac{x^4}{720} + \frac{x^6}{30240} - \frac{x^8}{1209600} + \cdots .$$

Sia $Q(x) = B(-x)$; provare che $B(x) - Q(x) = -x$ e dedurre che $B_n = 0$ per ogni n dispari e maggiore di 1; provare che $Q(x) = e^x B(x)$ e dedurre le equazioni ricorsive (2.6).

111 (☕). Sia B_n la successione dei numeri di Bernoulli. Provare che per ogni $n \geq 2$ vale la formula:

$$B_n = \frac{-1}{1 + n(-1)^n} \sum_{i=1}^{n-1} (-1)^i \binom{n}{i} B_i B_{n-i}\,.$$

Dedurre da ciò che se $n \geq 2$ allora:

(1) $B_n = 0$ se e solo se n è dispari,
(2) $B_n < 0$ se e solo se n è divisibile per 4,
(3) $B_n > 0$ se e solo se n è pari e non divisibile per 4.

(Suggerimento: nelle notazioni dell'Esercizio 110 provare che $xQ(x)' = Q(x) - Q(x)B(x)$.)

112 (Triangoli rovesciati di Tartaglia). Chiameremo triangolo rovesciato di Tartaglia una successione doppia $a_{i,j}$ di numeri razionali, dipendenti da due indici $i, j \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$ interi non negativi, tale che $a_{0,0} = 1$ e

$$a_{i,j} = a_{j,i}, \qquad a_{i,j} = a_{i+1,j} + a_{i,j+1},$$

per ogni coppia di indici $i, j \geq 0$. Per esempio, è un triangolo rovesciato la successione $a_{i,j} = 2^{-i-j}$. Dimostrare che:

(1) per ogni triangolo rovesciato $a_{i,j}$ si ha $a_{1,0} = a_{0,1} = 1/2$ e, per ogni $n > 0$, vale la formula
$$\sum_{i=0}^{n} \binom{n}{i} a_{i,n-i} = 1;$$
(2) per ogni $t \in \mathbb{Q}$ la successione
$$a_{i,j} = \frac{t^i(1-t)^j + t^j(1-t)^i}{2}$$
è un triangolo rovesciato (quando $t = 0$ oppure $t = 1$ si pone per convenzione $0^0 = 1$). In particolare esistono infiniti triangoli rovesciati;
(3) esiste un unico triangolo rovesciato $a_{i,j}$ tale che $a_{n,n} = 0$ per ogni intero positivo $n > 0$;
(4) (☕) esiste un unico triangolo rovesciato $a_{i,j}$ tale che $a_{n,0} = 0$ per ogni intero dispari $n \geq 3$;
(5) (Ⓐ☕) si consideri la funzione in due variabili
$$f(x, y) = \frac{x - y}{e^x - e^y}$$
assieme a tutte le sue derivate parziali calcolate nel punto $(0, 0)$:
$$B_{i,j} = \frac{\partial^{i+j} f}{\partial^i x\, \partial^j y}(0, 0)\,.$$
Provare che $B_{0,n} = B_n$ (numeri di Bernoulli) e che $a_{i,j} = (-1)^{i+j} B_{i,j}$ è un triangolo rovesciato di Tartaglia.

CAPITOLO 3

Numeri reali e complessi

Continuiamo la parte di algebretta iniziata nel capitolo precedente, studiando i numeri reali, i numeri complessi, i polinomi e le funzioni razionali; faremo anche una breve ma importante escursione nell'algebra astratta introducendo i concetti di campo e sottocampo. Per facilitare la comprensione di questi concetti studieremo, in aggiunta ai numeri reali e complessi, alcuni esempi di campi e sottocampi descrivibili in maniera molto concreta mediante numeri razionali e radici quadrate di interi positivi.

3.1. I numeri reali

È noto sin dai tempi di Pitagora che non tutte le radici quadrate di numeri interi sono razionali.

ESEMPIO 3.1.1. Non esiste alcun numero razionale r tale che $r^2 = 2$. Infatti, se esistesse un tale r potremmo scrivere

$$r = \frac{a}{b}$$

con $b > 0$ e $a, b \in \mathbb{Z}$ privi di fattori comuni. Si ha $2 = \dfrac{a^2}{b^2}$, da cui $2a^2 = b^2$. Pertanto 2 divide b^2 e quindi b è pari; allora scrivendo $b = 2c$ si ottiene $2a_2 = 4c^2$, da cui $a^2 = 2c^2$ e come prima deduciamo che a è pari. Così a e b hanno il fattore 2 in comune e questo è in contraddizione con l'asserto iniziale.

La necessità, tra le altre, di risolvere le equazioni del tipo $x^2 = n$ porta all'introduzione, in aggiunta ai numeri naturali, interi e razionali, del sistema numerico dei **numeri reali**, solitamente indicato con il simbolo $\mathbb{R}$.

Esistono diversi modi di costruire i numeri reali a partire dai numeri razionali, tutti abbastanza laboriosi, non banali e richiedenti una certa maturità matematica del lettore. Una possibile costruzione, basata sulla nozione di insieme quoziente, verrà data nella Sezione 13.3. Per il momento è preferibile accontentarsi di una descrizione intuitiva ed informale, sebbene imprecisa e lacunosa, secondo la quale i numeri reali positivi sono l'insieme delle possibili lunghezze dei segmenti nel piano euclideo e più in generale delle curve regolari contenute in esso: ad esempio è un numero reale la lunghezza della circonferenza di raggio 1. Esiste una bigezione tra l'insieme dei numeri reali e l'insieme dei punti della **retta reale**, ossia della retta Euclidea in cui sono stati fissati due punti distinti 0 e 1. I numeri razionali sono anche reali, un numero reale che non è razionale si dice **irrazionale**.

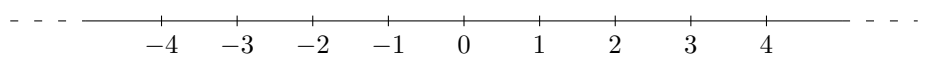


FIGURA 3.1. Inclusione degli interi nella retta reale.

Un altro modo, quasi perfetto, di descrivere i numeri reali è mediante sviluppi decimali del tipo

$$3{,}14159265358979323846264338327950288419716939937510582097494459 23\ldots .$$

Per sviluppo decimale si intende un'espressione

$$m{,}\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\ldots$$

dove $m \in \mathbb{Z}$ e $\alpha_1, \alpha_2, \ldots$ è una successione, possibilmente infinita,[1] di cifre decimali, ossia $\alpha_i = 0, \ldots, 9$ per ogni indice i; è ben noto che in tale rappresentazione i numeri razionali sono tutti e soli quelli il cui sviluppo decimale è finito oppure infinito periodico. Il difetto è che certi sviluppi decimali, in apparenza diversi, danno lo stesso numero reale: ad esempio $0{,}999999\ldots = 1$. Similmente, dato uno sviluppo decimale finito di un numero non intero, otteniamo uno sviluppo equivalente diminuendo di 1 l'ultima cifra decimale non nulla ed aggiungendo di seguito infiniti 9:

$$65{,}25299999\ldots = 65{,}253\,, \qquad -3{,}1699999\ldots = -3{,}17.$$

I numeri reali si possono sommare e moltiplicare; il prodotto di un numero finito di numeri reali si annulla se e soltanto se almeno uno dei fattori è uguale a 0. Inoltre, i numeri reali si possono ordinare, nel senso che dati due numeri reali a, b si può sempre dire se $a < b$ (a è minore di b), se $a = b$ oppure se $a > b$ (a è maggiore di b). Scriveremo $a \le b$ se $a < b$ oppure se $a = b$; similmente scriveremo $a \ge b$ se $a > b$ oppure se $a = b$. Se $a \le b$ allora $-a \ge -b$.

FIGURA 3.2. Quando la retta reale viene disegnata in orizzontale, di norma il punto 1 si trova a destra dello 0.

Abbiamo già detto che $\mathbb{Q} \subseteq \mathbb{R}$; ogni numero reale può essere approssimato per difetto e per eccesso con numeri razionali la cui differenza può essere scelta "piccola a piacere": con ciò intendiamo che per ogni numero reale t e per ogni numero razionale $\epsilon > 0$ possiamo trovare due numeri razionali a, b tali che

$$a \le t \le b, \qquad b - a \le \epsilon.$$

Ad esempio se $t \ge 0$ (rispettivamente: $t \le 0$) e $\epsilon = 10^{-n}$ si può prendere come a (rispettivamente: b) il troncamento alla n-esima cifra dopo la virgola dello sviluppo decimale di t e definire $b = a + \epsilon$.

Il **valore assoluto** $|a|$ di un numero reale a è definito mediante la formula

$$|a| = \begin{cases} a & \text{se } a \ge 0, \\ -a & \text{se } a \le 0. \end{cases}$$

In ogni caso vale $|a| = |-a| \ge 0$ e, per ogni $a, b \in \mathbb{R}$, si ha:

$$|a + b| \le |a| + |b|, \qquad |ab| = |a||b|.$$

Per induzione su n si deduce che per ogni $a_1, \ldots, a_n \in \mathbb{R}$ si ha

$$|a_1 a_2 \cdots a_n| = |a_1||a_2 \cdots a_n| = |a_1||a_2| \cdots |a_n|,$$

$$|a_1 + a_2 + \cdots + a_n| \le |a_1| + |a_2 + \cdots + a_n| \le |a_1| + |a_2| + \cdots + |a_n|. \tag{3.1}$$

La disguaglianza (3.1) viene detta **disuguaglianza triangolare**. Per ogni numero reale $a \in \mathbb{R}$ si ha $a^2 = |a|^2 \ge 0$ e vale $a^2 = 0$ se e solo se $a = 0$.

La **parte intera** $\lfloor x \rfloor$ di un numero reale x è il più grande intero minore o uguale a x. Equivalentemente $\lfloor x \rfloor$ è l'unico intero tale che $\lfloor x \rfloor \le x < \lfloor x \rfloor + 1$. Ad esempio:

$$\lfloor 1 \rfloor = 1, \quad \lfloor \sqrt{2} \rfloor = 1, \quad \lfloor 1 - \sqrt{2} \rfloor = -1, \quad \lfloor \pi \rfloor = 3, \quad \lfloor -\pi \rfloor = -4.$$

Altre importanti proprietà dei numeri reali, che dimostreremo in seguito, sono la densità dei numeri razionali (Proposizione 13.3.9) ed il principio di completezza (Teorema 13.3.12), e cioè:

Densità dei numeri razionali. *Siano $s, t \in \mathbb{R}$ con $s < t$. Allora esiste $r \in \mathbb{Q}$ tale che $s < r < t$.*

[1] In matematica l'avverbio possibilmente sta ad indicare una possibilità e non una preferenza né una probabilità né una intenzione.

Principio di completezza dei numeri reali. *Siano $A, B \subseteq \mathbb{R}$ due sottoinsiemi non vuoti tali che $a \le b$ per ogni $a \in A$ ed ogni $b \in B$. Esiste allora un numero reale $\xi \in \mathbb{R}$ tale che*

$$a \le \xi \le b \qquad \text{per ogni } a \in A,\ b \in B.$$

Se $a, b \in \mathbb{R}$ e $a \le b$, l'insieme $[a, b] = \{t \in \mathbb{R} \mid a \le t \le b\}$ viene detto **intervallo chiuso** di estremi a e b. Una delle prime applicazioni del principio di completezza è il teorema di esistenza degli zeri, di cui il prossimo teorema è un caso molto particolare, e tuttavia più che sufficiente per gli utilizzi di queste note.

Ricordiamo che una funzione $f\colon \mathbb{R} \to \mathbb{R}$ si dice polinomiale se esistono un intero $d \ge 0$ e $d+1$ numeri reali $a_0, \dots, a_d$ tali che

$$f(x) = a_0 + a_1 x + \cdots + a_d x^d, \qquad \text{per ogni } x \in \mathbb{R}. \tag{3.2}$$

Teorema 3.1.2. *Siano $f\colon \mathbb{R} \to \mathbb{R}$ una funzione polinomiale e $p \le q$ due numeri reali tali che $f(p)f(q) \le 0$. Allora esiste $\xi \in [p, q]$ tale che $f(\xi) = 0$.*

Dimostrazione. Supponiamo per fissare le idee che $f(p) \le f(q)$; se $f(p) \ge f(q)$ il ragionamento è del tutto simile, oppure più semplicemente basta considerare $-f$ al posto di f. Non è restrittivo supporre $p < q$: infatti se $p = q$ per ipotesi si ha $f(p)^2 \le 0$ e questo è possibile solo per $f(p) = 0$.

Definiamo il numero reale $M = \max(-p, q)$. Dal fatto che $p < q$ segue che $M > 0$ e che per ogni $x \in [p, q]$ vale $|x| \le M$. Dunque per ogni $x, y \in [p, q]$ ed ogni $a, b \in \mathbb{N}$ si ha $|x^a y^b| \le M^{a+b}$. In particolare, segue dalla disuguaglianza triangolare (3.1) che per ogni $x, y \in [p, q]$ ed ogni intero positivo n si hanno le disuguaglianze

$$|x^{n-1} + x^{n-2}y + \cdots + y^{n-1}| \le nM^{n-1},$$
$$|x^n - y^n| = |x - y|\,|x^{n-1} + x^{n-2}y + \cdots + y^{n-1}| \le nM^{n-1}|x - y|.$$

Supponiamo che f sia rappresentata come in (3.2), se $a_i = 0$ per ogni $i > 0$ allora la funzione f è costante e la condizione $f(p)f(q) \le 0$ è soddisfatta solo nel caso f identicamente nulla. Se invece $a_i \ne 0$ per qualche $i > 0$, definiamo

$$C = \sum_{n=1}^{d} n|a_n|M^{n-1}$$

osservando che $C > 0$. Per la disuguaglianza triangolare (3.1) si ha

$$|f(x) - f(y)| \le \sum_{n=1}^{d} |a_n|\,|x^n - y^n| \le \sum_{n=1}^{d} n|a_n|M^{n-1}|x - y| = C|x - y|.$$

Osserviamo inizialmente che $f(p) \le 0$ e $f(q) \ge 0$: infatti, se fosse $f(p) > 0$ allora $f(q) \ge f(p) > 0$ e quindi $f(p)f(q) > 0$; similmente se $f(q) < 0$ si ha $f(p) \le f(q) < 0$ e quindi $f(p)f(q) > 0$.

Consideriamo i seguenti sottoinsiemi di $\mathbb{R}$:

$$A = \{a \in [p, q] \mid f(a) \le 0\}, \qquad B = \{b \in [p, q] \mid a \le b \text{ per ogni } a \in A\}.$$

Chiaramente $p \in A$, $q \in B$ e pertanto A e B sono diversi dal vuoto; per costruzione $a \le b$ per ogni $a \in A$ ed ogni $b \in B$ e per il principio di completezza esiste $\xi \in \mathbb{R}$ tale che $a \le \xi \le b$ per ogni $a \in A$, $b \in B$. Dunque $\xi \in [p, q]$ e vogliamo dimostrare che $f(\xi) = 0$; lo faremo mostrando che entrambe le alternative $f(\xi) < 0$ e $f(\xi) > 0$ conducono ad una contraddizione.

Se fosse $f(\xi) < 0$, allora $\xi < q$ e denotando con t il minimo tra $q - \xi$ e $\dfrac{|f(\xi)|}{C}$ si ha

$$t > 0, \quad \xi + t \in [p, q], \quad Ct \le |f(\xi)|,$$
$$f(\xi + t) \le f(\xi) + |f(\xi + t) - f(\xi)| \le f(\xi) + Ct \le 0,$$

e quindi $\xi + t \in A$, in contraddizione con il fatto che $\xi \ge a$ per ogni $a \in A$.

Se invece fosse $f(\xi) > 0$, allora $\xi > p$. Denotando $t = \dfrac{f(\xi)}{2C}$ si ha $t > 0$, dunque $\xi - t \notin B$ e, per come è definito B, esiste $a \in A$ tale che $\xi - t < a$. D'altra parte $a \le \xi$ e quindi $|\xi - a| < t$. Però

$$f(a) \ge f(\xi) - |f(\xi) - f(a)| \ge f(\xi) - C|\xi - a| > f(\xi) - Ct = \frac{f(\xi)}{2} > 0$$

in contraddizione con la definizione di A. □

COROLLARIO 3.1.3. *Per ogni $t \in \mathbb{R}$ non negativo ed ogni intero positivo n esiste un unico numero reale non negativo $\sqrt[n]{t}$ tale che $(\sqrt[n]{t})^n = t$.*

DIMOSTRAZIONE. Mostriamo prima l'esistenza, a tal fine fissiamo un intero positivo m tale che $m^n \geq t$. Allora la funzione polinomiale

$$f\colon [0,m] \to \mathbb{R}, \qquad f(x) = x^n - t,$$

soddisfa le ipotesi del Teorema 3.1.2 e quindi esiste $\xi \in [0,m]$ tale che $f(\xi) = \xi^n - t = 0$, ossia $\xi^n = t$.

Per dimostrare l'unicità basta osservare che se $x, y \geq 0$ e $x^n = y^n$, allora

$$(x-y)(x^{n-1} + x^{n-2}y + \cdots + xy^{n-2} + y^{n-1}) = x^n - y^n = 0$$

e quindi $x = y$ oppure $x^{n-1} + x^{n-2}y + \cdots + y^{n-1} = 0$. Ma ogni addendo $x^i y^{n-1-i}$ è non negativo e quindi $x^{n-1} + x^{n-2}y + \cdots + y^{n-1} = 0$ implica necessariamente $x^{n-1} = y^{n-1} = 0$, da cui $x = y = 0$. □

Il numero $\sqrt[2]{t}$ viene detto radice quadrata di t, $\sqrt[3]{t}$ radice cubica, $\sqrt[4]{t}$ radice quarta, ..., $\sqrt[n]{t}$ radice n-esima; per le radici quadrate di solito viene omesso l'apice e si scrive semplicemente $\sqrt[2]{t} = \sqrt{t}$.

Un numero naturale n è un quadrato perfetto se esiste un altro numero naturale m tale che $n = m^2$ e l'Esempio 3.1.1 può essere facilmente esteso a tutti i numeri naturali che non sono quadrati perfetti.

TEOREMA 3.1.4. *Sia $n \in \mathbb{N}$, allora $\sqrt{n}$ è razionale se e solo se $\sqrt{n} \in \mathbb{N}$, ossia se e solo se n è un quadrato perfetto.*

DIMOSTRAZIONE. L'unica affermazione non banale da dimostrare è che se $\sqrt{n}$ non è intero, allora $\sqrt{n}$ non è nemmeno razionale.

Supponiamo per assurdo che $\sqrt{n}$ sia un numero razionale non intero; fra tutte le possibili coppie x, y di interi positivi tali che $\sqrt{n} = x/y$ consideriamo quella con il più piccolo valore di y: per definizione $x^2 = ny^2$. Indichiamo con $p = \lfloor \sqrt{n} \rfloor$ la parte intera di $\sqrt{n}$, per ipotesi si ha $p < \sqrt{n} < p+1$ e di conseguenza valgono le disuguaglianze

$$p < \frac{x}{y} < p+1, \qquad yp < x < y(p+1), \qquad 0 < x - py < y.$$

Osserviamo adesso che

$$\frac{x}{y} - \frac{ny - px}{x - py} = \frac{x(x-py) - y(ny-px)}{y(x-py)} = \frac{x^2 - ny^2}{y(x-py)} = 0,$$

e cioè che vale l'uguaglianza

$$\sqrt{n} = \frac{ny - px}{x - py}$$

che però entra in contraddizione con le disuguaglianze $0 < x - py < y$ e con la scelta della coppia x, y. □

Esercizi.

113. Siano a, b, c, d numeri razionali, con c, d non entrambi nulli, e sia α u numero irrazionale. Provare che $ad = bc$ se e solo se il numero $(a\alpha + b)/(c\alpha + d)$ è razionale.

114 (♡). Provare che per ogni intero positivo n esiste un intero a tale che

$$n \leq a^2 \leq n + 2\sqrt{n-1}.$$

115. Siano c, t due numeri reali tali che $0 \leq c$ e $c^2 \leq t$. Mostrare che

$$(c+\alpha)^2 \leq t \quad \text{per ogni} \quad 0 \leq \alpha \leq \min\left(c+1, \frac{t-c^2}{3c+1}\right).$$

Nel prossimo Esercizio 116 compare il cosiddetto "principio dei cassetti" o "del portalettere" (in inglese: *pigeonhole principle*, espressione spesso tradotta erroneamente come "principio delle gabbie dei piccioni", mentre per gli inglesi pigeonhole non è la "piccionaia", ma la cassetta delle lettere). Il principio dei cassetti afferma che se si ripartiscono più di k oggetti in k

cassetti, necessariamente almeno uno dei cassetti conterrà più di un oggetto; equivalentemente, una funzione da un insieme con più di k elementi in uno con k elementi non può essere iniettiva. Più in generale, se si ripartiscono almeno $nk+1$ oggetti in k cassetti, almeno uno dei cassetti dovrà contenere almeno $n+1$ oggetti.

116. 1) Dimostrare che per ogni terna di interi $a, b, n \in \mathbb{Z}$, con $0 < b \le n$, si ha:

$$\left|\frac{1}{n+1} - \frac{a}{b}\right| \ge \frac{1}{b(n+1)}.$$

Per quali valori di a e b vale l'uguaglianza?

2) Dato un un numero reale ξ ed un intero positivo n, consideriamo gli $n+2$ numeri reali

$$x_0 = 0, \quad x_1 = \xi - \lfloor \xi \rfloor, \quad x_2 = 2\xi - \lfloor 2\xi \rfloor, \quad \dots, \; x_n = n\xi - \lfloor n\xi \rfloor, \quad x_{n+1} = 1.$$

Dimostrare, usando il principio dei cassetti, che esistono due indici i, j con $0 \le i < j \le n+1$ tali che $|x_j - x_i| \le \dfrac{1}{n+1}$ e dedurre che esistono due interi a e b tali che

$$0 < b \le n, \qquad \left|\xi - \frac{a}{b}\right| \le \frac{1}{b(n+1)}.$$

117. Sia $a_0, a_1, a_2, \dots$ la successione di numeri reali:

$$a_0 = 0, \quad a_1 = 2, \quad a_2 = 2 + \sqrt{2}, \quad a_3 = 2 + \sqrt{2 + \sqrt{2}}, \dots, \quad a_n = 2 + \sqrt{a_{n-1}}, \dots \; .$$

Dimostrare, usando il principio di induzione, che per ogni $n > 0$ valgono le disuguaglianze $0 \le a_{n-1} < a_n < 4$.

118 (Principio di Cantor degli intervalli incapsulati). Sia $[a_n, b_n]$, $n \in \mathbb{N}$, una successione di intervalli chiusi tali che per ogni $n > 0$ si ha

$$\emptyset \ne [a_n, b_n] \subseteq [a_{n-1}, b_{n-1}].$$

Dedurre dal principio di completezza che $\bigcap_{n=0}^{\infty} [a_n, b_n] \ne \emptyset$, ossia che esiste almeno un numero reale contenuto in tutti gli intervalli.

119 (☕). Usare la disuguaglianza dell'Esempio 2.3.5 per dimostrare che per ogni intero $n > 0$ vale

$$\frac{\left(1 + \dfrac{1}{n+1}\right)^{n+1}}{\left(1 + \dfrac{1}{n}\right)^{n}} = \left(\frac{1 + \dfrac{1}{n+1}}{1 + \dfrac{1}{n}}\right)^{n} \left(1 + \frac{1}{n+1}\right) \ge 1,$$

e di conseguenza

$$\left(1 + \frac{1}{n+1}\right)^{n+1} \ge \left(1 + \frac{1}{n}\right)^{n} \ge 2$$

per ogni intero $n > 0$. Mostrare inoltre, con un ragionamento analogo, che per ogni $n > 0$ vale

$$\left(1 + \frac{1}{n+1}\right)^{n+2} \le \left(1 + \frac{1}{n}\right)^{n+1} \le 4$$

e dedurre dal principio degli intervalli incapsulati che esiste un unico numero reale e tale che per ogni intero positivo n vale

$$\left(1 + \frac{1}{n}\right)^{n} \le e \le \left(1 + \frac{1}{n}\right)^{n+1} = \frac{1}{\left(1 - \dfrac{1}{n+1}\right)^{n+1}}.$$

Nota: si può dimostrare che e è un numero irrazionale. Il suo valore approssimato alla ventesima cifra decimale è $2,71828182845904523536$, vedi anche l'Esercizio 91.

120 (☕, ♡). Siano α, β due numeri irrazionali positivi tali che

$$\frac{1}{\alpha} + \frac{1}{\beta} = 1.$$

Dimostrare che $\lfloor n\alpha \rfloor \ne \lfloor m\beta \rfloor$ per ogni coppia di interi positivi n, m e che ogni intero positivo è uguale a $\lfloor n\alpha \rfloor$ oppure a $\lfloor n\beta \rfloor$ per un opportuno intero n. (Suggerimento: per ogni intero N calcolare quanti sono gli interi $n > 0$ tali che $\lfloor n\alpha \rfloor \le N$.)

3.2. Estensioni quadratiche

Oltre ai numeri reali e razionali esistono altri insiemi di numeri interessanti: uno di questi è $\mathbb{Q}(\sqrt{2})$ definito come l'insieme dei numeri reali x che si possono scrivere come $x = a + b\sqrt{2}$, con a, b numeri razionali. Equivalentemente

$$\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}.$$

Poiché ogni numero razionale a si può scrivere nella forma $a = a + 0\sqrt{2}$, ne consegue che $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2})$; se $x, y \in \mathbb{Q}(\sqrt{2})$ allora anche

$$-x,\ x + y,\ xy \in \mathbb{Q}(\sqrt{2}).$$

Infatti, se $x = a + b\sqrt{2}$ e $y = c + d\sqrt{2}$ si ha $-x = (-a) + (-b)\sqrt{2}$, $x + y = (a + c) + (b + d)\sqrt{2}$ e

$$\begin{aligned} xy &= (a + b\sqrt{2})(c + d\sqrt{2}) = ac + ad\sqrt{2} + bc\sqrt{2} + bd(\sqrt{2})^2 \\ &= (ac + 2bd) + (bc + ad)\sqrt{2} \in \mathbb{Q}(\sqrt{2}). \end{aligned}$$

Mostriamo adesso che se $x \in \mathbb{Q}(\sqrt{2})$ e $x \neq 0$, allora anche $x^{-1} \in \mathbb{Q}(\sqrt{2})$; più precisamente mostriamo che se a, b sono due numeri razionali non entrambi nulli, allora $a + b\sqrt{2} \neq 0$ e vale

$$(a + b\sqrt{2})^{-1} = \frac{a - b\sqrt{2}}{a^2 - 2b^2} = \frac{a}{a^2 - 2b^2} - \frac{b}{a^2 - 2b^2}\sqrt{2}.$$

Infatti se fosse $a + b\sqrt{2} = 0$ allora $b \neq 0$ altrimenti $a = -b\sqrt{2} = 0$ contrariamente all'ipotesi che a e b non siano entrambi nulli. Dividendo per b si ottiene $\sqrt{2} = -a/b$, ossia che $\sqrt{2}$ è un numero razionale: assurdo! Abbiamo quindi dimostrato che i due numeri $a + b\sqrt{2}$ e $a - b\sqrt{2}$ sono non nulli (il secondo poiché a e $-b$ non sono entrambi nulli) e quindi

$$\frac{1}{a + b\sqrt{2}} = \frac{1}{a + b\sqrt{2}} \frac{a - b\sqrt{2}}{a - b\sqrt{2}} = \frac{a - b\sqrt{2}}{a^2 - 2b^2}.$$

La precedente costruzione funziona se poniamo al posto di 2 un qualsiasi numero primo p (abbiamo già dimostrato che $\sqrt{p}$ è irrazionale):

$$\mathbb{Q}(\sqrt{p}) = \{a + b\sqrt{p} \mid a, b \in \mathbb{Q}\}.$$

Si verifica rapidamente che $\mathbb{Q}(\sqrt{p})$ è chiuso rispetto alla somma e al prodotto, cioè, la somma e/o il prodotto di due elementi di $\mathbb{Q}(\sqrt{p})$ è ancora un elemento di $\mathbb{Q}(\sqrt{p})$. La dimostrazione dell'esistenza dell'inverso ricalca fedelmente quella per $p = 2$, ossia dato $x = a + b\sqrt{p}$ con a e b non entrambi nulli si ha:

$$x^{-1} = \frac{a - \sqrt{p}b}{a^2 - pb^2} = \frac{a}{a^2 - pb^2} + \frac{-b}{a^2 - pb^2}\sqrt{p}.$$

Ancora più in generale, se ξ è un qualsiasi numero irrazionale il cui quadrato ξ^2 è razionale, allora l'insieme

$$\mathbb{Q}(\xi) = \{a + b\xi \mid a, b \in \mathbb{Q}\}$$

si comporta allo stesso modo di $\mathbb{Q}(\sqrt{2})$, e cioè è chiuso per le operazioni di somma, prodotto, opposto e inverso di elementi non nulli: i dettagli sono lasciati per esercizio al lettore.

Esempio 3.2.1. Dimostriamo che $\mathbb{Q}(\sqrt{2}) \cap \mathbb{Q}(\sqrt{3}) = \mathbb{Q}$. Infatti, per ogni elemento x dell'intersezione possiamo scrivere

$$x = a + b\sqrt{2} = c + d\sqrt{3}, \qquad a, b, c, d \in \mathbb{Q},$$

e vogliamo dimostrare che x è razionale: a tal fine basta dimostrare che $b = 0$ oppure $d = 0$. Si ha $d\sqrt{3} = (a - c) + b\sqrt{2}$, ed elevando al quadrato

$$3d^2 = (a - c)^2 + 2b^2 + 2b(a - c)\sqrt{2}$$

e, siccome $\sqrt{2}$ non è razionale, deve essere $b(a - c) = 0$. Se $b = 0$ abbiamo $x = a \in \mathbb{Q}$, mentre se $a = c$ allora $3d^2 = 2b^2$ e questo è possibile solo se $d = b = 0$, non essendo $3/2$ il quadrato di un numero razionale.

Esercizi.

121. Scrivere i seguenti numeri nella forma $a + b\sqrt{2}$:

$$(1+\sqrt{2})^3, \quad \frac{2-\sqrt{2}}{3+2\sqrt{2}}, \quad \frac{1+\sqrt{2}}{1-\sqrt{2}}, \quad \frac{1-2\sqrt{2}}{1-\sqrt{2}}, \quad (1-\sqrt{2})^{200}(1+\sqrt{2})^{200}.$$

122. Scrivere i seguenti numeri nella forma $a + b\sqrt{5}$:

$$(1+2\sqrt{5})^2, \quad \frac{1-\sqrt{5}}{1+\sqrt{5}}, \quad \frac{\sqrt{5}}{5-\sqrt{5}}, \quad (2+2\sqrt{5})(2-2\sqrt{5}).$$

123 (♡). Determinare se $\sqrt{3+2\sqrt{2}} \in \mathbb{Q}(\sqrt{2})$.

124. Risolvere il seguente sistema lineare a coefficienti in $\mathbb{Q}(\sqrt{3})$:

$$\begin{cases} x + y + z = 1 \\ x + \sqrt{3}\, y + 2z = 0 \\ x + 3y + 4z = -1 \end{cases}$$

(si chiede di trovare le terne $x, y, z \in \mathbb{Q}(\sqrt{3})$ che soddisfano il precedente sistema lineare).

125. Risolvendo un opportuno sistema lineare, trovare tre numeri razionali x, y, z tali che

$$\frac{1}{1 - \sqrt[3]{2} + 2\sqrt[3]{4}} = x + y\sqrt[3]{2} + z\sqrt[3]{4}.$$

126. Siano p, q interi positivi senza fattori comuni. Provare che $\sqrt{\dfrac{p}{q}} = \dfrac{\sqrt{p}}{\sqrt{q}}$ è un numero razionale se e solo se p, q sono due quadrati.

127. Dati due numeri reali non nulli a, b, con $a \neq b$, definiamo per ogni intero non negativo n il numero

$$A_n = \frac{a^n - b^n}{a - b}.$$

Provare che valgono le formule

$$A_0 = 0, \quad A_1 = 1, \quad A_{n+1} = (a+b)A_n - abA_{n-1}, \quad n > 0,$$

e dedurre che se

$$x = \frac{1+\sqrt{5}}{2}, \qquad y = \frac{1-\sqrt{5}}{2},$$

allora

$$F_n = \frac{x^n - y^n}{x - y} = \frac{x^n - y^n}{\sqrt{5}} = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n$$

coincide con la successione dei numeri di Fibonacci:

$$F_0 = 0, \quad F_1 = 1, \quad F_{n+1} = F_n + F_{n-1}.$$

128. Si consideri la successione

$$a_1 = 2, \quad a_2 = 6, \quad a_3 = 14, \quad \dots \quad a_n = 2a_{n-1} + a_{n-2}, \dots$$

Dimostrare:

(1) $a_n = (1+\sqrt{2})^n + (1-\sqrt{2})^n$;
(2) la parte intera di $(1+\sqrt{2})^n$ è pari se e solo se n è dispari.

129. Usare il risultato dell'Esempio 3.2.1 per dimostrare che se vale

$$a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6} = 0$$

con $a, b, c, d \in \mathbb{Q}$, allora $a = b = c = d = 0$.

130 (☕). Quali sono i numeri del tipo

$$a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}, \qquad a, b, c, d \in \mathbb{Q},$$

che sono radici di un'equazione di secondo grado a coefficienti razionali?

3.3. I numeri complessi

Nella costruzione formale di $\mathbb{Q}(\sqrt{2})$ e nella definizione delle operazioni di somma e prodotto abbiamo avuto bisogno di sapere solo due cose: che $\sqrt{2} \notin \mathbb{Q}$ e che $(\sqrt{2})^2 \in \mathbb{Q}$. Siamo adesso pronti per ripetere la costruzione in una situazione più astratta dove il "protagonista" è *l'unità immaginaria* i che deve essere pensata come un puro *simbolo* dotato esclusivamente delle due proprietà formali:

$$i \notin \mathbb{R}, \qquad i^2 = -1.$$

A rigore, la prima proprietà segue dalla seconda in quanto ogni quadrato di un numero reale è maggiore o uguale a 0.

Definiamo l'insieme $\mathbb{C}$ dei **numeri complessi** come l'insieme formato dagli elementi $a+ib$, con a e b numeri reali:

$$\mathbb{C} = \{a + ib \mid a, b \in \mathbb{R}\}.$$

Ogni numero reale può essere pensato anche come un particolare numero complesso; per la precisione consideriamo $\mathbb{R}$ contenuto in $\mathbb{C}$ interpretando il numero complesso $a + i0$ come il numero reale a. In particolare quando scriviamo $0 \in \mathbb{C}$ intendiamo $0 = 0 + i0$ e quando scriviamo $1, i \in \mathbb{C}$ intendiamo rispettivamente $1 = 1 + i0$, $i = 0 + i1$. I numeri complessi del tipo $ib = 0 + ib$ si dicono **immaginari puri**; lo 0 è l'unico numero complesso ad essere contemporaneamente reale ed immaginario puro.

Definizione 3.3.1. Dato un numero complesso $a + ib$, i numeri reali a e b ne sono detti rispettivamente la **parte reale** e la **parte immaginaria**. Si scrive

$$\Re(a + ib) = a, \qquad \Im(a + ib) = b.$$

Dunque un numero complesso è reale se e solo se ha parte immaginaria uguale a 0 ed è immaginario puro se e solo se ha parte reale uguale a 0.

Vogliamo adesso definire le quattro operazioni sui numeri complessi. A differenza delle estensioni quadratiche, non abbiamo una rapprentazione di $a + ib$ come numero reale e allora si esegue il calcolo considerando i numeri complessi come pure espressioni algebriche, ricordandosi di mettere -1 al posto di i^2 ogni qualvolta quest'ultimo compare. Ad esempio si ha:

$$(1 + i)(1 - i) = 1 + i - i - i^2 = 1 - i^2 = 1 - (-1) = 2.$$

Equivalentemente possiamo definire le quattro operazioni in maniera assiomatica come nella definizione seguente.

Definizione 3.3.2. Il *campo dei numeri complessi* è l'insieme $\mathbb{C}$ dotato delle seguenti operazioni:

(1) $(a + ib) + (c + id) = (a + c) + i(b + d)$,

(2) $(a + ib) - (c + id) = (a - c) + i(b - d)$,

(3) $(a + ib)(c + id) = ac + i(ad + bc) + i^2bd = (ac - bd) + i(ad + bc)$,

(4) se $c + id \neq 0$, allora

$$\frac{a + ib}{c + id} = \frac{a + ib}{c + id}\,\frac{c - id}{c - id} = \frac{(a + ib)(c - id)}{c^2 + d^2} = \frac{ac + bd}{c^2 + d^2} + i\frac{bc - ad}{c^2 + d^2}.$$

Esempio 3.3.3.

$$\frac{1 + 2i}{1 - i} = \frac{1 + 2i}{1 - i}\,\frac{1 + i}{1 + i} = \frac{(1 + 2i)(1 + i)}{(1 - i)(1 + i)} = \frac{-1 + 3i}{2} = -\frac{1}{2} + i\frac{3}{2}.$$

Esempio 3.3.4. Si considerino i numeri complessi $z = 1 + i$ e $w = 1 - 2i$; si ha

$$z + w = (1 + i) + (1 - 2i) = 2 - i, \qquad z - w = (1 + i) - (1 - 2i) = 3i,$$

$$zw = (1 + i)(1 - 2i) = 3 - i, \qquad \frac{z}{w} = \frac{1 + i}{1 - 2i} = \frac{(1 + i)(1 + 2i)}{1 + 4} = \frac{-1}{5} + \frac{3}{5}i.$$

Abbiamo visto che per riportare una frazione $\dfrac{a + ib}{c + id}$ alla forma standard $x+iy$, è sufficiente moltiplicare a numeratore e denominatore per il numero complesso $c - id$.

DEFINIZIONE 3.3.5. Il **coniugato** di un numero complesso $a+ib$, $a,b\in\mathbb{R}$, è il numero complesso $a-ib$. Viene denotato con una sopralineatura, ossia

$$\overline{a+ib}=a-ib.$$

Ad esempio $\overline{1+2i}=1-2i$, $\overline{2-i}=2+i$, $\overline{3}=3$, $\overline{7i}=-7i$. Osserviamo che un numero complesso z è reale se e solo se è uguale al suo coniugato; più in generale si hanno le formule:

$$\overline{\overline{z}}=z,\qquad \Re(z)=\Re(\overline{z})=\frac{z+\overline{z}}{2},\qquad \Im(z)=-\Im(\overline{z})=\frac{z-\overline{z}}{2i}.$$

Moltiplicando un numero complesso diverso diverso da 0 per il suo coniugato si ottiene sempre un numero reale positivo:

$$(a+ib)(a-ib)=a^2+b^2.$$

Quindi se $z,w\in\mathbb{C}$ e $w\neq 0$ si ha

$$\frac{z}{w}=\frac{z\overline{w}}{w\overline{w}},\qquad w^{-1}=\frac{\overline{w}}{w\overline{w}}.$$

In particolare abbiamo dimostrato che ogni numero complesso $z\neq 0$ è invertibile ed il suo inverso è dato dalla formula

$$z^{-1}=\frac{1}{z}=\frac{1}{z}\frac{\overline{z}}{\overline{z}}=\frac{1}{z\,\overline{z}}\overline{z}\,,$$

che nella forma $a+ib$ diventa

$$\frac{1}{a+ib}=\frac{1}{a+ib}\frac{a-ib}{a-ib}=\frac{a-ib}{a^2+b^2}=\frac{a}{a^2+b^2}-i\frac{b}{a^2+b^2}.$$

ESEMPIO 3.3.6. Calcoliamo l'inverso del numero complesso $2-i$. Applicando la precedente formula si ha

$$\frac{1}{2-i}=\frac{1}{2-i}\frac{2+i}{2+i}=\frac{2+i}{5}=\frac{2}{5}+i\frac{1}{5}.$$

Il coniugio commuta con le operazioni di somma e prodotto, ossia per ogni coppia di numeri complessi z,w vale

$$\overline{z}+\overline{w}=\overline{z+w},\qquad \overline{z}\,\overline{w}=\overline{zw},$$

Infatti se $z=a+ib$ e $w=x+iy$ si ha $\overline{z}+\overline{w}=(a-ib)+(x-iy)=(a+x)-i(b+y)=\overline{z+w}$, $\overline{z}\,\overline{w}=(a-ib)(x-iy)=(ax-by)-i(bx+ay)=\overline{zw}$.

È importante osservare che le operazioni di somma e prodotto sul campo dei numeri complessi sono *associative*; ciò significa che per ogni terna $x,y,z\in\mathbb{C}$ si ha:

$$(x+y)+z=x+(y+z)\qquad\text{(associatività della somma)},$$
$$(xy)z=x(yz)\qquad\text{(associatività del prodotto)}.$$

Se z,w sono due numeri complessi non nulli, allora anche il loro prodotto zw è diverso da 0. Infatti se per assurdo fosse $zw=0$ allora si avrebbe

$$z=z(ww^{-1})=(zw)w^{-1}=0w^{-1}=0$$

contrariametne all'ipotesi che $z,w\neq 0$. L'associatività del prodotto permette in particolare di definire senza ambiguità le potenze z^n di un numero complesso z per ogni intero $n>0$.

ESEMPIO 3.3.7. Le potenze dell'unità immaginaria sono:

$$i^2=-1,\qquad i^3=i^2i=-i,\qquad i^4=(i^2)^2=1,\qquad i^5=i^4i=i,\qquad \dots$$

ESEMPIO 3.3.8. Le potenze del numero complesso $\xi=\dfrac{1+i}{\sqrt{2}}$ sono:

$$\xi^2=i,\quad \xi^3=\frac{-1+i}{\sqrt{2}},\quad \xi^4=-1,$$
$$\xi^5=-\xi,\quad \xi^6=-\xi^2,\quad \xi^7=-\xi^3,\quad \xi^8=-\xi^4=1,\quad \dots$$

ESEMPIO 3.3.9. Ogni numero reale negativo possiede una radice quadrata nel campo dei numeri complessi. Infatti se a è numero reale negativo, allora $\sqrt{-a}\in\mathbb{R}$ e si può considerare il numero immaginario puro $z=(\sqrt{-a})i$; chiaramente $z^2=i^2(\sqrt{-a})^2=-(\sqrt{-a})^2=a$.

TEOREMA 3.3.10. *Sia $z\in\mathbb{C}$ un qualunque numero complesso, allora esiste un numero complesso $w\in\mathbb{C}$ tale che $w^2=z$.*

DIMOSTRAZIONE. Già sappiamo che il teorema è vero se z è un numero reale. Se z non è reale scriviamo $z = a + ib$, con $a, b \in \mathbb{R}$, $b \neq 0$, e cerchiamo due numeri reali x, y tali che

$$a + ib = (x + iy)^2 = (x^2 - y^2) + 2ixy,$$

ossia cerchiamo di risolvere il sistema

$$\begin{cases} x^2 - y^2 = a \\ 2xy = b \end{cases}.$$

Siccome $b \neq 0$, la seconda equazione implica $x \neq 0$ e quindi possiamo moltiplicare la prima equazione per x^2 senza alcun rischio di introdurre soluzioni fantasma.

$$\begin{cases} x^4 - x^2y^2 = x^4 - \dfrac{b^2}{4} = ax^2 \\ 2xy = b \end{cases}.$$

Guardando all'equazione biquadratica $x^4 - ax^2 - \dfrac{b^2}{4} = 0$, siccome $b \neq 0$ l'equazione di secondo grado $t^2 - at - \dfrac{b^2}{4} = 0$ possiede una sola radice reale positiva uguale a

$$\frac{a + \sqrt{a^2 + b^2}}{2}$$

e quindi si ottiene

$$w = x + iy = x + i\frac{b}{2x}, \quad \text{dove} \quad x = \pm\sqrt{\frac{a + \sqrt{a^2 + b^2}}{2}}.$$

Riassumendo possiamo scrivere

$$\sqrt{z} = \pm\frac{1}{\sqrt{2}}\left(\sqrt{a + |z|} + i\frac{b}{\sqrt{a + |z|}}\right), \quad \text{dove } z = a + ib, \ \ |z| = \sqrt{a^2 + b^2}, \tag{3.3}$$

e dove per $\sqrt{z}$ si intende un qualunque numero complesso il cui quadrato è uguale a z. □

OSSERVAZIONE 3.3.11. Siccome $i^2 = -1$, sempre per $b \neq 0$, la Formula (3.3) è del tutto equivalente a

$$\sqrt{z} = i\sqrt{-z} = \pm\frac{1}{\sqrt{2}}\left(\frac{b}{\sqrt{|z| - a}} + i\sqrt{|z| - a}\right), \qquad z = a + ib. \tag{3.4}$$

Notiamo che la Formula (3.3) vale anche per numeri reali positivi e che la Formula (3.4) vale anche per numeri reali negativi. È possibile dimostrare che non esiste alcuna formula generale valida per tutti i numeri complessi diversi da 0, ma questo va (molto) al di là degli obiettivi di questo testo e richiede conoscenze non banali di topologia e variabile complessa.

ESEMPIO 3.3.12 (Radici dell'unità, prima parte). Diremo che un numero complesso z è una radice dell'unità se esiste un intero positivo $n > 0$ tale che $z^n = 1$; è ovvio che ogni radice dell'unità è un numero diverso da 0. Abbiamo già osservato che i numeri $\pm 1, \pm i$ e $\dfrac{1}{\sqrt{2}}(\pm 1 \pm i)$ sono radici dell'unità. Vedremo tra non molto che ne esistono molte altre, ed è chiaro che non tutti i numeri complessi sono radici dell'unità: ad esempio ogni numero reale diverso da ± 1 non è una radice dell'unità.

Se z è una radice dell'unità, allora lo è anche z^{-1} e più in gnerale lo sono tutte le potenze z^n, $n \in \mathbb{Z}$. Infatti se $r > 0$ è tale che $z^r = 1$, allora per ogni $n \in \mathbb{Z}$ si ha

$$(z^n)^r = z^{nr} = (z^r)^n = 1^n = 1.$$

Similmente, se z, u sono radici dell'unità, allora lo è anche zu. Infatti se r, s sono due interi positivi tali che $z^r = u^s = 1$ allora

$$(zu)^{rs} = z^{rs}u^{rs} = (z^r)^s(u^s)^r = 1^s 1^r = 1.$$

Sia z una radice dell'unità e sia $r > 0$ il *più piccolo* intero positivo tale che $z^r = 1$. Allora per un intero $n \in \mathbb{Z}$ vale $z^n = 1$ se e solo se n è divisibile per r. Infatti possiamo scrivere $n = qr + s$ con $0 \leq s < r$ e

$$z^s = z^{n-qr} = \frac{z^n}{(z^r)^q} = \frac{1}{1^q} = 1$$

e poiché per ipotesi r è il più piccolo intero positivo tale che $z^r = 1$ ne consegue che $s \leq 0$ oppure $s \geq r$. Dato che $0 \leq s < r$ l'unica possibilità è $s = 0$, ossia che r divide n.

Esercizi.

131. Scrivere i seguenti numeri complessi nella forma $a + ib$:

$$3(1-i) + i(2+i), \quad (2+4i)(1-2i), \quad \frac{1-i}{1+2i} + \frac{1-2i}{1-i}, \quad (1+i)^4, \quad \frac{(1+i)^2}{3-2i}.$$

132. Mostrare che per ogni numero complesso z vale

$$\Re(iz) = -\Im(z), \qquad \Im(iz) = \Re(z).$$

133. Provare che $z = \overline{z} \iff z \in \mathbb{R}$ e $z = -\overline{z} \iff iz \in \mathbb{R}$.

134. Trovare un numero complesso w tale che $w^2 = 2 + 2\sqrt{3}\,i$.

135 (♡). Trovare tutti i numeri complessi $z \in \mathbb{C}$ tali che $\overline{z}^2 + z = 0$.

136. Trovare tutti i numeri complessi z tali che $z(z + \overline{z}) = 4z$.

137. Siano z, w due numeri complessi tali che $z^2 = w^2$. Mostrare che $z = \pm w$ (suggerimento: considerare il prodotto $(z + w)(z - w)$).

138. Dire, motivando la risposta, se l'applicazione $f\colon \mathbb{C} \to \mathbb{C} \times \mathbb{C}$, $f(z) = (z^2, z^3)$, è iniettiva.

139. Sia z un numero complesso tale che $z^{64} = z^{81} = 1$. Chi è z?

140. Sia z una radice dell'unità e sia 15 il più piccolo intero positivo tale che $z^{15} = 1$. Dire, motivando la risposta se $u = z^3$ è una radice dell'unità e calcolare il più piccolo intero positivo r tale che $u^r = 1$.

141. Sia z un numero complesso tale che $z^{13783} = 1$. Dimostrare che $z^n \neq -1$ per ogni $n \in \mathbb{Z}$.

3.4. Rappresentazione geometrica dei numeri complessi

Poiché ogni numero complesso z è individuato in modo univoco dalla parte reale ed immaginaria, e cioè da una coppia di numeri reali, possiamo identificare $\mathbb{C}$ con il piano Cartesiano facendo corrispondere al numero complesso z il punto del piano di coordinate $(\Re z, \Im z)$. Quindi possiamo anche identificare i numeri complessi con l'insieme dei vettori nel piano aventi come punto iniziale l'origine.

Viceversa al punto del piano di coordinate cartesiane (x, y) associamo il numero complesso $z = x + iy$. Quando i punti del piano vengono identificati con i numeri complessi si parla di *piano di Gauss* anziché di piano cartesiano.

La distanza tra l'origine ed un punto $z = x + iy$ è calcolata usando il teorema di Pitagora ed è uguale a $|z| = \sqrt{x^2 + y^2}$.

Definizione 3.4.1. Il **modulo**, o **norma**, di un numero complesso $z = a + ib$ è il numero reale

$$|z| = \sqrt{a^2 + b^2}.$$

Si noti che per i numeri reali il modulo coincide con il valore assoluto, infatti se $z = a + i0 \in \mathbb{R}$, allora

$$|z| = \sqrt{a^2} = \begin{cases} a & \text{se } a \geq 0 \\ -a & \text{se } a \leq 0 \end{cases}$$

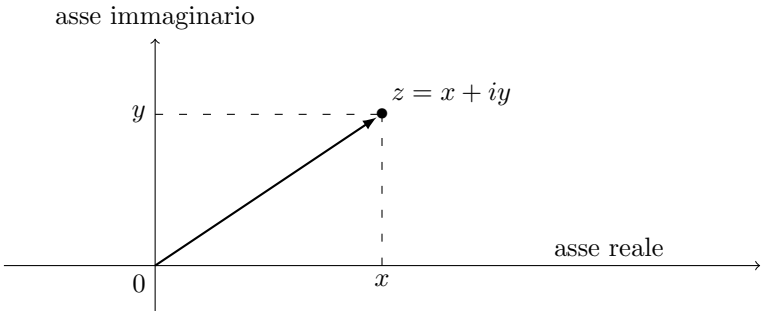


FIGURA 3.3. Il piano di Gauss, ossia la bigezione tra $\mathbb{C}$ ed i punti del piano.

ESEMPIO 3.4.2. Per ogni numero reale θ si ha

$$|\cos\theta + i\sin\theta| = \sqrt{\cos^2\theta + \sin^2\theta} = 1.$$

Più in generale, se $r \geq 0$ è un numero reale si ha

$$|r(\cos\theta + i\sin\theta)| = \sqrt{r^2\cos^2\theta + r^2\sin^2\theta} = r.$$

PROPOSIZIONE 3.4.3. *Per ogni numero complesso z si ha:*

(1) $|z| \geq 0$ *e l'uguaglianza vale se e solo se* $z = 0$,
(2) $|z| = |\overline{z}|$,
(3) $|\Re z| \leq |z|, \quad |\Im z| \leq |z|$,
(4) $z\overline{z} = |z|^2$,
(5) *Se* $z \neq 0$, *allora* $z^{-1} = \dfrac{\overline{z}}{|z|^2}$.

DIMOSTRAZIONE. Se $z = a + ib$ allora

$$|\Re z| = |a| = \sqrt{a^2} \leq \sqrt{a^2 + b^2} = |z|, \quad |\Im z| = |b| = \sqrt{b^2} \leq \sqrt{a^2 + b^2} = |z|.$$

Le altre proprietà sono di immediata verifica e lasciate per esercizio. □

La prossima proposizione elenca il comportamento del modulo rispetto alla somma ed al prodotto.

PROPOSIZIONE 3.4.4. *Siano z, w due numeri complessi, allora:*

$$|zw| = |z|\,|w|\,, \qquad |z + w| \leq |z| + |w|\,, \qquad |z| - |w| \leq |z - w|.$$

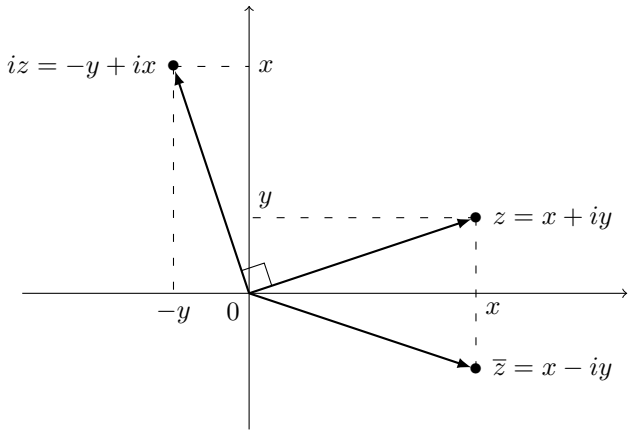


FIGURA 3.4. Geometricamente, il coniugio corrisponde alla riflessione rispetto all'asse reale e la moltiplicazione per l'unità immaginaria corrisponde alla rotazione di 90 gradi in senso antiorario.

DIMOSTRAZIONE. Per quanto riguarda le prime due relazioni, tutte le quantità sono numeri reali non negativi, possiamo quindi elevare tutto al quadrato e dimostrare che $|zw|^2 = |z|^2|w|^2$, $|z+w|^2 \leq |z|^2 + |w^2| + 2|z||w|$. Possiamo scrivere

$$|zw|^2 = zw\,\overline{zw} = zw\,\overline{z}\,\overline{w} = z\,\overline{z}\,w\overline{w} = |z|^2|w|^2,$$

$$|z+w|^2 = (z+w)(\overline{z}+\overline{w}) = |z|^2 + |w|^2 + z\overline{w} + \overline{z}w = |z|^2 + |w|^2 + 2\Re(z\overline{w}).$$

Dato che la parte reale è sempre minore od uguale del modulo si ha

$$\Re(z\overline{w}) \leq |z\overline{w}| = |z||\overline{w}| = |z||w|$$

e quindi

$$|z+w|^2 = |z|^2 + |w|^2 + 2\Re(z\overline{w}) \leq |z|^2 + |w|^2 + 2|z|w|.$$

Per dimostrare che $|z| - |w| \leq |z-w|$ basta scrivere $u = z - w$ e usare la relazione $|u+w| \leq |u| + |w|$.

□

Ogni numero complesso $z = x+iy$ è determinato dal suo modulo $|z|$ e dal suo **argomento** θ che indica l'angolo *in radianti* che il vettore z forma con l'asse delle ascisse (Figura 3.5). Più precisamente

$$x = |z|\cos\theta, \quad y = |z|\sin\theta.$$

Viceversa ogni numero complesso non nullo determina il suo argomento a meno di multipli interi di 2π.

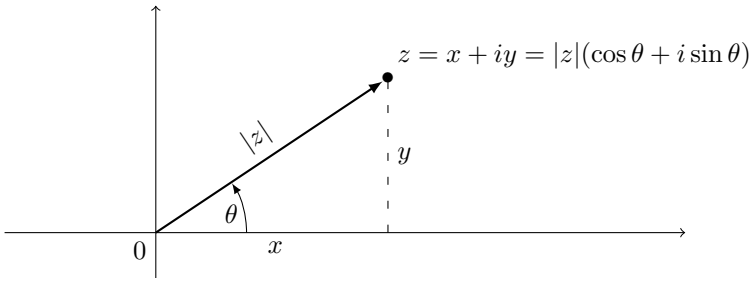


FIGURA 3.5. Modulo ed argomento di un numero complesso

DEFINIZIONE 3.4.5. I numeri reali $|z|, \theta$ sono dette le **coordinate polari** del del numero complesso z corrispondente, determinato dalla formula:

$$z = |z|(\cos\theta + i\sin\theta).$$

Possiamo adesso rappresentare graficamente le operazioni di somma e prodotto di numeri complessi (Figura 3.6): infatti alla somma di due numeri complessi corrisponde la somma delle parti reali e immaginarie, mentre per il prodotto di

$$z = |z|(\cos\theta + i\sin\theta), \qquad w = |w|(\cos\eta + i\sin\eta),$$

abbiamo, grazie alle ben note formule trigonometriche,

$$\begin{aligned} zw =& |z|(\cos\theta + i\sin\theta)|w|(\cos\eta + i\sin\eta) \\ =& |z||w|[(\cos\theta\cos\eta - i\sin\theta\sin\eta) + i(\cos\theta\sin\eta + \sin\theta\cos\eta)] \\ =& |z||w|[\cos(\theta+\eta) + i\sin(\theta+\eta)] \end{aligned} \tag{3.5}$$

Quindi il prodotto zw è il numero complesso che ha come modulo il prodotto dei moduli e come argomento la somma degli argomenti, a meno di multipli di 2π. Similmente si ha la formula

$$\frac{z}{w} = \frac{|z|}{|w|}[\cos(\theta-\eta) + i\sin(\theta-\eta)].$$

Notiamo che se $|z|, \theta$ sono le coordinate polari di z, allora $|z|, -\theta$ sono le coordinate polari di $\overline{z}$ e $|z|^{-1}, -\theta$ sono le coordinate polari di z^{-1}.

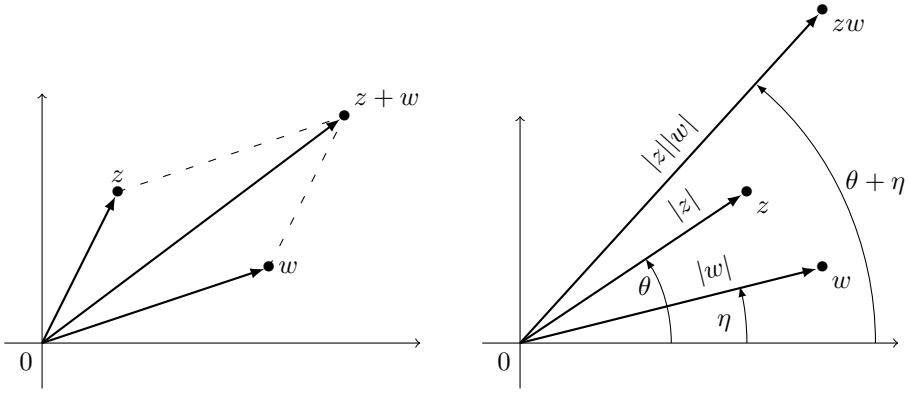


FIGURA 3.6. Rappresentazione geometrica di somma e prodotto; si noti come la disuguaglianza $|z+w| \le |z|+|w|$ equivale al fatto che in un triangolo la lunghezza di un lato non supera la somma delle lunghezze degli altri due.

Diremo che un numero complesso z è rappresentato in **forma polare** o **trigonometrica** se è definito in funzione di modulo e argomento, ossia $z = |z|(\cos\theta + i\sin\theta)$. Diremo che un numero complesso z è rappresentato in **forma algebrica** o **cartesiana** se è definito in funzione di parte reale ed immaginaria, ossia $z = a + ib$.

Esercizi.

142. Calcolare il modulo e la parte reale di

$$(1+i)^3(1-i)^3, \qquad (1+i)^{476}\left(\frac{1-i}{2}\right)^{476}.$$

143. Scrivere in coordinate polari i numeri $2 + i2\sqrt{3}$ e $3(1+i)$.

144. Descrivere l'insieme del piano formato dai numeri complessi z tali che $\dfrac{iz}{1+iz}$ è un numero reale.

145. Dati tre numeri complessi z_1, z_2, z_3 diversi da 0. Quanto vale la somma degli argomenti di

$$\frac{z_1}{z_2}, \quad \frac{z_2}{z_3}, \quad \frac{z_3}{z_1}.$$

146. Descrivere il luogo $H \subseteq \mathbb{C}$ dei numeri complessi z tali che $|z-i| < |z+i|$.

147. Descrivere il luogo $Q \subseteq \mathbb{C}$ dei numeri complessi $z = a + ib$ tali che $|z^2 - i| < |z^2 + i|$ e $b > 0$.

148. Siano $z_1, \dots, z_n$ numeri complessi. Dimostrare che si ha:

$$|z_1 + \cdots + z_n| \le |z_1| + \cdots + |z_n|.$$

149 (♡). Siano $z, a_1, \dots, a_n$ numeri complessi tali che

$$z^n + a_1 z^{n-1} + a_2 z^{n-2} + \cdots + a_n = 0.$$

Dimostrare che

$$|z| < 2\max\{\sqrt[k]{|a_k|} \mid k = 1, \dots, n\}.$$

150. Provare che per ogni numero complesso z ed ogni numero reale t il prodotto

$$(z - \overline{z})^2 (z-t)^2 (\overline{z} - t)^2$$

è un numero reale ≤ 0.

151 (Disuguaglianza di Abel). Siano z_1, z_2, z_3 numeri complessi e a_1, a_2, a_3 numeri reali tali che $a_1 \ge a_2 \ge a_3 \ge 0$. Dimostrare che vale la disuguaglianza

$$|a_1 z_1 + a_2 z_2 + a_3 z_3| \le a_1 \max(|z_1|, |z_1 + z_2|, |z_1 + z_2 + z_3|).$$

Più in generale, provare che se $z_1, \dots, z_n$ sono numeri complessi e $a_1 \geq a_2 \geq \cdots \geq a_n \geq 0$ sono numeri reali, allora

$$\left| \sum_{i=1}^{n} a_i z_i \right| \leq a_1 \max_{1 \leq k \leq n} \left| \sum_{i=1}^{k} z_i \right|.$$

3.5. Potenze e radici di numeri complessi

La rappresentazione in coordinate polari del prodotto di numeri complessi trova la sua più fruttifera applicazione nella descrizione delle potenze e delle radici di un numero complesso.

LEMMA 3.5.1 (Formula di de Moivre). *Sia $z = |z|(\cos\theta + i\sin\theta)$ un numero complesso; per ogni intero n si ha*

$$z^n = |z|^n(\cos(n\theta) + i\sin(n\theta)).$$

DIMOSTRAZIONE. È sufficiente iterare la formula (3.5) nel caso specifico $z = w$: più precisamente si può ragionare per induzione assumendo la formula vera per un dato esponente n ed osservando che

$$\begin{aligned} z^{n+1} &= z^n\, z = |z|^n(\cos(n\theta) + i\sin(n\theta)) \cdot |z|(\cos(\theta) + i\sin(\theta)) \\ &= |z|^n|z|(\cos(n\theta+\theta) + i\sin(n\theta+\theta)) = |z|^{n+1}(\cos((n+1)\theta) + i\sin((n+1)\theta)). \end{aligned}$$

□

DEFINIZIONE 3.5.2. Sia z un numero complesso e n un numero naturale positivo, una *radice n-esima di z* è un numero complesso w tale che $w^n = z$.

Nel caso reale noi sappiamo che ogni numero reale x, differente da 0, ha una radice n-esima se n è dispari e due radici n-esime se n è pari e x è positivo; nel caso complesso invece la situazione è, almeno concettualmente, più semplice. Prima di trattare il caso generale è però utile studiare il caso delle radici complesse di 1.

ESEMPIO 3.5.3 (Radici dell'unità, seconda parte). Per ogni intero positivo n denotiamo con μ_n l'insieme dei numeri complessi z tali che $z^n = 1$, ossia l'insieme delle radici n-esime di 1. L'insieme μ_n è certamente non vuoto in quanto contiene 1. Inoltre la formula di de Moivre ci permette di affermare che gli n numeri complessi

$$\xi_k = \cos\left(\frac{2k\pi}{n}\right) + i\sin\left(\frac{2k\pi}{n}\right), \qquad k = 0, 1, \dots, n-1,$$

appartengono tutti a μ_n: infatti per ogni k vale

$$\xi_k^n = \cos\left(n\frac{2k\pi}{n}\right) + i\sin\left(n\frac{2k\pi}{n}\right) = 1.$$

Si noti che nel piano di Gauss, i numeri complessi ξ_k corrispondono ai vertici di un poligono regolare di n lati inscritto nel cerchio unitario e con un vertice in 1.

Mostriamo adesso che ogni radice dell'unità è uguale a ξ_k per qualche k, ossia che

$$\mu_n = \{\xi_0, \dots, \xi_{n-1}\}.$$

Sia z un numero complesso tale che $z^n = 1$, allora $|z|^n = 1$ da cui segue che $|z| = 1$ e quindi possiamo scrivere

$$z = \cos(\theta) + i\sin(\theta), \qquad 0 \leq \theta < 2\pi.$$

Sia k l'unico intero tale che

$$\frac{2k\pi}{n} \leq \theta < \frac{2(k+1)\pi}{n};$$

siccome $0 \leq \theta < 2\pi$ si ha $0 \leq k < n$. Se consideriamo il numero complesso $w = z/\xi_k$ si ha

$$w = \frac{z}{\xi_k} = \cos(\alpha) + i\sin(\alpha), \qquad 0 \leq \alpha = \theta - \frac{2k\pi}{n} < \frac{2\pi}{n},$$

e quindi

$$w^n = \cos(n\alpha) + i\sin(n\alpha), \qquad 0 \leq n\alpha < 2\pi.$$

D'altra parte $w^n = z^n/\xi_k^n = 1/1 = 1$ e questo è possibile se e solo se $n\alpha = 0$, ossia se e solo se $\alpha = 0$ e $z = \xi_k$.

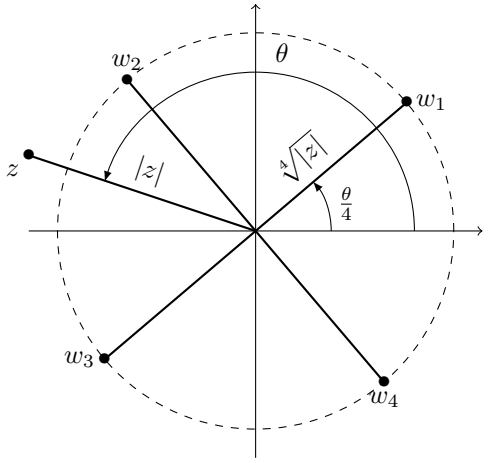


FIGURA 3.7. Un numero complesso z e le sue radici quarte w_1, w_2, w_3, w_4.

DEFINIZIONE 3.5.4. Sia n un intero positivo. Una radice dell'unità z di dice **primitiva di ordine** n se n è il più piccolo intero positivo tale che $z^n = 1$.

Lasciamo al lettore il compito di dimostrare che, nelle notazioni precedenti, la radice $\xi_k \in \mu_n$ è primitiva di ordine n se e solo se n e k non hanno fattori comuni maggiori di 1.

ESEMPIO 3.5.5. I numeri complessi 1, i, -1, $-i$ sono radici quarte dell'unità. Essi hanno tutti modulo 1 e corrispondono ai valori 0, $\dfrac{\pi}{2}$, π, $\dfrac{3\pi}{2}$ dell'angolo θ.

PROPOSIZIONE 3.5.6. *Sia $z \neq 0$ un numero complesso e n un numero naturale positivo, allora esistono esattamente n radici n-esime distinte di z. Se $z = |z|(\cos\theta + i\sin\theta)$ è la forma polare di z, allora le radici n-esime di z sono esattamente gli n numeri complessi*

$$w_k = |z|^{\frac{1}{n}} \left(\cos\frac{\theta + 2k\pi}{n} + i\sin\frac{\theta + 2k\pi}{n} \right), \quad k = 0, 1, \ldots, n-1. \tag{3.6}$$

DIMOSTRAZIONE. Che i numeri w_k siano radici n-esime di z segue immediatamente dalla formula di de Moivre:

$$w_k^n = (|z|^{\frac{1}{n}})^n \left(\cos n\frac{\theta + 2k\pi}{n} + i\sin n\frac{\theta + 2k\pi}{n} \right) = |z| \left(\cos(\theta + 2k\pi) + i\sin(\theta + 2k\pi) \right) = z.$$

Consideriamo gli n angoli

$$\theta_k = \frac{\theta + 2k\pi}{n}, \qquad k = 0, 1, \ldots, n-1.$$

Siccome $|\theta_h - \theta_k| < 2\pi$ per ogni $0 \leq h, k < n$ si ha che i numeri $w_0, \ldots, w_{n-1}$ sono tutti distinti tra loro.

Rimane da dimostrare che ogni radice n-esima di z è uguale a w_k per qualche k. Sia $w = |w|(\cos\phi + i\sin\phi)$, $0 \leq \phi < 2\pi$, una radice n-esima di z:

$$|z|(\cos\theta + i\sin\theta) = z = w^n = |w|^n(\cos n\phi + i\sin n\phi),$$

da cui segue che $|w| = |z|^{1/n}$ e che $n\phi - \theta$ è un multiplo intero di 2π. Siccome $-2\pi < n\phi - \theta < 2n\pi$ si deve avere $n\phi - \theta = 2k\pi$ per qualche $k = 0, \ldots, n-1$ e quindi $\phi = \theta_k$.

Per una dimostrazione alternativa si può osservare che se $w^n = z$ allora w/w_0 è una radice n-esima di 1, poiché $(w/w_0)^n = w^n/w_0^n = z/z = 1$. Nelle notazioni dell'Esempio 3.5.3 esiste $k = 0, \ldots, n-1$ tal che

$$w/w_0 = \xi_k = \cos\left(\frac{2k\pi}{n}\right) + i\sin\left(\frac{2k\pi}{n}\right)$$

e quindi $w = w_0\xi_k = w_k$. □

ESEMPIO 3.5.7. Calcoliamo le radici quadrate del numero complesso $z = 2 + 2\sqrt{3}i$. Il modulo e la forma polare di z è $|z| = \sqrt{4+12} = 4$ e quindi

$$\frac{z}{|z|} = \frac{1}{2} + i\frac{\sqrt{3}}{2}.$$

Riconosciamo immediatamente che $1/2$ e $\sqrt{3}/2$ sono rispettivamente coseno e seno di $\pi/3$. Applicando la formula (3.6) abbiamo quindi che le radici quadrate di z sono

$$w_k = |z|^{\frac{1}{2}} \left(\cos \frac{\pi/3 + 2k\pi}{2} + i \sin \frac{\pi/3 + 2k\pi}{2} \right), \quad k = 0, 1,$$

ossia

$$w_0 = 2(\cos \pi/6 + i \sin \pi/6) = \sqrt{3} + i, \qquad w_1 = -w_0 = -\sqrt{3} - i.$$

Esercizi.

152. Calcolare le radici terze dell'unità, ossia tutti i numeri complessi z tali che $z^3 = 1$.

153. Calcolare le radici terze del numero complesso $z = 8i$.

154. Per ogni intero positivo n, calcolare i numeri complessi

$$\frac{i^{4n}(1-i)^{4n}}{4^n}, \qquad \frac{i^{4n}(1+i)^{4n}}{4^n}.$$

155. Siano $z_1 \neq z_2$ le due radici quadrate del numero complesso $3 - 4i$. Scrivere nella forma $a + ib$ il numero

$$\frac{z_1 + z_2 + 1 + i}{1 + 2i} + z_1 z_2.$$

156. Risolvere le seguenti equazioni nella variabile complessa z:

$$z^2 - z - iz + i = 0, \quad z^2 = 5 + 12i, \quad |z|\overline{z} = 2i, \quad z^3 = 1 - i, \quad z^4 = \overline{z}^3,$$

$$z^3 = iz\overline{z}, \quad z^2 + |z|^2 = 1 + i, \quad z^4 - (2i+1)z^2 + 2i = 0, \quad \overline{z}^3 + z^2 = 0,$$

$$z^2 + iz + \overline{z} = 0, \quad \Re(z^2) = z + i, \quad z|z|^2 = 2\Re(z), \quad z^2 + 2\overline{z} = |z|^2, \quad z^2\overline{z} = 1 + i.$$

157. Siano $\xi_0, \dots, \xi_{n-1}$ le radici n-esime di 1. Provare che per ogni indice $h = 0, \dots, n-1$ esiste un'applicazione bigettiva $\sigma_h \colon \{0, \dots, n-1\} \to \{0, \dots, n-1\}$ tale che

$$\xi_h \xi_i = \xi_{\sigma_h(i)} \quad \text{per ogni} \quad i = 0, \dots, n-1.$$

158. Sia $\xi \in \mathbb{C}$ tale che $\xi^2 + \xi + 1 = 0$. Provare che:
(1) $\xi^3 = 1$, $(1+\xi)^3 = (1+\xi^2)^3 = -1$;
(2) dati due numeri complessi u, v, i tre numeri

$$x = u + v, \quad y = \xi u + \xi^2 v, \quad z = \xi^2 u + \xi v,$$

sono uguali se e solo se $u = v = 0$, mentre sono a due a due distinti se e solo se $u^3 \neq v^3$.

159 (☕). Sia G un insieme finito di n numeri complessi non nulli, con $n > 0$, e tali che per ogni $z, w \in G$ si ha $zw \in G$. Dimostrare che $G = \mu_n$.

Suggerimento: per ogni $z \in G$ esistono due interi $0 < a < b$ tali che $z^a = z^b \neq 0$. Dedurre che $|z| = 1$, che $1 \in G$ e che per ogni $z, w \in G$ si ha $z/w \in G$. Ordinare gli elementi di G per argomento crescente, ossia $G = \{z_0, z_1, \dots, z_{n-1}\}$ con

$$z_i = \cos(\alpha_i) + i \sin(\alpha_i), \qquad 0 = \alpha_0 < \alpha_1 < \alpha_2 < \cdots < \alpha_{n-1} < 2\pi,$$

e provare che $\alpha_k = k\alpha_1$ per ogni $0 \leq k < n$ e che $n\alpha_1 = 2\pi$.

3.6. Campi di numeri

DEFINIZIONE 3.6.1. Un **sottocampo** di $\mathbb{C}$, detto anche **campo di numeri**, è un sottoinsieme $\mathbb{K} \subseteq \mathbb{C}$ che contiene 0, contiene 1 e che è chiuso per le operazioni di somma, prodotto, opposto ed inverso di numeri diversi da 0.

In altri termini, un campo di numeri è un sottoinsieme $\mathbb{K} \subseteq \mathbb{C}$ che soddisfa le seguenti proprietà:

(1) $0 \in \mathbb{K}$ e $1 \in \mathbb{K}$.
(2) Se $z, w \in \mathbb{K}$, allora $z + w, zw \in \mathbb{K}$.
(3) Se $z \in \mathbb{K}$ allora $-z \in \mathbb{K}$ e, se $z \neq 0$, allora anche $z^{-1} \in \mathbb{K}$.

Ad esempio $\mathbb{Q}, \mathbb{R}$ e $\mathbb{C}$ sono campi di numeri e, se p è un numero primo, allora anche l'insieme

$$\mathbb{Q}(\sqrt{p}) = \{a + b\sqrt{p} \in \mathbb{C} \mid a, b \in \mathbb{Q}\}$$

è un campo di numeri.

ESEMPIO 3.6.2. L'insieme $\mathbb{Q}(i) = \{a + ib \mid a, b \in \mathbb{Q}\}$ dei numeri complessi aventi parte reale ed immaginaria razionali è un campo di numeri.

LEMMA 3.6.3. *Ogni sottocampo di $\mathbb{C}$ contiene tutti i numeri razionali e quindi contiene infiniti elementi.*

DIMOSTRAZIONE. Sia $\mathbb{K} \subseteq \mathbb{C}$ un sottocampo, siccome $1 \in \mathbb{K}$, allora per ogni intero positivo n si ha

$$n = \underbrace{1 + 1 + \cdots + 1}_{n \text{ addendi}} \in \mathbb{K}$$

e quindi per ogni coppia di interi positivi n, m si ha

$$-n \in \mathbb{K}, \quad \frac{n}{m} \in \mathbb{K}, \quad \frac{-n}{m} \in \mathbb{K}.$$

□

Possiamo generalizzare la costruzione di $\mathbb{Q}(\sqrt{p})$ nel modo seguente. Sia $\mathbb{K} \subseteq \mathbb{C}$ un campo di numeri e sia α un numero complesso tale che $\alpha^2 \in \mathbb{K}$. Allora l'insieme

$$\mathbb{K}(\alpha) = \{a + b\alpha \mid a, b \in \mathbb{K}\}$$

è ancora un campo di numeri; per dimostrarlo occorre trattare separatamente i due casi $\alpha \in \mathbb{K}$ e $\alpha \notin \mathbb{K}$. Se $\alpha \in \mathbb{K}$ allora, siccome $\mathbb{K}$ è un campo, ne segue che ogni elemento del tipo $a + b\alpha$ appartiene ancora a $\mathbb{K}$, ossia $\mathbb{K}(\alpha) = \mathbb{K}$. Se $\alpha \notin \mathbb{K}$ allora, per ogni $a, b \in \mathbb{K}$ non entrambi nulli i due numeri $a + b\alpha, a - b\alpha$ sono diversi da 0 e quindi anche il loro prodotto $(a + b\alpha)(a - b\alpha)$ non è nullo. Ne consegue che

$$\frac{1}{a + b\alpha} = \frac{a}{a^2 - b^2\alpha^2} - \frac{b}{a^2 - b^2\alpha^2}\alpha \in \mathbb{K}(\alpha).$$

DEFINIZIONE 3.6.4. Nelle notazioni precedenti se $\alpha^2 \in \mathbb{K}$ e $\alpha \notin \mathbb{K}$ il campo $\mathbb{K}(\alpha)$ viene detto **estensione quadratica** di $\mathbb{K}$.

ESEMPIO 3.6.5. Un numero intero positivo m si dice ridotto se non è divisibile per quadrati di numeri primi. Sono ridotti i numeri $2, 3, 5, 6, 7, 10, 11, 13, 14 \ldots$, mentre non sono ridotti i numeri $4, 8, 9, 12, \ldots$; se m è ridotto, allora $\sqrt{m}$ non è razionale e quindi

$$\mathbb{Q}(\sqrt{m}) = \{a + b\sqrt{m} \mid a, b \in \mathbb{Q}\}$$

è un'estensione quadratica di $\mathbb{Q}$.

ESEMPIO 3.6.6. Sia $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, $\ldots p_n$, $\ldots$ la successione di tutti i numeri primi in ordine crescente e consideriamo la successione di sottocampi:

$$F_0 = \mathbb{Q}, \quad F_1 = \mathbb{Q}(\sqrt{p_1}) = \mathbb{Q}(\sqrt{2}), \quad F_2 = F_1(\sqrt{p_2}) = \mathbb{Q}(\sqrt{2}, \sqrt{3}), \ \ldots \ ,$$
$$F_n = F_{n-1}(\sqrt{p_n}) \qquad \forall\, n > 0.$$

Allora $F_n \neq F_{n-1}$ per ogni $n > 0$, ossia $\sqrt{p_n} \notin F_{n-1}$ per ogni $n > 0$.

Volendo dimostrare che $\sqrt{p_n} \notin F_{n-1}$ per induzione, l'autore si è accorto, dopo alcuni infruttuosi tentativi, che la proprietà $\sqrt{p_n} \notin F_{n-1}$ non è abbastanza forte per dimostrare il

passo induttivo. Allora, in maniera piuttosto controintuitiva, dimostriamo per induzione un enunciato più forte ed apparentemente difficile da dimostrare. Più precisamente, dimostriamo per induzione su n che valgono le proprietà:

P_n**:** *$\sqrt{p_n} \notin F_{n-1}$ e se $x \in F_n$ è un numero tale che $x^2 \in \mathbb{Q}$, allora $x = a\sqrt{m}$ con $a \in \mathbb{Q}$ e m intero positivo che divide il prodotto $p_1 p_2 \cdots p_n$ dei primi n numeri primi.*

La validità di P_1 è semplice ed è lasciata per esercizio. Supponiamo quindi $n > 1$ e che valga la proprietà P_{n-1}. Se fosse $\sqrt{p_n} \in F_{n-1}$, allora per la proprietà P_{n-1} si avrebbe $\sqrt{p_n} = a\sqrt{m}$, dove $a \in \mathbb{Q}$ e m divide il prodotto $p_1 \cdots p_{n-1}$. Moltiplicando per $\sqrt{p_n}$ otteniamo $\sqrt{mp_n} \in \mathbb{Q}$ che sappiamo essere falso.

Sia adesso $x = u + v\sqrt{p_n} \in F_n$, dove $u, v \in F_{n-1}$, un numero tale che

$$x^2 = u^2 + v^2 p_n + 2uv\sqrt{p_n} \in \mathbb{Q}.$$

Abbiamo già provato che $\sqrt{p_n} \notin F_{n-1}$ e quindi $uv = 0$, altrimenti si avrebbe

$$\sqrt{p_n} = \frac{x^2 - u^2 - v^2}{2uv} \in F_{n-1}.$$

Se $v = 0$ allora $x = u \in F_{n-1}$ e per la proprietà P_{n-1} si avrebbe $u = a\sqrt{m}$. Se $u = 0$ allora $v^2 = x^2/p_n \in \mathbb{Q}$ e sempre per P_{n-1} si ha $v = a\sqrt{m}$ e $x = a\sqrt{mp_n}$.

Esercizi.

160. Dimostrare che un sottoinsieme $\mathbb{K}$ di $\mathbb{C}$ è un campo di numeri se e solo se soddisfa le seguenti condizioni:
   (1) $1 \in \mathbb{K}$;
   (2) se $z, w \in \mathbb{K}$ e $w \neq 0$ allora $z - w, \dfrac{z}{w} \in \mathbb{K}$.

161. Sia $\xi \in \mathbb{C}$ tale che $\xi^2 + \xi + 1 = 0$. Provare che

$$\mathbb{Q}(\xi) = \{a + b\xi \in \mathbb{C} \mid a, b \in \mathbb{Q}\}$$

è un campo di numeri. Scrivere l'inverso di $1 + \xi$ nella forma $a + b\xi$.

162. Mostrare che l'intersezione di sottocampi di $\mathbb{C}$ è ancora un sottocampo di $\mathbb{C}$.

163. Mostrare che $\mathbb{C}$ è l'unica estensione quadratica di $\mathbb{R}$.

164. Dimostrare che $\sqrt{2} + \sqrt{3} + \sqrt{5} + \sqrt{8}$ non è razionale.

3.7. Campi, polinomi e funzioni razionali

Abbiamo appena visto che i numeri razionali, reali e complessi sono dotati di due operazioni, la somma $+$ e il prodotto $\cdot$ che godono delle seguenti proprietà:

Proprietà della somma:

(S1) La somma è associativa: per ogni x, y, z si ha

$$(x + y) + z = x + (y + z).$$

(S2) La somma è commutativa: per ogni x, y si ha

$$x + y = y + x.$$

(S3) Esiste un elemento, denotato 0, che è neutro per la somma: ciò significa che per ogni x si ha

$$x + 0 = 0 + x = x.$$

(S4) Ogni elemento ha un opposto: per ogni x esiste $-x$ tale che

$$x + (-x) = 0.$$

Proprietà del prodotto:

(P1) Il prodotto è associativo: per ogni x, y, z si ha

$$(x \cdot y) \cdot z = x \cdot (y \cdot z).$$

(P2) Il prodotto è commutativo: per ogni x, y si ha

$$x \cdot y = y \cdot x.$$

(P3) Esiste un elemento, denotato 1, che è neutro per il prodotto, ossia per ogni x si ha
$$x \cdot 1 = 1 \cdot x = x.$$
(P4) Ogni elemento diverso da 0 ha un inverso: per ogni x esiste x^{-1} tale che
$$x \cdot x^{-1} = 1.$$

Proprietà distributiva:

(D) Il prodotto è distributivo rispetto alla somma: per ogni x, y, z si ha
$$x \cdot (y + z) = x \cdot y + x \cdot z.$$

DEFINIZIONE 3.7.1. Una terna formata da un insieme $\mathbb{K}$ e da due operazioni
$$\mathbb{K} \times \mathbb{K} \xrightarrow{+} \mathbb{K} \text{ (somma)}, \qquad \mathbb{K} \times \mathbb{K} \xrightarrow{\cdot} \mathbb{K} \text{ (prodotto)},$$
che godono delle precedenti 9 proprietà (S1)–(S4), (P1)–(P4), (D) si dice un **campo**.

Dalle precedenti 9 proprietà si deduce facilmente che:

(1) l'elemento neutro per la somma è unico. Infatti se $u \in \mathbb{K}$ è un elemento tale che $x + u = x$ per ogni x, in particolare si ha $0 + u = 0$, mentre dalla proprietà (S3) segue che $u + 0 = u$, da cui segue $u = u + 0 = 0$;
(2) l'opposto di un elemento è unico. Infatti se $u \in \mathbb{K}$ è un elemento tale che $x + u = 0$, allora $u = u + 0 = u + x + (-x) = 0 + (-x) = -x$. In particolare x è l'opposto di $-x$, ossia $x = -(-x)$;
(3) l'elemento neutro per il prodotto è unico. Infatti se $e \in \mathbb{K}$ è un elemento tale che $e \cdot x = x$ per ogni x, in particolare si ha $e \cdot 1 = 1$, mentre dalla proprietà (P3) segue che $e \cdot 1 = e$ e quindi $e = 1$;
(4) l'inverso di di un elemento $x \neq 0$ è unico. Infatti se $u \in \mathbb{K}$ è un elemento tale che $xu = 1$, allora $u = u(xx^{-1}) = (ux)x^{-1} = x^{-1}$. In particolare x è l'inverso di x^{-1}, ossia $x = (x^{-1})^{-1}$;
(5) $x \cdot 0 = 0$ per ogni x: infatti
$$0 = x + (-x) = x \cdot (1 + 0) + (-x) = x \cdot 1 + x \cdot 0 + (-x) = x + x \cdot 0 + (-x) = x \cdot 0;$$
(6) $(-x)y = -xy$ e $(-x)(-y) = xy$ per ogni x, y: infatti
$$-xy = -xy + (x + (-x))y = -xy + xy + (-x)y = (-x)y,$$
$$(-x)(-y) = -x(-y) = -(-xy) = xy;$$
(7) un prodotto si annulla, ossia è uguale a 0, se e solo se si annulla almeno uno dei fattori: se $x \cdot y = 0$ e ad esempio $y \neq 0$ si ha
$$0 = 0 \cdot y^{-1} = (x \cdot y) \cdot y^{-1} = x \cdot (y \cdot y^{-1}) = x \cdot 1 = x.$$
Più in generale, se $n > 2$ e $a_1, \dots, a_n \in \mathbb{K}$ sono tutti diversi da 0 allora $a_1 a_2 \neq 0$ ed un ragionamento per induzione su n mostra che $a_1 a_2 \cdots a_n = (a_1 a_2) a_3 \cdots a_n \neq 0$.

l'inverso del prodotto è uguale al prodotto degli inversi: dati $a_1, \dots, a_n \in \mathbb{K}$ non nulli si ha
$$a_n^{-1} \cdots a_2^{-1} \underbrace{a_1^{-1} a_1}_{=1} a_2 \cdots a_n = a_n^{-1} \cdots \underbrace{a_2^{-1} a_2}_{=1} \cdots a_n = \cdots = 1$$
e quindi $(a_1 a_2 \cdots a_n)^{-1} = a_n^{-1} \cdots a_2^{-1} a_1^{-1}$.

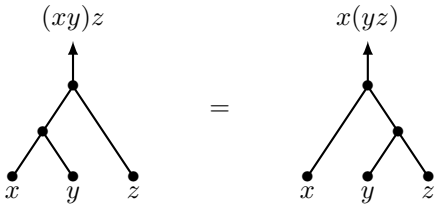


FIGURA 3.8. Rappresentazione grafica della proprietà associativa $(xy)z = x(yz)$.

Gli assiomi di campo non escludono l'ipotesi che $1 = 0$, ossia che l'elemento neutro per la somma coincide con l'elemento neutro per il prodotto; abbiamo però dimostrato in (5) che in tal caso $x = x \cdot 1 = x \cdot 0 = 0$ per ogni x, ossia che il campo contiene il solo elemento 0.

Per semplificare le notazioni, usualmente, il simbolo del prodotto viene omesso e si scrive xy per $x \cdot y$, oltre ad usare comunemente le notazioni:

$$x - y = x + (-y), \qquad \frac{x}{y} = x \cdot y^{-1}.$$

È chiaro che $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ sono campi, mentre $\mathbb{Z}$ non è un campo; infatti su $\mathbb{Z}$ la proprietà (P4) non è soddisfatta per ogni intero diverso da ± 1. Ogni campo di numeri è in particolare un campo. Esistono tuttavia moltissimi campi che non sono sottocampi di $\mathbb{C}$.

Esempio 3.7.2. L'insieme $\mathbb{F}_2 = \{0, 1\}$ formato dai due elementi 0 e 1, con, oltre le usuali, l'ulteriore proprietà che $1+1 = 0$, è un campo. Questo prova in particolare che esistono campi con un numero finito di elementi.

Ha quindi senso dividere i campi in *campi finiti*, che possiedono un numero finito di elementi, e *campi infiniti*. Ogni sottocampo di $\mathbb{C}$ contiene i numeri razionali ed è pertanto infinito.

Un'altra utile distinzione è quella tra i campi in cui $1 \neq 0$ e $1 + 1 = 0$, detti di **caratteristica** 2, ed i rimanenti, detti di caratteristica diversa da 2.

Avvertenza: Lo studente che studia per la prima volta algebra lineare può limitarsi a intendere i campi esclusivamente come sottocampi di $\mathbb{C}$, che sono tutti infiniti e di caratteristica diversa da 2. Tuttavia, al fine di avere un testo utilizzabile anche da lettori più esperti, già avvezzi al linguaggio algebrico abbiamo deciso di scrivere gran parte degli enunciati per un campo arbitrario, aggiungendo quando necessario le ipotesi sulla caratteristica e sul numero di elementi.

Sia x un simbolo formale, un **polinomio** nella variabile x a coefficienti in un campo $\mathbb{K}$ è un'espressione del tipo

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0, \quad a_i \in \mathbb{K},\ n \in \mathbb{N}.$$

Gli elementi $a_i \in \mathbb{K}$ sono i coefficienti del polinomio $p(x)$.

Il fatto che x sia considerato un simbolo formale implica che non esistono relazioni tra le potenze di x, diversamente da quello che accade con l'unità immaginaria i, soggetta alla relazione $i^2 + 1 = 0$. Due espressioni formali

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0, \quad q(x) = b_m x^m + b_{m-1} x^{m-1} + \cdots + b_1 x + b_0,$$

rappresentano lo stesso polinomio se e solo se $a_i = b_i$ per ogni i, dove si conviene che $a_i = b_j = 0$ per $i > n$ e $j > m$.

L'insieme dei polinomi a coefficienti in un campo $\mathbb{K}$ nella variabile x viene indicato con $\mathbb{K}[x]$. Con le usuali regole algebriche tra gli elementi di $\mathbb{K}[x]$ possono essere definite le operazioni di somma e prodotto. Ad esempio:

$$(x^3 + x^2 - x + 1) + (2x^2 - 3x - 6) = x^3 + 3x^2 - 4x - 5,$$

$$(x^2 + 2)(3x^3 - x + 2) = 3x^5 - x^3 + 2x^2 + 6x^3 - 2x + 4 = 3x^5 + 5x^3 + 2x^2 - 2x + 4.$$

Il polinomio nullo è il polinomio che ha tutti i coefficienti uguali a 0; ogni polinomio non nullo si scrive in maniera unica nella forma

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0, \quad \text{con} \quad a_n \neq 0. \tag{3.7}$$

In tal caso chiameremo l'intero n **grado** del polinomio $p(x)$, e si scrive $n = \deg(p(x))$, mentre il coefficiente a_n viene detto **coefficiente direttivo**. Osserviamo che il coefficiente direttivo del prodotto di due polinomi è uguale al prodotto dei coefficienti direttivi. In particolare *il prodotto di due polinomi non nulli è ancora non nullo.*

Un polinomio **monico** è un polinomio non nullo il cui coefficiente direttivo è uguale a 1: il polinomio in (3.7) è quindi monico se e solo se $a_n = 1$. Il polinomio monico associato ad un polinomio non nullo $p(t)$ è, per definizione, il quoziente di $p(t)$ per il suo coefficiente direttivo: ad esempio il polinomio monico associato al polinomio in (3.7) è

$$x^n + \frac{a_{n-1}}{a_n} x^{n-1} + \cdots + \frac{a_1}{a_n} x + \frac{a_0}{a_n}.$$

Per convenzione, il grado del polinomio nullo è $-\infty$, ossia quello "strano" simbolo matematico con le proprietà che $-\infty \leq a$ e $-\infty + a = -\infty$ per ogni numero $a \in \mathbb{R}$. Con tale convenzione, se $p(x)$ e $q(x)$ sono due polinomi a coefficienti nello stesso campo si hanno le formule

$$\deg(p(x) + q(x)) \leq \max(\deg(p(x)), \deg(q(x)),$$
$$\deg(p(x)q(x)) = \deg(p(x)) + \deg(q(x)).$$

Notiamo che tutte le proprietà dei campi, eccetto l'esistenza dell'inverso, sono soddisfatte: segue in fatti dalla formula $\deg(p(x)q(x)) = \deg(p(x)) + \deg(q(x))$ che i polinomi di grado 0 sono tutti e soli quelli che possiedono un inverso.

In particolare vale la formula distributiva $p(t)(q(t) + r(t)) = p(t)q(t) + p(t)q(t)$ e se $p(t)q(t) = p(t)r(t)$ allora o $p(t) = 0$ oppure $q(t) = r(t)$: infatti se $p(t)q(t) = p(t)r(t)$ allora $p(t)(q(t) - r(t)) = 0$ ed almeno uno dei due fattori si annulla.

Il lettore non deve fare confusione tra polinomi e **funzioni polinomiali**, definite come le applicazioni $p\colon \mathbb{K} \to \mathbb{K}$ per cui esiste una successione finita $a_0, \ldots, a_n \in \mathbb{K}$ tale che

$$p(\alpha) = a_n\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 \quad \text{per ogni} \quad \alpha \in \mathbb{K}.$$

Ad ogni polinomio corrisponde in modo naturale una funzione polinomiale.

Definizione 3.7.3. Sia $p(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 \in \mathbb{K}[x]$ un polinomio. Un elemento $\alpha \in \mathbb{K}$ si dice una **radice di** $p(x)$ se annulla la corrispondente funzione polinomiale, ossia se

$$p(\alpha) = a_n\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 = 0.$$

Esempio 3.7.4. Siano $\mathbb{K}$ un campo, $a_1, \ldots, a_n \in \mathbb{K}$ e si consideri il polinomio

$$p(x) = (x - a_1)(x - a_2)\cdots(x - a_n) \in \mathbb{K}[x].$$

Allora un elemento $\alpha \in \mathbb{K}$ è una radice di $p(x)$ se e solo se $\alpha = a_i$ per qualche indice i. Infatti, dato $a \in \mathbb{K}$ il prodotto $(a - a_1)(a - a_2)\cdots(a - a_n)$ si annulla se e solo se si annulla almeno uno dei fattori, ossia se e solo se $a - a_i = 0$ per qualche i.

Si noti che $p(x)$ è monico di grado n, e più precisamente

$$p(x) = x^n - \left(\sum\nolimits_{i=1}^{n} a_i\right) x^{n-1} + \cdots,$$

dove i puntini di sospensione stanno ad indicare "roba di grado minore", ossia una somma di addendi del tipo ax^i con $i \leq n - 2$.

Teorema 3.7.5. *Siano $\mathbb{K}$ un campo e $p(x) \in \mathbb{K}[x]$ un polinomio non nullo di grado n. Allora $p(x)$ possiede al più n radici distinte nel campo $\mathbb{K}$.*

Dimostrazione. Prima di procedere alla dimostrazione osserviamo che l'enunciato del teorema è del tutto equivalente a dire che se un polinomio $p(x)$ di grado $\leq n$ possiede almeno $n + 1$ radici distinte, allora $p(x) = 0$ è il polinomio nullo.

Il risultato è chiaramente vero se $n = 0$; dimostriamo il caso generale per induzione su n. Supponiamo che esistano $n + 1$ radici distinte $\alpha_0, \alpha_1, \ldots, \alpha_n$ del polinomio

$$p(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_0, \qquad a_n \neq 0,$$

e consideriamo il polinomio

$$q(x) = a_n(x - \alpha_1)(x - \alpha_2)\cdots(x - \alpha_n) = a_nx^n + \cdots.$$

Se $p(x) \neq q(x)$ allora il polinomio $p(x) - q(x)$ ha grado minore di n e ha $\alpha_1, \ldots, \alpha_n$ come radici, in contraddizione con l'ipotesi induttiva. Quindi $p(x) = q(x)$ e siccome $p(\alpha_0) = 0$ si ha

$$0 = p(\alpha_0) = q(\alpha_0) = a_n(\alpha_0 - \alpha_1)(\alpha_0 - \alpha_2)\cdots(\alpha_0 - \alpha_n).$$

Ma i fattori $a_n, \alpha_0 - \alpha_i$ sono tutti diversi da 0 e quindi otteniamo una contraddizione.

□

Corollario 3.7.6. *Sia $p(x) \in \mathbb{K}[x]$ un polinomio a coefficienti in un campo infinito $\mathbb{K}$. Se $p(a) = 0$ per ogni $a \in \mathbb{K}$, allora $p(x) = 0$ in $\mathbb{K}[x]$, ossia $p(x)$ è il polinomio nullo.*

Dimostrazione. Per ipotesi ogni elemento di $\mathbb{K}$ è una radice del polinomio e quindi, essendo $\mathbb{K}$ infinito, per il Teorema 3.7.5 il polinomio $p(x)$ deve essere nullo. □

Esempio 3.7.7. Ogni polinomio di secondo grado

$$ax^2 + bx + c, \qquad a, b, c \in \mathbb{C}, \quad a \neq 0,$$

a coefficienti complessi possiede radici complesse. Più precisamente esistono $\alpha_+, \alpha_- \in \mathbb{C}$ tali che

$$ax^2 + bx + c = a(x - \alpha_+)(x - \alpha_-).$$

Abbiamo già visto che ogni numero complesso possiede radici quadrate e per risolvere un'equazione di secondo grado basta applicare la formula risolutiva standard

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

Osservazione 3.7.8. Il precedente esempio è un caso particolare del *teorema fondamentale dell'algebra*, un risultato di grande importanza, che dimostremo nella Sezione 9.8 di queste note, il quale afferma che *ogni polinomio di grado positivo a coefficienti complessi possiede radici complesse.*

È utile osservare che se $\lambda \in \mathbb{C}$ è una radice di un polinomio a coefficienti reali, allora anche il numero complesso coniugato $\overline{\lambda}$ è una radice dello stesso polinomio. Infatti, se $p(t) \in \mathbb{R}[t]$, si dimostra immediatamente che per ogni numero complesso λ vale $p(\overline{\lambda}) = \overline{p(\lambda)}$ e quindi $p(\lambda) = 0$ se e solo se $p(\overline{\lambda}) = 0$.

La definizione di sottocampo di $\mathbb{C}$ si estende in modo ovvio a campi generici.

Definizione 3.7.9. Un **sottocampo** di un campo $\mathbb{K}$ è un sottoinsieme $F \subseteq \mathbb{K}$ che contiene 0, contiene 1 e che è chiuso per le operazioni di somma, prodotto, opposto ed inverso di numeri diversi da 0.

Esempio 3.7.10. È utile osservare che ogni campo $\mathbb{K}$ può essere visto come un sottocampo di un campo infinito: ad esempio possiamo ripetere la costruzione dei numeri razionali mettendo i polinomi a coefficienti in $\mathbb{K}$ al posto degli interi. Si definisce quindi il **campo delle funzioni razionali**, denotato $\mathbb{K}(x)$, come l'insieme di tutte la frazioni

$$\frac{p(x)}{q(x)}$$

con $p(x), q(x) \in \mathbb{K}[x]$ e $q(x) \neq 0$. Due frazioni $\dfrac{p(x)}{q(x)}, \dfrac{a(x)}{b(x)}$ definiscono la stessa funzione razionale se e solo se $p(x)b(x) = q(x)a(x)$. Le operazioni di somma e prodotto sono quelle imparate alle scuole superiori:

$$\frac{p(x)}{q(x)} + \frac{a(x)}{b(x)} = \frac{p(x)b(x) + a(x)q(x)}{q(x)b(x)}, \qquad \frac{p(x)}{q(x)}\,\frac{a(x)}{b(x)} = \frac{p(x)a(x)}{q(x)b(x)}.$$

Siccome

$$\mathbb{K} \subseteq \mathbb{K}[x] = \left\{ \frac{p(x)}{1} \right\} \subseteq \mathbb{K}(x)$$

si ha che il campo $\mathbb{K}(x)$ è infinito e contiene $\mathbb{K}$.

Se $\mathbb{K}$ è un campo qualsiasi, possiamo definire un'applicazione $\alpha\colon \mathbb{N} \to \mathbb{K}$ ponendo

$$\alpha(1) = 1, \quad \alpha(2) = 1 + 1, \quad \ldots, \alpha(n) = \underbrace{1 + \cdots + 1}_{n \text{ addendi}}, \ldots$$

Si noti che $\alpha(a + b) = \alpha(a) + \alpha(b)$ e quindi $\alpha(a + b) = \alpha(a)$ se e solo se $\alpha(b) = 0$.

Definizione 3.7.11. Un campo $\mathbb{K}$ si dice:

(1) di **caratteristica** 0 se l'applicazione α è iniettiva,
(2) di **caratteristica positiva** se l'applicazione α non è iniettiva.

Ad esempio, $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ e più in generale tutti i campi di numeri hanno caratteristica 0. Il campo $\mathbb{F}_2$, il campo $\mathbb{F}_2(x)$ e, più in generale, i campi di caratteristica 2 hanno caratteristica positiva.

Per semplicità notazionale, spesso si omette la lettera α e si scrive semplicemente

$$2 = 1 + 1 \in \mathbb{K}, \quad 3 = 1 + 1 + 1 \in \mathbb{K}, \ \ldots$$

È chiaro che $\alpha(n+m) = \alpha(n) + \alpha(m)$ e di conseguenza se $n \leq m$ vale anche $\alpha(m-n) = \alpha(m) - \alpha(n)$. Siccome il prodotto è distributivo rispetto alla somma si verifica immediatamente che $\alpha(n)\alpha(m) = \alpha(nm)$ per ogni coppia di interi non negativi.

Se α non è iniettiva, esistono $n < m$ tali che $\alpha(n) = \alpha(m)$ e quindi $\alpha(m-n) = 0$.

Definizione 3.7.12. Sia $\mathbb{K}$ un campo di caratteristica positiva: il più piccolo intero positivo p tale che $\alpha(p) = 0$ viene detto **caratteristica del campo**.

Lemma 3.7.13. *Sia $\mathbb{K}$ un campo non nullo di caratteristica positiva. Allora la caratteristica di $\mathbb{K}$ è un numero primo.*

Dimostrazione. Indichiamo con p la caratteristica di $\mathbb{K}$. Per ipotesi $1 \neq 0$ e dunque $p \geq 2$. Se $p = ab$ allora $\alpha(a)\alpha(b) = \alpha(p) = 0$ e da ciò segue che $\alpha(a) = 0$ oppure $\alpha(b) = 0$; nel primo caso questo implica che $a \geq p$ e nel secondo che $b \geq p$. □

Esempio 3.7.14. Non è difficile dimostrare che per ogni primo positivo p esistono campi di caratteristica p. L'esempio più semplice è dato dal campo $\mathbb{F}_p = \{0, 1, \dots, p-1\}$ delle classi di resto della divisione per p; le operazioni di somma e prodotto sono fatte prendendo il resto della divisione per p della somma e prodotto tradizionali. Ad esempio, in $\mathbb{F}_{11}$ si hanno le uguaglianze:

$$10 + 2 = 1, \quad 4 \cdot 5 = 9, \quad 6 = -5 \quad \text{ecc.}$$

La verifica che $\mathbb{F}_p$ è effettivamente un campo è omessa e viene rimandata ai corsi di algebra. A titolo esemplativo mostriamo che ogni elemento diverso da 0 possiede un inverso in $\mathbb{F}_p$: per ogni $n = 1, 2, \dots, p-1$ il piccolo teorema di Fermat (Corollario 2.6.4) afferma che p divide $n^p - n = n(n^{p-1} - 1)$. Siccome p non divide n allora deve dividere $n^{p-1} - 1$, ossia il resto della divisione per p di n^{p-1} è uguale ad 1. Abbiamo quindi dimostrato che per ogni $n \in \mathbb{F}_p$, $n \neq 0$, l'inverso n^{-1} esiste ed è uguale alla classe di resto modulo p di n^{p-2}.

Ovviamente ogni campo finito è di caratteristica positiva. Si può dimostrare che ogni campo finito di caratteristica p possiede p^n elementi, per un opportuno intero positivo n. È da notare che se $h > 1$ non è primo, allora l'insieme $\mathbb{Z}/(h) = \{0, 1, \dots, h-1\}$, dotato delle operazioni di somma e prodotto definite dal resto della divisione per h della somma e prodotto tradizionali, non è un campo: infatti se $h = ab$ allora il prodotto $a \cdot b$ si annulla in $\mathbb{Z}/(h)$.

Naturalmente esistono anche campi infiniti di caratteristica positiva, come ad esempio $\mathbb{F}_p(x)$, il campo delle funzioni razionali su $\mathbb{F}_p$.

Esercizi.

165 (morfismi di valutazione). Siano $\mathbb{K}$ un campo ed $a \in \mathbb{K}$ un elemento qualsiasi. L'applicazione

$$e_a \colon \mathbb{K}[x] \to \mathbb{K}, \qquad e_a(p(x)) = p(a),$$

viene detta *morfismo di valutazione* in a. Ad esempio, $e_0(x^2+1) = 1$, $e_1(x^2-2) = -1$ eccetera.

Convincetevi che i morfismi di valutazione commutano con somme e prodotti, ossia che valgono le formule

$$e_a(p(x) + q(x)) = p(a) + q(a) = e_a(p(x)) + e_a(q(x)),$$
$$e_a(p(x)q(x)) = p(a)q(a) = e_a(p(x))e_a(q(x)),$$

per ogni $p(x), q(x) \in \mathbb{K}[x]$ ed ogni $a \in \mathbb{K}$.

166. L'applicazione identità di un campo in sé è una funzione polinomiale? La funzione valore assoluto $\alpha \mapsto |\alpha|$ di $\mathbb{R}$ in sé è polinomiale?

167 (Il semianello tropicale). Per semianello tropicale si intende l'insieme $\mathbb{R}$ dei numeri reali dotato delle due operazioni

$$\oplus, \odot \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}, \qquad a \oplus b = \min(a, b), \qquad a \odot b = a + b.$$ [2]

Mostrare che le operazioni $\oplus, \odot$ soddisfano le proprietà S1,S2,P1,P2,P3,P4 e D, mentre $\oplus$ non soddisfa le proprietà S3 ed S4.

[2] Già noto come *semianello min-plus*, la comunità matematica gli ha cambiato recentemente nome in onore del matematico Brasiliano Imre Simon. Inizialmente studiato in relazione a problemi di combinatoria algebrica, gode attualmente di molta fama essendo alla base della cosiddetta geometria tropicale.

168. Trovare quattro numeri razionali a, b, c, d tali che

$$x^3 + x^2 - 1 = a(x-1)^3 + b(x-1)^2 + c(x-1) + d.$$

(Suggerimento: se $x - 1 = y$ allora $x = y + 1$.)

169. Trovare un polinomio $p(x) \in \mathbb{Q}[x]$ di terzo grado tale che $p(\sqrt{2} + \sqrt{3}) = \sqrt{2}$.

170. Trovare $a \in \mathbb{Q}$ tale che il polinomio $p(x) = x^3 + x^2 + ax - 1$ possiede come radice il numero 3.

171. Per quali valori di $a, b \in \mathbb{Q}$ il polinomio $p(x) = x^3 + bx^2 + ax - 1$ possiede come radice il numero -1?

172. Calcolare le radici complesse del polinomio $4x^2 + 4ix - (1 + 4i)$.

173. Provare che in un campo infinito a polinomi distinti corrispondono funzioni polinomiali distinte.

174. Calcolare le radici complesse del polinomio $x^4 - 2x^3 - 2x - 1$.

175. Sia $C^0(\mathbb{R})$ lo spazio di tutte le funzioni continue $f\colon \mathbb{R} \to \mathbb{R}$. Esibire una funzione continua $a_0\colon \mathbb{R} \to \mathbb{R}$ tale che l'equazione $x^2 + a_0 = 0$ abbia infinite soluzioni distinte in $C^0(\mathbb{R})$.

176 (☕, ♡).

(1) Siano $a_1, \dots, a_n \in \mathbb{C}$ tali che $\sum_{i=1}^n a_i^j = 0$ per ogni $j = 1, \dots, n$. Dimostrare che $a_1 a_2 \cdots a_n = 0$ (Suggerimento: calcolare $\sum_{i=1}^n p(a_i)$, dove $p = (a_1 - t) \cdots (a_n - t) \in \mathbb{C}[t]$.)

(2) Siano $a_1, \dots, a_n \in \mathbb{C}$ e $k \geq 0$ un intero tale che $\sum_{i=1}^n a_i^{k+j} = 0$ per ogni $j = 1, \dots, n$. Dimostrare che $a_1 = \cdots = a_n = 0$.

177 (☕). Dimostrare che il polinomio

$$p(x) = x^{128} - 2x^{127} + 4x^{113} - 8x - 88 \in \mathbb{Q}[x]$$

non possiede radici razionali.

178 (☕). Siano $\xi_k = \cos(2k\pi/n) + i \sin(2k\pi/n) \in \mathbb{C}$, $k = 0, \dots, n-1$, le radici n-esime di 1, ossia le soluzioni dell'equazione $z^n = 1$. Dimostrare che vale

$$z^n - 1 = (z - \xi_0)(z - \xi_1) \cdots (z - \xi_{n-1})$$

e determinare il valore dei prodotti

$$(z + \xi_0)(z + \xi_1) \cdots (z + \xi_{n-1}), \qquad (z - \xi_0^2)(z - \xi_1^2) \cdots (z - \xi_{n-1}^2).$$

179 (☕, ♡). Siano $\xi_0, \dots, \xi_{n-1}$ le radici complesse n-esime di 1. Provare che per ogni intero h non divisibile per n si ha $\xi_0^h + \xi_1^h + \cdots + \xi_{n-1}^h = 0$.

180 (♡). Risolvere il seguente quesito a risposta multipla, dove nulla è implicito e nulla è dato per scontato. Quanto vale 1 diviso 5?

- 2,
- 3,
- 8.

181. Sia $\mathbb{K}$ un campo di caratteristica $p > 0$. Dimostrare che

$$(x + y)^p = x^p + y^p, \qquad (x - y)^p = x^p - y^p,$$

per ogni $x, y \in \mathbb{K}$. Dedurre che l'applicazione $F\colon \mathbb{K} \to \mathbb{K}$, $x \mapsto x^p$, è iniettiva.

182. Mostrare che la funzione polinomiale associata al polinomio $t^p - t \in \mathbb{F}_p[t]$ è identicamente nulla.

183. Un campo si dice **perfetto** se ha caratteristica 0, oppure se ha caratteristica $p > 0$ e l'applicazione $x \mapsto x^p$ è surgettiva (vedi Esercizio 181). Sia p un numero primo: dimostrare che $\mathbb{F}_p$ è un campo perfetto e che $\mathbb{F}_p(x)$ non è perfetto.

184 (☕, ♡). Siano $p > 0$ un numero primo e $a \geq b > 0$ due interi positivi. Dimostrare che:

(1) p divide la differenza tra coefficienti binomiali $\binom{pa}{pb} - \binom{a}{b}$;
(2) se per qualche intero $k > 0$ si ha che p^k divide a ma non divide b, allora p divide il coefficiente binomiale $\binom{a}{b}$.

3.8. Complementi: la formula di Cardano

La formula di Cardano permette di ricondurre il calcolo delle radici di un polinomio di terzo grado alla soluzione di un'equazione di secondo grado ed al calcolo delle radici cubiche di un numero complesso. Consideriamo un polinomio monico

$$x^3 + 3a_1x^2 + 3a_2x + a_3, \qquad a_i \in \mathbb{C},$$

a coefficienti complessi. A meno di sostituire x con $x - a_1$ si può assumere $a_1 = 0$ e ci riconduciamo al calcolo delle radici di un polinomio del tipo

$$x^3 + 3ax + b.$$

Come prima cosa osserviamo che se $a = 0$ allora le radici del polinomio non sono altro che le radici cubiche di $-b$. Possiamo quindi limitarci a considerare il caso in cui $a \neq 0$. Vediamo due distinte procedure di risoluzione: la prima parte dalla semplice osservazione che

$$(u + v)^3 - 3uv(u + v) - (u^3 + v^3) = 0$$

e quindi se u, v risolvono le due equazioni $uv = -a$ e $u^3 + v^3 = -b$, allora $u + v$ è una radice di $x^3 + 3ax + b$. Elevando al cubo $uv = -a$ si ottiene

$$-a^3 = u^3v^3 = -u^3(b + u^3), \qquad u^6 + bu^3 - a^3 = 0, \quad u^3 = \frac{-b \pm \sqrt{b^2 + 4a^3}}{2},$$

e quindi il numero complesso

$$\sqrt[3]{\frac{-b + \sqrt{b^2 + 4a^3}}{2}} + \sqrt[3]{\frac{-b - \sqrt{b^2 + 4a^3}}{2}}$$

è una radice del polinomio $x^3 + 3ax + b$, a condizione che le due radici cubiche siano scelte in modo che il loro prodotto sia $-a$; senza questa condizione troviamo non solo le radici di $x^3 + 3ax + b$ ma anche le radici di $x^3 + 3\tilde{a}x + b$, dove $\tilde{a} \in \mathbb{C}$ è una qualsiasi radice cubica di a^3.

Viceversa, ogni radice è ottenuta in questo modo: infatti se $x^3 + 3ax + b = 0$ possiamo certamente trovare due numeri complessi u, v tali che $uv = -a$, $u + v = x$ che di conseguenza soddisfano la relazione $u^3 + v^3 = -b$. Lasciamo al lettore il compito di provare, usando la precedente formula e l'Esercizio 158, che $x^3 + 3ax + b$ possiede tre radici distinte se e solo se $u^3 \neq v^3$, ossia se e solo se $b^2 + 4a^3 \neq 0$.

Nella seconda procedura di risoluzione, sempre assumendo $a \neq 0$, l'idea è quella di trovare, se esistono, tre numeri complessi t, n, m tali che

$$x^3 + 3ax + b = t(x - n)^3 + (1 - t)(x - m)^3.$$

Ciò equivale a risolvere il sistema

$$\begin{cases} tn + (1-t)m &= 0 \\ tn^2 + (1-t)m^2 &= a \\ tn^3 + (1-t)m^3 &= -b \end{cases}$$

Dalla prima equazione segue che se $n = m$ allora $n = m = 0$ in contraddizione con la seconda equazione e con l'ipotesi $a \neq 0$. Quindi $n - m \neq 0$, possiamo ricavare il valore di t dalla prima equazione e sostituirlo nelle altre due; semplificando si ottiene

$$\begin{cases} \dfrac{m}{m-n} &= t \\ nm &= -a \\ \dfrac{mn^3 - nm^3}{m-n} &= -b. \end{cases}$$

Mettendo nella terza equazione $-a$ al posto di nm e semplificando si ottiene

$$\begin{cases} \dfrac{m}{m-n} &= t \\ nm &= -a \\ n+m &= -\dfrac{b}{a}. \end{cases}$$

Dunque possiamo calcolare n, m risolvendo l'equazione di secondo grado

$$a(x-n)(x-m) = ax^2 + bx - a^2,$$

e di conseguenza

$$n = \frac{-b+\sqrt{b^2+4a^3}}{2a}, \qquad m = \frac{-b-\sqrt{b^2+4a^3}}{2a},$$

osservando che $n \neq m$ se e solo se $b^2+4a^3 \neq 0$. Possiamo riassumere quanto dimostrato nella seguente proposizione.

PROPOSIZIONE 3.8.1. *Siano $a, b \in \mathbb{C}$ tali che $a(b^2+4a^3) \neq 0$. Allora vale la formula*

$$x^3 + 3ax + b = t(x-n)^3 + (1-t)(x-m)^3,$$

dove

$$n = \frac{-b+\sqrt{b^2+4a^3}}{2a}, \qquad m = \frac{-b-\sqrt{b^2+4a^3}}{2a}, \qquad t = \frac{m}{m-n}.$$

Da tale fatto è facile calcolare le radici $\alpha_1, \alpha_2, \alpha_3$ del polinomio

$$x^3 + 3ax + b = t(x-n)^3 + (1-t)(x-m)^3,$$

quando $a(b^2+4a^3) \neq 0$. Infatti l'equazione $t(x-n)^3 + (1-t)(x-m)^3 = 0$ diventa

$$\left(\frac{x-n}{x-m}\right)^3 = \frac{t-1}{t} = \frac{n}{m} = \frac{b-\sqrt{b^2+4a^3}}{b+\sqrt{b^2+4a^3}}$$

le cui soluzioni si calcolano facilmente: dette $\beta_1, \beta_2, \beta_3$ le radici cubiche di n/m si ha

$$\frac{\alpha_i - n}{\alpha_i - m} = \beta_i, \qquad \alpha_i = \frac{n - \beta_i m}{1-\beta_i} = -m(\beta_i + \beta_i^2), \qquad i = 1, 2, 3.$$

Rimane da considerare la situazione in cui $a \neq 0$ e $b^2 + 4a^3 = 0$; in questo caso basta applicare l'identità algebrica (esercizio: verificare)

$$x^3 + 3ax + b = \left(\left(x - \frac{b}{a}\right)\left(x + \frac{b}{2a}\right) + \frac{3}{4a^2}(b^2+4a^3)\right)\left(x + \frac{b}{2a}\right) - \frac{b}{8a^3}(b^2+4a^3)$$

per scoprire che quando $b^2+4a^3 = 0$ le radici sono esattamente $\dfrac{b}{a}, -\dfrac{b}{2a}, -\dfrac{b}{2a}$.

DEFINIZIONE 3.8.2. Il numero complesso $\Delta = -27(b^2+4a^3)$ viene detto **discriminante** del polinomio $x^3 + 3ax + b$.

Il ragionamento esposto ci mostra inoltre che il polinomio $x^3 + 3ax + b$ possiede radici multiple se e solo se $b^2+4a^3 = 0$. Assumendo valido il teorema fondamentale dell'algebra (che dimostreremo più avanti) lo stesso fatto può essere dimostrato più facilmente osservando che se $x^3 + 3ax + b = (x-\alpha_1)(x-\alpha_2)(x-\alpha_3)$, allora

$$\alpha_1 + \alpha_2 + \alpha_3 = 0, \quad \alpha_1\alpha_2 + \alpha_2\alpha_3 + \alpha_3\alpha_1 = 3a, \qquad \alpha_1\alpha_2\alpha_3 = -b,$$

e da tali uguaglianze si deduce facilmente (esercizio) che

$$\Delta = -27(b^2+4a^3) = (\alpha_1-\alpha_2)^2(\alpha_2-\alpha_3)^2(\alpha_3-\alpha_1)^2. \tag{3.8}$$

OSSERVAZIONE 3.8.3. La formula di Cardano sconfina nei numeri complessi anche quando le tre radici sono tutte reali. Ad esempio l'uso della formula di Cardano per il calcolo delle radici del polinomio $x(x-3)(x+3) = x^3 - 9x$ richiede il calcolo dei numeri

$$n = \frac{\sqrt{4(-3)^3}}{-6} = -i\sqrt{3}, \qquad m = \frac{-\sqrt{4(-3)^3}}{-6} = i\sqrt{3},$$

che sono immaginari puri.

Con ragionamenti analoghi è possibile dimostrare che ogni equazione di quarto grado si riconduce alla soluzione di un'equazione di terzo grado e tre equazioni di secondo grado, vedi Esercizio 185. È invece dimostrato che non esiste alcuna formula generale che permette di ricondurre equazioni di grado superiore al quarto ad equazioni di grado inferiore ed estrazioni di radici; tutto ciò va oltre gli obiettivi di questo volume e viene pertanto omesso.

Esercizi.

185. Dato un polinomio monico di quarto grado

$$p(x) = x^4 + 2ax^3 + bx^2 + cx + d$$

a coefficienti complessi, trovare un polinomio $q(x)$ di grado al più 3 tale che $q(\alpha) = 0$ se e solo se $(x^2 + ax + \alpha)^2 - p(x)$ è del tipo $(\beta x + \gamma)^2$ per opportuni $\beta, \gamma \in \mathbb{C}$. Trovata una radice α di $q(x)$ tramite la formula di Cardano ed i corrispondenti numeri β, γ, si ottiene

$$p(x) = (x^2 + ax + \alpha + \beta x + \gamma)(x^2 + ax + \alpha - \beta x - \gamma).$$

186 (☕). Provare che il polinomio $x^3 + 3ax + b$ possiede tre radici reali distinte se e solo se $a, b \in \mathbb{R}$ e $b^2 + 4a^3 < 0$. (Suggerimento: se possiede una radice reale e due complesse coniugate segue da (3.8) che $\Delta < 0$.)

CAPITOLO 4

Spazi e sottospazi vettoriali

Iniziamo la parte vera e propria di algebra lineare introducendo il concetto fondamentale di spazio vettoriale, per poi passare nel prossimo capitolo a quello di applicazione lineare.

Da adesso in poi il simbolo $\mathbb{K}$ indicherà un campo che potrà essere di qualsiasi natura, come ad esempio: un campo di numeri (Definizione 3.6.1), un campo di funzioni razionali (Esempio 3.7.10), un campo di classi di resto (Esempio 3.7.14) ecc. Per non fissare la natura del campo nemmeno a livello linguistico, chiameremo **scalari** gli elementi di $\mathbb{K}$.

I lettori che non hanno ancora maturato una sufficiente padronanza della nozione astratta di campo, potranno (temporaneamente) supporre che $\mathbb{K}$ sia sempre un campo di numeri, ossia un sottocampo di $\mathbb{C}$; a tal fine è utile ricordare che ogni campo di numeri contiene tutti i numeri razionali ed in particolare vale la disuguaglianza $1+1 \neq 0$. Negli esempi numerici e negli esercizi, in assenza di ulteriori indicazioni, supporremo che $\mathbb{K}$ sia un campo di numeri.

4.1. Vettori numerici

Dato un campo $\mathbb{K}$ possiamo considerare il prodotto cartesiano di $\mathbb{K}$ con se stesso un numero finito di volte

$$\mathbb{K}^{(1)} = \mathbb{K}, \quad \mathbb{K}^{(2)} = \mathbb{K} \times \mathbb{K}, \quad \mathbb{K}^{(3)} = \mathbb{K} \times \mathbb{K} \times \mathbb{K}, \quad \dots, \quad \mathbb{K}^{(n)} = \underbrace{\mathbb{K} \times \cdots \times \mathbb{K}}_{n \text{ fattori}}.$$

Ogni elemento di $\mathbb{K}^{(n)}$ è una successione $(a_1, \dots, a_n)$ di n elementi nel campo $\mathbb{K}$. Chiameremo $\mathbb{K}^{(n)}$ lo **spazio dei vettori riga ad n coordinate sul campo $\mathbb{K}$**.

Lo **spazio dei vettori colonna** $\mathbb{K}^n$ è definito in maniera del tutto simile; gli elementi di $\mathbb{K}^n$ sono le successioni in colonna di n scalari:

$$\mathbb{K}^n = \left\{ \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \;\middle|\; a_1, a_2, \dots, a_n \in \mathbb{K} \right\}.$$

Sia i vettori riga che i vettori colonna vengono chiamati genericamente **vettori numerici**, indipendentemente dalla natura del campo $\mathbb{K}$. Esiste una ovvia bigezione tra $\mathbb{K}^n$ e $\mathbb{K}^{(n)}$ che rende tali spazi indistinguibili nella sostanza. Tuttavia, per il momento è utile tenere in considerazione anche la forma e considerare $\mathbb{K}^n$ e $\mathbb{K}^{(n)}$ come due entità distinte.[1] Il passaggio da un vettore riga ad uno colonna, e viceversa, mediante la bigezione naturale[2] viene chiamato **trasposizione** e si indica graficamente con una T all'esponente:

$$(1,2,3)^T = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \qquad \begin{pmatrix} 4 \\ 0 \\ 2 \end{pmatrix}^T = (4,0,2),$$

e più in generale

$$(a_1, \dots, a_n)^T = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}, \qquad \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}^T = (a_1, \dots, a_n).$$

[1] Il perché di questa distinzione apparentemente insensata sarà chiaro più avanti; al lettore chiediamo un po' di pazienza.

[2] Naturale nella misura in cui le righe sono lette da sinistra a destra e le colonne dall'alto in basso: un ipotetico lettore extraterrestre potrebbe trovare tale bigezione poco naturale.

Nei vettori numerici le parentesi hanno funzione puramente decorativa e servono solo a separare graficamente il vettore da ciò che lo circonda. La scelta di usare le parentesi tonde è soggettiva; molti autori preferiscono usare le parentesi quadre e qualcuno preferisce usare due semplici barre verticali.

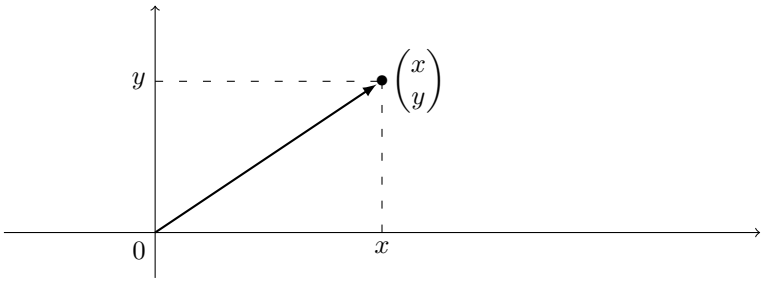


FIGURA 4.1. La bigezione tra $\mathbb{R}^2$ ed i vettori del piano applicati nell'origine.

La bigezione di $\mathbb{R}^2$ e $\mathbb{R}^{(2)}$ con lo spazio dei vettori del piano (Figura 4.1) permette di definire delle operazioni di somma e prodotto per scalare che, per analogia, vengono estese agli spazi $\mathbb{K}^n$ e $\mathbb{K}^{(n)}$ nel modo seguente:

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

$$t(a_1, \dots, a_n) = (ta_1, \dots, ta_n),$$

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{pmatrix}, \qquad t \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} ta_1 \\ \vdots \\ ta_n \end{pmatrix}.$$

Ad esempio, si ha:

(1) $(1, 2, 3) + (4, 5, 6) = (5, 7, 9)$,
(2) $3(1, 0, 0) = (3, 0, 0)$,
(3) $(1, 2) + (2, -1) + (0, 3) = (3, 4)$,
(4) $2(1, 2, 3) + 4(0, 0, 1) = (2, 4, 6) + (0, 0, 4) = (2, 4, 10)$.

Notiamo che la trasposizione commuta con le operazioni di somma e prodotto per scalare, ciò significa che valgono le formule

$$(v + w)^T = v^T + w^T, \qquad (tv)^T = tv^T,$$

per ogni coppia di vettori numerici v, w ed ogni scalare $t \in \mathbb{K}$.

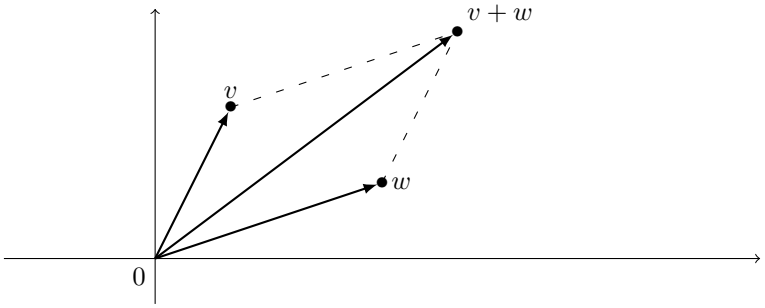


FIGURA 4.2. Somma di vettori in $\mathbb{R}^2$.

Esempio 4.1.1. Dato che ogni numero complesso è univocamente determinato dalla sua parte reale e dalla sua parte immaginaria, l'applicazione

$$f\colon \mathbb{R}^2 \to \mathbb{C}, \qquad f\begin{pmatrix} a \\ b \end{pmatrix} = a + ib\,,$$

è bigettiva e commuta con le operazioni di somma, ossia vale la formula:

$$f\left(\begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} c \\ d \end{pmatrix}\right) = f\begin{pmatrix} a \\ b \end{pmatrix} + f\begin{pmatrix} c \\ d \end{pmatrix}.$$

In maniera del tutto simile (esercizio), per ogni primo positivo p esiste una bigezione tra $\mathbb{Q}^2$ e $\mathbb{Q}[\sqrt{p}]$ che commuta con le operazioni di somma.

Ogni sistema lineare in m incognite ed n equazioni nel campo $\mathbb{K}$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m = b_1 \\ \quad\vdots \qquad\qquad \vdots \qquad\qquad\quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m = b_n \end{cases}$$

può essere interpretato come l'equazione

$$x_1\begin{pmatrix} a_{11} \\ \vdots \\ a_{n1} \end{pmatrix} + x_2\begin{pmatrix} a_{12} \\ \vdots \\ a_{n2} \end{pmatrix} + \cdots + x_m\begin{pmatrix} a_{1m} \\ \vdots \\ a_{nm} \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

ad m incognite $x_1, \ldots, x_m$ nello spazio dei vettori colonna $\mathbb{K}^n$. In effetti questa è una delle principali motivazioni per l'introduzione degli spazi $\mathbb{K}^n$.

Esercizi.

187. Determinare $v + w$, $v - w$ e $2v - 3w$ in ognuno dei seguenti casi:
   (1) $v = (1, 0, 0)$, $w = (0, 1, 0)$;
   (2) $v = (2, 3)$, $w = (1, -1)$;
   (3) $v = (1, 1, 1, 1)^T$, $w = (0, 1, 2, 3)^T$.

188. Trovare, se esistono, due numeri razionali t, s tali che

$$t\begin{pmatrix} 1 \\ 3 \end{pmatrix} + s\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

189. Trovare, se esistono, tre numeri razionali t, s, r tali che

$$t\begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix} + s\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ r \end{pmatrix}.$$

190. Descrivere un'applicazione iniettiva $f\colon \mathbb{Q}^2 \to \mathbb{R}$ tale che $f(v + w) = f(v) + f(w)$ per ogni $v, w \in \mathbb{Q}^2$.

191 (Somme di Minkowski). Dati due sottoinsiemi non vuoti $A, B \subset \mathbb{K}^n$, si definisce la somma di Minkowski come

$$A + B = \{a + b \mid a \in A,\ b \in B\} \subset \mathbb{K}^n.$$

Dimostrare che la somma di Minkowski gode della proprietà associativa, ossia che $(A+B)+C = A + (B + C)$.

192 (☕). Denotiamo con $\mathbb{Z}^n \subset \mathbb{R}^n$ il sottoinsieme dei vettori colonna a coordinate intere e con $\mathcal{P}$ la famiglia di tutti i sottoinsiemi finiti e non vuoti di $\mathbb{Z}^n$. È immediato verificare che $\mathcal{P}$ è chiuso rispetto alla somma di Minkowski, ossia $A + B \in \mathcal{P}$ per ogni $A, B \in \mathcal{P}$.

   (1) mostrare che se B contiene almeno due elementi allora il numero di elementi di $A+B$ è strettamente maggiore del numero di elementi di A;
   (2) un elemento $C \in \mathcal{P}$ si dice riducibile se si può scrivere $C = A + B$ con A e B che contengono ciascuno almeno due vettori;
   (3) un elemento di $\mathcal{P}$ si dice irriducibile se non è riducibile. Dimostrare che ogni elemento di $\mathcal{P}$ si può scrivere come somma finita di irriducibili;

(4) provare che i seguenti 5 sottoinsiemi di $\mathbb{Z}^2$:

$$A = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}, \quad B = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}, \quad C = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}, \quad A \cup C, \quad B \cup C,$$

sono irriducibili in $\mathcal{P}$ e che

$$A + B + C = (A \cup C) + (B \cup C).$$

Dedurre che la decomposizione in irriducibili non è unica in generale.

4.2. Spazi vettoriali

Uno **spazio vettoriale** su $\mathbb{K}$ è un insieme V i cui elementi, chiamati **vettori**, possono essere sommati tra loro e moltiplicati per elementi di $\mathbb{K}$ in modo che i risultati di tali operazioni siano ancora elementi di V:

$$v, w \in V, \quad t \in \mathbb{K} \quad \rightsquigarrow \quad v + w,\, tv \in V.$$

Le operazioni di somma $V \times V \xrightarrow{+} V$ e di prodotto per scalare $\mathbb{K} \times V \to V$ devono soddisfare le seguenti condizioni assiomatiche:

(1) *(Proprietà associativa della somma).* Comunque si prendano $v, w, u \in V$ vale

$$v + (w + u) = (v + w) + u.$$

(2) *(Proprietà commutativa della somma).* Comunque si prendano $v, w \in V$ vale

$$v + w = w + v.$$

(3) Esiste un elemento $0_V \in V$, detto **vettore nullo** di V, tale che per ogni $v \in V$ si ha

$$v + 0_V = 0_V + v = v.$$

(4) Per ogni vettore $v \in V$ esiste un vettore $-v \in V$, detto **opposto** di v, tale che

$$v + (-v) = -v + v = 0_V.$$

(5) Per ogni $v \in V$ vale $1v = v$, dove $1 \in \mathbb{K}$ è l'unità.

(6) *(Proprietà distributive).* Comunque si prendano $v, w \in V$ e $a, b \in \mathbb{K}$ si ha che

$$a(v + w) = av + aw, \qquad (a + b)v = av + bv.$$

(7) Comunque si prendano $v \in V$ e $a, b \in \mathbb{K}$ si ha che

$$a(bv) = (ab)v.$$

Esempio 4.2.1. Gli spazi vettoriali numerici $\mathbb{K}^n$ e $\mathbb{K}^{(n)}$, con le operazioni di somma e prodotto per scalare descritte nella sezione precedente, sono spazi vettoriali sul campo $\mathbb{K}$.

Esempio 4.2.2. L'insieme $\{0\}$ formato da un solo elemento è uno spazio vettoriale, con le operazioni di somma e prodotto per scalare definite nell'unico modo possibile, e cioè $0+0 = 0$, $a0 = 0$.

Altre proprietà della somma e del prodotto per scalare seguono come conseguenza dalle precedenti 7 condizioni assiomatiche. Ne elenchiamo subito le più rilevanti:

(8) *Ogni spazio vettoriale è un insieme non vuoto:* infatti per la condizione (3) deve contenere almeno il vettore nullo.

(9) *Il vettore nullo è unico:* infatti se $O \in V$ è un vettore con la proprietà che $v + O = v$ per ogni v si avrebbe

$$0_V = 0_V + O = O,$$

dove la prima uguaglianza vale perché O è un vettore nullo e la seconda uguaglianza vale perché 0_V è un vettore nullo.

(10) *Vale la proprietà di cancellazione della somma:* ciò significa che dati tre vettori $u, v, w \in V$, se vale $v + w = v + u$ allora $w = u$. Infatti, per l'esistenza dell'opposto e l'associatività della somma si ha

$$w = 0_V + w = (-v) + v + w = (-v) + v + u = 0_V + u = u.$$

(11) *L'opposto di un vettore è unico:* infatti se $v, w \in V$ sono tali che $v + w = 0_V$, allora dal fatto che $v + (-v) = 0_V$ e dalla proprietà di cancellazione segue che $w = -v$.

(12) *Per ogni $v \in V$ vale $0v = 0_V$, $(-1)v = -v$ e $-(-v) = v$, dove $0, -1 \in \mathbb{K}$:* infatti per ogni vettore v si ha

$$v + 0_V = v = 1v = (1+0)v = 1v + 0v = v + 0v$$

e quindi $0v = 0_V$ per la proprietà di cancellazione. Similmente per ogni vettore v si ha

$$v + (-v) = 0_V = 0v = (1-1)v = v + (-1)v$$

da cui $-v = (-1)v$. Infine $-(-v) = (-1)((-1)v) = (-1)^2 v = v$.

Da ciò segue anche che se $\mathbb{K} = 0$ è il campo banale (quello in cui $0 = 1$), allora ogni spazio vettoriale su $\mathbb{K}$ è anch'esso banale. Dunque, ogni volta che parleremo di spazi vettoriali supporremo implicitamente che i campi coinvolti siano non banali.

(13) *Per ogni $a \in \mathbb{K}$ vale $a0_V = 0_V$:* abbiamo già dimostrato che $0\,0_V = 0_V$ e quindi

$$a0_V = a(0\,0_V) = (a0)v = 0\,0_V = 0_V\,.$$

(14) *Dati $a \in \mathbb{K}$ e $v \in V$ vale $av = 0_V$ se e solo se $a = 0$ oppure $v = 0_V$:* infatti se $a \neq 0$ e $av = 0_V$ si ha $v = 1v = a^{-1}av = a^{-1}0_V = 0_V$.

La differenza di due vettori u, v in uno spazio vettoriale V si definisce nel modo ovvio tramite la formula

$$u - v = u + (-v) = u + (-1)v.$$

Notiamo che valgono le proprietà distributive

$$a(u-v) = au - av, \qquad (a-b)v = av - bv, \qquad a, b \in \mathbb{K}, \quad u, v \in V.$$

Esempio 4.2.3. Ogni campo di numeri $F \subset \mathbb{C}$, con le usuali operazioni di somma e prodotto è uno spazio vettoriale sul campo $\mathbb{K} = \mathbb{Q}$ dei numeri razionali.

Esempio 4.2.4. Il campo $\mathbb{C}$, con le usuali operazioni di somma e prodotto è uno spazio vettoriale su ogni suo sottocampo $\mathbb{K} \subseteq \mathbb{C}$.

Esempio 4.2.5. Siano $\mathbb{K}$ un campo e S un insieme qualunque. Indichiamo con $\mathbb{K}^S$ l'insieme di tutte le applicazioni $x\colon S \to \mathbb{K}$, $s \mapsto x_s$. Notiamo che $\mathbb{K}^S$ non è mai vuoto: infatti se $S = \emptyset$, allora l'insieme $\mathbb{K}^S$ contiene esattamente un elemento, mentre se $S \neq \emptyset$ allora $\mathbb{K}^S$ contiene, tra le altre, *l'applicazione nulla*

$$0\colon S \to \mathbb{K}, \qquad 0_s = 0 \text{ per ogni } s \in S.$$

Esiste una naturale operazione di somma tra gli elementi di $\mathbb{K}^S$: date due applicazioni $x, y \in \mathbb{K}^S$ possiamo definire

$$x + y\colon S \to \mathbb{K}, \qquad (x+y)_s = x_s + y_s \text{ per ogni } s \in S.$$

Se $x \in \mathbb{K}^S$ e $a \in \mathbb{K}$ è uno scalare, possiamo definire una nuova applicazione $ax \in \mathbb{K}^S$ data da:

$$(ax)_s = ax_s \text{ per ogni } s \in S.$$

Si verifica facilmente che le operazioni di somma e prodotto per scalare rendono $\mathbb{K}^S$ uno spazio vettoriale su $\mathbb{K}$. Se $S = \{1, 2, \dots, n\}$ allora $\mathbb{K}^S$ si identifica naturalmente con gli spazi vettoriali numerici $\mathbb{K}^n$ e $\mathbb{K}^{(n)}$.

Esempio 4.2.6. L'insieme $\mathbb{K}[x]$ dei polinomi a coefficienti in $\mathbb{K}$ nella variabile x, dotato delle usuali regole di somma e prodotto per scalare,

$$(\sum a_i x^i) + (\sum b_i x^i) = \sum (a_i + b_i) x^i, \qquad s(\sum a_i x^i) = \sum s a_i x^i,$$

è uno spazio vettoriale su $\mathbb{K}$. Il vettore nullo coincide con il polinomio nullo.

Esempio 4.2.7. Dotiamo l'insieme $V = \mathbb{R}$ dell'usuale operazione di somma e di un'operazione di prodotto per numeri complessi $*$ tale che $a * b = ab$ per ogni $a, b \in \mathbb{R}$ (ad esempio $(a+ib) * t = at$ oppure $(a+ib) * t = (a+b)t$). Con tali operazioni $\mathbb{R}$ **non è** uno spazio vettoriale su $\mathbb{C}$. Infatti se $a = i * 1 \in \mathbb{R}$ allora $(i-a) * 1 = 0$ e quindi $i = a$.[3]

[3] Si può dimostrare che esiste un prodotto per scalare che rende $\mathbb{R}$, con la usuale somma, uno spazio vettoriale su $\mathbb{C}$, vedi Esercizio 673. Tuttavia, tale dimostrazione non è costruttiva e nessuno, nel mondo sublunare, è in grado di descrivere esplicitamente un tale prodotto.

Se V e W sono spazi vettoriali su $\mathbb{K}$, allora anche il loro prodotto cartesiano

$$V \times W = \{(v, w) \mid v \in V,\ w \in W\}$$

è uno spazio vettoriale con le operazioni di somma e prodotto per scalare date da

$$(v_1, w_1) + (v_2, w_2) = (v_1 + v_2, w_1 + w_2), \qquad a(v, w) = (av, aw).$$

Notiamo che vale $-(v, w) = (-v, -w)$ e $0_{V\times W} = (0_V, 0_W)$.

In maniera simile si definisce il prodotto cartesiano di una qualsiasi successione finita $V_1, \ldots, V_n$ di spazi vettoriali:

$$V_1 \times \cdots \times V_n = \{(v_1, \ldots, v_n) \mid v_i \in V_i \text{ per ogni } i\}.$$

Le operazioni di somma e prodotto per scalare sono definite nel modo ovvio

$$(u_1, \ldots, u_n) + (v_1, \ldots, v_n) = (u_1 + v_1, \ldots, u_n + v_n), \qquad a(v_1, \ldots, v_n) = (av_1, \ldots, av_n).$$

Per ogni spazio vettoriale V e per ogni insieme S, l'insieme V^S di tutte le applicazioni $v\colon S \to V$, $s \mapsto v_s$, ha una naturale struttura di spazio vettoriale, con le operazioni definite allo stesso modo di $\mathbb{K}^S$, ossia

$$(v + w)_s = v_s + w_s, \qquad (av)_s = av_s, \qquad \text{per ogni} \quad s \in S.$$

Esercizi.

193. Sia V uno spazio vettoriale.
(1) Quanto vale $0_V + 0_V + 0_V$?
(2) Siano $u, v \in V$. Mostrare che $u - v = 0_V$ se e solo se $u = v$.
(3) Siano $u, v, x, y \in V$ tali che $x + y = 2u$ e $x - y = 2v$. Mostrare che $x = u + v$ e $y = u - v$.

194. Siano $V = \mathbb{C} \times \mathbb{C}$ e $\mathbb{K} = \mathbb{C}$. Per ciascuna delle seguenti 5 coppie di operazioni di "somma" $\oplus$ e "prodotto per scalare" $*$, determinare quali tra le 7 condizioni assiomatiche che definiscono lo spazio vettoriale sono soddisfatte e quali non lo sono.
(1) $(a, b) \oplus (c, d) = (a + c, b + d)$, $t * (a, b) = (ta, b)$;
(2) $(a, b) \oplus (c, d) = (a + c, b - d)$, $t * (a, b) = (ta, tb)$;
(3) $(a, b) \oplus (c, d) = (a + c, b + d)$, $t * (a, b) = (|t|a, |t|b)$;
(4) $(a, b) \oplus (c, d) = (a + c, b + d)$, $t * (a, b) = (ta, 0)$;
(5) $(a, b) \oplus (c, d) = (a + c, b + d)$, $t * (a, b) = (2ta, 2tb)$.

195. Mostrare che il sottoinsieme $V \subseteq \mathbb{R}$ dei numeri della forma $a + b\sqrt[3]{2}$, con $a, b \in \mathbb{Q}$, è uno spazio vettoriale su $\mathbb{Q}$ ma non è un campo di numeri.

196. Mostrare che se $\mathbb{K} \subseteq F \subseteq \mathbb{C}$ sono due campi di numeri, allora F è uno spazio vettoriale su $\mathbb{K}$.

197. Sia $V = \mathbb{Q}$ dotato dell'usuale somma. Per ciascuno dei seguenti campi $\mathbb{K}$ e prodotti per scalare $*$ dire quali tra le condizioni assiomatiche (5), (6) e (7) sono verificate:
(1) $\mathbb{K} = \mathbb{Q}$ e $a * v = 0$ per ogni $a \in \mathbb{K}$ e $v \in V$;
(2) $\mathbb{K} = \mathbb{Q}$ e $a * v = v$ per ogni $a \in \mathbb{K}$ e $v \in V$;
(3) $\mathbb{K} = \mathbb{Q}(\sqrt{2})$ e $(a + b\sqrt{2}) * v = av$ per ogni $a, b \in \mathbb{Q}$ e $v \in V$.

4.3. Sottospazi vettoriali

Sia V uno spazio vettoriale sul campo $\mathbb{K}$. Diremo che un sottoinsieme $U \subseteq V$ è un **sottospazio vettoriale** se soddisfa le seguenti condizioni:
(1) $0_V \in U$;
(2) U è chiuso per l'operazione di somma, ossia se $u_1, u_2 \in U$, allora $u_1 + u_2 \in U$;
(3) U è chiuso per l'operazione di prodotto per scalare, ossia se $u \in U$ e $a \in \mathbb{K}$, allora $au \in U$.

Notiamo che se U è un sottospazio vettoriale di V, allora per ogni vettore $u \in U$ si ha $-u = (-1)u \in U$. Ne segue che U è a sua volta uno spazio vettoriale, con le operazioni di somma e prodotto per scalare indotte da quelle di V.

ESEMPIO 4.3.1. Se V è uno spazio vettoriale, allora $\{0_V\}$ e V sono sottospazi vettoriali di V.

ESEMPIO 4.3.2. Dato un vettore riga $a = (a_1, \dots, a_n) \in \mathbb{K}^{(n)}$, il sottoinsieme

$$H_a = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{K}^n \;\middle|\; a_1x_1 + \cdots + a_nx_n = 0 \right\}$$

è un sottospazio vettoriale di $\mathbb{K}^n$.

LEMMA 4.3.3. *Siano U_1, U_2 due sottospazi vettoriali di uno spazio vettoriale V. Allora $U_1 \cap U_2$ è ancora un sottospazio vettoriale. Più in generale, l'intersezione di una famiglia qualunque di sottospazi vettoriali è ancora un sottospazio vettoriale.*

DIMOSTRAZIONE. Sia U_i una famiglia di sottospazi vettoriali di V e indichiamo con $U = \cap U_i$ la loro interesezione. Per definizione di sottospazio si ha $0_V \in U_i$ per ogni i e quindi $0_V \in U$. Se $u_1, u_2 \in U$, allora $u_1, u_2 \in U_i$ per ogni indice i; dunque $u_1 + u_2 \in U_i$ per ogni i e di conseguenza $u_1 + u_2 \in U$. La dimostrazione che U è chiusa rispetto al prodotto per scalare è del tutto simile ed è lasciata per esercizio al lettore. □

DEFINIZIONE 4.3.4. Dati due sottospazi vettoriali $U, W \subseteq V$ definiamo la loro **somma** $U + W$ come l'insieme formato da tutti i vettori della forma $u + w$, al variare di $u \in U$ e $w \in W$:

$$U + W = \{u + w \mid u \in U,\ w \in W\} \subseteq V.$$

Più in generale se $U_1, \dots, U_n \subseteq V$ sono sottospazi, la loro somma è definita come

$$U_1 + \cdots + U_n = \{u_1 + \cdots + u_n \mid u_i \in U_i,\ i = 1, \dots, n\} \subseteq V.$$

Osserviamo che $U \subseteq U + W$ (U è l'insieme dei vettori $u + 0$) e $W \subseteq U + W$.

Si verifica facilmente che la somma di sottospazi vettoriali è ancora un sottospazio vettoriale. Siano infatti $U_1, \dots, U_n$ sottospazi vettoriali di V e denotiamo $U = U_1 + \cdots + U_n$; siccome $0_V \in U_i$ per ogni indice i si ha

$$0_V + \cdots + 0_V = 0_V \in U.$$

Dati due vettori $u_1 + \cdots + u_n, v_1 + \cdots + v_n \in U$ ed uno scalare $a \in \mathbb{K}$ si ha

$$(u_1 + \cdots + u_n) + (v_1 + \cdots + v_n) = (u_1 + v_1) + \cdots + (u_n + v_n) \in U,$$

$$a(u_1 + \cdots + u_n) = au_1 + \cdots + au_n \in U.$$

LEMMA 4.3.5. *Dati $U_1, \dots, U_n \subseteq V$ sottospazi vettoriali, le seguenti condizioni sono equivalenti:*

1. *Ogni vettore $v \in U_1 + \cdots + U_n$ si scrive in modo unico nella forma $v = u_1 + \cdots + u_n$ con $u_i \in U_i$.*
2. *Dati n vettori $u_i \in U_i$, $i = 1, \dots, n$, se $u_1 + \cdots + u_n = 0_V$, allora $u_i = 0_V$ per ogni i.*

DIMOSTRAZIONE. [1 ⇒ 2] Siccome $0_V + \cdots + 0_V = 0_V$, se vale $u_1 + \cdots + u_n = 0_V$ per l'unicità della decomposizione vale $u_i = 0_V$ per ogni indice i.

[2 ⇒ 1] Se vale

$$v = u_1 + \cdots + u_n = w_1 + \cdots + w_n$$

con $u_i, w_i \in U_i$, allora si ha

$$0_V = v - v = (u_1 + \cdots + u_n) - (w_1 + \cdots + w_n) = (u_1 - w_1) + \cdots + (u_n - w_n)$$

e quindi $u_i - w_i = 0_V$ per ogni i. □

DEFINIZIONE 4.3.6. Se dei sottospazi U_i soddisfano le condizioni del Lemma 4.3.5 diremo che la loro somma $U = U_1 + \cdots + U_n$ è una **somma diretta** e scriveremo $U = U_1 \oplus \cdots \oplus U_n$.

LEMMA 4.3.7. *Dati due sottospazi $U_1, U_2 \subseteq V$, vale $U_1 + U_2 = U_1 \oplus U_2$ se e solo se $U_1 \cap U_2 = \{0_V\}$.*

DIMOSTRAZIONE. Mostriamo che esiste una bigezione tra $U_1 \cap U_2$ ed i modi di scrivere 0_V come somma di un vettore di U_1 ed un vettore di U_2. Dato $u \in U_1 \cap U_2$, per definizione di intersezione $u \in U_1$, $u \in U_2$ e quindi $-u \in U_2$ e $0_V = u + (-u)$. Viceversa se $0_V = u_1 + u_2$ con $u_i \in U_i$, allora $u_1 = -u_2 \in U_2$ e quindi $u_1 \in U_1 \cap U_2$.

□

Tra i risultati di algebra lineare più utili figura il fatto che ogni spazio vettoriale su di un campo infinito non può essere scritto come unione finita di sottospazi vettoriali propri. Più in generale si ha il seguente teorema.

TEOREMA 4.3.8. *Sia V uno spazio vettoriale su di un campo $\mathbb{K}$. Se $\mathbb{K}$ contiene almeno n elementi distinti, allora non si può scrivere V come unione di n sottospazi vettoriali propri. In particolare ogni spazio vettoriale su di un campo infinito non è unione finita di sottospazi vettoriali propri.*

DIMOSTRAZIONE. Siano $H_1, \dots, H_n \subset V$ sottospazi vettoriali propri. Dimostriamo per induzione su n che se $\mathbb{K}$ contiene n scalari distinti allora esiste un vettore $v \notin H_1 \cup \cdots \cup H_n$. Se $n = 1$ tutto segue dall'ipotesi che H_1 è un sottospazio proprio, ossia $H_1 \neq V$. Supponiamo $n > 1$ e per induzione che $H_1 \cup \cdots \cup H_{n-1} \neq V$. Siccome $H_n \neq V$ per ipotesi possiamo trovare due vettori $u, w \in V$ tali che

$$u \notin H_1 \cup \cdots \cup H_{n-1}, \qquad w \notin H_n.$$

Supponiamo per assurdo che $H_1 \cup \cdots \cup H_n = V$, allora necessariamente $u \in H_n$. Scegliamo n scalari distinti $a_1, \dots, a_n \in \mathbb{K}$ e consideriamo gli n vettori

$$v_1 = a_1 u + w, \quad v_2 = a_2 u + w, \quad \dots, \quad v_n = a_n u + w.$$

Notiamo che $v_i \notin H_n$ per ogni indice i: infatti, se fosse $v_i \in H_n$, allora anche $w = v_i - a_i u \in H_n$, in contraddizione con la scelta di w. Dunque $v_i \in H_1 \cup \cdots \cup H_{n-1}$ per ogni $i = 1, \dots, n$ e per il principio dei cassetti esistono due indici distinti i, j tali che $v_i, v_j \in H_k$ per qualche $k = 1, \dots, n-1$. Ma allora si ha

$$v_i - v_j = (a_i - a_j)u, \quad u = \frac{1}{a_i - a_j}(v_i - v_j) \in H_k,$$

in contraddizione con il fatto che $u \notin H_1 \cup \cdots \cup H_{n-1}$. □

Esercizi

198. Dimostrare le affermazioni fatte negli Esempi 4.3.1 e 4.3.2

199. Nello spazio vettoriale $\mathbb{K}[x]$, dire quali dei seguenti sottoinsiemi sono sottospazi vettoriali:

(1) $U = \{p(x) \in \mathbb{K}[x] \mid p(0) = 0\}$,
(2) $U = \{p(x) \in \mathbb{K}[x] \mid p(0) = 1\}$,
(3) $U = \{p(x) \in \mathbb{K}[x] \mid p(1) = 0\}$,
(4) $U = \{p(x) \in \mathbb{K}[x] \mid p(0) = p(1) = 0\}$,
(5) $U = \{p(x) \in \mathbb{K}[x] \mid p(0)p(1) = 0\}$.

200. Dati U, W sottospazi vettoriali di V, provare che $U \cap W$ è l'unico sottospazio vettoriale con le seguenti proprietà:

(1) $U \cap W \subseteq U$ e $U \cap W \subseteq W$;
(2) se A è un sottospazio vettoriale di V e $A \subseteq U$, $A \subseteq W$, allora $A \subseteq U \cap W$.

201. Dati U, W sottospazi vettoriali di V, provare che $U + W$ è l'unico sottospazio vettoriale con le seguenti proprietà:

(1) $U \subseteq U + W$ e $W \subseteq U + W$;
(2) se A è un sottospazio vettoriale di V e $U \subseteq A$, $W \subseteq A$, allora $U + W \subseteq A$.

202. Siano U, W sottospazi vettoriali di V. Mostrare che le seguenti quattro condizioni sono equivalenti:

(1) $U \subseteq W$,
(2) $U \cap W = U$,
(3) $U + W = W$,

(4) $W \cap (S+U) = (W \cap S) + U$ per ogni sottospazio vettoriale S.

203. Mostrare che la somma di sottospazi è associativa e simmetrica, ossia che

$$U + W = W + U, \qquad (U + W) + Z = U + (W + Z).$$

204. Trovare tre sottospazi vettoriali $U, V, W \subseteq \mathbb{R}^2$ tali che $U \cap V = U \cap W = V \cap W = \{0_{\mathbb{R}^2}\}$ e la cui somma $U + V + W$ non è diretta.

205. Trovare un esempio di due sottospazi vettoriali la cui unione non è un sottospazio vettoriale.

206. Dati tre sottospazi vettoriali A, B, C provare che valgono le inclusioni

$$(A \cap B) + (A \cap C) \subseteq A \cap (B + C), \qquad A + (B \cap C) \subseteq (A + B) \cap (A + C),$$

e trovare degli esempi in cui tali inclusioni sono strette.

207 (♡). Sia $\mathbb{K}$ un campo finito con q elementi. Provare che per ogni $n \geq 2$ lo spazio vettoriale $\mathbb{K}^n$ è unione di $q + 1$ sottospazi vettoriali propri.

4.4. Combinazioni lineari e generatori

Attenzione: per alleggerire la notazione, da ora in poi, quando il rischio di confusione sarà assente o improbabile indicheremo il vettore nullo ed il sottospazio nullo di uno spazio vettoriale V con il simbolo 0, anziché con i più precisi e pedanti 0_V e $\{0_V\}$.

Definizione 4.4.1. Siano $\mathbb{K}$ un campo, V uno spazio vettoriale su $\mathbb{K}$ e $v_1, \ldots, v_n$ vettori in V. Un vettore $v \in V$ si dice **combinazione lineare** di $v_1, \ldots, v_n$ a coefficienti in $\mathbb{K}$ se vale

$$v = a_1 v_1 + \cdots + a_n v_n$$

per opportuni scalari $a_1, \ldots, a_n \in \mathbb{K}$.

Ad esempio, il vettore $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \in \mathbb{K}^3$ è combinazione lineare di $\begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}$ e $\begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix}$ in quanto vale

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = 2 \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} - \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix}.$$

Indichiamo con $\mathrm{Span}(v_1, \ldots, v_n)$ l'insieme di tutte le possibili combinazioni lineari dei vettori $v_1, \ldots, v_n$, ossia:

$$\mathrm{Span}(v_1, \ldots, v_n) = \{v \in V \mid v = a_1 v_1 + \cdots + a_n v_n,\ a_i \in \mathbb{K},\ i = 1, \ldots, n\}.$$

È facile dimostrare che $\mathrm{Span}(v_1, \ldots, v_n)$ è un sottospazio vettoriale. Infatti contiene lo 0 (basta porre $a_i = 0$ per ogni i); se

$$v = a_1 v_1 + \cdots + a_n v_n, \qquad w = b_1 v_1 + \cdots + b_n v_n,$$

sono due combinazioni lineari, allora la somma

$$v + w = (a_1 + b_1) v_1 + \cdots + (a_n + b_n) v_n$$

è ancora una combinazione lineare; per ogni scalare $t \in \mathbb{K}$ si ha

$$t(a_1 v_1 + \cdots + a_n v_n) = t a_1 v_1 + \cdots + t a_n v_n.$$

Chiameremo $\mathrm{Span}(v_1, \ldots, v_n)$ **sottospazio vettoriale generato da** $v_1, \ldots, v_n$ **su** $\mathbb{K}$. Quando il campo $\mathbb{K}$ è chiaro dal contesto diremo più semplicemente **sottospazio generato da** $v_1, \ldots, v_n$, oppure **chiusura lineare di** $v_1, \ldots, v_n$, oppure ancora **span di** $v_1, \ldots, v_n$.[4]

Osserviamo che il sottospazio $\mathrm{Span}(v_1, \ldots, v_n)$ non dipende dall'ordine dei vettori v_i, ragion per cui, ad esempio vale $\mathrm{Span}(v, w) = \mathrm{Span}(w, v)$. Questo ci permette di definire, per ogni sottoinsieme finito e non vuoto $A \subseteq V$ la sua chiusura lineare $\mathrm{Span}(A)$ come

$$\mathrm{Span}(A) = \{ \text{ combinazioni lineari di vettori in } A\},$$

[4]Dall'inglese "to span" che si può tradurre come "estendere da parte a parte", da non confondersi con il passato di to spin.

e cioè

$$\mathrm{Span}(A) = \mathrm{Span}(v_1, \dots, v_n), \qquad \text{dove} \quad A = \{v_1, \dots, v_n\}.$$

Possiamo estendere tale definizione anche all'insieme vuoto ponendo $\mathrm{Span}(\emptyset) = \{0\}$.

ESEMPIO 4.4.2. Sia $V = \mathbb{K}[x]$ lo spazio vettoriale dei polinomi in x a coefficienti in $\mathbb{K}$ e sia $A = \{x, x^2\} \subseteq V$. Allora $\mathrm{Span}(A)$ è l'insieme dei polinomi di grado ≤ 2 senza termine noto.

ESEMPIO 4.4.3. Consideriamo i vettori

$$v_1 = \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 2 \\ 4 \\ 2 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad w = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} \in \mathbb{R}^3,$$

e chiediamoci se la relazione $w \in \mathrm{Span}(v_1, v_2, v_3)$ è vera o falsa, e cioè se l'equazione lineare vettoriale $av_1 + bv_2 + cv_3 = w$ possiede una soluzione a, b, c. Per rispondere occorre studiare il sistema lineare

$$\begin{cases} a + 2b = 4 \\ 3a + 4b + c = 5 \\ 2b + c = 6 \end{cases},$$

che ammettendo soluzioni, implica che $w \in \mathrm{Span}(v_1, v_2, v_3)$, ossia che w appartiene al sottospazio vettoriale generato da v_1, v_2, v_3.

DEFINIZIONE 4.4.4. Lo spazio vettoriale V si dice **di dimensione finita** su $\mathbb{K}$, o anche **finitamente generato**, se esistono vettori $v_1, \dots, v_n$ in V tali che $V = \mathrm{Span}(v_1, \dots, v_n)$. In questo caso diremo che $\{v_1, \dots, v_n\}$ è un **insieme di generatori** di V.

Uno spazio vettoriale che non è di dimensione finita si dice di **dimensione infinita**.

ESEMPIO 4.4.5. Lo spazio vettoriale numerico $\mathbb{K}^n$ ha dimensione finita. Consideriamo infatti la successione di vettori $e_1, \dots, e_n$, dove e_i è il vettore colonna con la i-esima coordinata uguale ad 1 e tutte le altre uguali a 0, ossia

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

Tale successione, che più avanti chiameremo "base canonica", è un insieme di generatori. Infatti, per ogni $a_1, \dots, a_n \in \mathbb{K}$, vale la formula

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = a_1 e_1 + \cdots + a_n e_n.$$

ESEMPIO 4.4.6. Lo spazio vettoriale $\mathbb{K}[x]$ ha dimensione infinita su $\mathbb{K}$. Infatti, per ogni sottoinsieme finito $A \subseteq \mathbb{K}[x]$ è possibile trovare un intero d con la proprietà che ogni polinomio in A ha grado minore di d. Dunque $\mathrm{Span}(A)$ contiene solamente polinomi di grado minore di d e quindi $\mathrm{Span}(A) \neq \mathbb{K}[x]$.

La seguente proposizione riepiloga le principali proprietà della chiusura lineare.

PROPOSIZIONE 4.4.7. *Sia A un sottoinsieme finito di uno spazio vettoriale V. Si ha:*

1. $A \subseteq \mathrm{Span}(A)$*;*
2. $\mathrm{Span}(A)$ *è un sottospazio vettoriale di V;*
3. *Sia $W \subseteq V$ un sottospazio vettoriale, allora $A \subseteq W$ se e solo se $\mathrm{Span}(A) \subseteq W$;*
4. *Dato un sottoinsieme finito $B \subseteq V$, vale $\mathrm{Span}(A) \subseteq \mathrm{Span}(B)$ se e solo se $A \subseteq \mathrm{Span}(B)$.*

DIMOSTRAZIONE. La prima proprietà è ovvia e la seconda è già stata dimostrata. La quarta segue dalla terza e dalla seconda ponendo $W = \mathrm{Span}(B)$; rimane solo da dimostrare la (3). Sia W un sottospazio vettoriale, se $\mathrm{Span}(A) \subseteq W$, dato che $A \subseteq \mathrm{Span}(A)$ ne segue $A \subseteq W$. Se $A \subseteq W$, e siccome W è chiuso per le operazioni di somma e prodotto per scalare, ed ogni combinazione lineare può essere pensata come una composizione di somme e prodotti

per scalare, ne segue che ogni combinazione lineare di elementi di A appartiene a W e quindi $\mathrm{Span}(A) \subseteq W$. □

Nel prosieguo useremo spesso la seguente semplice conseguenza della precedente proposizione.

LEMMA 4.4.8. *Siano A, B sottoinsiemi finiti di uno spazio vettoriale. Allora $B \subseteq \mathrm{Span}(A)$ se e solo se $\mathrm{Span}(A) = \mathrm{Span}(A \cup B)$.*

DIMOSTRAZIONE. Siccome $B \subset A \cup B \subset \mathrm{Span}(A \cup B)$, se $\mathrm{Span}(A) = \mathrm{Span}(A \cup B)$ allora $B \subseteq \mathrm{Span}(A)$. Viceversa, $B \subseteq \mathrm{Span}(A)$, dato che $A \subseteq \mathrm{Span}(A)$ si ha che $A \cup B$ è contenuto nel sottospazio vettoriale $\mathrm{Span}(A)$ e quindi $\mathrm{Span}(A \cup B) \subseteq \mathrm{Span}(A)$. Per finire osserviamo che $A \subseteq A \cup B$ e quindi $\mathrm{Span}(A) \subseteq \mathrm{Span}(A \cup B)$. □

In particolare, il Lemma 4.4.8 implica che se per una successione di vettori $v_1, \ldots, v_n$ si ha $v_n \in \mathrm{Span}(v_1, \ldots, v_{n-1})$, allora $\mathrm{Span}(v_1, \ldots, v_{n-1}) = \mathrm{Span}(v_1, \ldots, v_{n-1}, v_n)$ (basta prendere $A = \{v_1, \ldots, v_{n-1}\}$ e $B = \{v_n\}$).

ESEMPIO 4.4.9. Siano $v_1, \ldots, v_n$ generatori di uno spazio vettoriale V e sia $W \subset V$ un sottospazio vettoriale proprio. Allora esiste un indice i tale che $v_i \notin W$. Infatti se $v_i \in W$ per ogni i si avrebbe $V = \mathrm{Span}(v_1, \ldots, v_n) \subseteq W$ in contraddizione con il fatto che $W \neq V$.

ESEMPIO 4.4.10. Chiediamoci se i vettori v_1, v_2 e v_3 dell'Esempio 4.4.3 generano $\mathbb{K}^3$. Affinché ciò sia vero è necessario che i tre vettori della base canonica appartengano a $\mathrm{Span}(v_1, v_2, v_3)$. Tale condizione è anche sufficiente perché se $\{e_1, e_2, e_3\} \subseteq \mathrm{Span}(v_1, v_2, v_3)$ allora vale

$$\mathbb{K}^3 = \mathrm{Span}(e_1, e_2, e_3) \subseteq \mathrm{Span}(v_1, v_2, v_3).$$

Il problema si riconduce quindi allo studio dei tre sistemi lineari

$$\begin{cases} a + 2b = 1 \\ 3a + 4b + c = 0 \\ 2b + c = 0 \end{cases}, \quad \begin{cases} a + 2b = 0 \\ 3a + 4b + c = 1 \\ 2b + c = 0 \end{cases}, \quad \begin{cases} a + 2b = 0 \\ 3a + 4b + c = 0 \\ 2b + c = 1 \end{cases}.$$

Per determinare se un determinato insieme finito genera uno spazio vettoriale V possono essere utili le seguenti osservazioni:

a) Se $A \subseteq B$ sono sottoinsiemi finiti di V, e se A genera V, allora anche B genera V.

b) Siano A, B due sottoinsiemi finiti di V, se A genera V ed ogni elemento di A può essere scritto come combinazione lineare di elementi di B, allora anche B genera V. Infatti se $V = \mathrm{Span}(A)$ e $A \subseteq \mathrm{Span}(B)$; ne segue che $\mathrm{Span}(A) \subseteq \mathrm{Span}(B)$ e quindi $V = \mathrm{Span}(B)$. In particolare, se A è ottenuto da B aggiungendo un numero finito di combinazioni lineari di elementi di B, e se A è un insieme di generatori, allora anche B è un insieme di generatori.

ESEMPIO 4.4.11. Usiamo le precedenti osservazioni per mostrare che i vettori

$$u = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad v = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad w = \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix}$$

generano $\mathbb{K}^3$. Abbiamo visto che non è restrittivo aggiungere ai tre vettori u, v, w alcune loro combinazioni lineari. Ad esempio

$$a = v - u = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \quad b = w - u = \begin{pmatrix} 0 \\ 2 \\ 3 \end{pmatrix}.$$

(L'idea è chiara: far comparire quanti più zeri è possibile.) Ripetiamo la procedura aggiungendo conbinazioni lineari di vettori del nuovo insieme $\{u, v, w, a, b\}$:

$$c = b - 2a = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}, \quad d = a + 2c = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad e = u - d + c = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Abbiamo già osservato che i vettori $e_1 = e$, $e_2 = d$ e $e_3 = -c$ generano $\mathbb{K}^3$ e quindi anche u, v, w sono generatori.

LEMMA 4.4.12. *Siano A, B sottoinsiemi finiti di uno spazio vettoriale V. Allora vale*

$$\mathrm{Span}(A \cup B) = \mathrm{Span}(A) + \mathrm{Span}(B),$$

dove $\mathrm{Span}(A) + \mathrm{Span}(B)$ *denota la somma dei sottospazi vettoriali* $\mathrm{Span}(A)$ *e* $\mathrm{Span}(B)$

DIMOSTRAZIONE. Siccome il sottospazio vettoriale $\mathrm{Span}(A) + \mathrm{Span}(B)$ contiene sia A che B, esso contiene l'unione $A \cup B$ e quindi contiene il sottospazio da essa generato, ossia $\mathrm{Span}(A \cup B) \subseteq \mathrm{Span}(A) + \mathrm{Span}(B)$. Viceversa se $v \in \mathrm{Span}(A) + \mathrm{Span}(B)$, per definizione esistono $u \in \mathrm{Span}(A)$ e $w \in \mathrm{Span}(B)$ tali che $v = u + w$. Siccome u è una combinazione lineare di elementi di A e w è una combinazione lineare di elementi di B, la loro somma $u + w$ è una combinazione lineare di elementi di $A \cup B$. □

Esercizi.

208 (♡). Dire, motivando la risposta se il vettore $e_1 = (1, 0, 0, 0) \in \mathbb{R}^{(4)}$ è combinazione lineare dei vettori $u = (1, 0, 1, 2)$, $v = (3, 4, 2, 1)$ e $w = (5, 8, 3, 0)$.

209. Dimostrare che i vettori $(1, 2, 1)^T$, $(2, 1, 3)^T$ e $(3, 3, 3)^T$ generano $\mathbb{R}^3$.

210. Dimostrare che ogni insieme di generatori di $\mathbb{R}^2$ contiene almeno due vettori ed ogni insieme di generatori di $\mathbb{R}^3$ contiene almeno tre vettori.

211. Sia V uno spazio vettoriale e si assuma che esista un'applicazione surgettiva $f\colon V \to \mathbb{N}$ tale che per ogni $n \in \mathbb{N}$ il sottoinsieme $V_n = \{v \in V \mid f(v) \leq n\}$ è un sottospazio vettoriale di V. Provare che V ha dimensione infinita.

4.5. Indipendenza lineare e teorema di scambio

Siamo adesso pronti per definire i due concetti fondamentali di dipendenza ed indipendenza lineare di un insieme di vettori in uno spazio vettoriale.

DEFINIZIONE 4.5.1. Diremo che m vettori $w_1, \ldots, w_m$ in uno spazio vettoriale sul campo $\mathbb{K}$, sono **linearmente dipendenti** se esiste una loro combinazione lineare, con coefficienti non tutti nulli, che dà come risultato il vettore nullo:

$$a_1 w_1 + \cdots + a_m w_m = 0, \qquad \text{con} \quad a_i \in \mathbb{K} \quad \text{non tutti} = 0 \tag{4.1}$$

I vettori $w_1, \ldots, w_m$ si dicono **linearmente indipendenti** se non sono linearmente dipendenti.

Una combinazione lineare viene detta banale se tutti i coefficienti sono nulli e quindi, dei vettori risultano essere linearmente dipendenti se e solo se esiste una loro combinazione lineare nulla (ossia che ha come risultato il vettore nullo) ma non banale. Equivalentemente dei vettori sono linearmente indipendenti se e solo se l'unica combinazione lineare nulla tra loro è quella banale.

In pratica per stabilire se i vettori $w_1, \ldots, w_m$ sono o meno linearmente dipendenti occorre studiare l'equazione vettoriale

$$x_1 w_1 + \cdots + x_m w_m = 0$$

e determinare se esiste o meno una soluzione $(x_1, \ldots, x_m)$, con $x_i \in \mathbb{K}$ non tutti nulli.

ESEMPIO 4.5.2. Due vettori non nulli sono linearmente dipendenti se sono uno multiplo dell'altro (Figura 4.3).

ESEMPIO 4.5.3. I vettori v_1, v_2, v_3 dell'Esempio 4.4.3 sono linearmente indipendenti. Infatti l'equazione $av_1 + bv_2 + cv_3 = 0$, corrispondente al sistema lineare omogeneo

$$\begin{cases} a + 2b = 0 \\ 3a + 4b + c = 0 \\ 2b + c = 0 \end{cases},$$

ammette $a = b = c = 0$ come unica soluzione.

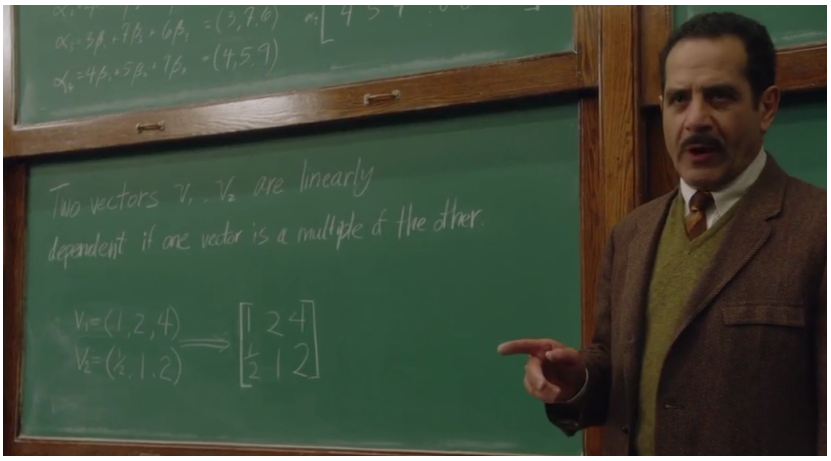


FIGURA 4.3. Tony Shalhoub interpreta un docente della Columbia University nella serie TV "The Marvelous Mrs. Maisel".

Osserviamo che se $w_1, \dots, w_m$ sono vettori linearmente indipendenti, allora i vettori w_i sono tutti diversi da 0 e distinti tra loro. Infatti se $w_i = 0$ si avrebbe la combinazione lineare non banale $1 \cdot w_i = 0$, mentre se $w_j = w_k$, con $j \neq k$ si avrebbe la combinazione lineare non banale $w_j - w_k = 0$. Un vettore è linearmente indipendente se e solo se è diverso da 0.

Il seguente risultato, che chiameremo *lemma di estensione*, è alla base della maggior parte dei risultati sull'indipendenza lineare e sarà usato più volte nelle pagine seguenti.

LEMMA 4.5.4 (di estensione). *Siano $v_1, \dots, v_n$ vettori in uno spazio vettoriale. Le seguenti condizioni sono equivalenti:*

(1) *$v_1, \dots, v_n$ sono linearmente indipendenti;*
(2) *$v_1, \dots, v_{n-1}$ sono linearmente indipendenti e $v_n \notin \mathrm{Span}(v_1, \dots, v_{n-1})$;*
(3) *$v_i \notin \mathrm{Span}(v_1, \dots, v_{i-1})$ per ogni $i = 1, \dots, n$ (per $i = 1$ la condizione $v_1 \notin \mathrm{Span}(\emptyset)$ equivale a dire $v_1 \neq 0$).*

DIMOSTRAZIONE. Per provare l'equivalenza delle tre condizioni dimostriamo prima che (1) implica (2), poi che (2) implica (3) ed infine che (3) implica (1).

Supponiamo che $v_1, \dots, v_n$ siano linearmente indipendenti. È chiaro che $v_1, \dots, v_{n-1}$ sono linearmente indipendenti, mentre se per assurdo $v_n \in \mathrm{Span}(v_1, \dots, v_{n-1})$ esisterebbero $n-1$ scalari $a_1, \dots, a_{n-1} \in \mathbb{K}$ tali che

$$v_n = a_1 v_1 + \cdots + a_{n-1} v_{n-1}.$$

In tal caso si avrebbe

$$a_1 v_1 + \cdots + a_{n-1} v_{n-1} + (-1) v_n = 0$$

in contraddizione con la lineare indipendenza di $v_1, \dots, v_n$.

Supponiamo adesso $v_1, \dots, v_{n-1}$ sono linearmente indipendenti e $v_n \notin \mathrm{Span}(v_1, \dots, v_{n-1})$. Avendo già dimostrato l'implicazione (1)⇒(2), si ha $v_1, \dots, v_{n-2}$ linearmente indipendenti e $v_{n-1} \notin \mathrm{Span}(v_1, \dots, v_{n-2})$. Ripetendo il ragionamento a ritroso si arriva alla conclusione che $v_i \notin \mathrm{Span}(v_1, \dots, v_{i-1})$ per ogni i.

Per finire, supponiamo che $v_i \notin \mathrm{Span}(v_1, \dots, v_{i-1})$ per ogni i e supponiamo per assurdo che $v_1, \dots, v_n$ siano linearmente dipendenti, ossia che esiste una relazione $a_1 v_1 + \cdots + a_n v_n = 0$, con gli scalari a_i non tutti nulli. indichiamo con r il massimo intero, compreso tra 1 ed n, tale che $a_r \neq 0$, allora $a_i = 0$ per ogni $i > r$ e dunque $a_1 v_1 + \cdots + a_r v_r = 0$, da cui si ricava

$$v_r = -\frac{a_1}{a_r} v_1 - \cdots - \frac{a_{r-1}}{a_r} v_{r-1} \in \mathrm{Span}(v_1, \dots, v_{r-1})$$

contrariamente a quanto ipotizzato. □

TEOREMA 4.5.5 (di scambio). *Sia A un sottoinsieme finito di uno spazio vettoriale. Se $\mathrm{Span}(A)$ contiene m vettori linearmente indipendenti, allora anche A contiene m vettori linearmente indipendenti.*

DIMOSTRAZIONE. Sia $B \subseteq \mathrm{Span}(A)$ un insieme di m vettori linearmente indipendenti e indichiamo con $\mathcal{F}$ la famiglia (finita) di tutti i sottoinsiemi di $A \cup B$ formati da m vettori linearmente indipendenti. La famiglia $\mathcal{F}$ non è vuota perché contiene B. Tra tutti i sottoinsiemi appartenenti alla famiglia $\mathcal{F}$ scegliamone uno, che chiameremo C, che ha il maggior numero di elementi in comune con A. Per dimostrare il teorema è sufficiente provare che $C \subseteq A$.

Supponiamo per assurdo che C non sia contenuto in A, possiamo allora scrivere

$$C = \{w_1, \dots, w_m\}, \qquad \text{con} \quad w_m \notin A.$$

Per quanto dimostrato nel Lemma 4.5.4 i vettori $w_1, \dots, w_{m-1}$ sono linearmente indipendenti e $w_m \notin \mathrm{Span}(w_1, \dots, w_{m-1})$. Poiché $w_n \in \mathrm{Span}(A)$, a maggior ragione $\mathrm{Span}(A)$ non è contenuto in $\mathrm{Span}(w_1, \dots, w_{m-1})$ e per la Proposizione 4.4.7 esiste un vettore $v \in A$ tale che $v \notin \mathrm{Span}(w_1, \dots, w_{m-1})$. Per il lemma di estensione 4.5.4 il sottoinsieme $D = \{w_1, \dots, w_{m-1}, v\}$ è ancora formato da m vettori indipendenti, ma ha in comune con A un vettore in più rispetto a C, in contraddizione con la scelta di C. □

COROLLARIO 4.5.6. *In uno spazio vettoriale generato da n vettori esistono al più n vettori linearmente indipendenti.*

DIMOSTRAZIONE. Sia V uno spazio vettoriale generato da $v_1, \dots, v_n$. Per definizione $V = \mathrm{Span}(v_1, \dots, v_n)$ e quindi se V contiene m vettori linearmente indipendenti, allora anche $\{v_1, \dots, v_n\}$ contiene m vettori linearmente indipendenti; dunque $m \leq n$. □

COROLLARIO 4.5.7. *Uno spazio vettoriale V è di dimensione infinita se e solo se per ogni intero positivo m esistono m vettori linearmente indipendenti in V.*

DIMOSTRAZIONE. Se V è di dimensione infinita, allora per ogni successione finita $v_1, \dots, v_n$ di vettori in V si ha $\mathrm{Span}(v_1, \dots, v_n) \neq V$. Possiamo quindi costruire per ricorrenza una successione infinita $\{v_i\}$, $i = 1, 2, \dots$, con le proprietà

$$v_1 \neq 0, \quad v_2 \notin \mathrm{Span}(v_1), \quad \dots, \quad v_{i+1} \notin \mathrm{Span}(v_1, \dots, v_i), \quad \dots$$

Qualunque sia $m > 0$ i primi m termini della successione sono linearmente indipendenti.

Viceversa, se V ha dimensione finita è possibile trovare un intero $n \geq 0$ ed n vettori che generano V. Per il teorema di scambio non esistono m vettori linearmente indipendenti per ogni $m > n$. □

ESEMPIO 4.5.8. Abbiamo visto nell'Esempio 3.6.6 che le radici quadrate dei numeri primi sono linearmente indipendenti su $\mathbb{Q}$. In maniera ancora più semplice si dimostra che i logaritmi (in qualsiasi base $b > 1$) dei numeri primi sono linearmente indipendenti su $\mathbb{Q}$: infatti siano $p_1, \dots, p_n$ numeri primi distinti e supponiamo

$$a_1 \log_b(p_1) + \cdots + a_n \log_b(p_n) = 0, \qquad a_i \in \mathbb{Q}\,.$$

Moltiplicando per un denominatore comune possiamo supporre $a_i \in \mathbb{Z}$ per ogni i e quindi

$$a_1 \log_b(p_1) + \cdots + a_n \log_b(p_n) = \log_b(p_1^{a_1} \cdots p_n^{a_n}) = 0$$

da cui segue $p_1^{a_1} \cdots p_n^{a_n} = 1$ che però è possibile solo se $a_i = 0$ per ogni i.

Siccome esistono infiniti numeri primi, segue dal Corollario 4.5.7 che $\mathbb{R}$ e $\mathbb{C}$ hanno dimensione infinita come spazi vettoriali su $\mathbb{Q}$.

ESEMPIO 4.5.9. Sia α un numero reale, allora gli $n + 1$ numeri $1, \alpha, \alpha^2, \dots, \alpha^n$ sono linearmente dipendenti su $\mathbb{Q}$ se e solo se α è la radice di un polinomio non nullo di grado $\leq n$. Più avanti dimostreremo (Teoremi 4.8.4 e 17.7.3) che il numero $\pi \in \mathbb{R}$ è trascendente, ossia non è radice di alcun polinomio a coefficienti razionali. Ne segue che per ogni $n > 0$ i numeri $1, \pi, \pi^2, \dots, \pi^n$ sono linearmente indipendenti su $\mathbb{Q}$ e ritroviamo il fatto che $\mathbb{R}$ è uno spazio vettoriale di dimensione infinita su $\mathbb{Q}$.

COROLLARIO 4.5.10. *Ogni sottospazio vettoriale di uno spazio di dimensione finita ha ancora dimensione finita.*

DIMOSTRAZIONE. Osserviamo che se W è un sottospazio di V e se $w_1, \dots, w_m \in W$ sono vettori linearmente indipendenti in W, allora sono linearmente indipendenti anche in V. Basta adesso applicare il Corollario 4.5.7. □

ESEMPIO 4.5.11. Siano $a_0, \dots, a_n \in \mathbb{K}$ scalari distinti e dimostriamo che gli $n+1$ vettori

$$v_i = \begin{pmatrix} 1 \\ a_i \\ a_i^2 \\ \vdots \\ a_i^n \end{pmatrix} \in \mathbb{K}^{n+1}, \qquad i = 0, \dots, n,$$

sono linearmente indipendenti. Per induzione su n possiamo supporre $v_1, \dots, v_n$ linearmente indipendenti (esercizio: perché?) e quindi basta dimostrare che se $\sum_{i=0}^n x_i v_i = 0$ con $x_0, \dots, x_n \in \mathbb{K}$ allora $x_0 = 0$. Consideriamo gli scalari $b_0, \dots, b_n \in \mathbb{K}$ definiti dall'identità di polinomi

$$p(t) = \prod_{i=1}^n \frac{t - a_i}{a_0 - a_i} = \sum_{j=0}^n b_j t^j.$$

Si vede subito che $p(a_0) = \sum_{j=0}^n b_j a_0^j = 1$ e $p(a_i) = \sum_{j=0}^n b_j a_i^j = 0$ per ogni $i > 0$. Dato che $\sum_{i=0}^n x_i a_i^j = 0$ per ogni $i = 0, \dots, n$ possiamo scrivere

$$0 = \sum_{j=0}^n b_j \sum_{i=0}^n x_i a_i^j = \sum_{i=0}^n x_i \sum_{j=0}^n b_j a_i^j = x_0.$$

Esercizi.

212. Per quali valori di $t \in \mathbb{R}$ i quattro vettori

$$v_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 3 \\ 2 \\ t \end{pmatrix}, \quad v_3 = \begin{pmatrix} 2 \\ 2 \\ t^2 \end{pmatrix}, \quad v_4 = \begin{pmatrix} 2 \\ 2 \\ t^3 \end{pmatrix} \in \mathbb{R}^3$$

sono linearmente dipendenti?

213. Siano $v_0, \dots, v_n$ vettori in uno spazio vettoriale V sul campo $\mathbb{K}$ con le seguenti proprietà:

(1) $v_1, \dots, v_n$ generano V;
(2) esiste un unico vettore riga $(a_1, \dots, a_n) \in \mathbb{K}^{(n)}$ tale che $a_1 v_1 + \cdots + a_n v_n = v_0$.

Dimostrare che $v_1, \dots, v_n$ sono linearmente indipendenti.

214. Ogni vettore di $\mathbb{R}^n$ può essere pensato come un vettore di $\mathbb{C}^n$ a coordinate reali. Dimostrare che $v_1, \dots, v_m \in \mathbb{R}^n$ sono linearmente indipendenti su $\mathbb{R}$ se e solo se, pensati come vettori di $\mathbb{C}^n$, sono linearmente indipendenti su $\mathbb{C}$.

215. Siano u, v, w tre vettori linearmente indipendenti in uno spazio vettoriale sul campo $\mathbb{K}$. Provare che per ogni scelta di $a, b, c \in \mathbb{K}$ i vettori $u, v + au, w + bv + cu$ sono ancora linearmente indipendenti.

216. Siano $v_0, v_1, \dots, v_n$ vettori linearmente indipendenti in uno spazio vettoriale sul campo $\mathbb{K}$. Provare che per ogni scelta di $a_1, \dots, a_n \in \mathbb{K}$ i vettori $v_0, v_1 + a_1 v_0, \dots, v_n + a_n v_0$ sono ancora linearmente indipendenti.

217. Dedurre dal teorema di scambio che ogni insieme di generatori di $\mathbb{R}^n$ contiene almeno n vettori.

218 (Indipendenza affine). Sia V uno spazio vettoriale su $\mathbb{K}$. Diremo che $p+1$ vettori $v_0, \dots, v_p \in V$ sono **affinemente dipendenti** se esistono $a_0, \dots, a_p \in \mathbb{K}$, non tutti nulli, e tali che:

$$a_0 v_0 + \cdots + a_p v_p = 0, \qquad a_0 + \cdots + a_p = 0.$$

I medesimi vettori si dicono **affinemente indipendenti** se non sono affinemente dipendenti.

Dimostrare che le seguenti condizioni sono equivalenti:

(1) i $p+1$ vettori $v_0, \dots, v_p \in V$ sono affinemente dipendenti;
(2) esiste un indice $i = 0, \dots, p$ tale che i p vettori $v_j - v_i$, $j \neq i$, sono linearmente dipendenti;
(3) per ogni $i = 0, \dots, p$ i p vettori $v_j - v_i$, $j \neq i$, sono linearmente dipendenti.

219 (Combinazioni affini). Sia V uno spazio vettoriale sul campo $\mathbb{K}$. Una combinazione lineare $a_0v_0 + \cdots + a_pv_p$ di vettori $v_i \in V$ e coefficienti $a_i \in \mathbb{K}$ si dice una **combinazione affine** se $\sum_{i=0}^{p} a_i = 1$.

Dimostrare che $p+1$ vettori $v_0, \ldots, v_p \in V$ sono affinemente indipendenti (Esercizio 218) se e solo se nessuno di essi può essere scritto come combinazione affine degli altri n.

220. Si assuma che il vettore nullo di uno spazio vettoriale V non sia combinazione affine dei vettori $v_0, \ldots, v_p \in V$. Dimostrare che $v_0, \ldots, v_p \in V$ sono affinemente indipendenti se e solo se sono linearmente indipendenti.

221 (♡). Sia $n > 0$ un intero positivo fissato e si denoti con $f\colon \mathbb{K}^n \to \mathbb{K}^{n+1}$ l'applicazione

$$f(v) = \begin{pmatrix} 1 \\ v \end{pmatrix}, \quad \text{ossia} \quad f\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Dimostrare che per una successione di $p+1$ vettori $v_0, \ldots, v_p \in \mathbb{K}^n$ le seguenti condizioni sono equivalenti:

(1) i vettori $v_1 - v_0, \ldots, v_p - v_0$ sono linearmente dipendenti in $\mathbb{K}^n$;
(2) i vettori $f(v_0), f(v_1), \ldots, f(v_p)$ sono linearmente dipendenti in $\mathbb{K}^{n+1}$.

222 (♡). Dati $v_1, \ldots, v_n \in \mathbb{R}^k$, con $n, k \geq 3$, provare che esiste un vettore $u \in \mathbb{R}^k$ tale che $v_1 + u$ non appartiene al sottospazio vettoriale generato da v_i e v_j, per ogni coppia di indici i, j.

223 (☕,Ⓐ,♡). Siano $p_1(x), \ldots, p_n(x) \in \mathbb{R}[x]$ polinomi distinti e tali che $p_i(0) = 0$ per ogni indice i. Provare che le n funzioni $f_1 = e^{p_1(x)}, \ldots, f_n = e^{p_n(x)}$ sono linearmente indipendenti nello spazio vettoriale su $\mathbb{R}$ delle funzioni continue sulla retta reale.

4.6. Basi e dimensione

In uno spazio vettoriale, i sottoinsiemi di vettori che sono sia generatori che linearmente indipendenti hanno un ruolo chiave in algebra lineare e pertanto meritano apposita definizione.

Definizione 4.6.1. Diremo che n vettori $v_1, \ldots, v_n$ **formano una base** di uno spazio vettoriale V se sono contemporaneamente generatori di V e linearmente indipendenti. Una **base** è una successione di generatori linearmente indipendenti.

Osservazione 4.6.2. Per un insieme finito di vettori, la proprietà di formare una base è indipendente dall'ordine in cui questi vettori sono considerati. Viceversa una base dipende dall'ordine in cui i vettori sono considerati. Dunque n generatori linearmente indipendenti formano esattamente $n!$ basi distinte.

Abbiamo osservato nell'Esempio 4.4.5 che gli n vettori di $\mathbb{K}^n$

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

formano un insieme di generatori. Mostriamo adesso che sono anche linearmente indipendenti e quindi che formano una base. Siano $a_1, \ldots, a_n \in \mathbb{K}$ tali che $\sum_{i=1}^{n} a_i e_i = 0$. Allora dalla formula

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = a_1e_1 + \cdots + a_ne_n$$

segue immediatamente che $a_1 = a_2 = \cdots = a_n = 0$.

Definizione 4.6.3. Nelle notazioni precedenti, la successione di vettori $e_1, \ldots, e_n$ viene detta **base canonica** di $\mathbb{K}^n$.

Ad esclusione del caso banale dello spazio vettoriale nullo, le basi non sono uniche. Ad esempio, se $v_1, \dots, v_n$ è una base, con $n > 0$, allora per ogni scalare $a \neq 0$ la successione di vettori $av_1, v_2, \dots, v_n$ è ancora una base (esercizio: perch é?). Come ulteriore esempio, il lettore può facilmente dimostrare come esercizio che una coppia di vettori $\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} c \\ d \end{pmatrix} \in \mathbb{K}^2$ è una base se e solo se $ad \neq cb$.

È utile avere condizioni necessarie e sufficienti su un insieme finito di vettori affinché formi una base: le due più importanti sono le caratterizzazioni come *insieme minimale di generatori* e come *insieme massimale di vettori linearmente indipendenti*; ma andiamo con ordine.

Siano V uno spazio vettoriale di dimensione finita e $v_1, \dots, v_n \subseteq V$ una successione finita di generatori. Diremo che un vettore v_i della successione è ridondante se i rimanenti vettori sono ancora un insieme di generatori. In altri termini, se $v_1, \dots, v_n$ generano V ed anche $v_2, \dots, v_n$ generano V, allora v_1 è un generatore ridondante (all'interno della successione $v_1, \dots, v_n$).

Una successione di generatori di dice **minimale** se non possiede elementi ridondanti, ossia se, comunque si tolga un vettore, i rimanenti cessano di generare lo spazio vettoriale.

Diremo invece che una successione $v_1, \dots, v_n$ di vettori linearmente indipendenti è **massimale** in V se, comunque si prenda un vettore $u \in V$, i vettori $v_1, \dots, v_n, u$ sono linearmente dipendenti.

TEOREMA 4.6.4. *Per una successione finita di vettori $v_1, \dots, v_n \in V$ le seguenti condizioni sono equivalenti:*

(1) *$v_1, \dots, v_n$ è una base;*
(2) *$v_1, \dots, v_n$ è una successione minimale di generatori;*
(3) *$v_1, \dots, v_n$ è una successione massimale di vettori linearmente indipendenti.*

DIMOSTRAZIONE. Per provare l'equivalenza delle tre condizioni dimostriamo prima che (1) implica (2), poi che (2) implica (3) ed infine che (3) implica (1).

Supponiamo che $v_1, \dots, v_n$ sia una base; per definizione tali vettori generano e vogliamo dimostrare che comunque se ne tolga uno, i rimanenti non generano. Per semplicità notazionale mostriamo che v_n non è ridondante, alla stessa maniera si dimostra che ciascun v_i non è ridondante. Siccome $v_1, \dots, v_n$ sono linearmemte indipendenti, per il Lemma 4.5.4 si ha $v_n \notin \mathrm{Span}(v_1, \dots, v_{n-1})$ e questo è possibile solo se $\mathrm{Span}(v_1, \dots, v_{n-1})$ è un sottoinsieme proprio di V.

Supponiamo adesso che $v_1, \dots, v_n$ sia una successione minimale di generatori e dimostriamo che sono anche una successione massimale di vettori linearmente indipendenti. Per il Lemma 4.4.8 la condizione di minimalità implica che $v_i \notin \mathrm{Span}(v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n)$ per ogni i. A maggior ragione $v_i \notin \mathrm{Span}(v_1, \dots, v_{i-1})$ per ogni i e quindi i vettori $v_1, \dots, v_n$ sono linearmente indipendenti per il lemma di estensione.

Dato che V è generato da n vettori, per il teorema di scambio in V esistono al massimo n vettori linearmente indipendenti e questo implica ovviamente che la successione di vettori linearmente indipendenti $v_1, \dots, v_n$ è massimale.

Per concludere, supponiamo $v_1, \dots, v_n$ una successione massimale di vettori linearmente indipendenti e proviamo che $v_1, \dots, v_n$ sono generatori. Se non lo fossero esisterebbe un vettore $u \in V$ tale che $u \notin \mathrm{Span}(v_1, \dots, v_n)$ e per il lemma di estensione $v_1, \dots, v_n, u$ sarebbero linearmente indipendenti, in contraddizione con la massimalità. □

COROLLARIO 4.6.5. *Sia A un insieme finito di generatori di uno spazio vettoriale V. Allora esiste una base $v_1, \dots, v_n$ di V con $v_i \in A$ per ogni indice i.*

DIMOSTRAZIONE. Tra tutti i sottoinsiemi $B \subseteq A$ che generano V scegliamone uno, diciamo $\{v_1, \dots, v_n\}$, con il minor numero di elementi. Chiaramente $v_1, \dots, v_n$ è una successione minimale di generatori, che è una base per il Teorema 4.6.4. □

COROLLARIO 4.6.6 (Esistenza ed equicardinalità delle basi). *Ogni spazio vettoriale di dimensione finita possiede basi. Tutte le basi sono formate dallo stesso numero di vettori.*

DIMOSTRAZIONE. Per definizione di spazio di dimensione finita esiste un insieme finito A di generatori. Per il Corollario 4.6.5 esiste una base $v_1, \dots, v_n$ con $v_i \in A$ per ogni i.

Siamo adesso $v_1, \dots, v_n$ e $w_1, \dots, w_m$ due basi dello stesso spazio vettoriale V. Siccome $v_1, \dots, v_n$ generano e $w_1, \dots, w_m$ sono linearmente indipendenti, per il teorema di scambio vale $m \leq n$. Per simmetria, ossia scambiando i ruoli, si ottiene $n \leq m$ e quindi $n = m$. □

Osservazione 4.6.7. Il concetto di base e relativo teorema di esistenza ed equicardinalità si può estendere agli spazi di dimensione infinita; ciò richiede strumenti matematici per nulla banali e verrà trattato nel Capitolo 12.

Definizione 4.6.8. Sia V uno spazio vettoriale di dimensione finita. La **dimensione** $\dim_{\mathbb{K}} V$ di V su $\mathbb{K}$ è il numero di elementi di una (qualunque) base di V. Scriveremo semplicemente $\dim V$ al posto di $\dim_{\mathbb{K}} V$ quando il campo $\mathbb{K}$ è chiaro dal contesto.

Esempio 4.6.9. Lo spazio vettoriale nullo (formato dal solo vettore nullo) è l'unico spazio vettoriale di dimensione 0.

Segue immediatamente dal Teorema 4.6.4 che la dimensione è uguale al massimo numero di vettori linearmente indipendenti ed anche al minimo numero di generatori.

Esempio 4.6.10. Si ha $\dim \mathbb{K}^n = n$, infatti la base canonica è formata da n vettori.

Esempio 4.6.11. Sia $V \subseteq \mathbb{K}[x]$ il sottospazio vettoriale dei polinomi di grado minore di n. Allora V ha dimensione n in quanto una base è data dai polinomi $1, x, x^2, \dots, x^{n-1}$.

Esempio 4.6.12. Si ha $\dim_{\mathbb{R}} \mathbb{C} = 2$ in quanto $1, i \in \mathbb{C}$ sono una base di $\mathbb{C}$ come spazio vettoriale su $\mathbb{R}$.

Lemma 4.6.13. *Per una successione $v_1, \dots, v_n$ di vettori in uno spazio vettoriale di dimensione n le seguenti condizioni sono equivalenti:*

(1) *$v_1, \dots, v_n$ è una base,*
(2) *$v_1, \dots, v_n$ sono linearmente indipendenti,*
(3) *$v_1, \dots, v_n$ sono generatori.*

Dimostrazione. Per il teorema di scambio, in uno spazio vettoriale di dimensione n, allora ogni insieme di generatori deve contenere almeno n vettori; ne segue che ogni insieme di n generatori è minimale e quindi una base.

Ancora per il teorema di scambio, in uno spazio vettoriale di dimensione n ogni insieme di n vettori linearmente indipendenti è necessariamente massimale e quindi una base. □

Esempio 4.6.14. Ogni insieme di n vettori linearmente indipendenti di $\mathbb{K}^n$ è una base.

Teorema 4.6.15 (di completamento). *Siano $v_1, \dots, v_m$ vettori linearmente indipendenti in uno spazio vettoriale V di dimensione n. Allora $m \leq n$ ed esistono $v_{m+1}, \dots, v_n \in V$ tali che $v_1, \dots, v_n$ è una base.*

Dimostrazione. Siccome V può essere generato da n vettori la disuguaglianza $m \leq n$ segue dal teorema di scambio. Dimostriamo l'esistenza dei vettori $v_{m+1}, \dots, v_n$ per induzione su $n - m$. Se $m = n$ allora $v_1, \dots, v_m$ è già una base per il Lemma 4.6.13. Se $m < n$ allora $v_1, \dots, v_m$ non sono generatori e quindi possiamo scegliere un vettore $v_{m+1} \notin \operatorname{Span}(v_1, \dots, v_m)$. Per il lemma di estensione i vettori $v_1, \dots, v_{m+1}$ sono linearmente indipendenti e per l'ipotesi induttiva esistono $v_{m+2}, \dots, v_n$ tali che $v_1, \dots, v_n$ è una base. □

Lemma 4.6.16. *Sia W un sottospazio vettoriale di uno spazio vettoriale V di dimensione finita. Allora $\dim W \leq \dim V$ e vale $\dim W = \dim V$ se e solo se $W = V$.*

Dimostrazione. Abbiamo già dimostrato che W ha dimensione finita. Se $\dim W = m$ allora W contiene m vettori $w_1, \dots, w_m$ linearmente indipendenti. Tali vettori sono linearmente indipendenti anche in V e quindi $m \leq \dim V$. Se $m = \dim V$ allora $w_1, \dots, w_m$ è una base di V e quindi $W = V$. □

Proposizione 4.6.17. *Sia $v_1, \dots, v_n$ una base di uno spazio vettoriale V. Allora per ogni vettore $v \in V$ esistono, e sono unici, dei coefficienti $a_1, \dots, a_n \in \mathbb{K}$ tali che*

$$v = a_1 v_1 + \cdots + a_n v_n.$$

DIMOSTRAZIONE. L'esistenza dei coefficienti a_i è del tutto equivalente al fatto che i vettori v_i generano V. Siccome $v_1, \dots, v_n$ sono linearmente indipendenti, se

$$v = a_1 v_1 + \cdots + a_n v_n, \qquad v = b_1 v_1 + \cdots + b_n v_n,$$

allora

$$0 = v - v = (a_1 - b_1)v_1 + \cdots + (a_n - b_n)v_n$$

da cui segue $a_i = b_i$ per ogni i, ossia l'unicità dei coefficienti. □

DEFINIZIONE 4.6.18. Si chiamano **coordinate** di un vettore v rispetto ad una base $v_1, \dots, v_n$ i coefficienti $a_1, \dots, a_n$ tali che

$$v = a_1 v_1 + \cdots + a_n v_n.$$

ESEMPIO 4.6.19. Calcoliamo le coordinate del vettore $(2, 0, 1) \in \mathbb{R}^{(3)}$ rispetto alla base $v_1 = (1, 0, 1)$, $v_2 = (0, 1, 1)$ e $v_3 = (1, 1, 0)$. In pratica dobbiamo trovare tre numeri x, y, z tali che $xv_1 + yv_2 + zv_3 = (1, 1, 1)$, ossia bisogna risolvere il sistema lineare

$$\begin{cases} x + z = 2 \\ y + z = 0 \\ x + y = 1 \end{cases}$$

la cui (unica) soluzione è $x = 3/2$, $y = -1/2$ e $z = 1/2$.

Esercizi.

224. Calcolare le coordinate del vettore $(1, 0, 0) \in \mathbb{K}^{(3)}$ rispetto alla base $v_1 = (1, 1, 1)$, $v_2 = (1, -1, 0)$, $v_3 = (0, 0, 1)$.

225. Sia $v_1, \dots, v_n$ una base di uno spazio vettoriale V sul campo $\mathbb{K}$. Dimostrare che per ogni vettore $v \in V$ esiste $t \in \mathbb{K}$ tale che i vettori

$$v_1, \dots, v_{n-1}, v + tv_n$$

sono linearmente dipendenti. Dire inoltre se un tale t è unico.

226. Siano $v_1, \dots, v_n$ una base di V e $u_1, \dots, u_m$ una base di U. Mostrare che le $n + m$ coppie

$$(v_1, 0), \dots, (v_n, 0), (0, u_1), \dots, (0, u_m)$$

sono una base di $V \times U$.

227. Siano dati $n+1$ vettori $v_0, \dots, v_n$ in uno spazio vettoriale ed $n+1$ scalari $a_0, \dots, a_n$. Provare che il sottospazio vettoriale generato dagli $n(n-1)/2$ vettori

$$v_{ij} = a_i v_j - a_j v_i, \qquad 0 \le i < j \le n,$$

ha dimensione al più n.

228. Siano V uno spazio vettoriale di dimensione finita n, $H \subset V$ un sottospazio vettoriale proprio e $v, w \in V$ vettori non appartenenti ad H. Dimostrare che esiste un sottospazio vettoriale $W \subset V$ di dimensione $n - 1$ tale che $H \subset W$ e $v, w \notin W$.

229 (Codimensione). Siano V uno spazio vettoriale ed $U \subseteq V$ un sottospazio. Diremo che U ha codimensione finita in V se esiste una successione finita di vettori $v_1, \dots, v_n \in V$ tali che $U + \operatorname{Span}(v_1, \dots, v_n) = V$. Se i vettori v_i sono linearmente indipendenti e $V = U \oplus \operatorname{Span}(v_1, \dots, v_n)$ diremo che $v_1, \dots, v_n$ è una cobase di U in V.

Sia $\mathcal{A} \subseteq V$ un insieme finito di vettori tale che $U + \operatorname{Span}(\mathcal{A}) = V$ e sia $v_1, \dots, v_n \in \mathcal{A}$ una successione di lunghezza minima tale che $U + \operatorname{Span}(v_1, \dots, v_n) = V$. Provare che $v_1, \dots, v_n$ è una cobase.

Dedurre dal teorema di scambio che due cobasi di uno spazio di codimensione finita hanno lo stesso numero di elementi; tale numero viene detto **codimensione** di U in V.

230 (☕, ♡). Siano $\mathbb{K}$ un campo e $F \subseteq \mathbb{K}$ un sottocampo; dunque $F^n \subseteq \mathbb{K}^n$ per ogni n. Dati m vettori $v_1, \dots, v_m \in F^n$ linearmente indipendenti su F, provare che sono linearmente indipendenti anche come vettori di $\mathbb{K}^n$.

4.7. Semisemplicità e formula di Grassmann

DEFINIZIONE 4.7.1. Siano U, W due sottospazi di V. Diremo che W è un **complementare di U in V** se vale $V = U \oplus W$.

Notiamo subito che, in generale, un sottospazio possiede più di un complementare. Ad esempio sullo spazio $\mathbb{R}^2$, ogni coppia di rette distinte passanti per l'origine sono una il complementare dell'altra. Gli spazi vettoriali godono di una importante proprietà, detta *semisemplicità*, non sempre valida in altre strutture algebriche ed espressa dal prossimo teorema.

TEOREMA 4.7.2 (Semisemplicità degli spazi vettoriali). *Sia U un sottospazio di uno spazio vettoriale V di dimensione finita. Allora esiste un complementare di U in V, ossia esiste un sottospazio vettoriale $W \subseteq V$ tale che $V = U \oplus W$. Inoltre vale* $\dim W = \dim V - \dim U$.

DIMOSTRAZIONE. Sia $v_1, \dots, v_r$ una base di U, allora i vettori v_i sono linearmente indipendenti in V e possono estesi ad una base $v_1, \dots, v_n$. Proviamo che il sottospazio $W = \mathrm{Span}(v_{r+1}, \dots, v_n)$ ha le proprietà richieste. Innanzitutto

$$U + W = \mathrm{Span}(v_1, \dots, v_r) + \mathrm{Span}(v_{r+1}, \dots, v_n) = \mathrm{Span}(v_1, \dots, v_n) = V$$

e se $x \in U \cap W$, dalla condizione $x \in U$ si ricava

$$x = a_1 v_1 + \cdots + a_r v_r, \qquad a_i \in \mathbb{K},$$

mentre dalla condizione $x \in W$ si ha

$$x = a_{r+1} v_{r+1} + \cdots + a_n v_n, \qquad a_i \in \mathbb{K},$$

$$a_1 v_1 + \cdots + a_r v_r - a_{r+1} v_{r+1} - \cdots - a_n v_n = x - x = 0$$

e quindi $a_i = 0$ per ogni i in quanto $v_1, \dots, v_n$ linearmente indipendenti. Ma questo implica in particolare che $x = 0$ e di conseguenza $U \cap W = 0$. Per finire basta osservare che $n = \dim V$, $r = \dim U$ e che $v_{r+1}, \dots, v_n$ formano una base di W. □

OSSERVAZIONE 4.7.3. La prima parte del Teorema 4.7.2 è vero anche se V ha dimensione infinita (vedi Teorema 12.5.5), ma in tal caso la dimostrazione è decisamente più complicata ed è posticipata al Capitolo 12.

TEOREMA 4.7.4 (Formula di Grassmann). *Siano U, W due sottospazi di dimensione finita di uno spazio vettoriale V. Allora vale la formula*

$$\dim(U + W) = \dim U + \dim W - \dim(U \cap W).$$

DIMOSTRAZIONE. Denotiamo con $n = \dim U$, $m = \dim W$ e con $p = \dim(U \cap W)$; siccome $U \cap W$ è un sottospazio di U vale $p \leq n$ e similmente $p \leq m$. Consideriamo una base $e_1, \dots, e_p$ di $U \cap W$; i vettori $e_1, \dots, e_p$ sono linearmente indipendenti in U e quindi possono essere completati ad una base $e_1, \dots, e_p, u_{p+1}, \dots, u_n$ di U. Similmente i vettori $e_1, \dots, e_p$ possono essere completati ad una base $e_1, \dots, e_p, w_{p+1}, \dots, w_m$ di W. Basta adesso dimostrare che gli $n + m - p$ vettori

$$e_1, \dots, e_p, u_{p+1}, \dots, u_n, w_{p+1}, \dots, w_m$$

formano una base di $U + W$. Abbiamo già visto che tali vettori generano $U + W$; resta quindi da dimostrare che sono linearmente indipendenti.

Siano dunque $a_1, \dots, a_p, b_{p+1}, \dots, b_n, c_{p+1}, \dots, c_m \in \mathbb{K}$ tali che

$$a_1 e_1 + \cdots + a_p e_p + b_{p+1} u_{p+1} + \cdots + b_n u_n + c_{p+1} w_{p+1} + \cdots + c_m w_m = 0;$$

dobbiamo dimostrare che $a_i = b_j = c_k = 0$ per ogni i, j, k.
Consideriamo il vettore di W

$$w = -(c_{p+1} w_{p+1} + \cdots + c_m w_m).$$

Dalla relazione precedente segue che

$$w = a_1 e_1 + \cdots + a_p e_p + b_{p+1} u_{p+1} + \cdots + b_n u_n \in U$$

e quindi $w \in U \cap W$. Se indichiamo con $d_1, \dots, d_p$ i coefficienti di w nella base $e_1, \dots, e_p$ si ha

$$0 = w - w = d_1 e_1 + \cdots + d_p e_p + c_{p+1} w_{p+1} + \cdots + c_m w_m$$

e siccome $e_1, \dots, e_p, w_{p+1}, \dots, w_m$ è una base di W deve essere $d_i = c_j = 0$ per ogni i, j e quindi $w = 0$, ossia

$$w = a_1 e_1 + \cdots + a_p e_p + b_{p+1} u_{p+1} + \cdots + b_n u_n = 0.$$

Siccome $e_1, \dots, e_p, u_{p+1}, \dots, u_n$ è una base di U deve essere $a_i = b_j = 0$ per ogni i, j, come volevasi dimostrare. □

Esercizi.

231. Trovare una bigezione (=applicazione bigettiva) tra l'insieme dei sottospazi complementari in $\mathbb{R}^3$ al piano π di equazione $x + y + z = 0$ e l'insieme dei vettori $(a, b, c) \in \mathbb{R}^3$ tali che $a + b + c = 1$.

232. Sia V uno spazio vettoriale di dimensione m e sia $U_1, U_2, \dots$ una successione di sottospazi vettoriali di dimensione $m - 1$. Dimostrare per induzione su n che la dimensione di $U_1 \cap U_2 \cap \cdots \cap U_n$ è maggiore od uguale a $m - n$.

233. Sia V uno spazio vettoriale di dimensione n e $U, W \subseteq V$ due sottospazi tali che $U \oplus W = V$. Dimostrare che per ogni sottospazio $U \subseteq L$ vale $L = (L \cap W) \oplus U$.

234. Sia V uno spazio vettoriale di dimensione n e $W \subseteq V$ un sottospazio di dimensione $m < n$. Dimostrare che W si può scrivere come intersezione di $n - m$ sottospazi vettoriali di dimensione $n - 1$. (Sugg.: estendere una base di W ad una base di V.)

235. Provare che se se $\dim V < \infty$ e $V = U_1 \oplus \cdots \oplus U_n$ allora $\dim V = \dim U_1 + \cdots + \dim U_n$.

236. Siano H, K sottospazi vettoriali di uno spazio vettoriale V di dimensione finita. Dimostrare che esiste un sottospazio vettoriale $L \subseteq V$ con le seguenti proprietà:

$$K \subseteq L, \qquad H + L = V, \qquad H \cap L = H \cap K\,.$$

237. Siano V uno spazio vettoriale ed $U \subset V$ un sottospazio. Diremo che una successione di vettori $v_1, \dots, v_n \in V$ è **omogenea rispetto ad** U se

$$U \cap \mathrm{Span}(v_1, \dots, v_n) = \mathrm{Span}(\{v_i \mid v_i \in U\}).$$

Supponendo V di dimensione finita, dire per quali valori di k è vera la seguente affermazione: dati k sottospazi vettoriali $U_1, \dots, U_k$ di V ed una successione di vettori linearmente indipendenti, che è omogenea rispetto a ciascun U_k, è possibile estendere tale successione ad una base omogenea rispetto a ciascun U_k.

Rispondere poi alla stessa domanda sotto l'ipotesi aggiuntiva che i sottospazi U_i formino una filtrazione crescente, ossia $U_1 \subseteq U_2 \subseteq \cdots \subseteq U_k$.

4.8. Complementi: i numeri algebrici

Anticipiamo in questa sezione alcuni argomenti che saranno trattati in maniera più approfondita nel Capitolo 17.

Definizione 4.8.1. Un **numero algebrico** è un numero complesso che è radice di un polinomio non nullo a coefficienti interi. Un numero complesso che non è algebrico viene detto **trascendente**.

Più precisamente, un numero $\alpha \in \mathbb{C}$ è algebrico se e solo se esistono un intero positivo n ed $n + 1$ numeri interi $a_0, a_1, \dots, a_n$ tali che

$$a_n \neq 0, \qquad a_0 + a_1 \alpha + \cdots + a_n \alpha^n = 0.$$

Ogni numero razionale $x = p/q$ è algebrico poiché è radice del polinomio $qx - p$. Il numero $\sqrt{2}$ è algebrico poiché è radice del polinomio $x^2 - 2$. L'unità immaginaria i è un numero algebrico poiché è radice del polinomio $x^2 + 1$.

È facilissimo dimostrare in maniera non costruttiva l'esistenza di numeri trascendenti: per ogni intero $n > 0$ indichiamo con $S_n \subset \mathbb{C}$ l'insieme dei numeri complessi che sono radice di un polinomio di grado $\leq n$ i cui coefficienti sono numeri interi di valore assoluto $\leq n$. Per definizione un numero è algebrico se e solo se appartiene all'unione degli S_n.

Siccome esistono al più $(2n+1)^{n+1}$ polinomi di grado $\leq n$ con coefficienti interi di valore assoluto $\leq n$ (esercizio: perché?), e siccome ogni polinomio siffatto possiede al più n radici complesse, ne consegue che ogni S_n è un sottoinsieme finito, con al più $n(2n+1)^{n+1}$ elementi. A maggior ragione, per ogni n ed ogni $a, b \in \mathbb{R}$, $a \leq b$, l'insieme $S_n \cap [a,b] = \{x \in \mathbb{R} \cap S_n \mid a \leq x \leq b\}$ è finito.

Possiamo quindi trovare due successioni $a_n, b_n \in \mathbb{R}$ tali che

$$a_n \leq a_{n+1} < b_{n+1} \leq b_n, \qquad [a_n, b_n] \cap S_n = \emptyset,$$

per ogni n. Siccome $a_n \leq b_m$ per ogni $n, m \in \mathbb{N}$, per il principio di completezza esiste $\xi \in \mathbb{R}$ tale che $a_n \leq \xi \leq b_m$ per ogni $n, m \in \mathbb{N}$. Di conseguenza $\xi \notin S_n$ per ogni n e quindi ξ non è algebrico.

Non è invece facile dimostrare che determinati numeri sono trascendenti: ad esempio non è affatto banale dimostrare i seguenti tre teoremi.

TEOREMA 4.8.2 (Liouville 1844). *Il numero di Liouville*

$$\sum_{n=1}^{+\infty} \frac{1}{10^{n!}} = 0.11000100000000000000000010000000\ldots$$

è trascendente.

TEOREMA 4.8.3 (Hermite 1873). *Il numero di Nepero*

$$e = \sum_{n=0}^{+\infty} \frac{1}{n!} = \lim_{n\to+\infty} \left(1 + \frac{1}{n}\right)^n = 2.71828182845904523536028 74\ldots$$

(base dei logaritmi naturali) è trascendente.

TEOREMA 4.8.4 (Lindemann 1882). *La costante di Archimede*

$$\pi = 3,14159265358979323846\ldots$$

è trascendente.

Le dimostrazioni dei Teoremi 4.8.2, 4.8.3 e 4.8.4 richiedono alcuni risultati di analisi matematica e sono riportate nel Capitolo 17.

Usando le proprietà base degli spazi vettoriali, possiamo dimostrare da subito che i numeri algebrici formano un campo di numeri. A tal fine è utile considerare $\mathbb{C}$ come uno spazio vettoriale su $\mathbb{Q}$.

LEMMA 4.8.5. *Sia $U \subseteq \mathbb{C}$ un sottospazio vettoriale di dimensione finita su $\mathbb{Q}$. Se $1 \in U$ e $uv \in U$ per ogni $u, v \in U$, allora U è un campo ed ogni elemento di U è un numero algebrico.*

DIMOSTRAZIONE. Per dimostrare che U è un sottocampo di $\mathbb{C}$ basta dimostrare che se $u \in U$, $u \neq 0$, allora $u^{-1} \in U$. Sia n la dimensione di U come spazio vettoriale su $\mathbb{Q}$ e prendiamo un qualsiasi sottoinsieme $v_1, \ldots, v_n \in U$ di generatori linearmente indipendenti. Allora gli n vettori $uv_1, \ldots, uv_n \in U$ sono ancora linearmente indipendenti su $\mathbb{Q}$: infatti se

$$a_1 uv_1 + \cdots + a_n uv_n = 0, \qquad a_1, \ldots, a_n \in \mathbb{Q},$$

possiamo scrivere $u(a_1 v_1 + \cdots + a_n v_n) = 0$ e poiché $u \neq 0$ si deve avere $a_1 v_1 + \cdots + a_n v_n = 0$, da cui $a_1 = \cdots = a_n = 0$. Per il Lemma 4.6.13 i vettori $uv_1, \ldots, uv_n$ sono un insieme di generatori e quindi esistono $b_1, \ldots, b_n \in \mathbb{Q}$ tali che

$$1 = b_1 uv_1 + \cdots + b_n uv_n = u(b_1 v_1 + \cdots + b_n v_n),$$

ossia $u^{-1} = b_1 v_1 + \cdots + b_n v_n$.

Dimostriamo adesso che ogni $u \in U$ è un numero algebrico. Se U ha dimensione n su $\mathbb{Q}$, allora gli $n+1$ vettori

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

sono linearmente dipendenti. Esistono quindi $n+1$ numeri razionali $a_0, \ldots, a_n$ non tutti nulli a tali che

$$a_0 + a_1 u + a_2 u^2 + \cdots + a_n u^n = 0.$$

Moltiplicando per un denominatore comune non è restrittivo supporre $a_i \in \mathbb{Z}$ per ogni i e quindi u è radice di un polinomio non nullo a coefficienti interi di grado $\leq n$. □

DEFINIZIONE 4.8.6. Il **grado** di un numero algebrico $\alpha \in \mathbb{C}$ è il più piccolo intero positivo d tale che x è radice di un polinomio di grado d a coefficienti interi.

ESEMPIO 4.8.7. Ogni numero razionale $x = p/q$ è algebrico di grado 1 poiché è radice del polinomio $qx - p$. Il numero $\sqrt{2}$ è algebrico di grado 2 poiché non è razionale ed è radice del polinomio $x^2 - 2$. L'unità immaginaria i è un numero algebrico di grado 2 poiché non è razionale (e nemmeno reale) ed è radice del polinomio $x^2 + 1$.

DEFINIZIONE 4.8.8. Sia $u \in \mathbb{C}$ un numero algebrico, e sia n il suo grado. Denotiamo con $\mathbb{Q}[u] \subseteq \mathbb{C}$ il sottospazio vettoriale

$$\mathbb{Q}[u] = \mathrm{Span}(1, u, \dots, u^{n-1}) = \{a_0 + a_1 u + a_2 u^2 + \cdots + a_{n-1}u^{n-1} \mid a_0, \dots, a_{n-1} \in \mathbb{Q}\}.$$

Se $w \in \mathbb{C}$ è un altro numero algebrico, di grado m, indichiamo $\mathbb{Q}[u, w] \subseteq \mathbb{C}$ il sottospazio vettoriale

$$\mathbb{Q}[u, w] = \mathrm{Span}(1, u, w, \dots, u^i w^j, \dots), \qquad 0 \le i < n,\ 0 \le j < m.$$

Notiamo che $\mathbb{Q}[u]$ e $\mathbb{Q}[u, w]$ sono finitamente generati e quindi di dimensione finita come spazi vettoriali su $\mathbb{Q}$.

LEMMA 4.8.9. *Siano u, w numeri algebrici di gradi n, m rispettivamente, allora $\mathbb{Q}[u, w]$ è un campo.*

DIMOSTRAZIONE. Sappiamo che $\mathbb{Q}[u, w]$ è un sottospazio vettoriale di $\mathbb{C}$ che contiene $\mathbb{Q}$ e di dimensione finita. Basta quindi dimostrare che se $v_1, v_2 \in \mathbb{Q}[u, w]$ allora $v_1 v_2 \in \mathbb{Q}[u, w]$. Dimostriamo come primo passo che $u^i w^j \in \mathbb{Q}[u, w]$ per ogni $i, j \ge 0$. Se $i < n$ e $j < m$ ciò è vero per definizione. Per induzione su $i + j$ basta dimostrare che se $i \ge n$ o $j \ge m$ possiamo scrivere $u^i w^j$ come combinazione lineare di monomi $u^a w^b$, con $a + b < i + j$. Supponiamo per fissare le idee che $i \ge n$; quando $j \ge m$ basterà ripetere il ragionamento con w al posto di u. Siccome u ha grado n vale una relazione del tipo

$$b_0 + b_1 u + \cdots + b_n u^n = 0, \qquad b_i \in \mathbb{Z},\ b_n \neq 0,$$

e quindi per ogni $i \ge n$ ed ogni j vale

$$u^i w^j = u^n u^{i-n} w^j = \left(-\frac{b_0}{b_n} - \frac{b_1}{b_n} u - \cdots - \frac{b_{n-1}}{b_n} u^{n-1}\right) u^{i-n} w^j.$$

Se $v_1, v_2 \in \mathbb{Q}[u, w]$, allora v_1, v_2 sono entrambi combinazioni lineari a coefficienti razionali di $u^i w^j$ e quindi il prodotto $v_1 v_2$ è una combinazione lineare di $u^i w^j$. Per quanto visto $u^i w^j \in \mathbb{Q}[u, w]$ per ogni $i, j \ge 0$ e dunque $v_1 v_2$ è combinazione lineare di vettori del sottospazio $\mathbb{Q}[u, w]$. □

TEOREMA 4.8.10. *Siano $u, w \in \mathbb{C}$ due numeri algebrici. Allora i numeri $-u, u + w, uw$ sono ancora algebrici e, se $u \neq 0$, allora anche u^{-1} è algebrico. In altre parole, l'insieme $\overline{\mathbb{Q}}$ dei numeri algebrici è un sottocampo di $\mathbb{C}$.*

DIMOSTRAZIONE. Per i due lemmi precedenti, se u, w sono algebrici il sottospazio vettoriale $\mathbb{Q}[u, w]$ ha dimensione finita su $\mathbb{Q}$, è un campo di numeri ed ogni elemento di $\mathbb{Q}[u, w]$ è algebrico. In particolare sono algebrici i numeri

$$-u, u^{-1}, u + w, uw \in \mathbb{Q}[u, w].$$

□

I numeri trascendenti sono "di più" dei numeri algebrici, nel senso indicato dal seguente teorema.

TEOREMA 4.8.11 (Cantor 1874). *Siano $\overline{\mathbb{Q}} \subseteq \mathbb{C}$ l'insieme dei numeri algebrici e $T \subseteq \mathbb{C}$ l'insieme dei numeri trascendenti. Allora esistono applicazioni iniettive $\overline{\mathbb{Q}} \to T$, ma **non esiste** alcuna applicazione surgettiva $\overline{\mathbb{Q}} \to T$.*

DIMOSTRAZIONE. Abbiamo già dimostrato che $T \neq \emptyset$ all'inizio della sezione come conseguenza dei seguenti fatti:

(1) $\overline{\mathbb{Q}}$ è unione di una successione $S_1, S_2, S_3, \dots$ di insiemi finiti;
(2) per ogni successione $F_1, F_2, F_3, \dots$ di sottoinsiemi finiti di $\mathbb{C}$ esiste un numero reale $\xi \in \mathbb{R}$ tale $\xi \notin F_n$ per ogni $n \in \mathbb{N}$.

Sia $\xi \in T$ un qualsiasi elemento, allora $\xi - a \in T$ per ogni $a \in \overline{\mathbb{Q}}$: infatti, se fosse $\xi - a \notin T$, allora $\xi - a \in \overline{\mathbb{Q}}$ e di conseguenza anche $\xi = (\xi - a) + a \in \overline{\mathbb{Q}}$ per il Teorema 4.8.10. L'applicazione $\overline{\mathbb{Q}} \to T$, $a \mapsto \xi - a$, è iniettiva.

Supponiamo per assurdo che esista un'applicazione surgettiva $f\colon \overline{\mathbb{Q}} \to T$ e scriviamo $\overline{\mathbb{Q}} = \bigcup_{i=1}^{\infty} S_i$ con S_n finito per ogni n. Allora $F_n = S_n \cup f(S_n)$ è un sottoinsieme finito e l'unione degli F_n coincide con l'unione $\overline{\mathbb{Q}} \cup T = \mathbb{C}$, che abbiamo visto essere falso. □

Esercizi.

238. Vero o falso?
(1) La somma di due numeri trascendenti è trascendente.
(2) L'inverso di un numero trascendente è trascendente.
(3) Il coniugato di un numero trascendente è trascendente.

239. Dimostrare che se $u \neq 0$ è un numero algebrico, allora anche $-u$ e u^{-1} sono algebrici.

240. Sapendo che esistono numeri trascendenti, dimostrare la prima parte del Teorema 4.8.11, ossia che esistono applicazioni iniettive $\overline{\mathbb{Q}} \to T$.

241. Sia $u \in \mathbb{C}$ un numero algebrico di grado $n > 0$. Dimostrare che i numeri

$$1, u, u^2, \dots, u^{n-1}$$

sono linearmente indipendenti su $\mathbb{Q}$.

242 (☕). Possiamo generalizzare la definizione di $\mathbb{Q}[u, w]$ ad una successione finita di $u_1, \dots, u_n$ di numeri algebrici, ponendo $\mathbb{Q}[u_1, \dots, u_n]$ uguale al sottospazio vettoriale generato da tutti i monomi

$$u_1^{i_1} \cdots u_n^{i_n}, \qquad 0 \leq i_j < \text{grado di } u_j.$$

Si assuma che $u_i^2 \in \mathbb{Q}$ per ogni i e che $\mathbb{Q}[u_1, \dots, u_n]$ abbia dimensione 2^n su $\mathbb{Q}$. Dimostrare che se $u \in \mathbb{Q}[u_1, \dots, u_n]$ e $u^2 \in \mathbb{Q}$, allora u è un multiplo razionale di un monomio $u_1^{i_1} \cdots u_n^{i_n}$.

243 (☕). Siano $\mathbb{K} \subseteq L \subseteq \mathbb{C}$ due sottocampi, con $\mathbb{K}$ di dimensione finita come spazio vettoriale su $\mathbb{Q}$ e L di dimensione finita come spazio vettoriale su $\mathbb{K}$. Dimostrare che L ha dimensione finita come spazio vettoriale su $\mathbb{Q}$ e vale la formula

$$\dim_{\mathbb{K}} L = \frac{\dim_{\mathbb{Q}} L}{\dim_{\mathbb{Q}} \mathbb{K}}.$$

Dedurre che ogni somma di radici quadrate di numeri razionali è un numero algebrico di grado uguale ad una potenza di 2.

CAPITOLO 5

Applicazioni lineari

Dopo aver studiato le principali proprietà di singoli spazi vettoriali, in questo capitolo inizieremo a studiare le possibili interazioni fra due o più spazi vettoriali introducendo la nozione di applicazione lineare. Rispetto alle generiche applicazioni di tipo insiemistico, le applicazioni lineari godono di notevoli proprietà che le rendono contemporaneamente utili e più semplici da studiare. Infine, l'insieme di tutte le applicazioni lineari tra due spazi fissati forma a sua volta uno spazio vettoriale, al quale si applicano quindi tutti i risultati visti nel capitolo precedente.

5.1. Applicazioni lineari

In parole semplici, un'applicazione tra spazi vettoriali si dice lineare se commuta con le combinazioni lineari.

DEFINIZIONE 5.1.1. Siano V, W due spazi vettoriali sullo stesso campo $\mathbb{K}$. Un'applicazione $f\colon V \to W$ si dice **lineare** (su $\mathbb{K}$) se commuta con le somme ed i prodotti per scalare, ossia se

$$f(u+v) = f(u) + f(v), \quad f(tv) = tf(v), \quad \text{per ogni} \quad u, v \in V,\ t \in \mathbb{K}.$$

Ad esempio:

(1) L'applicazione nulla, che manda ogni vettore nel vettore nullo, è lineare.
(2) Per ogni spazio vettoriale V, **l'identità** $\mathrm{Id}_V\colon V \to V$, $\mathrm{Id}_V(v) = v$, è lineare.

Notiamo subito che se $f\colon V \to W$ è lineare, allora $f(0) = 0$: infatti, siccome $0 + 0 = 0$ si ha $f(0) = f(0+0) = f(0) + f(0)$ da cui segue $f(0) = 0$. In particolare un'applicazione costante tra spazi vettoriali è lineare se e solo se è nulla.

ESEMPIO 5.1.2. L'applicazione

$$f\colon \mathbb{K} \to \mathbb{K}, \qquad f(u) = au + b, \quad a, b \in \mathbb{K},$$

è lineare se e solo se $b = 0$. Infatti se f è lineare allora $b = f(0)$ da cui segue $b = 0$. Viceversa se $f(u) = au$ per ogni $u \in \mathbb{K}$, allora

$$f(u+v) = a(u+v) = au + av = f(u) + f(v), \qquad f(tv) = atv = tav = tf(v),$$

per ogni $t, u, v \in \mathbb{K}$.

ESEMPIO 5.1.3. Per **omotetia** si intende un'applicazione di uno spazio vettoriale in sé ottenuta moltiplicando tutti i vettori per uno stesso scalare diverso da 0. Lo stesso ragionamento dell'Esempio 5.1.2 prova che le omotetie sono applicazioni lineari.

ESEMPIO 5.1.4. Le proiezioni $\pi_i\colon \mathbb{K}^n \to \mathbb{K}$, definite come

$$\pi_i \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_i, \qquad i = 1, \dots, n,$$

sono applicazioni lineari: ciò segue immediatamente dalla struttura di spazio vettoriale su $\mathbb{K}^n$.

Sia $f\colon V \to W$ un'applicazione lineare, allora per ogni vettore $v \in V$ vale

$$f(-v) = f((-1)v) = (-1)f(v) = -f(v)$$

e più in generale per ogni $u, v \in V$ vale

$$f(u - v) = f(u + (-v)) = f(u) + (-f(v)) = f(u) - f(v).$$

Infine, come detto in precedenza, *le applicazioni lineari commutano con le combinazioni lineari*, ossia per ogni $v_1, \dots, v_n \in V$ e per ogni $a_1, \dots, a_n \in \mathbb{K}$ si ha

$$f(a_1v_1 + \cdots + a_nv_n) = a_1f(v_1) + \cdots + a_nf(v_n).$$

Possiamo dimostrare la precedente relazione per induzione su n, essendo per definizione vera per $n = 1$: si può dunque scrivere

$$\begin{aligned} f(a_1v_1 + \cdots + a_nv_n) &= f(a_1v_1 + \cdots + a_{n-1}v_{n-1}) + f(a_nv_n) \\ &= (a_1f(v_1) + \cdots + a_{n-1}f(v_{n-1})) + a_nf(v_n). \end{aligned}$$

LEMMA 5.1.5. *Sia $f\colon V \to W$ un'applicazione lineare bigettiva. Allora l'applicazione inversa $f^{-1}\colon W \to V$ è lineare.*

DIMOSTRAZIONE. Per ogni $w_1, w_2 \in W$ si hanno le uguaglianze

$$\begin{aligned} w_1 + w_2 &= f(f^{-1}(w_1 + w_2)), \\ w_1 + w_2 &= f(f^{-1}(w_1)) + f(f^{-1}(w_2)) = f(f^{-1}(w_1) + f^{-1}(w_2)), \end{aligned}$$

dove nella seconda abbiamo utilizzato la linearità di f. Uguagliando i termini a destra otteniamo

$$f(f^{-1}(w_1 + w_2)) = f(f^{-1}(w_1) + f^{-1}(w_2))$$

e dall'iniettività di f segue $f^{-1}(w_1 + w_2) = f^{-1}(w_1) + f^{-1}(w_2)$. La dimostrazione che f^{-1} commuta con i prodotti per scalare è del tutto analoga ed è lasciata per esercizio. □

DEFINIZIONE 5.1.6. Un **isomorfismo** (lineare) di spazi vettoriali è un'applicazione lineare bigettiva. Due spazi vettoriali si dicono **isomorfi** se esiste un isomorfismo tra di loro.

Possiamo quindi riformulare il Lemma 5.1.5 dicendo che l'inverso di un isomorfismo lineare è ancora un isomorfismo lineare.

ESEMPIO 5.1.7. Siano X un insieme, V uno spazio vettoriale e $f\colon X \to V$ un'applicazione bigettiva. Possiamo allora usare f per definire su X una struttura di spazio vettoriale ponendo

$$x + y = f^{-1}(f(x) + f(y)), \qquad ax = f^{-1}(af(x)).$$

Tale struttura è l'unica che rende f un isomorfismo lineare.

PROPOSIZIONE 5.1.8. *Sia $f\colon V \to W$ un'applicazione lineare:*

1. *se f è iniettiva, allora trasforma vettori linearmente indipendenti in vettori linearmente indipendenti;*
2. *se f è surgettiva, allora trasforma generatori in generatori;*

In particolare, se f è un isomorfismo lineare tra spazi vettoriali di dimensione finita, allora trasforma basi in basi.

DIMOSTRAZIONE. Siano $v_1, \dots, v_n \in V$ vettori linearmente indipendenti. Siano $a_1, \dots, a_n \in \mathbb{K}$ tali che $a_1f(v_1) + \cdots + a_nf(v_n) = 0$, allora

$$f(a_1v_1 + \cdots + a_nv_n) = a_1f(v_1) + \cdots + a_nf(v_n) = 0.$$

Dunque $f(a_1v_1 + \cdots + a_nv_n) = f(0)$ e se f è iniettiva allora $a_1v_1 + \cdots + a_nv_n = 0$; per l'indipendenza lineare dei vettori v_i se ne deduce che $a_1 = \cdots = a_n = 0$.

Supponiamo f surgettiva e siano $v_1, \dots, v_n \in V$ un insieme di generatori di V. Dato un qualsiasi vettore $w \in W$ esiste $v \in V$ tale che $f(v) = w$ ed è possibile trovare $a_1, \dots, a_n \in \mathbb{K}$ tali che $a_1v_1 + \cdots + a_nv_n = v$. Ne segue che $w = f(v) = a_1f(v_1) + \cdots + a_nf(v_n)$ e quindi che $f(v_1), \dots, f(v_n)$ generano W. □

ESEMPIO 5.1.9. Utilizziamo la Proposizione 5.1.8 per determinare una base del sottospazio

$$V = \left\{ \begin{pmatrix} x \\ y \\ x \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; 2x + 3y - z = 0 \right\}.$$

Dato che V è definito dalla relazione $z = 2x + 3y$, è chiaro che l'applicazione lineare

$$f\colon \mathbb{R}^2 \to V, \qquad f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \\ 2x + 3y \end{pmatrix}$$

è bigettiva. Dunque l'immagine tramite f della base canonica di $\mathbb{R}^2$ è una base di V:

$$f\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}1\\0\\2\end{pmatrix}, \qquad f\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}0\\1\\3\end{pmatrix}.$$

COROLLARIO 5.1.10. *Sia $f\colon V \to W$ lineare:*

1. *se f è iniettiva e W ha dimensione finita, allora anche V ha dimensione finita e $\dim V \leq \dim W$;*
2. *se f è surgettiva e V ha dimensione finita, allora anche W ha dimensione finita e $\dim W \leq \dim V$;*
3. *se f è bigettiva, allora V e W hanno la stessa dimensione (che può essere finita o infinita).*

DIMOSTRAZIONE. Se f è iniettiva e se $\dim W = n < \infty$, allora per la Proposizione 5.1.8 esistono al più n vettori linearmente indipendenti in V e quindi $\dim V \leq \dim W$.

Se f è surgettiva e V ha dimensione finita, allora f trasforma basi in generatori e questo implica $\dim W \leq \dim V$. □

Se $f, g\colon V \to W$ sono due applicazioni lineari, possiamo definire la loro somma

$$f+g\colon V \to W, \qquad (f+g)(v) = f(v) + g(v) \text{ per ogni } v \in V,$$

che risulta ancora essere lineare. Allo stesso modo risultano lineari la differenza

$$f-g\colon V \to W, \qquad (f-g)(v) = f(v) - g(v) \text{ per ogni } v \in V,$$

e più in generale qualsiasi combinazione lineare

$$af+bg\colon V \to W, \qquad (af+bg)(v) = af(v) + bg(v), \qquad a, b \in \mathbb{K}.$$

ESEMPIO 5.1.11. L'applicazione

$$f\colon \mathbb{K}^2 \to \mathbb{K}, \qquad f\begin{pmatrix}x_1\\x_2\end{pmatrix} = x_1 + x_2,$$

è lineare ed è uguale alla somma delle due proiezioni: $f = \pi_1 + \pi_2$.

ESEMPIO 5.1.12. Fissato un qualsiasi scalare $a \in \mathbb{K}$, l'applicazione (vedi Esempio 4.2.6)

$$\mathbb{K}[x] \to \mathbb{K}, \qquad p(x) \mapsto p(a),$$

che ad ogni polinomio associa il valore in a della corrispondente funzione polinomiale, è un'applicazione lineare.

LEMMA 5.1.13. *La composizione di applicazioni lineari è ancora lineare. La composizione di isomorfismi è ancora un isomorfismo.*

DIMOSTRAZIONE. Siano $V \xrightarrow{f} W \xrightarrow{g} Z$ due applicazioni lineari. Per ogni $u, v \in V$ si ha

$$g \circ f(u+v) = g(f(u+v)) = g(f(u) + f(v)) = g(f(u)) + g(f(v)) = g \circ f(u) + g \circ f(v).$$

Similmente per ogni $v \in V$ ed ogni $a \in \mathbb{K}$ si ha

$$g \circ f(av) = g(f(av)) = g(af(v)) = ag(f(v)) = ag \circ f(v).$$

Se f e g sono isomorfismi, allora sono entrambe bigettive e quindi anche la loro composizione è bigettiva. □

OSSERVAZIONE 5.1.14. Tranne il caso banale in cui gli spazi hanno dimensione 0, se esiste un isomorfismo $f\colon V \to W$ tra spazi vettoriali, allora in generale ne esistono molti altri: ad esempio, se il campo base possiede più di due elementi possiamo considerare ad esempio i multipli λf, $\lambda \in \mathbb{K}$, $\lambda \neq 0, 1$. Tuttavia, in certi casi esiste un isomorfismo con caratteristiche tali da renderlo indipendente da scelte soggettive. In tal caso diremo che l'isomorfismo è *naturale* oppure *canonico*.

Se $f\colon V \to W$ è lineare, allora per ogni sottospazio $U \subseteq V$ la restrizione

$$f_{|U}\colon U \to W$$

è ancora lineare.

Una delle principali proprietà delle applicazioni lineari è che esse sono univocamente definite dai valori che assumono in una base.

TEOREMA 5.1.15. *Siano V, W spazi vettoriali, $(v_1, \dots, v_n)$ una base di V e $w_1, \dots, w_n$ vettori qualsiasi di W. Allora vi è un'unica applicazione lineare $f\colon V \to W$ tale che $f(v_i) = w_i$ per ogni indice i.*

DIMOSTRAZIONE. Se una tale f esiste allora è necessariamente unica: infatti ogni vettore $v \in V$ si scrive in maniera unica come $v = a_1 v_1 + \cdots + a_n v_n$ e quindi per linearità si ottiene

$$f(v) = f(a_1 v_1 + \cdots + a_n v_n) = a_1 f(v_1) + \cdots + a_n f(v_n) = a_1 w_1 + \cdots + a_n w_n.$$

Abbiamo quindi provato che

$$v = a_1 v_1 + \cdots + a_n v_n \quad \Longrightarrow \quad f(v) = a_1 w_1 + \cdots + a_n w_n. \tag{5.1}$$

Per dimostrare l'esistenza è sufficiente usare l'Equazione (5.1) per definire f. □

COROLLARIO 5.1.16. *Siano U un sottospazio di uno spazio vettoriale di dimensione finita V e $v \in V$ un vettore tale che $v \notin U$. Allora esiste $f\colon V \to \mathbb{K}$ lineare tale che $f(v) = 1$ e $f(u) = 0$ per ogni $u \in U$.*

DIMOSTRAZIONE. Denotiamo $v_1 = v$, $n = \dim V$, $s = 1 + \dim U$ e sia $v_2, \dots, v_s$ una qualunque base di U. Siccome $v_1 \notin \operatorname{Span}(v_2, \dots, v_s)$ i vettori $v_1, v_2, \dots, v_s$ sono linearmente indipendenti e possono essere completati ad una base $v_1, \dots, v_n$ di V. Basta allora usare il Teorema 5.1.15 e prendere come f l'applicazione lineare che nella base $v_1, \dots, v_n$ vale $f(v_1) = 1$ e $f(v_i) = 0$ per ogni $i > 1$. □

Il seguente risultato, analogo al Teorema 5.1.15, mostra che le applicazioni lineari sono univocamente definite dai valori che assumono in una coppia di sottospazi complementari.

PROPOSIZIONE 5.1.17. *Sia $V = H \oplus K$ uno spazio vettoriale somma diretta di due suoi sottospazi. Per ogni spazio vettoriale W ed ogni coppia di applicazioni lineari $h\colon H \to W$, $k\colon K \to W$ esiste, ed è unica, un'applicazione lineare $f\colon V \to W$ tale che $f_{|H} = h$, $f_{|K} = k$.*

DIMOSTRAZIONE. Ogni vettore v di V si scrive in maniera unica come somma $v = x + y$, com $x \in H$ e $y \in K$. Ne segue che

$$f(v) = f_{|H}(x) + f_{|K}(y)$$

e quindi che f è univocamente determinata dalle sue restrizioni. Date h, k come sopra, l'applicazione

$$f(x + y) = h(x) + k(y), \qquad x \in H,\ y \in K,$$

è ben definita, è lineare e ha come restrizioni h e k. □

Esercizi.

244. Dimostrare che ogni applicazione lineare trasforma vettori linearmente dipendenti in vettori linearmente dipendenti.

245. Sia $f\colon \mathbb{K}^n \to \mathbb{K}^m$ un'applicazione lineare. Dimostrare che l'applicazione

$$g\colon \mathbb{K}^{n+m} \to \mathbb{K}^{n+m}, \qquad g\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ f(x) + y \end{pmatrix}, \qquad x \in \mathbb{K}^n,\ y \in \mathbb{K}^m,$$

è lineare e bigettiva.

246. Verificare che l'applicazione

$$f\colon \mathbb{K}[x] \to \mathbb{K}[x], \qquad f(p(x)) = p(x + 1),$$

ossia $f(\sum_i a_i x^i) = \sum_i a_i (x + 1)^i$, è un isomorfismo lineare.

247. Come nell'Esempio 4.2.5, per ogni insieme S indichiamo con $\mathbb{K}^S$ lo spazio vettoriale di tutte le applicazioni $\alpha\colon S \to \mathbb{K}$. Data un'applicazione di insiemi $f\colon S \to T$ definiamo

$$f^*\colon \mathbb{K}^T \to \mathbb{K}^S, \qquad f^*(\alpha) = \alpha \circ f.$$

Provare che: f^* è lineare; f^* è iniettiva se e solo se f è surgettiva; f^* è surgettiva se e solo se f è iniettiva. Dedurre che esistono applicazioni lineari $\mathbb{K}^{\mathbb{N}} \to \mathbb{K}^{\mathbb{N}}$ che sono iniettive ma non surgettive, e viceversa.

248. Sia (u, v) una base di uno spazio vettoriale bidimensionale V e sia $f\colon V \to V$ lineare tale che $f(u) = v$ e $f(v) = u + av$ per uno scalare a. Mostrare che f è un isomorfismo.

249. Siano $(v_1, \dots, v_n)$ una base dello spazio vettoriale V ed $f\colon V \to V$ lineare tale che $f(v_i) = v_{i+1}$ per ogni $i < n$ e $f(v_n) = a_1 v_1 + \cdots + a_n v_n$. Provare che f è un isomorfismo se e solo se $a_1 \neq 0$.

250. Siano $V \xrightarrow{f} W \xrightarrow{g} Z$ applicazioni tra spazi vettoriali. Provare che se f è lineare surgettiva e gf è lineare, allora anche g è lineare.

251. Siano H, K due sottospazi vettoriali tali che $H \cap K = 0$, e quindi tali che $H + K = H \oplus K$. Mostrare che l'applicazione

$$f\colon H \times K \to H \oplus K, \qquad (h, k) \mapsto h + k,$$

è un isomorfismo di spazi vettoriali.

252. Siano $f\colon V \to W$ e $g\colon W \to V$ due applicazioni lineari. Cosa vuol dire che il diagramma

$$\begin{array}{ccc} V & \xrightarrow{\mathrm{Id}} & V \\ {\scriptstyle g}\uparrow & \nearrow_{f} & \uparrow{\scriptstyle g} \\ W & \xrightarrow[\mathrm{Id}]{} & W \end{array}$$

è commutativo?

253. Siano $H, K \subseteq V$ sottospazi vettoriali e siano $h\colon H \to W$, $k\colon K \to W$ applicazioni lineari tali che $h(v) = k(v)$ per ogni $v \in H \cap K$. Dimostrare che vi è un'unica applicazione lineare $f\colon H + K \to W$ tale che $f_{|H} = h$, $f_{|K} = k$.

254. Descrivere un'applicazione lineare $f\colon \mathbb{K}^2 \to \mathbb{K}^2$ tale che $f^2 = f \circ f = -\operatorname{Id}$.

255 (☕). Sia $f\colon V \to V$ applicazione lineare non nulla. Provare che f è una omotetia se e solo se per ogni $v \in V$ i vettori $v, f(v)$ sono linearmente dipendenti.

256 (☕). Sia V uno spazio vettoriale di dimensione finita su di un campo $\mathbb{K}$. Si assuma che esistano un elemento $a \in \mathbb{K}$ senza radici quadrate, ossia tale che $a \neq b^2$ per ogni $b \in \mathbb{K}$, ed un'applicazione lineare $f\colon V \to V$ tale che $f^2 = f \circ f = a \operatorname{Id}_V$. Dimostrare che V ha dimensione pari e, più precisamente, provare che $\dim V = 2s$, dove s è il massimo intero per cui esistono $v_1, \dots, v_s \in V$ tali che $v_1, \dots, v_s, f(v_1), \dots, f(v_s)$ siano linearmente indipendenti.

257 (☕). Sia $\mathbb{K} = \mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5}, \dots) \subset \mathbb{R}$ l'unione di tutti i sottocampi F_n, $n \geq 0$, descritti nell'Esempio 3.6.6. Provare che $\mathbb{K}$ è un sottocampo di $\mathbb{R}$ e che $\mathbb{K}$ è isomorfo a $\mathbb{K}^2$ come spazio vettoriale su $\mathbb{Q}$.

5.2. Nucleo, iperpiani e sistemi di coordinate

Sia $f\colon V \to W$ un'applicazione lineare; abbiamo già osservato che se f è iniettiva allora $0 \in V$ è l'unico vettore che viene mandato nel vettore nullo di W. Il bello delle applicazioni lineari è che vale anche il viceversa.

Definizione 5.2.1. Il **nucleo** di un'applicazione lineare $f\colon V \to W$ è l'insieme

$$\operatorname{Ker}(f) = \{v \in V \mid f(v) = 0\}.$$ [1]

Lemma 5.2.2. *Il nucleo di un'applicazione lineare $f\colon V \to W$ è un sottospazio vettoriale di V. L'applicazione f è iniettiva se e solo se* $\operatorname{Ker}(f) = 0$.

Dimostrazione. Siccome $f(0) = 0$ si ha $0 \in \operatorname{Ker}(f)$; se $u, v \in \operatorname{Ker}(f)$ allora $f(u + v) = f(u) + f(v) = 0 + 0 = 0$ e quindi $u + v \in \operatorname{Ker}(f)$; se $u \in \operatorname{Ker}(f)$ e $a \in \mathbb{K}$ si ha $f(au) = af(u) = a0 = 0$ e quindi $au \in \operatorname{Ker}(f)$; abbiamo quindi dimostrato che il nucleo è un sottospazio vettoriale.

[1] In inglese *kernel* (si pronuncia kèrnel).

Supponiamo f è inettiva e sia $u \in \mathrm{Ker}(f)$ allora $f(u) = 0 = f(0)$ e quindi $u = 0$. Questo prova che ogni vettore del nucleo è nullo e dunque $\mathrm{Ker}(f) = 0$. Viceversa supponiamo $\mathrm{Ker}(f) = 0$ e siano $u, v \in V$ due vettori tali che $f(u) = f(v)$. Allora $f(u-v) = f(u)-f(v) = 0$, ossia $u - v \in \mathrm{Ker}(f)$ e di conseguenza $u - v = 0$, $u = v$. □

Esempio 5.2.3. Ogni successione finita $v_1, \ldots, v_n$ di vettori in uno spazio vettoriale V definisce un *polivettore riga*

$$\mathbf{v} = (v_1, \ldots, v_n) \in \underbrace{V \times \cdots \times V}_{n \text{ volte}} = V^{\times n}$$

ed una applicazione (che indicheremo con lo stesso simbolo) $\mathbf{v} \colon \mathbb{K}^n \to V$ definita secondo la regola del *prodotto riga per colonna*:

$$\mathbf{v} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = (v_1, \ldots, v_n) \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n.$$

Si verifica immediatamente che $\mathbf{v}$ è lineare e che è surgettiva se e solo se i vettori $v_1, \ldots, v_n$ generano V, mentre il Lemma 5.2.2 dice che $\mathbf{v}$ è iniettiva se e solo se i vettori $v_1, \ldots, v_n$ sono linearmente indipendenti.

In conclusione, $\mathbf{v}$ è un isomorfismo lineare se e solo se $v_1, \ldots, v_n$ è una base. In tal caso $\mathbf{v}^{-1}(u)$ è il vettore colonna che ha come componenti le coordinate di u rispetto a tale base.

È interessante osservare che, nella situazione del Teorema 5.1.15, vale $f = \mathbf{w} \circ \mathbf{v}^{-1}$, dove $\mathbf{v} = (v_1, \ldots, v_n)$ e $\mathbf{w} = (w_1, \ldots, w_n)$.

Esempio 5.2.4. Siano $f \colon V \to W$ lineare e $U \subseteq V$ un sottospazio. Allora il nucleo della restrizione $f_{|U} \colon U \to W$ è uguale a $U \cap \mathrm{Ker}(f)$. In particolare $f_{|U}$ è iniettiva se e solo se $U \cap \mathrm{Ker}(f) = 0$.

Definizione 5.2.5. Sia V uno spazio vettoriale sul campo $\mathbb{K}$. Un sottospazio $H \subseteq V$ si dice un **iperpiano** se esiste un'applicazione lineare *non nulla* $f \colon V \to \mathbb{K}$ tale che $H = \mathrm{Ker}(f)$.

Se $H \subseteq V$ è un iperpiano e se il campo base ha almeno tre elementi, allora l'applicazione $f \colon V \to \mathbb{K}$ tale che $H = \mathrm{Ker}(f)$ non è unica. Infatti $\mathrm{Ker}(f) = \mathrm{Ker}(\lambda f)$ per ogni scalare $\lambda \neq 0$; tuttavia vale il seguente risultato di parziale unicità.

Proposizione 5.2.6. *Siano $f, g \colon V \to \mathbb{K}$ due applicazioni lineari. Allora $\mathrm{Ker}(f) \subseteq \mathrm{Ker}(g)$ se e solo se g è un multiplo scalare di f. In particolare, se H, K sono due iperpiani in V e $H \subseteq K$, allora $H = K$.*

Dimostrazione. Se $g = \lambda f$ con $\lambda \in \mathbb{K}$, per ogni vettore $v \in \mathrm{Ker}(f)$ si ha $g(v) = \lambda f(v) = 0$ e quindi $\mathrm{Ker}(f) \subseteq \mathrm{Ker}(g)$.

Supponiamo viceversa che $\mathrm{Ker}(f) \subseteq \mathrm{Ker}(g)$; se $f = 0$ allora $\mathrm{Ker}(f) = V$, a fortiori $\mathrm{Ker}(g) = V$ e quindi $g = f = 0$. Se f non è nulla fissiamo un vettore $v \in V$ tale che $f(v) \neq 0$. Posto $\lambda = g(v)/f(v)$ dimostriamo che $g = \lambda f$, ossia che per ogni vettore $u \in V$ vale $g(u) = \lambda f(u)$. A tale scopo introduciamo il vettore $w = u - \dfrac{f(u)}{f(v)} v$; allora $f(w) = 0$ e quindi $w \in \mathrm{Ker}(f)$. Per ipotesi $\mathrm{Ker}(f) = \mathrm{Ker}(g)$ e quindi

$$0 = g(w) = g(u) - \frac{f(u)}{f(v)} g(v) = g(u) - \lambda f(u).$$

Se H, K sono due iperpiani in V e $H \subseteq K$, prese due applicazioni lineari non nulle $f, g \colon V \to \mathbb{K}$ tali che $H = \mathrm{Ker}(f)$ e $K = \mathrm{Ker}(g)$, abbiamo appena dimostrato che $g = \lambda f$ per qualche $\lambda \in \mathbb{K}$. Dal fatto che $g \neq 0$ segue $\lambda \neq 0$ e quindi $\mathrm{Ker}(f) = \mathrm{Ker}(g)$. □

Corollario 5.2.7. *Siano $H \subseteq V$ un iperpiano e $v \in V$ un vettore non appartenente ad H. Vi è allora un'unica applicazione lineare $f \colon V \to \mathbb{K}$ tale che $H = \mathrm{Ker}(f)$ e $f(v) = 1$.*

Dimostrazione. Per definizione di iperpiano esiste $g \colon V \to \mathbb{K}$ lineare e non nulla talw che $H = \mathrm{Ker}(g)$. Siccome $v \notin H$ si ha $g(v) \neq 0$. Se poniamo $\lambda = g(v)^{-1}$ e $f = \lambda g$, allora $\mathrm{Ker}(f) = H$ e $f(v) = 1$. L'unicità segue dal fatto che tutte le applicazioni $V \to \mathbb{K}$ con nucleo H sono tutte multipli scalari di f. □

Negli spazi vettoriali di dimensione finita possiamo equivalentemente definire gli iperpiani in termini di dimensione.

TEOREMA 5.2.8. *Sia V uno spazio vettoriale di dimensione finita $n > 0$. Un sottospazio vettoriale $H \subseteq V$ è un iperpiano se e solo se ha dimensione $n-1$.*

DIMOSTRAZIONE. Siano $\mathbb{K}$ il campo base e $H \subseteq V$ un sottospazio vettoriale. Se $\dim H = n-1$ fissiamo una base $v_1, \dots, v_{n-1}$ di H e completiamola ad una base $v_1, \dots, v_n$ di V. Se $f\colon V \to \mathbb{K}$ è l'applicazione lineare tale che $f(v_i) = 0$ per ogni $i < n$ e $f(v_n) = 1$, si vede immediatamente che $\operatorname{Ker}(f) = H$.

Viceversa, supponiamo $H = \operatorname{Ker}(f)$ un iperpiano e sia $s = \dim H$. Siccome $H \neq V$ si ha $s < n$. Come sopra, prendiamo una base $v_1, \dots, v_n$ di V tale che i primi vettori $v_1, \dots, v_s$ siano una base di H. Siccome $v_i \notin H$ per ogni $i > s$ si ha $f(v_i) \neq 0$ per ogni $i > s$. Se fosse $s \leq n-2$ si avrebbe $w = f(v_n)v_{n-1} - f(v_{n-1})v_n \in \operatorname{Ker}(f)$ in contraddizione con il fatto che $\operatorname{Ker}(f)$ è il sottospazio generato da $v_1, \dots, v_s$.

Avendo escluso le ipotesi $s \geq n$ e $s \leq n-2$ non rimane che concludere $s = n-1$. Vedremo più avanti che la medesima conclusione si può ottenere in maniera più rapida usando il teorema del rango. □

DEFINIZIONE 5.2.9. Dato uno spazio vettoriale V di dimensione n, una successione di applicazioni lineari $\varphi_1, \dots, \varphi_n\colon V \to \mathbb{K}$ si dice un **sistema di coordinate** se l'applicazione

$$V \to \mathbb{K}^n, \qquad v \mapsto \begin{pmatrix} \varphi_1(v) \\ \vdots \\ \varphi_n(v) \end{pmatrix}$$

è un isomorfismo di spazi vettoriali.

Per uno spazio vettoriale di dimensione finita *dare una base è la stessa cosa che dare un sistema di coordinate*, nel senso descritto dal seguente teorema.

TEOREMA 5.2.10. *Sia V uno spazio vettoriale di dimensione finita n. Per ogni base $v_1, \dots, v_n$ di V esiste un unico sistema di coordinate $\varphi_1, \dots, \varphi_n$ tale che*

$$\varphi_i(v_j) = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}. \tag{5.2}$$

Viceversa per ogni sistema di coordinate $\varphi_1, \dots, \varphi_n$ esiste un'unica base $v_1, \dots, v_n$ che soddisfa (5.2).

DIMOSTRAZIONE. Se $\mathbf{v} = (v_1, \dots, v_n)$ è una base, ogni vettore di V si scrive in modo unico nella forma $a_1v_1 + \cdots + a_nv_n$, con $a_1, \dots, a_n \in \mathbb{K}$. Basta allora definire $\varphi_i(a_1v_1 + \cdots + a_nv_n) = a_i$ per ogni i. In altri termini, le funzioni φ_i sono le componenti dell'isomorfismo lineare $\mathbf{v}^{-1}\colon V \to \mathbb{K}^n$, vedi Esempio 5.2.3.

Viceversa, se $\varphi_1, \dots, \varphi_n$ è un sistema di coordinate, basta considerare l'immagine della base canonica di $\mathbb{K}^n$ mediante l'inverso dell'isomorfismo lineare

$$V \to \mathbb{K}^n, \qquad v \mapsto \begin{pmatrix} \varphi_1(v) \\ \vdots \\ \varphi_n(v) \end{pmatrix}.$$

□

Esercizi.

258. Siano $f\colon V \to W$ un'applicazione lineare surgettiva e $U \subseteq V$ un sottospazio vettoriale. Dimostrare che la restrizione $f_{|U}\colon U \to W$ è un isomorfismo se e solo se U è un complementare di $\operatorname{Ker}(f)$ in V.

259. Siano $f\colon V \to W$ un'applicazione lineare e $A, B \subseteq V$ due sottospazi tali che $A \cap B = 0$ e $(A+B) \cap \operatorname{Ker}(f) = 0$. Dimostrare che $f(A) \cap f(B) = 0$.

260 (♡). Sia $f\colon V \to W$ lineare tra spazi vettoriali di dimensione finita. Provare che $0 \neq \operatorname{Ker}(f) \neq V$ se e solo se esiste $g\colon W \to V$ lineare tale che $f \circ g = 0$ e $g \circ f \neq 0$.

261. Sia V spazio vettoriale di dimensione finita:
(1) Dimostrare che ogni sottospazio vettoriale di V è intersezione di iperpiani.
(2) Sia $v_1, \dots, v_n$ una base di V e denotiamo $v_0 = v_1 + v_2 + \cdots + v_n$. Sia $f\colon V \to V$ un'applicazione lineare tale che $f(v_i) = \lambda_i v_i$ per ogni $i = 0, \dots, n$ ed opportuni $\lambda_i \in \mathbb{K}$. Dimostrare che $\lambda_0 = \lambda_1 = \cdots = \lambda_n$.
(3) Sia $f\colon V \to V$ lineare tale che $f(L) \subseteq L$ per ogni retta $L \subseteq V$ (retta=sottospazio di dimensione 1). Dimostrare che f è un multiplo scalare dell'identità.
(4) Sia $f\colon V \to V$ lineare tale che $f(H) \subseteq H$ per ogni iperpiano $H \subseteq V$. Dimostrare che f è un multiplo scalare dell'identità.

262. Sia V spazio vettoriale di dimensione n e siano $H_1, \dots, H_n \subseteq V$ iperpiani fissati e tali che $H_1 \cap \cdots \cap H_n = 0$. Dimostrare che esiste una base $v_1, \dots, v_n$ di V tale che $v_i \in H_j$ per ogni $i \neq j$.

5.3. Immagine e teorema del rango

Ogni applicazione lineari è anche un'applicazione di insiemi ed ha quindi senso parlare delle sua immagine. Se $f\colon V \to W$ è un'applicazione lineare, chiameremo

$$f(V) = \{f(v) \mid v \in V\}$$

l'**immagine** di f, talvolta denotata con $\operatorname{Im}(f)$.

PROPOSIZIONE 5.3.1. *Siano $f\colon V \to W$ un'applicazione lineare e $U \subseteq V$ un sottospazio vettoriale. Allora la sua immagine*

$$f(U) = \{f(u) \mid u \in U\}$$

è un sottospazio vettoriale di W.

DIMOSTRAZIONE. Siccome $0 \in U$ si ha $0 = f(0) \in f(U)$. Se $w_1, w_2 \in f(U)$ allora esistono $u_1, u_2 \in U$ tali che $w_1 = f(u_1)$ e $w_2 = f(u_2)$ e dunque

$$w_1 + w_2 = f(u_1) + f(u_2) = f(u_1 + u_2).$$

Siccome $u_1 + u_2 \in U$ ne consegue che $w_1 + w_2 \in f(U)$. Similmente se $w = f(u) \in f(U)$ e $a \in \mathbb{K}$ si ha

$$aw = af(u) = f(au) \in f(U).$$

□

DEFINIZIONE 5.3.2. Il **rango** $\operatorname{rg}(f)$ di un'applicazione lineare $f\colon V \to W$ è la dimensione dell'immagine $f(V)$; in formule $\operatorname{rg}(f) = \dim f(V)$.

Naturalmente ha senso parlare di rango solo quando l'immagine di f è un sottospazio di dimensione finita di W; questo sicuramente accade se almeno uno dei due spazi vettoriali V, W ha dimensione finita.

Siano $(v_1, \dots, v_n)$ una base di V, $w_1, \dots, w_n$ vettori di W e denotiamo $f\colon V \to W$ l'applicazione lineare tale che $f(v_i) = w_i$ per ogni i. Allora $f(V) = \operatorname{Span}(w_1, \dots, w_n)$ e quindi il rango di f è uguale al massimo numero di vettori w_i linearmente indipendenti.

LEMMA 5.3.3. *Valgono le seguenti disuguaglianze:*
(1) *per ogni $f\colon V \to W$ lineare vale $\operatorname{rg}(f) \leq \min(\dim V, \dim W)$;*
(2) *date $U \xrightarrow{g} V \xrightarrow{f} W$ due applicazioni lineari tra spazi vettoriali di dimensione finita. Allora $\operatorname{rg}(fg) \leq \min(\operatorname{rg}(f), \operatorname{rg}(g))$;*
(3) *date $f, g\colon V \to W$ lineari, allora per ogni $a, b \in \mathbb{K}$ si ha $\operatorname{rg}(af + bg) \leq \operatorname{rg}(f) + \operatorname{rg}(g)$.*

DIMOSTRAZIONE. $f(V)$ è un sottospazio di W, quindi $\operatorname{rg}(f) \leq \dim W$, e l'applicazione $f\colon V \to f(V)$ è surgettiva, quindi $\operatorname{rg}(f) \leq \dim V$ per il Corollario 5.1.10.

L'immagine di fg è contenuta nell'immagine di f e quindi $\operatorname{rg}(fg) \leq \operatorname{rg}(f)$. D'altra parte le due applicazioni $fg\colon V \to W$ e $f_{|g(U)}\colon g(U) \to W$ hanno la stessa immagine, e per il punto precedente $\operatorname{rg}(gf) \leq \dim g(U) = \operatorname{rg}(g)$.

Per ogni $v \in V$, il vettore $(af + bg)(v) = af(v) + bg(v) = f(av) + g(bv)$ appartiene al sottospazio vettoriale $f(V) + g(V)$ e quindi

$$\dim(af + bg)(V) \leq \dim(f(V) + g(V)) \leq \dim(f(V)) + \dim(g(V)).$$

□

TEOREMA 5.3.4 (Teorema del rango). *Sia $f\colon V \to W$ un'applicazione lineare. Allora V ha dimensione finita se e solo se $f(V)$ e $\operatorname{Ker}(f)$ hanno entrambi dimensione finita; in tal caso vale la formula*

$$\dim V = \dim \operatorname{Ker}(f) + \operatorname{rg}(f).$$

DIMOSTRAZIONE. Se V ha dimensione finita, per il Corollario 4.5.10 anche $\operatorname{Ker}(f)$ ha dimensione finita; se $v_1, \dots, v_n$ sono generatori di V, per la Proposizione 5.1.8 i vettori $f(v_1), \dots, f(v_n)$ sono generatori dello spazio vettoriale $f(V)$ che pertanto ha dimensione finita.

Viceversa, se $f(V)$ e $\operatorname{Ker}(f)$ hanno entrambi dimensione finita, scegliamo una base $v_1, \dots, v_p$ di $\operatorname{Ker}(f)$ ed una base $w_1, \dots, w_q$ di $f(V)$ e poi scegliamo q vettori $u_1, \dots, u_q \in V$ tali che $f(u_i) = w_i$ per ogni indice i. Per concludere la dimostrazione basta dimostrare che i vettori $v_1, \dots, v_p, u_1, \dots, u_q$ formano una base di V.

Proviamo che $v_1, \dots, v_p, u_1, \dots, u_q$ sono linearmente indipendenti. Siano dati degli scalari $a_1, \dots, a_p, b_1, \dots, b_q \in \mathbb{K}$ tali che

$$a_1 v_1 + \cdots + a_p v_p + b_1 u_1 + \cdots + b_q u_q = 0 \tag{5.3}$$

e mostriamo che vale $a_i = b_j = 0$ per ogni i, j. Dato che $f(v_i) = 0$ per ogni i si ha

$$f(a_1 v_1 + \cdots + a_p v_p + b_1 u_1 + \cdots b_q u_q) = b_1 w_1 + \cdots b_q w_q = 0$$

da cui si deduce $b_1 = \cdots = b_q = 0$ in quanto i vettori w_j sono linearmente indipendenti. Dunque la relazione (5.3) diventa $a_1 v_1 + \cdots + a_p v_p = 0$ da cui segue $a_1 = \cdots = a_p = 0$ in quanto i vettori v_i sono linearmente indipendenti. □

ESEMPIO 5.3.5. Calcoliamo il rango dell'applicazione lineare

$$\mathbb{R}^3 \to \mathbb{R}^3, \qquad (x, y, z)^T \mapsto (x - y, y - z, z - x)^T.$$

Il nucleo è definito dalle equazioni $x - y = y - z = z - x = 0$ che equivalgono a $x = y = z$. In altri termini

$$\operatorname{Ker}(f) = \{(a, a, a)^T \in \mathbb{R}^3 \mid a \in \mathbb{R}\}$$

è un sottospazio di dimensione 1 e per il teorema del rango, l'immagine di f ha dimensione $3 - 1 = 2$.

COROLLARIO 5.3.6. *Siano V uno spazio vettoriale di dimensione finita e $f\colon V \to V$ un'applicazione lineare. Allora f è un isomorfismo se e solo se* $\operatorname{Ker}(f) = 0$.

DIMOSTRAZIONE. Se f è un isomorfismo, allora f è iniettiva e quindi $\operatorname{Ker}(f) = 0$. Viceversa, se $\operatorname{Ker}(f) = 0$ allora f è iniettiva per il Lemma 5.2.2. Per il teorema del rango $\dim f(V) = \dim V$ e quindi $f(V) = V$ per il Lemma 4.6.16, ossia f è anche surgettiva. □

Il Corollario 5.3.6 è falso senza l'ipotesi che lo spazio vettoriale V abbia dimensione finita. Ad esempio, nello spazio vettoriale $\mathbb{K}[x]$ dei polinomi a coefficienti nel campo $\mathbb{K}$, l'applicazione di moltiplicazione per x è lineare iniettiva ma non è surgettiva, cf. Esercizio 247.

Esercizi.

263 (♡). Calcolare il rango dell'applicazione lineare

$$f\colon \mathbb{R}^3 \to \mathbb{R}^2, \qquad (x, y, z)^T \mapsto (x - y, y - 2z)^T.$$

264. Sia $f\colon \mathbb{R}^{350} \to \mathbb{R}^{250}$ un'applicazione lineare e sia $V \subseteq \mathbb{R}^{350}$ un sottospazio vettoriale tale che

$$\dim V = 300, \qquad \dim(V \cap \operatorname{Ker}(f)) = 50.$$

Calcolare le dimensioni di $f(V)$ e di $V + \operatorname{Ker}(f)$. Dire se f è surgettiva.

265. Siano $f\colon V \to W$ lineare tra spazi vettoriali di dimensione finita e $U \subseteq V$ un sottospazio vettoriale. Si assuma che per ogni $v \in V$ esista un unico vettore $u \in U$ tale che $f(u + v) = 0$. Dimostrare che $\operatorname{rg}(f) = \dim V$.

266. Siano $\xi_0, \xi_1, \dots, \xi_{n-1} \in \mathbb{C}$ le radici n-esime di 1. Usare il teorema del rango, anziché il Lemma 4.8.5, per dimostrare che

$$\mathbb{K} = \{z \in \mathbb{C} \text{ tale che } z = a_0\xi_0 + a_1\xi_1 + \cdots + a_{n-1}\xi_{n-1}, \text{ con } a_k \in \mathbb{Q}\}$$

è un campo di numeri.

267 (Proiezioni). Un'applicazione lineare $f\colon V \to V$ da uno spazio vettoriale in sé si dice una **proiezione** se $f = f \circ f$. Provare che $f\colon V \to V$ è una proiezione se e solo se $f(v) = v$ per ogni $v \in f(V)$; provare inoltre che se f è una proiezione allora anche $\mathrm{id}_V - f$ è una proiezione e vale

$$V = \mathrm{Ker}(f) \oplus f(V), \qquad f(V) = \mathrm{Ker}(\mathrm{id}_V - f)\,.$$

268. Sia $f\colon V \to V$ lineare con V di dimensione finita. Provare che $V = \mathrm{Ker}(f) \oplus f(V)$ se e solo se f e $f \circ f$ hanno lo stesso rango.

269 (☕). Sia $f\colon V \to V$ lineare con V di dimensione finita. Provare che esiste $g\colon V \to V$ lineare invertibile tale che

$$f \circ g \circ f \circ g = f \circ g\,.$$

270. Si consideri il seguente quadrato commutativo di applicazioni lineari tra spazi vettoriali di dimensione finita

$$\begin{array}{ccc} A & \xrightarrow{f} & P \\ {\scriptstyle i}\downarrow & & \downarrow{\scriptstyle p} \\ B & \xrightarrow{g} & Q. \end{array}$$

Dimostrare che esiste un'applicazione lineare $h\colon B \to P$ che rende il diagramma

$$\begin{array}{ccc} A & \xrightarrow{f} & P \\ {\scriptstyle i}\downarrow & \overset{h}{\nearrow} & \downarrow{\scriptstyle p} \\ B & \xrightarrow{g} & Q \end{array}$$

commutativo se e solo se $\mathrm{Ker}(i) \subseteq \mathrm{Ker}(f)$ e $g(B) \subseteq p(P)$.

271. Siano $f\colon V \to W$ un'applicazione lineare ed $Z \subseteq W$ un sottospazio vettoriale. Dimostrare che il sottoinsieme $H = \{v \in V \mid f(v) \in Z\}$ è un sottospazio vettoriale di V.

272. Siano $f\colon V \to W$ un'applicazione lineare surgettiva tra spazi vettoriali di dimensione finita e $Z \subseteq W$ un sottospazio vettoriale. Dimostrare che la dimensione del sottospazio vettoriale $H = \{v \in V \mid f(v) \in Z\}$ è uguale a $\dim Z + \dim V - \dim W$.

273. Date due applicazioni lineari $U \xrightarrow{f} V \xrightarrow{g} W$ tra spazi vettoriali di dimensione finita, usare il risultato dell'Esercizio 272 per dimostrare che vale la disuguaglianza $\mathrm{rg}(gf) \geq \mathrm{rg}(f) + \mathrm{rg}(g) - \dim V$.

274 (Prospettive lineari). Sia V spazio vettoriale di dimensione finita e $f\colon V \to V$ un'applicazione lineare invertibile. Provare che le seguenti condizioni sono equivalenti:

(1) $\mathrm{rg}(f - I) \leq 1$;
(2) esiste un iperpiano $H \subseteq V$ tale che $f(v) = v$ per ogni $v \in H$;
(3) Esiste $w \in V$ tale che $f(v) - v \in \mathrm{Span}(w)$ per ogni $v \in V$.

Un'applicazione lineare invertibile f che soddisfa le precedenti condizioni viene detta prospettiva lineare.

275 (Omologie ed elazioni). Siano V uno spazio vettoriale di dimensione finita $n \geq 3$ e $f\colon V \to V$ una prospettiva lineare diversa dall'identità (Esercizio 274). Provare che esiste un unico sottospazio $L \subseteq V$ di dimensione 1 e tale che $f(v) \in L + \mathrm{Span}(v)$ per ogni $v \in V$.

Nota: la parte non banale dell'esercizio è dimostrare l'unicità di L; anticamente, una prospettiva come la suddetta veniva chiamata elazione se $L \subseteq \mathrm{Ker}(f - I)$ ed omologia se $L \not\subseteq \mathrm{Ker}(f - I)$.

5.4. Matrici ed applicazioni lineari

I vettori numerici introdotti nella Sezione 4.1 sono utilmente generalizzati ad una classe di oggetti più ampia, quella delle matrici. Una tabellina rettangolare

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}$$

è detta una **matrice** a n righe ed m colonne, o più brevemente una matrice $n \times m$. Gli elementi a_{ij} sono detti i **coefficienti** della matrice A ed è una consolidata convenzione che *il primo indice di ogni coefficiente rappresenta la posizione di riga ed il secondo indice la posizione di colonna*. Possiamo abbreviare la notazione scrivendo $A = (a_{ij})$, $i = 1, \dots, n$, $j = 1, \dots, m$. Talvolta, per evitare possibili ambiguità, scriveremo $a_{i,j}$ inserendo una virgola per separare l'indice di riga da quello di colonna.

Esempio 5.4.1. Un esempio di matrice 2×3 è

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix},$$

le cui righe sono $(1, 2, 3)$, $(4, 5, 6)$ e le cui colonne sono

$$\begin{pmatrix} 1 \\ 4 \end{pmatrix}, \quad \begin{pmatrix} 2 \\ 5 \end{pmatrix}, \quad \begin{pmatrix} 3 \\ 6 \end{pmatrix}.$$

Definizione 5.4.2. Sia $\mathbb{K}$ un campo: l'insieme di tutte le matrici $n \times m$ a coefficienti in $\mathbb{K}$ si denota $M_{n,m}(\mathbb{K})$.

In linea di principio possiamo considerare tabelle di enti algebrici più generali degli scalari; ad esempio ha senso considerare matrici i cui coefficienti sono vettori, applicazioni lineari, polinomi, matrici eccetera.

Dato che esiste una ovvia bigezione tra l'insieme $M_{n,m}(\mathbb{K})$ e lo spazio $\mathbb{K}^{nm}$ (basta mettere tutti coefficienti in una sola colonna) non è sorprendente scoprire che $M_{n,m}(\mathbb{K})$ è uno spazio vettoriale su $\mathbb{K}$ con le operazioni di somma e prodotto per scalare eseguite coefficiente per coefficiente:

$$\begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix} + \begin{pmatrix} b_{11} & \dots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{n1} & \dots & b_{nm} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & \dots & a_{1m} + b_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} + b_{n1} & \dots & a_{nm} + b_{nm} \end{pmatrix},$$

$$\lambda \begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix} = \begin{pmatrix} \lambda a_{11} & \dots & \lambda a_{1m} \\ \vdots & \ddots & \vdots \\ \lambda a_{n1} & \dots & \lambda a_{nm} \end{pmatrix}.$$

Ad esempio:

$$\begin{pmatrix} 2 & 9 \\ -1 & 5 \\ 4 & 7 \end{pmatrix} + \begin{pmatrix} 1 & -3 \\ 1 & 0 \\ 2 & 6 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 0 & 5 \\ 6 & 13 \end{pmatrix}, \qquad 3 \begin{pmatrix} 2 & 9 \\ -1 & 5 \\ 4 & 7 \end{pmatrix} = \begin{pmatrix} 6 & 27 \\ -3 & 15 \\ 12 & 21 \end{pmatrix}.$$

Il ruolo del vettore nullo nello spazio $M_{n,m}(\mathbb{K})$ è interpretato dalla **matrice nulla**, ossia dalla matrice che ha tutti i coefficienti uguali a 0. Per ovvi motivi (vedi Esempio 5.1.7) la bigezione appena descritta tra $M_{n,m}(\mathbb{K})$ e $\mathbb{K}^{nm}$ è un isomorfismo lineare, in particolare $M_{n,m}(\mathbb{K})$ ha dimensione nm come spazio vettoriale su $\mathbb{K}$.

Ogni matrice $n \times m$ può essere pensata sia come una successione (orizzontale) di m vettori colonna, sia come una successione (verticale) di n vettori riga. Viceversa ogni vettore colonna può essere pensato come una matrice con una sola colonna ed ogni vettore riga può essere pensato come una matrice con una sola riga, ossia

$$\mathbb{K}^n = M_{n,1}(\mathbb{K}), \qquad \mathbb{K}^{(m)} = M_{1,m}(\mathbb{K}).$$

Abbiamo detto che ogni matrice $A \in M_{n,m}(\mathbb{K})$ può essere pensata come una successione di m vettori colonna; possiamo quindi scrivere

$$A = (A^1, \dots, A^m), \qquad \text{con} \quad A^1, \dots, A^m \in \mathbb{K}^n.$$

Dato quindi un qualunque vettore

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \in \mathbb{K}^m$$

possiamo definire, come nell'Esempio 5.2.3, il prodotto riga per colonna

$$Ax = (A^1, \dots, A^m) \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = x_1 A^1 + \cdots + x_m A^m \in \mathbb{K}^n. \tag{5.4}$$

In funzione dei coefficienti della matrice tale prodotto diventa

$$Ax = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1m}x_m \\ \vdots \\ a_{n1}x_1 + \cdots + a_{nm}x_m \end{pmatrix} \tag{5.5}$$

e quindi lo i-esimo coefficiente di Ax è uguale al prodotto riga per colonna della i-esima riga di A con il vettore colonna x. Nel caso particolare in cui $x = e_i$ è lo i-esimo vettore della base canonica di $\mathbb{K}^m$, ossia $x_i = 1$ e $x_j = 0$ per ogni $j \neq i$, allora il prodotto Ae_i coincide con la i-esima colonna di A.

Esempio 5.4.3. Ecco 3 esempi numerici di prodotti riga per colonna:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 6 \end{pmatrix} = \begin{pmatrix} 13 \\ 27 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \\ 9 \end{pmatrix}.$$

Definizione 5.4.4. Sia $A \in M_{n,m}(\mathbb{K})$. L'applicazione lineare associata alla matrice A è definita come

$$L_A \colon \mathbb{K}^m \to \mathbb{K}^n, \qquad L_A(x) = Ax.$$

Da come è definito il prodotto righe per colonna segue immediatamente che L_A è lineare. Si faccia attenzione al fatto che nel passaggio da A ad L_A, e viceversa, le posizioni di n ed m si scambiano.

Proposizione 5.4.5. *Sia $A = (a_{ij}) \in M_{n,m}(\mathbb{K})$. Allora:*

1. *l'immagine di $L_A \colon \mathbb{K}^m \to \mathbb{K}^n$ è il sottospazio vettoriale generato dalle colonne di A;*
2. *il nucleo di L_A è l'insieme delle soluzioni del sistema lineare omogeneo*

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m = 0 \\ \qquad \cdots \qquad \cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m = 0. \end{cases}$$

Dimostrazione. L'enunciato sull'immagine di L_A segue immediatamente dalla Formula (5.4), mentre l'enunciato sul nucleo segue dalla Formula (5.5). □

Teorema 5.4.6. *Per ogni applicazione lineare $f \colon \mathbb{K}^m \to \mathbb{K}^n$ vi è un'unica matrice $A \in M_{n,m}(\mathbb{K})$ tale che $f = L_A$. In particolare $L_A = L_B$ se e solo se $A = B$.*

Inoltre, vale la formula $A = (f(e_1), \dots, f(e_m))$, dove $e_1, \dots, e_m \in \mathbb{K}^m$ è la base canonica di $\mathbb{K}^m$.

Dimostrazione. Abbiamo già notato che per ogni matrice $A \in M_{n,m}(\mathbb{K})$ i vettori Ae_i coincidono con le colonne di A, quindi $A = (L_A(e_1), \dots, L_A(e_n))$ e questo prova che la matrice A è univocamente determinata dall'applicazione L_A.

Sia $f \colon \mathbb{K}^m \to \mathbb{K}^n$ lineare e consideriamo la matrice che ha come colonne i valori di f nella base canonica di $\mathbb{K}^m$, ossia

$$A = (A^1, \dots, A^m), \qquad \text{dove} \quad A^i = f(e_i).$$

Dato un qualunque vettore

$$\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = x_1 e_1 + \cdots + x_m e_m \in \mathbb{K}^m$$

si ha

$$\begin{aligned} L_A(x) = Ax = x_1 A^1 + \cdots + x_m A^m = x_1 f(e_1) + \cdots + x_m f(e_m) \\ = f(x_1 e_1 + \cdots + x_m e_m) = f(x). \end{aligned}$$

□

Giova osservare che la Formula (5.5) e la dimostrazione del Teorema 5.4.6 danno due utili ricette per il calcolo della matrice associata ad un'applicazione lineare tra spazi vettoriali numerici.

ESEMPIO 5.4.7. L'identità $\mathrm{Id}\colon \mathbb{K}^n \to \mathbb{K}^n$ è lineare e la matrice corrispondente ha come coefficienti $a_{ii} = 1$ per ogni i e $a_{ij} = 0$ per ogni $i \neq j$.

ESEMPIO 5.4.8. Consideriamo l'applicazione lineare $f\colon \mathbb{K}^3 \to \mathbb{K}^3$ definita in coordinate da

$$f\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x + 2y + 3z \\ 4x + 5y + 6z \\ 7x + 8y + 9z \end{pmatrix}$$

e calcoliamo la matrice $A \in M_{3,3}(\mathbb{K})$ tale che $f = L_A$. Applicando la Formula (5.5) si ottiene immediatamente

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Equivalentemente possiamo usare la ricetta esposta nella dimostrazione del Teorema 5.4.6, e cioè interpretare le colonne di A come le immagini dei vettori della base canonica:

$$\begin{aligned} A = (f(e_1), f(e_2), f(e_3)) &= \left(f\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, f\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, f\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right) \\ &= \left(\begin{pmatrix} 1 \\ 4 \\ 7 \end{pmatrix}, \begin{pmatrix} 2 \\ 5 \\ 8 \end{pmatrix}, \begin{pmatrix} 3 \\ 6 \\ 9 \end{pmatrix} \right) = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}. \end{aligned}$$

ESEMPIO 5.4.9. Dato un vettore riga $w = (w_1, \ldots, w_n) \in \mathbb{K}^{(n)}$ definiamo

$$\begin{aligned} w^\perp &= \{x \in \mathbb{K}^n \mid w \cdot x = 0 \text{ (prodotto riga per colonna)}\} \\ &= \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \,\middle|\, w_1 x_1 + \cdots + w_n x_n = 0 \right\} \subseteq \mathbb{K}^n. \end{aligned}$$

Se interpretiamo w come una matrice $1 \times n$, allora $w^\perp = \mathrm{Ker}(L_w)$ e dunque, se $w \neq 0$ allora $w^\perp$ è un iperpiano di $\mathbb{K}^n$. Viceversa, poiché ogni applicazione lineare $\mathbb{K}^n \to \mathbb{K}$ è del tipo L_w per qualche $w \in M_{1,n}(\mathbb{K})$, si ha che ogni iperpiano di $\mathbb{K}^n$ è uguale a $w^\perp$ per qualche $w \neq 0$.

Esercizi.

276. Eseguire le seguenti operazioni fra matrici:

$$2\begin{pmatrix} 0 & 4 \\ 5 & -1 \\ 9 & 2 \end{pmatrix} - 3\begin{pmatrix} 0 & 3 \\ -4 & 4 \\ -3 & 3 \end{pmatrix}; \qquad \begin{pmatrix} 1 & 5 & 7 \\ 6 & 1 & 2 \end{pmatrix} + 5\begin{pmatrix} -1 & 3 & 0 \\ -2 & 4 & 3 \end{pmatrix}.$$

277. Trovare due applicazioni lineari $f, g\colon \mathbb{K}^2 \to \mathbb{K}^2$ tali che $f \circ g = 0$ e $g \circ f \neq 0$.

278. Si consideri l'applicazione lineare $f\colon \mathbb{R}^4 \to \mathbb{R}^3$ definita in coordinate da $f(x, y, z, w) = (x - y, y - z, z - w)$. Descrivere la matrice A tale che $f = L_A$.

279. Data la matrice

$$A = \begin{pmatrix} 1 & -1 & 3 \\ 2 & 0 & -1 \\ -1 & -1 & 4 \end{pmatrix} \in M_{3,3}(\mathbb{R}),$$

si dimostri che

$$V = \left\{ (x,y,z) \in \mathbb{R}^3 \;\middle|\; A \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0 \right\}$$

è un sottospazio vettoriale di $\mathbb{R}^3$, se ne calcoli la dimensione e se ne trovi una base.

280. Usare i numeri di Bernoulli per determinare il nucleo dell'applicazione lineare $L_A\colon \mathbb{R}^{n+1} \to \mathbb{R}^n$ associata alla matrice A di coefficienti $a_{ij} = \binom{i+1}{j-1}$, $i = 1,\dots,n$, $j = 1,\dots,n+1$.

5.5. Spazi di applicazioni lineari

Dati due spazi vettoriali V, W sullo stesso campo $\mathbb{K}$ indichiamo con $\operatorname{Hom}(V,W)$ l'insieme di tutte le applicazioni lineari $f\colon V \to W$; talvolta conviene anche tenere traccia del campo e scrivere $\operatorname{Hom}_{\mathbb{K}}(V,W)$.

L'insieme $\operatorname{Hom}(V,W)$ possiede una struttura naturale di spazio vettoriale dove la somma ed il prodotto per scalare sono definiti dalle regole:

$$f+g\colon V \to W, \qquad (f+g)(v) = f(v)+g(v), \qquad f,g \in \operatorname{Hom}(V,W),\ v \in V;$$

$$\lambda f\colon V \to W, \qquad (\lambda f)(v) = \lambda(f(v)), \qquad f \in \operatorname{Hom}(V,W),\ \lambda \in \mathbb{K},\ v \in V.$$

Segue immediatamente dalle definizioni e dal Teorema 5.4.6 che per ogni $n, m \geq 0$ l'applicazione

$$L\colon M_{n,m}(\mathbb{K}) \to \operatorname{Hom}_{\mathbb{K}}(\mathbb{K}^m, \mathbb{K}^n), \quad A \mapsto L_A,$$

è un isomorfismo di spazi vettoriali. In particolare, $\dim \operatorname{Hom}_{\mathbb{K}}(\mathbb{K}^m, \mathbb{K}^n) = nm$.

Abbiamo visto nell'Esempio 5.2.3 che la scelta di una base $\mathbf{v} = (v_1,\dots,v_m)$ di V e di una base $\mathbf{w} = (w_1,\dots,w_n)$ di W definisce due isomorfismi di spazi vettoriali $\mathbf{v}\colon \mathbb{K}^m \to V$ e $\mathbf{w}\colon \mathbb{K}^n \to W$ tramite le formule:

$$\mathbf{v}\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = x_1v_1 + x_2v_2 + \cdots + x_mv_m, \qquad \mathbf{w}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = y_1w_1 + y_2w_2 + \cdots + y_nw_n.$$

Passando agli spazi di applicazioni, la scelta delle basi determina due isomorfismi di spazi vettoriali

$$\Phi\colon \operatorname{Hom}_{\mathbb{K}}(\mathbb{K}^m, \mathbb{K}^n) \to \operatorname{Hom}_{\mathbb{K}}(V,W), \qquad \Phi(f) = \mathbf{w} \circ f \circ \mathbf{v}^{-1},$$

$$\Psi\colon \operatorname{Hom}_{\mathbb{K}}(V,W) \to \operatorname{Hom}_{\mathbb{K}}(\mathbb{K}^m, \mathbb{K}^n), \qquad \Psi(g) = \mathbf{w}^{-1} \circ g \circ \mathbf{v},$$

che sono uno l'inverso dell'altro. Componendo i due isomorfismi precedenti Φ ed L otteniamo il seguente risultato.

TEOREMA 5.5.1. *Siano V e W spazi vettoriali di dimensione finita sul campo $\mathbb{K}$, allora per ogni scelta di una base $\mathbf{v} = (v_1,\dots,v_m)$ di V ed ogni scelta di una base $\mathbf{w} = (w_1,\dots,w_n)$ di W è definito un isomorfismo di spazi vettoriali*

$$L^{\mathbf{v}}_{\mathbf{w}}\colon M_{n,m}(\mathbb{K}) \to \operatorname{Hom}_{\mathbb{K}}(V,W), \qquad L^{\mathbf{v}}_{\mathbf{w}}(A) = \mathbf{w} \circ L_A \circ \mathbf{v}^{-1}.$$

In particolare, $\dim \operatorname{Hom}(V,W) = \dim V \cdot \dim W$.

Vediamo adesso in maniera più esplicita e concreta cosa significa, per una matrice $A \in M_{n,m}$ ed un'applicazione lineare $f\colon V \to W$, essere legate dalla relazione $L^{\mathbf{v}}_{\mathbf{w}}(A) = f$. Per definizione, vale $L^{\mathbf{v}}_{\mathbf{w}}(A) = f$ se e solo se $f = \mathbf{w} \circ L_A \circ \mathbf{v}^{-1}$, e questo vale se e solo se le due applicazioni lineari

$$f\mathbf{v}, \mathbf{w}L_A\colon \mathbb{K}^m \to W$$

coincidono.

Indichiamo come al solito con $e_1, \dots, e_m$ la base canonica di $\mathbb{K}^m$ e con $A^1, \dots, A^m$ le colonne di A. Siccome $A^i = L_A(e_i)$ per ogni i si ha

$$f\mathbf{v}(e_i) = f(v_i), \qquad \mathbf{w}L_A(e_i) = \mathbf{w}(A^i).$$

In definitiva, i coefficienti della i-esima colonna di A sono le coordinate del vettore $f(v_i)$ nella base $w_1, \dots, w_n$; vale $L^{\mathbf{v}}_{\mathbf{w}}(A) = f$ se e solo se

$$(f(v_1), \dots, f(v_m)) = (w_1, \dots, w_n)A, \tag{5.6}$$

dove il prodotto a destra è il solito righe per colonne.

Definizione 5.5.2. Se $f \in \mathrm{Hom}_{\mathbb{K}}(V, W)$ e $A \in M_{n,m}(\mathbb{K})$ sono legate dalla relazione (5.6), ossia se $L^{\mathbf{v}}_{\mathbf{w}}(A) = f$, diremo che A è la **matrice associata** all'applicazione lineare f rispetto alle basi $\mathbf{v}$ e $\mathbf{w}$.

Osservazione 5.5.3. Assumendo V, W entrambi non nulli, è possibile dimostrare che $\mathrm{Hom}(V, W)$ ha dimensione finita se e solo se V e W hanno dimensione finita; è facile vedere che se V ha dimensione finita $n > 0$, allora $\mathrm{Hom}(V, W)$ ha dimensione finita se e solo se W ha dimensione finita (Esercizio 291); è invece molto più difficile dimostrare che se V ha dimensione infinita, allora anche $\mathrm{Hom}(V, W)$ ha dimensione infinita per ogni $W \neq 0$, cf. Esercizio 672.

Siano V, W spazi vettoriali di dimensione finita, abbiamo già osservato che la funzione rango

$$\mathrm{rg}\colon \mathrm{Hom}(V, W) \to \mathbb{Z}$$

è subadditiva, ossia per ogni $f, g \in \mathrm{Hom}(V, W)$ vale la disuguaglianza

$$\mathrm{rg}(f + g) \leq \mathrm{rg}(f) + \mathrm{rg}(g).$$

Per induzione su n si ha che per ogni successione finita di applicazioni lineari $f_1, \dots, f_n \colon V \to W$ tra spazi vettoriali di dimensione finita vale la disuguaglianza

$$\mathrm{rg}(f_1 + \cdots + f_n) \leq \mathrm{rg}(f_1) + \cdots + \mathrm{rg}(f_n).$$

In particolare, se $f_1, \dots, f_r \colon V \to W$ hanno rango 1, allora la somma $f_1 + \cdots + f_r$ ha rango $\leq r$.

Proposizione 5.5.4. *Sia $f\colon V \to W$ un'applicazione lineare tra spazi vettoriali di dimensione finita. Allora il rango di f è uguale al più piccolo intero r per cui è possibile scrivere f come somma di r applicazioni lineari di rango 1.*

Dimostrazione. Se $r = \mathrm{rg}(f)$ abbiamo già osservato precedentemente che non è possibile scrivere f come somma di s applicazioni di rango 1, con $s < r$. Per concludere la dimostrazione basta dimostrare che esistono $f_1, \dots, f_r \colon V \to W$ di rango 1 tali che $f = f_1 + \cdots + f_r$.

Sia $w_1 \dots, w_r$ una base del sottospazio $f(V)$ e sia $\varphi_1, \dots, \varphi_r \in \mathrm{Hom}(f(V), \mathbb{K})$ il corrispondente sistema di coordinate, ossia $w = \sum_i \varphi_i(w) w_i$ per ogni $w \in f(V)$. Allora le applicazioni

$$f_i \colon V \to W, \qquad f_i(v) = \varphi_i(f(v))w_i, \qquad i = 1, \dots, r,$$

hanno rango 1 e per costruzione vale $f = f_1 + \cdots + f_r$. □

Le matrici di cambio di base. Siano $\mathbf{v} = (v_1, \dots, v_n)$ e $\mathbf{w} = (w_1, \dots, w_n)$ due basi di un medesimo spazio vettoriale V. Per ogni indice j esistono, e sono unici, dei coefficienti $a_{ij} \in \mathbb{K}$, $i = 1, \dots, n$, tali che

$$w_j = v_1 a_{1j} + \cdots + v_n a_{nj}, \qquad j = 1, \dots, n, \tag{5.7}$$

e la matrice $A = (a_{ij}) \in M_{n,n}(\mathbb{K})$ viene detta **matrice del cambio di base**. Usando la regola del prodotto righe per colonne possiamo riscrivere le relazioni (5.7) nella forma

$$(w_1, \dots, w_n) = (v_1, \dots, v_n)A \iff \mathbf{w} = \mathbf{v}A.$$

Possiamo interpretare l'applicazione lineare associata $L_A \colon \mathbb{K}^n \to \mathbb{K}^n$ nel modo seguente: abbiamo già osservato che le due applicazioni lineari $\mathbf{v}, \mathbf{w} \colon \mathbb{K}^n \to V$

$$\mathbf{v}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1 v_1 + x_2 v_2 + \cdots + x_n v_n, \qquad \mathbf{w}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = y_1 w_1 + y_2 w_2 + \cdots + y_n w_n,$$

sono isomorfismi lineari e quindi $\mathbf{v}^{-1}\mathbf{w}\colon \mathbb{K}^n \to \mathbb{K}^n$ è ancora un isomorfismo lineare. Ebbene, le relazioni (5.7) sono del tutto equivalenti a dire che $\mathbf{v}^{-1}\mathbf{w} = L_A$. Infatti per ogni $y_1, \dots, y_n \in \mathbb{K}$ si ha

$$\mathbf{w}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \sum_{j=1}^{n} y_j w_j = \sum_{i,j=1}^{n} y_j v_i a_{ij} = \sum_{i=1}^{n} v_i \sum_{j=1}^{n} a_{ij} y_j = \mathbf{v} L_A \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

e quindi $\mathbf{w} = \mathbf{v}L_A$. Componendo a sinistra per $\mathbf{v}^{-1}$ si ottiene la relazione $\mathbf{v}^{-1}\mathbf{w} = L_A$.

Occorre fare attenzione che la matrice di cambio di base agisce in *maniera inversa* sui corrispondenti sistemi di coordinate. Più precisamente, siano $\varphi_1, \dots, \varphi_n$ il sistema di coordinate associato alla base $v_1, \dots, v_n$ e $\psi_1, \dots, \psi_n$ il sistema di coordinate associato alla base $w_1, \dots, w_n$, allora vale la formula

$$\begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_n \end{pmatrix} = A \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_n \end{pmatrix} \iff \varphi_i = \sum_{j=1}^{n} a_{ij}\psi_j.$$

Infatti, tenendo presente per ogni vettore $u \in V$ vale

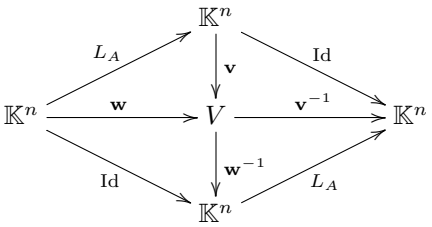
$$u = \sum_j w_j \psi_j(u) = \sum_{i,j} v_i a_{ij} \psi_j(u) = \sum_i v_i (\sum_j a_{ij}\psi_j)(u)$$

e questo è del tutto equivalente a dire che $\varphi_i = \sum_j a_{ij}\psi_j$ per ogni indice i.

Possiamo visualizzare mediante un diagramma commutativo le precedenti relazioni ricordando che, per definizione di sistemi di coordinate si ha

$$\mathbf{v}^{-1}(u) = \begin{pmatrix} \varphi_1(u) \\ \vdots \\ \varphi_n(u) \end{pmatrix} \in \mathbb{K}^n, \quad \mathbf{w}^{-1}(u) = \begin{pmatrix} \psi_1(u) \\ \vdots \\ \psi_n(u) \end{pmatrix} \in \mathbb{K}^n,$$

da cui segue il diagramma commutativo



Esercizi.

281. Determinare la matrice che rappresenta l'applicazione lineare

$$f\colon \mathbb{K}^3 \to \mathbb{K}^2, \qquad f\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x + 2y \\ x + y - z \end{pmatrix},$$

rispetto alle basi canoniche.

282. Determinare la matrice che rappresenta l'applicazione lineare $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ definita da

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

rispetto alla base

$$\mathbf{w} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right\}.$$

Nota: si considera la stessa base in partenza ed in arrivo, ossia si cerca la matrice A tale che $f(\mathbf{w}) = \mathbf{w}A$.

283. Sia $\mathbb{R}[x]_{\leq n}$ lo spazio vettoriale dei polinomi a coefficienti reali di grado minore o uguale a n.

(1) Mostrare che i tre vettori $e_1 = x + 1$, $e_2 = x + 2$, $e_3 = x^2 + x + 1$ formano una base di $\mathbb{R}[x]_{\leq 2}$.
(2) Mostrare che i due vettori $f_1 = x + 3$, $f_2 = x + 4$ formano una base di $\mathbb{R}[x]_{\leq 1}$.
(3) Scrivere la matrice che rappresenta l'applicazione lineare

$$\begin{aligned} \varphi\colon \mathbb{R}[x]_{\leq 2} &\to \mathbb{R}[x]_{\leq 1} \\ p(x) &\mapsto p(x+1) - p(x-1) \end{aligned}$$

rispetto alle basi $\{e_1, e_2, e_3\}$ e $\{f_1, f_2\}$.
(4) Determinare la dimensione del nucleo e dell'immagine di φ.

284 (☕, ♡). Siano $f, g\colon V \to W$ applicazioni lineari non nulle tra spazi vettoriali di dimensione finita. Provare che f, g sono linearmente dipendenti in $\mathrm{Hom}(V, W)$ se e solo se:

(1) $\mathrm{Ker}(f) = \mathrm{Ker}(g)$;
(2) per ogni $v \in V$ i vettori $f(v), g(v)$ sono linearmente dipendenti in W.

285. Siano V, W spazi vettoriali di dimensione finita e siano $A \subseteq V$ e $B \subseteq W$ due sottospazi. Provare che

$$H = \{f \in \mathrm{Hom}(V, W) \mid f(A) \subseteq B\}$$

è un sottospazio vettoriale di dimensione uguale a $(\dim V - \dim A)\dim W + \dim A \dim B$. (Suggerimento: scegliere basi di A e B ed estenderle a basi di V e W. Come sono fatte le matrici che rappresentano gli elementi di H in tali basi?).

286. Determinare tutte la applicazioni lineari $f\colon \mathbb{K} \to \mathbb{K}$ che rendono il diagramma

$$\begin{array}{ccc} \mathbb{K} & \xrightarrow{f} & \mathbb{K} \\ \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle f} \\ \mathbb{K} & \xleftarrow[f]{} & \mathbb{K} \end{array}$$

commutativo.

287. Sia $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ l'applicazione lineare definita in coordinate dalla formula

$$f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x - y \\ 4x - 4y \end{pmatrix}$$

e si consideri l'applicazione lineare

$$\Phi\colon \mathrm{Hom}(\mathbb{R}^2, \mathbb{R}^2) \to \mathrm{Hom}(\mathbb{R}^2, \mathbb{R}^2), \qquad \Phi(g) = f \circ g \quad \text{(composizione di } g \text{ e } f\text{)}.$$

Dimostrare che Φ è lineare, determinare una base di $\mathrm{Ker}(\Phi)$ e completarla ad una base di $\mathrm{Hom}(\mathbb{R}^2, \mathbb{R}^2)$.

288 (☕). Siano V uno spazio vettoriale ed H, K due suoi sottospazi complementari, ossia tali che $V = H \oplus K$. Mostrare che per ogni spazio vettoriale W si ha

$$\mathrm{Hom}(W, V) = \mathrm{Hom}(W, H) \oplus \mathrm{Hom}(W, K),$$

ed esiste un isomorfismo canonico

$$\mathrm{Hom}(V, W) \xrightarrow{\simeq} \mathrm{Hom}(H, W) \times \mathrm{Hom}(K, W).$$

289. Siano V, W spazi vettoriali e $F \subseteq \mathrm{Hom}(V, W)$ il sottoinsieme delle applicazioni lineari di rango finito. Dimostrare che F è un sottospazio vettoriale di $\mathrm{Hom}(V, W)$. (Nota: vedremo più avanti che se $V, W \neq 0$, allora $F \neq 0$ anche quando V ha dimensione infinita, cf. Corollario 12.5.6).

290 (☕). Siano V uno spazio vettoriale di dimensione finita e $F \subseteq \mathrm{Hom}(V, V)$ un sottospazio vettoriale tale che

$$\alpha \circ f \circ \beta \in F \quad \text{per ogni} \quad f \in F, \quad \alpha, \beta \in \mathrm{Hom}(V, V).$$

Dimostrare che $F = 0$ oppure $F = \mathrm{Hom}(V, V)$. Si può dire lo stesso se V ha dimensione infinita?

291. Siano V uno spazio vettoriale di dimensione finita $n > 0$ e W uno spazio vettoriale di dimensione infinita. Dimostrare che $\mathrm{Hom}(V, W)$ ha dimensione infinita.

5.6. Complementi: successioni esatte e caccia al diagramma

La caccia al diagramma, dove il termine caccia non è inteso in senso venatorio ma allo stesso modo di caccia al tesoro, è un utile metodo di dimostrazione usato specialmente in alcune branche dell'algebra. Dato un diagramma commutativo, la caccia al diagramma sfrutta in maniera formale alcune proprietà del diagramma stesso come l'iniettività o la surgettività di alcune applicazioni o come l'esattezza di alcune successioni.

Sappiamo già cosa sono le applicazioni iniettive e surgettive; introduciamo adesso il concetto di successione esatta: fra i tanti possibili diagrammi di spazi vettoriali ed applicazioni lineari, particolarmente importanti sono quelli a forma di stringa, ossia i diagrammi con le applicazioni disposte in serie:

$$\cdots \to V_n \xrightarrow{f_n} V_{n+1} \xrightarrow{f_{n+1}} V_{n+2} \to \cdots \tag{5.8}$$

Definizione 5.6.1. Una diagramma di applicazioni lineari disposte in serie come in (5.8) si dice un **complesso** di spazi vettoriali se $f_{n+1}f_n = 0$ per ogni n, ossia se la composizione di due applicazioni lineari contigue è sempre nulla.

Equivalentemente il diagramma (5.8) è un complesso se per ogni n l'immagine di f_n è contenuta nel nucleo di f_{n+1}. Un complesso si dice finito o limitato se contiene solo un numero finito di spazi vettoriali ed applicazioni lineari; tuttavia è utile, in vista di future applicazioni, considerare anche complessi infiniti o illimitati, nei quali gli indici n che compaiono nel diagramma sono tutti gli interi contenuti in un intervallo della retta reale.

Definizione 5.6.2. Una diagramma di applicazioni lineari disposte in serie come in (5.8) si dice una **successione esatta** di spazi vettoriali se per ogni n il nucleo di f_{n+1} è uguale all'immagine di f_n.

In particolare ogni successione esatta è anche un complesso, mentre il viceversa è generalmente falso: ad esempio, il diagramma

$$0 \to V \to 0$$

è un complesso qualunque sia lo spazio vettoriale V, mentre è una successione esatta se e solo se $V = 0$.

Esempio 5.6.3. Supponiamo che

$$V_0 \xrightarrow{f_0} V_1 \xrightarrow{f_1} V_2 \xrightarrow{f_2} V_3$$

sia una successione esatta. Allora le seguenti condizioni sono equivalenti:

(1) f_0 è surgettiva;
(2) $f_1 = 0$;
(3) f_2 è iniettiva.

Infatti, per l'esattezza in V_1 il nucleo di f_1 è uguale all'immagine di f_0; in particolare f_0 è surgettiva se e solo se $\operatorname{Ker} f_1 = V_1$, ossia se e solo se $f_1 = 0$. Similmente, per l'esattezza in V_2 il nucleo di f_2 è uguale all'immagine di f_1 ed in particolare $f_1 = 0$ se e solo se $\operatorname{Ker} f_2 = 0$.

Esempio 5.6.4. Supponiamo che

$$V_0 \xrightarrow{f_0} V_1 \xrightarrow{f_1} V_2 \xrightarrow{f_2} V_3 \xrightarrow{f_3} V_4$$

sia una successione esatta. Allora le seguenti condizioni sono equivalenti:

(1) f_0 è surgettiva e f_3 è iniettiva;
(2) $f_1 = f_2 = 0$;
(3) $V_2 = 0$.

I ragionamenti da fare sono analoghi a quelli dell'esempio precedente e lasciati per esercizio al lettore.

Definizione 5.6.5. Una **successione esatta corta** di spazi vettoriali è una successione esatta del tipo

$$0 \to U \xrightarrow{f} V \xrightarrow{g} W \to 0. \tag{5.9}$$

Dunque, la (5.9) è una successione esatta corta se e solo se f è iniettiva, g è surgettiva e $\operatorname{Ker} g = f(U)$.

In particolare, se (5.9) è una successione esatta si ha $W = g(V)$, $f\colon U \to \operatorname{Ker} g$ è un isomorfismo e, se V ha dimensione finita, per il teorema del rango si ha

$$\dim V = \dim \operatorname{Ker} g + \dim g(V) = \dim U + \dim W.$$

Esempio 5.6.6. Consideriamo una successione esatta

$$0 \to V_1 \xrightarrow{f_1} V_2 \xrightarrow{f_2} V_3 \xrightarrow{f_3} V_4 \to 0$$

e indichiamo con $U = \operatorname{Ker} f_3 = f_2(V_2)$ e con $i\colon U \to V_3$ il morfismo di inclusione. Allora la precedente successione si *spezza* in due successioni esatte corte

$$0 \to V_1 \xrightarrow{f_1} V_2 \xrightarrow{f_2} U \to 0, \qquad 0 \to U \xrightarrow{i} V_3 \xrightarrow{f_3} V_4 \to 0.$$

Se gli spazi vettoriali V_i hanno dimensione finita ricaviamo

$$\dim V_3 = \dim V_4 + \dim U, \qquad \dim V_2 = \dim V_1 + \dim U,$$

e quindi

$$\dim V_1 + \dim V_3 = \dim V_2 + \dim V_4\,.$$

Teorema 5.6.7. *Siano $U \xrightarrow{i} V \xrightarrow{p} W \to 0$ una successione esatta di spazi vettoriali e $f\colon V \to H$ un'applicazione lineare. Allora esiste un'applicazione lineare $g\colon W \to H$ tale che $f = gp$ se e solo se $fi = 0$.*

Dimostrazione. Siccome $pi = 0$ per definizione di complesso, e quindi a maggior ragione di successione esatta, se vale $f = gp$ allora $fi = gpi = g0 = 0$. Viceversa, supponiamo $fi = 0$ e usiamo la surgettività di p per definire $g\colon W \to H$ ponendo $g(w) = f(v)$, dove v è un *qualsiasi* vettore di V tale che $p(v) = w$.

Prima di proseguire dobbiamo dimostrare che g è *ben definita*, ossia che per ogni vettore $w \in W$ fissato il valore $g(w) = f(v)$ non dipende dalla scelta di v: se $p(v_1) = p(v_2) = w$, allora $p(v_1 - v_2) = 0$, ossia $v_1 - v_2 \in \operatorname{Ker} p$ e per esattezza esiste $u \in U$ tale che $v_1 - v_2 = i(u)$ e quindi

$$f(v_1) = f(v_2 + i(u)) = f(v_2) + f(i(u)) = f(v_2).$$

Dunque la precedente definizione di g è ben posta, in particolare dall'uguaglianza $p(v) = p(v)$ per ogni $v \in V$ segue che $g(p(v)) = f(v)$. La dimostrazione della linearità di g è lasciata per esercizio (cf. Esercizio 250). □

Teorema 5.6.8 (Lemma dei 5). *Sia dato il seguente diagramma commutativo di spazi vettoriali*

$$\begin{array}{ccccccccc}
E_1 & \xrightarrow{d_1} & E_2 & \xrightarrow{d_2} & E_3 & \xrightarrow{d_3} & E_4 & \xrightarrow{d_4} & E_5 \\
\downarrow{\scriptstyle \alpha_1} & & \downarrow{\scriptstyle \alpha_2} & & \downarrow{\scriptstyle \beta} & & \downarrow{\scriptstyle \alpha_4} & & \downarrow{\scriptstyle \alpha_5} \\
H_1 & \xrightarrow{h_1} & H_2 & \xrightarrow{h_2} & H_3 & \xrightarrow{h_3} & H_4 & \xrightarrow{h_4} & H_5
\end{array}$$

con entrambe le righe esatte.

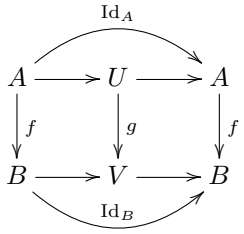
(1) *se α_1 è surgettiva e α_2, α_4 sono iniettive, allora β è iniettiva;*
(2) *se α_5 è iniettiva e α_2, α_4 sono surgettive, allora β è è surgettiva;*
(3) *se $\alpha_1, \alpha_2, \alpha_4, \alpha_5$ sono bigettive, allora β è è bigettiva.*

Dimostrazione. Dimostriamo solo il primo punto; la dimostrazione del secondo è del tutto simile ed è lasciata per esercizio. Il terzo punto segue banalmente dai primi due.

Sia $x \in E_3$ tale che $\beta(x) = 0$, allora $\alpha_4 d_3(x) = h_3\beta(x) = 0$ ed essendo per ipotesi α_4 iniettiva si ha $d_3(x) = 0$. La prima riga è esatta e quindi esiste $y \in E_2$ tale che $x = d_2(y)$; siccome $h_2\alpha_2(y) = \beta d_2(y) = \beta(x) = 0$ e la riga inferiore è esatta, esiste $z \in H_1$ tale che $h_1(z) = \alpha_2(y)$. Adesso usiamo la surgettività di α_1 per trovare $w \in E_1$ tale che $\alpha_1(w) = z$, quindi $\alpha_2 d_1(w) = h_1\alpha_1(w) = h_1(z) = \alpha_2(y)$. Per l'iniettività di α_2 si ha $y = d_1(w)$ e quindi $x = d_2(y) = d_2 d_1(w) = 0$. □

Esercizi.

292 (retrazioni). Con il termine retrazione si intende un diagramma commutativo di applicazioni lineari



ed in tal caso diremo che f è un retratto di g.

(1) Dimostrare che la relazione di retrazione gode della proprietà transitiva, ossia che se f è un retratto di g e se g è un retratto di h, allora f è un retratto di h.
(2) Sia f un retratto di g, dimostrare che se g è iniettiva (risp.: surgettiva), allora anche f è iniettiva (risp.: surgettiva).
(3) Sia $e\colon V \to V$ un'applicazione lineare tale che $e^2 = e$ e si denoti

$$F = \{v \in V \mid e(v) = v\}.$$

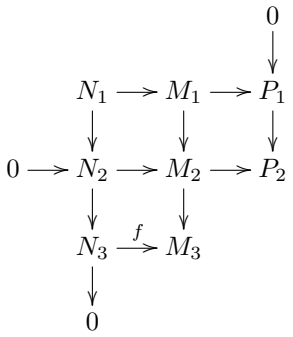
Dimostrare che:
(a) F è un sottospazio vettoriale di V che coincide con l'immagine di e;
(b) siano $i\colon F \to V$ il morfismo di inclusione e $p\colon V \to F$ l'unica applicazione lineare tale che $ip = e$. Allora i e p sono entrambe dei retratti di e.

293. Data una successione esatta di spazi vettoriali di dimensione finita

$$0 \to V_1 \to V_2 \to \cdots \to V_{n-1} \to V_n \to 0,$$

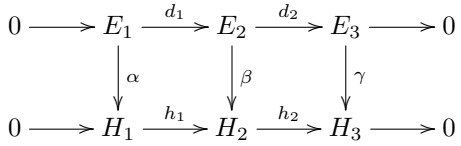
dimostrare che $\sum_{i=1}^{n} (-1)^i \dim V_i = 0$.

294. Si consideri il diagramma commutativo di spazi vettoriali



in cui tutte le righe e tutte le colonne sono successioni esatte. Provare che l'applicazione f è iniettiva.

295. Si abbia un diagramma commutativo di spazi vettoriali



con entrambe le righe esatte. Dimostrare che la successione indotta

$$0 \to \mathrm{Ker}(\alpha) \xrightarrow{d_1} \mathrm{Ker}(\beta) \xrightarrow{d_2} \mathrm{Ker}(\gamma)$$

è esatta. Mostrare inoltre che se α è surgettiva, allora anche $\mathrm{Ker}(\beta) \xrightarrow{d_2} \mathrm{Ker}(\gamma)$ è surgettiva.

296. Si consideri il seguente diagramma commutativo di spazi vettoriali:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & N_1 & \longrightarrow & M_1 & \longrightarrow & P_1 & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & N_2 & \xrightarrow{f} & M_2 & \xrightarrow{g} & P_2 & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & N_3 & \longrightarrow & M_3 & \longrightarrow & P_3 & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

Si assuma che le colonne siano esatte e che $gf = 0$. Provare che ogni riga è un complesso e che se due righe sono esatte allora è esatta anche la terza.

CAPITOLO 6

Operazioni con le matrici

Abbiamo già introdotto il concetto di matrice e lo spazio vettoriale $M_{n,m}(\mathbb{K})$ delle matrici $n \times m$ a coefficienti nel campo $\mathbb{K}$. Oltre alle tipiche operazioni dettate dalla struttura di spazio vettoriale (somma e prodotto per scalare), le matrici possiedono ulteriori caratteristiche che le rendono estremamente interessanti dal punto di vista matematico; in particolare è definito il prodotto "righe per colonne" che è la controparte algebrica del prodotto di composizione di applicazioni lineari.

6.1. Traccia e trasposta

Ricordiamo che una matrice è nulla se ha tutti i coefficienti uguali a 0.

DEFINIZIONE 6.1.1. La **trasposta** di una matrice $A \in M_{n,m}(\mathbb{K})$ è la matrice $A^T \in M_{m,n}(\mathbb{K})$ ottenuta scambiando l'indice di riga con quello di colonna ai coefficienti di A.

In altri termini, i coefficienti della prima riga di A^T (da sinistra a destra) sono uguali a quelli della prima colonna di A (dall'alto al basso) eccetera:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}^T = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{pmatrix}$$

Ad esempio:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 4 \\ 8 & 16 & 32 \end{pmatrix}^T = \begin{pmatrix} 1 & 8 \\ 2 & 16 \\ 4 & 32 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}^T = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix}.$$

La trasposizione commuta con le operazioni di somma e prodotto per scalare, ciò significa che valgono le formule:

$$(A+B)^T = A^T + B^T, \qquad \text{per ogni } A, B \in M_{n,m}(\mathbb{K}),$$
$$(\lambda A)^T = \lambda A^T, \qquad \text{per ogni } A \in M_{n,m}(\mathbb{K}),\ \lambda \in \mathbb{K}.$$

Anche lo spazio delle matrici possiede una base canonica (vedi Esempio 4.4.5) formata dalle matrici che hanno un solo coefficiente non nullo ed uguale ad 1. Più precisamente, per ogni $i = 1, \dots, n$ ed ogni $j = 1, \dots, m$ indichiamo con $E_{ij} \in M_{n,m}(\mathbb{K})$ la matrice che ha il coefficiente (i, j) uguale a 1 e tutti gli altri uguali a 0. Ad esempio per $n = 2$ e $m = 3$ si ha:

$$E_{11} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad E_{22} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad E_{23} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Dunque, per ogni matrice (a_{ij}) possiamo scrivere

$$\begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix} = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} E_{ij}$$

e da questo si deduce che le matrici E_{ij} sono generatori linearmente indipendenti. In particolare $M_{n,m}(\mathbb{K})$ ha dimensione nm come spazio vettoriale su $\mathbb{K}$.

DEFINIZIONE 6.1.2. Una matrice $n \times n$ si dice **quadrata di ordine** n. I coefficienti sulla *diagonale principale* di una matrice quadrata sono quelli che hanno indice di riga uguale a quello di colonna.

Ad esempio, nella seguente matrice quadrata di ordine 3

$$\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix},$$

i coefficienti sulla diagonale principale sono uguali a 2 e quelli al di fuori della diagonale principale sono uguali a 1.

DEFINIZIONE 6.1.3. Una matrice si dice **diagonale** se è quadrata ed è nulla al di fuori della diagonale principale. Equivalentemente una matrice (a_{ij}) è diagonale se è quadrata e $a_{ij} = 0$ per ogni $i \neq j$.

Ad esempio, delle seguenti quattro matrici, le prime tre sono diagonali mentre la quarta non è diagonale:

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

Si noti in particolare che ogni matrice 1×1 è diagonale, cosí come è diagonale ogni matrice quadrata con tutti i coefficienti uguali a 0. Il sottoinsieme di $M_{n,n}(\mathbb{K})$ formato da tutte le matrici diagonali è un sottospazio vettoriale di dimensione n: infatti è generato dagli elementi E_{ii}, $i = 1, \dots, n$, della base canonica.Ogni matrice diagonale è uguale alla sua trasposta.

DEFINIZIONE 6.1.4. Per ogni $n > 0$ denotiamo con $I_n \in M_{n,n}(\mathbb{K})$ la matrice che ha coefficienti uguali ad 1 sulla diagonale principale ed uguali a 0 al di fuori della diagonale principale. Chiameremo I_n **matrice identità** di ordine n.

Abbiamo visto nell'Esempio 5.4.7 che $I_n \in M_{n,n}(\mathbb{K})$ è l'unica matrice tale che $L_{I_n} = \mathrm{Id}_{\mathbb{K}^n}$. Per semplicità di notazione, quando l'ordine n è chiaro dal contesto scriveremo solamente I per indicare la matrice identità.

DEFINIZIONE 6.1.5. La **traccia** $\mathrm{Tr}(A)$ di una matrice quadrata A è la somma dei coefficienti sulla diagonale principale, ossia $\mathrm{Tr}(a_{ij}) = \sum_i a_{ii}$.

Ad esempio

$$\mathrm{Tr}\begin{pmatrix} 1 & 3 \\ 6 & -1 \end{pmatrix} = 0, \quad \mathrm{Tr}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 2, \quad \mathrm{Tr}\begin{pmatrix} 2 & 3 \\ 4 & 6 \end{pmatrix} = 8.$$

Notiamo che la traccia è un'applicazione lineare $\mathrm{Tr}\colon M_{n,n}(\mathbb{K}) \to \mathbb{K}$ non nulla e quindi il suo nucleo, lo spazio delle matrici a traccia nulla, ha dimensione $n^2 - 1$. Infine, ogni matrice quadrata ha traccia uguale alla sua trasposta: $\mathrm{Tr}(A) = \mathrm{Tr}(A^T)$.

Esercizi.

297. Siano

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

Calcolare $A + B^T, A^T + B, (A + B)^T$.

298. Trovare due matrici $S, E \in M_{2,2}(\mathbb{R})$ tali che

$$S = S^T, \qquad E = -E^T, \qquad S + E = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}.$$

299. Quali sono le matrici A tali che $A = 2A^T$?

300 (☕, ♡). Nello spazio vettoriale $M_{4,4}(\mathbb{K})$, per ogni coppia di indici $i, j = 1, \dots, 4$ indichiamo con $V_{ij} \subseteq M_{4,4}(\mathbb{K})$ l'insieme delle matrici tali che la somma dei coefficienti della riga i è uguale alla somma dei coefficienti della colonna j, ossia

$$V_{ij} = \left\{ (a_{hk}) \in M_{4,4}(\mathbb{K}) \;\middle|\; \sum_{k=1}^{4} a_{ik} = \sum_{h=1}^{4} a_{hj} \right\}.$$

Provare che ogni V_{ij} è un sottospazio vettoriale di dimensione 15 e calcolare la dimensione dell'intersezione dei 16 sottospazi V_{ij}.

301 (☕, ♡). Nello spazio vettoriale $M_{4,4}(\mathbb{K})$, per ogni coppia di indici $i, j = 1, \dots, 4$ indichiamo con $U_{ij} \subseteq M_{4,4}(\mathbb{K})$ l'insieme delle matrici tali che la somma dei coefficienti della riga i è uguale al doppio della somma dei coefficienti della colonna j, cioè

$$U_{ij} = \left\{ (a_{hk}) \in M_{4,4}(\mathbb{K}) \;\mid\; \sum_{k=1}^{4} a_{ik} = 2 \sum_{h=1}^{4} a_{hj} \right\}.$$

Provare che ogni U_{ij} è un sottospazio vettoriale di dimensione 15 e calcolare la dimensione dell'intersezione dei 16 sottospazi U_{ij}.

6.2. L'algebra delle matrici

Supponiamo di avere due matrici $A \in M_{n,m}(\mathbb{K})$ e $B \in M_{m,l}(\mathbb{K})$ tali che il numero di colonne di A sia uguale al numero di righe di B. In tal caso possiamo fare il prodotto righe per colonne di A con ciascuna colonna di B ed ottenere una successione di l vettori di $\mathbb{K}^n$. Se indichiamo con B^i le colonne di B, definiamo in prodotto righe per colonne

$$AB = A(B^1, \dots, B^l) = (AB^1, \dots, AB^l) \in M_{n,l}(\mathbb{K}).$$

In particolare ogni colonna di AB è combinazione lineare delle colonne di A.

Il **prodotto righe per colonne** AB è definito solo quando il numero di colonne di A è uguale al numero di righe di B e la matrice AB ha lo stesso numero di righe di A e di colonne di B.

Esempio 6.2.1. Siano

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & -1 \\ 0 & 2 \\ 1 & 0 \end{pmatrix}.$$

Allora possiamo effettuare il prodotto AB e le colonne di AB sono i due vettori

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 10 \\ 16 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \\ 9 \end{pmatrix},$$

e quindi

$$AB = \begin{pmatrix} 4 & 3 \\ 10 & 6 \\ 16 & 9 \end{pmatrix}.$$

Se indichiamo con a_{ij}, b_{ij} e c_{ij} rispettivamente i coefficienti delle matrici A, B e $AB = (c_{ij})$, si ha per definizione che c_{ij} è il prodotto della riga i della matrice A con la riga j della matrice B; in formule:

$$c_{ij} = \sum_h a_{ih} b_{hj} = a_{i1} b_{1j} + a_{i2} b_{2j} + \cdots + a_{im} b_{mj}.$$

Esempio 6.2.2.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} u & v \\ w & z \end{pmatrix} = \begin{pmatrix} au + bw & av + bz \\ cu + dw & cv + dz \end{pmatrix}.$$

Esempio 6.2.3.

$$\begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} 5 \\ 7 \end{pmatrix} = (19), \qquad \begin{pmatrix} 5 \\ 7 \end{pmatrix} \begin{pmatrix} 1 & 2 \end{pmatrix} = \begin{pmatrix} 5 & 10 \\ 7 & 14 \end{pmatrix}.$$

Esempio 6.2.4.

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

Esempio 6.2.5.

$$\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 5 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 9 \end{pmatrix} = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 21 & 0 \\ 0 & 0 & -9 \end{pmatrix}.$$

Le ragioni del perché il prodotto righe per colonne è importante, mentre altri possibili prodotti, come ad esempio quello coefficiente per coefficiente, non lo sono affatto deriva principalmente dal seguente risultato.

LEMMA 6.2.6. *Siano $A \in M_{n,m}(\mathbb{K})$ e $B \in M_{m,l}(\mathbb{K})$ due matrici. Allora vale*

$$L_{AB} = L_A \circ L_B \colon \mathbb{K}^l \to \mathbb{K}^n.$$

DIMOSTRAZIONE. Dato che $L_{AB}(x) = (AB)x$ e $L_A(L_B(x)) = A(Bx)$ bisogna dimostrare che per ogni $x \in \mathbb{K}^l$ vale $(AB)x = A(Bx)$. Siano $A = (a_{ij})$, $B = (b_{ij})$ $AB = (c_{ij})$ e $x \in \mathbb{K}^l$ fissato; denotiamo $y = Bx$, $z = Ay$ e $w = (AB)x$. Bisogna dimostrare che $z = w$. La coordinata i-esima di z è

$$z_i = \sum_h a_{ih} y_h = \sum_h a_{ih} (\sum_k b_{hk} x_k) = \sum_{h,k} a_{ih} b_{hk} x_k,$$

mentre la coordinata i-esima di w è

$$w_i = \sum_k c_{ik} x_k = \sum_k (\sum_h a_{ih} b_{hk}) x_k = \sum_{h,k} a_{ih} b_{hk} x_k.$$

□

Quindi, **il prodotto righe per colonne tra matrici corrisponde al prodotto di composizione tra le corrispondenti applicazioni lineari tra spazi vettoriali numerici.**

ESEMPIO 6.2.7. Moltiplicare per la matrice identità è un'operazione neutra, ossia

$$IA = A, \qquad BI = B,$$

per ogni scelta delle matrici A, B: ovviamente in entrambi i casi l'ordine della matrice I deve essere tale che i prodotti IA e BI siano definiti.

Infatti $L_{IA} = L_I \circ L_A = \mathrm{Id} \circ L_A = L_A$ e quindi $IA = A$.

TEOREMA 6.2.8. *Il prodotto di matrici è associativo, ossia per ogni $A \in M_{n,m}(\mathbb{K})$, $B \in M_{m,l}(\mathbb{K})$ e $C \in M_{l,p}(\mathbb{K})$ vale*

$$(AB)C = A(BC).$$

DIMOSTRAZIONE. Siccome il prodotto di matrici corrisponde al prodotto di composizione di applicazioni lineari, il teorema segue dall'associatività del prodotto di composizione. Possiamo comunque dare una diversa dimostrazione: infatti se $A = (a_{ij})$, $B = (b_{ij})$, $C = (c_{ij})$ e se denotiamo $AB = (d_{ij})$, $BC = (e_{ij})$, allora i coefficienti (i, j) delle matrici $(AB)C$ e $A(BC)$ sono rispettivamente

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

$$\sum_k a_{ik} e_{kj} = \sum_k a_{ik} (\sum_h b_{kh} c_{hj}) = \sum_{h,k} a_{ik} b_{kh} c_{hj}.$$

e dunque coincidenti. □

Come prima conseguenza dell'associatività del prodotto notiamo che per una qualsiasi matrice quadrata A se ne possono fare le potenze: $A^1 = A$, $A^2 = AA$, $A^3 = AAA$ ecc. È chiaro che per ogni $n, m > 0$ vale la regola $A^n A^m = A^{n+m}$.

Un'altra proprietà del prodotto di matrici, la cui semplice dimostrazione è lasciata per esercizio, è la proprietà distributiva: siano $A \in M_{n,m}(\mathbb{K})$ e $B, C \in M_{m,l}(\mathbb{K})$, allora vale $A(B + C) = AB + BC$. Similmente se $A, B \in M_{n,m}(\mathbb{K})$ e $C \in M_{m,l}(\mathbb{K})$ si ha $(A + B)C = AC + BC$.

LEMMA 6.2.9. *Per ogni $A = (a_{ij}) \in M_{n,m}(\mathbb{K})$ e $B = (b_{hk}) \in M_{m,n}(\mathbb{K})$ vale*

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA) = \sum_{i,j} a_{ij} b_{ji}.$$

DIMOSTRAZIONE. Basta sviluppare i conti. □

Il prodotto si comporta "bene" rispetto alla trasposizione.

LEMMA 6.2.10. *Per ogni $A \in M_{n,m}(\mathbb{K})$ e $B \in M_{m,l}(\mathbb{K})$ vale*

$$(AB)^T = B^T A^T.$$

DIMOSTRAZIONE. Siano $A = (a_{ij})$, $B = (b_{ij})$. Il coefficiente (i,j) di AB è $\sum_h a_{ih}b_{hj}$, mentre il coefficiente (j,i) di $B^T A^T$ è $\sum_h b_{hj}a_{ih}$ che coincide con il precedente. □

Abbiamo già visto nell'Esempio 6.2.4 che il prodotto di matrici non è commutativo, ossia in generale vale $AB \neq BA$. Quindi bisogna prestare molta attenzione allo svolgimento delle espressioni algebriche contenenti matrici. Ad esempio, se $A, B \in M_{n,n}(\mathbb{K})$ vale

$$(A+B)^2 = (A+B)(A+B) = A(A+B) + B(A+B) = A^2 + AB + BA + B^2$$

e non, come potremmo scrivere senza riflettere, $A^2 + 2AB + B^2$.

ESEMPIO 6.2.11. Sia $A \in M_{n,n}(\mathbb{K})$ una matrice quadrata, allora A commuta con tutte le matrici del tipo $\lambda_0 I + \lambda_1 A + \cdots + \lambda_m A^m$ al variare di $m \geq 0$ e $\lambda_0, \ldots, \lambda_m \in \mathbb{K}$. Infatti siccome il prodotto è associativo si ha $AA^i = A^{i+1} = A^i A$ e quindi

$$A(\lambda_0 I + \lambda_1 A + \cdots + \lambda_m A^m) = \lambda_0 AI + \lambda_1 AA + \cdots + \lambda_m AA^m$$

$$= \lambda_0 IA + \lambda_1 AA + \cdots + \lambda_m A^m A = (\lambda_0 I + \lambda_1 A + \cdots + \lambda_m A^m)A.$$

ESEMPIO 6.2.12. Sia $U \in M_{n,n}(\mathbb{K})$ una matrice che commuta con tutte le matrici $n \times n$, allora U è un multiplo scalare dell'identità. Siano u_{ij} i coefficienti di U e consideriamo i prodotti con le matrici E_{ij} della base canonica, al variare di tutti gli indici $i < j$. Un semplice conto dimostra che

$$E_{ij}U = \sum_k u_{jk}E_{ik}, \qquad UE_{ij} = \sum_k u_{ki}E_{kj}$$

e se $E_{ij}U = UE_{ij}$ allora

$$\sum_\beta u_{j\beta}E_{i\beta} = \sum_\alpha u_{\alpha i}E_{\alpha j}.$$

Da tale uguaglianza, ricordando che le matrici E_{hk} sono linearmente indipendenti, segue immediatamente che $u_{j\beta} = 0$ per ogni $\beta \neq j$, che $u_{\alpha i} = 0$ per ogni $\alpha \neq i$ e che $u_{ii} = u_{jj}$. Ma questo significa dire che U è un multiplo scalare dell'identità.

Talvolta conviene rappresentare una matrice $A \in M_{n,m}(\mathbb{K})$ sotto forma di **matrice a blocchi**

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1s} \\ A_{21} & A_{22} & \cdots & A_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1} & A_{r2} & \cdots & A_{rs} \end{pmatrix}, \qquad A_{ij} \in M_{k_i,h_j}(\mathbb{K}), \quad \sum k_i = n, \ \sum h_j = m.$$

Il prodotto righe per colonne di matrici a blocchi si può fare eseguendo i prodotti righe per colonne dei blocchi, qualora beninteso tali prodotti siano definiti. Ad esempio se

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \qquad A_{ij} \in M_{k_i,h_j}(\mathbb{K}),$$

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \qquad B_{ij} \in M_{h_i,l_j}(\mathbb{K}),$$

allora

$$AB = \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}.$$

Esercizi.

302. Date le matrici

$$A = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}, \quad C = \begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix},$$

calcolare AB, BA, BC, CB, AC e CA.

303. Trovare una maniera "furba" per calcolare le seguenti elevazioni a potenza:

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}^{2020}, \qquad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{16244}, \qquad \frac{1}{2^{40}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}^{80}.$$

304. Data la matrice $A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, determinare:

(1) tutte le matrici B quadrate di ordine 2 tali che $AB = 0$;
(2) tutte le matrici C quadrate di ordine 2 tali che $CA = 0$.

305. Calcolare in funzione di $A \in M_{n,n}(\mathbb{K})$ le potenze $B^2, B^3, \dots$ della matrice a blocchi

$$B = \begin{pmatrix} 0 & A \\ I & 0 \end{pmatrix} \in M_{2n,2n}(\mathbb{K}),$$

dove I denota la matrice identità $n \times n$.

306 (Il corpo dei quaternioni). Per ogni coppia di numeri complessi $a, b \in \mathbb{C}$ denotiamo

$$Q(a,b) = \begin{pmatrix} a & b \\ -\overline{b} & \overline{a} \end{pmatrix} \in M_{2,2}(\mathbb{C}).$$

Dimostrare che:

(1) $Q(1,0)Q(a,b) = Q(a,b)Q(1,0) = Q(a,b)$ per ogni $a, b \in \mathbb{C}$;
(2) $Q(a,b)Q(c,d) = Q(ac - b\overline{d}, ad + b\overline{c})$ per ogni $a, b, c, d \in \mathbb{C}$;
(3) se $a, b \in \mathbb{C}$ non sono entrambi nulli allora

$$Q(a,b)Q\left(\frac{\overline{a}}{|a|^2 + |b|^2}, \frac{-b}{|a|^2 + |b|^2}\right) = Q(1,0).$$

L'insieme $\mathbb{H} \subseteq M_{2,2}(\mathbb{C})$ formato da tutte le matrici del tipo $Q(a,b)$ viene detto **corpo dei quaternioni**. Si tratta di un sottospazio vettoriale reale di dimensione 4, chiuso per il prodotto, con elemento neutro per il prodotto $Q(1,0)$, in cui ogni elemento non nullo possiede un inverso. Dunque $\mathbb{H}$ soddisfa tutti gli assiomi di campo tranne la commutatività del prodotto, ragion per cui viene chiamato corpo, o campo non commutativo.

Gli otto elementi $\pm Q(1,0), \pm Q(i,0), \pm Q(0,1), \pm Q(0,i)$ sono dette **unità quaternionali**; provare che il prodotto di due unità quaternionali è ancora una unità quaternionale.

Infine, provare che il polinomio $t^2 - Q(-1,0)$ possiede infinite radici in $\mathbb{H}$, ragion per cui il Teorema 3.7.5 non vale nei campi non commutativi.

Sebbene $\mathbb{H}$ non è un sottospazio vettoriale complesso di $M_{2,2}(\mathbb{C})$, esso possiede una struttura di spazio vettoriale complesso di dimensione 2, con il prodotto per scalare definito dalla formula

$$z \cdot Q(a,b) = Q(a,b)Q(z,0) = Q(za, \overline{z}b).$$

307. Sia $A = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$. Descrivere tutte le matrici B tali che $AB = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

308. Calcolare

$$\begin{pmatrix} 2 & 0 & 1 \\ 2 & 3 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 5 & 4 & 3 \\ 0 & 7 & 2 \\ -1 & 0 & 9 \end{pmatrix}, \qquad \begin{pmatrix} 5 & 0 & -1 \\ 4 & 7 & 0 \\ 3 & 2 & 9 \end{pmatrix} \begin{pmatrix} 2 & 2 & 0 \\ 0 & 3 & 1 \\ 1 & 0 & -1 \end{pmatrix}$$

e dire se le due matrici prodotto sono una la trasposta dell'altra.

309. Indichiamo con E_{ij} la base canonica di $M_{n,n}(\mathbb{K})$ e con $E_1, \dots, E_n$ la base canonica di $M_{n,1}(\mathbb{K})$:

$$E_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots \quad E_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Convincetevi che $E_i E_j^T = E_{ij} \in M_{n,n}(\mathbb{K})$, $E_{ij}E_{jk} = E_{ik}$ e, se $j \neq h$ allora $E_{ij}E_{hk} = 0$.

310. Per ogni numero $\xi = a + b\sqrt{2} \in \mathbb{Q}(\sqrt{2})$ definiamo la matrice

$$R(\xi) = \begin{pmatrix} a & 2b \\ b & a \end{pmatrix} \in M_{2,2}(\mathbb{Q}).$$

Verificare che per ogni $\xi, \eta \in \mathbb{Q}(\sqrt{2})$ valgono le formule

$$R(\xi + \eta) = R(\xi) + R(\eta), \qquad R(\xi\eta) = R(\xi)R(\eta).$$

311. Denotando con $F_0 = 0$, $F_1 = 1$, $F_{n+1} = F_n + F_{n-1}$, i numeri di Fibonacci, provare che per ogni intero positivo n si ha

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^n = \begin{pmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{pmatrix}.$$

312. Per ogni numero complesso $z = a + ib \in \mathbb{C}$ definiamo la matrice

$$R(z) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \in M_{2,2}(\mathbb{R}).$$

Verificare che per ogni $z, u \in \mathbb{C}$ valgono le formule

$$R(z + u) = R(z) + R(u), \qquad R(zu) = R(z)R(u).$$

313. Per ogni numero reale $t \in \mathbb{R}$ definiamo la matrice

$$S(t) = \begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{pmatrix} \in M_{2,2}(\mathbb{R}).$$

Verificare che per ogni $a, b \in \mathbb{R}$ valgono le formule $S(-a) = S(a)^T$, $S(a + b) = S(a)S(b)$.

314. Calcolare i prodotti $(AB)^2$ e $(BA)^2$, dove

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

315. Calcolare il quadrato della matrice

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

316. Date le matrici quadrate

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

$$I = \begin{pmatrix} 1 & 3 & 0 & 1 \\ 0 & 1 & 2 & 2 \\ 4 & 1 & -1 & 0 \\ 0 & 1 & -1 & 1 \end{pmatrix}, \quad L = \begin{pmatrix} 0 & 1 & -1 & 1 \\ 1 & 1 & 1 & 0 \\ 2 & 2 & 2 & 0 \\ 2 & 1 & -1 & 3 \end{pmatrix},$$

quanto vale la traccia di $ALIBABA$? (Suggerimento: le matrici $ALIBABA$ e $AALIBAB$ hanno la stessa traccia? Quanto vale A^2?)

317. Sia
$$V = \left\{ \begin{pmatrix} a & 5b & 5c \\ c & a & 5b \\ b & c & a \end{pmatrix} \in M_{3,3}(\mathbb{R}) \;\middle|\; a, b, c \in \mathbb{R} \right\}.$$
(1) provare che $AB = BA \in V$ per ogni $A, B \in V$;
(2) trovare una matrice $X \in V$ tale che $X^3 = 5I$ e tale che I, X, X^2 sia una base di V come spazio vettoriale.

318. Dati tre numeri $a, b, c \in \mathbb{C}$, siano $\lambda = a^2 + b^2 + c^2$,
$$X = \begin{pmatrix} a \\ b \\ c \end{pmatrix}, \qquad A = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix}.$$
Verificare che $AX = 0$, $A^2 = XX^T - \lambda I$, $A^3 = -\lambda A$ e trovare una formula per A^n in funzione di X e λ per ogni intero pari $n \geq 4$.

319. Dato un intero $n > 1$ ed una matrice B quadrata $n \times n$ a coefficienti in un campo $\mathbb{K}$ si definisca
$$C(B) = \{A \in M_{n,n}(\mathbb{K}) \mid AB = BA\}.$$
(1) Provare che $C(B)$ è un sottospazio vettoriale di $M_{n,n}(\mathbb{K})$ di dimensione ≥ 2.
(2) Provare che $C(B) = C(B + \lambda I)$ per ogni $\lambda \in \mathbb{K}$.
(3) Determinare $C(B)$ nei casi seguenti:
$$B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

320. Il **commutatore** di due matrici $A, B \in M_{n,n}(\mathbb{K})$ è per definizione
$$[A, B] = AB - BA.$$
Dimostrare che per ogni $A, B, C \in M_{n,n}(\mathbb{K})$ vale:
(1) $[A, B] = -[B, A]$ e $[A, A] = 0$,
(2) $\mathrm{Tr}([A, B]) = 0$,
(3) (formula di Leibniz) $[A, BC] = [A, B]C + B[A, C]$,
(4) (identità di Jacobi) $[[A, B], C] = [A, [B, C]] - [B, [A, C]]$,
(5) al variare di $A, B \in M_{n,n}(\mathbb{K})$, le matrici del tipo $[A, B]$ generano lo spazio vettoriale delle matrici a traccia nulla. (Sugg.: studiare i commutatori delle matrici E_{ij} della base canonica.)

321. Tra dimensione finita ed infinita le cose possono andare molto diversamente:
(1) Sia V uno spazio vettoriale di dimensione finita su $\mathbb{R}$. Dimostrare che non esistono applicazioni lineari $f, g \colon V \to V$ tali che $gf - fg$ sia uguale all'identità.
(2) Si consideri l'applicazione lineare
$$f \colon \mathbb{R}[x] \to \mathbb{R}[x], \qquad f(p(x)) = xp(x)$$
data dalla moltiplicazione per x. Dimostrare che per ogni polinomio $q(x) \in \mathbb{R}[x]$ vi è un'unica applicazione lineare $g \colon \mathbb{R}[x] \to \mathbb{R}[x]$ tale che $g(1) = q(x)$ e $gf - fg$ è uguale all'identità su $\mathbb{R}[x]$. Se $h \colon \mathbb{R}[x] \to \mathbb{R}[x]$ denota la moltiplicazione per $q(x)$, chi è $g - h$?

322. Una *matrice Markoviana* è una matrice quadrata in cui la somma dei coefficienti di ciascuna colonna è uguale a 1. Dimostrare che le potenze delle matrici Markoviane sono ancora Markoviane.

323. Sia U una matrice quadrata che commuta con tutte le matrici diagonali dello stesso ordine. Dimostrare che U è diagonale.

324. Siano α, β, γ tre numeri complessi e definiamo $a, b, c \in \mathbb{C}$ mediante l'uguaglianza di polinomi
$$(t - \alpha)(t - \beta)(t - \gamma) = t^3 - at^2 - bt - c.$$
(1) descrivere le tre somme $\alpha^i + \beta^i + \gamma^i$, $i = 1, 2, 3$, come funzioni nelle variabili a, b, c;

(2) calcolare la traccia delle tre potenze di matrici

$$\begin{pmatrix} 0 & 0 & c \\ 1 & 0 & b \\ 0 & 1 & a \end{pmatrix}^i, \qquad i = 1, 2, 3;$$

(3) cosa lega i due punti precedenti?

325 (☕, ♡). Si consideri la matrice

$$A = \frac{1}{3}\begin{pmatrix} 1 & -2\sqrt{2} \\ 2\sqrt{2} & 1 \end{pmatrix} \in M_{2,2}(\mathbb{Q}(\sqrt{2})).$$

Dimostrare che per ogni intero positivo n la prima colonna di A^n è un vettore del tipo $\frac{1}{3^n}\begin{pmatrix} a \\ b\sqrt{2} \end{pmatrix}$ con a, b interi non divisibili per 3 e $a + b$ divisibile per 3.

6.3. Matrici invertibili

Date due matrici $A \in M_{n,m}(\mathbb{K}), B \in M_{m,l}(\mathbb{K})$ osserviamo che ogni vettore colonna del prodotto AB è una combinazione lineare dei vettori colonna di A: più precisamente la i-esima colonna di AB è la combinazione lineare delle colonne di A con coefficienti le coordinate dell'i-esimo vettore colonna di B. Similmente ogni riga di AB è combinazione lineare delle righe di B. Da questo ne deduciamo che:

(1) Date due matrici $A \in M_{n,m}(\mathbb{K}), D \in M_{n,l}(\mathbb{K})$, esiste una matrice $B \in M_{m,l}(\mathbb{K})$ tale che $AB = D$ se e solo se ogni colonna di D è combinazione lineare delle colonne di A.
(2) Date due matrici $A \in M_{n,m}(\mathbb{K}), D \in M_{l,m}(\mathbb{K})$, esiste una matrice $C \in M_{l,n}(\mathbb{K})$ tale che $CA = D$ se e solo se ogni riga di D è combinazione lineare delle righe di A.

Le precedenti considerazioni per $D = I$ matrice identità ci danno il seguente risultato.

LEMMA 6.3.1. *Sia $A \in M_{n,m}(\mathbb{K})$. Allora:*

(1) *Esiste $B \in M_{m,n}(\mathbb{K})$ tale che $AB = I_n$ se e solo se le colonne di A generano $\mathbb{K}^n$ (e quindi $m \geq n$).*
(2) *Esiste $C \in M_{m,n}(\mathbb{K})$ tale che $CA = I_m$ se e solo se le righe di A generano $\mathbb{K}^{(m)}$ (e quindi $m \leq n$).*

DIMOSTRAZIONE. Basta osservare che i vettori colonna di I_n sono la base canonica di $\mathbb{K}^n$, che i vettori riga di I_m sono una base di $\mathbb{K}^{(m)}$, ed applicare le osservazioni precedenti. □

LEMMA 6.3.2. *Sia $A \in M_{n,m}(\mathbb{K})$ e si assuma che esistano due matrici $B \in M_{m,n}(\mathbb{K})$ e $C \in M_{m,n}(\mathbb{K})$ tali che:*

$$AB = I_n, \quad CA = I_m. \tag{6.1}$$

Allora $n = m$ e $B = C$; in particolare, le matrici B, C sono le uniche che soddisfano (6.1).

DIMOSTRAZIONE. L'uguaglianza $n = m$ segue immediatamente dal Lemma 6.3.1. Si ha

$$C = CI_n = C(AB) = (CA)B = I_m B = B.$$

Se $\widetilde{B} \in M_{m,n}(\mathbb{K})$ e $\widetilde{C} \in M_{m,n}(\mathbb{K})$ sono tali che $A\widetilde{B} = I_n$ e $\widetilde{C}A = I_m$, lo stesso argomento applicato alle coppie $(B, \widetilde{C})$ e $(\widetilde{B}, C)$ mostra che $B = \widetilde{C}$, $\widetilde{B} = C$, e quindi che le quattro matrici $B, C, \widetilde{B}, \widetilde{C}$ coincidono. □

DEFINIZIONE 6.3.3. Una matrice quadrata $A \in M_{n,n}(\mathbb{K})$ si dice **invertibile** se esiste una matrice $A^{-1} \in M_{n,n}(\mathbb{K})$, tale che

$$A^{-1}A = AA^{-1} = I.$$

Per il Lemma 6.3.2, se A è invertibile allora A^{-1} è unica ed è anchessa invertibile con inversa $(A^{-1})^{-1} = A$. Se A è invertibile, allora anche A^T è invertibile con inversa $(A^T)^{-1} = (A^{-1})^T$: infatti

$$(A^{-1})^T A^T = (AA^{-1})^T = I^T = I, \qquad A^T(A^{-1})^T = (A^{-1}A)^T = I^T = I.$$

Se $A, B \in M_{n,n}(\mathbb{K})$ sono invertibili, allora anche AB è invertibile e vale $(AB)^{-1} = B^{-1}A^{-1}$.

TEOREMA 6.3.4. *Siano $A, B \in M_{n,n}(\mathbb{K})$ matrici* quadrate *tali che $AB = I$. Allora A e B sono invertibili e $B = A^{-1}$, $A = B^{-1}$.*

DIMOSTRAZIONE. Basta provare che esiste $C \in M_{n,n}(\mathbb{K})$ tale che $BC = I$; in tal caso, per il Lemma 6.3.2, si ha $A = C$ e dunque $AB = BA = I$.

L'esistenza di C equivale a dire che le colonne di B generano $\mathbb{K}^n$. Siccome B ha n colonne (qui stiamo usando l'ipotesi che B è quadrata), basta provare che le colonne di B sono linearmente indipendenti, ossia che l'applicazione L_B è iniettiva. L'iniettività di L_B segue immediatamente dal fatto che la composizione $L_A L_B = L_I = \mathrm{Id}$ è iniettiva. □

COROLLARIO 6.3.5. *Per una matrice quadrata A le seguenti condizioni sono equivalenti:*

(1) *A è invertibile.*
(2) *Le colonne di A sono linearmente indipendenti.*
(3) *Le righe di A sono linearmente indipendenti.*

DIMOSTRAZIONE. La condizione che le colonne siano indipendenti equivale al fatto che sono generatori, che a sua volta equivale all'esistenza di una matrice B tale che $AB = I$. Passando alla matrice trasposta otteniamo l'analogo risultato per le righe. □

Siano F un sottocampo di $\mathbb{K}$ e $A \in M_{n,n}(F) \subseteq M_{n,n}(\mathbb{K})$. Allora A è invertibile come matrice a coefficienti in $\mathbb{K}$ se e solo se è invertibile come matrice a coefficienti in F.

Infatti se A è invertibile come matrice a coefficienti in F allora $A^{-1} \in M_{n,n}(F) \subseteq M_{n,n}(\mathbb{K})$ e quindi A è invertibile anche come matrice a coefficienti in $\mathbb{K}$. Viceversa se A è invertibile come matrice a coefficienti in $\mathbb{K}$, allora le colonne sono linearmente indipendenti su $\mathbb{K}$ ed a maggior ragione sono linearmente indipendenti su F.

ESEMPIO 6.3.6 (La matrice di Vandermonde). Dati $n+1$ scalari $a_0, a_1, \dots, a_n$, la **matrice di Vandermonde** associata è la matrice $n \times n$

$$A = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ a_0 & a_1 & \cdots & a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_0^n & a_1^n & \cdots & a_n^n \end{pmatrix}$$

Dimostriamo che *la matrice A è invertibile se e solo se gli scalari a_i sono distinti.* Se $a_i = a_j$ per qualche coppia di indici $i \neq j$, allora le colonne di A non sono linearmente indipendenti e quindi A non è invertibile. Se A non è invertibile, allora le righe di A sono linearmente dipendenti, ossia esiste un vettore riga non nullo $(c_0, \dots, c_n)$ tale che $(c_0, \dots, c_n)A = 0$. Questo significa che per ogni $i = 0, \dots, n$ vale

$$c_0 + c_1 a_i + \cdots + c_n a_i^n = 0$$

e dunque che $a_0, a_1, \dots, a_n$ sono radici del polinomio non nullo $p(t) = c_0 + c_1 t + \cdots + c_n t^n$ che, avendo grado $\leq n$, possiede al più n radici distinte. Dunque $a_i = a_j$ per qualche coppia di indici $i \neq j$. Per il calcolo della matrice inversa A^{-1}, beninteso quando gli a_i sono distinti, rimandiamo all'Esercizio 332

ESEMPIO 6.3.7. Siano V uno spazio vettoriale su di un campo infinito $\mathbb{K}$, $U \subseteq V$ un sottospazio vettoriale e $v_0, \dots, v_n \in V$. Usiamo il risultato dell'Esempio 6.3.6 per dimostrare che, se

$$v_0 + t v_1 + t^2 v_2 + \cdots + t^n v_n \in U$$

per ogni $t \in \mathbb{K}$, allora $v_0, \dots, v_n \in U$.

Presi $n+1$ scalari distinti $a_0, \dots, a_n \in \mathbb{K}$, le $n+1$ relazioni $\sum_j a_i^j v_j \in U$ possono essere scritte in forma matriciale

$$(v_0, \dots, v_n) \begin{pmatrix} 1 & 1 & \cdots & 1 \\ a_0 & a_1 & \cdots & a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_0^n & a_1^n & \cdots & a_n^n \end{pmatrix} \in U^{(n+1)}.$$

Per l'Esempio 6.3.6 la matrice di Vandermonde è invertibile, moltiplicando a destra per l'inversa si ottiene

$$(v_0,\dots,v_n)=(v_0,\dots,v_n)\begin{pmatrix}1&1&\cdots&1\\a_0&a_1&\cdots&a_n\\\vdots&\vdots&\ddots&\vdots\\a_0^n&a_1^n&\cdots&a_n^n\end{pmatrix}\begin{pmatrix}1&1&\cdots&1\\a_0&a_1&\cdots&a_n\\\vdots&\vdots&\ddots&\vdots\\a_0^n&a_1^n&\cdots&a_n^n\end{pmatrix}^{-1}\in U^{(n+1)}.$$

COROLLARIO 6.3.8. *Sia $V\subseteq\mathbb{K}^n$ un sottospazio vettoriale di dimensione r. Allora esiste una matrice $A\in M_{n-r,n}(\mathbb{K})$ tale che $V=\mathrm{Ker}(L_A)=\{x\in\mathbb{K}^n\mid Ax=0\}$.*

DIMOSTRAZIONE. Sia $v_1,\dots,v_n\in\mathbb{K}^n$ una base i cui primi r vettori generano V e sia $B\in M_{n,n}(\mathbb{K})$ la matrice con vettori colonna $v_1,\dots,v_n$. Abbiamo dimostrato che B è invertibile e indichiamo con $A\in M_{n-r,n}(\mathbb{K})$ la matrice formata dalle ultime $n-r$ righe di B^{-1}, ossia $B^{-1}=\begin{pmatrix}C\\A\end{pmatrix}$, con $C\in M_{r,n}(\mathbb{K})$. Dalla formula $I=B^{-1}B=\begin{pmatrix}CB\\AB\end{pmatrix}$ ne consegue che AB è formata dalle ultime $n-r$ righe della matrice identità, e cioè,

$$Av_1=\cdots=Av_r=\begin{pmatrix}0\\\vdots\\0\end{pmatrix},\quad Av_{r+1}=\begin{pmatrix}1\\\vdots\\0\end{pmatrix},\ \dots,\ Av_n=\begin{pmatrix}0\\\vdots\\1\end{pmatrix}.$$

Dunque

$$A(t_1v_1+\cdots+t_nv_n)=\begin{pmatrix}t_{r+1}\\\vdots\\t_n\end{pmatrix}$$

e vale $Ax=0$ se e solo se $x\in V$. □

Esercizi.

326. Verificare se

$$\sqrt{2},\quad \sqrt{2}+\sqrt{5}+1,\quad \sqrt{10}+\sqrt{2},\quad \sqrt{5}+1,$$

formano una base di $\mathbb{Q}[\sqrt{2},\sqrt{5}]$ come spazio vettoriale su $\mathbb{Q}$.

327. Calcolare AB, dove

$$A=\begin{pmatrix}0&\sqrt{2}&\sqrt{3}\\-\sqrt{2}&0&\sqrt{5}\\-\sqrt{3}&-\sqrt{5}&0\end{pmatrix},\qquad B=\begin{pmatrix}\sqrt{5}\\-\sqrt{3}\\\sqrt{2}\end{pmatrix}.$$

Dire inoltre se i vettori colonna di A^{350} sono linearmente indipendenti.

328 (♡). Siano $A,B\in M_{n,n}(\mathbb{K})$ matrici invertibili tali che anche la loro somma $A+B$ è invertibile. Provare che anche $A^{-1}+B^{-1}$ è invertibile e valgono le formule

$$(A^{-1}+B^{-1})=A(A+B)^{-1}B=B(A+B)^{-1}A.$$

329. Data la matrice

$$A=\begin{pmatrix}1&1&-1\\0&0&1\\2&1&2\end{pmatrix}\in M_{3,3}(\mathbb{R}),$$

calcolare A^2, A^3 e verificare che

$$A^3=3A^2-3A+I. \tag{6.2}$$

Usando esclusivamente l'equazione (6.2), dimostrare che: $A^4=6A^2-8A+3I$, A è invertibile e $A^{-1}=A^2-3A+3I$.

330. Sia

$$D = \begin{pmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_n \end{pmatrix}$$

una matrice diagonale. Provare che se $d_i \neq d_j$ per ogni $i \neq j$, allora le matrici A che commutano con D (ossia $AD = DA$) sono tutte e sole le matrici diagonali. Provare inoltre che le n matrici $I, D, D^2, \ldots, D^{n-1}$ sono linearmente indipendenti in $M_{n,n}(\mathbb{K})$.

331. Provare che una matrice a blocchi $\begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$, dove $A \in M_{n,n}(\mathbb{K})$, $B \in M_{n,m}(\mathbb{K})$, $C \in M_{m,m}(\mathbb{K})$, è invertibile se e solo se A e C sono entrambe invertibili.

332 (Polinomi di Lagrange). Ad un insieme di $n+1$ scalari $a_0, a_1, \ldots, a_n \in \mathbb{K}$ distinti possiamo associare i corrispondenti **polinomi di Lagrange**:

$$L_i(t) = \prod_{j=0,\ldots,i-1,i+1,\ldots,n} \frac{t - a_j}{a_i - a_j}\,, \qquad i = 1, \ldots, n\,.$$

Dato che ciascun polinomio $L_i(t)$ ha grado n possiamo anche scrivere $L_i(t) = \sum_{j=0}^{n} b_{ij} t^j$ per opportuni coefficienti $b_{ij} \in \mathbb{K}$. Se consideriamo la matrice $B = (b_{ij}) \in M_{n+1,n+1}(\mathbb{K})$ si può dunque scrivere

$$B \begin{pmatrix} 1 \\ t \\ \vdots \\ t^n \end{pmatrix} = \begin{pmatrix} L_0(t) \\ L_1(t) \\ \vdots \\ L_n(t) \end{pmatrix}.$$

Dimostrare che:

(1) per ogni successione $c_0, \ldots, c_n \in \mathbb{K}$, il polinomio $p(t) = \sum_i c_i L_i(t)$ è l'unico polinomio di grado $\leq n$ tale che $p(a_i) = c_i$ per ogni i;
(2) il polinomio $\sum_i L_i(t)$ è invertibile in $\mathbb{K}[t]$;
(3) la matrice B è l'inversa della matrice di Vandermonde A dell'Esempio 6.3.6.

333. Sul campo dei numeri reali, calcolare le soluzioni del sistema lineare omogeneo

$$\begin{cases} x - y + z - w = 0 \\ x + y - z - w = 0 \\ x + y + z - 3w = 0 \end{cases}$$

e dedurre, senza fare conti, che il prodotto di matrici

$$\begin{pmatrix} 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 4 & 5 & 6 \\ 16 & 25 & 36 \\ 64 & 125 & 216 \end{pmatrix}$$

è invertibile.

334. Sia $p(x, y) \in \mathbb{C}[x, y]$ un polinomio in due variabili tale che $p(ax, ay) = a^n p(x, y)$ per ogni $a \in \mathbb{N}$. Provare che $p(x, y)$ è omogeneo di grado n, ossia che è combinazione lineare dei monomi $x^i y^j$, con $i + j = n$.

335. Dimostrare che due matrici $A \in M_{n,m}(\mathbb{K})$ e $B \in M_{m,n}(\mathbb{K})$ sono una l'inversa dell'altra, ossia $AB = I$ e $BA = I$, se e solo se le due applicazioni lineari

$$f \colon \mathbb{K}^n \times \mathbb{K}^m \to \mathbb{K}^n, \qquad f(x, y) = x + Ay,$$

$$g \colon \mathbb{K}^n \times \mathbb{K}^m \to \mathbb{K}^m, \qquad g(x, y) = Bx + y,$$

hanno lo stesso nucleo.

336. Sia $V = \mathbb{K}[t]_{\leq n}$ lo spazio vettoriale dei polinomi di grado $\leq n$ e siano $a_0, \ldots, a_n \in \mathbb{K}$ numeri distinti. Provare che:

(1) I polinomi

$$f_0 = 1, \quad f_1 = (t - a_1), \quad \dots, \ f_i = \prod_{j=1}^{i} (t - a_j), \quad i = 0, \dots, n,$$

formano una base di V.

(2) I polinomi

$$g_i = (t - a_i)^n, \qquad i = 0, \dots, n,$$

formano una base di V.

(3) I polinomi

$$g_i = (t - a_0) \cdots (t - a_{i-1})(t - a_{i+1}) \cdots (t - a_n), \qquad i = 0, \dots, n,$$

formano una base di V.

337 (♡). Sia $A \in M_{n,n}(\mathbb{K})$ una matrice i cui vettori colonna generano un sottospazio vettoriale di dimensione 1.

(1) Dimostrare che esistono $B \in M_{n,1}(\mathbb{K})$ e $C \in M_{1,n}(\mathbb{K})$ tali che $BC = A$.
(2) Dedurre dal punto precedente che $A^2 = \mathrm{Tr}(A)A$, dove $\mathrm{Tr}(A)$ indica la traccia di A.
(3) (☕) Provare che $A - tI$ è invertibile per ogni $t \in \mathbb{K}$, $t \neq 0, \mathrm{Tr}(A)$.

338. Diremo che una matrice quadrata A è invertibile in senso totalmente speciale (MTS), se è invertibile e se i coefficienti di A^{-1} sono uguali agli inversi dei coefficienti di A, ossia se, detto $A = (a_{ij})$ e $A^{-1} = (b_{ij})$, vale $a_{ij}b_{ij} = 1$ per ogni i, j. Dimostrare che:

(1) non esistono matrici 2×2 invertibili in senso MTS;
(2) non esistono matrici 3×3 reali ed invertibili in senso MTS.

339. Provare che il sottoinsieme $C \subset M_{n,n}(\mathbb{K})$ delle matrici non invertibili non è un sottospazio vettoriale. Provare che C è unione (possibilmente infinita) di sottospazi vettoriali di dimensione $n^2 - n$.

340 (☕,♡). Siano V, W due spazi vettoriali di dimensione n e $H \subset \mathrm{Hom}(V, W)$ un sottospazio vettoriale di dimensione $\geq n^2 - n + 1$. Dimostrare che H contiene un isomorfismo lineare $f\colon V \to W$.

6.4. Rango di una matrice

Iniziamo la sezione con una semplice applicazione del Corollario 5.1.16.

LEMMA 6.4.1. *Per una matrice $A \in M_{n,m}(\mathbb{K})$ le seguenti condizioni sono equivalenti:*

(1) *l'applicazione $L_A\colon \mathbb{K}^m \to \mathbb{K}^n$ è surgettiva;*
(2) *l'applicazione $L_{A^T}\colon \mathbb{K}^n \to \mathbb{K}^m$ è iniettiva.*

Prima di proseguire nella dimostrazione osserviamo che applicando il risultato del lemma alla matrice trasposta otteniamo l'enunciato duale, ossia che $L_A\colon \mathbb{K}^m \to \mathbb{K}^n$ è iniettiva se e solo se $L_{A^T}\colon \mathbb{K}^n \to \mathbb{K}^m$ è surgettiva.

DIMOSTRAZIONE. Dato che per ogni $x \in \mathbb{K}^n$ vale $(A^T x)^T = x^T A$ si ha che L_{A^T} è iniettiva se e solo se il vettore nullo è l'unico vettore riga $C \in \mathbb{K}^{(n)}$ tale che $CA = 0$.

Supponiamo L_A è surgettiva, allora le colonne di A generano $\mathbb{K}^n$ ed abbiamo visto che esiste una matrice B tale che $AB = I$. Dunque, se per un vettore riga $C \in \mathbb{K}^{(n)}$ vale $CA = 0$ si ha

$$0 = 0B = CAB = CI = C,$$

e quindi $C = 0$.

Viceversa, supponiamo L_A non surgettiva ed indichiamo con $U \subset \mathbb{K}^n$ la sua immagine; per il Corollario 5.1.16 esiste un'applicazione lineare non nulla $f\colon \mathbb{K}^n \to \mathbb{K}$ tale che $U \subseteq \mathrm{Ker}(f)$. Se $f = L_C$, con $0 \neq C = (c_1, \dots, c_n) \in M_{1,n}(\mathbb{K})$, allora ogni colonna di A appartiene al nucleo di L_C e quindi $CA = 0$, ossia $C^T \in \mathrm{Ker}(L_{A^T})$. □

DEFINIZIONE 6.4.2. Il **rango** $\mathrm{rg}(A)$ di una matrice A è la dimensione dell'immagine dell'applicazione lineare L_A, ossia la dimensione del sottospazio vettoriale generato dai vettori colonna. Equivalentemente, il rango è uguale al massimo numero di colonne linearmente indipendenti.

Osserviamo che per ogni matrice $A \in M_{n,m}(\mathbb{K})$ il rango $\mathrm{rg}(A)$ è sempre minore od uguale al minimo tra n ed m. Abbiamo dimostrato che una matrice quadrata $n \times n$ è invertibile se e solo se ha rango n.

Altre due semplici osservazioni:

(1) Se una matrice B è ottenuta da A eliminando una colonna, allora $\mathrm{rg}(B) \leq \mathrm{rg}(A)$. Questo è ovvio.
(2) Se una matrice C è ottenuta da A eliminando una riga, allora $\mathrm{rg}(C) \leq \mathrm{rg}(A)$. Infatti se $\mathrm{rg}(C) = r$ posso scegliere r colonne di C linearmente indipendenti e a maggior ragione le corrispondenti colonne di A sono ancora linearmente indipendenti.

Diremo che B è una **sottomatrice** di A se B si ottiene a partire da A eliminando alcune righe ed alcune colonne. Equivalentemente una sottomatrice è ottenuta scegliendo alcuni indici di riga, alcuni indici di colonna e prendendo i coefficienti con entrambi gli indici tra quelli scelti. Ad esempio le sottomatrici 2×2 di

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

sono

$$\begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix}, \begin{pmatrix} 1 & 3 \\ 4 & 6 \end{pmatrix}, \begin{pmatrix} 2 & 3 \\ 5 & 6 \end{pmatrix}. \tag{6.3}$$

Abbiamo visto che il rango di una sottomatrice di A è sempre minore od uguale al rango di A.

TEOREMA 6.4.3. *Per ogni matrice A vale $\mathrm{rg}(A) = \mathrm{rg}(A^T)$. Dunque, per ogni matrice, il massimo numero di colonne indipendenti è uguale al massimo numero di righe indipendenti.*

DIMOSTRAZIONE. Basta provare la disuguaglianza $\mathrm{rg}(A^T) \geq \mathrm{rg}(A)$; applicando la stessa disuguaglianza alla matrice trasposta otteniamo la disuguaglianza inversa $\mathrm{rg}(A) \geq \mathrm{rg}(A^T)$.

Supponiamo quindi $A \in M_{n,m}$ e sia $r = \mathrm{rg}(A)$. Allora è possibile scegliere r colonne di A linearmente indipedenti e, eliminando le rimanenti, trovare una sottomatrice $B \in M_{n,r}(\mathbb{K})$ di A le cui colonne sono linearmente indipendenti. Dunque L_B è iniettiva ed il Lemma 6.4.1 applicato alla matrice B^T implica che $L_{B^T} : \mathbb{K}^n \to \mathbb{K}^r$ è surgettiva, ossia $\mathrm{rg}(B^T) = r$. Siccome B^T è una sottomatrice di A^T, abbiamo provato che $\mathrm{rg}(A^T) \geq r$. □

DEFINIZIONE 6.4.4. Una sottomatrice quadrata viene anche detta **minore**. Un minore si dice **principale** se è ottenuto scegliendo righe e colonne con gli stessi indici.

Quindi, una matrice $n \times m$ possiede $\binom{n}{k}\binom{m}{k}$ minori di ordine k, dei quali $\min(\binom{n}{k}, \binom{m}{k})$ sono principali. Dei tre minori in (6.3) solo il primo è principale.

COROLLARIO 6.4.5. *Il rango di una matrice $A \in M_{n,m}(\mathbb{K})$ è uguale al massimo intero r tale che A possiede un minore di ordine r invertibile.*

DIMOSTRAZIONE. Sia r il rango di A. Siccome ogni sottomatrice ha rango $\leq r$ basta provare che A possiede una sottomatrice $r \times r$ invertibile. Sia $B \in M_{n,r}(\mathbb{K})$ la sottomatrice ottenuta scegliendo un insieme di r colonne linearmente indipendenti. Per costruzione $\mathrm{rg}(B) = r$ e per il Teorema 6.4.3 la matrice B possiede r righe linearmente indipendenti. Possiamo quindi trovare una sottomatrice C di B quadrata di ordine r con le righe indipendenti. Dunque C è invertibile. □

Il Corollario 6.4.5 implica che le considerazioni sul campo dei coefficienti fatte a proposito dell'invertibilità si applicano anche al rango: se F è un sottocampo di $\mathbb{K}$ e A è una matrice a coefficienti in F, allora il rango di A non cambia se la consideriamo come una matrice a coefficienti in $\mathbb{K}$.

Esercizi.

341. Il rango della matrice

$$\begin{pmatrix} 5 & 5 & 5 \\ 5 & 5 & 5 \\ 5 & 5 & 5 \end{pmatrix}$$

è uguale a 1, 3 o 5?

342. Calcolare il rango dela tabella Pitagorica

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 2 & 4 & 6 & 8 & 10 & 12 & 14 & 16 & 18 & 20 \\ 3 & 6 & 9 & 12 & 15 & 18 & 21 & 24 & 27 & 30 \\ 4 & 8 & 12 & 16 & 20 & 24 & 28 & 32 & 36 & 40 \\ 5 & 10 & 15 & 20 & 25 & 30 & 35 & 40 & 45 & 50 \\ 6 & 12 & 18 & 24 & 30 & 36 & 42 & 48 & 54 & 60 \\ 7 & 14 & 21 & 28 & 35 & 42 & 49 & 56 & 63 & 70 \\ 8 & 16 & 24 & 32 & 40 & 48 & 56 & 64 & 72 & 80 \\ 9 & 18 & 27 & 36 & 45 & 54 & 63 & 72 & 81 & 90 \\ 10 & 20 & 30 & 40 & 50 & 60 & 70 & 80 & 90 & 100 \end{pmatrix}.$$

343. Determinare una base dell'iperpiano di $\mathbb{K}^3$ di equazione $2x - y + 3z = 0$.

344. Calcolare il prodotto di matrici

$$\begin{pmatrix} 1 & 0 & 3 \\ -2 & 0 & 5 \\ -3 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 3 \\ 0 & 2 & 0 \end{pmatrix}$$

e determinarne il rango.

345. Dire, motivando la risposta, se i seguenti 5 vettori di $\mathbb{R}^3$:

$$v_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad v_4 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \quad v_5 = \begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix},$$

formano un insieme di generatori.

346. Per una matrice a blocchi

$$E = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix},$$

provare che $\operatorname{rg}(E) \geq \operatorname{rg}(A) + \operatorname{rg}(C)$.

347 (♡). Sia A una matrice $n \times m$, con $n, m \geq 2$, i cui coefficienti sono tutti e soli i numeri interi compresi tra 1 ed nm. Dimostrare che A ha rango maggiore o uguale a due.

348. Nello spazio vettoriale $\mathbb{K}[x]_{\leq 3}$ dei polinomi di grado minore od uguale a 3, calcolare la dimensione del sottospazio V formato dai polinomi tali che $p(0) = p(1) = p(2) = p(3)$.

349. Provare che per una matrice $A \in M_{n,n+1}(\mathbb{K})$ le seguenti condizioni sono equivalenti:

(1) A ha rango n ed una colonna nulla;
(2) tra i minori di ordine n di A, esattamente uno di loro è invertibile.

350 (♡). Sia $A = (a_{ij}) \in M_{n,n}(\mathbb{C})$ una matrice tale che per ogni indice $i = 1, \ldots, n$ si abbia

$$\sum_{j=1}^{n} |a_{ij}| < 2|a_{ii}|.$$

Provare che A è invertibile. Mostrare inoltre che per ogni $n > 0$ la matrice $B = (b_{ij}) \in M_{n,n}(\mathbb{C})$, data da $b_{ii} = n - 1$ per ogni i e $b_{ij} = -1$ per ogni $i \neq j$, non è invertibile.

351. Sia $A \in M_{n,m}(\mathbb{C})$. Provare che se il prodotto AA^T è invertibile allora A ha rango n; mostrare con un esempio che il viceversa è falso in generale.

352. Sia A una matrice $n \times n$ tale che $A^2 = I$. Dimostrare che:

(1) $A - I$ e $A + I$ non sono entrambe invertibili.
(2) $\operatorname{Ker} L_{A+I} \cap \operatorname{Ker} L_{A-I} = 0$.
(3) $Ax - x \in \operatorname{Ker} L_{A+I}$ per ogni $x \in \mathbb{K}^n$.
(4) $\operatorname{rg}(A - I) + \operatorname{rg}(A + I) = n$.

353. Dimostrare:

(1) Siano $v_1, \dots, v_n$ vettori in uno spazio vettoriale e sia $1 \leq r \leq n$ un indice tale che

$$\operatorname{Span}(v_1, \dots, v_r) = \operatorname{Span}(v_1, \dots, v_r, v_h)$$

per ogni $h = r + 1, \dots, n$. Allora

$$\operatorname{Span}(v_1, \dots, v_r) = \operatorname{Span}(v_1, \dots, v_r, v_{r+1}, \dots, v_n).$$

(2) (Teorema dell'orlare) Siano $A \in M_{n,m}(\mathbb{K})$ e B una sottomatrice di A invertibile di rango r. Allora il rango di A è uguale ad r se e solo se ogni sottomatrice quadrata di A, di ordine $(r + 1) \times (r + 1)$, contenente B come sottomatrice, non è invertibile.

354. Provare che ogni matrice di rango r può essere scritta come somma di r matrici di rango 1.

355 (☕). 1) Dati due numeri complessi a, b, determinare una formula generale per il calcolo delle potenze della matrice

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

e verificare che per $A^n - 3A^{n-1} + 3A^{n-2} - A^{n-3} = 0$ per ogni $n \geq 3$.

2) Usare il punto precedente per determinare tutte le successioni $x_0, x_1, x_2, \dots$ di numeri complessi che soddisfano, per ogni $n \geq 3$, l'uguaglianza $x_n - 3x_{n-1} + 3x_{n-2} - x_{n-3} = 0$.

356. Siano $A, B, C \in M_{4,4}(\mathbb{R})$ di rango rispettivamente $2, 4$ e 3. Quali sono i ranghi massimo e minimo possibili per la matrice $(A + B)C$.

357. Per una matrice A a coefficienti interi ed ogni campo $\mathbb{K}$ indichiamo con $\operatorname{rg}_{\mathbb{K}}(A)$ il suo rango calcolato interpretando A come matrice a coefficienti in $\mathbb{K}$ mediante l'applicazione $\mathbb{Z} \to \mathbb{K}$ che ad ogni intero n associa la somma di n volte l'elemento unità $1 \in \mathbb{K}$. Dimostrare che per ogni matrice A a coefficienti interi ed ogni sua sottomatrice B, possibilmente vuota, vale $\operatorname{rg}_{\mathbb{K}}(A) - \operatorname{rg}_{\mathbb{K}}(B) \leq \operatorname{rg}_{\mathbb{Q}}(A) - \operatorname{rg}_{\mathbb{Q}}(B)$. (Suggerimento: ricondursi al caso in cui B è ottenuta da A togliendo una riga o una colonna.)

6.5. Matrici speciali

Occupiamoci adesso di alcune classi di matrici quadrate che rivestono una particolare importanza in algebra lineare.

6.5.1. Matrici triangolari.

Definizione 6.5.1. Una matrice quadrata si dice **triangolare** (superiore) se tutti i coefficienti sotto la diagonale sono nulli, e cioè (a_{ij}) è triangolare se $a_{ij} = 0$ per ogni $i > j$.

Naturalmente si può dire che una matrice è triangolare inferiore se tutti i coefficienti sopra la diagonale sono nulli; una matrice è triangolare inferiore se e solo se la trasposta è triangolare superiore. salvo avviso contrario, per matrice triangolare si intende triangolare superiore.

Definizione 6.5.2. Una matrice quadrata si dice **strettamente triangolare** se è triangolare e se tutti i coefficienti sulla diagonale sono nulli.

Si dice **triangolare unipotente** se è triangolare e se tutti i coefficienti sulla diagonale sono uguali a 1.

Lemma 6.5.3. *Una matrice triangolare è invertibile se e solo se gli elementi sulla diagonale sono tutti diversi da* 0. *In tal caso l'inversa è ancora una matrice triangolare.*

DIMOSTRAZIONE. Sia $A = (a_{ij})$ una matrice triangolare $n \times n$. Supponiamo che A sia invertibile e denotiamo con b_{ij} i coefficienti di A^{-1}; dimostriamo per induzione su j che $b_{ij} = 0$ per ogni $i > j$ e $a_{jj}b_{jj} = 1$. Sviluppando il prodotto $A^{-1}A = I$, per $j = 1$ otteniamo

$$b_{11}a_{11} = 1, \qquad b_{i1}a_{11} = 0 \text{ per } i > 1,$$

da cui segue $a_{11} \neq 0$ e $b_{i1} = 0$ per $i > 1$. Per $i \geq j > 1$ otteniamo

$$\sum_{h=1}^{n} b_{ih}a_{hj} = \sum_{h=1}^{j-1} b_{ih}a_{hj} + b_{ij}a_{jj} + \sum_{h=j+1}^{n} b_{ih}a_{hj} = \delta_{ij}.$$

La sommatoria $\sum_{h=1}^{j-1} b_{ih}a_{hj}$ si annulla per l'ipotesi induttiva poiché $i \geq j > h$, mentre la sommatoria $\sum_{h=j+1}^{n} b_{ih}a_{hj}$ si annulla perché A è triangolare. Dunque $b_{ij}a_{jj} = \delta_{ij}$ da cui segue $a_{jj} \neq 0$ e $b_{ij} = 0$ per $i > j$.

Viceversa, supponiamo che $a_{ii} \neq 0$ per ogni i e mostriamo che le righe di A sono linearmente indipendenti, sapendo che tale condizione implica l'invertibilità della matrice. Consideriamo una combinazione lineare non banale di vettori riga

$$C = \lambda_1 A_1 + \cdots + \lambda_n A_n$$

e denotiamo con m il più piccolo indice tale che $\lambda_m \neq 0$. Si ha

$$C = (0, \ldots, 0, \lambda_m a_{mm}, *, \ldots, *)$$

e quindi $C \neq 0$. □

ESEMPIO 6.5.4. Se $U \in M_{n,n}(\mathbb{K})$ è una matrice che commuta con tutte le matrici invertibili $n \times n$ triangolari superiori, allora U è un multiplo scalare dell'identità. Per provarlo basta ripetere lo stesso ragionamento dell'Esempio 6.2.12 usando le uguaglianze

$$(I + E_{ij})U = U(I + E_{ij}), \qquad i < j\,,$$

ed il Lemma 6.5.3.

Le matrici triangolari formano un sottospazio vettoriale di $M_{n,n}(\mathbb{K})$ di dimensione uguale a $n(n+1)/2$. Infatti ogni matrice triangolare possiede n coefficienti liberi sulla prima riga, $n-1$ sulla seconda riga ecc.; basta adesso ricordare che

$$n + (n-1) + \cdots + 2 + 1 = \frac{n(n+1)}{2}\,.$$

Similmente si osserva che la matrici strettamente triangolari formano un sottospazio di dimensione

$$(n-1) + (n-2) + \cdots + 2 + 1 = \frac{n(n-1)}{2}\,.$$

6.5.2. Matrici simmetriche ed antisimmetriche.

DEFINIZIONE 6.5.5. Una matrice A si dice **simmetrica** se $A = A^T$, si dice **antisimmetrica** se $A = -A^T$.

Notiamo che una matrice simmetrica o antisimmetrica è necessariamente quadrata. Siccome una matrice quadrata ha la stessa diagonale principale della sua trasposta, ne segue che se $(a_{ij}) \in M_{n,n}(\mathbb{K})$ è antisimmetrica allora $a_{ii} = -a_{ii}$ per ogni $i = 1, \ldots, n$, ossia $2a_{ii} = 0$. Questo implica che se $\mathbb{K}$ è un campo di numeri o più in generale un campo in cui $2 \neq 0$, allora le matrici antisimmetriche sono nulle sulla diagonale principale. D'altra parte se $2 = 0$, allora $1 = -1$, ogni matrice antisimmetrica è simmetrica, e viceversa.

DEFINIZIONE 6.5.6. Una matrice si dice **alternante** se è antisimmetrica ed ha tutti i coefficienti sulla diagonale nulli.

Come abbiamo osservato nei campi di caratteristica diversa da 2, ogni matrice antisimmetrica è pure alternante.

LEMMA 6.5.7. *L'insieme $\mathcal{S}_n(\mathbb{K}) \subseteq M_{n,n}(\mathbb{K})$ delle matrici simmetriche è un sottospazio vettoriale di dimensione $n(n+1)/2$. L'insieme $\mathcal{A}_n(\mathbb{K}) \subseteq M_{n,n}(\mathbb{K})$ delle matrici alternanti è un sottospazio vettoriale di dimensione $n(n-1)/2$.*

DIMOSTRAZIONE. Ogni matrice simmetrica è univocamente determinata dai coefficienti sulla diagonale e sopra di essa, esiste quindi un isomorfismo naturale tra le lo spazio vettoriale delle matrici simmetriche e quello delle matrici triangolari. Similmente ogni matrice alternante è univocamente determinata dai coefficienti sopra la diagonale ed esiste quindi un isomorfismo naturale tra le lo spazio delle matrici alternanti e quello delle matrici strettamente triangolari. □

ESEMPIO 6.5.8. Sia $\mathbb{K}$ un campo di caratteristica diversa da 2, ossia un campo dove $1+1 \neq 0$. Allora ogni matrice quadrata si scrive in modo unico come somma di una matrice simmetrica e di una antisimmetrica.

Infatti, se A è una matrice quadrata possiamo scrivere

$$A = S + E, \qquad \text{dove } S = \frac{A + A^T}{2}, \quad E = \frac{A - A^T}{2},$$

ed è chiaro che S è simmetrica ed E antisimmetrica. Viceversa se si ha $A = C + D$ con $C = C^T$ e $D = -D^T$ allora

$$S = \frac{A + A^T}{2} = S = \frac{C + D + C - D}{2} = C, \quad E = \frac{A - A^T}{2} = \frac{C + D - C + D}{2} = D.$$

6.5.3. Matrici Hermitiane.

Sul campo dei numeri complessi, per ogni matrice A possiamo considerare la matrice coniugata $\overline{A}$ ottenuta prendendo il coniugio di tutti i coefficienti, ossia $\overline{(a_{ij})} = (\overline{a}_{ij})$. Come nel caso dei numeri complessi, il coniugio commuta con somme e prodotti di matrici:

$$\overline{A + B} = \overline{A} + \overline{B}\,, \qquad \overline{AB} = \overline{A}\,\overline{B},$$

e commuta anche con la trasposizione: $\overline{A^T} = \overline{A}^T$.

DEFINIZIONE 6.5.9. Una matrice $A = (a_{ij}) \in M_{n,n}(\mathbb{C})$ si dice **Hermitiana** se $A^T = \overline{A}$, ossia se $a_{ji} = \overline{a_{ij}}$ per ogni i, j.

Notiamo che i coefficienti sulla diagonale di una matrice Hermitiana sono numeri reali e che una matrice reale è Hermitiana se e solo se è simmetrica. Il prodotto di una matrice Hermitiana per un numero reale è ancora Hermitiana, mentre il prodotto di una matrice Hermitiana per un numero complesso può non essere Hermitiana e quindi l'insieme delle matrici Hermitiane **non è** un sottospazio vettoriale complesso, ossia sul campo $\mathbb{C}$.

Ogni matrice Hermitiana $n \times n$ è univocamente determinata da n numeri reali sulla diagonale e da $n(n-1)/2$ numeri complessi sopra la diagonale; considerando parte reale ed immaginaria di tali numeri complessi possiamo dedurre che ogni matrice Hermitiana $n \times n$ è univocamente determinata da $n + n(n-1) = n^2$ numeri reali. Dato che somma di Hermitiane è ancora Hermitiana se ne deduce che **l'insieme delle matrici Hermitiane $n \times n$ è uno spazio vettoriale sul campo dei numeri reali $\mathbb{R}$, di dimensione reale uguale a n^2.**

Per le matrici simmetriche, antisimmetriche ed Hermitiane possiamo migliorare il risultato del Corollario 6.4.5. Per semplicità espositiva tratteremo solo il caso delle matrici simmetriche ed antisimmetriche ma le stesse considerazioni valgono, mutatis mutandis, anche per le matrici Hermitiane.

Diremo che una sottomatrice A di una matrice quadrata è **principale** se la diagonale di A è contenuta nella diagonale di B o, equivalentemente, se è ottenuta eliminando righe e colonne con gli stessi indici.

TEOREMA 6.5.10. *Sia A una matrice tale che $A^T = \pm A$. Allora il rango di A è uguale al massimo intero r tale che A possiede una sottomatrice principale $r \times r$ invertibile.*

DIMOSTRAZIONE. Sia A una matrice $n \times n$ e sia $\lambda = \pm 1$ tale che $A^T = \lambda A$; se A è invertibile non c'è nulla da dimostrare. Per induzione su n basta quindi dimostrare che se A non è invertibile allora esiste un indice $i = 1, \dots, n$ tale che togliendo la riga i e la colonna i si ottiene una matrice dello stesso rango. Siccome le colonne di A sono linearmente dipendenti esiste un vettore non nullo

$$0 \neq x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \text{tale che} \quad Ax = 0\,.$$

Sia i un indice tale che $x_i \neq 0$, allora la i-esima colonna di A è combinazione lineare delle altre; se $B \in M_{n,n-1}(\mathbb{K})$ è ottenuta da A togliendo la colonna i si ha dunque $\mathrm{rg}(B) = \mathrm{rg}(A)$. Inoltre $x^T A = x^T A^T = \lambda(Ax)^T = 0$ ed a maggior ragione $x^T B = 0$. Ma anche la i-esima coordinata del vettore riga x^T è diversa da 0 e la relazione $x^T B = 0$ implica che la i-esima riga di B è combinazione lineare delle altre. Se $C \in M_{n-1,n-1}(\mathbb{K})$ è ottenuta da B togliendo la riga i si ha dunque

$$\mathrm{rg}(C) = \mathrm{rg}(B) = \mathrm{rg}(A) \ ,$$

e C è una sottomatrice principale di A. □

Esercizi.

358. Siano

$$A = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

matrici dello spazio vettoriale $\mathcal{S}_2(\mathbb{R})$ delle matrici simmetriche di ordine 2. Verificare che sono linearmente indipendenti e completarle ad una base di $\mathcal{S}_2(\mathbb{R})$.

359. Provare che il prodotto di due matrici triangolari superiori è ancora triangolare superiore.

360. Per ogni $z \in \mathbb{C} - \{0\}$ si consideri la matrice

$$L(z) = \begin{pmatrix} 1 & \log|z| \\ 0 & 1 \end{pmatrix} \in M_{2,2}(\mathbb{R})$$

e si verifichi che $L(zw) = L(z)L(w)$ per ogni coppia di numeri complessi non nulli z, w.

361. Mostrare con un esempio che il prodotto di matrici simmetriche non è simmetrico in generale.

362. Siano A, B due matrici simmetriche. Mostrare che le potenze A^n, $n > 0$, sono simmetriche e che vale $AB = BA$ se e solo se AB è simmetrica.

363. Determinare tutte le matrici strettamente triangolari A tali che $(I + A)^2 = I$. Si consiglia di fare i casi 2×2 e 3×3 prima di passare al caso generale.

364. Siano $A, B \in M_{n,n}(\mathbb{K})$ con A simmetrica e B alternante. Dimostrare che la matrice prodotto AB ha traccia nulla.

365. Mostrare che la parte reale di una matrice Hermitiana è simmetrica, mentre la parte immaginaria è antisimmetrica.

366. Dimostrare che per ogni $A \in M_{n,n}(\mathbb{K})$ la matrice AA^T è simmetrica.

367. Dimostrare che se $A, B \in M_{n,n}(\mathbb{K})$ e A è simmetrica, allora BAB^T è simmetrica.

368. Dimostrare che se $A \in M_{n,n}(\mathbb{C})$ è invertibile, allora anche la complessa coniugata $\overline{A}$ è invertibile e vale $\overline{A}^{-1} = \overline{A^{-1}}$.

369. Dimostrare che per $A \in M_{n,n}(\mathbb{R})$ vale $A\,A^T = 0$ se e solo se $A = 0$. Dedurre che se $A \in M_{n,n}(\mathbb{R})$ è simmetrica o antisimmetrica e $A^s = 0$ per qualche $s > 0$, allora $A = 0$.

370. Dimostrare che per ogni $A \in M_{n,n}(\mathbb{C})$ la matrice $A\overline{A}^T$ è Hermitiana.

371. Dimostrare che non esistono matrici antisimmetriche 3×3 invertibili.

372. Sia n un intero positivo fissato. Per ogni intero $0 \leq k \leq n$ indichiamo con T_k il sottospazio vettoriale di $M_{n,n}(\mathbb{K})$ formato dalle matrici (a_{ij}) tali che $a_{ij} = 0$ se $j - i < k$. Ad esempio T_0 è lo spazio delle matrici triangolari superiori e T_1 è lo spazio delle matrici strettamente triangolari superiori. Dimostrare:

(1) se $A \in T_a$ e $B \in T_b$, allora $AB \in T_{a+b}$;
(2) se $a > 0$ e $A \in T_a$, allora esiste $B \in T_a$ tale che $(I + B)^2 = I + A$ (suggerimento: mostrare per induzione che per ogni $k \geq 0$ esiste una matrice $B_k \in T_a$ tale che $(I + B_k)^2 - I - A \in T_{a+k}$).

373. Enunciare e dimostrare l'analogo del Teorema 6.5.10 per le matrici Hermitiane.

6.6. Complementi: attenti a chi si incontra in rete

Internet è una grande risorsa per chi studia matematica dove però è facile imbattersi in sedicenti esperti che alla prova dei fatti si rivelano palloni gonfiati. Un esempio di tronfiaggine matematica è illustrato del seguente dialogo riguardante l'Esempio 6.2.12 e trovato in un blog nel gennaio 2011. Per ovvi motivi, i nomi delle persone sono stati cambiati con nomi mitologici e riportati tutti al genere maschile.[1]

Efesto: Sia U una matrice che commuta con tutte le matrici. Dimostrare che U è un multiplo scalare dell'identità.
Dioniso: Questo è il Lemma di Schur: nella forma più semplice si può enunciare così: sia G un gruppo qualsiasi ed R una sua rappresentazione irriducibile, operante su uno spazio vettoriale V definito sul campo complesso, cioè G gruppo; $R\colon G \to GL(V)$; se T è un operatore lineare di V in sé $T\colon V \to V$ tale che $TR(g) = R(g)T$ qualunque sia $g \in G$ allora deve essere $T = hI$ con $h =$ lambda cioè, in breve, ogni operatore che commuta con tutte le matrici di una rappresentazione irriducibile, è necessariamente un multiplo dell'identità. dim: Sia v un autovettore di T con autovalore h (lambda) e sia V_h il sottospazio degli autovettori di T con autovalore h. Dunque V_h è diverso dal singleton di 0; preso un qualsiasi w appartenente a V_h, si ha $T(R(g)w) = R(g)Tw = h(R(g)w)$ dunque anche $R(g)w$ appartiene a V_h, qualunque sia g appartenente al gruppo G ma questo significa che V_h è un sottospazio invariante di R. Poiché, d'altronde R è irriducibile, ne segue che V_h deve coincidere con l'intero spazio base della rappresentazione, cioè $(T - hI) = 0$ su tutto lo spazio, cioè $T = hI$. c.v.d.
Efesto: Mi sono perso un po', poiché ci sono cose che non ho ancora fatto a lezione. Penso ci sia un modo diverso per cui sono richieste meno conoscenze avanzate per dimostrare questo quesito (insomma, è per Algebra Lineare, Algebra I ancora dobbiamo iniziarla). Non saprei. Comunque posso chiederti se sei laureato in matematica o qualcosa di simile?
Dioniso: Beh! Questo è un argomento di algebra superiore, che si utilizza essenzialmente in fisica quantistica. Non so se può essere semplificato, in quanto una matrice che commuti con tutte le altre matrici, non può essere una matrice generica, ma deve avere determinate proprietà: deve essere una matrice hermitiana irriducibile (diagonale con elementi sulla diagonale corrispondenti ad autovalori distinti ecc.). Sì, sono laureato in matematica. Ciao e in bocca al lupo per gli studi!
Apollo: ...confesso di aver sempre provato un rifiuto per l'algebra lineare, superiore ecc... sul tuo ragionamento Dioniso mi son proprio perso, sono argomenti che ho abbandonato quasi subito dopo aver fatto l'esame di Algebra senza poi mai più ritornarci ...avrei bisogno di ben più di una "spolverata"...
Dioniso: Pensa che è proprio l'esame di Algebra a differenziare un matematico da un ingenere, infatti quando mi iscrissi alla facoltà di Matematica, provenendo dalla facoltà di Ingeneria elettronica (sono per metà un ingenere), i colleghi matematici mi dicevano che io non conoscevo la matematica perché non avevo fatto l'esame di Algebra, io al momento non compresi questa affermazione, ma quando poi feci l'esame di algebra, capii, cosa volevano intendere: è vero, l'esame di Algebra ti crea un forma mentis matematica unica, con una logica affascinante ma estremamente rigida e difficile: senza quell'esame, non puoi definirti un matematico: i colleghi matematici avevano ragione!!!
Apollo: ...sono d'accordo Dioniso, pensa che Algebra è stato il mio primo esame...ma sarà che avevo una professoressa che non ce la faceva digerire in alcun modo, l'ho abbandonata da subito...
Ermete: Si può facilmente verificare: Consideriamo 2 matrici A e B (3×3) così fatte:

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad U = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

Dalla relazione $AU = UA$ si ottiene che U deve essere diagonale. Dalla relazione: $BU = UB$ si ottiene $a = e = i$. La matrice U deve essere quindi multiplo dell'identità.
Dioniso: Grazie! Hai preso una matrice A triangolare con gli elementi $a_{ij} = 0$ per i diverso da j; poi hai preso una matrice B che in pratica è un vettore riga. Cioè sono due particolari matrici. Devi generalizzare. Cmq sei arrivato alla conclusione che la matrice U che commuta tutte le altre matrici deve essere diagonale: però questa è un'ipotesi che è inclusa nel tipo di matrice da considerare, per questo ho detto che questo è un quesito di algebra superiore, infatti bisogna conoscere gli spazi vettoriali, le basi, gli autovalori, autovettori, matrici hermitiane, operatori irriducibili, nucleo (ker) di uno spazio vettoriale, polinomio caratteristico (con gli autovalori) etc.
Efesto: Però ripeto, il nostro professore ce l'ha dato per le vacanze, e noi non sappiamo cosa siano operatori irriducibili e matrici hermitiane.

[1] Con la lodevole eccezione degli interventi di Ermete, si tratta di un dialogo dai contenuti estremamente strampalati e/o incorretti: suggeriamo al lettore di non dare credito matematico a quanto riportato, soprattutto da Dioniso.

Ermete: Dioniso, sono arrivato alla conclusione che U deve essere multiplo dell'identità e non solo diagonale.
Dioniso: Si, ho visto! Ma partendo da particolari matrici. Cmq è un argomento troppo specialistico, per darlo come semplice esercizio di allenamento vacanziero!!! Ma chi è il prof, o la prof di Geometria? perché penso che stai facendo geometria e quindi stai studiando la prima parte della geometria, che è algebra lineare, e non "Algebra" che è un mondo molto più complesso (quel lineare semplifica tutto, rispetto all'Algebra dura e cruda, oserei dire).
Efesto: Il corso si chiama Algebra Lineare, Geometria la faccio al secondo anno. Il professore è Marco Manetti della Sapienza.
Dioniso: Ok. Quindi stai studiando Algebra lineare. Pertanto al massimo puoi conoscere gli autovalori e gli autovettori associati ad una matrice, la determinazione della base di uno spazio vettoriale, l'uso della matrice nella risoluzione di un sistema di equazioni, qualche operatore topologico e qualche struttura topologica (ma quelle più semplici); questo della matrice che commuta le altre, richiede conoscenze superiori. Non so perché alcuni prof strapazzino determinati argomenti rendendoli tanto semplici, dato che semplici non lo sono.

Esercizi.

374. Completare con tutti i dettagli la giusta dimostrazione di Ermete nel caso delle matrici 3×3 e generalizzarla al caso $n \times n$ considerando le matrici

$$A = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

375 (♡). Quale grande matematico inglese ha detto: "Only the mediocre are supremely confident of their ability."

376. A differenza del dialogo tra gli eroi mitologici che è del tutto reale, il seguente dialogo è di fantasia ma riporta in maniera fedele la struttura (o fallacia) logica di un sofisma utilizzato da due personaggi politici X e Y in un talk show televisivo nel novembre 2018 (povera Italia [nda]).

X: Non vorrà adesso raccontarmi la solita balla che se raddoppiamo le righe della matrice B allora la matrice AB raddoppia.
Y: Adesso le spiego: se moltiplichiamo B a destra per il doppio della matrice identità, per la proprietà associativa lo stesso accade alla matrice AB.
X: *Questo lo dice Lei!* Sta scritto chiaramente in questo libro che AB è il prodotto righe per colonne. Non serve una laurea in Matematica per capire che dipende dalle righe di A e dalle *colonne* di B. Basta bugie!, la smetta di dire che AB dipende dalle righe di B.

CAPITOLO 7

Riduzione a scala ed applicazioni

Qual è il metodo più rapido per calcolare il rango di una matrice? E per calcolare l'inversa di una matrice invertibile?

La risposta è quasi ovvia: dotarsi di computer ed installare uno dei tanti software disponibili ed in grado di rispondere alle precedenti domande (e non solo).

Tuttavia, pure il software ha bisogno di progettazione, e quindi è opportuno che gli specialisti della conoscenza siano in grado di portare avanti tali compiti anche se dotati solamente di un bastoncino ed una grande battigia sabbiosa.

L'applicazione organizzata e metodica del metodo di Gauss, brevemente accennato nella Sezione 1.4 come strumento per risolvere i sistemi lineari, fornisce un sistema concettualmente semplice e abbastanza rapido per il calcolo di ranghi, di matrici inverse ed altre quantità che scopriremo nel corso del capitolo.[1]

7.1. L'algoritmo di divisione

L'algoritmo di divisione Euclidea tra interi si estende alla divisione tra polinomi a coefficienti in un campo.

TEOREMA 7.1.1 (Divisione Euclidea tra polinomi). *Siano $p(x), q(x)$ polinomi a coefficienti in un campo $\mathbb{K}$, con $q(x) \neq 0$. Allora esistono, e sono unici, due polinomi $h(x), r(x) \in \mathbb{K}[x]$ tali che*

$$p(x) = h(x)q(x) + r(x), \qquad \deg(r(x)) < \deg(q(x)).$$

Il polinomio $r(x)$ viene chiamato il resto *della divisione.*

Notiamo che, essendo per convenzione il grado del polinomio nullo uguale a $-\infty$, il precedente teorema è vero anche se $r(x) = 0$, ossia se $q(x)$ divide $p(x)$.

DIMOSTRAZIONE. Dimostriamo prima l'unicità: se si ha

$$p(x) = h(x)q(x) + r(x) = k(x)q(x) + s(x)$$

e

$$\deg(r(x)) < \deg(q(x)), \qquad \deg(s(x)) < \deg(q(x)),$$

allora si ha $r(x) - s(x) = (k(x) - h(x))q(x)$ e siccome il grado di $r(x) - s(x)$ è strettamente minore del grado di $q(x)$ deve necessariamente essere $r(x) - s(x) = k(x) - h(x) = 0$.

Per l'esistenza daremo una dimostrazione algoritmica che illustra un metodo di calcolo di $h(x)$ e $r(x)$ ed anticipa il procedimento di riduzione a scala in un caso particolare. Siano n, m i gradi di $p(x)$ be $q(x)$ rispettivamente; per ipotesi $m \geq 0$ e si può scrivere

$$q(x) = b_0x^m + b_1x^{m-1} + \cdots + b_m, \qquad \text{con} \quad b_0 \neq 0.$$

Se se $n < m$ basta prendere $h(x) = 0$ e $r(x) = p(x)$. Se invece $n \geq m$ e

$$p(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n, \qquad \text{con} \quad a_0 \neq 0;$$

possiamo considerare il *primo resto parziale*

$$r_1(x) = p(x) - \frac{a_0}{b_0}x^{n-m}q(x) = \left(a_1 - \frac{a_0}{b_0}b_1\right)x^{n-1} + \cdots,$$

[1] Naturalmente esistono metodi più sofisticati dal punto di vista teorico ma più efficienti dal punto di vista computazionale: di tali metodi si occupa quella branca della matematica detta Algebra Lineare Numerica.

che ha la proprietà di avere grado strettamente minore di n. Se $\deg(r_1(x)) < m$ abbiamo finito, altrimenti si ripete la costruzione con $r_1(x)$ al posto di $p(x)$ e si ottiene il secondo resto parziale

$$r_1(x) = c_0x^r + c_1x^{r-1} + \cdots, \quad c_0 \neq 0,$$
$$r_2(x) = r_1(x) - \frac{c_0}{b_0}x^{r-m}q(x) = \left(c_1 - \frac{c_0}{b_0}b_1\right)x^{r-1} + \cdots .$$

Si prosegue calcolando i resti parziali $r_1(x), r_2(x), \ldots$ fino a quando si ottiene un polinomio $r_k(x)$ di grado strettamente minore di m.

Riepilogando, abbiamo trovato una successione finita di monomi $h_1(x) = \frac{a_0}{b_0}x^{n-m}, \ldots, h_k(x)$ ed una successione finita di polinomi $r_1(x), \ldots, r_k(x)$ tali che $\deg(r_k(x)) < m$ e

$$p(x) = h_1(x)q(x) + r_1(x), \quad r_1(x) = h_2(x)q(x) + r_2(x), \quad r_2(x) = h_3(x)q(x) + r_3(x), \ldots .$$

Da ciò ne consegue l'uguaglianza

$$p(x) = (h_1(x) + \cdots + h_k(x))q(x) + r_k(x)$$

e quindi l'esistenza della divisione euclidea. □

Ad esempio per $p(x) = x^5 + x^3 + x^2 - 17x + 3$ e $q(x) = x^2 + 5$ si ha $h(x) = x^3 - 4x + 1$ e $r(x) = 3x - 2$. L'esecuzione pratica (manuale) della divisione si può organizzare nel modo seguente:

$$x^5 + x^3 + x^2 - 17x + 3 : x^2 + 5\,,$$

dividendo $p(x)=$	x^5	$+x^3$	$+x^2$	$-17x$	$+3$	x^2+5	$= q(x)$ divisore
$x^3\, q(x) =$	x^5	$+5x^3$				$x^3 - 4x + 1$	$= h(x)$ quoziente
1^o resto parziale		$-4x^3$	$+x^2$	$-17x$	$+3$		
$-4x\, q(x) =$		$-4x^3$		$-20x$			
2^o resto parziale			x^2	$+3x$	$+3$		
$1\, q(x) =$			x^2		$+5$		
resto $r(x) =$				$+3x$	-2		

$$x^5 + x^3 + x^2 - 17x + 3 = (x^3 - 4x + 1)(x^2 + 5) + (3x - 2)\,.$$

Come prima applicazione della divisione Euclidea tra polinomi proviamo il risultato analogo al Lemma 2.4.2.

COROLLARIO 7.1.2. *Siano $h(x), p(x), q(x)$ polinomi. Se $h(x)$ divide il prodotto $p(x)q(x)$ allora esistono due polinomi $h_1(x), h_2(x)$ tali che: $h(x) = h_1(x)h_2(x)$, $h_1(x)$ divide $p(x)$ e $h_2(x)$ divide $q(x)$.*

DIMOSTRAZIONE. Innanzitutto osserviamo che il risultato è banalmente vero se $p(x)q(x) = 0$; infatti, se per fissare le idee $p(x) = 0$, basta prendere $h_1(x) = h(x)$ e $h_2(x) = 1$. Similmente il risultato è banalmente vero se $h(x)$ ha grado 0.

Ragioniamo per induzione sulla somma dei gradi $d = \deg(h(x)) + \deg(p(x)) + \deg(q(x))$, osservando che se $d = -\infty$ allora necessariamente $p(x)q(x) = 0$, e che se $d = 0$ allora $h(x), p(x), q(x)$ sono costanti non nulle.

Supponiamo quindi $d > 0$ ed il risultato vero qualora la somma dei gradi sia $< d$. Consideriamo separatamente i due casi $d \geq 3\deg(h(x))$ e $d < 3\deg(h(x))$.

Se $d \geq 3\deg(h(x))$ allora $\deg(p(x)) \geq \deg(h(x))$ oppure $\deg(q(x)) \geq \deg(h(x))$; i due sottocasi sono perfettamente speculari e per simmetria basta analizzare il primo. Per la divisione Euclidea possiamo trovare un polinomio $r(x)$ tale che $\deg(p(x) - h(x)r(x)) < \deg(h(x))$; siccome $h(x)$ divide il prodotto $(p(x) - h(x)r(x))q(x)$ per l'ipotesi induttiva si ha $h(x) = h_1(x)h_2(x)$, con $h_1(x)$ che divide $p(x) - h(x)r(x)$ e $h_2(x)$ che divide $q(x)$. Siccome $h_1(x)$ divide anche $h(x)r(x)$, se segue che $h_1(x)$ divide $p(x)$.

Se $d < 3\deg(h(x))$ allora il polinomio $k(x) = p(x)q(x)/h(x)$ ha grado minore del grado di $h(x)$ e divide il prodotto $p(x)q(x)$. Per l'ipotesi induttiva si ha $k(x) = k_1(x)k_2(x)$, con $k_1(x)$ che divide $p(x)$ e $k_2(x)$ che divide $q(x)$. Ma allora basta prendere $h_1(x) = p(x)/k_1(x)$ e $h_2(x) = q(x)/k_2(x)$. □

Una dimostrazione alternativa del Corollario 7.1.2 che utilizza le proprietà del massimo comune divisore sarà proposta nel Lemma 14.2.1.

Regola di Ruffini. Quando il divisore $q(x)$ è del tipo $x-b$ la divisione diventa particolarmente semplice, infatti se $p(x)=a_0x^n+\cdots+a_n$ e indichiamo $h(x)=c_0x^{n-1}+\cdots+c_{n-1}$ e $r(x)=c_n$ si hanno le uguaglianze

$$c_0=a_0,\quad a_i=c_i-bc_{i-1}\quad \text{per } i>0$$

da cui segue che i coefficienti c_i possono essere calcolati in maniera ricorsiva (regola di Ruffini)

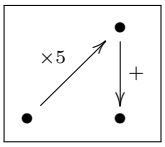
$$c_0=a_0,\quad c_1=a_1+bc_0,\quad c_2=a_2+bc_1,\quad \dots$$

Ad esempio per $p(x)=x^4-2x^3-3x^2-70x+3$ e $q(x)=x-5$ si ha $h(x)=x^3+3x^2+12x-10$ e $r(x)=-47$. L'esecuzione pratica (manuale) della divisione si può organizzare nel modo seguente:

primo passo ricorsivo

	1	-2	-3	-70	3
5					
	1				

$c_0=a_0=1$,



secondo passo ricorsivo

	1	-2	-3	-70	3
5		5			
	1	3			

$c_1=a_1+5c_0=-2+5=3$,

terzo passo

	1	-2	-3	-70	3
5		5	15		
	1	3	12		

$c_2=a_2+5c_1=-3+15=12$,

quarto passo

	1	-2	-3	-70	3
5		5	15	60	
	1	3	12	-10	

$c_3=a_3+5c_2=-70+60=-10$,

quinto passo

	1	-2	-3	-70	3
5		5	15	60	-50
	1	3	12	-10	-47

$r(x)=c_4=a_4+5c_3=3-50=-47$.

Teorema 7.1.3 (Teorema di Ruffini). *Siano $p(x)\in\mathbb{K}[x]$ e $\alpha\in\mathbb{K}$. Allora il resto della divisione di $p(x)$ per $x-\alpha$ è uguale a $p(\alpha)$; in particolare $x-\alpha$ divide $p(x)$ se e solo se α è una radice di $p(x)$.*

Dimostrazione. Per il teorema di divisione si ha $p(x)=h(x)(x-\alpha)+r(x)$; il resto $r(x)$ ha grado minore di 1 e quindi deve essere del tipo $r(x)=c$, con $c\in\mathbb{K}$. Calcolando le corrispondenti funzioni polinomiali in α si ha

$$p(\alpha)=h(\alpha)(\alpha-\alpha)+c=c$$

e quindi $r(x)=c=0$ se e solo se $p(\alpha)=0$. □

Il fatto che possiamo definire le radici di un polinomio $p(x)$ come gli elementi $a\in\mathbb{K}$ tali che $x-a$ divide $p(x)$ ci permette di introdurre facilmente il concetto di molteplicità di una radice.

Definizione 7.1.4. Siano $p(x)\in\mathbb{K}[x]$ polinomio non nullo e $a\in\mathbb{K}$. Chiameremo **molteplicità** di a come radice di $p(x)$ il massimo intero $\nu\geq 0$ tale che $(x-a)^\nu$ divide $p(x)$.

In particolare, $\nu=0$ se e solo se $p(a)\neq 0$ e $\nu>0$ se e solo se a è una radice di $p(x)$. Chiameremo a **radice semplice** di $p(x)$ se $\nu=1$, **radice multipla** se $\nu>1$.

Siccome $p(x)\neq 0$ la molteplicità ν di ogni radice è ben definita ed è minore od uguale al grado di $p(x)$, infatti $(x-a)^\nu$ può dividere $p(x)\neq 0$ solo se $\nu\leq\deg(p(x))$. Per definizione, se a è una radice di $p(x)$ di molteplicità ν allora esiste $q(x)\in\mathbb{K}[x]$ tale che $p(x)=(x-a)^\nu q(x)$; per il teorema di Ruffini $q(a)\neq 0$, altrimenti $q(x)=(x-a)r(x)$ per qualche $r(x)\in\mathbb{K}[x]$ e $p(x)=(x-a)^{\nu+1}r(x)$, in contraddizione con la definizione di molteplicità. In altri termini,

l'elemento $a \in \mathbb{K}$ *è una radice di molteplicità* ν *del polinomio non nullo* $p(x)$ *se e solo se si può scrivere* $p(x) = (x-a)^\nu q(x)$*, con* $q(a) \neq 0$*.*

LEMMA 7.1.5. *Siano* $a \in \mathbb{K}$ *e* $p_1(x), p_2(x) \in \mathbb{K}[x]$*. Allora la molteplicità di* a *come radice del prodotto* $p_1(x)p_2(x)$ *è uguale alla somma delle molteplicità di* a *come radice dei polinomi* $p_1(x)$ *e* $p_2(x)$*.*

DIMOSTRAZIONE. Siano ν_1, ν_2 le molteplicità di a come come radice dei polinomi $p_1(x)$ e $p_2(x)$, allora si può scrivere $p_1(x) = (x-a)^{\nu_1} q_1(x)$, $p_2(x) = (x-a)^{\nu_2} q_2(x)$ e dunque

$$p_1(x)p_2(x) = (x-a)^{\nu_1+\nu_2} q_1(x)q_2(x), \qquad q_1(a)q_2(a) \neq 0.$$

□

Nel prossimo corollario utilizzeremo il simbolo di produttoria $\prod$, che sta al prodotto come il simbolo di sommatoria sta alla somma.

COROLLARIO 7.1.6. *Siano* $p(x) \in \mathbb{K}[x]$ *un polinomio non nullo,* $a_1, \dots, a_n \in \mathbb{K}$ *elementi distinti e* $\beta_1, \dots, \beta_n$ *interi positivi. Allora il polinomio*

$$\prod_{i=1}^{n} (x-a_i)^{\beta_i} = (x-a_1)^{\beta_1}(x-a_2)^{\beta_2} \cdots (x-a_n)^{\beta_n}$$

divide $p(x)$ *se e solo se per ogni* i *il numero* β_i *è minore od uguale alla molteplicità di* a_i *come radice di* $p(x)$*. In particolare*

$$\prod_{i=1}^{n} (x-a_i) = (x-a_1)(x-a_2)\cdots(x-a_n)$$

divide $p(x)$ *se e solo se* $p(\alpha_i) = 0$ *per ogni* i*.*

DIMOSTRAZIONE. Denotiamo con ν_i la molteplicità di a_i come radice di $p(x)$. Se $\prod_{i=1}^{n}(x-a_i)^{\beta_i}$ divide $p(x)$, a maggior ragione $(x-a_i)^{\beta_i}$ divide $p(x)$ e quindi $\beta_i \leq \nu_i$ per definizione di molteplicità.

Viceversa, supponiamo $\beta_i \leq nu_i$ per ogni i e dimostriamo per induzione su $k = 1, \dots, n$ che $\prod_{i=1}^{k}(x-a_i)^{\beta_i}$ divide $p(x)$. Per $k = 1$ basta applicare la definizione di molteplicità. Assumiamo quindi che per qualche $k > 1$ si abbia

$$p(x) = q(x) \prod \prod_{i=1}^{k-1} (x-a_i)^{\beta_i}.$$

Siccome $a_k \neq a_i$ per ogni $i < k$ si ha $\prod_{i=1}^{k-1}(a_k - a_i)^{\beta_i} \neq 0$, quindi a_k ha molteplicità 0 come radice di $\prod_{i=1}^{k-1}(x-a_i)^{\beta_i}$ e per il Lemma 7.1.5 la molteplicità di a_k come radice di $q(x)$ è ν_k. Siccome $\beta_k \leq \nu_k$ si può dunque scrivere $q(x) = r(x)(x-a_k)_k^\beta$ e quindi

$$p(x) = r(x) \prod \prod_{i=1}^{k} (x-a_i)^{\beta_i}.$$

□

Il prossimo corollario del teorema di Ruffini è un caso particolare del teorema di fattorizzazione unica dei polinomi (Teorema 14.2.2).

COROLLARIO 7.1.7. *Sia* $p(t) \in \mathbb{K}[t]$ *un polinomio monico di grado* k *che divide un prodotto* $(t-\lambda_1)\cdots(t-\lambda_n)$ *di polinomi di primo grado. Allora* $k \leq n$ *ed esistono* k *indici* $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ *tali che*

$$p(t) = (t-\lambda_{i_1})\cdots(t-\lambda_{i_k}).$$

DIMOSTRAZIONE. Per ipotesi esiste un polinomio $h(t)$ tale che

$$p(t)h(t) = (t-\lambda_1)\cdots(t-\lambda_n),$$

e quindi $k = n - \deg(h(t)) \leq n$. Dimostriamo il corollario per induzione su n, essendo il caso $n = 1$ del tutto evidente. Se $n > 1$, siccome $p(\lambda_n)h(\lambda_n) = 0$ si ha $p(\lambda_n) = 0$ oppure $h(\lambda_n) = 0$. In entrambi i casi applichiamo il teorema di Ruffini: nel primo caso $p(t) = q(t)(t-\lambda_n)$ e quindi

$$q(t)(t-\lambda_n)h(t) = (t-\lambda_1)\cdots(t-\lambda_n),$$

da cui segue

$$q(t)h(t) = (t-\lambda_1)\cdots(t-\lambda_{n-1}),$$

e per l'ipotesi induttiva si ha

$$p(t) = q(t)(t-\lambda_n) = (t-\lambda_{i_1})\cdots(t-\lambda_{i_{k-1}})(t-\lambda_n).$$

Nel secondo caso $h(t) = r(t)(t-\lambda_n)$ e quindi

$$p(t)r(t)(t-\lambda_n) = (t-\lambda_1)\cdots(t-\lambda_n),$$

da cui segue

$$p(t)r(t) = (t-\lambda_1)\cdots(t-\lambda_{n-1}),$$

e per l'ipotesi induttiva

$$p(t) = (t-\lambda_n) = (t-\lambda_{i_1})\cdots(t-\lambda_{i_k}).$$

□

Esercizi.

377. Calcolare quoziente e resto delle seguenti divisioni tra polinomi a coefficienti razionali:

(1) $2x^8 + x^6 - x + 1 : x^3 - x^2 + 3$,
(2) $3x^5 - 2x^2 : x^2 + 1$,
(3) $x^5 - 5x : x - 2$.

378. Siano $a, b \in \mathbb{K}$ radici distinte di un polinomio $p(x) \in \mathbb{K}[x]$ di grado $n \geq 2$; per il Teorema di Ruffini si ha dunque

$$p(x) = (x-a)u(x) = (x-b)v(x), \qquad u(x), v(x) \in \mathbb{K}[x]\,.$$

Provare che il polinomio $q(x) = u(x) - v(x)$ ha grado $n-2$ e che ogni radice di $q(x)$ è anche radice di $p(x)$.

379. Trovare tutti i numeri primi p tali che $x^3 + x^2 + x + 1$ è divisible per $x^2 + 3x + 2$ come polinomio a coefficienti nel campo finito $\mathbb{F}_p$.

380. Sia $h(x)$ un polinomio di grado $d \geq 0$. Usare induzione su d ed il Corollario 7.1.2 per provare che il numero di polinomi monici distinti che dividono $h(x)$ è al più 2^d.

7.2. Matrici a scala

Una classe di matrici coinvolte nel procedimento di riduzione di Gauss, e meritevoli di essere palesate, è quella delle matrici a scala.

Definizione 7.2.1. Una matrice $n \times m$ si dice a **scala** se soddisfa la seguente proprietà: *per ogni* $0 < i < n$ *e* $0 \leq j < m$*, se la riga* i*-esima ha i primi* j *coefficienti nulli, allora la riga* $i+1$*-esima ha i primi* $j+1$ *coefficienti nulli.*

Detto in linguaggio meno rigoroso, una matrice è a scala quando soddisfa le seguenti due condizioni:

- le righe nulle sono raggruppate in fondo alla matrice;
- in due righe consecutive e non nulle, il primo coefficiente non nullo della riga superiore viene strettamente prima del primo coefficiente non nullo della riga inferiore.

Dal punto di vista grafico una matrice a scala è una matrice del tipo

(7.1)
$$\begin{pmatrix} 0 & \cdots & 0 & p_1 & ** & * & * & ** & * & * & * & \\ 0 & \cdots & 0 & 0 & \cdots & 0 & p_2 & ** & * & * & * & \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & p_3 & * & \\ & & & & & & & & & & & \ddots \end{pmatrix}$$

dove i p_i sono numeri diversi da 0 e sono detti i **perni** (o pivot) della matrice, mentre gli asterischi $*$ possono assumere qualsiasi valore. In riferimento alla Figura (7.1), la linea continua

a forma di scalinata ha tutti i gradini di altezza 1, mentre la larghezza di uno scalino può assumere qualunque valore positivo.

Notiamo che una matrice a scala possiede esattamente un perno per ogni riga diversa da 0 e che ogni colonna possiede al più un perno.

ESEMPIO 7.2.2. Ecco due matrici a scala:

$$\begin{pmatrix} \boxed{\mathbf{1}} & 2 & 3 & 4 \\ 0 & \boxed{\mathbf{5}} & 0 & 2 \\ 0 & 0 & 0 & \boxed{\mathbf{6}} \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & \boxed{\mathbf{4}} & 2 & 3 & 9 \\ 0 & 0 & 0 & \boxed{\mathbf{7}} & 0 \\ 0 & 0 & 0 & 0 & \boxed{\mathbf{8}} \end{pmatrix},$$

la prima con i tre perni $1, 5, 6$, e la seconda con i tre perni $4, 7, 8$.

ESEMPIO 7.2.3. Le matrici

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 5 & 6 & 7 \\ 0 & 0 & 8 & 9 \\ 0 & 0 & 0 & 10 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 & 3 & 4 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & -8 & 9 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

sono a scala, mentre le matrici

$$A + B = \begin{pmatrix} 2 & 2 & 6 & 8 \\ 0 & 10 & 6 & 7 \\ 0 & 0 & 0 & 18 \\ 0 & 0 & 0 & 10 \end{pmatrix}, \qquad A - B = \begin{pmatrix} 0 & 2 & 0 & 0 \\ 0 & 0 & 6 & 7 \\ 0 & 0 & 16 & 0 \\ 0 & 0 & 0 & 10 \end{pmatrix},$$

non sono a scala. Dunque le matrici a scala non formano un sottospazio vettoriale.

Ogni matrice quadrata a scala è anche triangolare, mentre l'Esempio 7.2.3 mostra che il viceversa è generalmente falso.

LEMMA 7.2.4. *Il rango di una matrice a scala è uguale al numero di righe diverse da 0. Equivalentemente, il rango di una matrice a scala è uguale al numero di perni. Inoltre, le colonne contenenti i perni sono linearmente indipendenti.*

DIMOSTRAZIONE. Siano r il rango ed s il numero di righe diverse da 0 di una matrice a scala A; chiaramente $r \leq s$. Sia B la sottomatrice $s \times s$ che contiene tutti i perni di A; la matrice B è triangolare con elementi sulla diagonale diversi da 0 ed è quindi invertibile, da cui segue $r \geq s$ per il Corollario 6.4.5. □

LEMMA 7.2.5. *Siano $A \in M_{n,m}(\mathbb{K})$ una matrice a scala e $e_1, \dots, e_m$ la base canonica di $\mathbb{K}^m$. Allora la i-esima colonna di A contiene un perno se e solo se*

$$\operatorname{Ker}(L_A) \cap \operatorname{Span}(e_1, \dots, e_{i-1}) = \operatorname{Ker}(L_A) \cap \operatorname{Span}(e_1, \dots, e_i).$$

DIMOSTRAZIONE. Indichiamo con $A(i)$ la sottomatrice formata dalle prime i colonne di A; ogni $A(i)$ è a scala e la i-esima colonna contiene un perno se e solo se $A(i)$ contiene una riga non nulla in più rispetto a $A(i-1)$. Per il Lemma 7.2.4 ciò equivale a dire che il rango di $A(i)$ è uguale al rango di $A(i-1)$ più 1, e per il teorema del rango 5.3.4 tale condizione equivale al fatto che le applicazioni lineari

$$L_{A(i)} \colon \operatorname{Span}(e_1, \dots, e_i) \to \mathbb{K}^n, \qquad L_{A(i-1)} \colon \operatorname{Span}(e_1, \dots, e_{i-1}) \to \mathbb{K}^n,$$

hanno nuclei della stessa dimensione. Basta adesso osservare che per ogni i si ha

$$\begin{aligned} \operatorname{Ker}(L_{A(i)}) &= \operatorname{Ker}(L_A) \cap \operatorname{Span}(e_1, \dots, e_i), \\ \operatorname{Ker}(L_A) \cap \operatorname{Span}(e_1, \dots, e_{i-1}) &\subseteq \operatorname{Ker}(L_A) \cap \operatorname{Span}(e_1, \dots, e_i). \end{aligned}$$

□

Supponiamo di avere un sistema lineare omogeneo di n equazioni in m incognite:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m = 0 \\ \vdots \qquad\qquad \vdots \qquad\qquad\quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m = 0 \end{cases}. \tag{7.2}$$

Possiamo associare ad esso la **matrice dei coefficienti**

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix} \in M_{n,m}(\mathbb{K})$$

e riscrivere il sistema (7.2) nella forma compatta

$$Ax = 0, \qquad x \in \mathbb{K}^m .$$

Dunque, l'insieme delle soluzioni del sistema (7.2) coincide con il nucleo dell'applicazione lineare $L_A \colon \mathbb{K}^m \to \mathbb{K}^n$ e di conseguenza formano un sottospazio vettoriale di dimensione uguale a $m - \operatorname{rg}(A)$.

Supponiamo adesso di conoscere, per qualsivoglia ragione, un insieme massimale di colonne linearmente indipendenti, e quindi anche il rango, della matrice $A \in M_{n,m}(\mathbb{K})$. Con tale informazione a disposizione possiamo calcolare facilmente una base di $\operatorname{Ker} L_A = \{x \mid Ax = 0\}$. Siano infatti $A^{i_1}, \dots, A^{i_r}$, dove r è il rango, colonne linearmente indipendenti e siano $A^{d_1}, \dots, A^{d_{m-r}}$ le colonne rimanenti. Se $V = L_A(\mathbb{K}^m) \subset \mathbb{K}^n$ è l'immagine dell'applicazione lineare L_A, allora $A^{i_1}, \dots, A^{i_r}$ sono una base di V e quindi per ogni $y \in \mathbb{K}^{m-r}$ vi è un unico vettore $x \in \mathbb{K}^r$ tale che

$$(A^{i_1}, \dots, A^{i_r})x = -(A^{d_1}, \dots, A^{d_{m-r}})y$$

ed è evidente che in questo modo, al variare di y, troviamo un isomorfismo lineare tra $\mathbb{K}^{m-r}$ e l'insieme delle soluzioni del sistema. Se facciamo variare y tra i vettori della base canonica di $\mathbb{K}^{m-r}$ troviamo una base dello spazio delle soluzioni. Vediamo alcuni esempi numerici:

Esempio 7.2.6. Calcoliamo una base dello spazio K delle soluzioni del sistema lineare

$$\begin{cases} x + 2y - z + w = 0 \\ \quad y + z - w = 0 \\ \quad\quad z + w = 0 \end{cases} .$$

La matrice dei coefficienti

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

è a scala con i perni nelle colonne 1,2 e 3, che quindi sono linearmente indipendenti. Dunque il sistema si risolve indicando w (colonna 4) come variabile indipendente e x, y, z come variabili dipendenti:

$$\begin{cases} x + 2y - z = -w \\ y + z = w \\ z = -w \end{cases}$$

che con il classico metodo di sostituzione diventa

$$x = -6w, \quad y = 2w, \quad z = -w,$$

ossia

$$K = \{(-6w, 2w, -w, w)^T \mid w \in \mathbb{K}\},$$

e ponendo $w = 1$ troviamo che il vettore colonna $(-6, 2, -1, 1)^T$ è una base di K.

Esercizi.

381. Sia $V \subset M_{3,5}(\mathbb{K})$ il più piccolo sottospazio vettoriale contenente tutte la matrici a scala. Calcolare la dimensione di V.

382. Sia $V \subset M_{5,6}(\mathbb{K})$ il più piccolo sottospazio vettoriale contenente tutte la matrici a scala di rango 3. Calcolare la dimensione di V.

383. Per ogni $0 \leq j \leq h$, indichiamo con $V_j \subseteq \mathbb{K}^h$ il sottospazio generato dai primi j vettori della base canonica. Ad ogni applicazione lineare $f\colon \mathbb{K}^m \to \mathbb{K}^n$ associamo una successione di interi non negativi

$$\alpha(f)_i = \min\{j \mid f(V_i) \subseteq V_j\}, \qquad i = 0, \dots, m.$$

Chiaramente $0 = \alpha(f)_0 \leq \alpha(f)_1 \leq \cdots \leq \alpha(f)_m \leq n$. Sia A una matrice, provare che A è a scala se e solo se per ogni $i \geq 0$ si ha

$$\alpha(L_A)_{i+1} \leq \alpha(L_A)_i + 1.$$

Dedurre che il prodotto di due matrici a scala è ancora una matrice a scala.

384. Siano $p_0(x), \dots, p_n(x) \in \mathbb{K}[x]$ polinomi tali che

$$\deg p_0(x) = 0, \qquad \deg p_i(x) \leq \deg p_{i+1}(x) \leq \deg p_i(x) + 1.$$

Sia $m = \deg p_n(x)$ e scriviamo $p_i(x) = \sum_{j=0}^{m} a_{ij} x^j$ per ogni i. Dire se la matrice (a_{ij}) è a scala e, in caso di risposta affermativa, calcolarne il rango.

7.3. Operazioni sulle righe e riduzione a scala

Riprendiamo in maniera più precisa e rigorosa le osservazioni fatte nella Sezione 1.4.

DEFINIZIONE 7.3.1. Le **operazioni elementari** sulle righe di una matrice sono:

(1) permutare le righe, ossia scambiarle di posto;
(2) moltiplicare una riga per uno scalare invertibile;
(3) sommare ad una riga un multiplo scalare di un'altra riga.

Una **operazione sulle righe** è una composizione finita di operazioni elementari sulle righe.

Adotteremo la seguente notazione per indicare le operazioni elementari sulle righe:

(1) $R_i \leftrightarrow R_j$ scambiare di posto le righe i e j;
(2) aR_i moltiplicare la riga i per uno scalare invertibile a;
(3) $R_i + aR_j$ sommare alla riga i la riga j moltiplicata per a, $(i \neq j)$.

È utile osservare che tutte le operazioni elementari sulle righe sono reversibili; l'inversa di $R_i \leftrightarrow R_j$ è $R_j \leftrightarrow R_i$; l'inversa di aR_i è $a^{-1}R_i$; l'inversa di $R_i + aR_j$ è $R_i - aR_j$.

ESEMPIO 7.3.2. Una successione di tre operazioni elementari sulle righe:

$$\begin{pmatrix} 0 & 1 \\ 2 & 2 \end{pmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{pmatrix} 2 & 2 \\ 0 & 1 \end{pmatrix} \xrightarrow{R_1 - 2R_2} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \xrightarrow{\frac{1}{2}R_1} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

e la sua successione inversa:

$$\begin{pmatrix} 0 & 1 \\ 2 & 2 \end{pmatrix} \xleftarrow{R_2 \leftrightarrow R_1} \begin{pmatrix} 2 & 2 \\ 0 & 1 \end{pmatrix} \xleftarrow{R_1 + 2R_2} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \xleftarrow{2R_1} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

La reversibilità delle operazioni elementari implica che se B si ottiene da A tramite alcune operazioni elementari sulle righe allora, prendendo il percorso inverso, anche A si ottiene da B tramite operazioni elementari sulle righe; diremo in tal caso che A e B sono **equivalenti per righe**. Lo stesso ragionamento implica che se A e B sono equivalenti per righe e se B e C sono equivalenti per righe, allora anche A e C sono equivalenti per righe.

LEMMA 7.3.3. *Siano $A, B \in M_{n,m}(\mathbb{K})$ due matrici equivalenti per righe, ossia ottenute l'una dall'altra mediante una successione finita di operazioni elementari sulle righe. Allora $\operatorname{Ker}(L_A) = \operatorname{Ker}(L_B)$; in particolare A e B hanno lo stesso rango.*

DIMOSTRAZIONE. Basta osservare che le operazioni elementari sulle righe di una matrice A lasciano invariate le soluzioni di un sistema lineare omogeneo $Ax = 0$. □

ESEMPIO 7.3.4. Ecco un esempio di operazione sulle righe formata dalla composizione di sette operazioni elementari:

$$\begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 2 & 3 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

le prime due righe sono scambiate,

$$\begin{pmatrix}1&2&3&0\\0&1&2&3\\1&1&1&1\end{pmatrix} \xrightarrow{R_3-R_1} \begin{pmatrix}1&2&3&0\\0&1&2&3\\0&-1&-2&1\end{pmatrix}$$ alla terza riga viene sottratta la prima,

$$\begin{pmatrix}1&2&3&0\\0&1&2&3\\0&-1&-2&1\end{pmatrix} \xrightarrow{R_3+R_2} \begin{pmatrix}1&2&3&0\\0&1&2&3\\0&0&0&4\end{pmatrix}$$ alla terza riga viene sommata la seconda.

Con le prime tre operazioni elementari abbiamo ottenuto una matrice a scala; con le ultime quattro (descritte a coppie) aumentiamo il numero dei coefficienti uguali a zero:

$$\begin{pmatrix}1&2&3&0\\0&1&2&3\\0&0&0&4\end{pmatrix} \xrightarrow{R_1-2R_2,\,\frac{1}{4}R_3} \begin{pmatrix}1&0&-1&-6\\0&1&2&3\\0&0&0&1\end{pmatrix}$$ riga 1 meno il doppio della 2, riga 3 moltiplicata per 1/4,

$$\begin{pmatrix}1&0&-1&-6\\0&1&2&3\\0&0&0&1\end{pmatrix} \xrightarrow{R_1+6R_3,\,R_2-3R_3} \begin{pmatrix}1&0&-1&0\\0&1&2&0\\0&0&0&1\end{pmatrix}$$ righe 1 e 2 più multipli della terza.

Grazie ai Lemmi 7.2.4 e 7.3.3, per calcolare in maniera efficiente il rango di una matrice, la si può trasformare mediante operazioni sulle righe in una matrice a scala e contare il numero di righe non nulle. La garanzia che questa ricetta funziona sempre è data dal seguente teorema.[2]

TEOREMA 7.3.5 (Eliminazione di Gauss, o Fangcheng). *Mediante una successione finita di operazioni elementari sulle righe è possibile trasformare qualsiasi matrice a coefficienti in un campo in una matrice a scala.*

Non daremo una dimostrazione formale dell'eliminazione di Gauss, ritenendo molto più utile dare una dimostrazione per esempi. In altri termini, mostreremo l'algoritmo che porta al Fangcheng in alcuni esempi numerici, in quantità più che sufficiente per illustrare completamente il procedimento e dare al lettore tutti gli strumenti per poter, se lo desidera, dimostrare il Teorema 7.3.5 in maniera rigorosa.

ESEMPIO 7.3.6. Effettuiamo l'eliminazione di Gauss della matrice

$$\begin{pmatrix}1&2&3\\4&5&6\\7&8&9\end{pmatrix}.$$

Per prima cosa annulliamo i coefficienti della prima colonna superiori al primo: nel caso in oggetto basterà sottrarre alla seconda riga 4 volte la prima ed alla terza riga 7 volte la prima

$$\begin{pmatrix}1&2&3\\4&5&6\\7&8&9\end{pmatrix} \xrightarrow{R_2-4R_1,R_3-7R_1} \begin{pmatrix}1&2&3\\0&-3&-6\\0&-6&-12\end{pmatrix}.$$

Adesso che la prima colonna è sistemata anche la prima riga è a posto e non deve essere coinvolta nelle successione operazioni elementari (esercizio: perché?). Per annullare l'ultimo coefficiente della seconda colonna sottraiamo alla terza riga il doppio della seconda

$$\begin{pmatrix}1&2&3\\0&-3&-6\\0&-6&-12\end{pmatrix} \xrightarrow{R_3-2R_2} \begin{pmatrix}1&2&3\\0&-3&-6\\0&0&0\end{pmatrix}.$$

ESEMPIO 7.3.7. Effettuiamo l'eliminazione di Gauss della matrice

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

Siccome la prima colonna non è nulla, per mettere la matrice in forma di scala occorre mettere un perno nel primo coefficiente; questo può essere fatto ad esempio scambiando la prima riga con la terza.

$$\begin{pmatrix}0&1&0\\2&1&1\\1&0&3\end{pmatrix} \xrightarrow{R_1\leftrightarrow R_3} \begin{pmatrix}1&0&3\\2&1&1\\0&1&0\end{pmatrix}.$$

[2] Attenti però a non sbagliare i conti: un corollario della legge di Murphy afferma che quando facciamo l'eliminazione di Gauss si commette un errore di calcolo nei primi tre passaggi.

Adesso, come nell'esempio precedente annulliamo i coefficienti della prima colonna superiori al primo;

$$\begin{pmatrix} 1 & 0 & 3 \\ 2 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \xrightarrow{R_2-2R_1} \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & -5 \\ 0 & 1 & 0 \end{pmatrix}.$$

Adesso che la prima colonna è sistemata occupiamoci della sottomatrice ottenuta togliendo la prima riga e la prima colonna.

$$\begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & -5 \\ 0 & 1 & 0 \end{pmatrix} \xrightarrow{R_3-R_2} \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & -5 \\ 0 & 0 & 5 \end{pmatrix} \xrightarrow{\frac{1}{5}R_3} \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & -5 \\ 0 & 0 & 1 \end{pmatrix}.$$

Adesso la matrice è a scala, ma possiamo "semplificarla" ulteriormente;

$$\begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & -5 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{R_1-3R_3,\, R_2+5R_3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

ESEMPIO 7.3.8. Per determinare il rango della matrice

$$\begin{pmatrix} 0 & 1 & 0 & 3 & 1 \\ 1 & 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 1 & 0 \\ 2 & 5 & 2 & 4 & 2 \end{pmatrix}$$

usiamo l'eliminazione di Gauss ed i Lemmi 7.2.4 e 7.3.3.

$$\begin{pmatrix} 0 & 1 & 0 & 3 & 1 \\ 1 & 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 1 & 0 \\ 2 & 5 & 2 & 4 & 2 \end{pmatrix} \xrightarrow{R_1\leftrightarrow R_2} \begin{pmatrix} 1 & 2 & 1 & 0 & 1 \\ 0 & 1 & 0 & 3 & 1 \\ 1 & 2 & 1 & 1 & 0 \\ 2 & 5 & 2 & 4 & 2 \end{pmatrix} \xrightarrow{R_3-R_1,\, R_4-2R_1} \begin{pmatrix} 1 & 2 & 1 & 0 & 1 \\ 0 & 1 & 0 & 3 & 1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 4 & 0 \end{pmatrix}$$

$$\xrightarrow{R_4-R_2} \begin{pmatrix} 1 & 2 & 1 & 0 & 1 \\ 0 & 1 & 0 & 3 & 1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix} \xrightarrow{R_4-R_3} \begin{pmatrix} 1 & 2 & 1 & 0 & 1 \\ 0 & 1 & 0 & 3 & 1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{scala!}$$

Dunque il rango è 3.

I precedenti esempi illustrano chiaramente l'algoritmo utilizzato. Sia $A \in M_{n,m}(\mathbb{K})$, se la prima colonna di A è nulla, ossia se $A = (0, B)$ con $B \in M_{n,m-1}(\mathbb{K})$, allora per ridurre A a scala é sufficiente fare lo stesso con B.

Se la prima colonna di $A = (a_{ij})$ non è nulla, allora a meno di uno scambio di righe possiamo supporre che $a_{11} \neq 0$; adesso applicando le operazioni elementari $R_i - \frac{a_{1i}}{a_{11}} R_1$ si ottiene una matrice del tipo $\begin{pmatrix} a_{11} & * \\ 0 & B \end{pmatrix}$, con $B \in M_{n-1,m-1}(\mathbb{K})$. Come prima, per ridurre A a scala é sufficiente fare lo stesso con B.

OSSERVAZIONE 7.3.9. L'eliminazione di Gauss, detta anche **riduzione a scala**, non è unica, tuttavia le colonne contenenti i perni non dipendono dal procedimento eseguito, purché questo sia corretto. Infatti se A è la riduzione a scala di una matrice B e indichiamo con $A(i)$ e $B(i)$ le sottomatrici formate dalla prime i colonne di A e B rispettivamente, allora $B(i)$ è una riduzione a scala di $A(i)$ e quindi ha lo stesso rango. Per il Lemma 7.2.5, la i-esima colonna di B contiene un perno se e solo se il rango di $A(i)$ è strettamente maggiore di quello di $A(i-1)$.

ESEMPIO 7.3.10. È possibile trovare una base dello spazio delle soluzioni di un sistema lineare omogeneo usando interamente il procedimento di riduzione a scala. Sia infatti $A \in M_{n,m}(\mathbb{K})$ la matrice dei coefficienti di un sistema lineare omogeneo, allora l'equazione $Ax = 0$ è del tutto equivalente a $x^T A^T = 0$. Consideriamo la matrice a blocchi

$$B = (A^T, I) \in M_{m,n+m}(\mathbb{K})$$

e effettuiamo su di essa operazioni sulle righe fin tanto che la sottomatrice formata dalle prime n colonne non diventi a scala. Se r è il rango di A^T, allora la matrice risultante avrà la forma

$$\begin{pmatrix} S & R \\ 0 & Q \end{pmatrix}, \quad \text{con} \quad S \in M_{r,n}(\mathbb{K}),\ R \in M_{r,m}(\mathbb{K}),\ Q \in M_{m-r,m}(\mathbb{K})$$

con la matrice S a scala e con tutte le righe diverse da 0. Dimostriamo adesso che:

$$\textit{le colonne di } Q^T \textit{ sono una base di } \operatorname{Ker}(L_A) = \{x \mid Ax = 0\}.$$

Il loro numero $m-r$ coincide con la dimensione del nucleo, inoltre B ha rango m e quindi gli $m-r$ vettori riga di Q devono essere linearmente indipendenti. Rimane da dimostrare che se $q \in \mathbb{K}^{(m)}$ è una riga di Q allora $qA^T = 0$: a tal fine basta provare $QA^T = 0$, ossia che per ogni $x \in \mathbb{K}^n$ vale $QA^T x = 0$. Sia dunque $x \in \mathbb{K}^n$ un qualsiasi vettore, allora

$$(A^T, I) \begin{pmatrix} -x \\ A^T x \end{pmatrix} = 0$$

e quindi vale anche

$$0 = \begin{pmatrix} S & R \\ 0 & Q \end{pmatrix} \begin{pmatrix} -x \\ A^T x \end{pmatrix} = \begin{pmatrix} -Sx + RA^T x \\ QA^T x \end{pmatrix}$$

da cui le uguaglianze $Sx = RA^T x$ e $QA^T x = 0$.

Esempio 7.3.11. Usiamo il procedimento descritto nell'Esempio 7.3.10 per calcolare una base dello spazio delle soluzioni del sistema lineare omogeneo

$$\begin{cases} x + 2y - z + w = 0 \\ x + y + z - 2w = 0 \end{cases} \iff (x, y, z, w) \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ -1 & 1 \\ 1 & -2 \end{pmatrix} = (0, 0).$$

Effettuiamo la riduzione a scala sulle prime due colonne:

$$\left(\begin{array}{cc|cccc} 1 & 1 & 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 & 0 \\ 1 & -2 & 0 & 0 & 0 & 1 \end{array}\right) \xrightarrow{R_2-2R_1, R_3+R_1, R_4-R_1} \left(\begin{array}{cc|cccc} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & -1 & -2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 & 1 & 0 \\ 0 & -3 & -1 & 0 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{R_3+2R_2, R_4-3R_2} \left(\begin{array}{cc|cccc} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & -1 & -2 & 1 & 0 & 0 \\ 0 & 0 & -3 & 2 & 1 & 0 \\ 0 & 0 & 5 & -3 & 0 & 1 \end{array}\right)$$

e le righe della matrice $\begin{pmatrix} -3 & 2 & 1 & 0 \\ 5 & -3 & 0 & 1 \end{pmatrix}$ sono dunque una base delle soluzioni del sistema.

Esercizi.

385. Usare la riduzione a scala per calcolare il rango delle matrici:

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 2 \\ 2 & 0 & 2 & -1 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 18 & 19 & 20 \end{pmatrix}.$$

386. Dimostrare che il polinomio $x^2 + x + 1$ divide il polinomio $ax^4 + bx^3 + cx^2 + dx + e$ se e solo se la matrice

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ a & b & c & d & e \end{pmatrix}$$

ha rango 3.

387. Calcolare, in funzione di a, b, c e d, il rango della matrice

$$\begin{pmatrix} 1 & a & 0 & 0 \\ -b & 1 & b & 0 \\ 0 & -c & 1 & c \\ 0 & -d & 1 & d \end{pmatrix}$$

388. Siano $A \in M_{n,n}(\mathbb{K})$ una matrice simmetrica e i, j due indici distinti compresi tra 1 ed n. Si agisca su A effettuando nell'ordine le seguenti operazioni:

(1) scambiare la colonna i con la colonna j;
(2) sulla matrice ottenuta dopo la prima operazione, scambiare la riga i con la riga j.

Provare che la matrice ottenuta è ancora simmetrica. Vale lo stesso per le matrici antisimmetriche?

389. Siano $A \in M_{n,n}(\mathbb{K})$ una matrice simmetrica, i, j due indici, non necessariamente distinti, compresi tra 1 ed n e $s \in \mathbb{K}$. Si agisca su A effettuando nell'ordine le seguenti operazioni:

(1) aggiungere alla colonna i la colonna j moltiplicata per s;
(2) sulla matrice ottenuta dopo la prima operazione, aggiungere alla riga i la riga j moltiplicata per s.

Provare che la matrice ottenuta è ancora simmetrica. Vale lo stesso per le matrici antisimmetriche?

390. Sia $S \in M_{n,m}(\mathbb{K})$ la matrice a blocchi

$$S = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \qquad \begin{array}{ll} A \in M_{r,r}(\mathbb{K}), & B \in M_{r,m-r}(\mathbb{K}), \\ C \in M_{n-r,r}(\mathbb{K}), & D \in M_{n-r,m-r}(\mathbb{K}), \end{array}$$

e si assuma la sottomatrice A invertibile, ossia di rango r.

Dimostrare che S ha rango $\geq r$ e che il rango di S è esattamente r se e solo se $D = CA^{-1}B$. (Suggerimento: moltiplicare a sinistra S per matrici del tipo $\begin{pmatrix} I & 0 \\ T & I \end{pmatrix}$.)

7.4. Il teorema di Rouché–Capelli

Supponiamo di avere un sistema lineare di n equazioni in m incognite

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m = b_1 \\ \vdots \qquad\quad \vdots \qquad\qquad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m = b_n \end{cases}$$

Possiamo associare ad esso la **matrice dei coefficienti**

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix} \in M_{n,m}(\mathbb{K})$$

e la **matrice completa**

$$C = \begin{pmatrix} a_{11} & \cdots & a_{1m} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nm} & b_n \end{pmatrix} \in M_{n,m+1}(\mathbb{K}).$$

Teorema 7.4.1 (Rouché–Capelli). *Un sistema lineare:*

- ☞ *ammette soluzioni se e soltanto se la matrice completa e la matrice dei coefficienti hanno lo stesso rango;*
- ☞ *ammette soluzione unica se e soltanto se la matrice completa e la matrice dei coefficienti hanno rango uguale al numero di incognite.*

Dimostrazione. Nelle notazioni precedenti, il sistema ammette soluzione se e soltanto se l'ultimo vettore colonna di C appartiene al sottospazio vettoriale generato dai vettori colonna di A e quindi se e soltanto se l'immagine di L_C è uguale all'immagine di $L_A\colon \mathbb{K}^m \to \mathbb{K}^n$. Dato che l'immagine di L_C contiene sempre l'immagine di L_A, tali immagini coincidono se e solo se sono sottospazi vettoriali della stessa dimensione.

La soluzione è unica se e solo se l'applicazione L_A è iniettiva, ossia se e solo se L_A ha rango m. □

L'Osservazione 7.3.9 ed il teorema di Rouché–Capelli ci danno un metodo pratico per decidere se un sistema lineare possiede soluzioni: basta applicare la riduzione a scala alla matrice completa e guardare se l'ultima colonna, quella dei termini noti, contiene un perno (nessuna soluzione) oppure se non contiene perni (sistema con soluzioni).

Esempio 7.4.2. La matrice completa del sistema lineare

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

è uguale a

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

e la riduzione a scala ci dà

$$\xrightarrow{R_2-R_1,R_3-R_1} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & -2 & 2 & 0 \\ 0 & 2 & -2 & 1 \end{pmatrix} \xrightarrow{R_3+R_2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & -2 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

La colonna dei termini noti contiene un perno e quindi il sistema non ammette soluzioni.

Esempio 7.4.3. La matrice completa del sistema lineare

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

è uguale a

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

e la riduzione a scala ci dà

$$\xrightarrow{R_2-R_1,R_3-R_1} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & -2 & 2 & 0 \\ 0 & 0 & -2 & 1 \end{pmatrix}.$$

La colonna dei termini noti non contiene perni e quindi il sistema ammette soluzioni.

Esempio 7.4.4. Calcoliamo per quali valori del parametro k il sistema lineare

$$\begin{cases} x+y+z=1 \\ ky+z=1 \\ (k-1)z=-1 \end{cases}$$

possiede soluzioni. La matrice completa è

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & k & 1 & 1 \\ 0 & 0 & k-1 & -1 \end{pmatrix}$$

che per $k \neq 0, 1$ è a scala senza perni nell'ultima colonna. Quindi per $k \neq 0, 1$ il sistema ammette soluzione unica. Per $k = 1$ la matrice completa

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

è ancora a scala con un perno nell'ultima colonna. Dunque per $k = 1$ il sistema non ammette soluzioni. Per $k = 0$ la matrice completa

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

non è a scala. La sua riduzione a scala è

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

e quindi per $k = 0$ il sistema possiede infinite soluzioni.

ESEMPIO 7.4.5. Vogliamo determinare per quali valori del parametro k il sistema lineare

$$\begin{cases} x + y + kz = 1 \\ 2x - y + 3z = k \\ x + 3y - z = 2 \end{cases}$$

possiede soluzioni. Eseguiamo il procedimento di riduzione a scala della matrice completa

$$\begin{pmatrix} 1 & 1 & k & 1 \\ 2 & -1 & 3 & k \\ 1 & 3 & -1 & 2 \end{pmatrix} \xrightarrow{R_2-2R_1, R_3-R_1} \begin{pmatrix} 1 & 1 & k & 1 \\ 0 & -3 & 3-2k & k-2 \\ 0 & 2 & -1-k & 1 \end{pmatrix}$$

$$\xrightarrow{R_3+\frac{2}{3}R_2} \begin{pmatrix} 1 & 1 & k & 1 \\ 0 & -3 & 3-2k & k-2 \\ 0 & 0 & \frac{3-7k}{3} & \frac{2k-1}{3} \end{pmatrix}$$

e quindi per $k \neq \dfrac{3}{7}$ l'ultima colonna non contiene perni ed il sistema ammette soluzione unica. Per $k = \dfrac{3}{7}$ la precedente matrice assume la forma

$$\begin{pmatrix} 1 & * & * & * \\ 0 & -3 & * & * \\ 0 & 0 & 0 & \frac{-1}{21} \end{pmatrix}$$

e quindi il sistema non possiede soluzioni.

Un'altra interessante conseguenza del teorema di Rouché–Capelli è che la risolubilità di un sistema lineare non dipende dalla scelta del campo.

COROLLARIO 7.4.6. *Sia*

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m = b_1 \\ \vdots \qquad\qquad \vdots \qquad\qquad\quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m = b_n \end{cases}$$

un sistema lineare a coefficienti in un campo $\mathbb{K}$. Se il sistema possiede soluzioni in un campo F contenente $\mathbb{K}$, allora possiede soluzioni anche in $\mathbb{K}$.

DIMOSTRAZIONE. Se il sistema possiede soluzioni in F, allora le matrici dei coefficienti e completa hanno lo stesso rango in F. Abbiamo però dimostrato che il rango di una matrice è invariante per estensione di campi e quindi le matrici dei coefficienti e completa hanno lo stesso rango anche su $\mathbb{K}$. □

Un sottospazio vettoriale di $\mathbb{K}^n$ può essere descritto in svariati modi, ciascuno dei quali può essere più o meno vantaggioso a seconda dei casi. Tra i modi possibili è doveroso citare le descrizione **parametrica** e la descrizione **cartesiana**.

Dal punto di vista più teorico, dare la descrizione parametrica di un sottospazio $V \subseteq \mathbb{K}^n$ significa dare una base di V, mentre dare la descrizione cartesiana significa dare un insieme finito di iperpiani $H_1, \ldots, H_s \subseteq \mathbb{K}^n$ tali che $V = H_1 \cap \cdots \cap H_s$.

Più concretamente, se V ha dimensione k, dare le equazioni parametriche di V significa trovare una matrice $A = (a_{ij}) \in M_{n,k}(\mathbb{K})$ tale che ogni $x \in V$ si scrive in modo unico come

$$x = A \begin{pmatrix} t_1 \\ \vdots \\ t_k \end{pmatrix} = \begin{pmatrix} a_{11}t_1 + a_{12}t_2 + \cdots + a_{1k}t_k \\ \vdots \\ a_{n1}t_1 + a_{n2}t_2 + \cdots + a_{nk}t_k \end{pmatrix}$$

al variare di $t_1, \dots, t_k \in \mathbb{K}$. Ciò equivale a dire che le colonne di A sono una base di V. In tal caso V coincide con l'immagine di $L_A \colon \mathbb{K}^k \to \mathbb{K}^n$.

Viceversa, dare le equazioni cartesiane significa dare una matrice $C = (c_{ij}) \in M_{n-k,n}(\mathbb{K})$ tale che V è l'insieme dei vettori $x \in \mathbb{K}^n$ che soddisfano il sistema lineare omogeneo $Cx = 0$.

Per passare dalla descrizione cartesiana a quella parametrica basta quindi risolvere un sistema lineare omogeneo con uno dei metodi descritti precedentemente e trovare una base dello spazio delle soluzioni.

Esempio 7.4.7. Troviamo l'equazione parametrica del sottospazio $W \subseteq \mathbb{K}^4$ di equazioni

$$x_1 + x_2 + x_3 + x_4 = x_1 - x_2 - x_3 + x_4 = 0.$$

Con il metodo di sostituzione ...

$$\begin{cases} x_4 = -x_1 - x_2 - x_3 \\ x_1 - x_2 - x_3 + (-x_1 - x_2 - x_3) = 0 \end{cases} \implies \begin{cases} x_4 = -x_1 - x_2 - x_3 \\ x_2 = -x_3 \end{cases}$$

$$\implies \begin{cases} x_4 = -x_1 \\ x_2 = -x_3 \end{cases}$$

... possiamo prendere $x_1 = t_1$ e $x_3 = t_2$ come variabili libere e di conseguenza $x_2 = -x_3 = -t_2$, $x_4 = -x_1 = -t_1$, ossia

$$W = \left\{ \begin{pmatrix} t_1 \\ -t_2 \\ t_2 \\ -t_1 \end{pmatrix} \;\middle|\; t_1, t_2 \in \mathbb{K} \right\}.$$

Grazie al teorema di Rouché–Capelli è altrettanto facile passare dalla descrizione parametrica a quella cartesiana. Sia infatti

$$v_1 = \begin{pmatrix} v_{11} \\ \vdots \\ v_{n1} \end{pmatrix}, \dots, v_k = \begin{pmatrix} v_{1k} \\ \vdots \\ v_{nk} \end{pmatrix}$$

una base di un sottospazio $V \subseteq \mathbb{K}^n$; allora un vettore $x = (x_1, \dots, x_n)^T$ appartiene a V se e solo se il sistema lineare

$$\begin{cases} v_{11}t_1 + v_{12}t_2 + \cdots + v_{1k}t_k = x_1 \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ v_{n1}t_1 + v_{n2}t_2 + \cdots + v_{nk}t_k = x_n \end{cases}$$

possiede soluzioni. Pensiamo tale sistema come dipendente dai parametri $x_1, \dots, x_n$ ed eseguiamo alcune operazioni sulle righe della matrice completa

$$\begin{pmatrix} v_{11} & \cdots & v_{1k} & x_1 \\ \vdots & \ddots & \vdots & \vdots \\ v_{n1} & \cdots & v_{nk} & x_n \end{pmatrix}$$

in modo tale che la sottomatrice $n \times k$ dei coefficienti sia a scala. Così facendo si ottiene una matrice in cui i primi k coefficienti delle ultime $n - k$ righe si annullano, poiché la matrice dei coefficienti ha rango esattamente k. Quindi, guardando agli ultimi $n-k$ coefficienti dell'ultima colonna si trovano esattamente $n - k$ combinazioni lineari nelle $x_1, \dots, x_n$ che devono essere annullate affinché la matrice completa abbia rango k.

Esempio 7.4.8. Troviamo l'equazione cartesiana del sottospazio di $\mathbb{K}^4$ generato dai vettori $v_1 = (1, 1, 1, 1)^T$ e $v_2 = (0, 1, 2, 3)^T$. A tal fine dobbiamo fare la riduzione a scala delle prime

due colonne della matrice (v_1, v_2, x) (per convenienza grafica omettiamo le parentesi tonde e disegniamo la linea separatrice tra la seconda e terza colonna):

$$\begin{array}{cc|c} 1 & 0 & x_1 \\ 1 & 1 & x_2 \\ 1 & 2 & x_3 \\ 1 & 3 & x_4 \end{array} \xrightarrow[R_4-R_1]{\substack{R_2-R_1 \\ R_3-R_1}} \begin{array}{cc|c} 1 & 0 & x_1 \\ 0 & 1 & x_2-x_1 \\ 0 & 2 & x_3-x_1 \\ 0 & 3 & x_4-x_1 \end{array} \xrightarrow[R_4-3R_2]{R_3-2R_2} \begin{array}{cc|c} 1 & 0 & x_1 \\ 0 & 1 & x_2-x_1 \\ 0 & 0 & x_3-2x_2+x_1 \\ 0 & 0 & x_4-3x_2+2x_1 \end{array}$$

Tale matrice non ha perni nella terza colonna se e solo se

$$x_3 - 2x_2 + x_1 = x_4 - 3x_2 + 2x_1 = 0,$$

e tali sono dunque le equazioni cartesiane di V.

Osservazione 7.4.9. Ritorneremo sul confronto parametrico vs cartesiano dopo aver fatto il principio di dualità, che se ben compreso, è un fighissimo strumento in grado di diminuire, e di molto, i conti da fare.

Esercizi.

391. Determinare, al variare dei parametri a e b, quando i sistemi

$$\begin{cases} x - 2y + 3z = 4 \\ 2x - 3y + az = 5 \\ 3x - 4y + 5z = b \end{cases}, \qquad \begin{cases} x + y + az = 1 \\ x + z = 0 \\ x + y + a^3 z = 3 \\ x + y + z = b \end{cases},$$

$$\begin{cases} (1+a)x + 2y + 2z = 1 \\ 2x + (1+a)y + 2z = 1 \\ (1-a)x - y + z = b \end{cases}, \qquad \begin{cases} (1+a)x + 2y + 2z = 1 \\ 2x + (1+a)y + 2z = b \\ x - y + az = a + 1 \end{cases},$$

possiedono soluzioni e quando sono uniche.

392. Provare che esistono numeri complessi $z_1, \ldots, z_4$ tali che

$$z_1^2 z_3 = 1, \quad z_4^2 z_2 = -1, \quad z_3^2 z_4 = i, \quad z_2^2 z_1 = -i, \quad z_1 z_4 + z_2 z_3 = 0.$$

(Sugg.: cercare soluzioni del tipo $z_i = \cos(\pi x_i) + i \sin(\pi x_i)$ con $x_1, \ldots, x_4$ numeri razionali).

393. Determinare l'equazioni parametriche dei seguenti sottospazi di $\mathbb{K}^4$:

(1) $x_1 + x_2 - x_4 = x_1 - x_2 - x_3 = 0$;
(2) $x_1 + x_2 - 3x_3 = 0$;
(3) $x_1 - x_4 = x_1 - x_2 = x_2 + x_3 + 2x_4 = 0$.

394. Determinare le equazioni cartesiane dei seguenti sottospazi di $\mathbb{K}^4$:

$$\left\{ \begin{pmatrix} t \\ t \\ t \\ t \end{pmatrix} \;\middle|\; t \in \mathbb{K} \right\}, \quad \left\{ \begin{pmatrix} t_1 + t_2 + t_3 \\ t_2 - t_3 \\ t_1 - t_2 \\ t_1 + t_3 \end{pmatrix} \;\middle|\; t_1, t_2, t_3 \in \mathbb{K} \right\}, \quad \left\{ \begin{pmatrix} t_1 + t_2 \\ t_1 - t_2 \\ t_1 - t_2 \\ t_1 + 3t_2 \end{pmatrix} \;\middle|\; t_1, t_2 \in \mathbb{K} \right\}.$$

395. In un campo di caratteristica $\neq 2, 3$, studiare in funzione del parametro t esistenza ed unicità delle soluzioni del sistema lineare

$$\begin{cases} tx + y + z = 1 \\ x + ty + z = t \\ x + y + tz = t^2 \end{cases}.$$

7.5. Riduzione di Gauss–Jordan e calcolo della matrice inversa

Le tecniche usate nel procedimento di riduzione a scala risultano utili non solo per il calcolo del rango ma anche per il calcolo esplicito dell'inversa di una matrice quadrata di rango massimo.

DEFINIZIONE 7.5.1. Una matrice a scala si dice **ridotta** se i suoi perni sono tutti uguali a 1 e se ogni colonna contenente un perno ha tutti gli altri coefficienti uguali a 0.

Ad esempio, delle seguenti tre matrici a scala, solo la prima è ridotta:

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 3 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 1 & 0 & 1 \\ 0 & 1 & 0 & 3 & 1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} -1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

LEMMA 7.5.2. *Siano $A, B \in M_{n,m}(\mathbb{K})$ matrici a scala ridotte. Se $\operatorname{Ker} L_A = \operatorname{Ker} L_B$ allora $A = B$.*

DIMOSTRAZIONE. Basta dimostrare che i coefficienti di una matrice a scala ridotta $A = (a_{ij})$ sono univocamente determinati da $\operatorname{Ker} L_A$. Sia $e_1, \dots, e_m$ la base canonica di $\mathbb{K}^m$; abbiamo già osservato che per una matrice a scala A, la colonna A^i contiene un perno se e solo se $\operatorname{Ker} L_A \cap \operatorname{Span}(e_1, \dots, e_{i-1}) = \operatorname{Ker} L_A \cap \operatorname{Span}(e_1, \dots, e_i)$.

Se $j_1 < j_2 < \cdots < j_r$ sono gli indici delle colonne che contengono i perni, essi sono univocamente determinati da $\operatorname{Ker} L_A$ e, per definizione di matrice a scala ridotta, le colonne $A^{j_1}, \dots, A^{j_r}$ sono i primi r vettori della base canonica e quindi $L_A(e_{j_i}) = e_i$ per ogni $i = 1, \dots, r$.

Sia j un indice di colonna fissato tale che $j \neq j_k$ per ogni k, e sia $d \geq 0$ il massimo intero tale che $j_d < j$ (si pone $d = 0$ se $j < j_1$). Dal fatto che A è a scala segue che $a_{ij} = 0$ per ogni $i > d$. Per mostrare che i coefficienti a_{ij}, $i = 1, \dots, d$, dipendono solo da $\operatorname{Ker} L_A$ basta osservare che per ogni $t_1, \dots, t_d \in \mathbb{K}$ si ha

$$L_A\left(e_j - \sum_{i=1}^{d} t_i e_{j_i}\right) = \sum_{i=1}^{n} a_{ij} e_i - \sum_{i=1}^{d} t_i e_i = \sum_{i=1}^{d} (a_{ij} - t_i) e_i,$$

e quindi $e_j - \sum_{i=1}^{d} t_i e_{j_i} \in \operatorname{Ker} L_A$ se e solo se $t_i = a_{ij}$ per ogni i. □

TEOREMA 7.5.3 (Riduzione di Gauss–Jordan). *Mediante una successione finita di operazioni elementari sulle righe è possibile trasformare qualsiasi matrice a coefficienti in un campo in una matrice a scala ridotta, univocamente determinata.*

Come per l'eliminazione di Gauss, la dimostrazione dell'esistenza è omessa e rimpiazzata con alcuni esempi numerici. Per quanto l'unicità della riduzione di Gauss–Jordan basta ricordare che le operazioni elementari sulle righe non modificano il nucleo della corrispondente applicazione lineare ed applicare il lemma precedente.

ESEMPIO 7.5.4. Calcoliamo le riduzioni di Gauss-Jordan delle matrici

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

La procedura è la solita: si effettuano le operazioni elementari sulle righe in modo da rendere uguali a 1 i perni e uguali a zero i coefficienti che devono essere annullati. Nel primo caso si ha

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{R_1 - R_2, R_2 - R_3} \begin{pmatrix} 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix},$$

e nel secondo

$$\begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 1 & 0 & 1 \end{pmatrix} \xrightarrow{R_3 - R_1} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 0 & -2 & 1 \end{pmatrix} \xrightarrow{R_3 + 2R_2} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 5 \end{pmatrix}$$

$$\xrightarrow{\frac{1}{5} R_3, R_1 - 2R_2} \begin{pmatrix} 1 & 0 & -4 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{R_1 + 4R_3, R_2 - 2R_3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Giova osservare che l'identità è l'unica matrice a scala ridotta quadrata di rango massimo; dunque come caso particolare della riduzione di Gauss–Jordan si ha che ogni matrice invertibile è equivalente per righe alla matrice identità.

TEOREMA 7.5.5. *Siano $A \in M_{n,n}(\mathbb{K})$ e $B, C \in M_{n,m}(\mathbb{K})$ tre matrici, e come al solito indichiamo con $I \in M_{n,n}(\mathbb{K})$ la matrice identità. Allora le due matrici a blocchi*

$$(A, C),\ (I, B) \in M_{n,n+m}(\mathbb{K})$$

sono equivalenti per righe se e solo se A è invertibile e $B = A^{-1}C$. Se ciò accade, allora (I, B) è la riduzione di Gauss–Jordan di (A, C).

DIMOSTRAZIONE. Supponiamo che (A, C) e (I, B) siano equivalenti per righe. In particolare A è equivalente per righe alla matrice identità I, in particolare la matrice A ha rango n ed è quindi invertibile. Tenendo presente che due matrici equivalenti per righe hanno lo stesso nucleo (Lemma 7.3.3) e che per ogni vettore $x \in \mathbb{K}^m$ si ha

$$(I, B)\begin{pmatrix} Bx \\ -x \end{pmatrix} = Bx - Bx = 0,$$

possiamo dedurre che

$$0 = (A, C)\begin{pmatrix} Bx \\ -x \end{pmatrix} = ABx - Cx\,.$$

Dunque per ogni vettore x si ha $ABx = Cx$ e questo equivale a dire $AB = C$.

Viceversa, supponiamo la matrice A invertibile e $AB = C$, allora la riduzione a scala ridotta di A è l'identità I e quindi la riduzione a scala ridotta (A, C) è una matrice della forma (I, D). Per quanto visto nella prima parte della dimostrazione si ha $AD = C$ e dunque $D = A^{-1}C = B$. □

ESEMPIO 7.5.6. Date le matrici

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \qquad C = \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix},$$

vogliamo mostrare che A è invertibile e calcolare il prodotto $A^{-1}C$. Per quanto visto nel Teorema 7.5.5 basta effettuare la riduzione di Gauss–Jordan della matrice

$$(A, C) = \begin{pmatrix} 1 & 2 & 5 & 6 \\ 3 & 4 & 7 & 8 \end{pmatrix}$$

che è

$$\begin{pmatrix} 1 & 2 & 5 & 6 \\ 3 & 4 & 7 & 8 \end{pmatrix} \xrightarrow{R_2 - 3R_1} \begin{pmatrix} 1 & 2 & 5 & 6 \\ 0 & -2 & -8 & -10 \end{pmatrix}$$

$$\xrightarrow{-\frac{1}{2}R_2} \begin{pmatrix} 1 & 2 & 5 & 6 \\ 0 & 1 & 4 & 5 \end{pmatrix} \xrightarrow{R_1 - 2R_2} \begin{pmatrix} 1 & 0 & -3 & -4 \\ 0 & 1 & 4 & 5 \end{pmatrix}.$$

Dunque A è invertibile e $A^{-1}C = \begin{pmatrix} -3 & -4 \\ 4 & 5 \end{pmatrix}$.

ESEMPIO 7.5.7. Consideriamo l'applicazione lineare $f\colon \mathbb{K}^2 \to \mathbb{K}^3$ definita da

$$f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x + 2y \\ 4x + 2y \\ 9x + 5y \end{pmatrix}.$$

Vogliamo determinare la matrice B che rappresenta f rispetto alle basi

$$\mathbf{v} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}, \qquad \mathbf{w} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} \right\}$$

di $\mathbb{K}^2$ ed $\mathbb{K}^3$ rispettivamente, ossia la matrice $B \in M_{3,2}(\mathbb{K})$ tale che $\mathbf{w}B = f(\mathbf{v})$, e cioè

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 3 \end{pmatrix} B = \left(f\begin{pmatrix} 1 \\ 1 \end{pmatrix}, f\begin{pmatrix} 1 \\ 2 \end{pmatrix} \right) = \begin{pmatrix} 4 & 6 \\ 6 & 8 \\ 14 & 19 \end{pmatrix}.$$

Per calcolare B possiamo applicare il Teorema 7.5.5, dal quale segue che (I, B) è uguale alla riduzione di Gauss-Jordan della matrice $(\mathbf{w}, f(\mathbf{v}))$:

$$\left(\begin{array}{ccc|cc} 1 & 0 & 1 & 4 & 6 \\ 0 & 1 & 1 & 6 & 8 \\ 1 & 1 & 3 & 14 & 19 \end{array}\right) \to \left(\begin{array}{ccc|cc} 1 & 0 & 1 & 4 & 6 \\ 0 & 1 & 1 & 6 & 8 \\ 0 & 1 & 2 & 10 & 13 \end{array}\right) \to \left(\begin{array}{ccc|cc} 1 & 0 & 1 & 4 & 6 \\ 0 & 1 & 1 & 6 & 8 \\ 0 & 0 & 1 & 4 & 5 \end{array}\right)$$
$$\to \left(\begin{array}{ccc|cc} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 & 3 \\ 0 & 0 & 1 & 4 & 5 \end{array}\right) \quad \text{e dunque} \quad B = \begin{pmatrix} 0 & 1 \\ 2 & 3 \\ 4 & 5 \end{pmatrix}.$$

COROLLARIO 7.5.8. *Siano $A, B \in M_{n,n}(\mathbb{K})$ due matrici. Se le due matrici a blocchi*

$$(A, I),\ (I, B) \in M_{n,2n}(\mathbb{K})$$

sono equivalenti per righe, allora le matrici A, B sono invertibili e sono una l'inversa dell'altra.

DIMOSTRAZIONE. Per il Teorema 7.5.5 si ha $AB = I$ e questo implica che A e B sono una l'inversa dell'altra. □

È chiaro come il Corollario 7.5.8 si presta al calcolo esplicito delle matrici inverse. Notiamo che dai conti svolti nell'Esempio 7.5.4 segue che le matrici

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

sono l'una inversa dell'altra.

ESEMPIO 7.5.9. Calcoliamo l'inversa della matrice $\begin{pmatrix} 1 & 2 \\ 3 & 7 \end{pmatrix}$ usando il Corollario 7.5.8:

$$\begin{pmatrix} 1 & 2 & 1 & 0 \\ 3 & 7 & 0 & 1 \end{pmatrix} \xrightarrow{R_2 - 3R_1} \begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & -3 & 1 \end{pmatrix} \xrightarrow{R_1 - 2R_2} \begin{pmatrix} 1 & 0 & 7 & -2 \\ 0 & 1 & -3 & 1 \end{pmatrix}$$

e dunque

$$\begin{pmatrix} 1 & 2 \\ 3 & 7 \end{pmatrix}^{-1} = \begin{pmatrix} 7 & -2 \\ -3 & 1 \end{pmatrix}.$$

PROPOSIZIONE 7.5.10. *Per due matrici $A, B \in M_{n,m}(\mathbb{K})$, sono fatti equivalenti:*

(1) *A e B sono equivalenti per righe;*
(2) *esiste una matrice $U \in M_{n,n}(\mathbb{K})$ invertibile tale che $B = UA$;*
(3) $\mathrm{Ker}(L_A) = \mathrm{Ker}(L_B)$.

DIMOSTRAZIONE. Diamo una dimostrazione che utilizza l'eliminazione di Gauss-Jordan; vedremo in seguito altre dimostrazioni più concettuali dello stesso risultato.

[1 ⇒ 2] Le stesse operazioni sulle righe che trasformano A in B, applicate alla matrice a blocchi $(I, A) \in M_{n,n+m}(\mathbb{K})$ ci danno la matrice (U, B), con $U \in M_{n,n}(\mathbb{K})$. Per il Teorema 7.5.5 la matrice U è invertibile e $UA = B$.

[2 ⇒ 3] Siccome U è invertibile si ha $Uy = 0$ se e solo se $y = 0$ e quindi, $UAx = 0$ se e solo se $Ax = 0$, ossia

$$\{x \in \mathbb{K}^m \mid Bx = UAx = 0\} = \{x \in \mathbb{K}^m \mid Ax = 0\}.$$

[3 ⇒ 1] Per il Lemma 7.5.2 le matrici A, B sono equivalenti per righe alla stessa matrice a scala ridotta. □

Esercizi.

396. Usando l'eliminazione di Gauss–Jordan, determinare se le seguenti matrici sono invertibili e, nel caso, determinarne l'inversa.

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 1 & -1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{pmatrix},$$
$$\begin{pmatrix} -2 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

397. Dire per quali valori di $a \in \mathbb{R}$ la matrice

$$\begin{pmatrix} a & 1 & 0 \\ 1 & 1 & a \\ 0 & a & 0 \end{pmatrix}$$

è invertibile e calcolarne l'inversa.

398. Risolvere le equazioni matriciali

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 1 & 1 \end{pmatrix} X = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 2 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 0 & 4 & 7 \end{pmatrix} Y = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 4 \\ 1 & 3 & 7 \end{pmatrix}.$$

399. Sia $A \in M_{n,m}(\mathbb{K})$ Descrivere esplicitamente le matrici invertibili U tali che la matrice UA è ottenuta da A mediante un'operazione elementare per righe (tali matrici esistono per la Proposizione 7.5.10).

400 (♡). Siano $A, B \in M_{n,n}(\mathbb{K})$. Dimostrare che le tre matrici a blocchi

$$\begin{pmatrix} I & B \\ 0 & I \end{pmatrix}, \begin{pmatrix} I & A \\ 0 & B \end{pmatrix}, \begin{pmatrix} I & A \\ B & 0 \end{pmatrix} \in M_{2n,2n}(\mathbb{K})$$

hanno rango rispettivamente uguale a $2n$, $n + \operatorname{rg}(B)$ e $n + \operatorname{rg}(BA)$.

401 (☕). Mostrare che se due sistemi lineari nelle stesse incognite sono risolubili ed hanno le stesse soluzioni, allora anche i sistemi omogenei associati hanno le stesse soluzioni. Usare questo fatto e la Proposizione 7.5.10 per dimostrare quanto è stato affermato nell'Osservazione 1.4.2.

7.6. Prodotto scalare e proiezioni ortogonali

In questa sezione tratteremo esclusivamente spazi vettoriali sul campo dei numeri reali, usando in più occasioni il fatto che per una successione $a_1, \dots, a_n \in \mathbb{R}$ vale $a_1^2 + \cdots + a_n^2 = 0$ se e soltanto se $a_1 = \cdots = a_n = 0$. Dati due punti nel piano reale

$$p = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}, \ q = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} \in \mathbb{R}^2,$$

la loro distanza (lunghezza del segmento che li unisce) è data dalla formula Pitagorica

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2}.$$

Interpretando i punti del piano come vettori applicati nell'origine, e definendo la norma di un vettore v come

$$v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \qquad \|v\| = \sqrt{v_1^2 + v_2^2},$$

è immediato osservare che la distanza è uguale alla norma della differenza, ossia

$$d(p,q) = \|p - q\|.$$

Risulta anche utile notare che, pensando ogni $v \in \mathbb{R}^2$ come una matrice reale 2×1, si ha $\|v\|^2 = v^T v$. Più in generale, dati due vettori $v, w \in \mathbb{R}^2$, il loro prodotto scalare, come definito in fisica, coincide con il prodotto righe per colonne

$$v^T w = v_1 w_1 + v_2 w_2 = w^T v.$$

È facile dimostrare che si ha $v^T w = 0$ se e solo se i vettori v, w sono perpendicolari: infatti v è perpendicolare a w se e solo se è equidistante da w e $-w$, ossia se e solo se $\|v - w\| = \|v - (-w)\|$. Per concludere basta osservare che $v^T w = w^T v$ e quindi

$$\|v + w\|^2 - \|v - w\|^2 = (v + w)^T (v + w) - (v - w)^T (v - w) = 4 v^T w.$$

Le precedenti nozioni si estendono immediatamente agli spazi $\mathbb{R}^n$ in maniera indolore.

DEFINIZIONE 7.6.1. Il **prodotto scalare** di due vettori $v, w \in \mathbb{R}^n$ è definito dalla formula

$$v \cdot w = v^T w = \sum_{i=1}^{n} v_i w_i \,.$$

La **norma** di un vettore $v \in \mathbb{R}^n$ è

$$\|v\| = \sqrt{v \cdot v} = \sqrt{v^T v} = \sqrt{v_1^2 + \cdots + v_n^2} \,.$$

Notiamo che il prodotto scalare è commutativo, ossia $v \cdot w = w \cdot v$ e che $\|v + w\|^2 = \|v\|^2 + 2v \cdot w + \|w\|^2$ (esercizio: perché?). Diremo che i vettori v e w sono **ortogonali** o **perpendicolari**, e scriveremo $v \perp w$, se $v \cdot w = 0$ (Figura 7.1).

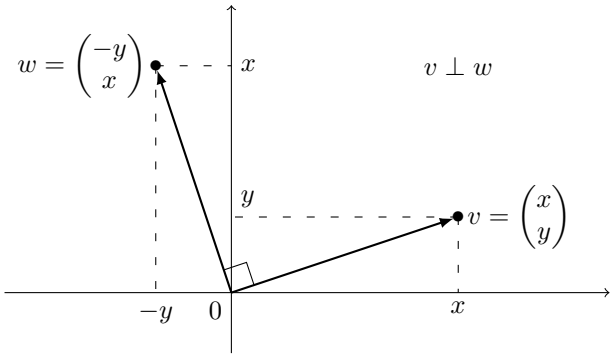


FIGURA 7.1

LEMMA 7.6.2. *Per un vettore $v \in \mathbb{R}^n$ le seguenti condizioni sono equivalenti:*

(1) $\|v\| = 0$,
(2) $v = 0$,
(3) $v \cdot w = 0$ *per ogni* $w \in \mathbb{R}^n$.

DIMOSTRAZIONE. Siano $v_1, \ldots, v_n$ le coordinate di v. Se $\|v\| = 0$ allora $\sum v_i^2 = 0$ e siccome $v_i \in \mathbb{R}$ per ogni i ne segue $v_i = 0$ per ogni i. Se $v = 0$ è chiaro che $v \cdot w = 0$ per ogni $w \in \mathbb{R}^n$. Se $v \cdot w = 0$ per ogni $w \in \mathbb{R}^n$, in particolare $v \cdot v = 0$ e quindi $\|v\| = \sqrt{v \cdot v} = 0$. □

TEOREMA 7.6.3. *Sia $A \in M_{n,m}(\mathbb{R})$, allora:*

(1) *il rango della matrice simmetrica $A^T A$ è uguale al rango di A;*
(2) *per ogni vettore $b \in \mathbb{R}^n$ l'equazione*

$$A^T A x = A^T b, \qquad x \in \mathbb{R}^m, \tag{7.3}$$

è risolubile e le sue soluzioni sono tutti e soli i punti di minimo assoluto della funzione

$$f \colon \mathbb{R}^m \to \mathbb{R}, \qquad f(x) = \|Ax - b\|^2.$$

DIMOSTRAZIONE. 1) Considerando i due sottospazi vettoriali

$$V = \operatorname{Ker} L_A = \{x \in \mathbb{R}^m \mid Ax = 0\}, \qquad W = \operatorname{Ker} L_{A^T A} = \{x \in \mathbb{R}^m \mid A^T A x = 0\},$$

per il teorema del rango basta dimostrare che V e W hanno la stessa dimensione. È chiaro che $V \subseteq W$, in quanto se $Ax = 0$ a maggior ragione vale $A^T A x = A^T 0 = 0$. Viceversa, se $x \in W$ si ha $\|Ax\|^2 = (Ax)^T (Ax) = x^T A^T A x = x^T 0 = 0$ e quindi $Ax = 0$ e $W \subseteq V$. Abbiamo dimostrato che $V = W$ e di conseguenza che A e $A^T A$ hanno lo stesso rango.

2) Per il punto precedente le matrici A e $A^T A$ hanno lo stesso rango e quindi anche le matrici A^T e $A^T A$ hanno lo stesso rango.

Consideriamo le due applicazioni lineari:

$$L_{A^T} \colon \mathbb{R}^n \to \mathbb{R}^m, \qquad L_{A^T A} \colon \mathbb{R}^m \to \mathbb{R}^m \,.$$

Siccome $L_{A^T A} = L_{A^T} \circ L_A$, l'immagine di $L_{A^T A}$ è contenuta nell'immagine di L_{A^T}. Siccome A^T e $A^T A$ hanno lo stesso rango le applicazioni $L_{A^T A}$ e L_{A^T} hanno la stessa immagine.

In particolare, per ogni $b \in \mathbb{R}^m$ il vettore $A^T b$ appartiene all'immagine di L_{A^TA} e quindi l'equazione $A^TAx = A^Tb$ è risolubile.

Per concludere resta da provare che se $A^TAx = A^Tb$, allora per ogni $y \in \mathbb{R}^m$ si ha $\|Ay - b\|^2 \geq \|Ax - b\|^2$. Siccome

$$(Ay - Ax)^T(Ax - b) = (y - x)^T A^T(Ax - b) = (y - x)^T(A^TAx - A^Tb) = 0$$

si ha:

$$\begin{aligned}\|Ay - b\|^2 &= \|(Ay - Ax) + Ax - b\|^2 \\ &= \|Ay - Ax\|^2 + 2(Ay - Ax)^T(Ax - b) + \|Ax - b\|^2 \\ &= \|Ay - Ax\|^2 + \|Ax - b\|^2 \\ &\geq \|Ax - b\|^2.\end{aligned}$$

Per uso futuro osserviamo che per il Lemma 7.6.2 l'equazione $A^TAx = A^Tb$ è soddisfatta se e solo se per ogni $v \in \mathbb{R}^m$ si ha

$$0 = v^T(A^TAx - A^Tb) = (Av)^T(Ax - b).$$

□

COROLLARIO 7.6.4. *Sia $H \subseteq \mathbb{R}^n$ un sottospazio vettoriale. Per ogni $v \in \mathbb{R}^n$ esiste un unico vettore $p(v) \in H$ tale che $(v - p(v)) \perp u$ per ogni $u \in H$. Inoltre, il punto $p(v)$ risulta essere l'unico punto di minimo della funzione*

$$f\colon H \to \mathbb{R}, \qquad f(u) = \|v - u\|^2.$$

L'applicazione

$$p\colon \mathbb{R}^n \to \mathbb{R}^n, \qquad v \mapsto p(v),$$

è lineare e viene detta ***proiezione ortogonale*** *su H.*

DIMOSTRAZIONE. Sia $m \leq n$ la dimensione di H e sia $A \in M_{n,m}(\mathbb{R})$ una matrice i cui vettori colonna formano una base di H; equivalentemente A è una matrice tale che l'applicazione lineare $L_A\colon \mathbb{R}^m \to \mathbb{R}^n$ induce una bigezione tra $\mathbb{R}^m$ ed H. Dunque i vettori di H sono tutti e soli i vettori del tipo Ax al variare di $x \in \mathbb{R}^m$, e bisogna dimostrare che per ogni $v \in \mathbb{R}^n$ esiste un unico $x \in \mathbb{R}^m$ tale che $(v - Ax) \perp Ay$ per ogni $y \in \mathbb{R}^m$. Equivalentemente, per quanto osservato al termine della dimostrazione del Teorema 7.6.3 bisogna provare che esiste un unico $x \in \mathbb{R}^m$ tale che $A^TAx = A^Tv$. Siccome la matrice A ha rango m, anche il rango di $A^TA \in M_{m,m}(\mathbb{R})$ è uguale a m; dunque A^TA è invertibile e l'unica soluzione dell'equazione $A^TAx = A^Tv$ è $x = (A^TA)^{-1}A^Tv \in \mathbb{R}^m$. Ne consegue che p è ben definita e lineare. Abbiamo inoltre provato che vale la formula

$$p(v) = A(A^TA)^{-1}A^Tv, \qquad \text{ossia} \quad p = L_{A(A^TA)^{-1}A^T}.$$

□

Sia $u_1, \ldots, u_m$ un insieme di generatori di un sottospazio vettoriale $H \subseteq \mathbb{R}^n$. Condizione necessaria e sufficiente affinché un vettore $p(v) \in H$ sia la proiezione ortogonale di un vettore v è che $u_i \cdot v = u_i \cdot p(v)$ per ogni $i = 1, \ldots, k$. La necessità è chiara; viceversa se $u_i \cdot v = u_i \cdot p(v)$ per ogni i, allora per ogni $u \in H$ esistono $t_1, \ldots, t_k \in \mathbb{R}$ tali che $u = t_1u_1 + \cdots + t_ku_k$ e di conseguenza

$$u^T(v - p(v)) = (\sum t_iu_i)^T(v - p(v)) = \sum t_iu_i^T(v - p(v)) = 0.$$

Osserviamo che la proiezione ortogonale su un sottospazio $H \subseteq \mathbb{R}^n$ è l'unica applicazione lineare $p\colon \mathbb{R}^n \to \mathbb{R}^n$ tale che

$$p(\mathbb{R}^n) = H, \qquad p(u) \perp (v - p(v)) \text{ per ogni } u, v \in \mathbb{R}^n. \tag{7.4}$$

Infatti, se $p\colon \mathbb{R}^n \to \mathbb{R}^n$ soddisfa (7.4), allora per ogni $v \in \mathbb{R}^n$ ed ogni $w \in H$ esiste $u \in \mathbb{R}^n$ tale che $p(u) = w$ e quindi $w \cdot (v - p(v)) = p(u) \cdot (v - p(v)) = 0$. Si deduce che, se $A \in M_{n,n}(\mathbb{R})$, allora L_A è la proiezione ortogonale su H se e solo se le colonne di A generano H e $A = A^TA$. Infatti, la prima condizione equivale a dire che $L_A(\mathbb{R}^n) = H$, mentre la seconda equivale a dire che

$$(u - L_A(u)) \cdot L_A(v) = u^T(A - A^TA)v = 0$$

per ogni $u, v \in \mathbb{R}^n$.

COROLLARIO 7.6.5 (Formule di regressione lineare). *1) Per ogni matrice $A \in M_{n,m}(\mathbb{R})$ ed ogni vettore $b \in \mathbb{R}^n$ vale l'uguaglianza*

$$A^T Ax = A^T b, \qquad x \in \mathbb{R}^m,$$

se e solo se Ax è la proiezione ortogonale di b sul sottospazio vettoriale generato dalle colonne della matrice A.

2) Per ogni matrice $B \in M_{k,n}(\mathbb{R})$ ed ogni vettore $c \in \mathbb{R}^n$ vale l'uguaglianza

$$BB^T x = Bc, \qquad x \in \mathbb{R}^k,$$

se e solo se $c - B^T x$ è la proiezione ortogonale di c sul nucleo di L_B.

DIMOSTRAZIONE. 1) Abbiamo già visto nella dimostrazione del Teorema 7.6.3 che $A^T Ax = A^T b$ se e solo se se per ogni $v \in \mathbb{R}^m$ si ha

$$0 = v^T(A^T Ax - A^T b) = (Av)^T(Ax - b)\,,$$

ossia se e solo se Ax è la proiezione ortogonale di b sul sottospazio $H = \{Av \mid v \in \mathbb{R}^m\}$.

2) Il Teorema 7.6.3 applicato alla matrice $A = B^T$ ci assicura che l'equazione $BB^T x = Bc$ è risolubile. Se $BB^T x = Bc$, allora $B(c - B^T x) = 0$ e $c - B^T x$ appartiene al nucleo di L_B; inoltre per ogni $y \in \operatorname{Ker} L_B$ si ha

$$y^T(c - (c - B^T x)) = y^T(B^T x) = (By)^T x = 0$$

e questo significa che $c - B^T x$ è la proiezione ortogonale di c su $\operatorname{Ker} L_B$. □

Notiamo incidentalmente che le formule di regressione lineare forniscono un metodo costruttivo per calcolare le proiezioni ortogonali su un sottospazio $H \subseteq \mathbb{R}^n$. La prima formula per quando il sottospazio è definito in forma parametrica, le seconda quando sottospazio è definito in forma cartesiana. Nei prossimi esempi vediamo più in dettaglio le regole di calcolo da seguire.

Avere un sottospazio H descritto in forma parametrica significa avere dato una base $u_1, \dots, u_m \in Hs$; denotiamo con $a_{ij} = u_i \cdot u_j$.

La proiezione ortogonale di $v \in \mathbb{R}^n$ su H è allora l'unico vettore $p(v) \in H$ tale che $u_i \cdot v = u_i \cdot p(v)$ per ogni $i = 1, \dots, m$. Scrivendo $p(v) = x_1 u_1 + \cdots + x_m u_m$, i coefficienti $x_1, \dots, x_m$ si calcolano risolvendo le m equazioni

$$u_i \cdot p(v) = a_{i1}x_1 + \cdots + a_{im}x_m = u_i \cdot v, \qquad i = 1, \dots, m\,.$$

Se indichiamo con $U \in M_{n,m}(\mathbb{R})$ la matrice le cui colonne sono i vettori $u_1, \dots, u_m$, allora $U^T U = (a_{ij})$ e le precedenti equazioni si possono scrivere nella forma compatta

$$U^T Ux = U^T v.$$

Se invece H è descritto in forma cartesiana

$$H = \{u \in \mathbb{R}^n \mid Cu = 0\}$$

per una opportuna matrice $C \in M_{n-m,n}(\mathbb{R})$, per calcolare $p(v)$ basta trovare un vettore $y \in \mathbb{R}^{n-m}$ tale che $CC^T y = Cv$ e porre $p(v) = v - C^T y$. Infatti, per ogni $u \in H$ ed ogni $y \in \mathbb{R}^{n-m}$ si ha $(C^T y) \cdot u = y^T Cu = 0$ e se $CC^T y = Cv$ e $p(v) = v - C^T y$, allora

$$Cp(v) = C(v - C^T y) = 0, \qquad p(v) \cdot u = (C^T y) \cdot u = 0, \quad \text{per ogni } u \in H.$$

ESEMPIO 7.6.6. Calcoliamo la proiezione ortogonale del vettore $v = (1,1,1)^T \in \mathbb{R}^3$ sul sottospazio generato dai vettori $u_1 = (2,-1,0)^T$ e $u_2 = (3,0,-1)^T$.

Abbiamo visto che $p(v)$ è determinato dalle equazioni

$$(v - p(v)) \cdot u_1 = (v - p(v)) \cdot u_2 = 0;$$

scrivendo $p(v)^T = a(2,-1,0) + b(3,0,-1)$, i coefficienti a, b si calcolano risolvendo il sistema

$$\begin{cases} (a(2,-1,0) + b(3,0,-1)) \cdot (2,-1,0)^T = (1,1,1) \cdot (2,-1,0)^T \\ \\ (a(2,-1,0) + b(3,0,-1)) \cdot (3,0,-1)^T = (1,1,1) \cdot (3,0,-1)^T, \end{cases}$$

e quindi

$$\begin{cases} 5a + 6b = 1 \\ 6a + 10b = 2 \end{cases} \qquad \Longrightarrow \quad a = -\frac{1}{7},\ b = \frac{2}{7}, \quad p(v) = \left(\frac{4}{7}, \frac{1}{7}, \frac{-2}{7}\right)^T.$$

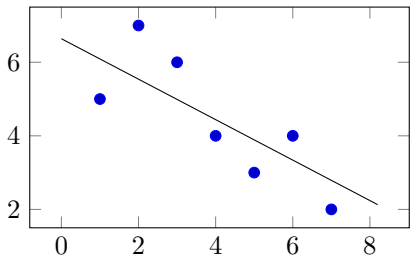


FIGURA 7.2. Diagramma di dispersione (scatterplot) e retta di regressione della successione $(1,5),(2,7),(3,6),(4,4),(5,3),(6,4),(7,2)$.

ESEMPIO 7.6.7. Calcoliamo la proiezione ortogonale del vettore $v = (1,1,1)^T \in \mathbb{R}^3$ sul sottospazio di equazione $x_1 + 2x_2 + 3x_3 = 0$. Scrivendo l'equazione del sottospazio come $Cx = 0$, dove $C = (1,2,3) \in M_{1,3}(\mathbb{R})$, abbiamo visto che $p(v) = v - C^T y$, dove $CC^T y = Cv$. Siccome $CC^T = 14$ e $Cv = 6$ si ha $y = 3/7$ e quindi

$$p(v) = v - \frac{3}{7}C^T = \left(\frac{4}{7}, \frac{1}{7}, \frac{-2}{7}\right)^T.$$

Tuttavia l'applicazione primaria delle formule di regressione è legata al calcolo della retta di regressione di un diagramma di dispersione bidimensionale. Supponiamo di avere una successione finita $(x_1,y_1),\dots,(x_n,y_n)$, di punti del piano, possibilmente ripetuti ma non tutti uguali tra loro. Se $n \geq 3$, in generale non esiste alcuna retta che li contiene tutti e può essere interessante, soprattutto in ambito applicativo, trovare una retta che comunque descrive in maniera approssimata l'eventuale relazione intercorrente tra le quantità x_i e le quantità y_i (Figura 7.2).

DEFINIZIONE 7.6.8. Per **retta di regressione** di una successione $(x_1,y_1),\dots,(x_n,y_n)$ di coppie di numeri reali, con gli x_i non tutti uguali tra loro, si intende la retta di equazione $y = ax + b$, con i coefficienti a, b che minimizzano la sommatoria

$$\sum_{i=1}^{n}(ax_i + b - y_i)^2.$$

Per quanto visto nel Teorema 7.6.3, trovare la retta di regressione di una successione $(x_1,y_1),\dots,(x_n,y_n)$ equivale a trovare la proiezione ortogonale del vettore $y = (y_1,\dots,y_n)^T \in \mathbb{R}^n$ sul sottospazio vettoriale generato dai vettori $x = (x_1,\dots,x_n)^T$ ed $e = (1,\dots,1)^T$. In particolare, poiché gli x_i non sono uguali tra loro, i due vettori x, e sono linearmente indipendenti e la retta di regressione esiste ed è unica.

Il calcolo di a, b si riconduce al sistema di due equazioni nelle incognite a, b:

$$(ax + be - y) \cdot x = (ax + be - y) \cdot e = 0$$

che sviluppato in coordinate diventa il sistema, detto delle **equazioni normali**:

$$a\sum x_i^2 + b\sum x_i = \sum x_i y_i, \qquad a\sum x_i + nb = \sum_i y_i. \tag{7.5}$$

OSSERVAZIONE 7.6.9. La definizione della retta di regressione non è simmetrica in x e y, nel senso che in generale, la retta di equazione $x = cy + d$ che minimizza la sommatoria $\sum_{i=1}^{n}(cy_i+d-x_i)^2$, non coincide con la retta di regressione $y = ax+b$ ed è possibile dimostrare che i punti del piano $(x_1,y_1),\dots,(x_n,y_n)$ sono allineati se e solo se $ac = 1$, vedi Esercizio 415.

Esercizi.

402. Per quali valori di $t \in \mathbb{R}$ i vettori $v = (-5,t,2-t)^T$ e $(t,t,4)^T$ sono ortogonali?

403. Determinare le proiezioni ortogonali del vettore $(1,1,1,1)^T \in \mathbb{R}^4$ sugli iperpiani H, K di equazioni $x_1 + x_2 + x_3 + x_4 = 0$, $x_1 - x_2 - x_3 + x_4 = 0$ e sulla loro intersezione.

404. Provare che una matrice $S \in M_{n,n}(\mathbb{R})$ è simmetrica se e solo se $Sx \cdot y = x \cdot Sy$ per ogni $x, y \in \mathbb{R}^n$.

405. Provare che una matrice $P \in M_{n,n}(\mathbb{R})$ rappresenta la proiezione ortogonale su un sottospazio se e soltanto se $P = P^T = P^2$. (Suggerimento: se $P = P^T = P^2$ allora per ogni $x, y \in \mathbb{R}^n$ si ha $(x - Px) \cdot Py = P(x - Px) \cdot y = 0$.)

406. Provare che per una matrice $A \in M_{n,n}(\mathbb{R})$ le seguenti condizioni sono equivalenti:

(1) $\|Ax\| = \|x\|$ per ogni $x \in \mathbb{R}^n$,
(2) $Ax \cdot Ay = x \cdot y$ per ogni $x, y \in \mathbb{R}^n$,
(3) $A^T A = I$,
(4) i vettori colonna di A hanno norma uno e sono ortogonali due a due.

Una matrice con tali caratteristiche viene detta **matrice ortogonale**.

407. Dimostrare la **disuguaglianza di Cauchy–Schwarz**: per ogni coppia di vettori $v, w \in \mathbb{R}^n$ vale

$$|v \cdot w| \leq \|v\| \, \|w\|,$$

e vale $|v \cdot w| = \|v\| \, \|w\|$ se e solo se v, w sono linearmente dipendenti. (Sugg.: se v, w sono linearmente indipendenti allora i due vettori v, $r = w\|v\|^2 - (v \cdot w)v$ sono diversi da 0 ed in particolare $\|r\|^2, \|v\|^2 > 0$.)

408 (♡). Dire se esiste una matrice $A \in M_{4,4}(\mathbb{R})$ take che

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

409. Abbiamo visto che ogni matrice reale B ha lo stesso rango di $B^T B$. Trovare una matrice 2×2 a coefficienti complessi per cui la precedente proprietà non vale.

410. Data una matrice $A \in M_{n,m}(\mathbb{R})$ di rango m, interpretare geometricamente l'applicazione lineare associata alla matrice $A(A^T A)^{-1} A^T$.

411. Dire se la matrice

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

rappresenta la proiezione ortogonale di $\mathbb{R}^4$ su un sottospazio vettoriale.

412. Determinare la retta di regressione della serie $(0, 0), (1, 1), (2, 3), (3, 2)$.

413. Scrivere tre coppie $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ di numeri reali tali che esistono infiniti valori di (a, b) che minimizzano la sommatoria

$$\sum_{i=1}^{3} |ax_i + b - y_i|.$$

414. Mostrare che la soluzione delle equazioni normali (7.5) è

$$a = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2}, \qquad b = \overline{y} - a\overline{x},$$

dove $\overline{x}$ e $\overline{y}$ denotano le medie aritmetiche di $x_1, \ldots, x_n$ e $y_1, \ldots, y_n$ rispettivamente.

415 (☕). Date due successioni finite e non costanti di numeri reali $x_1, \ldots, x_n$ e $y_1, \ldots, y_n$, già sappiamo che il sistema di quattro equazioni lineari

$$a \sum x_i^2 + b \sum x_i = \sum x_i y_i, \qquad a \sum x_i + nb = \sum_i y_i,$$

$$c \sum y_i^2 + d \sum y_i = \sum x_i y_i, \qquad c \sum y_i + nd = \sum_i x_i,$$

nelle incognite a, b, c, d possiede soluzione unica. Dimostrare che $0 \leq ac \leq 1$ e vale $ac = 1$ se e solo se gli n punti del piano $(x_1, y_1), \ldots, (x_n, y_n)$ sono allineati. (Suggerimento: mostrare che non è restrittivo assumere $\sum x_i = \sum y_i = 0$.)

7.7. Complementi: matrici a coefficienti interi e riduzione di Smith

Supponiamo di avere un sistema di equazioni lineari a coefficienti interi del quale ci interessa trovare, se esistono, le soluzioni anch'esse intere. A tal fine possiamo agire con delle operazioni elementari sulle righe, che per l'occasione dovranno essere:

(1) permutare le righe, ossia scambiarle di posto,
(2) moltiplicare una riga per un intero invertibile, ossia per ± 1,
(3) sommare ad una riga un multiplo intero di un'altra riga.

Avendo perso la possibilità di dividere per numeri diversi da 0, è ragionevole attendersi che in generale non valgono le riduzioni di Gauss e Gauss–Jordan. Tale problema è stato ampiamente studiato nel corso del XIX secolo da diversi matematici, in particolare da Hermite, e si colloca naturalmente nell'ambito dei corsi di algebra, e più precisamente nella teoria dei moduli su anelli ad ideali principali.

Qui invece ci occupiamo di vedere cosa succede se, oltre alle operazioni elementari sulle righe, possiamo agire sulla matrice anche con operazioni elementari sulle colonne, ossia con operazioni del tipo:

(1) permutare le colonne,
(2) moltiplicare una colonna per un intero invertibile, ossia per ± 1,
(3) sommare ad una colonna un multiplo intero di un'altra colonna.

In tal caso abbiamo il seguente risultato, dimostrato da Henry J. S. Smith nel 1861.

TEOREMA 7.7.1. *Sia $A \in M_{n,m}(\mathbb{Z})$ una matrice a coefficienti interi. Mediante un numero finito di operazioni elementari sulle righe e di operazioni elementari sulle colonne è possibile trasformare la matrice A in una matrice $B = (b_{ij})$ tale che:*

(1) *$b_{ij} = 0$ se $i \neq j$;*
(2) *b_{ii} divide b_{jj} per ogni $i \leq j$.*

Si può mostrare che i coefficienti b_{ii} sono univocamente determinati a meno del segno; la dimostrazione richiede l'uso del determinante ed è pertanto posticipata all'Esercizio 459.

La dimostrazione del Teorema 7.7.1 è ottenuta applicando un numero finito di volte il seguente Lemma 7.7.2; i dettagli sono lasciati per esercizio al lettore. Anticipiamo che il punto chiave di tutto il procedimento è la divisione Euclidea di interi, ed in particolare del fatto che che se n, m sono due interi con $m > 0$, allora esiste un intero s tale che $0 \leq n + sm < m$.

LEMMA 7.7.2. *Per ogni matrice non nulla A a coefficienti interi indichiamo con $\nu(A) > 0$ il minimo tra i valori assoluti dei coefficienti non nulli di A. Data una matrice non nulla $A \in M_{n,m}(\mathbb{Z})$, mediante un numero finito di operazioni elementari sulle righe e sulle colonne è possibile trasformare A in una matrice B che soddisfa una delle seguenti condizioni:*

(1) *$\nu(B) < \nu(A)$, oppure*

(2) *$B = \begin{pmatrix} b_{11} & 0 \\ 0 & C \end{pmatrix}$, con $C \in M_{n-1,m-1}(\mathbb{Z})$ e b_{11} divide tutti i coefficienti di C.*

DIMOSTRAZIONE. A meno di permutare righe e colonne non è restrittivo assumere $\nu(A) = |a_{11}|$ e moltiplicando la prima riga per ± 1 possiamo supporre $\nu(A) = a_{11}$. Per ogni indice $j > 1$ possiamo sommare alla colonna j un opportuno multiplo intero della prima colonna in modo tale che la risultante matrice B sia tale che $0 \leq b_{1j} < b_{11} = a_{11}$ per ogni $j > 1$. Se $b_{1j} > 0$ per qualche j allora $\nu(B) < \nu(A)$. Se invece $b_{1j} = 0$ per ogni $j > 0$, ripetiamo la procedura con le righe; per ogni indice $j > 0$ sommiamo alla riga j un opportuno multiplo intero della prima riga in modo tale che la risultante matrice C sia tale che $0 \leq c_{j1} < c_{11} = b_{11} = a_{11}$ per ogni $j > 1$. Se $c_{j1} > 0$ per qualche $j > 1$ allora $\nu(C) < \nu(A)$. Se invece $c_{j1} = 0$ per ogni j si ha

$$C = \begin{pmatrix} a_{11} & 0 \\ 0 & \hat{C} \end{pmatrix}, \qquad \hat{C} \in M_{n-1,m-1}(\mathbb{Z}),$$

e se esiste un coefficiente c_{ij}, $i, j > 1$, non divisibile per a_{11}, basterà sommare alla prima riga la riga i e poi sommare alla colonna j un opportuno multiplo intero della prima colonna in modo da ottenere una matrice D con $\nu(D) \leq |d_{1j}| < a_{11} = \nu(A)$. □

Esercizi.

416 (☕, ♡). Sia $A \in M_{n,n}(\mathbb{Z})$ tale che $(I + pA)^q = I$, per qualche coppia di numeri primi p, q. Provare che vale $A = 0$ oppure $p = q = 2$ e $A + A^2 = 0$.

CAPITOLO 8

Il determinante

Dopo aver introdotto il concetto di matrice invertibile abbiamo immediatamente osservato che, a differenza degli scalari, non tutte le matrici quadrate diverse da 0 possiedono un'inversa. Abbiamo dato alcune condizioni geometriche necessarie e sufficienti affinché una matrice quadrata risulti invertibile (righe o colonne linearmente indipendenti) ed abbiamo visto anche alcune relazioni che collegano l'invertibilità ed il calcolo dell'inversa con la riduzione di Gauss–Jordan.

In questo capitolo introdurremo il determinante, che tra le altre cose fornirà una condizione algebrica, di importanza teorica fondamentale, per stabilire se una matrice è invertibile. Per ogni matrice quadrata A a coefficienti in un campo $\mathbb{K}$ definiremo uno scalare $\det(A) \in \mathbb{K}$, detto *determinante di* A, in grado di dare molte informazioni sulla natura algebrica di A: in particolare risulterà che A è invertibile come matrice se e solo se il determinante è invertibile come scalare.

Osserveremo poi che le formule che definiscono il determinante funzionano pure per matrici a coefficienti polinomi, e questo sarà alla base di alcuni risultati di notevole importanza e profondità che affronteremo nei capitoli seguenti.

8.1. Una formula per il determinante

Iniziamo con il definire, per ogni $n \geq 0$ e per ogni matrice quadrata $A \in M_{n,n}(\mathbb{K})$ uno scalare $|A| \in \mathbb{K}$ detto **determinante** di A. Spesso si usa anche la notazione $\det(A)$ per indicare $|A|$, ed anche

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \quad \text{per indicare il determinante della matrice} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Se $n = 0$ e $A \in M_{0,0}(\mathbb{K})$ è la matrice vuota poniamo per convenzione $|A| = 1$. Se $n = 1$ ed $A = (a)$, con $a \in \mathbb{K}$ poniamo $|A| = a$. Se $n > 1$ definiamo $|A|$ in maniera ricorsiva, come una funzione polinomiale dei coefficienti di A e dei determinanti di opportune sottomatrici di A di ordine minore di n. Data una matrice $A = (a_{ij}) \in M_{n,n}(\mathbb{K})$ indichiamo con $A_{ij} \in M_{n-1,n-1}(\mathbb{K})$ la sottomatrice ottenuta cancellando la riga i e la colonna j; ad esempio:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \quad A_{11} = \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix}, \quad A_{23} = \begin{pmatrix} 1 & 2 \\ 7 & 8 \end{pmatrix}, \quad A_{31} = \begin{pmatrix} 2 & 3 \\ 5 & 6 \end{pmatrix} \quad \text{ecc.}$$

Definiamo poi il determinante di A tramite la formula ricorsiva:

$$|A| = a_{11}|A_{11}| - a_{12}|A_{12}| + a_{13}|A_{13}| - \cdots = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} |A_{1j}|. \tag{8.1}$$

Esempio 8.1.1. Per $n = 2$ si ha

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Esempio 8.1.2.

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - 2 \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + 3 \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} = (45 - 48) - 2(36 - 42) + 3(32 - 35) = 0.$$

ESEMPIO 8.1.3.

$$\begin{vmatrix} 1 & 2 & 3 & -1 \\ 4 & 5 & 6 & 0 \\ 7 & 8 & 9 & 1 \\ 0 & 1 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 5 & 6 & 0 \\ 8 & 9 & 1 \\ 1 & 0 & 1 \end{vmatrix} - 2 \begin{vmatrix} 4 & 6 & 0 \\ 7 & 9 & 1 \\ 0 & 0 & 1 \end{vmatrix} + 3 \begin{vmatrix} 4 & 5 & 0 \\ 7 & 8 & 1 \\ 0 & 1 & 1 \end{vmatrix} - (-1) \begin{vmatrix} 4 & 5 & 6 \\ 7 & 8 & 9 \\ 0 & 1 & 0 \end{vmatrix}.$$

OSSERVAZIONE 8.1.4. Nella Formula (8.1) abbiamo supposto che gli indici di riga e colonna della matrice A siano i numeri interi compresi tra 1 ed n, ed il fattore $(-1)^{1+j}$ sta a significare che si considera la somma alterna partendo da $+1$. Per essere ulteriormente chiari, per una matrice $B = (b_{ij})$ con $i, j = 0, \dots, n$ si ha

$$|B| = b_{00}|B_{00}| - b_{01}|B_{01}| + \cdots + (-1)^n b_{0n}|B_{0n}|.$$

ESEMPIO 8.1.5. *Il determinante di una matrice $A \in M_{n,n}(\mathbb{K})$ che possiede una colonna nulla è uguale a 0.* Dimostriamo tale fatto per induzione su n, essendo del tutto evidente per $n = 1$. Supponiamo $n > 1$ e che per un indice j si abbia $a_{ij} = 0$ per ogni $i = 1, \dots, n$. In particolare $a_{1j} = 0$ e siccome anche la matrice A_{1k} ha una colonna nulla per ogni $k \neq j$ si ha per l'ipotesi induttiva $|A_{1k}| = 0$ per $j \neq k$ e quindi

$$|A| = \sum_{k=1}^{n} (-1)^{k+1} a_{1k} |A_{1k}| = (-1)^{j+1} a_{1j} |A_{1j}| + \sum_{k \neq j} (-1)^{k+1} a_{1k} |A_{1k}| = 0.$$

ESEMPIO 8.1.6. *Il determinante di una matrice triangolare è uguale al prodotto degli elementi sulla diagonale principale.* È possibile dimostrare tale fatto per induzione sull'ordine della matrice, essendo lo stesso del tutto evidente per le matrici quadrate di ordine 1. Sia $A = (a_{ij}) \in M_{n,n}(\mathbb{K})$ triangolare, $n > 1$, allora per definizione

$$|A| = a_{11}|A_{11}| - a_{12}|A_{12}| + a_{13}|A_{13}| - \cdots = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} |A_{1j}|.$$

Se A è triangolare inferiore si ha $a_{1j} = 0$ per ogni $j > 1$, mentre se A è triangolare superiore, per ogni $j > 1$ la sottomatrice A_{1j} è triangolare superiore con la prima colonna nulla e per l'Esempio 8.1.5 $|A_{1j}| = 0$. In ogni caso, quando A è triangolare vale la formula $|A| = a_{11}|A_{11}|$ e basta applicare il principio di induzione per concludere la dimostrazione.

DEFINIZIONE 8.1.7. Sia V uno spazio vettoriale sul campo $\mathbb{K}$, un'applicazione

$$\Phi \colon \underbrace{V \times \cdots \times V}_{n \text{ fattori}} \to \mathbb{K}$$

si dice **multilineare**, o anche **separatamente lineare**, se è lineare in ognuna delle n variabili, quando le rimanenti $n - 1$ sono lasciate fisse.

In altri termini, Φ è multilineare se per ogni indice i vale

$$\Phi(v_1, \dots, \lambda v_i, \dots, v_n) = \lambda \Phi(v_1, \dots, v_i, \dots, v_n),$$
$$\Phi(v_1, \dots, v_i + w_i, \dots, v_n) = \Phi(v_1, \dots, v_i, \dots, v_n) + \Phi(v_1, \dots, w_i, \dots, v_n).$$

ESEMPIO 8.1.8. Se $f, g \colon V \to \mathbb{K}$ sono due applicazioni lineari, allora

$$\Phi \colon V \times V \to \mathbb{K}, \qquad \Phi(v, w) = f(v)g(w),$$

è multilineare. Infatti

$$\begin{aligned} \Phi(\lambda v, w) &= f(\lambda v)g(w) = \lambda f(v)g(w) = \lambda \Phi(v, w), \\ \Phi(v, \lambda w) &= f(v)g(\lambda w) = \lambda f(v)g(w) = \lambda \Phi(v, w), \\ \Phi(u + v, w) &= f(u + v)g(w) = (f(u) + f(v))g(w) = f(u)g(w) + f(v)g(w) \\ &= \Phi(u, w) + \Phi(v, w), \\ \Phi(v, w + z) &= f(v)g(w + z) = f(v)(g(w) + g(z)) = f(v)g(w) + f(v)g(z) \\ &= \Phi(v, w) + \Phi(v, z). \end{aligned}$$

TEOREMA 8.1.9. *L'applicazione $A \mapsto |A|$ definita in (8.1) ha le seguenti proprietà:*

(1) *L'applicazione $A \mapsto |A|$ è multilineare sulle colonne.*

(2) *Se la matrice A ha due colonne adiacenti uguali, allora $|A| = 0$.*
(3) *$|I| = 1$, dove I è la matrice identità.*

Per essere precisi, la condizione (1) equivale a dire che l'applicazione

$$\underbrace{\mathbb{K}^n \times \cdots \times \mathbb{K}^n}_{n \text{ fattori}} \to \mathbb{K}, \qquad (A^1, \dots, A^n) \mapsto |A^1, \dots, A^n|,$$

è multilineare, dove ogni A^i è un vettore colonna e $|A^1, \dots, A^n|$ è il determinante della matrice che ha come colonne $A^1, \dots, A^n$.

Dimostrazione. Dimostriamo la multilinearità per induzione sull'ordine delle matrici. Per semplicità espositiva mostriamo che il determinante è lineare rispetto alla prima colonna: la linearità rispetto alle altre colonne è del tutto simile. Consideriamo quindi una matrice $A = (A^1, \dots, A^n) = (a_{ij})$, uno scalare $\lambda \in \mathbb{K}$ ed un vettore colonna $B^1 = (b_{11}, \dots, b_{n1})^T$. Considerando le matrici

$$C = (\lambda A^1, A^2, \dots, A^n), \quad D = (B^1, A^2, \dots, A^n), \quad E = (A^1 + B^1, A^2, \dots, A^n),$$

occorre dimostrare che

$$|C| = \lambda |A|, \qquad |E| = |A| + |D|.$$

Si ha $C_{11} = A_{11}$, mentre per ogni $j > 1$ la matrice C_{1j} è ottenuta da A_{1j} moltiplicando la prima riga per λ; per l'ipotesi induttiva $|C_{1j}| = \lambda |A_{1j}|$ per ogni $j > 1$ e quindi per la formula (8.1) si ha

$$|C| = \lambda a_{11}|C_{11}| - a_{12}|C_{12}| + \cdots = \lambda a_{11}|A_{11}| - a_{12}\lambda|A_{12}| + \cdots = \lambda |A|.$$

Similmente si ha $E_{11} = A_{11} = D_{11}$ e per induzione $|E_{1j}| = |A_{1j}| + |D_{1j}|$ per ogni $j > 1$. Quindi

$$\begin{aligned} |E| &= (a_{11} + b_{11})|E_{11}| - a_{12}|E_{12}| + \cdots \\ &= (a_{11}|A_{11}| - a_{12}|A_{12}| + \cdots) + (b_{11}|D_{11}| - a_{12}|D_{12}| + \cdots) = |A| + |D|. \end{aligned}$$

Supponiamo adesso che la matrice $A = (A^1, \dots, A^n)$ abbia due colonne adiacenti uguali, diciamo $A^i = A^{i+1}$. Allora per ogni $j \neq i, i+1$ la matrice A_{1j} ha due colonne adiacenti uguali. Per induzione $|A_{1j}| = 0$ per ogni $j \neq i, i+1$ e quindi la formula (8.1) si riduce a

$$|A| = (-1)^{i+1} a_{1,i} |A_{1,i}| + (-1)^{i+2} a_{1,i+1} |A_{1,i+1}|$$

e basta osservare che $a_{1,i} = a_{1,i+1}$ e $A_{1,i} = A_{1,i+1}$ per avere $|A| = 0$. Il determinante della matrice identità si calcola facilmente per induzione. Infatti $|I| = |I_{11}|$ e la sottomatrice I_{11} è ancora una matrice identità. □

Osservazione 8.1.10. Per calcolare il determinante di una matrice abbiamo usato solo le operazioni, entrambe commutative ed associative, di somma e prodotto: non abbiamo mai dovuto dividere per alcun coefficiente. Se $A(x) = (a_{ij}(x))$ è una matrice i cui coefficienti sono polinomi $a_{ij}(x) \in \mathbb{K}[x]$ possiamo ancora calcolare il determinante, che continuerà ad essere un polinomio. Ad esempio

$$\begin{vmatrix} x & 2 \\ x^2 - x & x + 1 \end{vmatrix} = x(x+1) - 2(x^2 - x) = -x^2 + 3x \in \mathbb{K}[x].$$

Se valutiamo tutti i coefficienti di $A(x)$ in uno scalare $\lambda \in \mathbb{K}$ otteniamo una matrice $A(\lambda) \in M_{n,n}(\mathbb{K})$; siccome i morfismi di valutazione commutano con somme e prodotti, il determinante $|A(x)| \in \mathbb{K}[x]$ calcolato in $x = \lambda$ coincide con il determinante $|A(\lambda)|$.

Esercizi.

417. Calcolare i determinanti:

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

418. Calcolare i determinanti

$$\begin{vmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{vmatrix}, \qquad \begin{vmatrix} 1 & 1 & \omega \\ 1 & 1 & \omega^2 \\ \omega^2 & \omega & 1 \end{vmatrix}, \qquad \text{dove } \omega = \cos\frac{2\pi}{3} + i\sin\frac{2\pi}{3}.$$

419. Verificare la correttezza della seguente formula (scoperta da Lagrange nel 1773):

$$a_{11}\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} - \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}.$$

420. Sia A una matrice quadrata di ordine 2. Trovare una formula per il calcolo del determinante in funzione delle tracce di A ed A^2.

421. Per l'Esempio 8.1.6 il determinante di una matrice diagonale è uguale al prodotto degli elementi sulla diagonale principale. A cosa è uguale il determinante delle matrici antidiagonali, ossia delle matrici (a_{ij}), $i, j = 1, \dots, n$, tali che $a_{ij} = 0$ ogni volta che $i + j \neq n + 1$?

422 (♡). Tenendo presente l'Osservazione 8.1.10, si consideri il determinante

$$p(x) = \begin{vmatrix} x & x^2 & 2x & 3x \\ 1 & x^2 & 4 & x^3 \\ 1 & x^3 & 4x & 5 \\ 1 & x^4 & 16 & x^9 \end{vmatrix} \in \mathbb{K}[x].$$

Calcolare $p(0)$, $p(1)$ e $p(2)$.

423. Calcolare i determinanti

$$\begin{vmatrix} 1 & -x \\ a_0 & a_1 \end{vmatrix}, \quad \begin{vmatrix} 1 & -x & 0 \\ 0 & 1 & -x \\ a_0 & a_1 & a_2 \end{vmatrix}, \quad \begin{vmatrix} 1 & -x & 0 & 0 \\ 0 & 1 & -x & 0 \\ 0 & 0 & 1 & -x \\ a_0 & a_1 & a_2 & a_3 \end{vmatrix}, \quad \begin{vmatrix} 1 & -x & 0 & 0 & 0 \\ 0 & 1 & -x & 0 & 0 \\ 0 & 0 & 1 & -x & 0 \\ 0 & 0 & 0 & 1 & -x \\ a_0 & a_1 & a_2 & a_3 & a_4 \end{vmatrix}.$$

424. Imitare il ragionamento dell'Esempio 8.1.5 per dimostrare che il determinante di una matrice quadrata con una riga nulla è uguale a 0. Siano $A \in M_{n,n}(\mathbb{K})$, $B \in M_{m,m}(\mathbb{K})$ e $C \in M_{n,m}(\mathbb{K})$. Dimostrare per induzione su n le formule:

$$\begin{vmatrix} A & C \\ 0 & B \end{vmatrix} = |A||B|, \qquad \begin{vmatrix} 0 & B \\ A & C \end{vmatrix} = (-1)^{nm}|A||B|.$$

425. Calcolare il determinante

$$\begin{vmatrix} 1 & a & a^2 & a^3 \\ 0 & 1 & a & a^2 \\ b & 0 & 1 & a \\ c & e & 0 & 1 \end{vmatrix}$$

dove $a =$ il tarapia tapioco, $b =$ come fosse antani, $c =$ gli scribai con cofandina ed $e =$ le pastène soppaltate secondo l'articolo 12.

426 (♡). Dati tre scalari x, y, z, dire se è possibile trovare una matrice

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$$

tale che

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = x, \qquad \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} = y, \qquad \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} = z.$$

427 (☕,♡). Data una successione finita $d_1, \dots, d_n$ di scalari in un campo $\mathbb{K}$, denotiamo

$$\operatorname{diag}(d_1, \dots, d_n) = \begin{pmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_n \end{pmatrix}$$

la matrice diagonale che ha i coefficienti d_i sulla diagonale principale. Supponiamo che $2d_i \neq 0$ per ogni $i = 1, \dots, n$. Provare che per ogni matrice $A \in M_{n,n}(\mathbb{K})$ esiste una scelta dei segni $\epsilon_1, \dots, \epsilon_n = \pm 1$ tale che la matrice $|A + \operatorname{diag}(\epsilon_1 d_1, \dots, \epsilon_n d_n)| \neq 0$. Trovare un esempio di matrice A per cui la scelta dei segni è anche unica.

8.2. Segnatura delle permutazioni ed unicità del determinante

Vogliamo adesso dimostrare che le tre proprietà elencate nel Teorema 8.1.9 determinano univocamente il determinante.

DEFINIZIONE 8.2.1. Diremo che un'applicazione $d\colon M_{n,n}(\mathbb{K}) \to \mathbb{K}$ è **multilineare alternante sulle colonne** se soddisfa le seguenti due condizioni:

D1: L'applicazione d è multilineare sulle colonne.
D2: Se la matrice A ha due colonne adiacenti uguali, allora $d(A) = 0$.

ESEMPIO 8.2.2. Per il Teorema 8.1.9 l'applicazione determinante $A \mapsto |A|$ è multilineare alternante sulle colonne. Lo stesso vale per l'applicazione $d(A) = \lambda|A|$, dove $\lambda \in \mathbb{K}$ è uno scalare qualsiasi.

ESEMPIO 8.2.3. Fissata una matrice $B \in M_{n,n}(\mathbb{K})$ l'applicazione

$$d\colon M_{n,n}(\mathbb{K}) \to \mathbb{K}, \qquad d(A) = |BA|,$$

è multilineare alternante sulle colonne. Questo segue facilmente dal fatto che la i-esima colonna di BA è uguale a BA^i, dove A^i è la i-esima colonna di A.

LEMMA 8.2.4. *Sia $d\colon M_{n,n}(\mathbb{K}) \to \mathbb{K}$ multilineare alternante sulle colonne. Allora si ha:*

D3: *Se una matrice B è ottenuta da A scambiando tra loro le posizioni di due colonne vale $d(B) = -d(A)$.*
D4: *Se la matrice A ha due colonne uguali, allora $d(A) = 0$.*
D5: *Se A contiene una colonna nulla allora $d(A) = 0$.*
D6: *Se le colonne di A sono linearmente dipendenti allora $d(A) = 0$.*
D7: *Se $d(A) \neq 0$ allora la matrice A è invertibile.*
D8: *Se la matrice B è ottenuta da una matrice quadrata A aggiungendo ad una colonna una combinazione lineare delle altre colonne, allora $d(B) = d(A)$.*

In particolare tutte le precedenti proprietà valgono per la funzione $d(A) = |A|$.

DIMOSTRAZIONE. Sia B la matrice ottenuta da A scambiando tra di loro due colonne adiacenti, diciamo

$$A = (\dots, A^i, A^{i+1}, \dots), \qquad B = (\dots, A^{i+1}, A^i, \dots).$$

Per le proprietà D1 e D2 si ha l'uguaglianza

$$\begin{aligned} 0 &= d(\dots, A^i + A^{i+1}, A^i + A^{i+1}, \dots) \\ &= d(\dots, A^i, A^i + A^{i+1}, \dots) + d(\dots, A^{i+1}, A^i + A^{i+1}, \dots) \\ &= d(\dots, A^i, A^i, \dots) + d(\dots, A^i, A^{i+1}, \dots) + d(\dots, A^{i+1}, A^i, \dots) + d(\dots, A^{i+1}, A^{i+1}, \dots) \end{aligned}$$

che, sempre per D2, si riduce a

$$0 = d(\dots, A^i, A^{i+1}, \dots) + d(\dots, A^{i+1}, A^i, \dots),$$

e cioè $d(B) = -d(A)$.

Se adesso la matrice A ha due colonne uguali, diciamo A^i ed A^j con $i < j$ possiamo scambiare la colonna i con la colonna $i+1$, poi la colonna $i+1$ con la $i+2$ e si prosegue fino a quando la colonna i si trova nella posizione $j-1$. Si ha quindi $d(A) = (-1)^{j-i-1} d(B)$ dove B è una matrice con le colonne $j-1$ e j uguali. Dunque $d(B) = 0$, $d(A) = 0$ e questo prova D4.

Chiamiano scambio elementare lo scambio di due colonne adiacenti; per provare D3 è sufficiente dimostrare che ogni scambio di due colonne si può ottenere come composizione di un numero dispari di scambi elementari. Indichiamo con

$$\tau_i\colon M_{n,n}(\mathbb{K}) \to M_{n,n}(\mathbb{K})$$

l'applicazione che scambia la colonna i con la colonna $i+1$. Se $i<j$ si vede facilmente che la composizione di $2(j-i)-1$ scambi elementari

$$\underbrace{\tau_i\circ\tau_{i+1}\circ\cdots\circ\tau_{j-2}}_{\text{indici crescenti}}\circ\tau_{j-1}\circ\underbrace{\tau_{j-2}\circ\tau_{j-3}\circ\cdots\circ\tau_i}_{\text{indici decrescenti}}$$

scambia le colonne i,j e lascia le altre al loro posto.

Sia adesso $A=(A^1,\dots,A^n)$ una matrice con una colonna nulla, che per semplicità notazionale supporremo essere la prima. Allora $0=A^1=0A^1$ e quindi

$$d(A)=d(A^1,\dots,A^n)=d(0A^1,\dots,A^n)=0d(A^1,\dots,A^n)=0.$$

Resta da dimostrare la proprietà D6: supponiamo che le colonne $A^1,\dots,A^n$ siano linearmente dipendenti e, per fissare le idee che l'ultima colonna sia combinazione lineare delle precedenti: $A^n=a_1A^1+\cdots+a_{n-1}A^{n-1}$. Allora si ha

$$d(A^1,\dots,A^n)=d\left(A^1,\dots,A^{n-1},\sum_{i=1}^{n-1}a_iA^i\right)=\sum_{i=1}^{n-1}a_id(A^1,\dots,A^{n-1},A^i)=0.$$

La D7 segue immediatamente da D6 e dal Corollario 6.3.5.

Per quanto riguarda D8, per ipotesi $A=(A^1,\dots,A^n)$ e $B=(B^1,\dots,B^n)$, dove per un qualche indice fissato i si ha

$$B^i=A^i+\sum_{j\neq i}\alpha_jA^j,\qquad B^j=A^j\text{ per ogni }j\neq i,$$

per opportuni scalari α_j, $j\neq i$. Se indichiamo con $A(j)$ la matrice ottenuta sostituendo la colonna A^i con la colonna A^j, per multilinearità si ottiene

$$d(B)=d(A)+\sum_{j\neq i}\alpha_jd(A(j)).$$

Adesso basta osservare che per $i\neq j$ la matrice $A(j)$ ha le colonne i e j identiche e per D4 vale $d(A(j))=0$. □

Esempio 8.2.5. Consideriamo una matrice $A\in M_{n,n}(\mathbb{K})$ che sia della forma

$$A=\begin{pmatrix}B&C\\0&D\end{pmatrix}$$

con $B\in M_{r,s}(\mathbb{K})$, $C\in M_{r,n-s}(\mathbb{K})$ e $D\in M_{n-r,n-s}(\mathbb{K})$. Se $r<s$ allora $|A|=0$ in quanto le prime s colonne di A sono linearmente dipendenti.

Esempio 8.2.6. Per ogni $n>0$ indichiamo con

$$B_n=\begin{pmatrix}1&1&1&\cdots\\1&2&3&\cdots\\1&3&6&\cdots\\\vdots&\vdots&\vdots&\ddots\end{pmatrix}\in M_{n,n}(\mathbb{Q})$$

la matrice di coefficienti $b_{ij}=\binom{i+j-2}{i-1}=\dfrac{(i+j-2)!}{(i-1)!\,(j-1)!}$. Dimostriamo per induzione su n che $|B_n|=1$ per ogni n. Supponiamo $n>1$ e denotiamo con D_n la matrice ottenuta da B_n eseguendo nell'ordine le seguenti $n-1$ operazioni:

- sottrarre alla n-esima colonna la $(n-1)$-esima colonna,
- sottrarre alla $n-1$-esima colonna la $(n-2)$-esima colonna,

$\vdots$

- sottrarre alla seconda colonna la prima colonna.

Ciascuna di tali operazioni non cambia il determinante, ossia $|B_n|=|D_n|$, mentre dalle formule

$$\binom{i+j-2}{i-1}-\binom{i+j-3}{i-2}=\binom{i+j-3}{i-1},\qquad i\geq 2,j\geq 1,$$

ne consegue che

$$D_n = \begin{pmatrix} 1 & 0 \\ v & B_{n-1} \end{pmatrix}, \qquad \text{dove} \quad v = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix},$$

e quindi $|D_n| = |B_{n-1}|$.

Indichiamo con Σ_n l'insieme di tutte le **permutazioni** di $\{1, \dots, n\}$, ossia l'insieme di tutte le applicazioni bigettive $\sigma\colon \{1, \dots, n\} \to \{1, \dots, n\}$. Una **trasposizione** è una permutazione che scambia due elementi e lascia fissi i rimanenti: è intuitivo e facilmente dimostrabile (vedi Esercizio 431) che ogni permutazione si può scrivere come composizione di trasposizioni. Va osservato che non c'è un modo unico di scrivere una permutazione come composizione di trasposizioni: ad esempio, se per ogni coppia di interi $1 \le i < j \le 3$ denotiamo con $(ij) \in \Sigma_3$ la trasposizione che scambia i con j, allora si ha $(12) \circ (23) = (23) \circ (13)$.

Data una matrice $A \in M_{n,n}(\mathbb{K})$ ed una permutazione $\sigma \in \Sigma_n$ denotiamo con A^σ la matrice ottenuta da A permutando le colonne secondo quanto dettato da σ, ossia $(A^\sigma)^i = A^{\sigma(i)}$. In altri termini, se $A = (A^1, \dots, A^n)$ allora $A^\sigma = (A^{\sigma(1)}, \dots, A^{\sigma(n)})$.

Osservazione 8.2.7. Si noti che se σ, τ sono due permutazioni, allora vale la formula $A^{\sigma\circ\tau} = (A^\sigma)^\tau$. Infatti, se pensiamo la matrice $A = (A^1, \dots, A^n)$ come l'applicazione $A\colon \{1, \dots, n\} \to \mathbb{K}^n$, $i \mapsto A^i$, allora A^σ equivale alla composizione $A \circ \sigma$ e quindi $A^{\sigma\circ\tau} = A \circ (\sigma \circ \tau) = (A \circ \sigma) \circ \tau = (A^\sigma)^\tau$.

Definizione 8.2.8. La **segnatura** $(-1)^\sigma$ di una permutazione $\sigma \in \Sigma_n$ è definita tramite la formula

$$(-1)^\sigma = |I^\sigma|,$$

dove $I \in M_{n,n}(\mathbb{Q})$ è la matrice identità.

Se la permutazione σ è ottenuta come composizione di k trasposizioni, allora la matrice I^σ è ottenuta dall'identità con k scambi di colonne e quindi

$$(-1)^\sigma = |I^\sigma| = (-1)^k |I| = (-1)^k.$$

Ne consegue in particolare che se una permutazione σ è scritta come composizione di k trasposizioni, allora $(-1)^k$ dipende solo da σ e non dalla scelta delle trasposizioni. Una permutazione si dice **pari** se ha segnatura 1, o equivalentemente se può essere scritta come composizione di un numero pari di trasposizioni. Si dice **dispari** se ha segnatura -1.

Lemma 8.2.9. *Sia $d\colon M_{n,n}(\mathbb{K}) \to \mathbb{K}$ multilineare alternante sulle colonne. Allora si ha:*

D9: *Per ogni matrice A ed ogni permutazione σ vale $d(A^\sigma) = (-1)^\sigma d(A)$.*

Dimostrazione. Scriviamo σ come composizione di k trasposizioni. Per D3 ogni trasposizione fa cambiare segno e quindi $d(A^\sigma) = (-1)^k d(A) = (-1)^\sigma d(A)$. □

Siamo adesso in grado di dimostrare il teorema di unicità del determinante.

Teorema 8.2.10. *Sia $d\colon M_{n,n}(\mathbb{K}) \to \mathbb{K}$ multilineare alternante sulle colonne. Allora per ogni matrice A si ha*

$$d(A) = |A| d(I).$$

In particolare se $d(I) = 1$ allora d è uguale al determinante.

Dimostrazione. Se $A = (A^1, \dots, A^n) = (a_{ij})$ allora per ogni i vale $A^i = \sum_{j=1}^n a_{ji} e_j$, dove $e_1, \dots, e_n$ è la base canonica di $\mathbb{K}^n$. Per linearità rispetto alla prima colonna si ha

$$d(A) = d\left(\sum_{j=1}^n a_{j1} e_j, A^2, \dots, A^n\right) = \sum_{j=1}^n a_{j1} d(e_j, A^2, \dots, A^n).$$

Ripetendo la procedura per la seconda colonna

$$d(A) = \sum_{j=1}^n a_{j1} d(e_j, \sum_{k=1}^n a_{k2} e_k, A^3, \dots, A^n) = \sum_{j=1}^n \sum_{k=1}^n a_{j1} a_{k2} d(e_j, e_k, A^3, \dots, A^n)$$

e procedendo fino alla n-esima si ottiene

$$d(A) = \sum_{j_1=1}^{n} \cdots \sum_{j_n=1}^{n} a_{j_1 1} \cdots a_{j_n n} d(e_{j_1}, \ldots, e_{j_n}).$$

Siccome $d(e_{j_1}, \ldots, e_{j_n}) = 0$ quando due indici j_h coincidono la precedente formula si riduce a

$$d(A) = \sum_{\sigma \in \Sigma_n} a_{\sigma(1),1} \cdots a_{\sigma(n),n} d(e_{\sigma(1)}, \ldots, e_{\sigma(n)}).$$

e tenendo presente che

$$d(e_{\sigma(1)}, \ldots, e_{\sigma(n)}) = d(I^\sigma) = (-1)^\sigma d(I)$$

arriviamo alla formula

$$d(A) = \sum_{\sigma \in \Sigma_n} (-1)^\sigma a_{\sigma(1),1} \cdots a_{\sigma(n),n} d(I), \tag{8.2}$$

che per d uguale al determinante diventa

$$|A| = \sum_{\sigma \in \Sigma_n} (-1)^\sigma a_{\sigma(1),1} \cdots a_{\sigma(n),n}. \tag{8.3}$$

Dal confronto delle equazioni (8.2) e (8.3) si arriva alla conclusione. □

TEOREMA 8.2.11 (Binet). *Date due matrici $A, B \in M_{n,n}(\mathbb{K})$ si ha*

$$|AB| = |BA| = |A||B|.$$

DIMOSTRAZIONE. Abbiamo già osservato che l'applicazione

$$d\colon M_{n,n}(\mathbb{K}) \to \mathbb{K}, \qquad d(A) = |BA|,$$

è multilineare alternante sulle colonne e quindi vale

$$|BA| = d(A) = |A| d(I) = |A||BI| = |A||B|.$$

Per simmetria $|AB| = |B||A| = |A||B| = |BA|$. □

ESEMPIO 8.2.12. Usiamo il teorema di Binet per dimostrare che se esiste una matrice $A \in M_{n,n}(\mathbb{R})$ tale che $A^2 = -I$, allora n è pari. Infatti $|A|$ è un numero reale e vale $|A|^2 = |-I| = (-1)^n$.

COROLLARIO 8.2.13. *Una matrice quadrata A a coefficienti in un campo è invertibile se e solo se $|A| \neq 0$; in tal caso il determinante dell'inversa è uguale all'inverso del determinante.*

DIMOSTRAZIONE. Se A è invertibile allora $AA^{-1} = I$ e per il teorema di Binet $|A||A^{-1}| = |I| = 1$ da cui segue $|A| \neq 0$ e $|A^{-1}| = |A|^{-1}$. Viceversa, per il Corollario 6.3.5, se A non è invertibile le sue colonne sono linearmente dipendenti e quindi $|A| = 0$ per il Lemma 8.2.4. □

COROLLARIO 8.2.14. *Il rango di una matrice $A \in M_{n,m}(\mathbb{K})$ è uguale al più grande intero r per cui A contiene una sottomatrice quadrata di ordine r con determinante diverso da 0.*

DIMOSTRAZIONE. Abbiamo già dimostrato nel Corollario 6.4.5 che il rango di una matrice A è uguale al più grande intero r per cui A contiene una sottomatrice quadrata e invertibile di ordine r. □

ESEMPIO 8.2.15. Sia

$$A = \begin{pmatrix} 1 & 1 & 2 & \cdots \\ 1 & 2 & 6 & \cdots \\ 2 & 6 & 24 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \in M_{n,n}(\mathbb{Q})$$

la matrice di coefficienti $a_{ij} = (i + j - 2)!$, $1 \leq i, j \leq n$. Usiamo il teorema di Binet per dimostrare che vale la formula

$$|A| = \prod_{i=0}^{n-1} (i!)^2.$$

Notiamo che vale $A = CBC$, dove

$$B = (b_{ij}), \qquad b_{ij} = \binom{i+j-2}{i-1} = \frac{a_{ij}}{(i-1)!(j-1)!},$$

e C è la matrice diagonale di coefficienti $c_{ii} = (i-1)!$. Si ha $|C| = \prod_{i=0}^{n-1} i!$, per il teorema di Binet si ha $|A| = |B|(\prod_{i=0}^{n-1} i!)^2$ e per concludere la dimostrazione basta osservare che $|B| = 1$, come dimostrato nell'Esempio 8.2.6.

Osservazione 8.2.16. In determinati contesti, soprattutto fisici, la segnatura delle permutazioni viene sostituita dal **simbolo di Levi–Civita**. Fissato un intero positivo n ed n numeri $i_1, \dots, i_n \in \{1, \dots, n\}$ si definisce

$$\varepsilon_{i_1 i_2 \cdots i_n} = \det(e_{i_1}, \dots, e_{i_n}),$$

dove come al solito $e_1, \dots, e_n$ indica la base canonica di $\mathbb{Q}^n$. A differenza della segnatura, il simbolo di Levi–Civita $\varepsilon_{i_1 i_2 \cdots i_n}$ è definito, e vale 0, anche quando vi sono ripetizioni negli indici, ossia quando l'applicazione $k \mapsto i_k$ non è una permutazione.

Esercizi.

428. I numeri $2418, 1395, 8091, 8339$ sono divisibili per 31. Dimostrare senza effettuare il conto esplicito che il determinante

$$\begin{vmatrix} 2 & 4 & 1 & 8 \\ 1 & 3 & 9 & 5 \\ 8 & 0 & 9 & 1 \\ 8 & 3 & 3 & 9 \end{vmatrix}$$

è divisibile per 31.

429. Provare che $|\lambda A| = \lambda^n |A|$ per ogni matrice $A \in M_{n,n}(\mathbb{K})$ ed ogni scalare $\lambda \in \mathbb{K}$.

430. Provare che se $|A| \neq 0$ allora esiste una permutazione σ tale che la matrice A^σ non possiede elementi nulli sulla diagonale.

431. Sia $\sigma \in \Sigma_n$ una permutazione fissata; si assuma che $\sigma(n) \neq n$, si denoti τ la trasposizione che scambia n con $\sigma(n)$. Si provi che la permutazione $\eta = \tau \circ \sigma$ lascia fisso n e che $\sigma = \tau \circ \eta$. Dedurre per induzione su n che ogni permutazione di n elementi è composizione di al più $n-1$ trasposizioni.

432. Siano $A, B \in M_{n,n}(\mathbb{Z})$ matrici a coefficienti interi e sia $p > 0$ un intero positivo. Dimostrare che $|A + pB| - |A|$ è un intero divisibile per p. Suggerimento: scrivere $B = B_1 + \cdots + B_n$ dove ciascuna matrice B_i è nulla al di fuori della colonna i-esima.

433. Determinare tutte le radici complesse del polinomio

$$p(x) = \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1-x & 1 & \cdots & 1 \\ 1 & 1 & 2-x & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & n-x \end{vmatrix}.$$

434. Provare che

$$\begin{vmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 6 \\ 1 & 0 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 5 \\ 1 & 0 & 0 & 0 & 2 \end{vmatrix} = 0.$$

435 (♡). Sia $A \in M_{3,3}(\mathbb{Q})$ una matrice in cui ogni coefficiente è uguale a $+1$ o -1. Provare che il determinante di A assume uno dei tre valori $-4,\ 0,\ +4$.

436. Siano $f_1, \dots, f_n \colon \mathbb{K}^n \to \mathbb{K}$ applicazioni lineari. Provare che l'applicazione

$$d \colon M_{n,n}(\mathbb{K}) \to \mathbb{K}, \qquad d(A^1, \dots, A^n) = \sum_{\sigma \in \Sigma_n} (-1)^\sigma f_{\sigma(1)}(A^1) f_{\sigma(2)}(A^2) \cdots f_{\sigma(n)}(A^n),$$

è multilineare alternante sulle colonne.

437. Completare la seguente traccia di dimostrazione alternativa del teorema di Binet:

(1) Siano $C, D \in M_{m,m}(\mathbb{K})$, con D triangolare unipotente, ossia con i coefficienti sulla diagonale principale uguali ad 1. Allora CD è ottenuta da C tramite una successione finita di operazioni consistenti nel sommare ad una colonna opportuni multipli scalari delle altre colonne: quindi $|C| = |CD|$.

(2) Siano $A, B \in M_{n,n}(\mathbb{K})$ e $I \in M_{n,n}(\mathbb{K})$ la matrice identità, allora

$$\det \begin{pmatrix} I & -B \\ A & 0 \end{pmatrix} = \det \begin{pmatrix} B & I \\ 0 & A \end{pmatrix} = |A||B|;$$

(3) Usando i due punti precedenti, dedurre il Teorema di Binet dalla formula:

$$\begin{pmatrix} I & -B \\ A & 0 \end{pmatrix} \begin{pmatrix} I & B \\ 0 & I \end{pmatrix} = \det \begin{pmatrix} I & 0 \\ A & AB \end{pmatrix}.$$

438 (♡). Calcolare, per ogni intero $n > 0$ il determinante della matrice

$$A_n = \begin{pmatrix} 1 & 2 & \cdots & n-1 & n \\ 2 & 3 & \cdots & n & n \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ n-1 & n & \cdots & n & n \\ n & n & \cdots & n & n \end{pmatrix} \in M_{n,n}(\mathbb{R})$$

di coefficienti $a_{ij} = \max(i + j - 1, n)$.

8.3. Incroci e segnatura

Un modo di rappresentare una permutazione σ dell'insieme $\{1, \ldots, n\}$ è mediante una matrice $2 \times n$ in cui la prima riga contiene i numeri da 1 a n e la seconda riga le rispettive immagini, ossia

$$\begin{bmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{bmatrix}.$$

Ricordiamo che le trasposizioni sono permutazioni che scambiano di posizione due elementi e lasciano invariati i rimanenti. Ad esempio, la permutazione

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{bmatrix}$$

è una trasposizione. Per ogni $i < n$, denotiamo con $\tau_i \in \Sigma_n$ la trasposizione che scambia i con $i + 1$:

$$\tau_i(i) = i + 1, \qquad \tau_i(i + 1) = i, \qquad \tau_i(a) = a \quad \forall\, a \neq i, i + 1.$$

Definizione 8.3.1. Diremo che un sottoinsieme $A \subseteq \{1, \ldots, n\}$ di **due** elementi è un **incrocio** della permutazione σ se la restrizione di σ ad A è decrescente; in altri termini, un sottoinsieme

$$A = \{i, j\}, \quad \text{con} \quad i < j,$$

è un incrocio di σ se $\sigma(i) > \sigma(j)$.

Indichiamo con $\delta(\sigma)$ il numero di incroci di σ. Ad esempio, l'identità ha zero incroci, le trasposizioni τ_i hanno un solo incrocio, mentre la permutazione

$$\sigma(i) = n + 1 - i$$

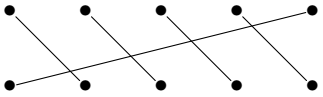
ha $n(n-1)/2$ incroci. In un certo senso, il numero di incroci è una misura della complessità della permutazione.

Osservazione 8.3.2. Una maniera di contare il numero di incroci di σ è la seguente. Per ogni $i = 1, \ldots, n$ si disegna nel piano il segmento che unisce il punto di coordinate $(i, 1)$ con il punto di coordinate $(\sigma(i), 0)$ e poi si conta il numero di punti di intersezione dei vari segmenti. Bisogna però fare attenzione al fatto che, se per un punto passano h segmenti, con

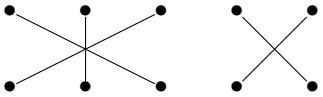
FIGURA 8.1. La copertina di Ummagumma rappresenta artisticamente una permutazione e le sue potenze.

$h > 2$, allora ci troviamo di fronte ad una intersezione multipla ed a tale punto corrispondono $h(h-1)/2$ incroci. Ad esempio, le due permutazioni

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 1 \end{bmatrix}$$



$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 1 & 5 & 4 \end{bmatrix}$$



hanno entrambe 4 incroci.

ESEMPIO 8.3.3. Siano $i < j$ e sia σ la trasposizione che scambia i e j. Un sottoinsieme

$$A = \{a, b\}, \quad \text{con} \quad a < b,$$

è un incrocio se e solo se $a = i$ e $b \leq j$, oppure se $a \geq i$ e $b = j$; quindi $\delta(\sigma) = 2(j-i) - 1$. Ne consegue che ogni trasposizione ha un numero dispari di incroci.

Le permutazioni, in quanto applicazioni di un insieme in sé, possono essere composte tra loro. Se $\sigma, \eta \in \Sigma_n$ definiamo il prodotto $\sigma\eta \in \Sigma_n$ come

$$\sigma\eta(i) = \sigma(\eta(i)), \qquad i = 1, \ldots, n.$$

Ad esempio

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{bmatrix}.$$

Notiamo che il prodotto di ogni trasposizione con sé stessa è uguale all'identità; in particolare $\tau_i \tau_i =$ identità per ogni i.

LEMMA 8.3.4. *Date due permutazioni $\sigma, \eta \in \Sigma_n$, indichiamo con a il numero di sottoinsiemi $A \subseteq \{1, \ldots, n\}$ di due elementi che soddisfano le due condizioni:*

(1) *A è un incrocio di η.*
(2) *$\eta(A)$ è un incrocio di σ.*

Allora vale la formula

$$\delta(\sigma\eta) = \delta(\sigma) + \delta(\eta) - 2a$$

DIMOSTRAZIONE. Per ogni insieme finito X indichiamo con $|X|$ la sua cardinalità, ossia il numero di elementi che contiene. Indichiamo con $\mathcal{P}$ la collezione di tutti i sottoinsiemi di $\{1, \ldots, n\}$ di cardinalità 2. Notiamo che $A \in \mathcal{P}$ è un incrocio di $\sigma\eta$ se e soltanto se vale una delle seguenti due condizioni:

(1) A è un incrocio di η e $\eta(A)$ non è un incrocio di σ.
(2) A non è un incrocio di η e $\eta(A)$ è un incrocio di σ.

Indichiamo adesso con

$$\mathcal{C} = \{A \in \mathcal{P} \mid A \text{ è incrocio di } \eta\}, \qquad \mathcal{D} = \{A \in \mathcal{P} \mid \eta(A) \text{ è incrocio di } \sigma\}.$$

Chiaramente $|\mathcal{C}| = \delta(\eta)$ e, siccome $\eta\colon \mathcal{P} \to \mathcal{P}$ è bigettiva, vale anche $|\mathcal{D}| = \delta(\sigma)$. Per definizione a è il numero di elementi di $\mathcal{C} \cap \mathcal{D}$. Denotiamo con c il numero di elementi di $\mathcal{C}$ che non appartengono a $\mathcal{D}$ e con d il numero di elementi di $\mathcal{D}$ che non appartengono a $\mathcal{C}$.
Abbiamo visto che valgono le uguaglianze

$$a + c = \delta(\eta), \qquad a + d = \delta(\sigma), \qquad c + d = \delta(\sigma\eta).$$

Da tali uguaglianze segue che

$$\delta(\sigma\eta) = \delta(\sigma) + \delta(\eta) - 2a.$$

□

Teorema 8.3.5. *Sia $\varepsilon\colon \Sigma_n \to \{\pm 1\}$ l'applicazione definita dalla formula*

$$\varepsilon(\sigma) = (-1)^{\delta(\sigma)}, \qquad \delta(\sigma) = \textit{numero di incroci di } \sigma.$$

Allora, per ogni $\sigma, \eta \in \Sigma_n$ vale

$$\varepsilon(\sigma\eta) = \varepsilon(\sigma)\varepsilon(\eta)$$

ed in particolare $\varepsilon(\sigma) = \varepsilon(\sigma^{-1})$. Se σ è uguale al prodotto di k trasposizioni, allora $\varepsilon(\sigma) = (-1)^k$ e quindi $\varepsilon(\sigma)$ è uguale alla segnatura $(-1)^\sigma$.

Dimostrazione. La prima uguaglianza segue immediatamente dal Lemma 8.3.4. Per la seconda basta osservare che

$$\varepsilon(\sigma)\varepsilon(\sigma^{-1}) = \varepsilon(\sigma\sigma^{-1}) = \varepsilon(\text{identità}) = 1.$$

Infine, sappiamo che ogni trasposizione ha un numero dispari di incroci. □

Corollario 8.3.6. *Ogni permutazione σ si può scrivere come prodotto di $\delta(\sigma)$ trasposizioni τ_i.*

Dimostrazione. Induzione su $\delta(\sigma)$. Se σ non ha incroci, allora σ è l'identità. Se invece σ è diversa dall'identità, allora l'applicazione bigettiva

$$\sigma\colon \{1, \dots, n\} \to \{1, \dots, n\}$$

non può essere crescente e dunque esiste almeno un indice $h < n$ tale che $\sigma(h) > \sigma(h+1)$. Dimostriamo adesso che

$$\delta(\sigma\tau_h) = \delta(\sigma) - 1.$$

Infatti la trasposizione τ_h ha un unico incrocio $\{h, h+1\}$ che, per come abbiamo scelto h, è anche un incrocio di σ. Quindi per il Lemma 8.3.4

$$\delta(\sigma\tau_h) = \delta(\sigma) + \delta(\tau_h) - 2 = \delta(\sigma) - 1.$$

Per l'ipotesi induttiva la permutazione $\sigma\tau_h$ è prodotto di $\delta(\sigma) - 1$ trasposizioni τ_i e quindi

$$\sigma = \sigma(\tau_h\tau_h) = (\sigma\tau_h)\tau_h$$

è prodotto di $\delta(\sigma)$ trasposizioni τ_i. □

Supponiamo adesso di avere un insieme finito X e di considerare una permutazione di X, ossia un'applicazione bigettiva $f\colon X \to X$. In questo caso non possiamo definire il numero di incroci (per fare ciò bisognerebbe che X fosse ordinato) ma possiamo ugualmente definire la segnatura nel modo seguente:

Supponiamo che X abbia esattamente n elementi e scegliamo un'applicazione bigettiva

$$h\colon \{1, \dots, n\} \to X.$$

Allora l'applicazione

$$h^{-1}fh\colon \{1, \dots, n\} \to \{1, \dots, n\}$$

è bigettiva e possiamo definire

$$\varepsilon(f) = \varepsilon(h^{-1}fh).$$

FIGURA 8.2. Il gioco del 15 nella configurazione iniziale.

Bisogna dimostrare che si tratta di una **buona definizione**, ossia che $\varepsilon(f)$ non dipende dalla scelta di h. Se prendiamo un'altra applicazione bigettiva

$$k\colon \{1,\dots,n\} \to X,$$

allora $\sigma = k^{-1}h$ è una permutazione di $\{1,\dots,n\}$ con inversa $\sigma^{-1} = h^{-1}k$ e quindi

$$\varepsilon(h^{-1}fh) = \varepsilon(\sigma^{-1}k^{-1}fk\sigma) = \varepsilon(\sigma^{-1})\varepsilon(k^{-1}fk)\varepsilon(\sigma) = \varepsilon(k^{-1}fk).$$

Esercizi.

439. Siano $r < n$ e $\sigma \in \Sigma_n$ la permutazione

$$\sigma(i) = \begin{cases} i + r & \text{se } i \le n - r \\ i - (n - r) & \text{se } i > n - r \end{cases}$$

Calcolare la segnatura di σ.

440 (Il gioco del 15). Il gioco del quindici è un rompicapo classico inventato nel XIX secolo. Il gioco consiste di una tabellina di forma quadrata, solitamente di plastica, divisa in quattro righe e quattro colonne (quindi 16 posizioni), su cui sono posizionate 15 tessere quadrate, numerate progressivamente a partire da 1 (vedi Figura 8.2). Le tessere possono scorrere in orizzontale o verticale, ma il loro spostamento è ovviamente limitato dall'esistenza di un singolo spazio vuoto. Lo scopo del gioco è riordinare le tessere dopo averle "mescolate" in modo casuale (la posizione da raggiungere è quella con il numero 1 in alto a sinistra e gli altri numeri a seguire da sinistra a destra e dall'alto in basso, fino al 15 seguito dalla casella vuota).

Dopo aver mescolato le tessere indichiamo con i, $1 \le i \le 4$, il numero di riga contenente lo spazio vuoto e con $\sigma \in \Sigma_{15}$ la permutazione ottenuta leggendo i numeri allo stile occidentale, ossia da sinistra a destra e dall'alto in basso, ignorando lo spazio vuoto. Dimostrare che $(-1)^\sigma = (-1)^i$.

441 (☕). Siano dati un campo $\mathbb{K}$, un insieme X ed un'applicazione $f_1\colon X \times X \to \mathbb{K}$ tale che $f_1(x,y) = -f_1(y,x)$ per ogni $x, y \in X$. Si considerino le applicazioni

$$f_n\colon \underbrace{X \times \cdots \times X}_{2n \text{ fattori}} \to \mathbb{K}, \qquad n > 0,$$

definite per $n > 1$ mediante la formula ricorsiva

$$f_n(x_1,\dots,x_{2n}) = (-1)^{n+1} \sum_{i=1}^{2n} (-1)^i f_1(x_1,x_i) f_{n-1}(x_2,\dots,x_{i-1},x_{i+1},\dots,x_{2n})\,.$$

Provare che per ogni permutazione σ di $\{1,\dots,2n\}$ si ha

$$f_n(x_{\sigma(1)},\dots,x_{\sigma(2n)}) = (-1)^\sigma f_n(x_1,\dots,x_{2n})\,.$$

Nel caso particolare in cui $X = \{x_1,\dots,x_n,y_1,\dots,y_n\}$ e $f_1(x_i,x_j) = f_1(y_i,y_j) = 0$ per ogni i, j, calcolare $f_n(x_1,\dots,x_n,y_1,\dots,y_n)$ in funzione della matrice di coefficienti $a_{ij} = f_1(x_i,y_j)$.

442 (Ⓐ). Per risolvere questo esercizio sono necessarie alcune nozioni di teoria delle serie di potenze. Per ogni intero $n \ge 0$ indichiamo con d_n il numero di permutazioni senza punti fissi di n elementi: $d_0 = 1$, $d_1 = 0$, $d_2 = 1$, $d_3 = 3$ eccetera. Per ogni $0 \le k \le n$ indichiamo inoltre $D_n(k)$ il numero di permutazioni con esattamente k punti fissi di un insieme di n elementi: chiaramente $d_n = D_n(0)$ e $n! = \sum_{k=0}^{n} D_n(k)$. Dimostrare che:

(1) $D_n(k) = \binom{n}{k} d_{n-k}$;

(2) $\sum_{k=0}^{n} \frac{d_k}{k!} \frac{1}{(n-k)!} = 1$;

(3) nell'intervallo aperto $(-1, 1)$ la serie di potenze $f(t) = \sum_{k=0}^{\infty} \frac{d_k}{k!} t^n$ è convergente e vale $f(t)e^t = \sum_{n=0}^{\infty} t^n$;

(4) per ogni n vale $\frac{d_n}{n!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}$.

8.4. Sviluppi di Laplace

In gergo matematico la formula per il determinante data nella Sezione 8.1 viene detta "sviluppo di Laplace rispetto alla prima riga" e, come vedremo in questa sezione, ammette formule analoghe per ogni riga ed ogni colonna.

PROPOSIZIONE 8.4.1. *Sia $A = (a_{ij})$, $i, j = 1, \dots, n$ una matrice quadrata. Allora per ogni indice $i = 1, \dots, n$ fissato vale lo* sviluppo di Laplace rispetto alla riga i*:*

$$|A| = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} |A_{ij}|.$$

DIMOSTRAZIONE. Per $i = 1$ la formula è vera per definizione. Definiamo per ogni i l'applicazione

$$d_i \colon M_{n,n}(\mathbb{K}) \to \mathbb{K}, \qquad d_i(A) = \sum_{j=1}^{n} (-1)^{j+1} a_{ij} |A_{ij}|;$$

e dimostriamo per induzione su i che $d_i(A) = (-1)^{i-1}|A|$ per ogni matrice A. Sia τ_i la trasposizione semplice che scambia gli indici $i, i+1$ e sia B la matrice ottenuta da A scambiando tra loro le righe i e $i+1$. Valgono allora le formule

$$d_{i+1}(A) = d_i(B), \qquad B = I^{\tau_i} A.$$

Per il teorema di Binet e per l'ipotesi induttiva si ha:

$$d_{i+1}(A) = d_i(B) = (-1)^{i-1}|B| = (-1)^{i-1}|I^{\tau_i}||A| = (-1)^i |A|.$$

□

ESEMPIO 8.4.2. Calcoliamo il determinante della matrice

$$\begin{pmatrix} 1 & 3 & 5 \\ 6 & 7 & 0 \\ 2 & 0 & 0 \end{pmatrix}$$

Dallo sviluppo di Laplace rispetto all'ultima riga segue

$$\begin{vmatrix} 1 & 3 & 5 \\ 6 & 7 & 0 \\ 2 & 0 & 0 \end{vmatrix} = 2 \begin{vmatrix} 3 & 5 \\ 7 & 0 \end{vmatrix} = -70.$$

LEMMA 8.4.3 (determinante della trasposta). *Per ogni matrice $A \in M_{n,n}(\mathbb{K})$ vale $|A^T| = |A|$.*

DIMOSTRAZIONE. Siccome $|I^T| = 1$ basta dimostrare che l'applicazione

$$d \colon M_{n,n}(\mathbb{K}) \to \mathbb{K}, \qquad d(A) = |A^T|,$$

è multilineare alternante sulle colonne.

Indicati con a_{ij} e coefficienti di A, fissato un indice i, per lo sviluppo di Laplace rispetto alla riga i si ha:

$$d(A) = |A^T| = \sum_{j=1}^{n} a_{ji} (-1)^{i+j} |(A^T)_{ij}|$$

e da tale formula segue immediatamente che $d(A)$ è lineare rispetto alla colonna i. Infine se A ha le colonne $i, i+1$ uguali allora ogni vettore colonna di A^T è contenuto nel sottospazio

vettoriale di equazione $x_i - x_{i+1} = 0$; dunque le colonne di A^T sono linearmente dipendenti e quindi $|A^T| = 0$. □

COROLLARIO 8.4.4. *Sia $A = (a_{ij}) \in M_{n,n}(\mathbb{K})$, allora per ogni indice $i = 1, \dots, n$ fissato vale lo* Sviluppo di Laplace rispetto alla colonna *i:*

$$|A| = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |A_{ij}|.$$

DIMOSTRAZIONE. Prendendo lo sviluppo di Laplace rispetto alla riga i della matrice trasposta si ha

$$|A^T| = \sum_{j=1}^{n} a_{ji} (-1)^{i+j} |A_{ij}^T|.$$

Siccome $(A^T)_{ij} = (A_{ji})^T$ ed il determinante di una matrice è uguale al determinante della propria trasposta si ha

$$|A| = |A^T| = \sum_{j=1}^{n} a_{ji} (-1)^{i+j} |(A^T)_{ij}| = \sum_{j=1}^{n} a_{ji} (-1)^{i+j} |A_{ji}|.$$

□

Dal fatto che il determinante di una matrice è uguale al determinante della trasposta, segue che il determinante è multilineare alternante sulle righe. In particolare:

(1) Scambiando due righe il determinante cambia di segno.
(2) Moltiplicando una riga per uno scalare λ, anche il determinante viene moltiplicato per λ.
(3) Aggiungendo ad una riga una combinazione lineare delle altre, il determinante non cambia.
(4) Se le righe sono linearmente dipendenti il determinante si annulla.

COROLLARIO 8.4.5. *Sia $A \in M_{n,n}(\mathbb{K})$ una matrice alternante, ossia antisimmetrica con i coefficienti sulla diagonale principale nulli. Allora A ha rango pari. In particolare, se n è dispari allora $|A| = 0$.*

DIMOSTRAZIONE. Dimostriamo prima che se n è dispari, allora $|A| = 0$. Siccome $A^T = -A$ si ha $|A| = |A^T| = |-A| = (-1)^n |A| = -|A|$ e quindi $2|A| = 0$. Se il campo $\mathbb{K}$ ha caratteristica $\neq 2$ questo basta per affermare che $|A| = 0$, mentre per campi di caratteristica 2 occorre fare un ragionamento per induzione su n. Se $n = 1$ allora $A = 0$; se $n \geq 3$ e $A = (a_{ij})$, per ogni coppia di indici $i, j = 2, \dots, n$, indichiamo con $A_{ij} \in M_{n-2,n-2}(\mathbb{K})$ la matrice ottenuta togliendo le righe $1, i$ e le colonne $1, j$. Siccome $a_{11} = 0$, effettuando lo sviluppo di Laplace rispetto alla prima colonna e successivamente rispetto alla prima riga otteniamo la formula

$$|A| = \sum_{i,j=2}^{n} (-1)^{i+j+1} a_{i1} a_{1j} |A_{ij}| .$$

Siccome $A_{ij}^T = -A_{ji}$ e A_{ii} è alternante per ogni i, possiamo assumere per induzione che $|A_{ij}| = -|A_{ji}|$ e $|A_{ii}| = 0$ per ogni i, j e quindi

$$\begin{aligned} |A| &= \sum_{i,j=2}^{n} (-1)^{i+j+1} a_{i1} a_{1j} |A_{ij}| = - \sum_{i,j=2}^{n} (-1)^{i+j+1} a_{1i} a_{1j} |A_{ij}| \\ &= \sum_{2 \leq i < j \leq n} (-1)^{i+j} a_{1i} a_{1j} (|A_{ij}| + |A_{ji}|) = 0. \end{aligned}$$

Per il Teorema 6.5.10 sappiamo che il rango di una matrice antisimmetrica è uguale al massimo intero r tale che esiste una sottomatrice principale $r \times r$ invertibile. Per quanto visto sopra l'invertibilità della sottomatrice principale implica che r è pari. □

Giova osservare che una matrice alternante di ordine pari può avere determinante non nullo, come ad esempio la matrice $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.

ESEMPIO 8.4.6. Le precedenti regole permettono di calcolare il determinante con un misto di eliminazione di Gauss e sviluppo di Laplace. Supponiamo ad esempio di voler calcolare il determinante

$$\lambda = \begin{vmatrix} 10 & 20 & 32 \\ 4 & 2 & 25 \\ 3 & 0 & 9 \end{vmatrix}$$

Togliamo alla terza colonna il doppio della prima; il determinante non cambia:

$$\lambda = \begin{vmatrix} 10 & 20 & 12 \\ 4 & 2 & 17 \\ 3 & 0 & 0 \end{vmatrix} = 3 \begin{vmatrix} 20 & 12 \\ 2 & 17 \end{vmatrix} = 3(340 - 24) = 632.$$

ESEMPIO 8.4.7. Per il calcolo del determinante

$$\Delta = \begin{vmatrix} a-b-c & 2a & 2a \\ 2b & b-c-a & 2b \\ 2c & 2c & c-a-b \end{vmatrix}$$

possiamo sostituire alla prima riga la somma delle tre righe

$$\Delta = \begin{vmatrix} a+b+c & a+b+c & a+b+c \\ 2b & b-c-a & 2b \\ 2c & 2c & c-a-b \end{vmatrix} = (a+b+c) \begin{vmatrix} 1 & 1 & 1 \\ 2b & b-c-a & 2b \\ 2c & 2c & c-a-b \end{vmatrix}$$

e poi sottrarre la prima colonna alle altre due

$$\Delta = (a+b+c) \begin{vmatrix} 1 & 0 & 0 \\ 2b & -b-c-a & 0 \\ 2c & 0 & -c-a-b \end{vmatrix} = (a+b+c)^3.$$

ESEMPIO 8.4.8. Calcoliamo il determinante della matrice di Vandermonde; più precisamente proviamo che per ogni $x_0, \dots, x_n \in \mathbb{K}$ si ha

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_0 & x_1 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_0^{n-1} & x_1^{n-1} & \cdots & x_n^{n-1} \\ x_0^n & x_1^n & \cdots & x_n^n \end{vmatrix} = \prod_{i>j} (x_i - x_j).$$

Ragioniamo per induzione su n, considerando il polinomio

$$p(t) = \prod_{j=0}^{n-1} (t - x_j) = t^n + \sum_{i=0}^{n-1} a_i t^i.$$

Sommando all'ultima riga della matrice di Vandermonde la combinazione lineare a coefficienti a_i delle rimanenti righe si ottiene

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_0 & x_1 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_0^{n-1} & x_1^{n-1} & \cdots & x_n^{n-1} \\ x_0^n & x_1^n & \cdots & x_n^n \end{vmatrix} = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_0 & x_1 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_0^{n-1} & x_1^{n-1} & \cdots & x_n^{n-1} \\ p(x_0) & p(x_1) & \dots & p(x_n) \end{vmatrix}$$

Dato che $p(x_i) = 0$ per ogni $i < n$ e $p(x_n) = \prod_{n>j} (x_n - x_j)$ si ha

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_0 & x_1 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_0^{n-1} & x_1^{n-1} & \cdots & x_n^{n-1} \\ x_0^n & x_1^n & \cdots & x_n^n \end{vmatrix} = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_0 & x_1 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_0^{n-1} & x_1^{n-1} & \cdots & x_n^{n-1} \\ 0 & 0 & \cdots & p(x_n) \end{vmatrix}$$

$$= p(x_n) \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_0 & x_1 & \cdots & x_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_0^{n-1} & x_1^{n-1} & \cdots & x_{n-1}^{n-1} \end{vmatrix} = \prod_{n>j} (x_n - x_j) \prod_{n>i>j} (x_i - x_j).$$

Se $x_i \neq x_j$ per ogni $i \neq j$ il determinante è diverso da 0 ed abbiamo quindi ridimostrato che in tal caso la matrice di Vandermonde

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_0 & x_1 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_0^{n-1} & x_1^{n-1} & \cdots & x_n^{n-1} \\ x_0^n & x_1^n & \cdots & x_n^n \end{pmatrix} \tag{8.4}$$

è invertibile.

Esercizi.

443. Sia

$$f(x) = \det \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & x \end{pmatrix}.$$

Osservando la matrice, e senza eseguire alcun calcolo, trovare un intero $x \in \mathbb{Z}$ tale che $f(x) \neq 0$. Successivamente, usare la riduzione di Gauss per determinare tutti gli x tali che $f(x) = 0$.

444. Calcolare i determinanti:

$$\begin{vmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 1 \\ 2 & 2 & 1 & 0 \\ -2 & 1 & 1 & 0 \end{vmatrix}, \qquad \begin{vmatrix} 13547 & 13647 \\ 22311 & 22411 \end{vmatrix}, \qquad \begin{vmatrix} 5 & 6 & 0 & 0 \\ 1 & 5 & 6 & 0 \\ 0 & 1 & 5 & 6 \\ 0 & 0 & 1 & 5 \end{vmatrix}.$$

445. Ridimostrare il risultato dell'Esercizio 424, ossia che per $A \in M_{n,n}(\mathbb{K})$, $B \in M_{m,m}(\mathbb{K})$ e $C \in M_{n,m}(\mathbb{K})$ vale

$$\begin{vmatrix} A & C \\ 0 & B \end{vmatrix} = |A||B|, \qquad \begin{vmatrix} 0 & B \\ A & C \end{vmatrix} = (-1)^{nm}|A||B|,$$

usando gli sviluppi di Laplace e le proprietà del determinante.

446. Usare lo sviluppo di Laplace rispetto all'ultima riga per dimostrare la formula

$$\begin{vmatrix} 1 & -\lambda & 0 & \dots & 0 \\ 0 & 1 & -\lambda & \dots & 0 \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & 0 & \dots & -\lambda \\ a_0 & a_1 & a_2 & \dots & a_n \end{vmatrix} = a_0\lambda^n + a_1\lambda^{n-1} + \cdots + a_n,$$

cf. Esercizio 423. Ridimostrare la stessa formula per induzione su n utilizzando lo sviluppo di Laplace rispetto alla prima colonna.

447 (♡). Dimostrare che

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_0 & x_1 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_0^{n-1} & x_1^{n-1} & \cdots & x_n^{n-1} \\ x_0^{n+1} & x_1^{n+1} & \cdots & x_n^{n+1} \end{vmatrix} = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_0 & x_1 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_0^{n-1} & x_1^{n-1} & \cdots & x_n^{n-1} \\ x_0^n & x_1^n & \cdots & x_n^n \end{vmatrix} \left(\sum_{i=0}^{n} x_i \right).$$

448. Dati due interi positivi n, p si consideri la matrice $A = (a_{ij})$, dove

$$a_{ij} = (ni + j)p + 1, \qquad i, j = 1, \dots, n.$$

Per quali valori di n, p il determinante di A è uguale a -1250?

449. Siano n un intero positivo dispari e $A \in M_{n,n}(\mathbb{Z})$ una matrice simmetrica con tutti gli elementi sulla diagonale principale numeri pari. Usare l'Esercizio 432 per dimostrare che il determinante di A è un numero pari. Dedurre che ogni matrice alternante di ordine dispari a coefficienti nel campo $\mathbb{F}_2 = \{0, 1\}$ ha determinante nullo.

450. Dimostrare, usando l'eliminazione di Gauss, che il determinante della matrice

$$\begin{pmatrix} 4 & 2 & 2 & 2 & 8 & 6 & 6 \\ 1 & -1 & -1 & 3 & 0 & 2 & 4 \\ 2 & 1 & -1 & 3 & 5 & 7 & -1 \\ 2 & 1 & 6 & 0 & 3 & -8 & 3 \\ 2 & 1 & 1 & 0 & -2 & 7 & 3 \\ 2 & 1 & 1 & 0 & 0 & 7 & 3 \\ 2 & 1 & 1 & 0 & 2 & 7 & 3 \end{pmatrix}$$

è uguale a 0.

451. Sia A una matrice 10×10. Calcolare, in funzione di $|A|$, il determinante della seguente matrice 20×20

$$\begin{pmatrix} 6A & 5A \\ A & 2A \end{pmatrix}.$$

452. Calcolare, in funzione dell'ordine n, il determinante delle matrici "epantemiche":

$$A_3 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad A_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \cdots \quad A_n = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{pmatrix},$$

cf. Esercizio 27.

453. Calcolare i determinanti

$$\begin{vmatrix} 1 & a_1 & a_2 & \cdots & a_n \\ 1 & a_1+b_1 & a_2 & \cdots & a_n \\ 1 & a_1 & a_2+b_2 & \cdots & a_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_1 & a_2 & \cdots & a_n+b_n \end{vmatrix}, \qquad \begin{vmatrix} 1 & 2 & 2 & \cdots & 2 \\ 2 & 2 & 2 & \cdots & 2 \\ 2 & 2 & 3 & \cdots & 2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 2 & 2 & 2 & \cdots & n \end{vmatrix}.$$

454 (♡). Sia $A \in M_{9,9}(\mathbb{Z})$ una matrice Sudoku, o più generalmente una matrice 9×9 a coefficienti interi tale che:

(1) la somma dei coefficienti di ogni riga è 45;
(2) la somma dei coefficienti di ogni colonna è 45.

Dimostrare che il determinante di A è divisibile per 420.[1] Generalizzare il risultato alle matrici $A \in M_{n,n}(\mathbb{Z})$ in cui la somma dei coefficienti di ogni riga e colonna è uguale a nm, con $m \in \mathbb{Z}$.

455 (♡). Provare la formula

$$\frac{1}{n!}\begin{vmatrix} 2 & 1 & 1 & \cdots & 1 \\ 1 & 3 & 1 & \cdots & 1 \\ 1 & 1 & 4 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & n+1 \end{vmatrix} = 1 + \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n}.$$

456. La formula per il determinante della matrice di Vandermonde può essere usata per dimostrare il seguente fatto non banale. Siano $x_1, \ldots, x_n$ numeri interi, allora il prodotto $\prod_{i<j}(x_j - x_i)$ è divisibile, negli interi, per il prodotto

$$\prod_{1\le i<j\le n}(j-i) = \prod_{2\le j\le n}(j-1)! = \prod_{1\le h<n} h^{n-h}.$$

Diamo solamente il punto chiave della dimostrazione, lasciando per esercizio il compito di aggiungere i passaggi mancanti. Per ogni coppia di interi x, n con $n > 0$ vale

$$\binom{x}{n} = \frac{1}{n!}x(x-1)\cdots(x-n+1) \in \mathbb{Z}.$$

[1] Contrariamente a quanto affermato in alcune leggende metropolitane, una matrice Sudoku può avere determinante diverso da 0.

Dedurre che il determinante della matrice $A = (a_{ij})$, $a_{ij} = \dfrac{x_i^{j-1}}{(j-1)!}$, $i, j = 1, \dots, n$ è un numero intero.

457 (☕). Indichiamo con d_k, $k \geq 1$, il determinante della matrice $k \times k$

$$\begin{pmatrix} 6 & 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 6 & 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 6 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & 6 & 1 & \dots & \\ & & & \dots & & & \\ 0 & 0 & 0 & 0 & 0 & \dots & 6 \end{pmatrix}$$

($d_1 = 6$, $d_2 = 35$ eccetera). Dimostrare che per ogni $k \geq 3$ vale $d_k = 6d_{k-1} - d_{k-2}$.
Siano x, y le radici del polinomio $t^2 - 6t + 1$. Dimostrare che per ogni $k \geq 3$ vale

$$x^k = 6x^{k-1} - x^{k-2}, \qquad y^k = 6y^{k-1} - y^{k-2}.$$

Determinare due numeri reali a, b tali che

$$d_k = ax^k + by^k$$

per ogni $k \geq 1$.

458 (☕). Usare lo stesso ragionamento dell'Esercizio 457 per calcolare il determinante

$$\begin{vmatrix} 1+x^2 & x & 0 & 0 & 0 & \dots & 0 \\ x & 1+x^2 & x & 0 & 0 & \dots & 0 \\ 0 & x & 1+x^2 & x & 0 & \dots & 0 \\ 0 & 0 & x & 1+x^2 & x & \dots & \\ & & & \dots & & & \\ 0 & 0 & 0 & 0 & 0 & \dots & 1+x^2 \end{vmatrix} \in \mathbb{Q}[x].$$

459 (☕, Unicità della riduzione di Smith). Sia $A \in M_{n,m}(\mathbb{Z})$ una matrice a coefficienti interi; per ogni intero $k = 1, \dots, \min(n, m)$ indichiamo con d_k il massimo comune divisore di tutti i determinanti delle sottomatrici quadrate di ordine k di A. Dimostrare che d_k è invariante per le operazioni elementari sulle righe e sulle colonne descritte nella Sezione 7.7. Calcolare i coefficienti d_k per le matrici (a_{ij}) tali che $a_{ij} = 0$ se $i \neq j$ e a_{ii} divide a_{jj} se $i \leq j$.

8.5. Aggiunta classica e regola di Cramer

Data una matrice quadrata A di ordine n, per ogni coppia di indici $1 \leq i, j \leq n$ lo scalare $(-1)^{i+j}|A_{ij}|$ viene detto **cofattore** di A alla posizione (i, j) o, più brevemente, (i, j)-cofattore di A. I cofattori formano a loro volta una matrice quadrata di ordine n chiamata per l'appunto matrice dei cofattori che assume un certo interesse, anche se si rivelerà molto più utile la sua trasposta.

Definizione 8.5.1. Data una matrice quadrata $A \in M_{n,n}(\mathbb{K})$, la **matrice dei cofattori** $\operatorname{cof}(A)$ e l'**aggiunta classica**[2] $\operatorname{adj}(A)$ di A sono definite mediante la formula:

$$\operatorname{cof}(A) = (c_{ij}), \qquad c_{ij} = (-1)^{i+j}|A_{ij}|, \qquad \operatorname{adj}(A) = \operatorname{cof}(A)^T = \operatorname{cof}(A^T).$$

L'uguaglianza $\operatorname{cof}(A)^T = \operatorname{cof}(A^T)$ è perché ogni matrice quadrata ha determinante uguale alla sua trasposta, e quindi per ogni i, j si ha $|A_{ji}| = |(A_{ji})^T| = |(A^T)_{ij}|$.

Esempio 8.5.2. L'aggiunta classica di $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ è uguale a $\begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix}$. L'aggiunta classica di una matrice 1×1 è sempre uguale alla matrice identità.

Teorema 8.5.3. *Per ogni matrice quadrata A vale*

$$A \operatorname{adj}(A) = \operatorname{adj}(A) A = |A| I.$$

[2] In alcuni testi l'aggiunta classica viene chiamata semplicemente aggiunta: noi sconsigliamo questa terminologia in quanto esiste un'altra nozione di matrice aggiunta che non ha nulla a che vedere con la presente.

DIMOSTRAZIONE. Se $A = (a_{ij})$, tenendo presente la definizione di $\mathrm{adj}(A)$ e del prodotto di matrici, la formula $A\,\mathrm{adj}(A) = |A|I$ equivale alle relazioni

$$\sum_{k=1}^{n}(-1)^{k+j}a_{ik}|A_{jk}| = \begin{cases}|A| & \text{se } i = j\\ 0 & \text{se } i \neq j\end{cases}. \tag{8.5}$$

Per $i = j$ la formula (8.5) coincide con lo sviluppo di Laplace rispetto alla riga i. Se invece $i \neq j$ indichiamo con B la matrice ottenuta da A mettendo la riga i al posto della riga j; la matrice B ha dunque due righe uguali (la i e la j) e vale $a_{ik} = b_{ik} = b_{jk}$, $A_{jk} = B_{jk}$ per ogni k. Ne segue che

$$0 = |B| = \sum_{k=1}^{n}(-1)^{k+j}b_{jk}|B_{jk}| = \sum_{k=1}^{n}(-1)^{k+j}a_{ik}|A_{jk}|.$$

La formula $\mathrm{adj}(A)A = |A|I$ si dimostra allo stesso modo utilizzando gli sviluppi di Laplace rispetto alle colonne. □

COROLLARIO 8.5.4. *Una matrice quadrata A è invertibile se e solo se $|A| \neq 0$; in tal caso l'inversa è uguale a $A^{-1} = \dfrac{\mathrm{adj}(A)}{|A|}$.*

DIMOSTRAZIONE. Abbiamo già dimostrato che A è invertibile se e solo se $|A| \neq 0$ ed in tal caso $|A^{-1}| = |A|^{-1}$. Se $|A| \neq 0$ segue dal Teorema 8.5.3 che la matrice $\dfrac{\mathrm{adj}(A)}{|A|}$ è l'inversa di A. □

TEOREMA 8.5.5 (regola di Cramer). *Sia $x_1, \dots, x_n$ una soluzione di un sistema lineare di n equazioni in n incognite*

$$\begin{cases}a_{11}x_1 + \cdots + a_{1n}x_n = b_1\\ \quad\cdots \qquad \cdots\\ a_{n1}x_1 + \cdots + a_{nn}x_n = b_n\end{cases}.$$

Allora per ogni i vale

$$x_i|A| = |B_i|$$

dove A è la matrice dei coefficienti del sistema e B_i è la matrice ottenuta sostituendo la i-esima colonna di A con la colonna $b = (b_1, \dots, b_n)^T$ dei termini noti.

DIMOSTRAZIONE. Indichiamo con $A^i = (a_{1i}, \dots, a_{ni})^T$ la i-esima colonna di A. Dire che $x_1, \dots, x_n$ è una soluzione del sistema equivale a dire che

$$x_1A^1 + \cdots + x_nA^n = b.$$

Dunque per la multilineare alternanza del determinante si ha

$$|B_1| = |x_1A^1 + \cdots + x_nA^n, A^2, \dots, A^n| = \sum_{i=1}^{n} x_i|A^i, A^2, \dots, A^n| = x_1|A|.$$

In maniera del tutto simile si prova che $|B_i| = x_i|A|$ per ogni indice i.

Se sappiamo già che A è una matrice invertibile e scriviamo il sistema nella forma $Ax = b$, con $x, b \in \mathbb{K}^n$, allora possiamo dimostrare la regola di Cramer anche nel modo seguente: dallo sviluppo di Laplace di $|B_i|$ rispetto alla colonna i segue che $|B_i| = b_1c_{1i} + \cdots + b_nc_{ni}$, dove c_{ij} sono i cofattori di A. Dunque $\mathrm{adj}(A)b$ è il vettore di coordinate $|B_1|, \dots, |B_n|$ e quindi

$$x = A^{-1}b = \frac{\mathrm{adj}(A)}{|A|}b = \frac{1}{|A|}(|B_1|, \dots, |B_n|)^T.$$

□

ESEMPIO 8.5.6. Sia $A \in M_{n,n}(\mathbb{K})$ una matrice quadrata di rango r. Dimostriamo che il rango della sua aggiunta classica $\mathrm{adj}(A)$ è uguale a:

(1) n se $r = n$;
(2) 1 se $r = n - 1$;
(3) 0 se $r \leq n - 2$.

Se $r = n$ allora $|A| \neq 0$ e il prodotto $\operatorname{adj}(A)\, A = |A| I$ è una matrice invertibile; dunque anche $\operatorname{adj}(A)$ deve essere invertibile. Se $r \leq n-2$, per il Corollario 8.2.14 la matrice dei cofattori è nulla. Sempre per il Corollario 8.2.14, se $r = n-1$ si ha $\operatorname{adj}(A) \neq 0$ e $A \operatorname{adj}(A) = 0$. Dunque l'immagine dell'applicazione lineare $L_{\operatorname{adj}(A)}$ è contenuta nel nucleo di L_A che per la formula di Grassmann ha dimensione 1. Abbiamo quindi provato che il rango dell'aggiunta classica è minore od uguale a 1 e diverso da 0.

Esercizi.

460. Calcolare l'inversa della matrice $\begin{pmatrix} 1 & -3 & 0 \\ 1 & 0 & 1 \\ 0 & 2 & 1 \end{pmatrix}$.

461. Calcolare le inverse delle matrici

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & -1 \\ 0 & -1 & 1 \\ 1 & 1 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

462. Vero o falso? Ogni matrice 2×4 nella quale i determinanti dei minori 2×2 formati da due colonne adiacenti si annullano ha rango minore di 2.

463. Provare che l'inversa di una matrice triangolare invertibile è ancora triangolare.

464. Sia $B \in M_{n,n+1}(\mathbb{K})$ e denotiamo con x_j il determinante della matrice $n \times n$ ottenuta togliendo a B la j-esima colonna. Dimostrare che

$$B \begin{pmatrix} x_1 \\ -x_2 \\ x_3 \\ \vdots \\ (-1)^n x_{n+1} \end{pmatrix} = 0\,.$$

465. Risolvere, usando la regola di Cramer, il sistema

$$\begin{cases} x + y + z = 1 \\ x + 2y + 3z = 4 \\ x + 4y + 9z = 16 \end{cases}.$$

466. Provare che l'aggiunta classica di una matrice simmetrica è ancora simmetrica. Cosa si può dire dell'aggiunta classica di una matrice antisimmetrica?

467. Siano $A \in M_{n,n}(\mathbb{K})$ una matrice qualsiasi e $\widetilde{A} = \operatorname{cof}(A)^T$ la sua aggiunta classica. Provare che per ogni $v, w \in \mathbb{K}^n$ si ha

$$v^T \widetilde{A} w = -\det \begin{pmatrix} A & w \\ v^T & 0 \end{pmatrix}.$$

468. Calcolare le inverse delle seguenti matrici:

$$\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix}, \qquad \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{pmatrix}.$$

Osservazione 8.5.7. Le matrici quadrate di coefficienti $a_{ij} = \frac{1}{i+j-1}$ vengono chiamate *matrici di Hilbert* ed hanno la curiosa proprietà di avere l'inversa a coefficienti interi. Ad esempio

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \end{pmatrix}^{-1} = \begin{pmatrix} 16 & -120 & 240 & -140 \\ -120 & 1200 & -2700 & 1680 \\ 240 & -2700 & 6480 & -4200 \\ -140 & 1680 & -4200 & 2800 \end{pmatrix}.$$

La dimostrazione di questo fatto va al di là degli scopi di queste note e viene pertanto omessa. Sarà invece molto facile, dopo che avremo dimostrato alcuni risultati sulle matrici

simmetriche reali, dimostrare che le matrici di Hilbert hanno tutte determinante positivo (Esercizio 600).

469 (☕, ♡). Dimostrare la seguente generalizzazione del teorema di Binet. Siano $n \le m$ due interi positivi e si considerino due matrici

$$A = (A^1, \dots, A^m) \in M_{n,m}(\mathbb{K}), \qquad B = \begin{pmatrix} B_1 \\ \vdots \\ B_m \end{pmatrix} \in M_{m,n}(\mathbb{K}),$$

dove gli $A^i \in \mathbb{K}^n$ sono i vettori colonna di A ed i $B_i \in \mathbb{K}^{(n)}$ sono i vettori riga di B; si noti che $AB \in M_{n,n}(\mathbb{K})$. Dimostrare che vale la formula

$$\det(AB) = \sum_{1 \le i_1 < \cdots < i_n \le m} \det(A^{i_1}, \dots, A^{i_n}) \cdot \det \begin{pmatrix} B_{i_1} \\ \vdots \\ B_{i_n} \end{pmatrix}.$$

470. È possibile dimostrare che il polinomio $t^6 - t + 1$ possiede 6 radici complesse distinte $a_1, \dots, a_6$, ossia con determinante di Vandermode diverso da 0:

$$\Delta = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_6 \\ \vdots & \vdots & \ddots & \vdots \\ a_1^5 & a_2^5 & \cdots & a_6^5 \end{vmatrix} \neq 0\,.$$

Calcolare i sette rapporti tra i determinanti dei minori 6×6 della matrice

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_6 \\ \vdots & \vdots & \ddots & \vdots \\ a_1^6 & a_2^6 & \cdots & a_6^6 \end{vmatrix}$$

ed il determinante Δ della matrice di Vandermonde. (Non si chiede di calcolare le radici $a_1, \dots, a_6$ ma di fare un ragionamento astratto sulla regola di Cramer).

471 (☕). Ricordiamo da (2.6) che la successione $B_0, B_1, B_2, \dots$ dei numeri di Bernoulli può essere definita in maniera ricorsiva dalle equazioni

$$B_n \in \mathbb{Q}, \quad B_0 = 1, \qquad \sum_{i=0}^{n} \binom{n+1}{i} B_i = 0, \quad n > 0.$$

Usare lo sviluppo di Laplace rispetto all'ultima colonna per dimostrare induttivamente la seguente rappresentazione determinantale dei numeri di Bernoulli

$$B_n = \frac{(-1)^n}{(n-1)!} \begin{vmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots & \frac{1}{n} & \frac{1}{n+1} \\ 1 & 1 & 1 & \cdots & 1 & 1 \\ 0 & 2 & 3 & \cdots & n-1 & n \\ 0 & 0 & \binom{3}{2} & \cdots & \binom{n-1}{2} & \binom{n}{2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \binom{n-1}{n-2} & \binom{n}{n-2} \end{vmatrix}, \qquad n > 0.$$

(Per maggior chiarezza, i coefficienti a_{ij}, $i, j = 1, \dots, n$, della matrice nella precedente formula sono uguali a $a_{1j} = 1/(j+1)$ e $a_{ij} = \binom{j}{i-2}$ per $i > 1$.)

8.6. Complementi: lo Pfaffiano

Abbiamo visto nel Corollario 8.4.5 che le matrici alternanti di ordine dispari hanno determinante nullo. In questa sezione studieremo le matrici alternanti di ordine pari, dimostrando in particolare che il loro determinante è un quadrato in $\mathbb{K}$.

Per ogni matrice alternante A ed ogni coppia di indici i, j, indichiamo come al solito con A_{ij} la matrice ottenuta cancellando la riga i e la colonna j; notiamo che A_{ii} è ancora alternante per ogni i.

DEFINIZIONE 8.6.1 (Lo Pfaffiano). Per ogni matrice $A \in M_{n,n}(\mathbb{K})$ alternante di ordine pari, definiamo $\mathrm{Pf}(A) \in \mathbb{K}$ ricorsivamente nel modo seguente:

(1) per $n = 2$ si ha $\mathrm{Pf}\begin{pmatrix} 0 & a \\ -a & 0 \end{pmatrix} = a$;

(2) per $n \geq 4$ consideriamo la matrice alternante $B = A_{nn}$ di ordine (dispari) $n-1$ e poniamo

$$\mathrm{Pf}(A) = \sum_{i=1}^{n-1} (-1)^{i+1} \mathrm{Pf}(B_{ii}) a_{in}.$$

Quando $n = 0$ si pone per convenzione uguale ad 1 lo Pfaffiano della matrice vuota.

Ad esempio, per la generica matrice alternante 4×4

$$A = \begin{pmatrix} 0 & a & b & c \\ -a & 0 & d & e \\ -b & -d & 0 & f \\ -c & -e & -f & 0 \end{pmatrix}$$

si ha

$$B = \begin{pmatrix} 0 & a & b \\ -a & 0 & d \\ -b & -d & 0 \end{pmatrix}, \quad B_{11} = \begin{pmatrix} 0 & d \\ -d & 0 \end{pmatrix}, \quad B_{22} = \begin{pmatrix} 0 & b \\ -b & 0 \end{pmatrix}, \quad B_{33} = \begin{pmatrix} 0 & a \\ -a & 0 \end{pmatrix},$$

e quindi lo Pfaffiano è dato dalla formula

$$\mathrm{Pf}\begin{pmatrix} 0 & a & b & c \\ -a & 0 & d & e \\ -b & -d & 0 & f \\ -c & -e & -f & 0 \end{pmatrix} = dc - be + af.$$

Si dimostra facilmente per induzione che lo Pfaffiano di una matrice alternante di ordine $n = 2m$ è un polinomio di grado m nei coefficienti.

LEMMA 8.6.2. *Siano $A \in M_{n,n}(\mathbb{K})$ e $B \in M_{m,m}(\mathbb{K})$ due matrici alternanti. Se gli interi n e m sono entrambi pari, allora*

$$\mathrm{Pf}\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = \mathrm{Pf}(A)\,\mathrm{Pf}(B).$$

DIMOSTRAZIONE. Lasciata per esercizio (suggerimento: induzione su $m/2$). □

TEOREMA 8.6.3. *Sia $A \in M_{n,n}(\mathbb{K})$, $n \geq 2$, una matrice alternante:*

(1) *se n è pari vale $|A| = \mathrm{Pf}(A)^2$ e per ogni matrice $H \in M_{n,n}(\mathbb{K})$ vale*

$$\mathrm{Pf}(HAH^T) = |H|\,\mathrm{Pf}(A)\ ;$$

(2) *se n è dispari, allora $|A_{ij}| = \mathrm{Pf}(A_{ii})\,\mathrm{Pf}(A_{jj})$ per ogni i, j e vale*

$$(\mathrm{Pf}(A_{11}), -\mathrm{Pf}(A_{22}), \mathrm{Pf}(A_{33}), \ldots, (-1)^{n-1}\mathrm{Pf}(A_{nn}))A = 0.$$

Osserviamo, prima della dimostrazione, che quando $H = I^\sigma$ è una matrice di permutazione, dalla relazione $\mathrm{Pf}(HAH^T) = |H|\,\mathrm{Pf}(A)$ segue che permutando allo stesso modo gli indici di riga e colonna di una matrice alternante di ordine pari, lo Pfaffiano viene moltiplicato per la segnatura della permutazione.

DIMOSTRAZIONE. Per semplificare la dimostrazione supponiamo che il campo $\mathbb{K}$ contenga infiniti elementi, pur essendo tale ipotesi non necessaria (vedi Esempio 3.7.10). Siccome

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0 & x \\ -x & 0 \end{pmatrix}\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} 0 & x(ad-bc) \\ x(bc-ad) & 0 \end{pmatrix}$$

il teorema è certamente vero per $n = 2$. Supponiamo quindi il teorema vero per tutte le matrici alternanti di ordine minore od uguale ad n e dimostriamo che vale per quelle di ordine $n+1$; occorre distinguere il caso n pari dal caso n dispari.

Caso n pari. Sia A una matrice alternante di ordine $n+1$; abbiamo dimostrato che $|A| = 0$ e quindi che il rango della sua aggiunta classica $\mathrm{adj}(A) = (c_{ij})$ è al più 1. Quindi il determinante

di ogni minore 2×2 di $\operatorname{adj}(A)$ si annulla ed in particolare, per ogni coppia di indici i, j si ha $c_{ii}c_{jj} - c_{ij}c_{ji} = 0$. Ricordando che $c_{ij} = (-1)^{i+j}|A_{ji}|$ si ottiene

$$|A_{ij}|\,|A_{ji}| = |A_{ii}|\,|A_{jj}|.$$

Notiamo che $(A_{ij})^T = (A^T)_{ji} = -A_{ji}$ e quindi $|A_{ij}| = |A_{ji}|$. Usando l'ipotesi induttiva otteniamo quindi

$$|A_{ij}|^2 = |A_{ii}|\,|A_{jj}| = \operatorname{Pf}(A_{ii})^2 \operatorname{Pf}(A_{jj})^2.$$

Siamo quindi in grado di dire che $|A_{ij}| = \pm \operatorname{Pf}(A_{ii}) \operatorname{Pf}(A_{jj})$ dove il segno $\pm$ dipende a priori da A e dalla coppia di indici i, j. Resta da far vedere che tale segno è sempre uguale a $+1$. Data una seconda matrice alternante B, indicando $C(t) = tA + (1-t)B$ si ha

$$\Big(|C(t)_{ij}| - \operatorname{Pf}(C(t)_{ii}) \operatorname{Pf}(C(t)_{jj}) \Big) \Big(|C(t)_{ij}| + \operatorname{Pf}(C(t)_{ii}) \operatorname{Pf}(C(t)_{jj}) \Big) = 0,$$

che trattandosi di un prodotto di polinomi in t può annullarsi solo se si annulla identicamente almeno uno dei due fattori. Ne consegue che, per ogni coppia di indici i, j basta trovare una matrice alternante B tale che $|B_{ij}| + \operatorname{Pf}(B_{ii}) \operatorname{Pf}(B_{jj}) \neq 0$: infatti, ciò implica che il polinomio

$$|C(t)_{ij}| + \operatorname{Pf}(C(t)_{ii}) \operatorname{Pf}(C(t)_{jj})$$

non è nullo in quanto la corrispondente funzione polinomiale non si annulla per $t = 0$. Questo implica

$$|C(t)_{ij}| - \operatorname{Pf}(C(t)_{ii}) \operatorname{Pf}(C(t)_{jj}) = 0$$

per ogni t ed in particolare per $t = 1$ e $C(1) = A$.

Per semplicità espositiva consideriamo solamente il caso $i = 1, j = 2$: i rimanenti casi possono essere dimostrati similmente oppure ricondotti al caso particolare mediante opportune permutazioni degli indici. Se $n + 1 = 2k + 3$ possiamo ad esempio considerare la matrice diagonale a blocchi

$$B = \begin{pmatrix} U & 0 & \cdots & 0 \\ 0 & J_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_k \end{pmatrix}, \qquad U = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ -1 & -1 & 0 \end{pmatrix}, \qquad J_1 = \cdots = J_k = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

per la quale vale

$$|B_{12}| = |U_{12}| = 1, \qquad \operatorname{Pf}(B_{11}) = \operatorname{Pf}(U_{11}) = 1, \qquad \operatorname{Pf}(B_{22}) = \operatorname{Pf}(U_{22}) = 1.$$

Per dimostrare l'uguaglianza

$$(\operatorname{Pf}(A_{11}), -\operatorname{Pf}(A_{22}), \ldots, (-1)^{n-1} \operatorname{Pf}(A_{nn}))A = 0$$

non è restrittivo supporre $\operatorname{Pf}(A_{hh}) \neq 0$ per un indice fissato h. Sia $\operatorname{adj}(A) = (c_{ij})$ l'aggiunta classica di A, dal Corollario 8.4.5 segue che $\operatorname{adj}(A)\,A = 0$, mentre dalla relazione $|A_{ij}| = \operatorname{Pf}(A_{ii}) \operatorname{Pf}(A_{jj})$ segue che $c_{ij} = (-1)^{i+j} \operatorname{Pf}(A_{ii}) \operatorname{Pf}(A_{jj})$. In particolare

$$0 = (c_{h1}, c_{h2}, \ldots, c_{hn})A = (-1)^{h+1} \operatorname{Pf}(A_{hh})(\operatorname{Pf}(A_{11}), -\operatorname{Pf}(A_{22}), \ldots, (-1)^{n-1} \operatorname{Pf}(A_{nn}))A.$$

Caso n dispari. Sia A una matrice alternante di ordine (pari) $n + 1$ e scriviamo

$$A = \begin{pmatrix} B & x^T \\ -x & 0 \end{pmatrix}$$

dove $x = (a_{1,n+1}, \ldots, a_{n,n+1})$. Per lo sviluppo di Laplace rispetto all'ultima colonna otteniamo

$$|A| = \sum_{i=1}^{n} (-1)^{n+i+1} a_{i,n+1} |A_{i,n+1}|$$

e calcolando ciascun determinante $|A_{i,n+1}|$ mediante lo sviluppo di Laplace rispetto all'ultima riga otteniamo

$$|A_{i,n+1}| = \sum_{j=1}^{n} (-1)^{n+j} (-a_{j,n+1}) |B_{ij}| = \sum_{j=1}^{n} (-1)^{n+j+1} a_{j,n+1} |B_{ij}|$$

da cui, utilizzando l'ipotesi induttiva

$$\begin{aligned}|A| &= \sum_{i,j=1}^{n} (-1)^{i+j+2} a_{i,n+1} a_{j,n+1} |B_{ij}| \\ &= \sum_{i,j=1}^{n} (-1)^{i+j+2} a_{i,n+1} a_{j,n+1} \operatorname{Pf}(B_{ii}) \operatorname{Pf}(B_{jj}) \\ &= \left(\sum_{i=1}^{n} (-1)^{i+1} a_{i,n+1} \operatorname{Pf}(B_{ii}) \right) \left(\sum_{j=1}^{n} (-1)^{j+1} a_{j,n+1} \operatorname{Pf}(B_{jj}) \right) \\ &= \operatorname{Pf}(A)^2.\end{aligned}$$

Se H è una qualunque matrice, siccome $|HAH^T| = |H|^2|A|$ si ha $\operatorname{Pf}(HAH^T) = \pm|H| \operatorname{Pf}(A)$ e quindi la relazione $\operatorname{Pf}(HAH^T) = |H| \operatorname{Pf}(A)$ è certamente vera se $\operatorname{Pf}(A) = 0$ oppure se $H = I$. ed il segno è certamente uguale a $+1$ quando $H = I$ oppure quando $1 = -1$. Per dimostrare che quando $\operatorname{Pf}(A) \neq 0$ e $1 \neq -1$ il segno non dipende da H scriviamo $C(t) = tH + (1-t)I$ ottenendo

$$\big(\operatorname{Pf}(C(t)AC(t)^T) - |C(t)| \operatorname{Pf}(A)\big) \big(\operatorname{Pf}(C(t)AC(t)^T) + |C(t)| \operatorname{Pf}(A)\big) = 0.$$

Come sopra almeno uno dei due polinomi in t si annulla identicamente e basta osservare che

$$\operatorname{Pf}(C(0)AC(0)^T) + |C(0)| \operatorname{Pf}(A) = 2 \operatorname{Pf}(A) \neq 0.$$

□

Esercizi.

472 (☕, ♡). Sia $A \in M_{n,n}(\mathbb{K})$ una matrice alternante di rango r. Dimostrare che:

(1) esiste un matrice $H \in M_{r,n}(\mathbb{K})$ tale che il prodotto $HA \in M_{r,n}(\mathbb{K})$ ha ancora rango r;
(2) per ogni matrice H come al punto precedente, la matrice $HAH^T \in M_{r,r}(\mathbb{K})$ ha ancora rango r;
(3) dedurre dai punti precedenti che r è pari.

CAPITOLO 9

Endomorfismi e polinomio caratteristico

Dopo aver studiato nel Capitolo 5 le applicazioni lineari tra due spazi vettoriali ed aver introdotto il determinante, iniziamo in questo capitolo lo studio delle applicazioni lineari da uno spazio vettoriale in sé, che costituisce la parte più profonda ed importante dell'algebra lineare. In questo studio, che proseguirà nei Capitoli 10, 11 e 14, gran parte dei risultati vale ed assume significato solamente per spazi vettoriali di dimensione finita; pertanto in questo capitolo ogni spazio vettoriale è assunto di norma, e salvo avviso contrario, di dimensione finita.

In particolare, ricordiamo dal Corollario 5.3.6 che se V è uno spazio vettoriale di dimensione finita e $f\colon V \to V$ è lineare, allora f è un isomorfismo se e solo se $\mathrm{Ker}(f) = 0$.

9.1. Matrici simili

Nei capitoli precedenti abbiamo studiato le matrici invertibili prese singolarmente; da adesso consideriamo anche l'insieme delle matrici invertibili di ordine fissato come un unico soggetto matematico.

Definizione 9.1.1. Il **gruppo lineare** $\mathrm{GL}_n(\mathbb{K})$ è l'insieme delle matrici $A \in M_{n,n}(\mathbb{K})$ che sono invertibili. Equivalentemente

$$\mathrm{GL}_n(\mathbb{K}) = \{A \in M_{n,n}(\mathbb{K}) \mid \det(A) \neq 0\}.$$

Abbiamo visto nella Sezione 6.3 che, se $A, B \in \mathrm{GL}_n(\mathbb{K})$ allora anche $A^{-1}, A^T, AB \in \mathrm{GL}_n(\mathbb{K})$. Si noti che $\mathrm{GL}_n(\mathbb{K})$ non contiene la matrice nulla e quindi non è un sottospazio vettoriale di $M_{n,n}(\mathbb{K})$.

Definizione 9.1.2. Diremo che due matrici quadrate $A, B \in M_{n,n}(\mathbb{K})$ sono **simili**, e scriveremo $A \sim B$, se esiste $C \in \mathrm{GL}_n(\mathbb{K})$ tale che $A = CBC^{-1}$.

La similitudine gode delle seguenti proprietà:

Proprietà riflessiva: $A \sim A$ per ogni $A \in M_{n,n}(\mathbb{K})$.
Proprietà simmetrica: Se $A \sim B$ allora $B \sim A$.
Proprietà transitiva: Se $A \sim B$ e $B \sim H$, allora $A \sim H$.

La verifica di tali proprietà è immediata: infatti $A = IAI^{-1}$; se $A = CBC^{-1}$ allora $B = C^{-1}A(C^{-1})^{-1}$; se $A = CBC^{-1}$ e $B = DHD^{-1}$ allora $A = (CD)H(CD)^{-1}$.

L'importanza della similitudine risiede nel fatto che spesso la risposta a molti problemi di algebra lineare non cambia se sostituiamo una matrice con un'altra ad essa simile. Tanto per fare un esempio, siano $A \in M_{n,n}(\mathbb{K})$, $H \in M_{m,m}(\mathbb{K})$ e consideriamo il problema di determinare se esiste una matrice $X \in M_{n,m}(\mathbb{K})$ di rango fissato k tale che $AX = XH$. La soluzione al problema non cambia se alla matrice A sostituiamo una matrice ad essa simile $B = CAC^{-1}$: infatti se poniamo $Y = CX$, le matrici X e Y hanno lo stesso rango,

$$AX = C^{-1}BCX = C^{-1}BY, \qquad C^{-1}YH = XH,$$

e quindi $AX = XH$ se e solo se $C^{-1}BY = C^{-1}YH$. Siccome C è invertibile ne segue che X è una soluzione di $AX = XH$ se e solo se Y è una soluzione di $BY = YH$. Un ragionamento simile mostra che possiamo sostituire H con una matrice simile.

In questo e nei prossimi capitoli dimostreremo che ogni matrice è simile ad un'altra di forma abbastanza semplice che quindi può essere usata per semplificare i conti e la teoria; introdurremo i concetti di matrice diagonalizzabile (simile ad una diagonale), triangolabile (simile ad una triangolare), ciclica (simile ad una matrice compagna) ecc. e studieremo le condizioni che rendono possibile o impossibile tali similitudini.

Iniziamo con alcuni semplici esempi di similitudine.

ESEMPIO 9.1.3. Le due matrici diagonali

$$A = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, \qquad B = \begin{pmatrix} b & 0 \\ 0 & a \end{pmatrix},$$

sono simili: infatti $A = CBC^{-1}$ dove

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

Più in generale due matrici diagonali ottenute l'una dall'altra mediante una permutazione degli elementi sulla diagonale sono simili. Infatti se indichiamo con $C_k \in M_{n,n}(\mathbb{K})$, $n \geq 2$, la matrice a blocchi

$$C_k = \begin{pmatrix} I_k & 0 & 0 \\ 0 & C & 0 \\ 0 & 0 & I_{n-k-2} \end{pmatrix}, \qquad \text{dove} \quad C = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

e I_k è la matrice identità di ordine k, allora per ogni matrice diagonale A la matrice CAC^{-1} è ancora diagonale ed è ottenuta da A scambiando tra loro il $k+1$-esimo ed il $k+2$-esimo coefficiente sulla diagonale. Basta adesso osservare che ogni permutazione di $\{1, \ldots, n\}$ si ottiene per composizione di trasposizioni che scambiano elementi adiacenti.

ESEMPIO 9.1.4. Ogni matrice triangolare superiore è simile ad una matrice triangolare inferiore. Più in generale ogni matrice $A = (a_{i,j}) \in M_{n,n}(\mathbb{K})$ è simile alla matrice ottenuta per "rotazione di 180 gradi" $B = (b_{i,j})$, dove $b_{i,j} = a_{n-i+1,n-j+1}$. Indicando con $C \in M_{n,n}(\mathbb{K})$ la matrice che ha tutti i coefficienti nulli tranne quelli sull'antidiagonale principale che sono uguali ad 1, vale la $BC = CA$; ad esempio si ha

$$\begin{pmatrix} 9 & 0 & 0 \\ 6 & 5 & 0 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 5 & 6 \\ 0 & 0 & 9 \end{pmatrix}.$$

In generale, BC si ottiene da B scambiando la colonna i con la colonna $n-i+1$, mentre CA si ottiene da A scambiando la riga i con la riga $n-i+1$, per ogni $i = 1, \ldots, n$.

ESEMPIO 9.1.5. Per ogni scalare $t \neq 0$ le matrici

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad \begin{pmatrix} a & bt^{-1} \\ ct & d \end{pmatrix},$$

sono simili. Infatti si ha

$$\begin{pmatrix} 1 & 0 \\ 0 & t \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & t \end{pmatrix}^{-1} = \begin{pmatrix} a & bt^{-1} \\ ct & d \end{pmatrix}.$$

TEOREMA 9.1.6. *Matrici simili hanno la stessa traccia, lo stesso determinante e lo stesso rango.*

DIMOSTRAZIONE. Siano A e $B = CAC^{-1}$ due matrici simili. Applicando la formula $\mathrm{Tr}(DE) = \mathrm{Tr}(ED)$ alle matrici quadrate $D = CA$ e $E = C^{-1}$ si ottiene

$$\mathrm{Tr}(B) = \mathrm{Tr}(CAC^{-1}) = \mathrm{Tr}(C^{-1}CA) = \mathrm{Tr}(A).$$

Per il teorema di Binet

$$\det(B) = \det(CAC^{-1}) = \det(C)\det(A)\det(C)^{-1} = \det(A).$$

Infine sappiamo che il rango di una matrice non viene cambiato dalla moltiplicazione per matrici invertibili, quindi il rango di $B = CAC^{-1}$ è uguale al rango di A. □

ESEMPIO 9.1.7. Le matrici

$$\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 4 & 1 \\ 0 & 1 \end{pmatrix},$$

non sono simili in quanto aventi traccia diversa, mentre le matrici

$$\begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix}, \quad \begin{pmatrix} 4 & 1 \\ 0 & 1 \end{pmatrix},$$

non sono simili in quanto hanno diverso determinante. Anche le matrici

$$\begin{pmatrix}1&1\\0&1\end{pmatrix},\quad \begin{pmatrix}1&0\\0&1\end{pmatrix},$$

non sono simili pur avendo stessa traccia, rango e determinante: infatti la matrice identità è simile solo a se stessa.

LEMMA 9.1.8. *Se $A\sim B$, allora $A-\lambda I\sim B-\lambda I$ per ogni scalare $\lambda\in\mathbb{K}$ e $A^h\sim B^h$ per ogni intero positivo h.*

DIMOSTRAZIONE. Se $A=CBC^{-1}$, per ogni $\lambda\in\mathbb{K}$ si ha

$$C(B-\lambda I)C^{-1}=CBC^{-1}-C(\lambda I)C^{-1}=A-\lambda I.$$

Dimostriamo per induzione su h che $A^h=CB^hC^{-1}$: per $h=1$ non c'è nulla da dimostrare, mentre per $h>1$, supponendo vero $A^{h-1}=CB^{h-1}C^{-1}$ si ottiene

$$A^h=AA^{h-1}=(CBC^{-1})(CB^{h-1}C^{-1})=CBB^{h-1}C^{-1}=CB^hC^{-1}.$$

□

ESEMPIO 9.1.9. Le matrici

$$A=\begin{pmatrix}1&1&0\\0&1&1\\0&0&2\end{pmatrix},\qquad B=\begin{pmatrix}1&0&1\\0&1&1\\0&0&2\end{pmatrix},$$

non sono simili pur avendo stessa traccia, stesso rango e stesso determinante. Infatti il rango di $A-I$ è 2, mentre il rango di $B-I$ è uguale ad 1.

ESEMPIO 9.1.10. Le matrici

$$A=\begin{pmatrix}0&1&0&0\\0&0&0&0\\0&0&0&1\\0&0&0&0\end{pmatrix},\qquad B=\begin{pmatrix}0&1&0&0\\0&0&1&0\\0&0&0&0\\0&0&0&0\end{pmatrix},$$

non sono simili pur avendo stessa traccia, stesso rango e stesso determinante. Infatti il rango di A^2 è 0, mentre il rango di B^2 è uguale ad 1.

Esercizi.

473. Usare il risultato dell'Esempio 6.5.4 per determinare le matrici che sono simili solamente a sé stesse.

474. Siano A,B due matrici simili, provare che la traccia di A^h è uguale alla traccia di B^h per ogni intero $h>0$.

475. Provare che

$$\begin{pmatrix}1&2\\3&4\end{pmatrix},\ \begin{pmatrix}4&3\\2&1\end{pmatrix};\qquad \begin{pmatrix}1&2&3\\4&5&6\\7&8&9\end{pmatrix},\ \begin{pmatrix}9&8&7\\6&5&4\\3&2&1\end{pmatrix};$$

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

sono tre coppie di matrici simili.

476. Dopo aver studiato e compreso l'Esempio 9.1.9 dire, motivando la risposta, se la due matrici

$$\begin{pmatrix}1&1&0\\-1&3&0\\0&0&1\end{pmatrix},\qquad \begin{pmatrix}2&0&0\\0&2&0\\0&0&1\end{pmatrix},$$

sono simili.

477. Tra tutte le 21 possibili coppie (A_i, A_j), $i \leq j$, delle seguenti 6 matrici, dire quali sono coppie di matrici simili e quali non lo sono:

$$A_1 = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix},$$

$$A_4 = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \quad A_5 = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}, \quad A_6 = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$

478. Mostrare che le matrici

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

non sono simili pur avendo lo stesso determinante e pur avendo A^h e B^h la stessa traccia per ogni $h > 0$.

479. Mostrare che per ogni $a \in \mathbb{K}$, $a \neq 0$, le matrici

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix},$$

sono simili.

480. Mostrare che per ogni $a \in \mathbb{K}$, $a \neq 0$, le matrici

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & a & a^2 \\ 0 & 1 & a \\ 0 & 0 & 1 \end{pmatrix},$$

sono simili.

481. Date tre matrici $A, B, C \in M_{n,n}(\mathbb{K})$, provare che $\mathrm{Tr}(ABC) = \mathrm{Tr}(BCA) = \mathrm{Tr}(CAB)$. Più in generale, date n matrici $A_1, \ldots, A_n \in M_{n,n}(\mathbb{K})$, $n \geq 2$, provare che vale la formula $\mathrm{Tr}(A_1 A_2 \cdots A_n) = \mathrm{Tr}(A_2 \cdots A_n A_1)$.

482. Siano

$$A = \begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 5 \\ -2 & 1 & 6 \\ 0 & 2 & 7 \end{pmatrix}.$$

Calcolare la traccia di ABA (riflettere sulle proprietà della traccia prima di mettersi a fare i conti). Cosa si può dire della traccia di $AB^{350}A$?

483. Siano $A, B \in M_{n,n}(\mathbb{K})$, con A matrice simmetrica e B matrice invertibile. Provare che la matrice ABB^T è simile ad una matrice simmetrica.

9.2. Spettro e polinomio caratteristico

Si definisce lo **spettro** di una matrice quadrata $A \in M_{n,n}(\mathbb{K})$ come l'insieme degli scalari $\lambda \in \mathbb{K}$ tali che il sistema lineare omogeneo

$$Ax = \lambda x, \qquad x \in \mathbb{K}^n = M_{n,1}(\mathbb{K}),$$

possiede soluzioni non banali. Gli elementi dello spettro vengono chiamati **autovalori** della matrice.

Per determinare lo spettro osserviamo che $Ax = \lambda x$ possiede soluzioni non banali il $\mathbb{K}^n$ se e solo se la matrice $A - \lambda I$ non è invertibile e quindi gli autovalori sono tutte e sole le soluzioni $\lambda \in \mathbb{K}$ dell'equazione, detta **equazione caratteristica** o, nei testi più antichi, **equazione secolare**:

$$\det(A - \lambda I) = 0.$$

Il ragionamento che abbiamo fatto nella sezione precedente mostra che matrici simili hanno lo stesso spettro, ossia gli stessi autovalori. Possiamo ridimostrare lo stesso fatto utilizzando

il teorema di Binet. Supponiamo di avere $B = CAC^{-1}$, con $\det(C) \neq 0$. Allora, siccome $I = CC^{-1} = CIC^{-1}$, per ogni $\lambda \in \mathbb{K}$ si ha:

$$\begin{aligned}\det(B - \lambda I) &= \det(CAC^{-1} - \lambda CIC^{-1}) = \det(C(A - \lambda I)C^{-1}) \\ &= \det(C)\det(A - \lambda I)\det(C^{-1}) = \det(A - \lambda I).\end{aligned}$$

Abbiamo già visto che lo spettro di una matrice A è l'insieme delle soluzioni dell'equazione caratteristica $\det(A - \lambda I) = 0$. Equivalentemente, introducendo una indeterminata t, lo spettro di A è l'insieme delle radici del polinomio $p_A(t) = \det(A - tI) \in \mathbb{K}[t]$, vedi Osservazione 8.1.10.

Definizione 9.2.1. Il polinomio $p_A(t) = \det(A - tI) \in \mathbb{K}[t]$ si chiama **polinomio caratteristico** della matrice $A \in M_{n,n}(\mathbb{K})$.

Esempio 9.2.2. Il polinomio caratteristico della matrice

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

è uguale a

$$p_A(t) = \begin{vmatrix} 1-t & 2 \\ 3 & 4-t \end{vmatrix} = (1-t)(4-t) - 6 = t^2 - 5t - 2.$$

Esempio 9.2.3. Il polinomio caratteristico della matrice

$$A = \begin{pmatrix} 1 & 2 & 8 \\ 0 & 3 & 4 \\ 0 & 0 & 5 \end{pmatrix}$$

è uguale a

$$p_A(t) = \begin{vmatrix} 1-t & 2 & 8 \\ 0 & 3-t & 4 \\ 0 & 0 & 5-t \end{vmatrix} = (1-t)(3-t)(5-t).$$

Se $A \in M_{n,n}(\mathbb{K})$, allora il determinante di $A - tI$ è un polinomio in t di grado n, e più precisamente

$$\det(A - tI) = (-1)^n t^n + (-1)^{n-1}\operatorname{Tr}(A)t^{n-1} + \cdots + \det(A).$$

In particolare, sia la traccia che il determinante di una matrice quadrata possono essere ricavati dal polinomio caratteristico. Infatti se

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

si ha

$$A - tI = \begin{pmatrix} a_{11} - t & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} - t \end{pmatrix}$$

e quindi, estraendo dalla sommatoria della formula (8.3) l'addendo relativo alla permutazione identità, si ha

$$\det(A - tI) = (a_{11} - t)\cdots(a_{nn} - t) + \sum \text{polinomi di grado} \leq n-2 \text{ in } t,$$

$$\det(A - tI) = (-t)^n + (a_{11} + \cdots + a_{nn})(-t)^{n-1} + \text{polinomio di grado} \leq n-2.$$

Infine il termine costante di $\det(A - tI)$ coincide con $\det(A - 0I) = \det(A)$.

Esempio 9.2.4. Il polinomio caratteristico di una matrice triangolare

$$\begin{pmatrix} a_{11} & * & ** & * \\ 0 & a_{22} & ** & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$$

è uguale a

$$p_A(t) = \begin{vmatrix} a_{11}-t & * & ** & * \\ 0 & a_{22}-t & ** & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn}-t \end{vmatrix} = (a_{11}-t)(a_{22}-t)\cdots(a_{nn}-t)$$

e quindi gli autovalori di A sono tutti e soli i coefficienti sulla diagonale.

ESEMPIO 9.2.5. Se $A \in M_{n,n}(\mathbb{K})$ è una matrice triangolare a blocchi, ossia

$$A = \begin{pmatrix} C & D \\ 0 & E \end{pmatrix}, \qquad C \in M_{p,p}(\mathbb{K}),\ E \in M_{n-p,n-p}(\mathbb{K})$$

allora

$$p_A(t) = \begin{vmatrix} C-tI & D \\ 0 & E-tI \end{vmatrix} = |C-tI||E-tI| = p_C(t)p_E(t).$$

ESEMPIO 9.2.6. Una matrice quadrata e la sua trasposta hanno lo stesso polinomio caratteristico. Infatti

$$p_{A^T}(t) = |A^T - tI| = |(A-tI)^T| = |A-tI| = p_A(t).$$

TEOREMA 9.2.7. *Il polinomio caratteristico è un invariante per similitudine, e cioè, se $A, B \in M_{n,n}(\mathbb{K})$ sono matrici simili, allora $p_A(t) = p_B(t) \in \mathbb{K}[t]$.*

DIMOSTRAZIONE. Se $B = CAC^{-1}$, allora

$$B - tI = CAC^{-1} - tCIC^{-1} = C(A-tI)C^{-1}$$

e quindi

$$\det(B-tI) = \det(C(A-tI)C^{-1}) = \det(C)\det(A-tI)\det(C)^{-1} = \det(A-tI).$$

□

ESEMPIO 9.2.8. L'uguaglianza del polinomio caratteristico è una condizione necessaria ma non sufficiente per la similitudine. Ad esempio, le matrici

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

hanno lo stesso polinomio caratteristico $p_A(t) = p_B(t) = t^2$ ma non sono simili. Infatti ogni matrice simile a B deve essere nulla.

Nell'Esempio 9.1.9 abbiamo usato il fatto che se $A, B \in M_{n,n}(\mathbb{K})$ sono simili, allora per ogni $\lambda \in \mathbb{K}$ le matrici $A - \lambda I$ e $B - \lambda I$ hanno lo stesso rango. Per verificare la validità o meno di tale condizione non è necessario calcolare i ranghi per ogni valore di λ ma è sufficiente restringere l'attenzione alle radici del polinomio caratteristico. È infatti chiaro dalla definizione che $p_A(\lambda) = 0$ se e solo se il rango di $A - \lambda I$ è minore di n.

ESEMPIO 9.2.9. Vogliamo dimostrare che le due matrici

$$A = \begin{pmatrix} 1 & 1 & -2 \\ 0 & 2 & -2 \\ 0 & 1 & -1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & 2 \\ 0 & -1 & -1 \end{pmatrix},$$

non sono simili, sebbene abbiamo lo stesso polinomio caratteristico

$$p_A(t) = p_B(t) = -t(t-1)^2$$

le cui radici sono $\lambda = 0$ e $\lambda = 1$. La teoria ci dice che per ogni $\lambda \neq 0, 1$ si ha

$$\operatorname{rg}(A - \lambda I) = \operatorname{rg}(B - \lambda I) = 3$$

ed un semplice conto, che lasciamo per esercizio al lettore, mostra che

$$\operatorname{rg}(A) = \operatorname{rg}(B) = 2, \qquad \operatorname{rg}(A - I) = 1, \quad \operatorname{rg}(B - I) = 2.$$

OSSERVAZIONE 9.2.10. In relazione al Lemma 9.1.8, per evitare inutili conteggi al lettore, anticipiamo il fatto (Esercizi 522 e 712) che se due matrici A, B hanno lo stesso polinomio caratteristico, allora le matrici A^h, B^h hanno lo stesso polinomio caratteristico per ogni intero positivo h.

Esercizi.

484. Calcolare i polinomi caratteristici delle seguenti matrici:

$$\begin{pmatrix}1&2\\1&2\end{pmatrix},\quad \begin{pmatrix}1&1&2\\0&1&2\\3&0&1\end{pmatrix},\quad \begin{pmatrix}2&0&2\\0&1&2\\0&0&1\end{pmatrix},\quad \begin{pmatrix}0&1&1\\1&0&1\\2&1&1\end{pmatrix},\quad \begin{pmatrix}1&0&1&0\\0&1&0&1\\1&0&-1&0\\0&-1&0&1\end{pmatrix}.$$

485. Date le due matrici:

$$A=\begin{pmatrix}1&2\\7&-1\end{pmatrix},\qquad B=\begin{pmatrix}2&1&0\\1&3&1\\0&1&2\end{pmatrix},$$

calcolare:

(1) le matrici A^2-15I e $-B^3+7B^2-14B+8I$;
(2) i polinomi caratteristici $p_A(t)$ e $p_B(t)$ di A e B rispettivamente.

486. Determinare gli autovalori della matrice

$$\begin{pmatrix}a&b&0\\b&a&b\\0&b&a\end{pmatrix}\qquad a,b\in\mathbb{R}.$$

487. Dimostrare che ogni matrice a coefficienti reali si può scrivere come somma di due matrici invertibili che commutano tra loro.

488. Sia $p(t)$ il polinomio caratteristico di una matrice antisimmetrica di ordine n. Provare che $p(-t)=(-1)^n p(t)$.

489. Provare che per ogni matrice $A\in M_{n,n}(\mathbb{K})$ si ha

$$p_{A^2}(t^2)=p_A(t)p_A(-t)\,.$$

490. Siano $A\in M_{n,n}(\mathbb{R})$ e $\lambda\in\mathbb{R}$ tali che $\det(A-\lambda I)\neq 0$. Provare che λ non è un autovalore della matrice $B=(I+\lambda A)(A-\lambda I)^{-1}$.

491. Sia $A\in M_{5,5}(\mathbb{C})$ tale che $p_{2A}(t)=p_A(t)+1$. Calcolare il determinante di A.

492. Ricordiamo che il commutatore di due matrici quadrate A,B è definito come $[A,B]=AB-BA$. Provare che per ogni intero positivo n valgono le formule

$$[A^n,B]=A^{n-1}[A,B]+[A^{n-1},B]A=\sum_{i=0}^{n-1}A^i[A,B]A^{n-i-1}\,,$$

dove A^0 è posta per convenzione uguale alla matrice identità. Dedurre che

$$\operatorname{Tr}(A+t[A,B])^n=\operatorname{Tr}A^n+t^2(\cdots)\,.$$

493 (☕). Siano $f,g\colon V\to W$ due applicazioni lineari tra spazi vettoriali di dimensione finita sul campo $\mathbb{K}$ e sia r il rango di f. Si assuma inoltre che il campo $\mathbb{K}$ contenga almeno $r+2$ elementi distinti. Provare che le seguenti condizioni sono equivalenti:

(1) $\operatorname{rg}(f+tg)\le r$ per ogni $t\in\mathbb{K}$;
(2) $\operatorname{rg}(f+tg)\le r$ per almeno $r+2$ valori distinti di $t\in\mathbb{K}$;
(3) $\operatorname{Ker}(f)\subseteq\operatorname{Ker}(g)$ oppure $g(V)\subseteq f(V)$.

9.3. Matrici compagne

Nella precedente sezione abbiamo associato ad ogni matrice quadrata un polinomio, monico a meno del segno, detto polinomio caratteristico. Viceversa, ad ogni polinomio monico possiamo associare in maniera canonica una matrice quadrata detta matrice compagna.

DEFINIZIONE 9.3.1. Una matrice quadrata della forma

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & a_n \\ 1 & 0 & \cdots & 0 & a_{n-1} \\ 0 & 1 & \cdots & 0 & a_{n-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & a_1 \end{pmatrix}$$

viene detta **matrice compagna** del polinomio

$$t^n - a_1 t^{n-1} - \cdots - a_{n-1} t - a_n \, .$$

Per esempio, sono compagne le matrici

$$(2), \qquad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix},$$

rispettivamente dei polinomi $t-2, t^2-1, t^2-2t-1, t^3-t^2-t-1$, mentre non sono compagne le matrici

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

PROPOSIZIONE 9.3.2. *Sia*

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & a_n \\ 1 & 0 & \cdots & 0 & a_{n-1} \\ 0 & 1 & \cdots & 0 & a_{n-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & a_1 \end{pmatrix} \tag{9.1}$$

la matrice compagna del polinomio monico $q(t) = t^n - a_1 t^{n-1} - \cdots - a_{n-1} t - a_n$. Allora, il polinomio caratteristico di A è uguale a

$$p_A(t) = (-1)^n (t^n - a_1 t^{n-1} - \cdots - a_{n-1} t - a_n) = (-1)^n q(t) \, .$$

DIMOSTRAZIONE. Per definizione

$$p_A(t) = \begin{vmatrix} -t & 0 & \cdots & 0 & a_n \\ 1 & -t & \cdots & 0 & a_{n-1} \\ 0 & 1 & \cdots & 0 & a_{n-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & a_1 - t \end{vmatrix}$$

e dallo sviluppo di Laplace rispetto alla prima riga si ottiene

$$p_A(t) = (-t) \begin{vmatrix} -t & 0 & \cdots & 0 & a_{n-1} \\ 1 & -t & \cdots & 0 & a_{n-2} \\ 0 & 1 & \cdots & 0 & a_{n-3} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & a_1 - t \end{vmatrix} + (-1)^{n+1} a_n \begin{vmatrix} 1 & -t & 0 & \cdots & 0 \\ 0 & 1 & -t & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{vmatrix} .$$

Per induzione su n, assumendo vero il risultato per le matrici compagne di ordine $n-1$, si ha

$$\begin{aligned} p_A(t) &= (-t)(-1)^{n-1}(t^{n-1} - a_1 t^{n-2} - \cdots - a_{n-1}) - (-1)^n a_n \\ &= (-1)^n (t^n - a_1 t^{n-1} - \cdots - a_{n-1} t - a_n). \end{aligned}$$

□

COROLLARIO 9.3.3. *Sia $p(t) \in \mathbb{K}[t]$ un polinomio di grado n. Esistono allora una matrice $A \in M_{n,n}(\mathbb{K})$ ed uno scalare $a \in \mathbb{K}$ tali che $p(t) = a p_A(t)$.*

DIMOSTRAZIONE. Se $p(t) = ct^n + \cdots$, $c \neq 0$, possiamo scrivere

$$\frac{p(t)}{(-1)^n c} = (-1)^n (t^n - a_1 t^{n-1} - \cdots - a_{n-1}t - a_n)$$

per opportuni coefficienti $a_1, \ldots, a_n \in \mathbb{K}$. È quindi sufficiente considerare $a = (-1)^n c$ e A la matrice compagna (9.1). □

Esercizi.

494. Scrivere le matrici compagne dei polinomi $t + 2$, $t^2 - t + 1$ e $t^3 + 4t - 3$.

495. Sia $(e_1, \ldots, e_n)$ la base canonica di $\mathbb{K}^n$ e sia $L_A \colon \mathbb{K}^n \to \mathbb{K}^n$ l'applicazione lineare associata ad una matrice quadrata A. Allora A è una matrice compagna se e solo se $L_A(e_i) = e_{i+1}$ per ogni $i = 1, \ldots, n-1$. Equivalentemente, A è compagna se e solo se $L_{A^i}(e_1) = e_{i+1}$ per ogni $i = 1, \ldots, n-1$.

496. Sia $A \in M_{n,n}(\mathbb{K})$ una matrice compagna di ordine n. Provare che:
(1) le n matrici $I, A, A^2, \ldots, A^{n-1}$ sono linearmente indipendenti in $M_{n,n}(\mathbb{K})$;
(2) se $B \in M_{n,n}(\mathbb{K})$ ha la prima colonna nulla e vale $AB = BA$, allora $B = 0$;
(3) le n matrici $I, A, A^2, \ldots, A^{n-1}$ generano il sottospazio delle matrici $B \in M_{n,n}(\mathbb{K})$ tali che $AB = BA$.

497. Sia A la matrice compagna del polinomio $t^n - a_1 t^{-1} - \cdots - a_n$. Dimostrare che $A^{n-1} \neq 0$ e che $A^n = 0$ se e solo se $a_1 = \cdots = a_n = 0$.

498. Si consideri la matrice compagna

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & a_n \\ 1 & 0 & \cdots & 0 & a_{n-1} \\ 0 & 1 & \cdots & 0 & a_{n-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & a_1 \end{pmatrix}$$

e denotiamo con $0 \leq r \leq n$ il più piccolo indice tale che $a_i = 0$ per ogni $r < i \leq n$. Provare che per ogni $m \geq n - r$ la matrice A^m ha rango r. In particolare, vale $A^m = 0$ per qualche intero $m > 0$ se e solo se i coefficienti a_i sono tutti nulli.

499. Provare che per ogni permutazione $\sigma \in \Sigma_n$ la matrice I^σ è simile ad una matrice diagonale a blocchi, in cui ciascun blocco sulla diagonale è la matrice compagna di un polinomio del tipo $t^m - 1$.

9.4. Endomorfismi ed autovalori

DEFINIZIONE 9.4.1. Sia V uno spazio vettoriale su $\mathbb{K}$. Un **endomorfismo** (lineare) di V è una qualsiasi applicazione lineare

$$f \colon V \longrightarrow V.$$

In dimensione finita, ogni endomorfismo lineare può essere rappresentato con una matrice secondo le regole descritte nella Sezione 5.5; siccome lo spazio di partenza coincide con lo spazio di arrivo possiamo scegliere come la stessa base di V sia nel dominio che nel codominio.

Dunque la matrice A che rappresenta un endomorfismo $f \colon V \to V$ in una base $(v_1, \ldots, v_n)$ è determinata dalla formula

$$(f(v_1), \ldots, f(v_n)) = (v_1, \ldots, v_n)A.$$

È importante ribadire ancora una volta che A rappresenta f nella *stessa base* in partenza ed in arrivo.

Se $x_1, \ldots, x_n$ sono le coordinate di un vettore v nella base $(v_1, \ldots, v_n)$, e se $y_1, \ldots, y_n$ sono le coordinate del vettore $f(v)$ nella stessa base, vale la formula:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Infatti si ha

$$f(v) = (v_1, \dots, v_n) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \qquad v = (v_1, \dots, v_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

e quindi

$$f(v) = (f(v_1), \dots, f(v_n)) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = (v_1, \dots, v_n) A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

In pratica, la colonna i-esima di A è formata dalle coordinate del vettore $f(v_i)$ nella base $(v_1, \dots, v_n)$.

Se $(w_1, \dots, w_n)$ è un'altra base e se scriviamo

$$(w_1, \dots, w_n) = (v_1, \dots, v_n) C$$

con C matrice invertibile $n \times n$, allora si ha

$$(v_1, \dots, v_n) = (v_1, \dots, v_n) C C^{-1} = (w_1, \dots, w_n) C^{-1},$$

e di conseguenza

$$(f(w_1), \dots, f(w_n)) = (f(v_1), \dots, f(v_n)) C = (v_1, \dots, v_n) AC = (w_1, \dots, w_n) C^{-1} AC.$$

Abbiamo quindi dimostrato il seguente lemma.

LEMMA 9.4.2. *Siano A, B le matrici che rappresentano un endomorfismo nelle basi $(v_1, \dots, v_n)$ e $(w_1, \dots, w_n)$ rispettivamente. Allora vale $A = CBC^{-1}$, dove C è la matrice "di cambio di base" determinata dalla relazione*

$$(w_1, \dots, w_n) = (v_1, \dots, v_n) C\,.$$

In particolare, cambiando base, la matrice che rappresenta un endomorfismo f si trasforma in una matrice ad essa simile.

Il lemma precedente ci permette, ad esempio, di *definire il determinante* $\det(f)$ di un endomorfismo $f\colon V \to V$ tramite la formula $\det(f) = |A|$, dove A è la matrice che rappresenta f in una qualunque base. Siccome il determinante è invariante per similitudine, il determinante di A non dipende dalla scelta della base.

Similmente possiamo definire la traccia $\operatorname{Tr}(f)$ ed il polinomio caratteristico $p_f(t)$ di un endomorfismo $f\colon V \to V$ tramite le formule $\operatorname{Tr}(f) = \operatorname{Tr}(A)$ e $p_f(t) = p_A(t) = |A - tI|$, dove A è la matrice che rappresenta f in una qualunque base.

Riepilogando, per **calcolare traccia, determinante e polinomio caratteristico** di un endomorfismo si calcolano le rispettive quantità per la matrice che rappresenta f in una qualunque base di V, che come detto più volte, deve essere la stessa in partenza ed in arrivo. Più in generale possiamo dire che **qualunque attributo delle matrici quadrate che sia invariante per similitudine definisce un attributo degli endomorfismi.**

ESEMPIO 9.4.3. Sia V lo spazio vettoriale dei polinomi di grado ≤ 2 nella variabile x; vogliamo calcolare traccia, determinante e polinomio caratteristico dell'endomorfismo $f\colon V \to V$ che trasforma il polinomio $q(x)$ nel polinomio $q(x+1)$. Prendendo come base di V la terna $(1, x, x^2)$ otteniamo

$$f(1) = 1, \qquad f(x) = x + 1, \qquad f(x^2) = (x+1)^2 = x^2 + 2x + 1,$$

e quindi f è rappresentato dalla matrice

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

Ne consegue che:

$$\det(f) = \det(A) = 1, \qquad \operatorname{Tr}(f) = \operatorname{Tr}(A) = 3, \qquad p_f(t) = p_A(t) = (1-t)^3.$$

ESEMPIO 9.4.4. Sia $(v_1, \dots, v_n)$ una base di uno spazio vettoriale V e consideriamo l'endomorfismo $f\colon V \to V$ definito dalle relazioni

$$f(v_i) = v_{i+1} \qquad 1 \le i \le n-1,$$
$$f(v_n) = a_1 v_n + \cdots + a_{n-1} v_2 + a_n v_1,$$

dove $a_1, \dots, a_n \in \mathbb{K}$ sono numeri qualunque. Tale endomorfismo è rappresentato nella base $(v_1, \dots, v_n)$ dalla matrice compagna

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & a_n \\ 1 & 0 & \cdots & 0 & a_{n-1} \\ 0 & 1 & \cdots & 0 & a_{n-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & a_1 \end{pmatrix}.$$

In particolare, per la Proposizione 9.3.2 il polinomio caratteristico di f è uguale a

$$p_f(t) = p_A(t) = (-1)^n (t^n - a_1 t^{n-1} - \cdots - a_{n-1} t - a_n).$$

Per concludere, elenchiamo alcune semplici ma utili osservazioni:

(1) Se A, B sono le matrici che rappresentano, rispettivamente, due endomorfismi f, g in una stessa base, allora $f+g$ è rappresentato dalla matrice $A+B$. Similmente λA rappresenta λf per ogni $\lambda \in \mathbb{K}$.
(2) L'applicazione **identità** $\mathrm{Id}\colon V \to V$ è rappresentata dalla matrice identità, qualunque sia la scelta della base.
(3) se A rappresenta $f\colon V \to V$ in una base fissata, allora $A - \lambda I$ rappresenta $f - \lambda I$ per ogni $\lambda \in \mathbb{K}$.

La nozione di autovalore di una matrice quadrata si estende senza fatica agli endomorfismi.

DEFINIZIONE 9.4.5. Siano V uno spazio vettoriale di dimensione finita sul campo $\mathbb{K}$ e $f\colon V \to V$ un endomorfismo lineare. Uno scalare $\lambda \in \mathbb{K}$ si dice un **autovalore** per f se l'endomorfismo $f - \lambda I\colon V \to V$ non è invertibile.

Dunque, se la matrice A rappresenta $f\colon V \to V$ in una base fissata, allora $A - \lambda I$ rappresenta $f - \lambda I$ e quindi λ è un autovalore per f se e solo se $\det(f - \lambda I) = 0$, se e solo se λ è una radice del polinomio caratteristico di f.

Ad ogni autovalore di un endomorfismo f si possono associare alcuni invarianti numerici, ciascuno dotato di significato algebrico e/o geometrico. I due principali sono riassunti nella prossima definizione.

DEFINIZIONE 9.4.6. Sia λ è un autovalore di un endomorfsmo $f\colon V \to V$:

(1) la molteplicità di λ come radice del polinomio caratteristico viene detta **molteplicità algebrica** dell'autovalore;
(2) la dimensione del nucleo di $f - \lambda I\colon V \to V$ viene detta **molteplicità geometrica** dell'autovalore.

Le due molteplicità sono generalmente diverse, tuttavia vale la disuguaglianza descritta nel prossimo lemma.

LEMMA 9.4.7. *La molteplicità algebrica di un autovalore è sempre maggiore od uguale alla molteplicità geometrica.*

DIMOSTRAZIONE. Sia λ è un autovalore di un endomorfsmo $f\colon V \to V$ e denotiamo rispettivamente con n e m le molteplicità algebrica e geometrica di λ. Osserviamo che $m > 0$ perché, siccome V ha dimensione finita, l'endomorfismo $f - \lambda I$ è iniettivo se e solo se è invertibile. Sia dunque $v_1, \dots, v_m$ una base di $\mathrm{Ker}(f - \lambda I)$ ed estendiamola ad una base $v_1, \dots, v_m, \dots, v_N$ di V. Per ogni $i = 1, \dots, m$ si ha $f(v_i) - \lambda v_i = 0$, ossia $f(v_i) = \lambda v_i$. Dunque nella base $v_1, \dots, v_N$ l'endomorfismo f si rappresenta con una matrice a blocchi

$$\begin{pmatrix} \lambda I_m & B \\ 0 & C \end{pmatrix}, \quad I_m \in M_{m,m}(\mathbb{K}),\ C \in M_{N-m,N-m}(\mathbb{K}).$$

Ma allora $p_f(t) = p_{\lambda I}(t) p_C(t) = (\lambda - t)^m p_C(t)$ e quindi m è minore od uguale alla molteplicità di λ come radice di $p_f(t)$. □

Esercizi.

500. Scrivere le matrici che rappresentano l'endomorfismo

$$f\colon \mathbb{K}^2 \to \mathbb{K}^2, \qquad f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+y \\ 2x-y \end{pmatrix},$$

nella base canonica e nella base $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Verificare che le due matrici ottenute hanno lo stesso polinomio caratteristico.

501. Sia $f\colon V \to V$ un endomorfismo. Provare che per ogni coppia di scalari distinti $\lambda, \mu \in \mathbb{K}$, $\lambda \neq \mu$, si ha $\operatorname{Ker}(f - \lambda I) \subseteq (f - \mu I)(V)$.

502 (♡). Sia $A \in M_{n,n}(\mathbb{K})$ tale che $BAB = 0$ per ogni matrice $B \in M_{n,n}(\mathbb{K})$ di rango 1. È vero che allora $A = 0$?

503 (Ⓐ). Sia $V = \mathbb{R}[x]_{\leq 3}$ lo spazio vettoriale dei polinomi di grado ≤ 3 a coefficienti reali. Calcolare il determinante dell'applicazione lineare $f\colon V \to V$ definita da

$$f(p(x)) = \frac{d^2p(x)}{d^2x} + 5\frac{dp(x)}{dx}.$$

504. Siano $f\colon V \to W$ e $g\colon W \to V$ due applicazioni lineari tra spazi vettoriali di dimensione finita. Provare che i due endomorfismi fg e gf hanno la stessa traccia e, se $\dim V = \dim W$, allora hanno anche lo stesso determinante. Descrivere un esempio in cui $\dim V \neq \dim W$ e $\det(fg) \neq \det(gf)$.

505. Sia A una matrice compagna invertibile. Mostrare che A^{-1} è simile ad una matrice compagna.

506. Sia V uno spazio vettoriale sul campo $\mathbb{K}$ e denotiamo con D l'insieme di tutte le coppie (h, f) tali che:

(1) $h\colon \mathbb{K} \to \mathbb{K}$, $f\colon V \to V$ sono applicazioni additive, ossia
$$h(0) = 0, \quad h(a+b) = h(a) + h(b), \qquad f(0) = 0, \quad f(v+u) = f(u) + f(v).$$
(2) $f(av) = af(v) + h(a)v$ per ogni $a \in \mathbb{K}$ ed ogni $v \in V$. In particolare f è lineare se e solo se $h = 0$.

Provare che per ogni (h, f) e $(k, g) \in D$, la coppia $(hk - kh, fg - gf)$ appartiene ancora a D.

9.5. Autovettori e sottospazi invarianti

Sia $f\colon V \to V$ un endomorfismo lineare di uno spazio vettoriale V di dimensione finita. Se $v \in V$ è un vettore non nullo tale che $f(v) = \lambda v$ allora λ è un autovalore. Infatti la condizione $f(v) = \lambda v$ è del tutto equivalente a $v \in \operatorname{Ker}(f - \lambda I)$ e abbiamo visto che $\operatorname{Ker}(f - \lambda I) \neq 0$ se e solo se λ è un autovalore.

Definizione 9.5.1. Siano V uno spazio vettoriale e $f\colon V \to V$ un endomorfismo lineare. Un **autovettore** per f è un vettore *non nullo* $v \in V$ tale che

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

per qualche $\lambda \in \mathbb{K}$. In questo caso si dice che v è un autovettore relativo all'autovalore λ di f.

Ricorda: gli autovettori non sono mai nulli.

Segue immediatamente dalle definizioni che $\operatorname{Ker}(f - \lambda I) - \{0\}$ coincide con l'insieme degli autovettori relativi all'autovalore λ.

Esempio 9.5.2. Se $f\colon V \to V$ non è iniettiva, allora ogni vettore non nullo del nucleo è un autovettore relativo all'autovalore 0.

Proposizione 9.5.3. *Un endomorfismo f di uno spazio vettoriale di dimensione finita V sul campo $\mathbb{K}$ possiede autovettori se e solo se il polinomio caratteristico $p_f(t)$ possiede radici nel campo $\mathbb{K}$.*

Dimostrazione. Abbiamo visto che gli autovalori sono tutte e sole le radici del polinomio caratteristico nel campo $\mathbb{K}$ e che non c'è autovalore senza autovettore, e viceversa. □

Esempio 9.5.4. Non tutti gli endomorfismi possiedono autovettori; si consideri ad esempio la rotazione in $\mathbb{R}^2$ di un angolo α che non sia multiplo intero di π radianti.

Nella base canonica di $\mathbb{R}^2$ la rotazione di un angolo α è rappresentata dalla matrice

$$\begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}$$

il cui polinomio caratteristico

$$\begin{vmatrix} \cos(\alpha) - t & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) - t \end{vmatrix} = t^2 - 2\cos(\alpha)t + 1$$

possiede radici reali se e solo se $\cos(\alpha) = \pm 1$.

Il concetto di autovettore di una matrice $A \in M_{n,n}(\mathbb{K})$ si definisce nel modo ovvio interpretando la matrice come un endomorfismo di $\mathbb{K}^n$. Equivalentemente, un vettore colonna $x \in \mathbb{K}^n$ è un autovettore per una matrice $A \in M_{n,n}(\mathbb{K})$ se $x \neq 0$ e se $Ax = \lambda x$ per qualche $\lambda \in \mathbb{K}$.

Esempio 9.5.5. Calcoliamo autovalori ed autovettori della matrice "dei segni"

$$A = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix} \in M_{3,3}(\mathbb{R}).$$

Il polinomio caratteristico

$$\begin{aligned} p_A(t) &= \begin{vmatrix} 1-t & -1 & 1 \\ -1 & 1-t & -1 \\ 1 & -1 & 1-t \end{vmatrix} \\ &= (1-t)\begin{vmatrix} 1-t & -1 \\ -1 & 1-t \end{vmatrix} + \begin{vmatrix} -1 & -1 \\ 1 & 1-t \end{vmatrix} + \begin{vmatrix} -1 & 1-t \\ 1 & -1 \end{vmatrix} \\ &= 3t^2 - t^3, \end{aligned}$$

ha come radici 0 (con moltelicità 2) e 3. Gli autovettori relativi a 0 corrispondono alle soluzioni non nulle del sistema lineare

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

mentre gli autovettori relativi a 3 corrispondono alle soluzioni non nulle del sistema lineare

$$\begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 3 \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Notiamo che i tre vettori

$$\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix},$$

formano una base di autovettori: i primi due relativi all'autovalore 0 ed il terzo relativo all'autovalore 3.

Abbiamo già visto che gli endomorfismi di uno spazio vettoriale V formano uno spazio vettoriale $\mathrm{Hom}(V, V)$ e che la composizione di due endomorfismi di V è ancora un endomorfismo di V.

Se $f, g \colon V \to V$ chiameremo semplicemente *prodotto* il prodotto di composizione e scriveremo fg per indicare $f \circ g$, ossia $fg(v) = f(g(v))$ per ogni vettore $v \in V$. Naturalmente, sono endomorfismi tutte le potenze di f:

$$f^k \colon V \longrightarrow V, \qquad k \geq 0,$$

dove si pone per convenzione $f^0 = I$ il morfismo identità, ossia $f^0(v) = v$ per ogni $v \in V$; con tale convenzione vale la formula $f^h f^k = f^{h+k}$ per ogni $h, k \geq 0$.

DEFINIZIONE 9.5.6. Siano $f\colon V \to V$ un endomorfismo e $U \subseteq V$ un sottospazio vettoriale. Diremo che U è un sottospazio **f-invariante**, o anche **invariante per** f, se $f(U) \subseteq U$, ossia se $f(u) \in U$ per ogni $u \in U$.

Chiaramente i sottospazi 0 e V sono invarianti per qualunque endomorfismo $f\colon V \to V$, e l'Esempio 9.5.4 mostra che per opportuni endomorfismi non vi sono altri sottospazi invarianti.

È del tutto evidente che se un sottospazio è invariante per un endomorfismo f, allora è invariante anche per ogni potenza f^k, $k \geq 0$.

ESEMPIO 9.5.7. Siano $f, g\colon \mathbb{K}^3 \to \mathbb{K}^3$ rappresentati nella base canonica e_1, e_2, e_3, rispettivamente dalle matrici

$$f = \begin{pmatrix} 1 & 2 & 0 \\ 3 & 4 & 0 \\ 0 & 0 & 5 \end{pmatrix}, \qquad g = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 4 & 1 \\ 0 & 0 & 5 \end{pmatrix}.$$

Allora i sottospazi $\mathrm{Span}(e_1, e_2)$ e $\mathrm{Span}(e_3)$ sono f-invarianti, mentre i sottospazi $\mathrm{Span}(e_1, e_2)$ e $\mathrm{Span}(e_1)$ sono g-invarianti.

ESEMPIO 9.5.8. Sia $f\colon V \to V$ un endomorfismo, allora per ogni $k \geq 0$ i sottospazi $\mathrm{Ker}(f^k)$ e $f^k(V)$ sono f-invarianti. Sia infatti $v \in \mathrm{Ker}(f^k)$, allora la formula

$$f^k(f(v)) = f^{k+1}(v) = f(f^k(v)) = f(0) = 0$$

prova che anche $f(v) \in \mathrm{Ker}(f^k)$. Similmente se $v \in f^k(V)$ vuol dire che esiste $w \in V$ tale che $v = f^k(w)$ e quindi

$$f(v) = f(f^k(w)) = f^k(f(w)) \in f^k(V).$$

Siccome $f^k f = f f^k = f^{k+1}$, il precedente esempio può essere visto come un caso particolare del seguente lemma.

LEMMA 9.5.9. *Siano $f, g\colon V \to V$ due endomorfismi che commutano tra loro, ossia tali che $fg = gf$. Allora* $\mathrm{Ker}(f)$ *e* $f(V)$ *sono sottospazi g-invarianti.*

DIMOSTRAZIONE. Sia $v \in \mathrm{Ker}(f)$, allora $f(v) = 0$ e quindi $f(g(v)) = g(f(v)) = g(0) = 0$ che implica $g(v) \in \mathrm{Ker}(f)$. Similmente, se $v = f(w)$ allora $g(v) = g(f(w)) = f(g(w)) \in f(V)$. □

ESEMPIO 9.5.10. Siano $f\colon V \to V$ un endomorfismo e $\lambda \in \mathbb{K}$ uno scalare, se $I\colon V \to V$ denota l'identità, allora si ha $f(f - \lambda I) = f^2 - \lambda f = (f - \lambda I)f$ e dal lemma precedente segue che $\mathrm{Ker}(f - \lambda I)$ e $(f - \lambda I)(V)$ sono sottospazi invarianti per f.

ESEMPIO 9.5.11. Vogliamo calcolare gli autovalori della matrice $A \in M_{n,n}(\mathbb{R})$ di coefficienti $a_{ij} = 1$ se $|i - j| = 1$ e $a_{ij} = 0$ altrimenti:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots \\ 1 & 0 & 1 & 0 & \dots \\ 0 & 1 & 0 & 1 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Dire che $(x_1, \dots, x_n)^T \neq 0$ è un autovettore con autovalore λ equivale a dire che

$$x_2 = \lambda x_1, \qquad x_{n-1} = \lambda x_n, \qquad x_{i-1} + x_{i+1} = \lambda x_i, \quad 1 < i < n. \tag{9.2}$$

Dalla prima delle odiatissime formule di prostaferesi

$$\sin(p) + \sin(q) = 2\cos\frac{p-q}{2}\sin\frac{p+q}{2},$$

ricaviamo che, fissando $\beta \in \mathbb{R}$, $0 < \beta < \pi$, e ponendo $x_i = \sin(i\beta)$ si ha $x_1 \neq 0$ e $x_{i-1} + x_{i+1} = 2\cos(\beta)x_i$ per ogni i; quindi le equazioni (9.2) sono verificate se e solo se $\sin((n+1)\beta) = 0$. Abbiamo quindi dimostrato che gli n numeri reali

$$2\cos\frac{k\pi}{n+1}, \qquad k = 1, \dots, n,$$

sono autovalori distinti della matrice A e per ovvi motivi di ordine e grado non esistono altri autovalori.

PROPOSIZIONE 9.5.12. *Siano $f\colon V \to V$ un endomorfismo e $U \subseteq V$ un sottospazio f-invariante, ossia tale che $f(U) \subseteq U$. Si consideri l'applicazione lineare $f_{|U}\colon U \to U$ ottenuta restringendo f al sottospazio U. Allora il polinomio caratteristico di $f_{|U}$ divide il polinomio caratteristico di f:*

$$p_f(t) = p_{f_{|U}}(t)q(t)\,.$$

DIMOSTRAZIONE. Sia $u_1, \dots, u_m$ una base di U e la si completi a una base $u_1, \dots, u_m, v_1, \dots, v_h$ di V. Poiché $f(U) \subseteq U$, la matrice di f in questa base è del tipo

$$A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix} \tag{9.3}$$

dove B è la matrice di $f_{|U}$ nella base $u_1, \dots, u_m$. Dunque:

$$p_f(t) = \det(A - tI) = \det(B - tI)\det(D - tI) = p_{f_{|U}}(t)\det(D - tI)\,.$$

□

Per uso futuro enunciamo e dimostriamo la versione "duale" della proposizione precedente.

LEMMA 9.5.13. *Siano $p\colon U \to V$ un'applicazione lineare surgettiva tra spazi vettoriali di dimensione finita, e $f\colon V \to V$, $g\colon U \to U$ due endomorfismi tali che $pg = fp$. Allora il polinomio caratteristico di f divide il polinomio caratteristico di g.*

DIMOSTRAZIONE. Siano m, n le dimensioni di U, V rispettivamente; allora $\operatorname{Ker} p$ è un sottospazio g-invariante di dimensione $m - n$. Scegliamo una base $u_1, \dots, u_m$ di U tale che $u_1, \dots, u_{m-n} \in \operatorname{Ker} p$; la matrice di g in questa base è del tipo

$$A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}, \quad B \in M_{m-n,m-n}(\mathbb{K}),\ D \in M_{n,n}(\mathbb{K}).$$

Osserviamo adesso che $p(u_{m-n+1}), \dots, p(u_m)$ è una base di V e che la condizione $f(p(u_i)) = p(g(u_i))$ per ogni $i > m-n$ equivale a dire che D è la matrice che rappresenta f nella suddetta base. Dunque:

$$p_g(t) = \det(A - tI) = \det(B - tI)\det(D - tI) = \det(B - tI)p_f(t).$$

□

Esercizi.

507. Siano e_1, e_2 la base canonica di $\mathbb{R}^2$, λ un numero reale. Dimostrare che la matrice

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

possiede un autovettore del tipo $e_1 + ae_2$, con $a \in \mathbb{R}$, se e solo se $\lambda \neq 1$.

508. Calcolare polinomio caratteristico, autovalori reali e rispettivi autovettori delle seguenti matrici:

$$\begin{pmatrix} 3 & 4 \\ 5 & 2 \end{pmatrix}, \quad \begin{pmatrix} 0 & 4 \\ -4 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 2 & 1 \\ -2 & 0 & 3 \\ -1 & -3 & 0 \end{pmatrix}, \quad \begin{pmatrix} 13 & 1 & -11 \\ 8 & 0 & -8 \\ 11 & -1 & -13 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 5 & 6 & -3 \\ -1 & 0 & 1 \\ 1 & 2 & -1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

509. Dire, motivando la risposta, per quali valori di $r \in \mathbb{Z}$ esiste una matrice $A \in M_{3,3}(\mathbb{R})$ di rango r che ha come autovalore complesso il numero $1 + i$.

510. Siano $A, B \in M_{n,n}(\mathbb{K})$. Dimostrare che AB e BA hanno gli stessi autovalori ma non necessariamente gli stessi autovettori. Dimostrare inoltre che se A è invertibile allora AB e BA hanno lo stesso polinomio caratteristico (questo vale anche se A e B sono entrambe non invertibili, ma è più difficile da dimostrare).

511. Siano $A \in M_{n,m}(\mathbb{K})$, $B \in M_{m,n}(\mathbb{K})$ e $k > 0$. Dimostrare che se $\lambda \neq 0$ è un autovalore di $(AB)^k$, allora è anche un autovalore di $(BA)^k$. Mostrare con un esempio che, se $n \neq m$, ciò non vale per l'autovalore nullo.

512. Sia $V \subseteq \mathbb{K}^4$ il sottospazio di equazione $x+y-z-t = 0$ e sia $f\colon \mathbb{K}^4 \to \mathbb{K}^4$ l'applicazione lineare tale che $f(x,y,z,t) = (x+x, y+x, z+x, t+x)$. Dimostrare che $f(V) \subseteq V$ e calcolare il polinomio caratteristico della restrizione $f_{|V}\colon V \to V$.

513. Siano $f\colon V \to V$ un endomorfismo ed $U \subseteq V$ un sottospazio vettoriale tale che $f(V) \subseteq U$. Dimostrare che U è un sottospazio f-invariante e che, se V ha dimensione finita, allora la traccia di f è uguale alla traccia della restrizione $f_{|U}\colon U \to U$.

514. Siano $f\colon V \to V$ un endomorfismo tale che $f^m = 0$ per qualche $m > 0$ e $v \in V$ un vettore. Detta $n \leq m$ la dimensione del sottospazio vettoriale generato da $v, f(v), f^2(v), \dots, f^{m-1}(v)$, dimostrare che gli n vettori $v, f(v), \dots, f^{n-1}(v)$ sono linearmente indipendenti.

515. Provare che la matrice dell'Esempio 9.5.11 è invertibile se e solo se n è pari.

516. Siano $a, b \in \mathbb{C}$. Usare il risultato dell'Esempio 9.5.11 per determinare una formula generale per il calcolo degli autovalori della matrice

$$\begin{pmatrix} a & b & 0 & 0 & \dots \\ b & a & b & 0 & \dots \\ 0 & b & a & b & \dots \\ 0 & 0 & b & a & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

517. 1) Determinare, se esiste, una base di $\mathbb{C}^n$ formata da autovettori della matrice compagna di $t^n - 1$.

2) Sia $R \subseteq M_{n,n}(\mathbb{C})$ il sottospazio vettoriale formato dalle matrici $A = (a_{ij})$ tali che $a_{ij} = a_{hk}$ ogniqualvolta $(i-j) - (h-k)$ è un multiplo intero di n. Calcolare la dimensione di R e provare che $AB = BA \in R$ per ogni $A, B \in R$.

3) Provare che tutte le matrici di R sono simultaneamente diagonalizzabili, ossia che esiste una matrice invertibile P tale che $P^{-1}AP$ è diagonale per ogni $A \in R$.

518. Un endomorfismo $f\colon V \to V$ si dice **semisemplice** se per ogni sottospazio f-invariante $U \subseteq V$ esiste un sottospazio f-invariante $W \subseteq V$ tale che $V = U \oplus W$. Dimostrare che:

(1) i multipli dell'identità sono semisemplici;
(2) l'endomorfismo

$$h\colon \mathbb{K}^2 \to \mathbb{K}^2, \qquad h\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 0 \end{pmatrix},$$

è semisemplice;
(3) l'endomorfismo

$$g\colon \mathbb{K}^2 \to \mathbb{K}^2, \qquad g\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ x \end{pmatrix},$$

non è semisemplice;
(4) (☕) se V ha dimensione finita su un campo di numeri (o più generalmente su un campo di caratteristica 0) e $f^p = I$ per qualche intero positivo p, allora f è semisemplice.

519 (♡). Siano $p, f\colon V \to V$ due endomorfismi lineari, con $p^2 = p$. Provare che $\operatorname{Ker} p$ è un sottospazio f-invariante se e solo se $pfp = pf$.

520. Siano $f, g\colon V \to V$ due endomorfismi lineari, con $fg = gf$ e V di dimensione finita. Si assuma che i polinomi caratteristici di f e g non abbiano fattori comuni di grado positivo. Dimostrare che $f - g$ è un isomorfismo.

9.6. Endomorfismi triangolabili

In questa e nella prossima sezione studieremo i criteri base per stabilire se un endomorfismo è triangolabile e/o diagonalizzabile. Ulteriori criteri saranno analizzati nei prossimi capitoli.

DEFINIZIONE 9.6.1. Sia V uno spazio vettoriale di dimensione finita. Un endomorfismo $f\colon V \to V$ si dice **triangolabile** se, in una base opportuna, si rappresenta con una matrice triangolare superiore. Una matrice triangolabile è una matrice quadrata simile ad una matrice triangolare superiore.

In altri termini, f è triangolabile se e solo se esiste una base $v_1, \dots, v_n$ tale che per ogni indice i il sottospazio $\operatorname{Span}(v_1, \dots, v_i)$ è f-invariante. Infatti, ciò significa che

$$f(v_i) \in \operatorname{Span}(v_1, \dots, v_i), \quad \text{per ogni} \quad i = 1, \dots, n,$$

cosicché la matrice di f nella base $v_1, \dots, v_n$ è triangolare superiore.

Se un endomorfismo $f\colon V \to V$ si rappresenta nella base $u_1, \dots, u_n$ con una matrice triangolare inferiore, allora f è triangolabile. Infatti basta considerare la base $v_i = u_{n-i+1}$, $i = 1, \dots, n$, per rappresentare f con una matrice triangolare superiore.

Una filtrazione crescente di k sottospazi vettoriali

$$0 = V_0 \subseteq V_1 \subseteq V_2 \subseteq \cdots \subseteq V_k \subseteq V \tag{9.4}$$

si dice una **bandiera** se $\dim V_i = i$ per ogni $i = 1, \dots, k$. La bandiera (9.4) è detta **completa** se $V_k = V$, o equivalentemente se $\dim V = k$.

Notiamo che ogni base $(v_1, \dots, v_n)$ determina in maniera canonica una bandiera completa

$$\operatorname{Span}(v_1) \subseteq \operatorname{Span}(v_1, v_2) \subseteq \cdots \subseteq V_i = \operatorname{Span}(v_1, \dots, v_i) \subseteq \cdots,$$

ed ogni bandiera completa è ottenuta in questo modo. Infatti se $V_0 \subseteq V_1 \subseteq V_2 \subseteq \cdots \subseteq V_n = V$ è una bandiera completa, basta scegliere un vettore $v_i \in V_i - V_{i-1}$ per ogni $i = 1, \dots, n$ per avere la base richiesta.

LEMMA 9.6.2. *Siano V uno spazio vettoriale di dimensione finita ed $f\colon V \to V$ un endomorfismo. Allora f è triangolabile se e solo se esiste una bandiera completa $V_0 \subseteq V_1 \subseteq V_2 \subseteq \cdots \subseteq V_n = V$ di sottospazi f-invarianti.*

DIMOSTRAZIONE. Basta osservare che f si triangolarizza in una base $v_1, \dots, v_n$ se e solo se $\operatorname{Span}(v_1, \dots, v_i)$ è un sottospazio f-invariante per ogni i. □

TEOREMA 9.6.3. *Siano V uno spazio vettoriale di dimensione finita ed $f\colon V \to V$ un endomorfismo lineare. Allora f è triangolabile se e solo se ogni sottospazio f-invariante non nullo possiede autovettori.*

DIMOSTRAZIONE. Indichiamo con n la dimensione di V. Supponiamo che f sia triangolabile e sia $0 \neq U \subseteq V$ un sottospazio f-invariante non nullo. Se $\dim V = n$ esiste una bandiera completa

$$0 = V_0 \subseteq V_1 \subseteq \cdots \subseteq V_n = V, \qquad \dim V_i = i,$$

di sottospazi f-invarianti e si hanno le disuguaglianze

$$\dim(V_i \cap U) \leq \dim(V_{i+1} \cap U) \leq \dim(V_i \cap U) + 1.$$

La prima disuguagliaza segue dal fatto che $V_i \cap U \subseteq V_{i+1} \cap U$, mentre per la seconda basta osservare che $V_i + U \subseteq V_{i+1} + U$ e dalla formula di Grassmann segue che

$$\begin{aligned} \dim(V_{i+1} \cap U) &= \dim V_{i+1} + \dim U - \dim(V_{i+1} + U) \\ &= \dim V_i + 1 + \dim U - \dim(V_{i+1} + U) \\ &\leq 1 + \dim V_i + \dim U - \dim(V_i + U) = \dim(V_i \cap U) + 1. \end{aligned}$$

In particolare, se i è il più piccolo indice tale che $V_i \cap U \neq 0$ allora il sottospazio $V_i \cap U$ è f-invariante e di dimensione 1. Dunque ogni vettore non nullo di $V_i \cap U$ è un autovettore.

Viceversa, dimostriamo per induzione sulla dimensione che se ogni sottospazio f-invariante non nullo possiede autovettori, allora f è triangolabile. Se V ha dimensione 0 non c'è nulla da dimostrare. Se $V \neq 0$ allora esiste un autovalore λ, ed il sottospazio $U = (f - \lambda I)(V)$ è f-invariante e proprio. Siccome ogni sottospazio invariante e non nullo di U possiede autovettori,

per l'ipotesi induttiva la restrizione di f ad U è triangolabile. Si può quindi trovare una base $u_1, \dots, u_m$ di U tale che

$$f_{|U}(u_i) = f(u_i) \in \operatorname{Span}(u_1, \dots, u_i), \quad \text{per ogni } i = 1, \dots, m.$$

Si completi ora la base $u_1, \dots, u_m$ di U a una base $u_1, \dots, u_m, u_{m+1}, \dots, u_n$ di V. In questa base f è in forma triangolare. Infatti per ogni indice $i > m$ vale

$$f(u_i) = f(u_i) - \lambda u_i + \lambda u_i = (f - \lambda I)u_i + \lambda u_i \in U + \operatorname{Span}(u_i),$$

e quindi a maggior ragione vale anche

$$f(u_i) \in \operatorname{Span}(u_1, \dots, u_m, u_{m+1}, \dots, u_i).$$

□

Se un endomorfismo f è triangolabile, il suo polinomio caratteristico è un prodotto di polinomi di primo grado e quindi f possiede tutti gli autovalori nel campo $\mathbb{K}$. Vale anche il vivecersa:

TEOREMA 9.6.4. *Sia $f\colon V \to V$ un endomorfismo di uno spazio vettoriale di dimensione finita su di un campo $\mathbb{K}$. Allora f è triangolabile se e solo se il suo polinomio caratteristico ha tutte le radici in $\mathbb{K}$, ossia se e solo se il suo polinomio caratteristico è un prodotto di polinomi di primo grado a coefficienti in $\mathbb{K}$.*

DIMOSTRAZIONE. Se f è triangolabile, ossia rappresentabile con una matrice triangolare, per l'Esempio 9.2.4 il suo polinomio caratteristico è prodotto di polinomi di primo grado.

Viceversa, se il suo polinomio caratteristico è prodotto di polinomi di primo grado, segue dalla Proposizione 9.5.12 e dal Corollario 7.1.7 che per ogni sottospazio f-invariante non nullo U il polinomio caratteristico della restrizione di f ad U ha grado positivo ed è un prodotto di polinomi di primo grado. In particolare la restrizione di f ad U possiede almeno un autovalore e quindi almeno un autovettore. □

Il Teorema 9.6.4 può essere enunciato in maniera equivalente dicendo che un endomorfismo $f\colon V \to V$ è triangolabile sul campo $\mathbb{K}$ se e solo se

$$p_f(t) = (\lambda_1 - t)^{\nu_1}(\lambda_2 - t)^{\nu_2} \cdots (\lambda_s - t)^{\nu_s}, \qquad \lambda_1, \dots, \lambda_s \in \mathbb{K} \text{ e } \lambda_i \neq \lambda_j \text{ se } i \neq j.$$

Dal fatto che $\lambda_i \neq \lambda_j$ per ogni $i \neq j$ segue che ν_i è la molteplicità algebrica dell'autovalore λ_i.

Un'altra formulazione equivalente del Teorema 9.6.4 dice che un endomorfismo di uno spazio vettoriale di dimensione n è triangolabile se e solo se la somma delle molteplicità algebriche dei suoi autovalori è uguale ad n.

COROLLARIO 9.6.5. *Sia $f\colon V \to V$ un endomorfismo triangolabile di uno spazio vettoriale di dimensione finita e sia $U \subseteq V$ un sottospazio f-invariante. Allora la restrizione $f_{|U}\colon U \to U$ di f ad U è triangolabile.*

DIMOSTRAZIONE. Basta osservare che ogni sottospazio f-invariante di U è anche un sottospazio f-invariante di V ed applicare il criterio del Teorema 9.6.3.

Alternativamente, per ipotesi esiste una bandiera completa $0 = V_0 \subseteq V_1 \subseteq \cdots \subseteq V_n = V$ di sottospazi f-invarianti. Poiché $V_i + U \subseteq V_{i+1} + U$ dalla formula di Grassmann segue

$$\begin{aligned}\dim(V_{i+1} \cap U) &= \dim V_{i+1} + \dim U - \dim(V_{i+1} + U) \\ &\leq \dim V_i + 1 + \dim U - \dim(V_i + U) = \dim(V_i \cap U) + 1.\end{aligned}$$

Considerando le intersezioni $V_i \cap U$ troviamo una filtrazione di sottospazi f-invarianti

$$0 = V_0 \cap U \subseteq V_1 \cap U \subseteq \cdots \subseteq V_n \cap U = U$$

che si riduce ad una bandiera completa dopo aver cancellato i doppioni.

□

Esercizi.

521. Delle seguenti matrici a coefficienti razionali,

$$\begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 1 \\ 3 & 2 \end{pmatrix}, \qquad \begin{pmatrix} 1 & -2 \\ 3 & 2 \end{pmatrix}, \qquad \begin{pmatrix} 1 & -2 & -1 \\ 3 & 2 & 5 \\ 0 & 2 & 2 \end{pmatrix},$$

dire quali sono triangolabili su $\mathbb{Q}$, quali su $\mathbb{R}$ e quali su $\mathbb{C}$.

522. Siano A una matrice quadrata triangolabile ed h un intero positivo. Provare che il polinomio caratteristico di A^h dipende solo da h e dal polinomio caratteristico di A.

523. Sia $A \in M_{n,n}(\mathbb{R})$ una matrice triangolabile con tutti gli autovalori negativi. Provare che se n è dispari, allora non esiste alcuna matrice $B \in M_{n,n}(\mathbb{R})$ tale che $B^2 = A$.

524. Sia $f\colon V \to V$ un endomorfismo tale che $f^2 = f^3$. Mostrare che per ogni sottospazio f-invariante non nullo $U \subseteq V$ gli endomorfismi $f\colon V \to V$ e $f - I\colon U \to U$ non possono essere entrambi invertibili e dedurre che f è triangolabile.

9.7. Endomorfismi diagonalizzabili

Fissato un endomorfismo $f\colon V \to V$, per ogni scalare $\lambda \in \mathbb{K}$ denotiamo

$$V_\lambda = \operatorname{Ker}(f - \lambda I).$$

Notiamo che V_λ è un sottospazio vettoriale che dipende da λ e da f. Tuttavia, per semplicità notazionale, la dipendenza da f rimane sottointesa. Siccome $v \in V_\lambda$ se e solo se $f(v) - \lambda v = 0$, si deduce immediatamente che λ è un autovalore se e solo se $V_\lambda \neq 0$.

Notiamo che ogni V_λ è un sottospazio f-invariante e che la restrizione $f\colon V_\lambda \to V_\lambda$ è uguale alla restrizione di λI.

Definizione 9.7.1. Se λ è un autovalore di un endomorfsmo $f\colon V \to V$, il sottospazio

$$V_\lambda = \operatorname{Ker}(f - \lambda I),$$

è detto **autospazio** relativo a λ.

In altri termini, l'autospazio relativo ad un autovalore è il sottospazio formato dal vettore nullo e da tutti gli autovalori corrispondenti. Inoltre, la molteplicità geometrica di un autovalore λ è uguale alla dimensione dell'autospazio corrispondente.

Lemma 9.7.2. *Sia $f\colon V \to V$ un endomorfismo lineare e siano $v_1, \dots, v_s \in V$ autovettori relativi ad autovalori distinti. Allora $v_1, \dots, v_s$ sono linearmente indipendenti.*

Dimostrazione. Induzione su s, non dovendo dimostrare nulla per $s = 1$. Supponiamo quindi il risultato vero per $s - 1$ autovettori relativi ad autovalori distinti.

Indichiamo con $\lambda_1, \dots, \lambda_s$ gli autovalori corrispondenti agli autovettori $v_1, \dots, v_s$; per ipotesi si ha $f(v_i) = \lambda_i v_i$ e $\lambda_i \neq \lambda_j$ per ogni $i \neq j$.

Data una combinazione lineare nulla $a_1 v_1 + \cdots + a_s v_s = 0$, applicando l'endomorfismo $f - \lambda_s I$ si ottiene

$$0 = (f - \lambda_s I)(a_1 v_1 + \cdots + a_s v_s) = a_1(\lambda_1 - \lambda_s)v_1 + \cdots + a_{s-1}(\lambda_{s-1} - \lambda_s)v_{s-1}.$$

Per l'ipotesi induttiva gli autovettori $v_1, \dots, v_{s-1}$ sono linearmente indipendenti e quindi

$$a_1(\lambda_1 - \lambda_s) = \cdots = a_{s-1}(\lambda_{s-1} - \lambda_s) = 0$$

e siccome $\lambda_i - \lambda_s \neq 0$ per ogni $i = 1, \dots, s-1$ si ottiene $a_1 = \cdots = a_{s-1} = 0$. Per finire la relazione $a_1 v_1 + \cdots + a_s v_s = 0$ si riduce a $a_s v_s = 0$ e siccome $v_s \neq 0$ si ha $a_s = 0$. □

Lemma 9.7.3. *Siano $\lambda_1, \dots, \lambda_s$ autovalori distinti di un endomorfismo f e si considerino i relativi autospazi $V_{\lambda_i} = \operatorname{Ker}(f - \lambda_i I)$. Allora esiste una decomposizione in somma diretta:*

$$V_{\lambda_1} + \cdots + V_{\lambda_s} = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_s} \subseteq V.$$

DIMOSTRAZIONE. Bisogna dimostrare che ogni vettore $w \in V_{\lambda_1} + \cdots + V_{\lambda_s}$ si scrive in modo unico come somma di vettori in ciascun V_{λ_i}. Basta quindi verificare che se

$$v_1 + \cdots + v_s = 0\,, \qquad v_i \in V_{\lambda_i}\,, \quad i = 1, \ldots, s,$$

allora $v_i = 0$ per ogni $i = 1, \ldots, s$. Questo segue immediatamente dal Lemma 9.7.2 applicato ai vettori v_i diversi da 0. □

Un endomorfismo $f\colon V \to V$ di uno spazio vettoriale di dimensione finita si dice **diagonalizzabile** se esiste una base rispetto alla quale f si rappresenta con una matrice diagonale. Equivalentemente $f\colon V \to V$ è diagonalizzabile se e solo se esiste una base di V fatta con autovettori per f.

TEOREMA 9.7.4. *Sia* $\dim V = n$ *e siano* $\lambda_1, \ldots, \lambda_s$ *gli autovalori nel campo* $\mathbb{K}$ *di un endomorfismo* $f\colon V \to V$. *Se indichiamo con* μ_i *la molteplicità geometrica di* λ_i, *allora* $\mu_1 + \cdots + \mu_s \leq n$ *e vale*

$$\mu_1 + \cdots + \mu_s = n$$

se e soltanto se f *è diagonalizzabile.*

DIMOSTRAZIONE. Per il Lemma 9.7.3 la somma dei sottospazi V_{λ_i} è diretta e per la formula di Grassmann il sottospazio $V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_s}$ ha dimensione $\mu_1 + \cdots + \mu_s$. Per finire basta osservare che f è diagonalizzabile se e solo se

$$V = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_s}$$

□

Siamo adesso in grado di determinare tre condizioni, la prima e la terza necessarie e sufficienti, la seconda solamente sufficiente (ma di più facile verifica), affinché un endomorfismo risulti diagonalizzabile.

COROLLARIO 9.7.5. *Siano* V *uno spazio vettoriale di dimensione finita sul campo* $\mathbb{K}$ *e* $f\colon V \to V$ *un endomorfismo lineare. Allora* f *è diagonalizzabile se e solo se il polinomio caratteristico di* f *si scrive come prodotto di polinomi di primo grado,*

$$p_f(t) = (\lambda_1 - t)^{\nu_1}(\lambda_2 - t)^{\nu_2} \cdots (\lambda_s - t)^{\nu_s}, \qquad \lambda_1, \ldots, \lambda_s \in \mathbb{K},\ \lambda_i \neq \lambda_j,$$

e per ogni autovalore λ_i *la molteplicità geometrica è uguale alla molteplicità algebrica* ν_i.

DIMOSTRAZIONE. Se in un'opportuna base l'endomorfismo f si rappresenta con una matrice diagonale

$$A = \begin{pmatrix} a_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{nn} \end{pmatrix}, \qquad n = \dim V,$$

allora il polinomio caratteristico di f è uguale a

$$p_f(t) = p_A(t) = (a_{11} - t) \cdots (a_{nn} - t) = (\lambda_1 - t)^{\nu_1}(\lambda_2 - t)^{\nu_2} \cdots (\lambda_s - t)^{\nu_s}\,,$$

dove ν_i è uguale al numero dei coefficienti a_{jj} uguali a λ_i. Per ogni scalare $\lambda \in \mathbb{K}$ la matrice $A - \lambda I$ è ancora diagonale e quindi il suo rango è uguale al numero dei coefficienti a_{ii} diversi da λ; equivalentemente, la molteplicità geometrica di ciascun autovalore λ_i è uguale a ν_i.

Viceversa, se $n = \dim V$ ed il polinomio caratteristico è come nell'enunciato, siccome il suo grado è n, ne segue che la somma delle molteplicità geometriche degli autovalori è n e per il Teorema 9.7.4 l'endomorfismo risulta diagonalizzabile. □

Grazie ai precedenti risultati, per determinare se un endomorfismo f è diagonalizzabile possiamo procedere nel modo seguente:

(1) si calcola il polinomio caratteristico $p_f(t)$;
(2) se il polinomio caratteristico non si fattorizza come prodotto di fattori lineari allora f non è diagonalizzabile;
(3) se $p_f(t) = (\lambda_1 - t)^{\nu_1}(\lambda_2 - t)^{\nu_2} \cdots (\lambda_s - t)^{\nu_s}$, allora f è diagonalizzabile se $\nu_i \leq \dim \operatorname{Ker}(f - \lambda_i I)$ per ogni i tale che $\nu_i > 1$ (questa condizione assicura che la somma delle molteplicità geometriche è almeno n, mentre per il Teorema 9.7.4 tale somma è sempre al più n).

COROLLARIO 9.7.6. *Siano V uno spazio vettoriale di dimensione finita n sul campo $\mathbb{K}$ e $f\colon V \to V$ un endomorfismo lineare. Se il polinomio caratteristico di f possiede n radici distinte sul campo $\mathbb{K}$, ossia se*

$$p_f(t) = (\lambda_1 - t)(\lambda_2 - t)\cdots(\lambda_n - t), \qquad \lambda_1, \ldots, \lambda_n \in \mathbb{K},\ \lambda_i \neq \lambda_j,$$

allora f è diagonalizzabile.

DIMOSTRAZIONE. Per definizione, ogni autovalore ha molteplicità geometrica positiva ed il corollario segue quindi immediatamente dal Teorema 9.7.4. □

TEOREMA 9.7.7. *Un endomorfismo $f\colon V \to V$ è diagonalizzabile se e solo se è triangolabile e per ogni autovalore λ i due endomorfismi $f - \lambda I$ e $(f - \lambda I)^2$ hanno lo stesso rango.*

DIMOSTRAZIONE. Osserviamo preliminarmente che per ogni autovalore λ si ha l'inclusione $\operatorname{Ker}(f - \lambda I) \subseteq \operatorname{Ker}((f - \lambda I)^2)$: infatti se $(f - \lambda I)v = 0$, a maggior ragione si ha

$$(f - \lambda I)^2 v = (f - \lambda I)(f(v) - \lambda v) = (f - \lambda I)(0) = 0.$$

Una implicazione è immediata: se f si rappresenta con una matrice diagonale, allora anche $f - \lambda I$ si rappresenta con una matrice diagonale per ogni λ e di conseguenza tutte le potenze $(f - \lambda I)^s$, $s > 0$, hanno stesso rango, stesso nucleo e stessa immagine.

Viceversa, supponiamo che f sia triangolabile e che $\operatorname{Ker}(f - \lambda I) = \operatorname{Ker}((f - \lambda I)^2)$ per ogni autovalore λ. Scegliamo un autovalore μ, denotiamo con $V_\mu = \operatorname{Ker}(f - \mu I)$ il corrispondente autospazio e consideriamo il sottospazio f-invariante $U = (f - \mu I)(V)$.

Per ogni $w \in U \cap V_\mu$ esiste $v \in V$ tale che $w = (f - \mu I)v$ e di conseguenza $(f - \mu I)^2 v = 0$. Siccome per ipotesi $\operatorname{Ker}(f - \mu I) = \operatorname{Ker}((f - \mu I)^2)$ si ha $v \in \operatorname{Ker}(f - \mu I)$, ossia $w = 0$. Dunque $U \cap V_\mu = 0$, $U + V_\mu = U \oplus V_\mu$ e per la formula di Grassmann $V = U \oplus V_\mu$.

Sia adesso $g = f_{|U}\colon U \to U$ la restrizione di f ad U, per il Corollario 9.6.5 l'endomorfismo g è triangolabile e per ogni autovalore λ si ha

$$\operatorname{Ker}(g - \lambda I) = \operatorname{Ker}(f - \lambda I) \cap U = \operatorname{Ker}((f - \lambda I)^2) \cap U = \operatorname{Ker}((g - \lambda I)^2)\,.$$

Per induzione sulla dimensione l'endomorfismo g è diagonalizzabile e di conseguenza anche f è diagonalizzabile.

□

ESEMPIO 9.7.8. Sia $f\colon V \to V$ un endomorfismo tale che $f^3 = f$, su di un campo di caratteristica diversa da 2. Allora f è diagonalizzabile. Infatti $(f - I)(f + I)f = 0$ e quindi per ogni sottospazio f-invariante $U \neq 0$, preso un qualunque vettore non nullo $u \in U$, $u \neq 0$, si ha $(f - I)(f + I)f(u) = 0$ e quindi:

(1) se $f(u) = 0$, allora u è un autovettore (con autovalore 0);
(2) se $f(u) \neq 0$ e $(f + I)f(u) = 0$, allora $f(u)$ è un autovettore (con autovalore -1);
(3) se $(f + I)f(u) \neq 0$, allora $(f + I)f(u)$ è un autovettore (con autovalore 1).

In ogni caso U possiede un autovettore e questo prova che f è triangolabile. La stessa formula $(f - I)(f + I)f = 0$ mostra che gli unici possibili autovalori di f sono $0, 1, -1$. Infatti se $f(v) = \lambda v$, $v \neq 0$, allora $0 = (f - I)(f + I)f(v) = (\lambda - 1)(\lambda + 1)\lambda v$ e quindi $\lambda = 0, \pm 1$. Per concludere basta osservare che

$$f \circ f^2 = f^3 = f \Rightarrow \operatorname{Ker}(f^2) \subseteq \operatorname{Ker}(f),$$
$$(f + 2I) \circ (f - I)^2 = -2(f - I) \Rightarrow \operatorname{Ker}((f - I)^2) \subseteq \operatorname{Ker}(f - I),$$
$$(f - 2I) \circ (f + I)^2 = -2(f + I) \Rightarrow \operatorname{Ker}((f + I)^2) \subseteq \operatorname{Ker}(f + I),$$

ed applicare il Teorema 9.7.7.

COROLLARIO 9.7.9. *Sia $f\colon V \to V$ un endomorfismo diagonalizzabile di uno spazio vettoriale di dimensione finita e sia $U \subseteq V$ un sottospazio f-invariante. Allora la restrizione di f ad U è diagonalizzabile.*

DIMOSTRAZIONE. Per il Corollario 9.6.5 la restrizione $g = f_{|U}\colon U \to U$ è triangolabile e come osservato nella dimostrazione del Teorema 9.7.7 per ogni autovalore λ si ha

$$\operatorname{Ker}(g - \lambda I) = \operatorname{Ker}(f - \lambda I) \cap U = \operatorname{Ker}((f - \lambda I)^2) \cap U = \operatorname{Ker}((g - \lambda I)^2).$$

□

Ritorneremo sull'Esempio 9.7.8 e sul Corollario 9.7.9 dopo aver sviluppato la teoria del polinomio minimo, che permetterà di ridimostrare i medesimi risultati in maniera più agile e veloce.

Esercizi.

525. Mostrare che due matrici diagonalizzabili sono simili se e solo se hanno lo stesso polinomio caratteristico.

526. Determinare, preferibilmente senza calcolare i polinomi caratteristici, una base di $\mathbb{R}^3$ formata da autovettori delle matrici:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

527. Determinare quali delle seguenti matrici sono diagonalizzabili su $\mathbb{R}$:

$$\begin{pmatrix} 1 & 2 & -1 \\ 1 & 0 & 1 \\ 4 & -4 & 5 \end{pmatrix}, \quad \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}, \quad \begin{pmatrix} -1 & -1 & 10 \\ -1 & -1 & 6 \\ -1 & -1 & 6 \end{pmatrix}, \quad \begin{pmatrix} 7 & -24 & -6 \\ 2 & -7 & -2 \\ 0 & 0 & 1 \end{pmatrix}.$$

528 (♡). Calcolare il polinomio caratteristico della matrice

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

e dire, motivando la risposta, se A è diagonalizzabile su $\mathbb{R}$ e su $\mathbb{C}$.

529. Dimostrare la seguente generalizzazione del Lemma 9.7.2: siano $f\colon V \to V$ un endomorfismo lineare, $U \subseteq V$ un sottospazio f-invariante e $v_1, \dots, v_s \in V$ autovettori relativi ad autovalori distinti. Se $v_1 + \cdots + v_s \in U$ allora $v_i \in U$ per ogni $i = 1, \dots, s$.

530. Per ogni matrice dell'Esercizio 508, dire se è diagonalizzabile su $\mathbb{Q}$, su $\mathbb{R}$ e su $\mathbb{C}$.

531. Sia $f\colon V \to V$ lineare di rango 1. Mostrare che f è diagonalizzabile se e solo se $f \circ f \neq 0$. Dedurre che le omologie lineari (Esercizio 275) sono diagonalizzabili, mentre le elazioni (ibidem) non lo sono.

532. *Melancolia I* è un'incisione di Albrecht Dürer del 1514 che simbolicamente rappresenta i pericoli dello studio ossessivo, mostrando le difficoltà che si incontrano nel tentativo di tramutare il piombo in oro, ed i conti in teoremi. Tra le altre figure, nell'opera compare anche un celebre quadrato magico i cui numeri sono riportati nella matrice

$$M = \begin{pmatrix} 16 & 3 & 2 & 13 \\ 5 & 10 & 11 & 8 \\ 9 & 6 & 7 & 12 \\ 4 & 15 & 14 & 1 \end{pmatrix}.$$

Dopo aver verificato che la somma dei 4 coefficienti su ogni riga, su ogni colonna, sulle due diagonali, su ciascuno dei quattro settori quadrati in cui si può dividere il quadrato è uguale a 34, così come la somma dei quattro numeri al centro e dei quattro numeri agli angoli, rispondere alle seguenti domande:

(1) il vettore $v_1 = (1, 1, 1, 1)^T \in \mathbb{R}^4$ è un autovettore?
(2) il sottospazio $U \subset \mathbb{R}^4$ di equazione $x_1 + x_2 + x_3 + x_4 = 0$ è invariante?
(3) indichiamo come al solito con $e_1, \dots, e_4$ la base canonica. Qual è la matrice che rappresenta l'endomorfismo L_M nella base

$$v_1 = e_1 + e_2 + e_3 + e_4, \quad v_2 = e_1 - e_4, \quad v_3 = e_2 - e_4, \quad v_4 = e_3 - e_4\,?$$

(4) la matrice M è diagonalizzabile su $\mathbb{R}$?

533 (♡). Un semplice gioco di magia da palcoscenico consiste nel farsi dire dal pubblico un numero intero n maggiore o uguale a 30 e scrivere all'istante un quadrato come nell'Esercizio 532 ma con le varie somme non più 34 ma uguali ad n; per contro non si richiede che il quadrato contenga tutti i numeri interi compresi tra 1 e 16. Scovate il trucco!

534. Siano $\xi_k = \cos(2k\pi/n)+i\sin(2k\pi/n) \in \mathbb{C}$, $k = 0, \dots, n-1$, le radici n-esime di 1, ossia le soluzioni dell'equazione $z^n = 1$. Usare il determinante di Vandermonde per provare che gli n vettori $v_k = (1, \xi_k, \xi_k^2, \dots, \xi_k^{n-1})^T$, $k = 0, \dots, n-1$, formano una base di $\mathbb{C}^n$. Caratterizzare, in funzione dei coefficienti, tutte e sole le matrici $C \in M_{n,n}(\mathbb{C})$ che possiedono $v_0, \dots, v_{n-1}$ come base di autovettori. (Si consiglia di fissare i coefficienti $c_1, \dots, c_n$ della prima riga di C e di determinare gli autovalori ed i rimanenti coefficienti in funzione di $c_1, \dots, c_n$.)

535. Provare che per ogni permutazione $\sigma \in \Sigma_n$ la matrice I^σ è diagonalizzabile su $\mathbb{C}$.

9.8. Il teorema fondamentale dell'algebra

Abbiamo visto che ogni polinomio di secondo grado a coefficienti complessi possiede sempre una radice complessa. L'obiettivo di questa sezione è generalizzare tale risultato a polinomi di qualsiasi grado positivo.

TEOREMA 9.8.1 (Teorema fondamentale dell'algebra). *Ogni polinomio di grado positivo a coefficienti complessi possiede radici complesse.*

Da tale risultato, applicando il teorema di Ruffini un numero finito di volte, si ricava che ogni polinomio a coefficienti complessi si scrive come prodotto di polinomi di primo grado.

La dimostrazione consiste nel dimostrare inizialmente due casi particolari: il primo, tipico dei corsi di analisi matematica, è che ogni polinomio a coefficienti reali di grado dispari possiede almeno una radice reale; il secondo è che ogni polinomio di secondo grado a coefficienti complessi possiede radici complesse. Poi, usando per intero tecniche di algebra lineare, dedurremo il caso generale da questi due casi particolari.

Per semplicità espositiva premettiamo alla dimostrazione una serie di risultati preliminari.

LEMMA 9.8.2. *Ogni polinomio di grado dispari a coefficienti reali possiede almeno una radice reale.*

DIMOSTRAZIONE. Si tratta di un risultato semplice e ben noto di analisi matematica che utilizza la continuità delle funzioni polinomiali; per completezza diamo un cenno di dimostrazione. Sia $p(t)$ un polinomio di grado dispari. A meno di moltiplicare $p(t)$ per una costante non nulla possiamo supporre senza perdita di generalità che il polinomio sia monico, e quindi

$$p(t) = a_0 t^n + a_1 t^{n-1} + \cdots + a_n \in \mathbb{R}[t], \qquad a_0 = 1,\ n = 2k+1.$$

Il polinomio $q(t) = -p(t)p(-t)$ è monico di grado pari $2n$ e quindi esiste $m > 0$ tale che $-p(m)p(-m) \geq 0$ (Esercizio: perché?), ossia $p(-m)p(m) \leq 0$. Per il Teorema 3.1.2 esiste $\xi \in [-m, m]$ tale che $p(\xi) = 0$. □

DEFINIZIONE 9.8.3. Sia k un intero positivo. Diremo che un campo $\mathbb{K}$ ha la proprietà P_k se ogni polinomio $p(t) \in \mathbb{K}[t]$ di grado positivo e non divisibile per 2^k possiede almeno una radice in $\mathbb{K}$.

Ad esempio, il Lemma 9.8.2 è del tutto equivalente a dire che il campo $\mathbb{R}$ ha la proprietà P_1.

LEMMA 9.8.4. *Un campo $\mathbb{K}$ ha la proprietà P_k se e solo se ogni matrice $A \in M_{n,n}(\mathbb{K})$, con n non divisibile per 2^k, possiede un autovettore.*

DIMOSTRAZIONE. Basta osservare che se $A \in M_{n,n}(\mathbb{K})$, allora il polinomio caratteristico di A ha grado n. Viceversa, per il Corollario 9.3.3, ogni polinomio di grado n è un multiplo scalare del polinomio caratteristico di una opportuna matrice quadrata di ordine n. □

Ricorda: per definizione, gli autovettori sono diversi da 0.

LEMMA 9.8.5. *Siano $\mathbb{K}$ un campo con la proprietà P_k, V uno spazio vettoriale su $\mathbb{K}$ di dimensione n e $f, g\colon V \to V$ due endomorfismi lineari. Se $fg = gf$ e 2^k non divide n, allora f e g hanno un autovettore in comune.*

DIMOSTRAZIONE. Il risultato è certamente vero per $n = 1$. Per induzione lo possiamo supporre vero per tutti gli interi $m < n$ che non sono divisibili per 2^k. Siccome i polinomi caratteristici di f e g hanno grado n, sia f che g possiedono autovalori in $\mathbb{K}$ ed autovettori in V. Se f è un multiplo dell'identità, allora ogni autovettore di g è un autovettore comune. Se f non è un multiplo dell'identità, scegliamo un autovalore $\lambda \in \mathbb{K}$ per f e consideriamo i due sottospazi propri $U = \operatorname{Ker}(f - \lambda I)$ e $W = \operatorname{Im}(f - \lambda I)$. Per il Lemma 9.5.9 i sottospazi U e W sono invarianti per f e g e, siccome $\dim U + \dim W = \dim V$, almeno uno di essi ha dimensione non divisibile per 2^k. Per l'ipotesi induttiva f e g possiedono un autovettore in comune in U oppure in W. □

LEMMA 9.8.6. *Il campo $\mathbb{C}$ ha la proprietà P_1.*

DIMOSTRAZIONE. Occorre dimostrare che se n è dispari, allora ogni matrice $A \in M_{n,n}(\mathbb{C})$ possiede un autovettore. Sia V lo spazio vettoriale reale delle matrici Hermitiane $n \times n$; ricordiamo che

$$\dim_{\mathbb{R}} V = n^2.$$

I due endomorfismi

$$f, g \colon V \to V, \qquad f(B) = \frac{AB + B\overline{A}^T}{2}, \qquad g(B) = \frac{AB - B\overline{A}^T}{2i},$$

sono $\mathbb{R}$-lineari e commutano tra loro. Siccome $\mathbb{R}$ ha la proprietà P_1 e n^2 è dispari, segue dal Lemma 9.8.5 che esiste $B \in V - \{0\}$ che è un autovettore comune per f e g. Se $f(B) = aB$ e $g(B) = bB$ con $a, b \in \mathbb{R}$, allora

$$(a + ib)B = f(B) + ig(B) = AB$$

ed ogni colonna non nulla di B è un autovettore per A. □

LEMMA 9.8.7. *Per ogni $k > 0$ il campo $\mathbb{C}$ ha la proprietà P_k.*

DIMOSTRAZIONE. Ragioniamo per induzione su k, avendo già dimostrato il lemma per $k = 1$. Supponiamo $k > 1$ e sia $A \in M_{n,n}(\mathbb{C})$ con n non divisibile per 2^k. Se 2^{k-1} non divide n allora per l'ipotesi induttiva A possiede un autovettore. Se 2^{k-1} divide n consideriamo il sottospazio vettoriale $V \subseteq M_{n,n}(\mathbb{C})$ delle matrici antisimmetriche. Notiamo che $\dim_{\mathbb{C}} V = \frac{n(n-1)}{2}$ non è divisibile per 2^{k-1}. Allora i due endomorfismi $f, g \colon V \to V$

$$f(B) = AB + BA^T, \qquad g(B) = ABA^T,$$

commutano tra loro e quindi possiedono un autovettore comune B, ossia

$$f(B) = \lambda B, \qquad g(B) = \mu B, \qquad \lambda, \mu \in \mathbb{C}.$$

Quindi

$$0 = \mu B - ABA^T = \mu B - A(f(B) - AB) = A^2 B - \lambda AB + \mu B = (A^2 - \lambda A + \mu I)B.$$

Se $\alpha, \beta \in \mathbb{C}$ sono le radici di $t^2 - \lambda t + \mu$ si ha

$$(A - \alpha I)(A - \beta I)B = 0.$$

Se $(A - \beta I)B = 0$ allora ogni colonna non nulla di B è un autovettore di A con autovalore β, altrimenti ogni colonna non nulla di $(A - \beta I)B$ è un autovettore di A con autovalore α. □

DIMOSTRAZIONE DEL TEOREMA 9.8.1. Sia $p(t) \in \mathbb{C}[t]$ un polinomio di grado $n > 0$; scegliamo un intero k tale che $2^k > n$ e usiamo il fatto che $\mathbb{C}$ ha la proprietà P_k. □

COROLLARIO 9.8.8. *Ogni polinomio a coefficienti complessi di grado $n > 0$ si può scrivere come prodotto di n polinomi di grado 1.*

DIMOSTRAZIONE. Conseguenza immediata del teorema di Ruffini e del teorema fondamentale dell'algebra. □

COROLLARIO 9.8.9. *Sul campo dei numeri complessi $\mathbb{C}$, ogni endomorfismo lineare è triangolabile.*

DIMOSTRAZIONE. Immediata conseguenza del Teorema 9.6.4 poiché per il teorema fondamentale dell'algebra ogni polinomio a coefficienti complessi si scrive come prodotto di polinomi di primo grado. □

Esercizi.

536. Verificare che gli endomorfismi f, g introdotti nelle dimostrazioni dei lemmi commutano tra loro.

537. Sia $a \in \mathbb{C}$ una radice di un polinomio $p(t) \in \mathbb{R}[t]$ a coefficienti reali. Provare che anche il coniugato $\overline{a}$ è una radice di $p(t)$, che $(t-a)(t-\overline{a}) \in \mathbb{R}[t]$ e che $p(t)$ si scrive come prodotto di polinomi reali di grado ≤ 2.

9.9. Complementi: similitudine complessa di matrici reali

Siano A, B due matrici $n \times n$ a coefficienti reali e supponiamo che siano simili sul campo dei numeri complessi, ossia supponiamo che esista una matrice invertibile $C \in M_{n,n}(\mathbb{C})$ tale che $A = CBC^{-1}$. Ci chiediamo se A e B sono simili anche sul campo dei numeri reali.

Siano F e G le matrici delle parti reali e immaginarie dei coefficienti di C, allora vale $F, G \in M_{n,n}(\mathbb{R})$, $C = F + iG$ e quindi

$$AF + iAG = A(F + iG) = AC = CB = (F + iG)B = FB + iGB.$$

Uguagliando parte reale ed immaginaria otteniamo le due uguaglianze

$$AF = FB, \qquad AG = GB.$$

Se F è invertibile si avrebbe $FBF^{-1} = AFF^{-1} = A$; similmente se G è invertibile si avrebbe $GBG^{-1} = A$ e quindi se almeno una tra F e G è invertibile abbiamo dimostrato che A e B sono simili sui reali. Purtroppo, può succedere che F e G siano entrambe non invertibili (vedi Esercizio 538). In tal caso, osserviamo che per ogni $\lambda \in \mathbb{R}$ si ha

$$A(F + \lambda G) = (F + \lambda G)B$$

e quindi basta dimostrare che esiste almeno un numero reale $\lambda \in \mathbb{R}$ tale che la matrice $F + \lambda G$ abbia determinante diverso da 0. Sia t un'indeterminata e consideriamo il polinomio

$$h(t) = |F + tG| \in \mathbb{R}[t].$$

Notiamo che $h(i) = |C| \neq 0$ e quindi $h(t)$ non è il polinomio nullo; se $\lambda \in \mathbb{R}$ è un qualunque numero reale tale che $h(\lambda) \neq 0$, allora $|F + \lambda G| = h(\lambda) \neq 0$.

Se al posto dei numeri reali consideriamo un qualsiasi sottocampo $\mathbb{K} \subseteq \mathbb{C}$, il precedente risultato è ancora vero, ma la dimostrazione diventa leggermente più complessa.

LEMMA 9.9.1. *Sia $\mathbb{K}$ un sottocampo di $\mathbb{C}$, e siano $A, B \in M_{n,n}(\mathbb{K})$, $C \in M_{n,n}(\mathbb{C})$ tre matrici. Si assuma che esista un numero complesso $\alpha \in \mathbb{C}$ tale che la matrice*

$$A + \alpha B + C$$

è invertibile. Allora esiste $\beta \in \mathbb{K}$ tale che la matrice

$$A + \beta B + C$$

è ancora invertibile.

DIMOSTRAZIONE. Sia t una indeterminata e si condideri il polinomio

$$p(t) = |A + tB + C| \in \mathbb{C}[t].$$

Il polinomio $p(t)$ ha grado $\leq n$ e non è nullo poiché $p(\alpha) \neq 0$. Dunque p possiede un numero finito di radici e basta prendere $\beta \in \mathbb{K}$ che non sia radice di $p(t)$. □

LEMMA 9.9.2. *Siano $\alpha_1, \ldots, \alpha_r \in \mathbb{C}$ linearmente indipendenti su $\mathbb{K}$ e siano $A_1, \ldots, A_r \in M_{n,n}(\mathbb{K})$ tali che*

$$\alpha_1 A_1 + \cdots + \alpha_r A_r = 0.$$

Allora $A_1 = \cdots = A_r = 0$.

DIMOSTRAZIONE. Basta osservare che ogni coefficiente della matrice $\alpha_1 A_1 + \cdots + \alpha_r A_r$ è una combinazione lineare su $\mathbb{K}$ dei numeri $\alpha_1, \ldots, \alpha_r$. □

TEOREMA 9.9.3. *Siano $A, B \in M_{n,n}(\mathbb{K})$. Se esiste una matrice $C \in M_{n,n}(\mathbb{C})$ invertibile tale che $AC = CB$, allora esiste una matrice $D \in M_{n,n}(\mathbb{K})$ invertibile tale che $AD = DB$.*

DIMOSTRAZIONE. Consideriamo $\mathbb{C}$ come spazio vettoriale su $\mathbb{K}$ e sia $V \subseteq \mathbb{C}$ il sottospazio vettoriale generato dai coefficienti di C; chiaramente V ha dimensione finita e minore od uguale a n^2. Sia $\alpha_1, \dots, \alpha_r$ una base di V, allora possiamo scrivere

$$C = \alpha_1 C_1 + \cdots + \alpha_r C_r$$

con $C_i \in M_{n,n}(\mathbb{K})$ per ogni i. Dunque

$$0 = AC - CB = \alpha_1(AC_1 - C_1B) + \cdots + \alpha_r(AC_r - C_rB)$$

e quindi $AC_i = C_iB$ per ogni i. Se C_i è invertibile per qualche i abbiamo finito. Altrimenti possiamo usare ripetutamente il Lemma 9.9.1 per dimostrare induttivamente che esistono $\beta_1, \dots, \beta_r \in \mathbb{K}$ tali che per ogni i la matrice

$$\beta_1 C_1 + \cdots + \beta_i C_i + \alpha_{i+1} C_{i+1} + \cdots + \alpha_r C_r$$

è invertibile. Alla fine possiamo prendere

$$D = \beta_1 C_1 + \cdots + \beta_r C_r.$$

□

COROLLARIO 9.9.4. *Sia $\mathbb{K}$ un campo di numeri. Due matrici $A, B \in M_{n,n}(\mathbb{K})$ sono simili in $M_{n,n}(\mathbb{K})$ se e solo se sono simili in $M_{n,n}(\mathbb{C})$.*

DIMOSTRAZIONE. Ovvia conseguenza del teorema precedente. □

Esercizi.

538. Trovare una matrice $A \in M_{2,2}(\mathbb{C})$ invertibile le cui parti reale ed immaginaria non siano invertibili.

539. Sia $C \in M_{n,m}(\mathbb{R})$. Dimostrare che:
(1) è possibile scrivere, in maniera non unica, $C = \sum_{i=1}^{r} a_i A_i$, dove $a_i \in \mathbb{R}$, $A_i \in M_{n,m}(\mathbb{Q})$ e $r \leq nm$.
(2) nella situazione del punto precedente, provare che se r è il minimo possibile, allora le matrici A_i sono linearmente indipendenti in $M_{n,m}(\mathbb{Q})$ ed i numeri a_i sono linearmente indipendenti su $\mathbb{Q}$.

540. Sia $A \in M_{n,n}(\mathbb{Q})$ una matrice a coefficienti razionali e sia r la dimensione su $\mathbb{Q}$ del sottospazio

$$V = \{B \in M_{n,n}(\mathbb{Q}) \mid AB = BA\}.$$

Sia $C \in M_{n,n}(\mathbb{R})$ tale che $AC = CA$. Dimostrare che esistono al più r coefficienti di C linearmente indipendenti su $\mathbb{Q}$.

CAPITOLO 10

Polinomio minimo

Dopo aver introdotto il polinomio caratteristico di un endomorfismo sotto forma di un attributo delle matrici quadrate invariante per similitudine, in questo capitolo affiancheremo ad ogni endomorfismo un altro polinomio, detto polinomio minimo, che risulterà molto utile nello studio delle applicazioni lineari. La definizione del polinomio minimo è semplice, naturale e geometrica ma non fornisce, tranne casi particolari, informazioni utili al calcolo del medesimo; fortunatamente, uno dei teoremi più importanti di tutta l'algebra lineare, il teorema di Cayley–Hamilton, ci dirà che il polinomio minimo divide il polinomio caratteristico, fornendoci così un formidabile metodo di calcolo.

Nella seconda parte del capitolo ci occuperemo di alcuni risultati che sono validi solamente sul campo $\mathbb{R}$ dei numeri reali. Proveremo in particolare che ogni matrice simmetrica reale si diagonalizza (su $\mathbb{R}$) e che è molto facile determinare se i suoi autovalori sono tutti positivi.

10.1. Il polinomio minimo

Sia V uno spazio vettoriale sul campo $\mathbb{K}$. Per un endomorfismo lineare $f\colon V \to V$ ha senso considerare le combinazioni lineari delle potenze di f, ossia gli endomorfismi lineari della forma

$$b_0 I + b_1 f + b_2 f^2 + \cdots + b_{k-1} f^{k-1} + b_k f^k \colon V \to V, \qquad k \geq 0,\ b_i \in \mathbb{K}.$$

Dunque, dato un qualsiasi polinomio

$$p(t) \in \mathbb{K}[t], \qquad p(t) = a_0 t^n + a_1 t^{n-1} + \cdots + a_n,$$

ha senso considerare l'endomorfismo

$$p(f)\colon V \to V, \qquad p(f) = a_0 f^n + a_1 f^{n-1} + \cdots + a_n I.$$

In questo modo abbiamo definito un'applicazione

$$\mathbb{K}[t] \to \operatorname{Hom}(V, V), \qquad p(t) \mapsto p(f),$$

che commuta con le operazioni di somma e prodotto, e cioè, per ogni coppia di polinomi $p, q \in \mathbb{K}[t]$ vale

$$(p+q)(f) = p(f) + q(f), \qquad pq(f) = p(f)q(f).$$

Se $f = L_A\colon \mathbb{K}^n \to \mathbb{K}^n$ per una opportuna matrice, si ha $f^n = L_{A^n}$ per ogni $n \geq 0$ e più in generale, per ogni polinomio $p(t) = a_0 + a_1 t + \cdots + a_k t^k \in \mathbb{K}[t]$, si ha $p(f) = L_{p(A)}$ $p(A) = a_0 I + a_1 A + \cdots + a_k A^k \in M_{n,n}(\mathbb{K})$. Ad esempio:

$$A = \begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix}, \quad p(t) = t^2 - t - 2, \quad \Rightarrow \quad p(A) = A^2 - A - 2I = \begin{pmatrix} 0 & 2 \\ 4 & 6 \end{pmatrix}.$$

LEMMA 10.1.1. *Sia $f\colon V \to V$ un endomorfismo e sia $p(t) \in \mathbb{K}[t]$. Se $v \in V$ è un autovettore per f con autovalore λ, allora v è anche un autovettore per $p(f)$ con autovalore $p(\lambda)$.*

DIMOSTRAZIONE. Se $f(v) = \lambda v$, allora

$$f^2(v) = f(f(v)) = f(\lambda v) = \lambda f(v) = \lambda^2 v.$$

Più in generale, si dimostra per induzione su k che $f^k(v) = \lambda^k v$: infatti

$$f^k(v) = f(f^{k-1}(v)) = f(\lambda^{k-1} v) = \lambda^{k-1} f(v) = \lambda^k v.$$

Quindi, se $p(t) = a_k t^k + \cdots + a_1 t + a_0$ si ha

$$p(f)(v) = a_k f^k(v) + \cdots + a_1 f(v) + a_0 v = (a_k \lambda^k + \cdots + a_1 \lambda + a_0) v = p(\lambda) v.$$

□

Se V ha dimensione finita, dal momento che

$$\dim \operatorname{Hom}(V,V) = (\dim V)^2\,,$$

non appena $k \geq (\dim V)^2$, i $k+1$ endomorfismi $f^k, f^{k-1}, \dots, f^2, f, I$ sono linearmente dipendenti in $\operatorname{Hom}(V,V)$ e quindi esiste una relazione lineare non banale

$$a_0 f^k + a_1 f^{k-1} + a_2 f^{k-2} + \cdots + a_{k-1} f + a_k I = 0, \qquad (a_0, \dots, a_k) \neq (0, \dots, 0).$$

In altri temini esiste un polinomio $p(t) = a_0 t^k + \cdots + a_k \in \mathbb{K}[t]$, $p(t) \neq 0$, tale che

$$p(f) = 0.$$

Viceversa, dato un qualunque polinomio non nullo $p(t) \in \mathbb{K}[t]$ tale che $p(f) = 0$, a meno di dividere $p(t)$ per il suo coefficiente direttivo, si può sempre assumere che un tale polinomio sia monico, e cioè che il coefficiente del suo termine di grado massimo sia uguale a 1:

$$p(t) = t^h + a_1 t^{h-1} + \cdots + a_h.$$

Abbiamo quindi provato che per ogni endomorfismo f di uno spazio vettoriale di dimensione finita, esiste almeno un polinomio monico $p(t)$ tale che $p(f) = 0$.

Tra tutti i polinomi monici (e quindi non nulli) che si annullano in f scegliamone uno di grado minimo. Un tale polinomio è unico perché se ce ne fossero due

$$p(t) = t^h + a_1 t^{h-1} + \cdots + a_h, \quad q(t) = t^h + b_1 t^{h-1} + \cdots + b_h, \quad p(f) = q(f) = 0,$$

dello stesso grado minimo h, posto

$$r(t) = p(t) - q(t) = ct^s + \cdots, \qquad c \neq 0,$$

risulterebbe $s < h$, $r(f) = p(f) = q(f) = 0$, ed il polinomio $r(t)/c$ sarebbe un polinomio monico di grado strettamente inferiore al grado di $p(t)$, che si annulla in f. Il che è assurdo a meno che non sia $q(t) = p(t)$.

Definizione 10.1.2. Il **polinomio minimo** $q_f(t) \in \mathbb{K}[t]$ di un endomorfismo $f\colon V \to V$ è il polinomio monico di grado minimo che si annulla in f.

Le precedenti considerazioni mostrano che se V ha dimensione finita allora il polinomio minimo esiste ed è unico. Osserviamo che, se $V \neq 0$, allora il polinomio minimo ha sempre grado maggiore di 0. Infatti l'unico polinomio monico di grado 0 è $p(t) = 1$ e quindi $p(f) = I \neq 0$.

Osservazione 10.1.3. La dimostrazione dell'esistenza del polinomio minimo di un endomorfismo $f\colon V \to V$ fornisce anche una dimostrazione del fatto che il grado di $q_f(t)$ è sempre minore od uguale a $(\dim V)^2$. Questa stima è ben lungi dall'essere ottimale e dimostreremo ben presto (Esercizio 553 e Corollario 10.2.3) che il grado del polinomio minimo è sempre minore od uguale alla dimensione $\dim V$ dello spazio vettoriale.

Esempio 10.1.4. Un endomorfismo $f\colon V \to V$, $V \neq 0$, è un multiplo dell'identità se e solo se il suo polinomio minimo ha grado 1. Infatti, siccome $I \neq 0$ il polinomio minimo deve avere grado positivo. Se $f = \lambda I$, allora $t - \lambda$ si annulla in f. Viceversa se $q_f(t) = t - \lambda$ è il polinomio minimo allora $0 = q_f(f) = f - \lambda I$ e dunque $f = \lambda I$.

In maniera del tutto simile si definisce il polinomio minimo di una matrice quadrata A, che coincide con il polinomio minimo dell'endomorfismo L_A.

Esempio 10.1.5. Calcoliamo il polinomio minimo $q_A(t)$ della matrice $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Dato che A non è un multiplo dell'identità, $q_A(t)$ deve avere grado > 1, mentre per l'Osservazione 10.1.3 $q_A(t)$ ha grado ≤ 2. Calcolare i coefficienti di $q_A(t) = t^2 + at + b$ significa trovare due scalari a, b tali che $A^2 + aA + bI = 0$, ossia

$$a \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + b \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a+b & 2a \\ 3a & 4a+b \end{pmatrix} = - \begin{pmatrix} 7 & 10 \\ 15 & 22 \end{pmatrix}$$

da cui si ricava $a = -5$, $b = -2$, $q_A(t) = t^2 - 5t - 2$.

TEOREMA 10.1.6. *Siano f un endomorfismo di uno spazio vettoriale di dimensione finita e $p(t)$ un polinomio tale che $p(f) = 0$. Allora il polinomio minimo di f divide $p(t)$, ossia esiste un polinomio $h(t) \in \mathbb{K}[t]$ tale che $p(t) = q_f(t)h(t)$.*

DIMOSTRAZIONE. Per la divisione euclidea tra polinomi esistono, e sono unici, due polinomi $h(t)$ e $r(t)$ tali che $p(t) = h(t)q_f(t) + r(t)$ e $\deg r(t) < \deg q_f(t)$. Poiché

$$0 = p(f) = h(f)q_f(f) + r(f) = r(f)$$

si deve avere $r(t) = 0$, per la minimalità del grado di $q_f(t)$. □

COROLLARIO 10.1.7. *Siano V spazio vettoriale di dimensione finita, $f\colon V \to V$ un endomorfismo e $U \subseteq V$ un sottospazio f-invariante, ossia tale che $f(U) \subseteq U$. Allora il polinomio minimo della restrizione $f_{|U}\colon U \to U$ divide il polinomio minimo di f.*

DIMOSTRAZIONE. Per il Teorema 10.1.6 basta dimostrare che $q_f(t)$ si annulla in $f_{|U}$. Per ogni vettore $u \in U$ si ha

$$0 = q_f(f)u = q_f(f_{|U})u$$

e quindi il polinomio q_f annulla l'endomorfismo $f_{|U}$. □

COROLLARIO 10.1.8. *Siano V di dimensione finita e $\lambda \in \mathbb{K}$ un autovalore di un endomorfismo $f\colon V \to V$. Allora $t - \lambda$ divide il polinomio minimo $q_f(t)$, ossia $q_f(\lambda) = 0$.*

DIMOSTRAZIONE. Per definizione di autovalore, l'autospazio $V_\lambda = \operatorname{Ker}(f - \lambda I)$ è diverso da 0 ed è un sottospazio f-invariante. Abbiamo visto nell'Esempio 10.1.4 che il polinomio minimo della restrizione di f a V_λ è uguale a $t - \lambda$ e per concludere basta applicare il Corollario 10.1.7.

Per una dimostrazione alternativa, sia v un autovettore con autovalore λ, allora v è anche un autovettore per $q_f(f)$ con autovalore $q_f(\lambda)$; ma allora $q_f(\lambda)v = q_f(f)v = 0$ e siccome $v \neq 0$ deve essere $q_f(\lambda) = 0$. □

LEMMA 10.1.9. *Sia $q_f(t) \in \mathbb{K}[t]$ il polinomio minimo di un endomorfismo $f\colon V \to V$ e sia $p(t) \in \mathbb{K}[t]$ un polinomio di grado positivo che divide $q_f(t)$. Allora l'endomorfismo $p(f)\colon V \to V$ non è invertibile.*

DIMOSTRAZIONE. Supponiamo per assurdo che l'endomorfismo $p(f)$ sia invertibile e scriviamo $q_f(t) = p(t)h(t)$. Allora

$$h(f) = p(f)^{-1}p(f)h(f) = p(f)^{-1}q_f(f) = 0$$

e siccome il grado di $h(t)$ è minore del grado di $q_f(t)$ otteniamo una contraddizione. □

TEOREMA 10.1.10. *Gli autovalori di un endomorfismo f di uno spazio di dimensione finita sono tutte e sole le radici in $\mathbb{K}$ del polinomio minimo di f.*

DIMOSTRAZIONE. Sia $q_f(t)$ il polinomio minimo e sia $\lambda \in \mathbb{K}$ un autovalore. Abbiamo dimostrato nel Corollario 10.1.8 che $q_f(\lambda) = 0$. Viceversa, se $q_f(\lambda) = 0$, per il teorema di Ruffini possiamo scrivere

$$q_f(t) = (t - \lambda)h(t)$$

e per il Lemma 10.1.9 l'endomorfismo $f - \lambda I$ non è invertibile, ossia λ è un autovalore. □

COROLLARIO 10.1.11. *Il polinomio minimo di un endomorfismo diagonalizzabile è il polinomio monico di grado minimo che annulla tutti gli autovalori. In altri termini se $\lambda_1, \dots, \lambda_s$ sono gli autovalori distinti di un endomorfismo diagnalizzabile $f\colon V \to V$, allora $q_f(t) = (t - \lambda_1) \cdots (t - \lambda_s)$.*

DIMOSTRAZIONE. Abbiamo già dimostrato che $q_f(t)$ annulla tutti gli autovalori. Viceversa, se $p(t)$ è un polinomio monico che annulla tutti gli autovalori, allora per ogni autovettore v con autovalore λ si ha $p(f)v = p(\lambda)v = 0$. Se esiste una base di autovettori allora $p(f) = 0$, per il Teorema 10.1.6 $q_f(t)$ divide $p(t)$ e quindi il grado di $q_f(t)$ è minore od uguale al grado di $p(t)$. □

Esercizi.

541. Calcolare i polinomi minimi delle matrici a coefficienti reali:

$$\begin{pmatrix}1 & 0\\ 0 & 3\end{pmatrix}, \quad \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}, \quad \begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}, \quad \begin{pmatrix}1 & 2\\ 0 & 1\end{pmatrix}, \quad \begin{pmatrix}0 & 1\\ 2 & 3\end{pmatrix}.$$

Verificare inoltre che tali polinomi hanno tutti grado 2 e calcolarne le radici complesse.

542. Sia $p(t)$ il polinomio minimo di una matrice $A \in M_{n,n}(\mathbb{C})$ antisimmetrica. Provare che il polinomio p è pari ($p(-t) = p(t)$) oppure dispari ($p(-t) = -p(t)$).

543. Si consideri una matrice $A \in M_{3,3}(\mathbb{C})$ antisimmetrica non nulla e sia k la somma dei quadrati dei tre coefficienti di A sopra la diagonale principale. Provare che il polinomio minimo di A è $t^3 + kt$.

544. Calcolare il polinimio minimo dell'endomorfismo

$$T\colon M_{n,n}(\mathbb{K}) \to M_{n,n}(\mathbb{K}), \qquad T(A) = A^T.$$

545. Sia f un endomorfismo tale che $f^2 = I$ e $f \neq \pm I$. Calcolare il polinomio minimo di f.

546. Siano $m, n > 0$ e $F \in M_{n,n}(\mathbb{K})$. Dimostrare che il polinomio minimo dell'endomorfismo

$$f\colon M_{n,m}(\mathbb{K}) \to M_{n,m}(\mathbb{K}), \qquad f(A) = FA,$$

è uguale al polinomio minimo di F. Cosa si può dire del polinomio caratteristico?

547. Siano V spazio vettoriale di dimensione $n > 1$ e $f\colon V \to V$ lineare di rango 1. Provare che in una base opportuna f si rappresenta con una matrice con le prime $n-1$ colonne nulle e dedurre che i polinomi minimo e caratteristico di f sono

$$q_f(t) = t(t - \operatorname{Tr}(f)), \qquad p_f(t) = (-1)^n t^{n-1}(t - \operatorname{Tr}(f)).$$

548. Consideriamo $\mathbb{C}[t]$ come uno spazio vettoriale su $\mathbb{C}$ e denotiamo con $f\colon \mathbb{C}[t] \to \mathbb{C}[t]$ l'endomorfismo dato dalla moltiplicazione per t: $f(p(t)) = tp(t)$. Dato $q(t) \in \mathbb{C}[t]$, descrivere in concreto l'endomorfismo $q(f)$ e dedurre che f non possiede polinomio minimo.

549. Sia $T \subseteq M_{n,n}(\mathbb{K})$ il sottospazio vettoriale delle matrici triangolari superiori. Sia $A \in T$ una matrice strettamente triangolare e denotiamo con $f\colon T \to T$ la moltiplicazione a sinistra per A, ossia $f(B) = AB$. Provare che il polinomio minimo di f è uguale a t^s per qualche intero $s \leq n$.

550. Provare che il grado del polinomio minimo di una matrice A è uguale al più piccolo intero k tale che le matrici $I, A, A^2, \ldots, A^k$ sono linearmente dipendenti.

551. Un polinomio si dice irriducibile se ha grado positivo e non può essere scritto come prodotto di due polinomi di grado positivo. Sia f un endomorfismo di uno spazio vettoriale V di dimensione finita e siano $p_1(t), \ldots, p_n(t) \in \mathbb{K}[t]$ polinomi monici irriducibili distinti. Provare per induzione su n che

$$\operatorname{Ker}(p_1(f)) + \cdots + \operatorname{Ker}(p_n(f)) = \operatorname{Ker}(p_1(f)) \oplus \cdots \oplus \operatorname{Ker}(p_n(f)).$$

Dedurre che esistono al più un numero finito di polinomi monici irriducibili che dividono il polinomio minimo di f.

552. Provare che il polinomio minimo è invariante per estensioni di campi. Più precisamente siano F è un campo, $\mathbb{K} \subseteq F$ è un sottocampo ed $A \in M_{n,n}(\mathbb{K})$ una matrice a coefficienti in $\mathbb{K}$. Denotiamo come al solito con $q_A(t) \in \mathbb{K}[t]$ il polinomio minimo di A e con $p(t) \in F[t]$ il polinomio minimo della stessa matrice, pensata come una matrice a coefficienti in F. Allora $p(t) = q_A(t)$. (Suggerimento: si tratta di applicare il fatto che il rango di una matrice non cambia per estensioni di campi: sia d il grado di $q_A(t)$, siccome $q_A(t) \in F[t]$ annulla A, ne segue che $p(t)$ divide $q_A(t)$ e basta dimostrare che d è minore od uguale al grado di $p(t)$, ossia che le matrici $I, A, \ldots, A^{d-1}$ sono linearmente indipendenti su F.)

553 (☕). Sia $f\colon V \to V$ un endomorfismo lineare di uno spazio vettoriale di dimensione finita n. I seguenti punti, svolti nell'ordine proposto, forniranno una dimostrazione del fatto che il polinomio minimo di f ha grado $\leq n$.

(1) Siano $U \subseteq V$ un sottospazio vettoriale di dimensione $m < n$ e $v \in V - U$. Provare che esiste un polinomio $p(t) \in \mathbb{K}[t]$ non nullo di grado $\leq n - m$ tale che $p(f)v \in U$.
(2) Siano $p(t) \in \mathbb{K}[t]$ un polimomio non nullo e $U \subseteq V$ un sottospazio f-invariante. Provare che
$$H = \{v \in V \mid p(f)v \in U\}$$
è un sottospazio f-invariante.
(3) Siano $U \subseteq H \subseteq V$ due sottospazi f-invarianti. Provare che esiste un polinomio $p(t) \in \mathbb{K}[t]$ non nullo di grado $\leq \dim H - \dim U$ tale che $p(f)(H) \subseteq U$. (Suggerimento: induzione su $\dim H - \dim U$, considerando separatamente i casi in cui esiste oppure non esiste un sottospazio invariante strettamente compreso tra U ed H.)
(4) Dedurre dal punto precedente che il polinomio minimo di f ha grado $\leq n$.

554. Siano V uno spazio vettoriale di dimensione n sul campo $\mathbb{K}$, $f\colon V \to V$ un endomorfismo con n autovalori distinti e $g\colon V \to V$ un endomorfismo che commuta con f, ossia tale che $gf = fg$. Provare che esiste un polinomio $p(t) \in \mathbb{K}[t]$ tale che $g = p(f)$.

555 (☕). Determinare l'insieme dei possibili autovalori delle matrici $A \in M_{n,n}(\mathbb{C})$ tali che $A^T = A^2 - I$. Cosa cambia se restringiamo l'attenzione alle matrici reali?

556 (☕, ♡). Siano $q_f(t) \in \mathbb{K}[t]$ il polinomio minimo di un endomorfismo f e $p(t) \in \mathbb{K}[t]$ un qualsiasi polinomio. Dimostrare che il grado del polinomio minimo di $p(f)$ è minore od uguale al grado di $q_f(t)$.

10.2. Il teorema di Cayley–Hamilton

In questa sezione, con V indicheremo sempre uno spazio vettoriale di dimensione finita sul campo $\mathbb{K}$.

LEMMA 10.2.1. *Sia $f\colon V \to V$ un endomorfismo che si rappresenta, in una opportuna base, con una matrice compagna. Allora*
$$p_f(t) = (-1)^{\dim V} q_f(t) \,.$$
In particolare $p_f(f) = 0$ ed il grado del polinomio minimo $q_f(t)$ è uguale alla dimensione di V.

DIMOSTRAZIONE. Sia $(v_1, \dots, v_n)$ una base di V nella quale f è rappresentato dalla matrice
$$\begin{pmatrix} 0 & 0 & \cdots & 0 & a_n \\ 1 & 0 & \cdots & 0 & a_{n-1} \\ 0 & 1 & \cdots & 0 & a_{n-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & a_1 \end{pmatrix}$$
Per quanto dimostrato nella Proposizione 9.3.2 il polinomio caratteristico è
$$p_f(t) = (-1)^n(t^n - a_1 t^{n-1} - \cdots - a_n).$$
Dobbiamo quindi dimostrare che il polinomio monico $p(t) = t^n - a_1 t^{n-1} - \cdots - a_n$ è uguale al polinomio minimo di f. Denotando per semplicità $v = v_1$, si ha $v_{i+1} = f^i(v)$ per ogni $i = 0, \dots, n-1$ e quindi
$$f^n(v) = f(f^{n-1}(v)) = f(v_n) = a_1 v_n + \cdots + a_n v_1 = a_1 f^{n-1}(v) + \cdots + a_n v.$$
Dunque
$$p(f)(v) = (f^n - a_1 f^{n-1} - \cdots - a_n I)v = f^n(v) - a_1 f^{n-1}(v) - \cdots - a_n v = 0,$$
e quindi, per ogni $i = 1, \dots, n$ si ha
$$p(f)(v_i) = p(f)(f^{i-1}(v)) = f^{i-1}(p(f)v) = f^{i-1}(0) = 0 \,.$$
Abbiamo dimostrato che $p(f)$ annulla tutti i vettori di una base e di conseguenza $p(f) = 0$. Per mostrare che $p(t) = q_f(t)$ bisogna mostrare che f non è annullato da alcun polinomio di grado $m < n$. Se $q(t) = \sum_{i=0}^{m} a_i t^i$, $a_m \neq 0$, si ha:
$$q(f)(v) = a_0 v + a_1 f(v) + \cdots + a_m f^m(v) \neq 0$$

in quanto i vettori $v, f(v), \dots, f^m(v)$ sono linearmente indipendenti; a maggior ragione $q(f)$ non è l'endomorfismo nullo. □

TEOREMA 10.2.2 (Cayley–Hamilton). *Sia $f\colon V \to V$ un endomorfismo lineare di uno spazio di dimensione finita. Allora*

$$p_f(f) = 0.$$

DIMOSTRAZIONE. Dobbiamo dimostrare che per ogni vettore $v \in V$ vale $p_f(f)(v) = 0$. Sia dunque $v \in V$ un vettore fissato; se $v = 0$ allora $p_f(f)(v) = 0$ per ovvi motivi, se invece $v \neq 0$ indichiamo con $k > 0$ il più grande intero tale che i k vettori $v, f(v), \dots, f^{k-1}(v)$ siano linearmente indipendenti. Dunque il vettore $f^k(v)$ appartiene alla chiusura lineare di $v, \dots, f^{k-1}(v)$, ossia esistono $a_1, \dots, a_k \in \mathbb{K}$ tali che

$$f^k(v) = a_1 f^{k-1}(v) + a_2 f^{k-2}(v) + \cdots + a_{k-1} f(v) + a_k v.$$

Completiamo i vettori $v, f(v), \dots, f^{k-1}(v)$ ad una base $v_1, \dots, v_n$ di V tale che

$$v_1 = v, \quad v_2 = f(v), \quad \dots, \quad v_k = f^{k-1}(v).$$

Allora vale $f(v_i) = v_{i+1}$ per $i < k$ e

$$f(v_k) = f^k(v) = a_1 v_k + a_2 v_{k-1} + \cdots + a_{k-1} v_2 + a_k v_1.$$

La matrice di f in questa base è del tipo

$$\begin{pmatrix} A & C \\ 0 & B \end{pmatrix}$$

dove A è la matrice compagna

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & a_k \\ 1 & 0 & \cdots & 0 & a_{k-1} \\ 0 & 1 & \cdots & 0 & a_{k-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & a_1 \end{pmatrix}$$

Per quanto visto nell'Esempio 9.2.5 e nella Proposizione 9.3.2 vale

$$p_f(t) = p_B(t) p_A(t) = p_B(t)\ (-1)^k (t^k - a_1 t^{k-1} - \cdots - a_k)$$

e di conseguenza

$$p_f(f)(v) = (-1)^k p_B(f)(f^k(v) - a_1 f^{k-1}(v) - \cdots - a_k v) = (-1)^k p_B(f)(0) = 0.$$

□

COROLLARIO 10.2.3. *Sia $f\colon V \to V$ un endomorfismo lineare di uno spazio vettoriale di dimensione finita. Allora il polinomio minimo $q_f(t)$ divide il polinomio caratteristico $p_f(t)$. In particolare $q_f(t)$ ha grado $\leq \dim V$.*

DIMOSTRAZIONE. Per il teorema di Cayley–Hamilton il polinomio caratteristico annulla l'endomorfismo e si può applicare il Teorema 10.1.6. □

TEOREMA 10.2.4. *Sia $f\colon V \to V$ un endomorfismo lineare di uno spazio vettoriale di dimensione n con polinomio minimo $q_f(t)$. Allora il polinomio caratteristico $p_f(t)$ divide $q_f(t)^n$.*

DIMOSTRAZIONE. La dimostrazione è molto simile a quella del teorema di Cayley–Hamilton, ma leggermente più complessa. Scriviamo il polinomio minimo come $q_f(t) = t^d - a_1 t^{d-1} - \cdots - a_d$ e consideriamo la sua matrice compagna

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & a_d \\ 1 & 0 & \cdots & 0 & a_{d-1} \\ 0 & 1 & \cdots & 0 & a_{d-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & a_1 \end{pmatrix},$$

della quale conosciamo il polinomio caratteristico $p_A(t) = (-1)^d q_f(t)$. Indichiamo come al solito con $e_1, \dots, e_d$ la base canonica di $\mathbb{K}^d$.

Per ogni vettore $v \in V$ consideriamo la seguente applicazione lineare

$$p_v \colon \mathbb{K}^d \to V, \qquad p_v(e_i) = f^{i-1}(v), \quad i = 1, \dots, d,$$

dove come al solito si intende $f^0(v) = v$. Si osserva che $fp_v = p_v L_A$; infatti per ogni $i = 1, \dots, d-1$ vale $fp_v(e_i) = f(f^{i-1}(v)) = f^i(v)$ e $p_v L_A(e_i) = p_v(e_{i+1}) = f^i(v)$, mentre

$$p_v L_A(e_d) = \sum_{i=1}^{d} p_v(a_i e_{d+1-i}) = \sum_{i=1}^{d} a_i f^{d-i}(v)$$
$$fp_v(e_d) = f^d(v),$$

e l'uguaglianza tra le due espressioni segue da fatto che $f^d = \sum_{i=1}^{d} a_i f^{d-i}$ in virtù dell'espressione del polinomio minimo. Sia adesso $v_1, \dots, v_k \in V$ una successione di $k \leq n$ vettori tali che $\operatorname{Im}(p_{v_1}) + \cdots + \operatorname{Im}(p_{v_k}) = V$; una tale successione esiste sempre, ad esempio una qualsiasi base di V, dato che $v \in \operatorname{Im}(p_v)$ per ogni v.

Consideriamo adesso lo spazio vettoriale $U = \underbrace{\mathbb{K}^d \times \cdots \times \mathbb{K}^d}_{k \text{ fattori}}$ e le applicazioni lineari

$$p \colon U \to V, \qquad p(u_1, \dots, u_k) = p_{v_1}(u_1) + \cdots + p_{v_k}(u_k),$$
$$g \colon U \to U, \qquad g(u_1, \dots, u_k) = (L_A(u_1), \dots, L_A(u_k)).$$

Allora p è surgettiva e $pg = fp$. Per il Lemma 9.5.13, il polinomio caratteristico di f divide il polinomio caratteristico di g che è $p_A(t)^k = (-1)^{dk} q_f(t)^k$. □

Siccome ogni polinomio non nullo è divisibile solamente per un numero finito di polinomi monici (Esercizio 380), abbiamo a disposizione un metodo di calcolo del polinomio minimo: a tal fine consideriamo i polinomi monici che dividono il polinomio caratteristico e la cui n-esima potenza è divisa dal polinomio caratteristico. Tra questi prendiamo quello di grado minimo che annulla l'endomorfismo.

Esempio 10.2.5. Per il teorema di Cayley–Hamilton il polinomio minimo $q_A(t)$ della matrice

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

divide il polinomio caratteristico $p_A(t) = (1-t)^3$ e quindi (vedi Corollario 7.1.7) deve essere uno tra i seguenti tre:

$$t - 1, \qquad (t-1)^2, \qquad (t-1)^3.$$

Partendo da quello di grado più basso, determiniamo quali di loro si annullano in A. Siccome

$$A - I = \begin{pmatrix} 0 & 1 & -1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad (A-I)^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

non rimane che l'ultima possibilità, ossia $q_A(t) = (t-1)^3$.

Esempio 10.2.6. Per il teorema di Cayley–Hamilton il polinomio minimo $q_B(t)$ della matrice

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

divide il polinomio caratteristico $p_B(t) = (1-t)^2(2-t)$ e quindi deve essere uno tra i seguenti:

$$t - 1, \qquad t - 2, \qquad (t-1)^2, \qquad (t-1)(t-2), \qquad (1-t)^2(2-t).$$

D'altra parte sappiamo che ogni autovalore è una radice del polinomio minimo e quindi la scelta si riduce a

$$(t-1)(t-2), \qquad (1-t)^2(2-t)\,.$$

Abbiamo

$$B - I = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \qquad B - 2I = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad (B-I)(B-2I) = 0,$$

e quindi $q_B(t) = (t-1)(t-2)$.

ESEMPIO 10.2.7. Sia $f\colon V \to V$ un endomorfismo di uno spazio vettoriale di dimensione n e denotiamo $C(f) = \{g \in \mathrm{Hom}(V,V) \mid fg = gf\}$; si verifica immediatamente che $C(f)$ è un sottospazio vettoriale che contiene tutte le potenze di f; questo implica immediatamente che la dimensione di $C(f)$ è maggiore od uguale al grado del polinomio minimo di f.

Gli argomenti usati nella dimostrazione del teorema di Cayley–Hamilton possono essere usati per dimostrare che $\dim C(f) \geq n$, anche nel caso in cui il polinomio minimo abbia grado minore. Diamo solo una traccia della dimostrazione lasciando al lettore il compito di completare i dettagli mancanti.

In una opportuna base l'endomorfismo f si rappresenta con una matrice a blocchi

$$F = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$$

con A matrice compagna di ordine $0 < r \leq n$. Se $r = n$ si usa il Lemma 10.2.1, mentre se $r < n$ possiamo supporre per induzione che i due sottospazi

$$U = \{H \in M_{r,r}(\mathbb{K}) \mid AH = HA\}, \qquad W = \{K \in M_{n-r,n-r}(\mathbb{K}) \mid CK = KC\},$$

abbiano dimensione $\geq r$ e $\geq n-r$ rispettivamente. Allora il sottospazio $Q \subset M_{r,r}(\mathbb{K})$ delle matrici a blocchi del tipo

$$\begin{pmatrix} H & L \\ 0 & K \end{pmatrix}, \quad H \in U,\ K \in W,\ L \in M_{r,n-r}(\mathbb{K}),$$

ha dimensione $\geq n + r(n-r)$, mentre l'applicazione lineare

$$Q \to Q, \quad X \mapsto FX - XF,$$

ha rango $\leq r(n-r)$.

Esercizi.

557. Sia V uno spazio vettoriale di dimensione finita sul campo $\mathbb{K}$ e siano $f, g\colon V \to V$ due endomorfismi coniugati, ossia tali che $f = hgh^{-1}$ per un opportuno endomorfismo invertibile $h\colon V \to V$. Dimostrare che:

(1) $f^2 = hg^2h^{-1}$, $f^3 = hg^3h^{-1}$ ecc.;
(2) per ogni polinomio $p(t) \in \mathbb{K}[t]$ si ha $p(f) = h\,p(g)\,h^{-1}$. In particolare gli endomorfismi $p(f)$ e $p(g)$ hanno lo stesso rango.

OSSERVAZIONE 10.2.8. Anche se dobbiamo aspettare il Capitolo 14 per poterlo dimostrare, anticipiamo al lettore il seguente bellissimo risultato (Teorema 14.5.3). *Siano $f, g\colon V \to V$ due endomorfismi con lo stesso polinomio caratteristico $p(t) = p_f(t) = p_g(t)$; allora f e g sono coniugati se e solo se, per ogni polinomio $q(t)$ che divide $p(t)$, i due endomorfismi $q(f)$ e $q(g)$ hanno lo stesso rango.*

558. Siano $f, g\colon V \to V$ due endomorfismi che commutano tra loro, ossia tali che $fg = gf$. Dimostrare che:

(1) Per ogni $h \geq 0$ vale $f^h g = g f^h$ (sugg.: induzione su h).
(2) Per ogni $h, k \geq 0$ vale $f^h g^k = g^k f^h$ (sugg.: induzione su k).
(3) Per ogni coppia di polinomi $p, q \in \mathbb{K}[t]$ vale $p(f)q(g) = q(g)p(f)$.

559. Calcolare polinomio caratteristico, polinomio minimo, autovalori ed autovettori delle matrici (intese come endomorfismi di $\mathbb{C}^n$)

$$\begin{pmatrix} 4 & -5 \\ 2 & -3 \end{pmatrix}, \quad \begin{pmatrix} 2 & 1 \\ -1 & 4 \end{pmatrix}, \quad \begin{pmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{pmatrix}, \quad \begin{pmatrix} 5 & 6 & -3 \\ -1 & 0 & 1 \\ 2 & 2 & -1 \end{pmatrix}.$$

560. Sia $f\colon V \to V$ un endomorfismo di uno spazio vettoriale di dimensione finita. Si assuma che V non abbia sottospazi f-invarianti eccetto 0 e V stesso. Dimostrare che il polinomio minimo di f è irriducibile, ossia non è uguale al prodotto di polinomi di grado positivo. (Nota: dimostreremo nel Corollario 14.4.5 che sotto le stesse ipotesi anche il polinomio caratteristico di f è irriducibile.)

561. Dimostrare che due matrici simili possiedono lo stesso polinomio minimo. Calcolare i polinomi minimo e caratteristico della matrice

$$A = \begin{pmatrix} i & 1 & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & 0 & 0 & -i \end{pmatrix} \in M_{4,4}(\mathbb{C})$$

e utilizzare l'Esercizio 552 per dedurre che A non è simile ad una matrice a coefficienti reali.

562. Siano $A, B \in M_{n,n}(\mathbb{C})$ due matrici fissate e si consideri l'applicazione lineare

$$f\colon M_{n,n}(\mathbb{C}) \to M_{n,n}(\mathbb{C}), \qquad f(X) = AX + XB.$$

Provare che:

(1) se $\det(A) = \det(B) = 0$ allora esiste una matrice non nulla $X \in\in M_{n,n}(\mathbb{C})$ tale che $AX = XB = 0$;
(2) siano λ autovalore di A e η autovalore di B, allora $\lambda + \eta$ è un autovalore di f;
(3) sia $M \in M_{n,n}(\mathbb{C})$ tale che $AM = MB$, allora $p_B(A)M = 0$ e dedurre che se $M \neq 0$ allora A e B hanno un autovalore comune;
(4) ogni autovalore di f è uguale a $\lambda + \eta$, con λ autovalore di A e η autovalore di B.

563 (☕). Sia $p\colon V \to W$ un'applicazione lineare surgettiva tra spazi vettoriali di dimensione finita e siano $f\colon V \to V$, $g\colon W \to W$ due endomorfismi tali che $pf = gp$. Dimostrare che il polinomio caratteristico di g divide il polinomio caratteristico di f e che il polinomio minimo di g divide il polinomio minimo di f.

564. Usando che ogni polinomio di grado positivo è divisibile solo per un numero finito di polinomi monici, dimostrare che se il campo $\mathbb{K}$ contiene infiniti elementi allora vale anche il viceversa del Lemma 10.2.1, ossia che se il grado del polinomio minimo di un endomorfismo $f\colon V \to V$ è uguale alla dimensione di V, allora f si rappresenta con una matrice compagna in una opportuna base. Il risultato è vero anche su campi finiti, ma in tal caso serve una diversa dimostrazione, vedi Esercizio 719.

10.3. Polinomio minimo e diagonalizzazione

In questa sezione studieremo gli analoghi dei criteri di triangolabilità e diagonalizzabilità visti nelle Sezioni 9.6 ed 9.7.

Abbiamo già dimostrato (Teorema 9.6.4) che un endomorfismo è triangolabile su un campo $\mathbb{K}$ se e solo se possiede tutti gli autovalori nel campo $\mathbb{K}$, ossia se e solo se il suo polinomio caratteristico è un prodotto di polinomi di primo grado a coefficienti in $\mathbb{K}$. Vale un risultato analogo per il polinomio minimo.

Teorema 10.3.1. *Sia $f\colon V \to V$ un endomorfismo di uno spazio vettoriale di dimensione finita su di un campo $\mathbb{K}$. Allora f è triangolabile se e solo se il suo polinomio minimo ha tutte le radici in $\mathbb{K}$, ossia se e solo se è un prodotto di polinomi di primo grado a coefficienti in $\mathbb{K}$.*

Dimostrazione. Se f è triangolabile allora il polinomio caratteristico è un prodotto di polinomi di primo grado e per il teorema di Cayley–Hamilton la medesima conclusione vale per il polinomio minimo.

Viceversa, se il polinomio minimo è un prodotto di polinomi di primo grado allora per il Teorema 10.1.10 l'endomorfismo f possiede autovettori. Per il Corollario 10.1.7 le stesse conclusioni valgono per la restrizione di f a ciascun sottospazio f-invariante non nullo e quindi la dimostrazione segue dal Teorema 9.6.3. □

È chiaro come segua anche dal precedente teorema, e dal teorema fondamentale dell'algebra, che ogni endomorfismo definito sul campo dei numeri complessi è triangolabile.

Osservazione 10.3.2. Per gli endomorfismi triangolabili esiste una ulteriore dimostrazione del teorema di Cayley–Hamilton. Sia $f\colon V \to V$ triangolabile e sia $v_1, \dots, v_n$ una base di V in cui f si rappresenta con una matrice triangolare superiore (a_{ij}). Per dimostrare Cayley–Hamilton basta dimostrare che per ogni $i = 1, \dots, n$ si ha

$$p_f(f)v_i = 0.$$

Per come abbiamo scelto la base, per ogni $j = 1, \dots, n$ si ha

$$f(v_j) = a_{jj}v_j + \sum_{s=1}^{j-1} a_{sj}v_s. \tag{10.1}$$

Notiamo che $a_{11}, \dots, a_{nn}$ è la lista, con possibili ripetizioni, degli autovalori di f. Dimostriamo per induzione su j che

$$(f - a_{11}I) \cdots (f - a_{jj}I)v_j = 0.$$

Questo, ovviamente, implicherà che $p_f(f)v_i = 0$, per ogni $i = 1, \dots, n$. Il caso $j = 1$ non è che la (10.1) scritta per $j = 1$. Supponiamo dunque che

$$(f - a_{11}I) \cdots (f - a_{ss}I)v_s = 0, \qquad s = 1, \dots, j-1.$$

Usando l'ipotesi induttiva, considerando che gli endomorfismi $f - a_{ii}I$, $i = 1, \dots, n$, commutano tra loro, si ha:

$$\begin{aligned}
&(f - a_{11}I) \cdots (f - a_{jj}I)v_j \\
&\quad = (f - a_{11}I) \cdots (f - a_{j-1,j-1}I) \left(\sum_{s=1}^{j-1} a_{sj}v_s \right) \\
&\quad = \sum_{s=1}^{j-1} a_{sj}(f - a_{s+1,s+1}I) \cdots (f - a_{j-1,j-1}I)((f - a_{11}I) \cdots (f - a_{s,s}I)v_s) \\
&\quad = 0.
\end{aligned}$$

TEOREMA 10.3.3. *Sia $f\colon V \to V$ un endomorfismo triangolabile di uno spazio vettoriale di dimensione finita con polinomio minimo*

$$q_f(t) = (t - \lambda_1)^{\sigma_1}(t - \lambda_2)^{\sigma_2} \cdots (t - \lambda_s)^{\sigma_s}, \qquad \sigma_i > 0, \quad \lambda_i \in \mathbb{K}, \quad \lambda_i \neq \lambda_j \text{ per } i \neq j.$$

Allora f è diagonalizzabile se e solo se $\sigma_i = 1$ per ogni indice i, ossia se e solo se $q_f(t)$ non ha radici multiple.

DIMOSTRAZIONE. L'implicazione "facile" è già stata vista nel Corollario 10.1.11 e ci basta dimostrare che se $\sigma_i = 1$ allora $\operatorname{Ker}((f - \lambda_i I)^2) = \operatorname{Ker}(f - \lambda_i I)$; la conclusione seguirà quindi dal Teorema 9.7.7.

Supponiamo per assurdo che esista un vettore $v \in \operatorname{Ker}((f - \lambda_i I)^2)$, $v \notin \operatorname{Ker}(f - \lambda_i I)$, allora si ha $u = (f - \lambda_i I)v \neq 0$; siccome $(f - \lambda_i I)u = (f - \lambda_i I)^2 v = 0$, ne segue che u è un autovettore con autovalore λ_i.

Se $\sigma_i = 1$ possiamo scrivere $q_f(t) = p(t)(t - \lambda_i)$ con $p(\lambda_i) \neq 0$ e, per il Lemma 10.1.1 abbiamo la seguente contraddizione:

$$0 = q_f(f)v = p(f)(f - \lambda_i I)v = p(f)u = p(\lambda_i)u \neq 0\,.$$

□

ESEMPIO 10.3.4. Sia $A \in M_{n,n}(\mathbb{K})$ una matrice triangolare con tutti i coefficienti uguali ad 1 sulla diagonale principale. Allora A è diagonalizzabile se e solo se è uguale all'identità. Infatti il polinomio caratteristico è $p_A(t) = (1 - t)^n$, dunque il polinomio minimo è $(t - 1)^\sigma$ per qualche $1 \leq \sigma \leq n$ e la matrice risulta diagonalizzabile se e solo se $\sigma = 1$, ossia se e solo se $f - I = 0$.

ESEMPIO 10.3.5. Siano V uno spazio vettoriale reale di dimensione finita e $f\colon V \to V$ una proiezione, ossia un endomorfismo tale che $f^2 = f$. Allora f è diagonalizzabile. Infatti $f^2 - f{=}0$, dunque il polinomio minimo divide $t^2 - t$ e deve pertanto essere uno dei seguenti:

$$q_f(t) = t, \qquad q_f(t) = t - 1, \qquad q_f(t) = t(t - 1).$$

In ciascun caso $q_f(t)$ possiede solo radici semplici e f risulta diagonalizzabile per il Teorema 10.3.3.

I precedenti risultati forniscono una dimostrazione alternativa del Corollario 9.7.9:

COROLLARIO 10.3.6. *Sia $f\colon V \to V$ un endomorfismo diagonalizzabile di uno spazio vettoriale di dimensione finita e sia $U \subseteq V$ un sottospazio f-invariante. Allora la restrizione $f_{|U}\colon U \to U$ di f ad U è diagonalizzabile.*

DIMOSTRAZIONE. Basta osservare che il polinomio minimo di $f_{|U}$ divide il polinomio minimo di f ed applicare il Teorema 10.3.3. □

PROPOSIZIONE 10.3.7 (Diagonalizzazione simultanea). *Siano V spazio vettoriale di dimensione finita e $\mathcal{A} \subseteq \mathrm{Hom}(V,V)$ un sottoinsieme, non necessariamente finito. Allora esiste una base di V in cui ogni elemento di $\mathcal{A}$ si rappresenta con una matrice diagonale se e solo se:*

(1) *ogni $f \in \mathcal{A}$ è diagonalizzabile;*
(2) *$fg = gf$ per ogni $f, g \in \mathcal{A}$.*

DIMOSTRAZIONE. Le condizioni sono ovviamente necessarie in quanto il prodotto di matrici diagonali è commutativo. Dimostriamo che le condizioni sono anche sufficienti per induzione sulla dimensione di V, considerando che se $\dim V \leq 1$ non c'è nulla da dimostrare. Supponiamo quindi $\dim V > 1$ ed il teorema vero per spazi vettoriali di dimensione inferiore. Se ogni elemento di $\mathcal{A}$ è un multiplo dell'identità allora ogni base di V va bene. Se esiste $f \in \mathcal{A}$ che non è un multiplo dell'identità e denotiamo con $\lambda_1, \dots, \lambda_s$ i suoi autovalori, si ha $s \geq 2$ e

$$V = V_1 \oplus \cdots \oplus V_s, \qquad V_i = \mathrm{Ker}(f - \lambda_i I) \neq 0.$$

Per ogni $g \in \mathcal{A}$, siccome $gf = fg$ vale anche $g(f - \lambda_i I) = (f - \lambda_i I)g$ e il Lemma 9.5.9 implica che ogni sottospazio V_i è g-invariante. Siccome le restrizioni degli endomorfismi di $\mathcal{A}$ ad ogni V_i commutano tra loro, per l'ipotesi induttiva ogni sottospazio V_i possiede una base di autovettori comuni a tutti gli endomorfismi di $\mathcal{A}$. L'unione delle basi dei V_i fornisce una base di V formata da autovettori comuni. □

Esercizi.

565. Calcolare autovalori e rispettive molteplicità algebriche e geometriche della matrice

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

Detto e_2 il secondo vettore della base canonica, provare che i 4 vettori $e_2, Ae_2, A^2e_2, A^3e_2$ sono linearmente indipendenti e calcolare il polinomio minimo di A.

566. Calcolare il polinomio minimo della matrice M dell'Esercizio 532. Rispondere alla stessa domanda interpretando però la matrice a coefficienti nel campo $\mathbb{F}_{17}$ delle classi di resto modulo 17 (Esempio 3.7.14).

567. Calcolare il polinomio minimo della matrice

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 5 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \end{pmatrix}.$$

Dire inoltre se tale matrice è diagonalizzabile su $\mathbb{R}$ e su $\mathbb{C}$.

568. Mostrare che per qualunque scelta dei coefficienti complessi al posto di $*$, le matrici

$$\begin{pmatrix} 1 & 1 & * \\ 0 & 1 & * \\ 0 & 0 & * \end{pmatrix}, \qquad \begin{pmatrix} 1 & * & 1 & * & * \\ 0 & * & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & * \\ 0 & 0 & 0 & * & 0 \\ 0 & * & 0 & 4 & * \end{pmatrix},$$

non sono diagonalizzabili su $\mathbb{C}$.

10.4. Matrici ed endomorfismi nilpotenti

Assieme alle matrici ed agli endomorfismi diagonalizzabili, un ruolo fondamentale in algebra lineare è interpretato dalle matrici e dagli endomorfismi nilpotenti.

DEFINIZIONE 10.4.1. Una matrice quadrata $A \in M_{n,n}(\mathbb{K})$ si dice **nilpotente** se $A^m = 0$, per qualche $m \geq 1$. Il più piccolo intero positivo s tale che $A^s = 0$ viene detto **indice di nilpotenza**.

Ad esempio le matrici

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \in M_{2,2}(\mathbb{Q}), \qquad \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix} \in M_{2,2}(\mathbb{C}),$$

sono nilpotenti con indice di nilpotenza 2 (esercizio: verificare), mentre le matrici

$$\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \in M_{2,2}(\mathbb{Q}), \qquad \begin{pmatrix} 1 & i \\ 0 & 1 \end{pmatrix} \in M_{2,2}(\mathbb{C}),$$

non sono nilpotenti (esercizio: perché?).

ESEMPIO 10.4.2. La matrice $J_n = (a_{ij}) \in M_{n,n}(\mathbb{K})$ tale che $a_{ij} = 1$ se $j = i+1$ e $a_{ij} = 0$ se $j \neq i+1$, detta **blocco di Jordan**[1] **nilpotente di ordine** n, è nilpotente con indice di nilpotenza uguale a n. Infatti, nella base canonica $e_1, \ldots, e_n$ di $\mathbb{K}^n$ si ha

$$J_n e_1 = 0, \qquad J_n e_i = e_{i-1}, \quad i > 1.$$

$$J_1 = (0), \quad J_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad J_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad J_4 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \ldots$$

Per induzione su $m > 0$ si prova che

$$J_n^m e_j = 0, \quad j \leq m, \qquad J_n^m e_i = e_{i-m}, \quad i > m,$$

e quindi per ogni $m \leq n$ il nucleo di J_n^m ha dimensione m.

Si noti che ogni blocco di Jordan J_n è simile al suo trasposto J_n^T, ossia alla matrice compagna di t^n: basta infatti considerare il cambio di base $e_i \mapsto e_{n-i+1}$ descritto nell'Esempio 9.1.4.

ESEMPIO 10.4.3. Sia A una matrice triangolare $n \times n$ con tutti zeri sulla diagonale principale, allora A è nilpotente. Infatti il polinomio caratteristico è $p_A(t) = (-t)^n$ e per il teorema di Cayley–Hamilton $A^n = 0$ (per una dimostrazione più diretta vedi Esercizio 574).

ESEMPIO 10.4.4. Date $A \in M_{n,n}(\mathbb{K})$, $B \in M_{n,m}(\mathbb{K})$ e $C \in M_{m,m}(\mathbb{K})$, la matrice

$$D = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} \in M_{n+m,n+m}(\mathbb{K}),$$

è nilpotente se e solo se A e C sono nilpotenti. Infatti, si dimostra facilmente per induzione su $k > 0$ che

$$D^k = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}^k = \begin{pmatrix} A^k & B_k \\ 0 & C^k \end{pmatrix}, \quad \text{dove} \quad B_k = \sum_{i=0}^{k-1} A^i B C^{k-i-1},$$

e di conseguenza se $D^p = 0$ allora $A^p = 0$ e $C^p = 0$; viceversa se $A^q = 0$ e $C^r = 0$ allora $D^{q+r} = 0$.

ESEMPIO 10.4.5. Sia $A \in M_{n,n}(\mathbb{R})$ una matrice simmetrica e nilpotente a coefficienti reali. Allora $A = 0$. Infatti, poiché $A^2 = A^T A$ è ancora simmetrica e $A^{2^k} = 0$ per k sufficientemente grande, ragionando per induzione su k basta dimostrare che se A è simmetrica e $A^2 = 0$, allora $A = 0$. A tal fine è sufficiente osservare che la traccia di $A^2 = A^T A$ è uguale alla somma dei quadrati dei coefficienti di A.

Notiamo che il precedente fatto non è vero per matrici a coefficienti complessi: ad esempio la matrice simmetrica

$$\begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix} \in M_{2,2}(\mathbb{C})$$

ha quadrato nullo.

[1] Leggasi giordàn.

Se A e B sono matrici simili, abbiamo già osservato che anche A^h e B^h sono simili per ogni $h > 0$; dunque se A è nilpotente anche B è nilpotente e viceversa. Essendo la nozione di nilpotenza invariante per similitudine possiamo estenderla agli endomorfismi, per i quali esiste però una definizione più diretta.

La nozione di nilpotenza si estende in maniera naturale agli endomorfismi.

Definizione 10.4.6. Un endomorfismo $f\colon V \to V$ si dice **nilpotente** se $f^m = 0$, per qualche $m \geq 1$. Il più piccolo intero positivo s tale che $f^s = 0$ viene detto **indice di nilpotenza**.

Ad esempio, un endomorfismo $f\colon V \to V$ è nilpotente con indice di nilpotenza 2 se e solo se $f \neq 0$ e $f(V) \subseteq \mathrm{Ker}(f)$. Equivalentemente un endomorfismo $f\colon V \to V$ è nilpotente se esiste un intero positivo $m > 0$ tale che $\mathrm{Ker}(f^m) = V$.

Teorema 10.4.7. *Sia f un endomorfismo di uno spazio vettoriale V di dimensione finita n. Allora le seguenti condizioni sono equivalenti:*

(1) $f^n = 0$;
(2) *f è nilpotente;*
(3) *la restrizione di f ad ogni sottospazio f-invariante non nullo possiede nucleo non banale;*
(4) *in una base opportuna f si rappresenta con una matrice triangolare strettamente superiore;*
(5) $p_f(t) = (-1)^n t^n$;
(6) $q_f(t) = t^s$, *con* $s \leq n$.

In particolare ogni endomorfismo nilpotente possiede $\lambda = 0$ come unico autovalore.

Dimostrazione. Iniziamo con l'osservare che se $V \neq 0$ e $f\colon V \to V$ è nilpotente, allora $\mathrm{Ker}(f) \neq 0$: infatti se k è il più piccolo intero tale che $f^k = 0$, allora $k > 0$ perché $V \neq 0$, esiste un vettore $v \in V$ tale che $w = f^{k-1}(v) \neq 0$ e quindi $f(w) = f^k(v) = 0$, ossia $w \in \mathrm{Ker}(f)$. Alternativamente, se fosse $\mathrm{Ker}(f) = 0$ allora sia f che tutte le potenze f^k risultano applicazioni iniettive.

L'implicazione (1)⇒(2) segue immediatamente dalla definizione. Supponiamo f nilpotente e sia $U \subseteq V$ un sottospazio f-invariante non nullo. Allora pure la restrizione $f_{|U}\colon U \to U$ è nilpotente e per quanto visto sopra $\mathrm{Ker}(f_{|U}) = U \cap \mathrm{Ker}(f) \neq 0$; abbiamo quindi dimostrato l'implicazione (2)⇒(3).

Se vale (3) allora ogni sottospazio invariante non nullo possiede un autovettore con autovalore 0 e per il Teorema 9.6.3 l'endomorfismo f è triangolabile, ossia si rappresenta in una opportuna base con una matrice triangolare. Se tale matrice avesse un coefficiente non nullo $\lambda \neq 0$ sulla diagonale principale, allora λ sarebbe un autovalore, ed il sottospazio $U = \mathrm{Ker}(f - \lambda I)$ non nullo e f-invariante. Però la restrizione di f ad U è iniettiva, in quanto coincidente con λI e quindi l'ipotesi $\lambda \neq 0$ porta ad una contraddizione. Abbiamo quindi dimostrato l'implicazione (3)⇒(4).

L'implicazione (4)⇒(5) segue dal calcolo del polinomio caratteristico di matrici triangolari, mentre (5)⇒(6) segue da Cayley–Hamilton. Infine, se il polinomio minimo è t^s con $s \leq n$ si ha $f^s = 0$ ed a maggior ragione $f^n = 0$, provando quindi l'implicazione (6)⇒(1) □

Osservazione 10.4.8. Sia $f\colon V \to V$ un endomorfismo nilpotente, è allora chiaro che anche λf è nilpotente per ogni $\lambda \in \mathbb{K}$. Inoltre, dal fatto che f non possiede autovalori diversi da 0 segue che $f - \lambda I$ e $I - \lambda f$ sono invertibili per ogni $\lambda \neq 0$.

Una dimostrazione alternativa dell'invertibilità di $I - \lambda f$ si può ricavare dalle ben note uguaglianze polinomiali

$$1 - t^m = (1 - t)(1 + t + t^2 + \cdots + t^{m-1}), \qquad m \geq 1.$$

Infatti, se $f^m = 0$ per qualche $m > 0$ allora

$$I = I - (\lambda f)^m = (I - \lambda f)(I + \lambda f + \cdots + (\lambda f)^{m-1}).$$

Esercizi.

569. Trovare due matrici 4×4, nilpotenti con lo stesso indice di nilpotenza, che non sono simili.

570. Dimostrare che una matrice $A \in M_{2,2}(\mathbb{C})$ è nilpotente se e soltanto se $\mathrm{Tr}(A) = \mathrm{Tr}(A^2) = 0$.

571. Sia $A \in M_{2,2}(\mathbb{K})$ nilpotente. Dimostrare che esistono $a, b, c \in \mathbb{K}$ tali che

$$A = a \begin{pmatrix} bc & b^2 \\ -c^2 & -bc \end{pmatrix}.$$

572. Siano $A, B \in M_{n,n}(\mathbb{K})$ tali che $A^2 = B^2 = 0$. Provare che le matrici A e B sono simili se e solo se hanno lo stesso rango.

573. Mostrare che una matrice triangolare è nilpotente se e solo se tutti i coefficienti sulla diagonale principale sono uguali a 0.

574. Sia $A = (a_{ij})$ una matrice $n \times n$ e sia k un intero tale che $0 \leq k < n$ e $a_{ij} = 0$ ogniqualvolta $j - i \leq k$. Dimostrare che $A^{n-k} = 0$.

575. Siano V spazio vettoriale su $\mathbb{C}$ e $f\colon V \to V$ un endomorfismo nilpotente. Dimostrare che $2I + 2f + f^2$ è invertibile e che per ogni $a, b \in V$ esistono due vettori $x, y \in V$ tali che

$$f(x) + x + y = a, \quad f(y) + y - x = b.$$

576. Sia $f\colon V \to V$ un endomorfismo nilpotente e non nullo. Dimostrare che non esiste alcun sottospazio f-invariante $U \subseteq V$ tale che $V = \mathrm{Ker}(f) \oplus U$.

577. Provare che una matrice $A \in M_{n,n}(\mathbb{C})$ è nilpotente se e solo se le matrici A e $2A$ hanno lo stesso polinomio caratteristico.

578 (☕). Siano V spazio vettoriale di dimensione finita e $f, g\colon V \to V$ endomorfismi nilpotenti tali che $fg = gf$. Provare che esiste una base rispetto alla quale f e g sono rappresentate da matrici triangolari strettamente superiori.

579 (☕). Sia $H \subseteq M_{n,n}(\mathbb{K})$ un sottospazio vettoriale tale che $A^2 = 0$ per ogni $A \in H$. Dimostrare che per ogni $A, B \in H$ si ha

$$AB + BA = 0, \qquad (AB - BA)^2 = 0.$$

10.5. Matrici simmetriche ed antisimmetriche reali

Siamo adesso in grado di dimostrare un risultato di grande importanza sia teorica che applicativa, ossia che ogni matrice simmetrica $A \in M_{n,n}(\mathbb{R})$ si diagonalizza su $\mathbb{R}$.

Tale risultato dipende in maniera fondamentale dalle proprietà dei numeri reali ed è falso ad esempio sia sul campo $\mathbb{Q}$ dei razionali che sul campo dei complessi. Infatti la matrice simmetrica

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \in M_{2,2}(\mathbb{Q})$$

non si diagonalizza su $\mathbb{Q}$ (ma si diagonalizza su $\mathbb{R}$), mentre la matrice simmetrica

$$\begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix} \in M_{2,2}(\mathbb{C})$$

ha quadrato nullo e quindi non è diagonalizzabile.

Conviene pensare ogni matrice $A \in M_{n,n}(\mathbb{R})$ come un endomorfismo dello spazio vettoriale $\mathbb{R}^n$ dei vettori colonna. Il fatto chiave è dato dalle proprietà della norma $\|v\| = \sqrt{v^T v}$, $v \in \mathbb{R}^n$, ed in particolare dal fatto che $\|v\|^2 = v^T v \geq 0$ e vale $v^T v = 0$ se e solo se $v = 0$.

LEMMA 10.5.1. *Per ogni matrice $B \in M_{n,n}(\mathbb{R})$ ed ogni numero reale $a > 0$, la matrice $B^T B + aI$ è invertibile.*

DIMOSTRAZIONE. Dimostriamo che $\mathrm{Ker}(B^TB + aI) = 0$ per ogni $a > 0$. Sia $v \in \mathbb{R}^n$ un vettore tale che $B^TBv + av = 0$, allora

$$0 = v^T(B^TBv + av) = v^TB^TBv + av^Tv = (Bv)^T(Bv) + av^Tv \geq av^Tv \geq 0.$$

Dunque $av^Tv = 0$ e quindi $v = 0$. □

LEMMA 10.5.2. *Sia $B \in M_{n,n}(\mathbb{R})$ una matrice tale che $BB^T = B^TB$ (ad esempio B simmetrica oppure antisimmetrica). Allora* $\mathrm{Ker}(B) = \mathrm{Ker}(B^2)$.

DIMOSTRAZIONE. Basta dimostrare che $\mathrm{Ker}(B^2) \subseteq \mathrm{Ker}(B)$. Consideriamo prima il caso particolare in cui $B = B^T$ è una matrice simmetrica; dato $v \in \mathrm{Ker}(B^2)$ si ha

$$0 = v^T(B^2v) = v^TBBv = v^TB^TBv = (Bv)^T(Bv) = \|Bv\|^2$$

e quindi $Bv = 0$. In generale, se $B^TB = BB^T$ e $v \in \mathrm{Ker}(B^2)$, a maggior ragione $v \in \mathrm{Ker}((B^T)^2B^2) = \mathrm{Ker}((B^TB)^2)$ e siccome B^TB è simmetrica si ottiene $v \in \mathrm{Ker}(B^TB)$. Moltiplicando la relazione $B^TBv = 0$ a sinistra per v^T si ottiene $\|Bv\|^2 = 0$ da cui $Bv = 0$. □

TEOREMA 10.5.3. *Ogni matrice simmetrica reale si diagonalizza su $\mathbb{R}$. In particolare, due matrici simmetriche reali sono simili se e solo se hanno lo stesso polinomio caratteristico.*

DIMOSTRAZIONE. Sia $A \in M_{n,n}(\mathbb{R})$ simmetrica. Per il teorema fondamentale dell'algebra la matrice A possiede tutti gli autovalori su $\mathbb{C}$ ed è quindi sufficiente dimostrare che:

(1) Se $\lambda \in \mathbb{C}$ è un autovalore di A, pensata come matrice a coefficienti complessi, allora $\lambda \in \mathbb{R}$.
(2) Per ogni autovalore λ vale $\mathrm{Ker}(A - \lambda I) = \mathrm{Ker}(A - \lambda I)^2$.

Sia dunque $\lambda = a + ib \in \mathbb{C}$, $a, b \in \mathbb{R}$, un autovalore complesso di A e consideriamo la matrice simmetrica reale $B = A - aI$. Allora $\det(B - ibI) = \det(A - \lambda I) = 0$ e quindi

$$\det(B^2 + b^2I) = \det((B + ibI)(B - ibI)) = \det(B + ibI)\det(B - ibI) = 0.$$

Per il Lemma 10.5.1 deve essere $b = 0$, ossia $\lambda = a \in \mathbb{R}$. Il secondo punto segue applicando il Lemma 10.5.2 alla matrice simmetrica reale $A - \lambda I$. □

OSSERVAZIONE 10.5.4. È possibile dimostrare che se A è simmetrica reale, allora esiste una base di autovettori $v_1, \ldots, v_n$ tale che $v_i^Tv_j = \delta_{ij}$ (delta di Kronecker) per ogni i, j.

COROLLARIO 10.5.5. *Siano $\lambda_1 < \lambda_2 < \cdots < \lambda_k$ gli autovalori, in ordine crescente, di una matrice simmetrica reale $A \in M_{n,n}(\mathbb{R})$. Allora per ogni vettore $x \in \mathbb{R}^n$ valgono le disuguaglianze*

$$\lambda_1\|x\|^2 \leq x^TAx \leq \lambda_k\|x\|^2.$$

DIMOSTRAZIONE. Osserviamo preliminarmente che se u è un autovettore per λ_i e v è un autovettore per λ_j, con $i \neq j$, allora vale $u^Tv = u^TAv = 0$. Infatti

$$u^TAv = u^T(\lambda_jv) = \lambda_ju^Tv, \qquad u^TAv = (u^TAv)^T = v^TA^Tu = v^TAu = \lambda_iv^Tu$$

da cui segue $(\lambda_i - \lambda_j)u^Tv = 0$. Abbiamo dimostrato che A è diagonalizzabile e quindi ogni vettore $x \in \mathbb{R}^n$ si scrive in modo unico come $x = v_1 + \cdots + v_k$, con $Av_i = \lambda_iv_i$. Dunque si ha

$$\|x\|^2 = x^Tx = \sum_{i,j} v_i^Tv_j = \sum_i v_i^Tv_i = \sum_i \|v_i\|^2,$$
$$x^TAx = \sum_{i,j} v_i^TAv_j = \sum_i v_i^TAv_i = \sum_i \lambda_iv_i^Tv_i = \sum_i \lambda_i\|v_i\|^2,$$

da cui segue

$$\lambda_1\|x\|^2 = \sum_i \lambda_1\|v_i\|^2 \leq \sum_i \lambda_i\|v_i\|^2 = x^TAx$$
$$x^TAx = \sum_i \lambda_i\|v_i\|^2 \leq \sum_i \lambda_k\|v_i\|^2 = \lambda_k\|x\|^2.$$

□

10.5.1. Matrici antisimmetriche. In generale le matrici antisimmetriche reali non si diagonalizzano su $\mathbb{R}$: più precisamente l'unica matrice antisimmetrica reale che si diagonalizza su $\mathbb{R}$ è quella nulla. Infatti, se $\lambda \in \mathbb{R}$ è un autovalore reale di una matrice antisimmetrica $A \in M_{n,n}(\mathbb{R})$ e v è un autovettore corrispondente, allora $A^2 v = \lambda^2 v$ e

$$\lambda^2 \|v\|^2 = (\lambda v^T)(\lambda v) = (Av)^T (Av) = v^T(-A^2)v = -\lambda^2 \|v\|^2,$$

da cui segue $\lambda^2 = 0$, ossia $\lambda = 0$.

DEFINIZIONE 10.5.6. Una matrice antisimmetrica $A \in M_{n,n}(\mathbb{K})$ si dice **normalizzata**, o in **forma normale**, se è diagonale a blocchi del tipo:

$$\begin{pmatrix} H_{a_1} & 0 & \dots & 0 & 0 \\ 0 & H_{a_2} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & H_{a_s} & 0 \\ 0 & 0 & \dots & 0 & N \end{pmatrix} \tag{10.2}$$

dove

$$H_{a_i} = \begin{pmatrix} 0 & -a_i \\ a_i & 0 \end{pmatrix} \in M_{2,2}(\mathbb{K}), \qquad a_i \neq 0,$$

per ogni i e N è la matrice nulla $(n-2s) \times (n-2s)$.

Dunque ogni matrice antisimmetrica normalizzata è una matrice diagonale a blocchi, con blocchi 1×1 nulli e 2×2 invertibili. Tali blocchi sono univocamente determinati a meno dell'ordine e del segno dal polinomio caratteristico: infatti il polinomio caratteristico della matrice in forma normale (10.2) è uguale a $(-t)^{n-2s} \prod_{i=1}^{s} (t^2 + a_i^2)$. Ricordiamo che per l'Esempio 9.1.4 ogni matrice H_a è simile a H_{-a} e quindi due matrici antisimmetriche normalizzate sono simili se e solo se hanno lo stesso polinomio caratteristico.

TEOREMA 10.5.7. *Ogni matrice antisimmetrica reale è simile ad una matrice antisimmetrica normalizzata. In particolare, due matrici antisimmetriche reali sono simili se e solo se hanno lo stesso polinomio caratteristico.*

DIMOSTRAZIONE. Sia $A \in M_{n,n}(\mathbb{R})$ antisimmetrica, allora A^2 è simmetrica e quindi diagonalizzabile. Denotando con $f = L_A \colon \mathbb{R}^n \to \mathbb{R}^n$ l'endomorfismo associato, si ha

$$v^T f(w) = v^T A w = (v^T A w)^T = -w^T A v = -w^T f(v) = -f(v)^T w$$

per ogni $v, w \in \mathbb{R}^n$. Per il Lemma 10.5.2 si ha $\operatorname{Ker} f = \operatorname{Ker} f^2$ ed infine, per il Corollario 10.3.6 la restrizione di f^2 ad ogni sottospazio f-invariante è diagonalizzabile.

Dimostriamo per induzione su m la seguente asserzione: *sia $V \subseteq \mathbb{R}^n$ uno sottospazio f-invariante di dimensione $\dim V = m$. Allora esiste una base $(v_1, \dots, v_m)$ di V nella quale $f_{|V}$ si rappresenta con una matrice antisimmetrica normalizzata.*

Il caso $V \subseteq \operatorname{Ker} f$ è del tutto ovvio; supponiamo quindi $V \not\subseteq \operatorname{Ker} f$. Siccome $\operatorname{Ker} f = \operatorname{Ker} f^2$ si ha $f^2(V) \neq 0$ e quindi l'endomorfismo diagonalizzabile $f^2 \colon V \to V$ possiede un autovettore v_1, con autovalore reale $\lambda \neq 0$. Dimostriamo che $\lambda < 0$ e che i vettori $v_1, f(v_1)$ sono linearmente indipendenti: si ha

$$\lambda \|v_1\|^2 = v_1^T f^2(v_1) = v_1^T f(f(v_1)) = -f(v_1)^T f(v_1) = -\|f(v_1)\|^2$$

da cui segue $\lambda < 0$. Se $v_1, f(v_1)$ fossero linearmente dipendenti si avrebbe $f(v_1) = \gamma v_1$ per qualche $\gamma \in \mathbb{R}$ e quindi $\lambda v_1 = f^2(v_1) = \gamma^2 v_1$ da cui $\lambda = \gamma^2$, in contraddizione con la negatività di λ. Detto $a_1 = \sqrt{-\lambda} > 0$ e ponendo $v_2 = f(v_1)/a_1$ si ha

$$f(v_2) = \frac{f^2(v_1)}{a_1} = \frac{\lambda}{a_1} v_1 = -a_1 v_1.$$

Dunque il sottospazio $U = \operatorname{Span}(v_1, v_2)$ è f-invariante e la restrizione di f ad U è rappresentata, nella base (v_1, v_2) dalla matrice H_{a_1}. Si noti che

$$v_1^T v_2 = \frac{1}{a_1} v_1^T f(v_1) = -\frac{1}{a_1} v_1^f(v_1)$$

da cui segue $v_1^T v_2 = 0$ e $v_2^T v_1 = (v_1^T v_2)^T = 0$.

Consideriamo adesso il sottospazio vettoriale $W = \{v \in V \mid v^T v_1 = v^T v_2 = 0\}$ e dimostriamo che $V = U \oplus W$ è una somma diretta f-invariante. Se $v \in W$ allora

$$f(v)^T v_1 = -v^T f(v_1) = -a_1 v^T v_2 = 0, \qquad f(v)^T v_2 = -v^T f(v_2) = a_1 v^T v_1 = 0,$$

e quindi $f(v) \in W$ per la definizione di W. Se $w \in U \cap W$, allora esistono $\alpha, \beta \in \mathbb{R}$ tali che $w = \alpha v_1 + \beta v_2$ e siccome $w \in W$ si ha

$$0 = w^T v_1 = \alpha \|v_1\|^2, \qquad 0 = w^T v_2 = \beta \|v_2\|^2,$$

per cui $\alpha = \beta = 0$, e cioè $w = 0$. D'altra parte $\dim W \geq m - 2$ in quanto W coincide con il nucleo dell'applicazione lineare

$$V \to \mathbb{R}^2, \qquad v \mapsto \begin{pmatrix} v^T v_1 \\ v^T v_2 \end{pmatrix},$$

e di conseguenza $\dim U \oplus W \geq \dim V$. Per concludere la dimostrazione basta applicare l'ipotesi induttiva alla restrizione di f al sottospazio W. □

Esempio 10.5.8. Calcoliamo la forma normale della matrice antisimmetrica reale

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

Abbiamo visto che basta calcolare il polinomio caratteristico, che nella fattispecie è uguale a

$$p_A(t) = t^4 + 3t^2 + 1 = \left(t^2 + \frac{3+\sqrt{5}}{2}\right)\left(t^2 + \frac{3-\sqrt{5}}{2}\right).$$

La forma normale, a meno di permutazioni e cambi di segno dei blocchi, è quindi uguale a:

$$\begin{pmatrix} 0 & a & 0 & 0 \\ -a & 0 & 0 & 0 \\ 0 & 0 & 0 & b \\ 0 & 0 & -b & 0 \end{pmatrix}, \qquad a = \sqrt{\frac{3+\sqrt{5}}{2}}, \quad b = \sqrt{\frac{3-\sqrt{5}}{2}}.$$

Esercizi.

580. Sia $B \in M_{n,n}(\mathbb{R})$ una matrice tale che $BB^T = B^T B$. Provare che per ogni intero positivo si ha $(B^n)^T B^n = B^n (B^n)^T$ e $\mathrm{Ker}(B^n) = \mathrm{Ker}(B^{n+1})$.

581. Calcolare autovalori ed autovettori della matrice simmetrica reale

$$\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}.$$

582. Sia $A \in M_{n,n}(\mathbb{R})$ simmetrica e siano $v, w \in \mathbb{R}^n$ autovettori di A relativi ad autovalori distinti. Mostrare che $v^T w = 0$.

583. Sia $A \in M_{n,n}(\mathbb{R})$ antisimmetrica. Usare un argomento simile a quello usato nella dimostrazione del Teorema 10.5.3 per dimostrare che ogni autovalore complesso di A è immaginario puro. Dimostrare inoltre che A è diagonalizzabile su $\mathbb{C}$.

584. Determinare la matrice antisimmetrica normalizzata simile, su $\mathbb{R}$, alla matrice

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

585. Sia $A = (a_{ij})$ una matrice simmetrica reale $n \times n$ e siano $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ i suoi autovalori. Dimostrare che per ogni indice i vale

$$a_{ii} \geq \min\{\lambda_1, \ldots, \lambda_n\}.$$

586. Una matrice simmetrica reale $A \in M_{n,n}(\mathbb{R})$ è detta **semidefinita positiva** se $x^T A x \geq 0$ per ogni $x \in \mathbb{R}^n$. Sia $A \in M_{n,n}(\mathbb{R})$ una matrice simmetrica; dimostrare che:

(1) A è semidefinita positiva se e solo se tutti i suoi autovalori sono ≥ 0;

(2) se A è semidefinita positiva allora $\det(I+A) \geq 1$;
(3) se A è semidefinita positiva e $x^T Ax = 0$, allora $Ax = 0$;
(4) se A è semidefinita positiva di rango 1, allora esistono $a_1, \dots, a_n \in \mathbb{R}$ tali che $A_{ij} = a_i a_j$ per ogni i, j.

587. Sia A una matrice antisimmetrica reale. Dimostrare che $-A^2$ è una matrice simmetrica semidefinita positiva.

588 (☕). Siano $A, B \in M_{n,n}(\mathbb{R})$ matrici simmetriche semidefinite positive (Esercizio 586). Dimostrare che la matrice $AB + I$ è invertibile e che ogni autovalore di AB è non negativo.

589. Siano p, n interi positivi, con $n > 2p$, e sia $U \in M_{p,p}(\mathbb{R})$ la matrice con tutti i coefficienti uguali ad 1. Calcolare autovalori ed una base di autovettori della matrice a blocchi

$$\begin{pmatrix} U & 0 & U \\ 0 & 0 & 0 \\ U & 0 & U \end{pmatrix} \in M_{n,n}(\mathbb{R}) .$$

590. Sia $q(t)$ il polinomio minimo di una matrice antisimmetrica a coefficienti reali. Provare che $q(0) \geq 0$.

10.6. Criterio di Sylvester e regola dei segni di Cartesio

Sappiamo che ogni polinomio a coefficienti reali di grado n possiede al più n radici reali. In questa sezione illustreremo un semplice criterio per determinare quante radici reali positive e quante radici reali negative può avere al massimo un determinato polinomio a coefficienti reali.

TEOREMA 10.6.1. *Sia $p(t)$ un polinomio di grado positivo a coefficienti reali. Allora il numero di radici reali positive di $p(t)$, contate con molteplicità, è minore od uguale al numero dei cambiamenti di segno nella successione ordinata dei coefficienti non nulli di $p(t)$.*

Ad esempio, la successione dei coefficienti non nulli di $t^9+2t^6-3t^4+2t^3-1$ è $1, 2, -3, 2, -1$ ed il numero dei cambiamenti di segno è quindi 3. Similmente il numero dei cambiamenti di segno di $t^3 + t^2 - t + 1$ è 2, quello di $t^4 - t^2 - 1$ è 1, mentre quello di $t^4 + t^2 + 2$ è 0. Dunque per il Teorema 10.6.1 il polinomio $t^4 - t^2 - 1$ possiede al più una radice reale positiva.

DIMOSTRAZIONE. Denotiamo con $s(p(t))$ il numero dei cambiamenti di segno nella successione dei coefficienti non nulli di un polinomio $p(t) \in \mathbb{R}[t]$. Per dimostrare il teorema è sufficiente dimostrare la formula

$$s((t-c)p(t)) > s(p(t)), \qquad \text{per ogni} \quad p(t) \in \mathbb{R}[t], \quad c \in \mathbb{R}\,, \quad c > 0\,. \tag{10.3}$$

Infatti se $c_1, \dots, c_k$ sono le radici reali positive, contate con molteplicità, di $p(t)$, è possibile scrivere

$$p(t) = (t-c_1)(t-c_2)\cdots(t-c_k)q(t)$$

e quindi $s(p(t)) \geq k + s(q(t)) \geq k$.

Per dimostrare la formula (10.3) supponiamo il polinomio $p(t)$ di grado n, con $t = 0$ radice di molteplicità m e scriviamo

$$p(t) = a_n t^n + a_{n-1} t^{n-1} + \cdots + a_m t^m, \qquad a_m, a_n \neq 0.$$

Denotiamo con $s = s(p(t))$ il numero di cambiamenti di segno e consideriamo la successione strettamente decrescente $n = i_0 > i_1 > \cdots > i_s \geq m$ definita ricorsivamente come

$$i_0 = n, \qquad i_{k+1} = \max\{j \mid j < i_k,\ a_j a_{i_k} < 0\}, \quad 0 \leq k < s.$$

Scrivendo $(t-c)p(t) = b_{n+1}t^{n+1} + \cdots + b_m t^m$, per concludere la dimostrazione è sufficiente dimostrare che

$$b_{n+1} = b_{i_0+1},\ b_{i_1+1}, \dots,\ b_{i_s+1},\ b_m,$$

è una successione di numeri reali a segni alterni. Dato che, per ogni $k = 0, \dots, s$, il coefficiente a_{i_k+1} è nullo oppure di segno opposto a a_{i_k}, ne segue che $b_{i_k+1} = a_{i_k} - c a_{i_k+1}$ ha lo stesso segno di a_{i_k}. Infine, siccome a_{i_s} ha lo stesso segno di a_m ne segue che b_{i_s+1} e $b_m = -c a_m$ hanno segni opposti. □

COROLLARIO 10.6.2 (Regola dei segni di Cartesio). *Sia $p(t)$ un polinomio di grado positivo a coefficienti reali. Se tutte le radici di $p(t)$ sono reali, allora il numero di radici positive, contate con molteplicità, è uguale al numero dei cambiamenti di segno della successione ordinata dei coefficienti non nulli.*

DIMOSTRAZIONE. Dividendo il polinomio per potenze di t il numero dei cambi di segno resta invariato; non è quindi restrittivo supporre che $p(0) \neq 0$ e quindi che tutte le radici siano reali e non nulle.

Sia n il grado di $p(t)$ e denotiamo con $c_1, \dots, c_k$ le radici positive e $-d_{k+1}, \dots, -d_n$ quelle negative, contate con molteplicità.

Le radici del polinomio $p(-t)$ sono pertanto $-c_1, \dots, -c_k, d_{k+1}, \dots, d_n$ e quindi per il Teorema 10.6.1 si hanno le disuguaglianze $s(p(t)) \geq k$, $s(p(-t)) \geq n-k$.

Dunque è sufficiente dimostrare che vale $s(p(t)) + s(p(-t)) \leq n$. I coefficienti dei due polinomi $p(t) = \sum_{i=0}^{n} a_i t^i$, $p(-t) = \sum_{i=0}^{n} b_i t^i$ sono legati dalle relazioni $b_i = (-1)^i a_i$. Se i coefficienti a_i sono tutti diversi da 0 allora a_i ed a_{i+1} hanno lo stesso segno se e solo se b_i e b_{i+1} hanno segni opposti ed è immediato osservare che $s(p(t)) + s(p(-t)) = n$. Se p ha qualche coefficiente nullo consideriamo un nuovo polinomio $q(t)$ ottenuto da $p(t)$ sostituendo ad ogni coefficiente nullo un qualsiasi numero reale $\neq 0$. Siccome $s(q(t)) \geq s(p(t))$ e $s(q(-t)) \geq s(p(-t))$, la dimostrazione è conclusa. □

Tornando all'algebra lineare, se sappiamo che una matrice $A \in M_{n,n}(\mathbb{R})$ è triangolabile, o meglio ancora diagonalizzabile, e $p_A(t)$ è il suo polinomio caratteristico, allora:

(1) la molteplicità algebrica dell'autovalore 0 è la più piccola potenza di t che compare in $p_A(t)$ con coefficiente non nullo;
(2) la somma delle molteplicità algebriche degli autovalori positivi è uguale al numero dei cambiamenti di segno della successione ordinata dei coefficienti non nulli di $p_A(t)$;
(3) la somma delle molteplicità algebriche degli autovalori negativi è la differenza tra n e la somma dei due numeri precedenti.

Ad esempio, se sappiamo che il polinomio caratteristico di una matrice triangolabile reale è $t^4 - t^3 - 7t^2 + t + 6$, allora tale matrice ha zero autovalori nulli, due autovalori positivi e due autovalori negativi, tutti contati con molteplicità.

Per importanti motivi riguardanti le forme quadratiche, i prodotti scalari ed altre faccende che vanno al di là degli obiettivi di queste note, è molto utile avere dei criteri, in aggiunta alla regola dei segni di Cartesio, per stabilire se gli autovalori di una matrice simmetrica reale sono tutti positivi. Uno di questi, detto criterio di Sylvester è particolarmente interessante in quanto richiede un numero di calcoli sensibilmente inferiore a quello necessario per determinare il polinomio caratteristico.

A tale scopo, per ogni matrice $A \in M_{n,n}(\mathbb{R})$ e per ogni $k = 1, \dots, n$ denotiamo con $A[k] \in M_{k,k}(\mathbb{R})$ la sottomatrice formata dai coefficienti contenuti nelle prime k righe e k colonne. Ad esempio, se

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

allora

$$A[1] = (1), \qquad A[2] = \begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix}, \qquad A[3] = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Alcuni autori chiamano le sottomatrici $A[k]$ *sottomatrici principali di nord-ovest*.

DEFINIZIONE 10.6.3. Diremo che una una matrice simmetrica reale $A \in M_{n,n}(\mathbb{R})$ è **definita positiva** se $x^T A x > 0$ per ogni vettore $x \in \mathbb{R}^n$ diverso da 0.

Notiamo che in una matrice definita positiva $A = (a_{ij})$ i coefficienti sulla diagonale principale sono tutti positivi: infatti si ha $a_{ii} = e_i^T A e_i$ dove, come al solito $e_1, \dots, e_n$ indica la base canonica di $\mathbb{R}^n$. Il viceversa è generalmente falso per matrici di ordine $n \geq 2$. Ad esempio la matrice $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ non è definita positiva in quanto $(1, -1) \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -2$.

TEOREMA 10.6.4 (Criterio di Sylvester). *Per una matrice simmetrica reale $A \in M_{n,n}(\mathbb{R})$ le seguenti condizioni sono equivalenti:*

(1) *gli autovalori di A sono tutti positivi;*
(2) *A è definita positiva;*
(3) $\det(A[k]) > 0$ *per ogni* $k = 1, \dots, n$.

DIMOSTRAZIONE. Proviamo prima che una matrice simmetrica reale A è definita positiva se e solo se tutti i suoi autovalori sono positivi. Se A è definita positiva e λ è un autovalore con autovettore x si ha

$$\lambda \|x\|^2 = x^T(\lambda x) = x^T A x > 0$$

da cui segue $\lambda > 0$. Viceversa, se tutti gli autovalori sono positivi, segue dal Corollario 10.5.5 che $x^T A x > 0$ per ogni $x \neq 0$.

Proviamo adesso che le prime due condizioni implicano la terza. Per ogni $1 \leq k \leq n$ sia $i\colon \mathbb{R}^k \to \mathbb{R}^n$ l'applicazione $i(x_1, \dots, x_k)^T = (x_1, \dots, x_k, 0, \dots, 0)^T$; in particolare il sottospazio $i(\mathbb{R}^k)$ è quello dei vettori con le ultime $n-k$ coordinate uguali a 0. Notiamo che $i(x)^T A i(x) = x^T A[k] x$ per ogni $x \in \mathbb{R}^k$. Dunque se A è definita positiva, anche le matrici $A[k]$ sono definite positive, quindi hanno tutti gli autovalori positivi e di conseguenza i loro determinanti sono tutti positivi.

Per concludere, supponiamo $\det(A[k]) > 0$ per ogni k e proviamo che gli autovalori di A sono tutti positivi; per induzione su n possiamo assumere già dimostrato che la sottomatrice $A[n-1]$ è definita positiva. Siano adesso $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ gli autovalori di A contati con molteplicità e supponiamo per assurdo $\lambda_1 \leq 0$; siccome per ipotesi $\det(A) = \det(A[n]) = \lambda_1 \lambda_2 \cdots \lambda_n > 0$ deve essere necessariamente $\lambda_1 \leq \lambda_2 < 0$, ossia λ_1 e λ_2 entrambi negativi. Siano adesso u, v autovettori linearmente indipendenti per λ_1 e λ_2 rispettivamente. Per la formula di Grassmann $i(\mathbb{R}^{n-1}) \cap \mathrm{Span}(u, v) \neq 0$ ed esistono $x \in \mathbb{R}^{n-1}$, $x \neq 0$, $a, b \in \mathbb{R}$ tali che $i(x) = au + bv$. Si ha dunque

$$0 < x^T A[n-1] x = (au + bv)^T A (au + bv)$$

Se $\lambda_1 = \lambda_2 = \lambda$, si ha $A(au + bv) = \lambda(au + bv)$ e quindi

$$(au + bv)^T A (au + bv) = \lambda \|au + bv\|^2 \leq 0.$$

Se $\lambda_1 \neq \lambda_2$ abbiamo visto nella dimostrazione del Corollario 10.5.5 che $u^T A v = v^T A u = 0$ e quindi

$$(au + bv)^T A (au + bv) = \lambda_1 \|au\|^2 + \lambda_2 \|bv\|^2 \leq 0.$$

In entrambi i casi si arriva ad una contraddizione. □

Il criterio di Sylvester si presta ad un approccio algoritmico per stabilire se una matrice simmetrica reale A è definita positiva. Infatti se ad una riga aggiungiamo dei multipli scalari delle righe precedenti, i determinanti delle matrici $A[k]$ non cambiano. Di conseguenza si può procedere nel modo seguente:

(1) Si guarda il coefficiente a_{11}; se $a_{11} \leq 0$ allora la matrice non è definita positiva ed il processo si ferma.
(2) Se $a_{11} > 0$, aggiungendo alle righe 2,3,... opportuni multipli della prima riga si annullano i coefficienti a_{i1} per ogni $i > 1$.
(3) Nella matrice ottenuta si guarda il coefficiente a_{22}; se è ≤ 0 allora la matrice non è definita positiva.
(4) Se $a_{22} > 0$, aggiungendo alle righe 3,4,... opportuni multipli della seconda riga si annullano i coefficienti a_{i2} per ogni $i > 2$.
(5) e così via, mutatis mutandis, con le righe dalla tre in poi.

Se il processo non si interrompe e si arriva alla fine ad una matrice triangolare con tutti i coefficienti sulla diagonale positivi, allora la matrice è definita positiva.

OSSERVAZIONE 10.6.5. Il Teorema 10.6.4 contiene solo la prima parte del criterio di Sylvester. La seconda parte, che svilupperemo nel Capitolo 15 (vedi Esercizio 779), afferma che se i determinanti delle sottomatrici $A[k]$ sono tutti $\neq 0$, allora il numero di autovalori negativi della matrice A è uguale al numero dei cambiamenti di segno della successione di $n+1$ numeri reali

$$1,\ \det(A[1]),\ \det(A[2]),\ \dots,\ \det(A[n]).$$

Esercizi.

591. Mostrare con un esempio che il Corollario 10.6.2 è in generale falso se il polinomio $p(t)$ possiede alcune radici complesse non reali.

592. Sia $p(t) = \sum_{i=0}^{n} a_i t^i$ un polinomio di grado n con tutte le radici reali non nulle. Provare che se $a_i = 0$ allora a_{i-1} e a_{i+1} sono $\neq 0$ ed hanno segni opposti. (Suggerimento: guardare alla dimostrazione del Corollario 10.6.2.)

593. Mostrare che:
(1) l'equazione $x^{1901} = 10 + \sum_{i=1}^{935} x^{2i}$ possiede una sola soluzione reale;
(2) il polinomio $t^7 + t^6 + t^4 + t^2$ possiede esattamente 3 radici reali contate con molteplicità.

594. Sia $p(t)$ un polinomio avente tutte le radici reali e siano $a < b \in \mathbb{R}$ tali che $p(a), p(b) \neq 0$. Determinare una formula per il numero di radici comprese tra a e b.

595. Usando il Corollario 10.6.2 e l'Esercizio 149, determinare una condizione necessaria affinché un polinomio in $\mathbb{R}[t]$ abbia tutte le radici reali.

596. Provare che, nella dimostrazione del Teorema 10.6.1 la differenza $s((t-c)p(t)) - s(p(t))$ è un numero dispari.

597. Usando il criterio di Sylvester, provare che la matrice simmetrica

$$\begin{pmatrix} 1 & 1 & a \\ 1 & 2 & b \\ a & b & -3 \end{pmatrix}$$

ha determinante negativo per qualunque scelta di $a, b \in \mathbb{R}$.

598 (Decomposizione LU). Sia $A \in M_{n,n}(\mathbb{K})$, non necessariamente simmetrica, che abbia tutte le sottomatrici principali di nord–ovest invertibili. Provare che A si fattorizza in modo unico come $A = LU$, con L triangolare inferiore e U triangolare superiore con i coefficienti sulla diagonale principale uguali a 1. (Sugg.: denotiamo con $e_1, \dots, e_n$ la base canonica e con $\pi_i \colon \mathbb{K}^n \to \mathbb{K}^i$ la proiezione sulle prime coordinate. Provare che per ogni $i = 0, \dots, n-1$ esiste unico un vettore $v_i \in \operatorname{Span}(e_1, \dots, e_i)$ tale che $\pi_i L_A(e_{i+1} + v_i) = 0$.)

599. Siano $A, B \in M_{n,n}(\mathbb{R})$ matrici simmetriche definite positive. È vero o falso che la matrice simmetrica $AB + BA$ è definita positiva?

600 (Ⓐ). Per ogni Siano $n > 0$; provare che la matrice simmetrica reale di coefficienti

$$a_{ij} = \frac{1}{i+j-1} = \int_0^1 t^{i-1} t^{j-1}\, dt, \qquad 1 \leq i, j \leq n.$$

è definita positiva. Suggerimento leggermente criptico:

$$x^T A y = \int_0^1 \left(\sum_{i=1}^{n} x_i t^{i-1} \right) \left(\sum_{j=1}^{n} y_j t^{j-1} \right) dt.$$

10.7. Complementi: il teorema di Cayley–Hamilton–Frobenius

In questa sezione daremo una diversa dimostrazione, ed al tempo stesso una estensione, del teorema di Cayley–Hamilton. Più precisamente, oltre a dimostrare che il polinomio minimo $q(t)$ di una matrice divide il polinomio caratteristico $p(t)$, scopriremo alcune interessanti proprietà del quoziente $p(t)/q(t)$.

Ogni matrice a coefficienti polinomi $B(t) \in M_{n,n}(\mathbb{K}[t])$ può anche essere pensata come un polinomio a coefficienti matrici, ossia possiamo scrivere

$$B(t) = \sum_{i=0}^{N} B_i t^i, \qquad B_i \in M_{n,n}(\mathbb{K}).$$

La sostituzione dell'indeterminata t con una qualunque matrice $A \in M_{n,n}(\mathbb{K})$, definisce un'applicazione lineare

$$\varphi_A \colon M_{n,n}(\mathbb{K}[t]) \to M_{n,n}(\mathbb{K}), \qquad \varphi_A\left(\sum_i B_i t^i\right) = \sum_i B_i A^i.$$

Notiamo in particolare che per ogni polinomio $h(t) \in \mathbb{K}[t]$ si ha $\varphi_A(h(t)I) = h(A)$.

LEMMA 10.7.1. *Siano $B(t), C(t) \in M_{n,n}(\mathbb{K}[t])$. Per ogni matrice $A \in M_{n,n}(\mathbb{K})$ tale che $AC(t) = C(t)A$ vale la formula:*

$$\varphi_A(B(t)C(t)) = \varphi_A(B(t))\,\varphi_A(C(t)).$$

DIMOSTRAZIONE. Se

$$B = \sum_i B_i t^i, \qquad C = \sum_j C_j t^j,$$

la condizione $AC(t) = C(t)A$ equivale a $AC_j = C_j A$ per ogni indice j. Allora si ha anche $A^i C_j = C_j A^i$ per ogni i, j e quindi

$$\varphi_A(B(t)C(t)) = \varphi_A(\sum_{i,j} B_i C_j t^{i+j}) = \sum_{i,j} B_i C_j A^{i+j},$$

$$\varphi_A(B(t))\varphi_A(C(t)) = (\sum_i B_i A^i)(\sum_j C_j A^j) = \sum_{i,j} B_i A^i C_j A^j = \sum_{i,j} B_i C_j A^{i+j}.$$

□

LEMMA 10.7.2. *Siano $A \in M_{n,n}(\mathbb{K})$ e $B(t) \in M_{n,n}(\mathbb{K}[t])$ tali che $B(t)(A - tI) = h(t)I$ con $h(t) \in \mathbb{K}[t]$. Allora vale $h(A) = 0$.*

DIMOSTRAZIONE. Siccome A commuta con $A - tI$ e $\varphi_A(A - tI) = A - AI = 0$, dal Lemma 10.7.1 segue immediatamente che

$$h(A) = \varphi_A(h(t)I) = \varphi_A(B(t)(A - tI)) = \varphi_A(B(t))\varphi_A(A - tI) = 0.$$

□

TEOREMA 10.7.3 (Cayley–Hamilton–Frobenius). *Siano $A \in M_{n,n}(\mathbb{K})$, $p_A(t) = \det(A - tI)$ il suo polinomio caratteristico, $q_A(t)$ il suo polinomio minimo e $B(t) \in M_{n,n}(\mathbb{K}[t])$ l'aggiunta classica di $A - tI$. Allora:*

(1) *$p_A(A) = 0$, e quindi $q_A(t)$ divide $p_A(t)$;*
(2) *il polinomio $p_A(t)/q_A(t)$ divide tutti i coefficienti di $B(t)$;*
(3) *se $s(t) \in \mathbb{K}[t]$ divide tutti i coefficienti di $B(t)$, allora $s(t)$ divide $p_A(t)/q_A(t)$.*

DIMOSTRAZIONE. Siccome $B(t)(A - tI) = p_A(t)I$, il primo punto segue immediatamente dal Lemma 10.7.2. Se

$$q_A(t) = a_0 + a_1 t + \cdots + a_d t^d, \qquad a_d = 1,$$

poiché tI commuta con $A - tI$ possiamo scrivere:

$$\begin{aligned}
0 &= \sum_{i=0}^{d} a_i A^i = \sum_{i=0}^{d} a_i ((A - tI) + tI)^i \\
&= \sum_{i=0}^{d} a_i \sum_{j=0}^{i} \binom{i}{j} (tI)^{i-j} (A - tI)^j = \sum_{j=0}^{d} (A - tI)^j \sum_{i=j}^{d} a_i \binom{i}{j} t^{i-j} \\
&= \sum_{i=0}^{d} a_i \binom{i}{0} t^i + (A - tI) \sum_{j=1}^{d} (A - tI)^{j-1} \sum_{i=j}^{d} a_i \binom{i}{j} t^{i-j} \\
&= q_A(t)I - (A - tI)D(t), \qquad D(t) \in M_{n,n}(\mathbb{K}[t]).
\end{aligned}$$

Moltiplicando prima per $B(t)$ e dividendo poi per $q_A(t)$ la relazione $q_A(t)I = (A - tI)D(t)$ si ottiene

$$q_A(t)B(t) = B(t)(A - tI)D(t) = p_A(t)D(t), \qquad B(t) = \frac{p_A(t)}{q_A(t)} D(t),$$

e questo dimostra che $p_A(t)/q_A(t)$ divide tutti i coefficienti di $B(t)$. Viceversa, se vale una formula del tipo $B(t) = s(t)C(t)$, con $s(t) \in \mathbb{K}[t]$ e $C(t) \in M_{n,n}(\mathbb{K}[t])$, ossia se $s(t)$ divide in $\mathbb{K}[t]$ tutti i coefficienti dell'aggiunta classica $B(t)$, allora

$$s(t)C(t)(A - tI) = B(t)(A - tI) = p_A(t)I$$

e dunque $s(t)$ divide $p_A(t)$; se denotiamo $h(t) = p_A(t)/s(t)$ vale $C(t)(A - tI) = h(t)I$, e dal Lemma 10.7.2 segue che $h(A) = 0$ e di conseguenza $q_A(t)$ divide $h(t)$, ossia $q_A(t)s(t)$ divide $p_A(t)$. □

Esercizi.

601. Sia $A \in M_{n,n}(\mathbb{K})$ una matrice nilpotente. Calcolare il polinomio $\det(I - tA) \in \mathbb{K}[t]$.

602. Date $A, B \in M_{n,n}(\mathbb{K})$, provare che la matrice $A + tB \in M_{n,n}(\mathbb{K}[t])$ è nilpotente se e solo se A è invertibile e BA^{-1} è nilpotente.

603. Sia $A \in M_{n,n}(\mathbb{K})$ una matrice fissata e denotiamo

$$R_A = \{B(t) \in M_{n,n}(\mathbb{K}[t]) \mid AB(t) = B(t)A\}.$$

Dimostrare che: R_A è chiuso per prodotto, che $\sum B_i t^i \in R_A$ se e solo se $AB_i = B_i A$ per ogni i e che l'aggiunta classica di $A - tI$ appartiene ad R_A.

604. Si consideri una matrice compagna a blocchi

$$B = \begin{pmatrix} 0 & 0 & \cdots & 0 & A_n \\ I & 0 & \cdots & 0 & A_{n-1} \\ 0 & I & \cdots & 0 & A_{n-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & A_1 \end{pmatrix}$$

dove $I, A_1, \ldots, A_n \in M_{p,p}(\mathbb{K})$ e I è la matrice identità. Dimostrare che gli autovalori di B coincidono con le radici del polinomio $\det(L(t))$, dove

$$L(t) = It^n - A_1 t^{n-1} - \cdots - A_n \in M_{p,p}(\mathbb{K}[t]).$$

CAPITOLO 11

Autospazi generalizzati e forma canonica di Jordan

Adesso che sappiamo che le proprietà, per un endomorfismo, di essere diagonalizzabile e/o triangolabile dipendono solamente dal polinomio minimo, possiamo fare un passo ulteriore nella teoria, dimostrando che ogni endomorfismo triangolabile si può rappresentare con una matrice, detta di Jordan, che ha una forma particolarmente semplice, con molti coefficienti uguali a zero e che, per gli endomorfismi diagonalizzabili si riduce ad una matrice diagonale.

In tutto il capitolo, il simbolo V denoterà uno spazio vettoriale di dimensione finita su di un campo $\mathbb{K}$.

11.1. La decomposizione di Fitting

Dato un endomorfismo $f\colon V \to V$, ha senso considerare nucleo ed immagine delle potenze di f; ricordiamo che per convenzione si pone f^0 uguale al morfismo identità. Per ogni intero $h \geq 0$ si ha

$$\operatorname{Ker}(f^h) \subseteq \operatorname{Ker}(f^{h+1}), \qquad f^{h+1}(V) \subseteq f^h(V).$$

Infatti, se $v \in \operatorname{Ker}(f^h)$, allora $f^{h+1}(v) = f(f^h(v)) = f(0) = 0$ e quindi $v \in \operatorname{Ker}(f^{h+1})$. Similmente, se $v \in f^{h+1}(V)$, allora esiste $u \in V$ tale che $v = f^{h+1}(u) = f^h(f(u))$ e quindi $v \in f^h(V)$.

Dunque, ad ogni endomorfismo f possiamo associare la **filtrazione (crescente) dei nuclei**

$$0 = \operatorname{Ker}(f^0) \subseteq \operatorname{Ker}(f) \subseteq \operatorname{Ker}(f^2) \subseteq \operatorname{Ker}(f^3) \subseteq \cdots,$$

e la **filtrazione (decrescente) delle immagini**

$$V = f^0(V) \supseteq f(V) \supseteq f^2(V) \supseteq f^3(V) \supseteq \cdots.$$

Se V ha dimensione finita, allora le dimensioni dei sottospazi $\operatorname{Ker}(f^h)$ possono assumere al più un numero finito dei valori e per il principio dei cassetti esiste un intero $0 \leq k \leq \dim V$ tale che $\operatorname{Ker}(f^k) = \operatorname{Ker}(f^{k+1})$. Inoltre, per il teorema del rango si ha

$$\dim \operatorname{Ker}(f^h) + \dim f^h(V) = \dim V = \dim \operatorname{Ker}(f^{h+1}) + \dim f^{h+1}(V),$$

di conseguenza

$$\dim \operatorname{Ker}(f^{h+1}) - \dim \operatorname{Ker}(f^h) = \dim f^h(V) - \dim f^{h+1}(V) \geq 0$$

ed in particolare

$$\operatorname{Ker}(f^k) = \operatorname{Ker}(f^{k+1}) \quad \text{se e solo se} \quad f^k(V) = f^{k+1}(V).$$

Lemma 11.1.1. *Sia f un endomorfismo lineare di uno spazio vettoriale V di dimensione finita. Se $\operatorname{Ker}(f^k) = \operatorname{Ker}(f^{k+1})$ per qualche $k \geq 0$, allora $\operatorname{Ker}(f^h) = \operatorname{Ker}(f^{h+1})$ e $f^h(V) = f^{h+1}(V)$ per ogni $h \geq k$. Inoltre, la successione di interi non negativi*

$$\alpha_h = \dim \operatorname{Ker}(f^h) - \dim \operatorname{Ker}(f^{h-1}) = \dim f^{h-1}(V) - \dim f^h(V) \geq 0, \qquad h > 0\,,$$

è monotona decrescente, ossia

$$\alpha_1 \geq \alpha_2 \geq \alpha_3 \geq \cdots.$$

Dimostrazione. Per dimostrare che $\alpha_h \geq \alpha_{h+1}$ consideriamo la restrizione di f al sottospazio $f^{h-1}(V)$:

$$f_{|f^{h-1}(V)}\colon f^{h-1}(V) \to V.$$

L'immagine di tale applicazione è il sottospazio $f(f^{h-1}(V)) = f^h(V)$, mentre il nucleo è uguale a $\operatorname{Ker}(f) \cap f^{h-1}(V)$. Per la formula di Grassmann si ha

$$\alpha_h = \dim f^{h-1}(V) - \dim f^h(V) = \dim \operatorname{Ker}(f_{|f^{h-1}(V)})$$

e quindi

$$\alpha_h = \dim(\operatorname{Ker}(f) \cap f^{h-1}(V)), \qquad h > 0\,. \tag{11.1}$$

Siccome $f^h(V) \subseteq f^{h-1}(V)$ si ha

$$\operatorname{Ker}(f) \cap f^h(V) \subseteq \operatorname{Ker}(f) \cap f^{h-1}(V)$$

e dunque $\alpha_{h+1} \leq \alpha_h$.

Se si ha $\operatorname{Ker}(f^k) = \operatorname{Ker}(f^{k+1})$ (ossia $\alpha_{k+1} = 0$), allora $\operatorname{Ker}(f^h) = \operatorname{Ker}(f^{h+1})$ (ossia $\alpha_{h+1} = 0$) per ogni $h \geq k$ e quindi che $\operatorname{Ker}(f^k) = \operatorname{Ker}(f^{k+1}) = \cdots = \operatorname{Ker}(f^h)$ per ogni $h \geq k$. □

Nelle notazioni del Lemma 11.1.1, notiamo che per ogni $h < k$ vale

$$\alpha_{h+1} + \alpha_{h+1} + \cdots + \alpha_k = \dim \operatorname{Ker}(f^k) - \dim \operatorname{Ker}(f^h),$$

ed in particolare per ogni $h > 0$ si ha

$$\alpha_1 + \alpha_2 + \cdots + \alpha_h = \dim \operatorname{Ker}(f^h) \leq \dim V.$$

Esempio 11.1.2. Per ogni endomorfismo $f\colon V \to V$ di uno spazio di dimensione finita e per ogni intero $h > 0$ vale la formula

$$\operatorname{rg}(f^h) \geq h \operatorname{rg}(f) - (h-1) \dim V.$$

Infatti, per la monotonia della successione $\alpha_1 \geq \alpha_2 \geq \cdots$ si ha

$$\operatorname{rg}(f) - \operatorname{rg}(f^h) = \alpha_2 + \cdots + \alpha_h \leq (h-1)\alpha_1 = (h-1)(\dim V - \operatorname{rg}(f)).$$

Quindi, se ad esempio $\operatorname{rg}(f) = \dim(V) - 1$, allora $\operatorname{rg}(f^h) \geq \dim(V) - h$ per ogni $h > 0$.

Abbiamo visto che in dimensione finita i nuclei delle potenze di f si stabilizzano e quindi esiste un intero $h \leq \dim V$, dipendente da f, tale che $\operatorname{Ker}(f^h) = \operatorname{Ker}(f^k)$ e $f^h(V) = f^k(V)$ per ogni $k \geq h$.

Definizione 11.1.3. Sia $f\colon V \to V$ un endomorfismo lineare di uno spazio vettoriale di dimensione finita e sia $k \geq 0$ un intero tale che $\operatorname{Ker}(f^h) = \operatorname{Ker}(f^{h+1})$ per ogni $h \geq k$. I sottospazi vettoriali

$$F_0(f) = \operatorname{Ker}(f^k) = \operatorname{Ker}(f^{k+1}) = \operatorname{Ker}(f^{k+2}) = \cdots,$$
$$F_1(f) = f^k(V) = f^{k+1}(V) = f^{k+2}(V) = \cdots,$$

vengono detti rispettivamente le **componenti zero-Fitting** e **uno-Fitting** di f.[1] Per semplicità notazionale scriveremo semplicemente F_0, F_1 quando non vi sono ambiguità sull'endomorfismo.

Segue immediatamente dalle definizioni che f è nilpotente se e solo se $F_0 = V$.

Teorema 11.1.4 (Decomposizione di Fitting). *Siano F_0, F_1 le componenti di Fitting di un endomorfismo $f\colon V \to V$ di uno spazio vettoriale di dimensione finita. Allora:*

1. *$V = F_0 \oplus F_1$;*
2. *$f(F_0) \subseteq F_0$, $f(F_1) \subseteq F_1$, ossia le componenti di Fitting sono f-invarianti;*
3. *$f_{|F_0}\colon F_0 \to F_0$ è nilpotente;*
4. *$f_{|F_1}\colon F_1 \to F_1$ è un isomorfismo.*

Dimostrazione. Fissiamo un intero $k \geq 0$ abbastanza grande tale che $F_0 = \operatorname{Ker}(f^h)$ e $F_1 = f^h(V)$ per ogni $h \geq k$. Per il teorema del rango

$$\dim F_0 + \dim F_1 = \dim \operatorname{Ker}(f^k) + \dim f^k(V) = \dim V$$

e quindi, per dimostrare che $V = F_0 \oplus F_1$ basta provare che $F_0 \cap F_1 = 0$. Se $v \in F_0 \cap F_1$, allora $f^k(v) = 0$ ed esiste $u \in V$ tale che $v = f^k(u)$. Ma allora $f^{2k}(u) = f^k(v) = 0$ e quindi $u \in \operatorname{Ker}(f^{2k})$. Per come abbiamo scelto k, si ha $\operatorname{Ker}(f^{2k}) = \operatorname{Ker}(f^k)$ e quindi $v = f^k(u) = 0$.

La f-invarianza di nucleo ed immagine di f^k sono entrambe dimostrate nell'Esempio 9.5.8.

Essendo l'applicazione $f\colon f^k(V) \to f^{k+1}(V)$ surgettiva e $f^k(V) = f^{k+1}(V) = F_1$, si ha che $f_{|F_1}\colon F_1 \to F_1$ è surgettiva e quindi anche un isomorfismo. Per costruzione $f^k(F_0) = 0$ e quindi $f_{|F_0}$ è nilpotente. □

[1] Fitting richiede la maiuscola perché si riferisce al matematico tedesco Hans Fitting (1906-1938) e non ad un termine inglese.

COROLLARIO 11.1.5. *Siano V uno spazio vettoriale di dimensione n, $f\colon V \to V$ un endomorfismo e $r = \dim F_0(f)$. Allora esiste una base di V rispetto alla quale f si rappresenta con una matrice diagonale a blocchi*

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

con A matrice $r \times r$ triangolare strettamente superiore e B matrice $(n-r)\times(n-r)$ invertibile. Inoltre vale $A^k = 0$ se e solo se $F_0(f) = \operatorname{Ker}(f^k)$.

DIMOSTRAZIONE. Siano F_0, F_1 le componenti di Fitting di f. Per il Teorema 11.1.4 è sufficiente prendere una base $v_1, \dots, v_n$ di V tale che $v_{r+1}, \dots, v_n \in F_1$ e $v_1, \dots, v_r$ è una base di F_0 rispetto alla quale l'endomorfismo nilpotente $f_{|F_0}$ si rappresenta con una matrice triangolare strettamente superiore. □

COROLLARIO 11.1.6. *Sia f un endomorfismo di uno spazio vettoriale V di dimensione n su di un campo $\mathbb{K}$ di caratteristica 0 oppure di caratteristica $p > n$. Allora f è nilpotente se e solo se $\operatorname{Tr}(f) = \operatorname{Tr}(f^2) = \cdots = \operatorname{Tr}(f^n) = 0$.*

DIMOSTRAZIONE. Se f è nilpotente, abbiamo già visto che in una opportuna base f si rappresenta con una matrice strettamente triangolare e quindi f^k ha traccia nulla per ogni $k > 0$; questa implicazione è vera su ogni campo.

Viceversa, supponiamo che $\operatorname{Tr}(f) = \operatorname{Tr}(f^2) = \cdots = \operatorname{Tr}(f^n) = 0$; allora per la decomposizione di Fitting possiamo rappresentare f con una matrice diagonale a blocchi

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

con A nilpotente e B invertibile. Dunque f^k è rappresentata dalla matrice

$$\begin{pmatrix} A^k & 0 \\ 0 & B^k \end{pmatrix}.$$

Dunque $\operatorname{Tr}(B^k) = \operatorname{Tr}(f^k) - \operatorname{Tr}(A^k) = 0$ per ogni $k = 1, \dots, n$. Sia $m \le n$ l'ordine della matrice B; se $m > 0$ per Cayley–Hamilton si ha

$$0 = p_B(B) = \det(B)I + a_1 B + \cdots + a_m B^m$$

per opportuni coefficienti $a_1, \dots, a_m \in \mathbb{K}$. Quindi

$$\det(B)I = -a_1 B - \cdots - a_m B^n$$

e per la linearità della traccia

$$0 \neq \det(B)\operatorname{Tr}(I) = -a_1 \operatorname{Tr}(B) - \cdots - a_m \operatorname{Tr}(B^m).$$

Abbiamo quindi una contraddizione e dunque $m = 0$. Si noti che se il campo ha caratteristica $p > 0$, tutte le potenze dell'identità su $\mathbb{K}^p$ hanno traccia nulla. □

ESEMPIO 11.1.7. La decomposizione di Fitting permette di ricondurre il problema dell'estrazione della radice quadrata di un endomorfismo ai casi nilpotente (vedi Esercizio 623) ed invertibile. Più precisamente, siano $f, g\colon V \to V$ endomorfismi tali che $g^2 = f$, allora $gf = fg$ e quindi i sottospazi $\operatorname{Ker} f^k$, $f^k(V)$ sono g-invarianti per ogni $k > 0$. In particolare

$$g\colon F_0(f) \to F_0(f), \qquad g\colon F_1(f) \to F_1(f).$$

Esercizi.

605. Dato un endomorfismo $f\colon V \to V$ e due interi positivi a, b, provare che

$$f^a(\operatorname{Ker}(f^{a+b})) \subseteq \operatorname{Ker}(f^b).$$

606. Sia $f\colon \mathbb{K}^5 \to \mathbb{K}^5$ endomorfismo nilpotente di rango 2. Provare che se $f = g^k$ per qualche intero $k > 1$ e qualche $g\colon \mathbb{K}^5 \to \mathbb{K}^5$, allora $f^2 = 0$ e $k \le 3$.

607 (Unicità della decomposizione di Fitting). Sia $f\colon V \to V$ un endomorfismo di uno spazio vettoriale di dimensione finita e sia $U \oplus W = V$ una decomposizione in somma diretta di sottospazi f-invarianti tali che

(1) $f_{|U}\colon U \to U$ è nilpotente,
(2) $f_{|W}\colon W \to W$ è un isomorfismo.

Provare che $U = F_0(f)$ e $V = F_1(f)$.

608. Siano $f\colon V \to V$ un endomorfismo nilpotente e $U \subseteq V$ un sottospazio f-invariante. Denotiamo $U_0 = U$ e per ogni intero $i \geq 0$

$$U_i = \{v \in V \mid f^i(v) \in U\}.$$

Provare che ogni U_i è un sottospazio vettoriale e che $U_{i-1} \subseteq U_i$, $f(U_i) \subseteq U_{i-1}$ per ogni $i > 0$.

609. Usare la disuguaglianza dell'Esempio 11.1.2 per dimostrare che se $f\colon V \to V$ è nilpotente con indice di nilpotenza r, allora

$$\dim V - r + 1 \geq \dim(\mathrm{Ker}(f)) \geq \frac{\dim V}{r}.$$

610. Sia V spazio vettoriale di dimensione finita e siano $f, g\colon V \to V$ endomorfismmi nilpotenti con indici di nipotenza a, b. Provare che se $fg = 0$ allora $a + b \leq \dim V + 2$.

611. Trovare una matrice invertibile $B \in M_{n,n}(\mathbb{K})$ tale che $\mathrm{Tr}(B^i) = 0$ per ogni $1 \leq i < n$.

612. Sia $f\colon V \to V$ un endomorfismo. Mostrare con un esempio che in generale $V \neq \mathrm{Ker}(f) \oplus f(V)$.

613. Sia $f\colon V \to V$ endomorfismo nilpotente con indice di nilpotenza σ e sia $v \in V$ un vettore tale che $f^{\sigma-1}(v) \neq 0$. Dimostrare che il sottospazio

$$U = \mathrm{Span}(v, f(v), \dots, f^{\sigma-1}(v))$$

è f-invariante di dimensione σ. Calcolare inoltre le dimensioni di $U \cap \mathrm{Ker}(f^i)$ e $U \cap f^i(V)$ per ogni intero $i \geq 0$.

614. Sia $f\colon V \to V$ endomorfismo nilpotente con indice di nilpotenza σ. Provare che

$$\frac{\dim V}{\dim \mathrm{Ker}(f)} \leq \sigma \leq \dim V$$

e che vale $\sigma = \dim V$ se e solo se esiste una base in cui f si rappresenta con il blocco di Jordan J_n.

615 (☕, ♡). Sia $f\colon \mathbb{C}^n \to \mathbb{C}^n$ un endomorfismo lineare. Dimostrare che f è nilpotente se e solo se la sua immagine è contenuta nell'immagine di $f - \lambda I$, per ogni $\lambda \in \mathbb{C}$.

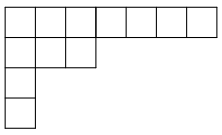
11.2. Diagrammi di Young

Una **partizione** è una successione $(\alpha_1, \alpha_2, \dots)$ di interi non negativi tale che $\alpha_i \geq \alpha_{i+1}$ per ogni indice i e $\alpha_j = 0$ per ogni j sufficientemente grande. Per partizione di un intero positivo n si intende una partizione $(\alpha_1, \alpha_2, \dots)$ tale che $\sum_i \alpha_i = n$.

Ad esempio le partizioni di 4 sono (vengono riportati solo i termini positivi):

$$(4), \qquad (3,1), \qquad (2,2), \qquad (2,1,1) \qquad (1,1,1,1).$$

Un altro modo di scrivere le partizioni è mediante i cosiddetti **diagrammi di Young**. Il diagramma di Young della partizione $(\alpha_1, \alpha_2, \dots)$ è semplicemente un diagramma a forma di scala, in cui la prima riga è costituita da α_1 quadretti, la seconda riga da α_2 quadretti, e così via.[2] Ad esempio, il diagramma di Young della partizione di 12 data da $(7, 3, 1, 1)$ è



.

Si noti che il diagramma di Young di una qualunque partizione di n contiene esattamente n quadretti.

Sia V spazio vettoriale di dimensione finita, ad ogni endomorfismo $f\colon V \to V$ possiamo associare in maniera canonica una partizione di $r = \dim F_0(f)$ e di conseguenza un diagramma di Young.

[2] La scelta dei quadretti, piuttosto che pallini, cuoricini o tazzine da caffè, è dovuta al fatto che i diagrammi di Young servono da base per i cosiddetti tableaux di Young, usati in teoria delle rappresentazioni. Un tableau di Young è un diagramma di Young in cui ogni quadrato contiene al suo interno un intero positivo.

Infatti per il Lemma 11.1.1 la successione degli interi

$$\alpha_h = \dim \mathrm{Ker}(f^h) - \dim \mathrm{Ker}(f^{h-1}) = \dim f^{h-1}(V) - \dim f^h(V) \geq 0, \qquad h > 0,$$

è monotona decrescente e siccome $\mathrm{Ker}(f^h) = F_0(f)$ per h sufficientemente grande si ha $\alpha_1 + \alpha_2 + \cdots = r$.

Allo stesso modo, per ogni matrice $A \in M_{n,n}(\mathbb{K})$ possiamo associare una partizione e quindi un diagramma di Young in cui, per costruzione, il numero di caselle nella riga i è uguale a $\mathrm{rg}(A^{i-1}) - \mathrm{rg}(A^i)$, $i > 0$.

Esempio 11.2.1. Se

$$J_n = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \in M_{n,n}(\mathbb{K})$$

indica il blocco di Jordan nilpotente di ordine n, abbiamo già dimostrato che per ogni $m \leq n$ il rango di J_n^m è uguale a $n - m$ e quindi la partizione associata è $(1, 1, \ldots, 1)$.

Dunque se prendiamo un matrice diagonale a blocchi

$$A = \begin{pmatrix} J_{k_1} & 0 & \cdots & 0 \\ 0 & J_{k_2} & \vdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{k_s} \end{pmatrix} \tag{11.2}$$

dove ciascun J_{k_i} è un blocco di Jordan nilpotente di ordine k_i si ha che per ogni m il rango di A^m è uguale alla somma dei ranghi di $J_{k_i}^m$ ed è quindi uguale a

$$\mathrm{rg}(A^m) = \sum_{i=1}^{s} \max(0, k_i - m).$$

Di conseguenza, la partizione associata alla matrice (11.2) è:

- $\alpha_1 = \mathrm{rg}(A^0) - \mathrm{rg}(A^1) = s =$ numero dei blocchi di Jordan,
- $\alpha_2 = \mathrm{rg}(A^1) - \mathrm{rg}(A^2) =$ numero di indici i tali che $k_i \geq 2$,

$\vdots$

- $\alpha_m = \mathrm{rg}(A^{m-1}) - \mathrm{rg}(A^m) =$ numero di indici i tali che $k_i \geq m$.

Lasciamo come semplice esercizio per il lettore la dimostrazione che gli interi $k_1, \ldots, k_s > 0$ sono esattamente le lunghezze, contate con ripetizioni, delle colonne del diagramma di Young (cf. Esercizio 50).

L'Esempio 11.2.1 implica in particolare che ogni diagramma di Young si può ricavare da una matrice nilpotente secondo le regole descritte precedentemente. Ad esempio per le partizioni di $1, 2, 3, 4$ si ha:

$$(0) \mapsto \square \qquad \left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & 0 \end{array}\right) \mapsto \square\square \qquad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \mapsto \begin{array}{l}\square\\\square\end{array}$$

$$\left(\begin{array}{c|c|c} 0 & 0 & 0 \\ \hline 0 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array}\right) \mapsto \square\square\square \qquad \left(\begin{array}{cc|c} 0 & 1 & 0 \\ 0 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array}\right) \mapsto \begin{array}{ll}\square&\square\\\square\end{array} \qquad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \mapsto \begin{array}{l}\square\\\square\\\square\end{array}$$

$$\left(\begin{array}{c|c|c|c} 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right) \mapsto \square\square\square\square \qquad \left(\begin{array}{cc|c|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right) \mapsto \begin{array}{lll}\square&\square&\square\\\square\end{array}$$

$$\left(\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right) \mapsto \begin{array}{|c|c|}\hline & \\ \hline & \\ \hline \end{array} \qquad \left(\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right) \mapsto \begin{array}{|c|c|}\hline & \\ \hline \\ \cline{1-1} \\ \cline{1-1} \end{array} \qquad \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \mapsto \begin{array}{|c|}\hline \\ \hline \\ \hline \\ \hline \\ \hline \end{array}$$

Se due matrici nilpotenti A, B sono simili, allora A^h e B^h sono simili per ogni $h \geq 0$ e, poiché matrici simili hanno lo stesso rango, deduciamo che matrici simili hanno lo stesso diagramma di Young. Vale anche il viceversa, ossia **due matrici nilpotenti hanno lo stesso diagramma di Young se e solo se sono simili**, come segue immediatamente dalle precedenti osservazioni sulle matrici a blocchi di Jordan e dal seguente teorema.

TEOREMA 11.2.2. *Per ogni endomorfismo nilpotente $f\colon V \to V$ di uno spazio vettoriale di dimensione finita, esiste una base rispetto alla quale f si rappresenta con una matrice diagonale a blocchi*

$$\begin{pmatrix} J_{k_1} & 0 & \cdots & 0 \\ 0 & J_{k_2} & \vdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{k_s} \end{pmatrix}$$

dove ogni J_{k_i} è un blocco di Jordan nilpotente e $k_1 \geq k_2 \geq \cdots \geq k_s > 0$.

DIMOSTRAZIONE. Sia $n = \dim V$; abbiamo già osservato che il diagramma di Young di f contiene esattamente n caselle, $\dim \operatorname{Ker}(f)$ colonne e τ righe, dove τ è l'indice di nilpotenza di f. Per la Formula (11.1) le lunghezze delle righe del diagramma, dall'alto verso il basso, sono uguali ai valori non nulli della successione:

$$\alpha_1 = \dim \operatorname{Ker}(f) \ \geq \cdots \geq \ \alpha_s = \dim(f^{s-1}(V) \cap \operatorname{Ker}(f)) \ \geq \cdots .$$

Partendo da una base di $f^{\tau-1}(V) \cap \operatorname{Ker}(f)$, si può completare ad una base di $f^{\tau-2}(V) \cap \operatorname{Ker}(f)$, completarla ancora ad una base di $f^{\tau-3}(V) \cap \operatorname{Ker}(f)$ e così via. Abbiamo quindi costruito una base $v_1, \ldots, v_{\alpha_1}$ di $\operatorname{Ker}(f)$ in modo tale che per ogni s i vettori $v_1, \ldots, v_{\alpha_s}$ siano una base di $f^{s-1}(V) \cap \operatorname{Ker}(f)$.

Per ogni indice $i = 1, \ldots, \alpha_1$ esiste un intero $h > 0$ tale che $\alpha_h \geq i > \alpha_{h+1}$ e quindi tale che $v_i \in f^{h-1}(V) \cap \operatorname{Ker}(f)$ e $v_i \notin f^h(V)$. Possiamo allora scegliere un vettore $u_i \in V$ tale che

$$f^{h-1}(u_i) = v_i, \qquad u_i \notin f(V), \qquad f^h(u_i) = f(v_i) = 0.$$

Possiamo raffigurare graficamente la situazione infilando i vettori v_i nelle ultime caselle ed i vettori u_i nelle prime caselle delle colonne del diagramma di Young:

$$\begin{array}{|c|c|c|c|c|}\hline & & & & v_5 \\ \hline & & v_3 & v_4 \\ \cline{1-4} v_1 & v_2 \\ \cline{1-2} \end{array} \qquad \begin{array}{|c|c|c|c|c|}\hline u_1 & u_2 & u_3 & u_4 & u_5 \\ \hline & & & \\ \cline{1-4} & \\ \cline{1-2} \end{array} \qquad \begin{array}{l} u_5 = v_5 \\ f(u_3) = v_3,\ f(u_4) = v_4 \\ f^2(u_1) = v_1,\ f^2(u_2) = v_2 \end{array}$$

Per concludere la dimostrazione basta provare che la successione di vettori

$$\begin{array}{l} v_1, \ldots, f^2(u_1), f(u_1), u_1, \\ v_2, \ldots, f^2(u_2), f(u_2), u_2, \\ \cdots \\ v_{\alpha_1}, \ldots, f^2(u_{\alpha_1}), f(u_{\alpha_1}), u_{\alpha_1}, \end{array} \tag{11.3}$$

è una base di V: infatti in tale base f viene rappresentata con una matrice diagonale a blocchi di Jordan. Siccome i precedenti vettori vengono inseriti bigettivamente ed in maniera naturale nelle caselle del diagramma di Young ($u_1, u_2, \ldots$ nella prima riga da sinistra a destra; per spostarsi di una casella verso il basso si applica f) il loro numero è uguale a n e basta dimostrare che sono linearmente indipendenti.

Data una qualunque combinazione lineare non banale dei vettori in (11.3), applicando l'operatore f^h per un opportuno $h \geq 0$ otteniamo una combinazione lineare non banale dei vettori $v_1, \ldots, v_{\alpha_1}$ che è quindi non nulla, essendo tali vettori linearmente indipendenti. □

Più avanti daremo una diversa dimostrazione del Teorema 11.2.2 come conseguenza del teorema di esistenza delle basi di persistenza.

Esercizi.

616. In un palazzo ci sono 10 appartamenti. In ognuno c'è almeno una persona. In otto di questi appartamenti ci sono almeno 2 persone, in sei appartamenti ci sono almeno 4 persone ed in un appartamento ci sono almeno 5 persone. Quante persone, come minimo, ospita il palazzo?

617. Sia $A \in M_{n,n}(\mathbb{C})$ una matrice nilpotente e sia $t \in \mathbb{C}$ un numero diverso da 0. Provare che tA è simile ad A.

618. Mostrare che la matrice

$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \in M_{4,4}(\mathbb{K})$$

è nilpotente qualunque sia il campo $\mathbb{K}$ e determinare il suo diagramma di Young nei campi $\mathbb{K} = \mathbb{R}$ e $\mathbb{K} = \mathbb{F}_2 = \{0, 1\}$.

619. Sia V lo spazio vettoriale dei polinomi a coefficienti reali di grado $< n$ nella variabile x e sia $D\colon V \to V$ l'operatore di derivazione. Dire se i seguenti tre operatori sono nilpotenti e, nel caso, calcolarne le partizioni di n associate:

$$D,\ D - xD^2,\ D - xD^2 + \frac{x^2}{2}D^3 \colon V \to V.$$

620. Sia A una matrice di $M_{n,n}(\mathbb{C})$ con un unico autovalore λ. Dimostrare che $A - \lambda I$ è nilpotente.

621. Definire in maniera ricorsiva una successione $a_1, a_2, \dots$ di numeri razionali tali che se $A \in M_{n,n}(\mathbb{C})$ è una matrice nilpotente con indice di nilpotenza $s + 1 > 0$, allora

$$(I + a_1 A + a_2 A^2 + \cdots + a_s A^s)^2 = I + A.$$

622 (☕, ♡). Siano V uno spazio vettoriale di dimensione finita ed $f\colon V \to V$ un endomorfismo nilpotente con indice di nilpotenza $s \geq 2$, ossia $f^s = 0$ e $f^{s-1} \neq 0$. Sia poi $v \in V$ un vettore tale che $f^{s-1}(v) \neq 0$. Dimostrare che:

(1) $v, f(v), \dots, f^{s-1}(v)$ sono linearmente indipendenti;
(2) $f^h(V) \cap \operatorname{Span}(v, f(v), \dots, f^{h-1}(v)) = 0$ per ogni intero positivo $h = 1, \dots, s-1$;
(3) $\operatorname{Ker}(f^h) \cap \operatorname{Span}(v, f(v), \dots, f^{s-h-1}(v)) = 0$ per ogni intero positivo $h = 1, \dots, s-1$;
(4) esiste un sottospazio vettoriale f-invariante $U \subseteq V$ tale che

$$V = U \oplus \operatorname{Span}(v, f(v), \dots, f^{s-1}(v)).$$

623 (☕). Sia $f\colon V \to V$ un endomorfismo nilpotente di uno spazio vettoriale di dimensione finita con partizione associata

$$\alpha_h = \dim \operatorname{Ker}(f^h) - \dim \operatorname{Ker}(f^{h-1}) = \dim f^{h-1}(V) - \dim f^h(V) \geq 0, \qquad h > 0.$$

Dato un intero $p > 1$, dimostrare che esiste un endomorfismo $g\colon V \to V$ tale che $f = g^p$ se e solo se per ogni $h > 0$ vale

$$\left\lfloor \frac{\alpha_h}{p} \right\rfloor + \left\lfloor -\frac{\alpha_{h+1}}{p} \right\rfloor \geq 0.$$

(Ricordiamo che il simbolo $\lfloor x \rfloor \in \mathbb{Z}$ denota la parte intera del numero reale x.)

11.3. Autospazi generalizzati

Dato un endomorfismo $f\colon V \to V$ ed un suo autovalore λ, abbiamo introdotto a suo tempo l'autospazio corrispondente $V_\lambda = \operatorname{Ker}(f - \lambda I)$, di dimensione uguale alla molteplicità geometrica di λ. Gli autospazi generalizzati si introducono in maniera analoga, con la componente zero-Fitting al posto del nucleo.

DEFINIZIONE 11.3.1. Sia $\lambda \in \mathbb{K}$ un autovalore per un endomorfismo $f\colon V \to V$. L'**autospazio generalizzato** di λ (relativo ad f) è il sottospazio vettoriale

$$E_\lambda = F_0(f - \lambda I) = \operatorname{Ker}((f - \lambda I)^k), \qquad k >> 0.$$

Si noti che l'autospazio generalizzato E_λ contiene sempre l'autospazio (usuale) V_λ. La prossima proposizione ci fornisce l'interpretazione geometrica della molteplicità algebrica degli autovettori.

PROPOSIZIONE 11.3.2. *Sia $f\colon V \to V$ un endomorfismo di uno spazio vettoriale di dimensione finita. Allora la molteplicità algebrica di un autovalore λ è uguale alla dimensione dell'autospazio generalizzato $E_\lambda = F_0(f - \lambda I)$.*

DIMOSTRAZIONE. Consideriamo prima il caso in cui $\lambda = 0$. Sia r la dimensione della componente zero-Fitting di f; per il Corollario 11.1.5, in una opportuna base di V l'applicazione f è rappresentata da una matrice a blocchi

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

con A matrice $r \times r$ triangolare strettamente superiore e B matrice $(n-r)\times(n-r)$ invertibile. Dunque $p_f(t) = p_A(t)p_B(t) = (-t)^r p_B(t)$. Siccome B è invertibile si ha $p_B(0) = \det(B) \neq 0$ e quindi r coincide con la molteplicità della radice 0 in $p_f(t)$.

Se $\lambda \neq 0$ basta sostituire f con l'endomorfismo $g = f - \lambda I$; infatti, $p_f(t) = \det(f - tI) = \det(g - (t - \lambda)I) = p_g(t - \lambda)$ e quindi la molteplicità algebrica di λ come autovalore di f è uguale alla molteplicità algebrica di 0 come autovalore di g. □

TEOREMA 11.3.3. *Siano $f\colon V \to V$ un endomorfismo e λ un suo autovalore. Allora l'autospazio generalizzato E_λ è un sottospazio f-invariante.*

Dato un polinomio $p(t) \in \mathbb{K}[t]$, l'applicazione lineare $p(f)\colon E_\lambda \to E_\lambda$ è: un isomorfismo se $p(\lambda) \neq 0$, nilpotente se $p(\lambda) = 0$.

DIMOSTRAZIONE. Sia $k >> 0$ tale che $E_\lambda = \operatorname{Ker}((f - \lambda I)^k)$; ponendo $g = (f - \lambda I)^k$ si ha $fg = gf$ e la f-invarianza di $E_\lambda = \operatorname{Ker}(g)$ segue immediatamente dal Lemma 9.5.9.

Per il teorema di Ruffini $t-\lambda$ divide $p(t)-p(\lambda)$, dunque $(t-\lambda)^k$ divide $(p(t)-p(\lambda))^k$, quindi $(f - \lambda I)^k$ divide $(p(f) - p(\lambda)I)^k$ e ne consegue che $(p(f) - p(\lambda)I)^k$ si annulla identicamente su E_λ.

Questo prova immediatamente che se $p(\lambda) = 0$ allora $p(f)\colon E_\lambda \to E_\lambda$ è nilpotente. Se invece $p(\lambda) \neq 0$, guardando allo sviluppo di Newton delle potenze del binomio, troviamo un polinomio $q(t)$ tale che

$$(p(t) - p(\lambda))^k = p(t)q(t) + (-1)^k p(\lambda)^k.$$

Ne segue che la restrizione di $p(f)q(f)$ a E_λ coincide con l'isomorfismo $p(\lambda)^k I$ e dunque entrambe le restrizioni a E_λ di $p(f)$ e $q(f)$ sono isomorfismi, cf. Osservazione 10.4.8. □

LEMMA 11.3.4. *Siano $\lambda_1, \dots, \lambda_s$ autovalori distinti di un endomorfismo $f\colon V \to V$ e si considerino i relativi autospazi generalizzati $E_{\lambda_i} = F_0(f - \lambda_i I)$. Allora esiste una decomposizione in somma diretta:*

$$E_{\lambda_1} + \cdots + E_{\lambda_s} = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_s} \subseteq V.$$

DIMOSTRAZIONE. Bisogna dimostrare che ogni vettore $w \in E_{\lambda_1} + \cdots + E_{\lambda_s}$ si scrive in modo unico come somma di vettori in ciascun E_{λ_i}. Basta quindi verificare che se

$$v_1 + \cdots + v_s = 0\,, \qquad v_i \in E_{\lambda_i}\,, \quad i = 1, \dots, s,$$

allora $v_i = 0$ per ogni $i = 1, \dots, s$. Per induzione su s possiamo assumere il risultato vero se $v_s = 0$. Se $v_s \neq 0$ consideriamo l'endomorfismo $g = f - \lambda_s I$, allora esiste $k > 0$ tale che

$g^k(v_s) = 0$, mentre per il Teorema 11.3.3 ogni E_{λ_i} è g-invariante e l'endomorfismo $g\colon E_{\lambda_i} \to E_{\lambda_i}$ è invertibile per ogni $i < s$. Si ha dunque

$$0 = g^k(0) = g^k(v_1) + \cdots + g^k(v_{s-1})$$

e per l'ipotesi induttiva $g^k(v_1) = \cdots = g^k(v_{s-1}) = 0$ da cui segue $v_1 = \cdots = v_{s-1} = 0$. □

Abbiamo già dimostrato che ogni autovalore è anche radice del polinomio minimo; il prossimo lemma ci dice qual è la sua molteplicità.

LEMMA 11.3.5. *Sia $\lambda \in \mathbb{K}$ un autovalore dell'endomorfismo $f\colon V \to V$. Allora la molteplicità di λ come radice del polinomio minimo $q_f(t)$ è uguale al più piccolo intero positivo τ tale che*

$$E_\lambda = \operatorname{Ker}(f - \lambda I)^\tau.$$

DIMOSTRAZIONE. A meno di sostituire f con $f - \lambda I$ non è restrittivo supporre $\lambda = 0$ e quindi $E_\lambda = E_0 = F_0(f)$. Scriviamo il polinomio minimo di f nella forma $q_f(t) = h(t)t^\sigma$, con $h(0) \neq 0$ e dimostriamo che valgono le disuguaglianze $\sigma \geq \tau$ e $\sigma \leq \tau$.

Nella decomposizione di Fitting 11.1.5 $V = F_0 \oplus F_1$, abbiamo le due restrizioni $f_0 = f_{|F_0}\colon F_0 \to F_0$ che è nilpotente con indice di nilpotenza τ, e $f_1 = f_{|F_1}\colon F_1 \to F_1$ che è invertibile.

Dunque t^τ è il polinomio minimo di f_0 che deve dividere il polinomio minimo di f e quindi $\sigma \geq \tau$. Se $q(t)$ è il polinomio caratteristico di f_1 si ha $q(0) = \det(f_1) \neq 0$, mentre per Cayley–Hamilton l'endomorfismo $f^\tau q(f)$ annulla sia i vettori di F_0 che i vettori di F_1. Quindi $q_f(t)$ divide $t^\tau q(t)$ e questo implica $\tau \geq \sigma$. □

TEOREMA 11.3.6. *Per un autovalore λ di un endomorfismo $f\colon V \to V$ le seguenti condizioni sono equivalenti:*

(1) *La molteplicità geometrica di λ è uguale alla molteplicità algebrica.*
(2) $\operatorname{Ker}(f - \lambda I) = E_\lambda = F_0(f - \lambda I)$.
(3) $\operatorname{Ker}(f - \lambda I) = \operatorname{Ker}(f - \lambda I)^2$.
(4) *λ è una radice semplice del polinomio minimo di f.*

DIMOSTRAZIONE. L'equivalenza [1 ⇔ 2] segue dalle definizioni e dalla Proposizione 11.3.2, mentre l'equivalenza [2 ⇔ 4] è una diretta conseguenza del Lemma 11.3.5.

L'implicazione [2 ⇒ 3] segue dal fatto che

$$\operatorname{Ker}(f - \lambda I) \subseteq \operatorname{Ker}(f - \lambda I)^2 \subseteq F_0(f - \lambda I).$$

Viceversa, sappiamo dalla teoria della filtrazione dei nuclei che se $\operatorname{Ker}(f - \lambda I) = \operatorname{Ker}(f - \lambda I)^2$ allora $\operatorname{Ker}(f - \lambda I)^k = \operatorname{Ker}(f - \lambda I)^{k+1}$ per ogni $k > 0$ e questo prova l'implicazione [3 ⇒ 2]. □

COROLLARIO 11.3.7. *Sia*

$$p_f(t) = (\lambda_1 - t)^{\nu_1}(\lambda_2 - t)^{\nu_2} \cdots (\lambda_s - t)^{\nu_s}, \qquad \lambda_1, \ldots, \lambda_s \in \mathbb{K},\ \lambda_i \neq \lambda_j,$$

il polinomio caratteristico di un endomorfismo $f\colon V \to V$. Allora f è diagonalizzabile se e solo se per ogni i vale $\nu_i = \dim \operatorname{Ker}(f - \lambda_i I)$, ossia se e solo se per ogni autovalore la molteplicità geometrica è uguale alla molteplicità algebrica.

DIMOSTRAZIONE. Valgono le diseguaglianze $\nu_i \geq \dim \operatorname{Ker}(f - \lambda_i I)$ e per ipotesi si ha $\nu_1 + \cdots + \nu_s = \deg p_f(t) = \dim V$. Di conseguenza, la somma delle molteplicità geometriche degli autovalori è uguale ad n se e solo se $\nu_i = \dim \operatorname{Ker}(f - \lambda_i I)$ per ogni i. La conclusione segue dal Teorema 9.7.4. □

COROLLARIO 11.3.8. *Sia $f\colon V \to V$ un endomorfismo con polinomio minimo*

$$q_f(t) = (t - \lambda_1)^{\sigma_1}(t - \lambda_2)^{\sigma_2} \cdots (t - \lambda_s)^{\sigma_s}, \qquad \sigma_i > 0, \quad \lambda_i \in \mathbb{K}, \quad \lambda_i \neq \lambda_j \text{ per } i \neq j.$$

Allora f è diagonalizzabile se e solo se $\sigma_i = 1$ per ogni indice i.

DIMOSTRAZIONE. Per il Teorema 11.3.6 vale $\sigma_i = 1$ se e solo se le molteplicità algebrica e geometrica di λ_i sono uguali. □

Esercizi.

624. Calcolare i polinomi minimo e caratteristico delle matrici

$$\begin{pmatrix} 2 & -2 & 1 \\ 1 & 1 & -1 \\ 1 & 2 & -2 \end{pmatrix}, \qquad \begin{pmatrix} 2 & 1 & -4 \\ 0 & 3 & 0 \\ 1 & 2 & -2 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 1 & 1 \\ 2 & 1 & 2 \\ 1 & 1 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 2 & 2 \\ 1 & 1 & 1 \\ 0 & -2 & -2 \end{pmatrix},$$

e dire, motivando la risposta, quali sono diagonalizzabili.

625. Data la matrice

$$A = \begin{pmatrix} 2 & 4 & -5 \\ 1 & 5 & -5 \\ 1 & 4 & -4 \end{pmatrix}$$

Calcolare:

(1) il polinomio caratteristico e gli autovalori;
(2) le molteplicità algebrica e geometrica di ciascun autovalore;
(3) il polinomio minimo.

626. Siano V uno spazio vettoriale reale di dimensione finita e $f\colon V \to V$ un endomorfismo che soddisfa una delle seguenti condizioni:

$$f^2 = I, \qquad f^3 = f, \qquad f^3 = 2f^2 + f - 2I.$$

Dimostrare che f è diagonalizzabile.

627. Sia $A \in M_{n,n}(\mathbb{C})$ una matrice tale che $A^3 = -A$. Provare che A è diagonalizzabile.

628. Sia $A \in M_{n,n}(\mathbb{C})$ invertibile e tale che A^k è diagonalizzabile per qualche $k > 0$. Provare che anche A è diagonalizzabile.

629. Sia I_n la matrice identità $n \times n$. Determinare i polinomi minimo e caratteristico della matrice

$$\begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix} \in M_{2n,2n}(\mathbb{K}).$$

630 (♡). Sia $A \in M_{n,n}(\mathbb{R})$ una matrice tale che $A^2 + I = 0$. Dimostrare che la traccia di A è uguale a 0, che il determinante di A è uguale a 1 e che n è pari.

631. Per quali valori di $a \in \mathbb{R}$ il polinomo minimo della matrice

$$\begin{pmatrix} 1 & a & 0 \\ a & a & 1 \\ a & a & -1 \end{pmatrix}$$

ha grado 3?

632. Di una matrice $A \in M_{3,3}(\mathbb{R})$ sappiamo che:

(1) La prima riga è $(1, -1, 1)$.
(2) A è diagonalizzabile.
(3) La traccia di A è uguale a 2.
(4) I due vettori $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ e $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ sono autovettori di A.

Determinare tutti i coefficienti di A.

633. Siano V uno spazio vettoriale di dimensione finita e $W = \mathrm{Hom}(V, V)$ lo spazio vettoriale degli endomorfismi di V. Dato un elemento $A \in W$ denotiamo

$$R_A\colon W \to W, \qquad R_A(B) = AB + BA.$$

(1) Se il polinomio caratteristico di A è $t(t-1)(t-2)$ provare che A e R_A sono diagonalizzabili e si calcoli il polinomio caratteristico di R_A.
(2) Provare che se A è diagonalizzabile allora R_A è diagonalizzabile e che se A è nilpotente allora R_A è nilpotente.
(3) Se il polinomio minimo di A è $t(t-1)(t-2)$ si calcolino gli autovalori di R_A.

(4) Se A è diagonalizzabile e $\dim V = n$, provare che il polinomio minimo di R_A ha grado minore od uguale a $\dfrac{n(n+1)}{2}$. Trovare un esempio in cui $n = 3$ ed il polinomio minimo di R_A ha grado 6.

634 (☕). Siano $A, B, M \in M_{n,n}(\mathbb{C})$ tali che $AM = MB$. Mostrare che per ogni $\lambda \in \mathbb{C}$ ed ogni $m > 0$ si ha $(A-\lambda I)^m M = M(B-\lambda I)^m$. Dedurre che i polinomi caratteristici di A e B hanno in comune un numero di radici, contate con molteplicità, maggiore o uguale al rango di M.

635 (☕). Siano $A, B \in M_{2,2}(\mathbb{C})$ matrici tali che

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

Dimostrare che A e B non sono diagonalizzabili e dedurre che $\det(A) = 1$, $\det(B) = -1$.

636. Sia p un numero primo positivo. Ricordando l'Esercizio 182 ed il teorema di Ruffini, dimostrare che una matrice quadrata A a coefficienti nel campo $\mathbb{F}_p$ è diagonalizzabile se e solo se $A^p = A$.

637. Sia $A \in M_{n,n}(\mathbb{R})$ triangolabile con tutti gli autovalori ≥ 0. Provare che se A^2 è diagonale, allora anche A^3 è diagonale.

638 (☕). Sia $\mathbb{K} \subseteq \mathbb{C}$ un campo di numeri. Dimostrare nell'ordine:

(1) Se $A, H \in M_{n,n}(\mathbb{K})$ sono matrici simili, allora
$$\dim_{\mathbb{K}}\{B \in M_{n,n}(\mathbb{K}) \mid AB = BA\} = \dim_{\mathbb{K}}\{B \in M_{n,n}(\mathbb{K}) \mid HB = BH\}.$$
(2) Per ogni matrice $A \in M_{n,n}(\mathbb{K})$ vale
$$\dim_{\mathbb{K}}\{B \in M_{n,n}(\mathbb{K}) \mid AB = BA\} = \dim_{\mathbb{C}}\{B \in M_{n,n}(\mathbb{C}) \mid AB = BA\}.$$
(3) Siano $T \subseteq M_{n,n}(\mathbb{K})$ lo spazio delle matrici triangolari e $N \subseteq T$ il sottospazio delle matrici triangolari con diagonale nulla. Allora per ogni $A, B \in T$ si ha $AB - BA \in N$.
(4) Usare i punti precedenti per ridimostrare che per ogni $A \in M_{n,n}(\mathbb{K})$ vale la disuguaglianza
$$\dim_{\mathbb{K}}\{B \in M_{n,n}(\mathbb{K}) \mid AB = BA\} \geq n.$$

639 (☕). Siano $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$ le radici, contate con molteplicità, del polinomio caratteristico di una matrice $n \times n$ a coefficienti interi; si assuma che $|\alpha_i| \leq 1$ per ogni i. Dimostrare che ogni α_i è uguale a 0 oppure è una radice dell'unità.

11.4. La forma canonica di Jordan

Per forma canonica di Jordan si intende l'estensione del Teorema 11.2.2 ad una generica matrice triangolabile. Ricordiamo che una matrice si dice triangolabile se è simile ad una matrice triangolare o, equivalentemente, se il polinomio caratteristico si scrive come un prodotto di fattori di primo grado.

Per ogni $\lambda \in \mathbb{K}$ ed ogni intero positivo n definiamo il **blocco di Jordan** di ordine n ed autovalore λ come la matrice

$$J_n(\lambda) = \lambda I + J_n(0) = \begin{pmatrix} \lambda & 1 & \cdots & 0 & 0 \\ 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & \cdots & 0 & \lambda \end{pmatrix} \in M_{n,n}(\mathbb{K}).$$

Per **matrice di Jordan** si intende una matrice diagonale a blocchi di Jordan:

$$\begin{pmatrix} J_{k_1}(\lambda_1) & 0 & \cdots & 0 \\ 0 & J_{k_2}(\lambda_2) & \vdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{k_s}(\lambda_s) \end{pmatrix}, \qquad k_i > 0, \quad \lambda_i \in \mathbb{K}.$$

Si noti che le matrici di Jordan sono tutte triangolari (inferiori).

OSSERVAZIONE 11.4.1. In letteratura vengono talvolta definiti i blocchi di Jordan come i trasposti dei blocchi precedentemente definiti, ossia come le matrici:

$$\begin{pmatrix} \lambda & 0 & \cdots & 0 & 0 \\ 1 & \lambda & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \lambda \end{pmatrix} = \begin{pmatrix} \lambda & 1 & \cdots & 0 & 0 \\ 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & \cdots & 0 & \lambda \end{pmatrix}^T .$$

È quindi importante ricordare che le matrici $J_n(\lambda)^T$ e $J_n(\lambda)$ sono simili ed ottenute l'una dall'altra mediante il cambio di base $(e_1, \dots, e_n) \mapsto (e_n, \dots, e_1)$, vedi Esempio 9.1.4.

TEOREMA 11.4.2 (Forma canonica di Jordan). *Ogni matrice triangolabile è simile ad una matrice di Jordan. Due matrici di Jordan sono simili se e solo se si ottengono l'una dall'altra mediante una permutazione dei blocchi di Jordan.*

DIMOSTRAZIONE. *(Esistenza.)* Bisogna dimostrare che se $f\colon V \to V$ è un endomorfismo triangolabile, allora esiste una base nella quale f si rappresenta con una matrice di Jordan. Siano $\lambda_1, \dots, \lambda_s$ gli autovalori di f contati senza molteplicità, ossia $\lambda_i \neq \lambda_j$ per ogni $i \neq j$. Siccome f è triangolabile la somma delle molteplicità algebriche è uguale alla dimensione di V e quindi per il Lemma 11.3.4 si ha una decomposizione in somma diretta di sottospazi f-invarianti.

$$V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_s}.$$

Considerando le restrizioni di f agli autospazi generalizzati ci possiamo ricondurre al caso in cui esiste uno scalare λ tale che $f - \lambda I$ è nipotente. Ma allora basta applicare il Teorema 11.2.2 per trovare una base in cui $f - \lambda I$ si rappresenta con una matrice di Jordan nilpotente.

(Unicità.) Lo stesso ragionamento del caso nilpotente mostra che, per ogni autovalore λ il numero di blocchi di Jordan $J_k(\lambda)$ è uguale al numero di colonne di lunghezza k nel diagramma di Young associato all'endomorfismo $f - \lambda I$. □

Le stesse considerazioni del caso nilpotente ci forniscono le seguenti regole pratiche per il calcolo della forma di Jordan:

- *la differenza* $\operatorname{rg}(f - \lambda I)^{a-1} - \operatorname{rg}(f - \lambda I)^a$, $a > 0$, *è uguale al numero di blocchi di Jordan* $J_k(\lambda)$ *con* $k \geq a$.
- *per ogni* $\lambda \in \mathbb{K}$ *e* $a \in \mathbb{N}$, *il numero di blocchi di Jordan* $J_a(\lambda)$ *è uguale a*

$$\operatorname{rg}(f - \lambda I)^{a-1} - 2\operatorname{rg}(f - \lambda I)^a + \operatorname{rg}(f - \lambda I)^{a+1}.$$

ESEMPIO 11.4.3. Calcoliamo la forma canonica di Jordan della matrice

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

La matrice è triangolare unipotente, il suo polinomio caratteristico è uguale a $(1-t)^3$ ed ogni blocco di Jordan è del tipo $J_k(1)$, $k > 0$. Poiché $A - I$ ha rango 2, esiste un solo blocco di Jordan e quindi la matrice di Jordan associata è

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

ESEMPIO 11.4.4. Calcoliamo la forma canonica di Jordan delle matrici

$$A = \begin{pmatrix} 4 & 0 & 0 \\ 1 & 8 & 0 \\ 0 & 0 & 4 \end{pmatrix}, \qquad B = \begin{pmatrix} 7 & -1 & -1 \\ -6 & 6 & 2 \\ 11 & 3 & 3 \end{pmatrix}.$$

Il polinomio caratteristico è uguale per entrambe a:

$$p_A(t) = p_B(t) = -t^3 + 16t^2 - 80t + 128 = -(t-4)^2(t-8).$$

Deduciamo che gli autovalori sono $\lambda_1 = 4$, con molteplicità algebrica 2 e $\lambda = 8$ con molteplicità algebrica 1. Sappiamo quindi che i sottospazi $\operatorname{Ker}(A-4I)^2$ e $\operatorname{Ker}(B-4I)^2$ hanno dimensione 2,

e che i sottospazi $\mathrm{Ker}(A-8I)$, $\mathrm{Ker}(B-8I)$ hanno dimensione 1. Inoltre, si verifica facilmente che:

(1) $\mathrm{Ker}(A-4I)$ ha dimensione 2. Ne consegue che l'autovalore $\lambda=4$ ha molteplicità geometrica 2, la matrice A è diagonalizzabile e la sua forma canonica di Jordan è
$$\begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 8 \end{pmatrix}.$$

(2) $\mathrm{Ker}(B-4I)$ ha dimensione 1. Dunque l'autovalore $\lambda=4$ ha molteplicità geometrica 1, la matrice B non è diagonalizzabile e siccome $\mathrm{rg}(B-4I)-\mathrm{rg}(B-4I)^2=1$ esiste un blocco di Jordan $J_k(4)$, con $k\geq 2$. Per motivi dimensionali si deduce che la forma canonica di Jordan di B è
$$\begin{pmatrix} 4 & 1 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 8 \end{pmatrix}.$$

ESEMPIO 11.4.5. Calcoliamo la forma canonica di Jordan della matrice
$$A=\begin{pmatrix} 2 & 0 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -3 & 2 & 1 \\ 0 & 3 & -2 & -1 \end{pmatrix}.$$
Per prima cosa calcoliamo il polinomio caratteristico; lasciamo al lettore il poco piacevole compito di verificare che $p_A(t)=(t-1)^4$. Dunque $\lambda=1$ è l'unico autovalore di A e pertanto la forma canonica di Jordan è del tipo
$$\begin{pmatrix} J_{k_1}(1) & 0 & \cdots & 0 \\ 0 & J_{k_2}(1) & \vdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{k_s}(1) \end{pmatrix}, \qquad k_i>0, \quad k_1+\cdots+k_s=4.$$
Il secondo passo consiste nel calcolo del numero s dei blocchi di Jordan, che sappiamo essere uguale alla dimensione di $\mathrm{Ker}(A-I)$; lasciamo di nuovo al lettore il compito di verificare che la matrice $A-I$ ha rango 2, quindi $\dim\mathrm{Ker}(A-I)=2$ e la forma di Jordan è una delle seguenti:
$$\begin{pmatrix} J_1(1) & 0 \\ 0 & J_3(1) \end{pmatrix}, \qquad \begin{pmatrix} J_2(1) & 0 \\ 0 & J_2(1) \end{pmatrix}.$$
Il terzo, e nella fattispecie, ultimo passo consiste nel calcolare il rango di $(A-I)^2$, che risulta uguale a 1 ed implica che la forma di Jordan della matrice A è pertanto
$$\begin{pmatrix} J_1(1) & 0 \\ 0 & J_3(1) \end{pmatrix}=\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Esercizi.

640 (♡). Calcolare la forma canonica di Jordan delle matrici
$$\begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 5 & 4 & 2 & 1 \\ 0 & 1 & -1 & -1 \\ -1 & -1 & 3 & 0 \\ 1 & 1 & -1 & 2 \end{pmatrix}, \qquad \begin{pmatrix} 4 & 1 & 0 & -1 \\ 0 & 3 & 0 & 1 \\ 0 & 0 & 4 & 0 \\ 1 & 0 & 0 & 5 \end{pmatrix}.$$

641. Calcolare la forma canonica di Jordan delle matrici
$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} -2 & 0 & 2 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 2 & 1 & 1 & 1 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & -1 & -1 & -1 & 0 \end{pmatrix}.$$

642. Calcolare la forma canonica di Jordan delle seguenti matrici definite sul campo dei numeri complessi:

$$\begin{pmatrix} 39 & -64 \\ 25 & -41 \end{pmatrix}, \quad \begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 3 & -2 \\ 0 & 7 & -4 \\ 0 & 9 & -5 \end{pmatrix}, \quad \begin{pmatrix} 3 & 0 & 1 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix},$$

$$\begin{pmatrix} 22 & -2 & -12 \\ 20 & 0 & -12 \\ 30 & -3 & -16 \end{pmatrix}, \quad \begin{pmatrix} -13 & 8 & 1 & 2 \\ -22 & 13 & 0 & 3 \\ 8 & -5 & 0 & -1 \\ -22 & 13 & 5 & 5 \end{pmatrix}, \quad \begin{pmatrix} 2 & 1 & 1 & 1 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & -1 & -1 & -1 & 0 \end{pmatrix},$$

$$\begin{pmatrix} 1 & -1 & -2 \\ 0 & 3 & 0 \\ 0 & 1 & 3 \end{pmatrix}, \quad \begin{pmatrix} -6 & -5 & -8 \\ -2 & -2 & -3 \\ 6 & 5 & 8 \end{pmatrix}, \quad \begin{pmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 1 & 3 & -2 \end{pmatrix},$$

$$\begin{pmatrix} 2 & -1 \\ 2 & 4 \end{pmatrix}, \quad \begin{pmatrix} 2 & -1 & 0 \\ 2 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 2 & -1 & 0 & 0 \\ 2 & 4 & 0 & 2 \\ 0 & 0 & 2 & -1 \\ 0 & 0 & 2 & 4 \end{pmatrix}, \quad \begin{pmatrix} 2 & -1 & 2 & 0 \\ 2 & 4 & 0 & 2 \\ 0 & 0 & 2 & -1 \\ 0 & 0 & 2 & 4 \end{pmatrix},$$

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

$$\begin{pmatrix} 2 & 5 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 4 & 2 & 0 \\ 0 & 0 & 3 & 5 & 0 \\ 0 & 0 & 0 & 0 & 7 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ -1 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

643. Determinare la forma canonica di Jordan della matrice complessa

$$\begin{pmatrix} -2 & -2 & -4 \\ 1 & b & 2 \\ 1 & 1 & 2 \end{pmatrix}$$

al variare del parametro b.

644. Sia A una matrice quadrata complessa tale che $A^3 = A^2$. Dimostrare che A^2 è sempre diagonalizzabile e mostrare con un esempio che A può non essere diagonalizzabile.

645. Sia $f\colon \mathbb{C}^{12} \to \mathbb{C}^{12}$ un endomorfismo i cui polinomi minimo e caratteristico sono

$$p_f(t) = t^3(t-1)^4(t-2)^5, \qquad q_f(t) = t^2(t-1)^3(t-2)^4.$$

Determinare la forma canonica di Jordan di f.

646. Sia A una matrice 6×6 con polinomio caratteristico $p_A(t) = t(t-1)^5$. Se il rango della matrice $A - I$ è 3, quali sono le possibili forme di Jordan di A?

647. Sia A una matrice con polinomio caratteristico $p_A(t) = (t-1)^2(t-2)^3$. Se l'autovalore 1 ha molteplicità geometrica 1 e l'autovalore 2 ha molteplicità geometrica 2, qual è la forma di Jordan di A?

648. Sia $A \in M_{n,n}(\mathbb{C})$ una matrice invertibile. Provare che esiste una matrice $B \in M_{n,n}(\mathbb{C})$ tale che $B^2 = A$.

649 (☕). Sia $A \in M_{n,n}(\mathbb{K})$ una matrice triangolabile. Dimostrare che se per ogni intero $a > 0$ ed ogni $\lambda \in \mathbb{K}$ il nucleo di $(A - \lambda I)^a$ ha dimensione pari, allora esiste una matrice $B \in M_{n,n}(\mathbb{K})$ tale che $B^2 = A$.

650 (☕). Sia $A \in M_{n,n}(\mathbb{R})$ una matrice triangolabile invertibile. Provare che esiste una matrice $B \in M_{n,n}(\mathbb{R})$ tale che $B^2 = A$ se e solo se per ogni intero positivo k ed ogni numero reale negativo $\lambda < 0$ il nucleo di $(A - \lambda I)^k$ ha dimensione pari.

11.5. Moduli di persistenza

I moduli di persistenza sono uno strumento dell'algebra lineare che trova applicazioni in alcune recenti tecniche di matematica applicata alla teoria del riconoscimento automatico delle immagini ed allo studio topologico di grandi basi di dati.

DEFINIZIONE 11.5.1. Un **modulo di persistenza** sul campo $\mathbb{K}$ è un diagramma in serie di spazi vettoriali di dimensione finita ed applicazioni lineari:

$$V_1 \xrightarrow{f_1} V_2 \xrightarrow{f_2} \cdots \longrightarrow V_i \xrightarrow{f_i} \cdots \longrightarrow V_j \xrightarrow{f_j} \cdots \xrightarrow{f_{k-1}} V_k\,, \quad k \geq 1\,.$$

$$V_i \xrightarrow{f_{i,j}} V_j$$

In riferimento al precedente diagramma, per ogni $i < j$ denoteremo $f_{i,j}$ la composizione delle applicazioni da V_i a V_j. Per convezione poniamo $f_{i,i}$ uguale all'identità su V_i. Per convenienza notazionale conviene estendere il diagramma ponendo $V_i = 0$ per ogni $i \leq 0$ (e di conseguenza $f_i = 0$ per ogni $i \leq 0$).

Sia $\mathcal{B}_i$ una base (non ordinata) dello spazio vettoriale V_i per ogni indice i. L'unione $\mathcal{B} = \cup_i \mathcal{B}_i$ viene detta una **base di persistenza** se soddisfa le seguenti due condizioni:

(1) per ogni $1 \leq i < k$ ed ogni $e \in \mathcal{B}_i$ si ha $f_i(e) = 0$ oppure $f_i(e) \in \mathcal{B}_{i+1}$.
(2) per ogni $1 \leq i < k$ l'applicazione ritretta

$$f_i \colon \{e \in \mathcal{B}_i \mid f_i(e) \neq 0\} \to V_{i+1}$$

è iniettiva.

Prima di dimostrare che le basi di persistenza esistono, mostriamo che sono omogenee rispetto ai nuclei ed alle immagini delle applicazioni $f_{i,j}$, ossia che vale la seguente proposizione.

PROPOSIZIONE 11.5.2. *Nelle notazioni precedenti, per ogni base di persistenza $\mathcal{B}$ ed ogni $i \leq j$ valgono le uguaglianze:*

$$\operatorname{Ker}(f_{i,j}) = \operatorname{Span}(\mathcal{B}_i \cap \operatorname{Ker}(f_{i,j})), \quad \operatorname{Im}(f_{i,j}) = \operatorname{Span}(\mathcal{B}_j \cap \operatorname{Im}(f_{i,j})). \tag{11.4}$$

DIMOSTRAZIONE. Siano $i \leq j$ fissati; in entrambi i casi le inclusioni $\supseteq$ sono ovvie. Sia n la dimensione di V_i e ordiniamo la base $\mathcal{B}_i$ in modo tale che per un intero $r \leq n$ si abbia

$$\mathcal{B}_i = \{e_1, \ldots, e_n\}, \quad f_{i,j}(e_h) = 0 \iff h \leq r.$$

In particolare,

$$\dim \operatorname{Ker}(f_{i,j}) \geq \dim \operatorname{Span}(\mathcal{B}_i \cap \operatorname{Ker}(f_{i,j})) = r.$$

D'altra parte, segue immediatamente dalla definizione di base di persistenza che l'applicazione

$$f_{i,j} \colon \{e \in \mathcal{B}_i \mid f_{i,j}(e) \neq 0\} \to \mathcal{B}_j$$

è iniettiva e quindi

$$\operatorname{rg}(f_{i,j}) \geq \dim \operatorname{Span}(\mathcal{B}_j \cap \operatorname{Im}(f_{i,j})) \geq n - r.$$

Basta adesso applicare il teorema del rango per concludere che $f_{i,j}$ ha rango $n - r$ e che valgono le due uguaglianze (11.4). □

LEMMA 11.5.3. *Siano $g \colon V \to W$ un'applicazione lineare, $v \in V$ tale che $g(v) \neq 0$ e $H \subseteq W$ un sottospazio tale che $W = \operatorname{Span}(g(v)) \oplus H$. Allora*

$$V = \operatorname{Span}(v) \oplus g^{-1}(H), \quad \textit{dove} \quad g^{-1}(H) = \{u \in V \mid g(u) \in H\}.$$

DIMOSTRAZIONE. Semplice e lasciata per esercizio. □

TEOREMA 11.5.4. *Ogni modulo si persistenza possiede una base di persistenza.*

DIMOSTRAZIONE. Nelle notazioni della Definizione 11.5.1, dimostriamo il teorema per induzione sulla somma delle dimensioni degli spazi V_i. Se $V_i = 0$ per ogni i non c'è nulla da dimostrare, altrimenti a meno di scorrere gli indici possiamo supporre $V_1 \neq 0$. Sia $1 \leq j \leq k$ in più grande indice tale che $f_{1,k} \neq 0$; scegliamo un vettore $e \in V_1$ tale che $f_{1,j}(e) \neq 0$ ed una decomposizione in somma diretta $V_j = \operatorname{Span}(f_{1,j}(e)) \oplus H_j$.

Per il Lemma 11.5.3 possiamo trovare, per ogni $i = 1, \ldots, j$, una decomposizione in somma diretta $V_i = \operatorname{Span}(f_{1,i}(e)) \oplus H_i$ tale che $f_i(H_i) \subseteq H_{i+1}$ per ogni $i < j$: ad esempio si

può prendere, in maniera ricorsiva, $H_{j-1} = f_{j-1}^{-1}(H_j)$, $H_{j-2} = f_{j-2}^{-1}(H_{j-1})$ e così via. Abbiamo quindi un nuovo modulo di persistenza

$$H_1 \xrightarrow{f_1} \cdots \longrightarrow H_j \xrightarrow{f_j} V_{j+1} \longrightarrow \cdots \xrightarrow{f_{k-1}} V_k \ ,$$

che per l'ipotesi induttiva possiede una base di persistenza $\mathcal{B}$. Allora l'unione di $\mathcal{B}$ con i vettori $e, f_1(e), f_{1,3}(e), \dots, f_{1,j}(e)$ è una base di persistenza del modulo di partenza. □

Data una base di persistenza $\mathcal{B} = \cup_i \mathcal{B}_i$, dati $i < j$ diremo che un elemento $e \in \mathcal{B}$ *nasce in* i *e muore in* j se:

(1) $e \in \mathcal{B}_i$ e $f_{i,j-1}(e) \in \mathcal{B}_{j-1}$;
(2) $f_{i,j}(e) = 0$;
(3) $e \notin f_{i-1}(\mathcal{B}_{i-1})$.

Abbiamo chiaramente usato la convezione che $f_i = 0$ se $i < 1$ oppure se $i \geq k$.

Per ogni $i < j$ denoteremo con $\beta_{i,j}$ il numero di elementi di $\mathcal{B}$ che nascono in i e muoiono in j. Il prossimo teorema mostra in particolare che tali numeri non dipendono dalla particolare base di persistenza scelta.

TEOREMA 11.5.5. *Nelle notazioni precedenti, sia* $r_{i,j} = \operatorname{rg}(f_{i,j})$. *Allora, per ogni* $i < j$ *si ha*

$$\beta_{i,j} = r_{i,j-1} + r_{i-1,j} - r_{i,j} - r_{i-1,j-1}.$$

DIMOSTRAZIONE. Se denotiamo con $\gamma_{i,j}$ il numero di elementi che nascono in i e sono ancora vivi in $j-1$, ossia coloro che soddisfano le precedenti condizioni (1) e (3), allora vale $\beta_{i,j} = \gamma_{i,j} - \gamma_{i,j+1}$. Dalla Proposizione 11.5.2 segue che $r_{i,j-1}$ è esattamente uguale al numero di elementi di $\mathcal{B}_i$ che sono ancora vivi in $j-1$. Da ciò segue che $\gamma_{i,j} = r_{i,j-1} - r_{i-1,j-1}$. □

L'insieme dei numeri $\beta_{i,j}$ viene detto *diagramma di persistenza*.[3]

Esercizi.

651. Sia $f\colon V \to V$ nilpotente con indice di nilpotenza k e si consideri il modulo di persistenza

$$V \xrightarrow{f} f(V) \xrightarrow{f} f^2(V) \xrightarrow{f} \cdots \xrightarrow{f} f^{k-1}(V) \ .$$

Calcolare il diagramma di persistenza in funzione del diagramma di Young di f.

652. Sia $f\colon V \to V$ idempotente, ossia $f^2 = f$, di rango r. Calcolare il diagramma di persistenza di

$$V \xrightarrow{f} V \xrightarrow{f} V \xrightarrow{f} V \xrightarrow{f} V \ .$$

11.6. Complementi: la decomposizione di Dunford

Talvolta, al posto della forma canonica di Jordan può essere utile usare un risultato conosciuto come decomposizione di Dunford. Per semplicità espositiva assumeremo che gli spazi vettoriali siano definiti su di un campo infinito $\mathbb{K}$: il risultato è vero anche sui campi finiti, ma in tal caso è richiesta una diversa dimostrazione.

LEMMA 11.6.1. *Siano* $f, g\colon V \to V$ *due endomorfismi tali che* $fg = gf$:

(1) *se* f *è invertibile, allora* $f^{-1}g = gf^{-1}$;
(2) *se* g *è nilpotente, allora* fg *è nilpotente;*
(3) *se* f *e* g *sono nilpotenti, allora* $f + g$ *è nilpotente.*

[3]Per amor di precisione, in letteratura il diagramma di persistenza viene definito come l'applicazione $D = \{(i,j) \in \mathbb{Z}^2 \mid 0 < i < j\} \to \mathbb{Z}$, $(i,j) \mapsto \beta_{i,j}$.

DIMOSTRAZIONE. 1) Si ha $gf^{-1} = f^{-1}fgf^{-1} = f^{-1}gff^{-1} = f^{-1}g$.

2) Basta osservare che $(fg)^k = f^k g^k$. Si noti che se f è invertibile, per il punto precedente $f^{-1}g = gf^{-1}$ e quindi anche $f^{-1}g$ è nilpotente.

3) Siano $h, k > 0$ tali che $f^h = g^k = 0$ e proviamo che $(f+g)^{h+k-1} = 0$. Siccome f e g commutano tra loro, vale lo sviluppo di Newton del binomio

$$(f+g)^n = \sum_{i=0}^{n} \binom{n}{i} f^i g^{n-i}.$$

Se $n \geq h+k-1$ possiamo scrivere

$$(f+g)^n = \sum_{i=0}^{h-1} \binom{n}{i} f^i g^{n-i} + \sum_{i=h}^{n} \binom{n}{i} g^{n-i} f^i.$$

e ogni addendo di $(f+g)^n$ nella precedente espressione è nullo. □

ESEMPIO 11.6.2. Si consideri lo spazio vettoriale V delle combinazioni lineari a coefficienti reali dei monomi $1, x, y, xy$ e si considerino i due operatori $\partial_x, \partial_y \colon V \to V$ definiti dalle relazioni

$$\partial_x(1) = \partial_x(y) = 0, \quad \partial_x(x) = 1, \quad \partial_x(xy) = y,$$
$$\partial_y(1) = \partial_y(x) = 0, \quad \partial_y(y) = 1, \quad \partial_y(xy) = x.$$

Allora $\partial_x^2 = \partial_y^2 = 0$, $\partial_x \partial_y = \partial_y \partial_x$ e per il lemma precedente $(\partial_x + \partial_y)^3 = 0$. Merita osservare che $(\partial_x + \partial_y)^2 \neq 0$.

LEMMA 11.6.3. *Siano V uno spazio vettoriale di dimensione finita e $f, g \colon V \to V$ due endomorfismi tali che $fg = gf$. Se g è nilpotente, allora f e $f+g$ hanno lo stesso determinante, lo stesso polinomio caratteristico e la stessa decomposizione di Fitting. In particolare f è triangolabile se e solo se $f+g$ è triangolabile.*

DIMOSTRAZIONE. Sia $s+1 > 0$ l'indice di nilpotenza di g. Dato un vettore $v \in F_0(f)$ scegliamo un intero $k > 0$ tale che $f^k(v) = 0$; per lo sviluppo del binomio di Newton segue che $(f+g)^{k+s}(v) = 0$ e quindi $v \in F_0(f+g)$. Abbiamo quindi provato che $F_0(f) \subseteq F_0(f+g)$. Siccome $f = (f+g) - g$ e l'endomorfismo nilpotente $-g$ commuta con $f+g$, lo stesso argomento prova che $F_0(f+g) \subseteq F_0(f)$. segue dalla formula

$$(f+g)^n = \sum_{i=0}^{n} \binom{n}{i} f^i g^{n-i}$$

che

$$(f+g)^n(V) \subseteq f^n(V) + f^{n-1}(g(V)) + \cdots + f^{n-s}(g^s(V)) \subseteq f^{n-s}(V)$$

e quindi, per n sufficientemente grande,

$$F_1(f+g) = (f+g)^n(V) \subseteq f^{n-s}(V) = F_1(f)\,.$$

Per simmetria vale anche l'inclusione opposta $F_1(f) \subseteq F_1(f+g)$.

Proviamo adesso che f e $f+g$ hanno lo stesso determinante. Se $\det(f) = 0$ allora $F_0(f) \neq 0$, dunque $F_0(f+g) \neq 0$ e $f+g$ non è invertibile. Se invece f è invertibile, allora $f+g = f(I + f^{-1}g)$. Adesso $f^{-1}g$ è nilpotente ed in una opportuna base si rappresenta con una matrice triangolare strettamente superiore. Dunque $\det(I + f^{-1}g) = 1$ in quanto determinante di una matrice triangolare con tutti 1 sulla diagonale; la conclusione segue dal teorema di Binet.

Per ogni $\lambda \in \mathbb{K}$ l'endomorfismo g commuta con $f - \lambda I$; si ha

$$p_{f+g}(\lambda) = \det(f+g-\lambda I) = \det((f-\lambda I)+g) = \det(f-\lambda I) = p_f(\lambda)\,.$$

Essendo il campo $\mathbb{K}$ infinito per ipotesi, ne consegue che $p_f(t) = p_{f+g}(t)$. Per concludere la dimostrazione basta osservare che un endomorfismo è triangolabile se e solo se il polinomio caratteristico si decompone in un prodotto di polinomi di primo grado. □

TEOREMA 11.6.4 (Decomposizione di Dunford). *Siano V uno spazio vettoriale di dimensione finita e $f \colon V \to V$ un endomorfismo triangolabile. Allora esistono, e sono unici, due endomorfismi $h, l \colon V \to V$ tali che:*

(1) *$f = h + l$,*
(2) *$lh = hl$,*
(3) *h è diagonalizzabile,*

(4) *l è nilpotente.*

DIMOSTRAZIONE. Per provare l'esistenza, siano $\lambda_1, \dots, \lambda_s$ gli autovalori di f, com molteplicità algebriche $a_1, \dots, a_s$. Siccome f è triangolabile la somma delle molteplicità algebriche è uguale alla dimensione di V e quindi si ha una decomposizione in somma diretta di sottospazi f-invarianti

$$V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_s}, \qquad E_{\lambda_i} = F_0(f - \lambda_i I).$$

Per ogni i definiamo

$$h, l \colon E_{\lambda_i} \to E_{\lambda_i}; \qquad h(v) = \lambda_i v, \quad l(v) = (f - \lambda_i I)v, \quad v \in E_{\lambda_i},$$

ed è immediato verificare che d, n soddisfano le 4 condizioni del teorema.

Per dimostrare l'unicità, siano f, h, l come nell'enunciato del teorema e siano $\lambda_1, \dots, \lambda_s$ gli autovalori di f. Il Lemma 11.6.3 applicato agli endomorfismi $f - \lambda_i I$ e $-l$ ci dice che $f - \lambda_i I$ e $h - \lambda_i I$ hanno la stessa componente zero-Fitting E_{λ_i}. Siccome h è diagonalizzabile si ha $F_0(h - \lambda_i I) = \operatorname{Ker}(h - \lambda_i I)$ e quindi, necessariamente, $h(v) = \lambda_i v$ per ogni $v \in E_{\lambda_i}$. Basta adesso ricordare che V è somma diretta degli autospazi generalizzati E_{λ_j}. □

Esercizi.

653. Mostrare con alcuni esempi che, se $f, g \colon V \to V$ sono endomorfismi nilpotenti, allora fg e $f + g$ possono essere non nilpotenti.

CAPITOLO 12

Spazi duali

Nella prima parte di questo capitolo tratteremo gli spazi duali: costoro sono, tra le altre cose, l'ambiente naturale dove collocare i sistemi di coordinate e consentono di trattare in maniera concettualmente più chiara molte proprietà viste nei capitoli precedenti. Parenti stretti degli spazi duali sono gli spazi di forme alternanti, che consentono una trattazione della teoria dei determinanti libera dalla scelta di una base e per tale ragione sono ampiamente usati in molte situazioni, specialmente quelle correlate alla geometria delle varietà algebriche e differenziabili.

Nella seconda parte del capitolo tratteremo invece il principio del massimo, il quale ci consentirà di estendere agli spazi vettoriali di dimensione infinita buona parte dei risultati già dimostrati in dimensione finita.

12.1. Spazi duali

Gli spazi duali sono particolari tipi di spazi di applicazioni e sono l'ambiente naturale dove collocare i sistemi di coordinate.

DEFINIZIONE 12.1.1. Un **funzionale lineare** su di uno spazio vettoriale V sul campo $\mathbb{K}$ è un'applicazione lineare $\varphi\colon V \to \mathbb{K}$. Lo spazio vettoriale dei funzionali lineari viene chiamato **duale** di V e viene indicato con $V^\vee$; nelle notazioni della Sezione 5.5 si ha dunque

$$V^\vee = \mathrm{Hom}(V, \mathbb{K}).$$

Dato un funzionale lineare $\varphi \in V^\vee$ ed un vettore $v \in V$, risulta utile adottare la notazione

$$\langle \varphi, v \rangle = \varphi(v) \in \mathbb{K}$$

per indicare il valore di φ calcolato nel vettore v. In questo modo risulta definita un'applicazione

$$\langle -, - \rangle\colon V^\vee \times V \to \mathbb{K},$$

detta **bracket** o **accoppiamento di dualità** che soddisfa le seguenti proprietà:

(1) $\langle \varphi, v_1 + v_2 \rangle = \langle \varphi, v_1 \rangle + \langle \varphi, v_2 \rangle$ per ogni $\varphi \in V^\vee$, $v_1, v_2 \in V$;
(2) $\langle \varphi_1 + \varphi_2, v \rangle = \langle \varphi_1, v \rangle + \langle \varphi_2, v \rangle$ per ogni $\varphi_1, \varphi_2 \in V^\vee$, $v \in V$;
(3) $\langle a\varphi, v \rangle = \langle \varphi, av \rangle = a\langle \varphi, v \rangle$ per ogni $\varphi \in V^\vee$, $v \in V$ e $a \in \mathbb{K}$.

Infatti le uguaglianze $\langle \varphi, v_1 + v_2 \rangle = \langle \varphi, v_1 \rangle + \langle \varphi, v_2 \rangle$ e $\langle \varphi, av \rangle = a\langle \varphi, v \rangle$ corrispondono, nella notazione usuale alle uguaglianze $\varphi(v_1 + v_2) = \varphi(v_1) + \varphi(v_1)$ e $\varphi(av) = a\varphi(v)$ e quindi seguono dalla linearità di φ. Similmente le rimanenti uguaglianze corrispondono a $(\varphi_1 + \varphi_2)(v) = \varphi_1(v) + \varphi_2(v)$ e $(a\varphi)(v) = a(\varphi(v))$ e quindi seguono dalla struttura di spazio vettoriale sul duale.

ESEMPIO 12.1.2. Segue dal Teorema 5.5.1 che se V ha dimensione finita n, allora ogni scelta di una base determina un isomorfismo tra lo spazio vettoriale $V^\vee = \mathrm{Hom}(V, \mathbb{K})$ e lo spazio $M_{1,n}(\mathbb{K})$, che ha la stessa dimensione n.

In dimensione finita, uno spazio ed il suo duale hanno la stessa dimensione.

Nel caso dello spazio vettoriale $V = \mathbb{K}^n$ con la base canonica, la regola del prodotto riga per colonna permette di definire un'applicazione lineare dallo spazio dei vettori riga

$$\mathbb{K}^{(n)} = \{(a_1, \dots, a_n) \mid a_1, \dots, a_n \in \mathbb{K}\}$$

al duale dello spazio $\mathbb{K}^n$: più precisamente per ogni vettore riga $a = (a_1, \dots, a_n)$ consideriamo l'applicazione lineare

$$\phi_a \colon \mathbb{K}^n \to \mathbb{K}, \qquad \phi_a \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = (a_1, \dots, a_n) \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = a_1 b_1 + \cdots + a_n b_n,$$

e definiamo

$$\phi \colon \mathbb{K}^{(n)} \to (\mathbb{K}^n)^\vee, \qquad \phi(a) = \phi_a,$$

ossia per ogni $a \in \mathbb{K}^{(n)}$ e $b \in \mathbb{K}^n$ si ha

$$\langle \phi(a), b \rangle = a \cdot b \quad \text{(prodotto riga per colonna)}.$$

Si verifica facilmente che ϕ è un isomorfismo di spazi vettoriali.

Osservazione 12.1.3. In letteratura si trovano diversi simboli per indicare il duale di uno spazio vettoriale V, i più noti sono nell'ordine $V^*, V^\vee$ e V'. La notazione che usiamo in queste note è stata adottata da Grothendieck negli *Eléments de Géométrie Algébrique* (1960) ed è quella attualmente più comune in geometria algebrica. La scelta più convenzionale di usare V^* è stata scartata senza rimpianto per gli innumerevoli conflitti notazionali che porterebbe in geometria algebrica, topologia algebrica ed algebra omologica.

L'indipendenza lineare in $V^\vee$ ha una chiara interpretazione geometrica:

Teorema 12.1.4. *Sia V uno spazio vettoriale sul campo $\mathbb{K}$ e siano $\varphi_1, \dots, \varphi_r \in V^\vee$. Consideriamo l'applicazione lineare*

$$\varphi \colon V \to \mathbb{K}^r, \qquad \varphi(v) = \begin{pmatrix} \varphi_1(v) \\ \vdots \\ \varphi_r(v) \end{pmatrix},$$

di componenti $\varphi_1, \dots, \varphi_r$. Allora φ è surgettiva se e solo se $\varphi_1, \dots, \varphi_r$ sono vettori linearmente indipendenti in $V^\vee$.

Dimostrazione. Se φ non è surgettiva allora l'immagine è un sottospazio vettoriale proprio e quindi contenuto nel nucleo di un funzionale lineare non nullo su $\mathbb{K}^r$. Per quanto già dimostrato esiste quindi un vettore riga non nullo $(a_1, \dots, a_r)$ tale che

$$a_1 \varphi_1(v) + \cdots + a_r \varphi_r(v) = 0, \qquad \text{per ogni } v \in V,$$

e quindi vale $a_1 \varphi_1 + \cdots + a_r \varphi_r = 0$ in $V^\vee$. Viceversa se i φ_i sono linearmente dipendenti e $a_1 \varphi_1 + \cdots + a_r \varphi_r = 0$ allora l'immagine di φ è contenuta nell'iperpiano di equazione $a_1 x_1 + \cdots + a_r x_r = 0$.

□

Corollario 12.1.5. *Sia V uno spazio vettoriale e siano $\varphi_1, \dots, \varphi_r, \psi \in V^\vee$. Allora vale $\psi \in \operatorname{Span}(\varphi_1, \dots, \varphi_r)$ se e solo se*

$$\operatorname{Ker} \varphi_1 \cap \cdots \cap \operatorname{Ker} \varphi_r \subseteq \operatorname{Ker} \psi.$$

Dimostrazione. Se ψ è una combinazione lineare di $\varphi_1, \dots, \varphi_r$, diciamo

$$\psi = a_1 \varphi_1 + \cdots + a_r \varphi_r,$$

allora per ogni $v \in \operatorname{Ker} \varphi_1 \cap \cdots \cap \operatorname{Ker} \varphi_r$ vale

$$\psi(v) = a_1 \varphi_1(v) + \cdots + a_r \varphi_r(v) = 0$$

e quindi $v \in \operatorname{Ker} \psi$. Viceversa, se

$$\operatorname{Ker} \varphi_1 \cap \cdots \cap \operatorname{Ker} \varphi_r \subseteq \operatorname{Ker} \psi,$$

e $\dim \operatorname{Span}(\varphi_1, \dots, \varphi_r) = h$, allora, a meno di permutazioni degli indici possiamo supporre $\varphi_1, \dots, \varphi_h$ linearmente indipendenti e

$$\varphi_i \in \operatorname{Span}(\varphi_1, \dots, \varphi_h)$$

per ogni i. Per la prima parte del corollario si ha quindi

$$\operatorname{Ker} \varphi_1 \cap \cdots \cap \operatorname{Ker} \varphi_h \subseteq \operatorname{Ker} \varphi_i$$

per ogni i e di conseguenza

$$\operatorname{Ker}\varphi_1 \cap \cdots \cap \operatorname{Ker}\varphi_h = \operatorname{Ker}\varphi_1 \cap \cdots \cap \operatorname{Ker}\varphi_r \subseteq \operatorname{Ker}\psi.$$

Supponiamo per assurdo che $\psi \notin \operatorname{Span}(\varphi_1, \ldots, \varphi_h)$, allora l'applicazione

$$\Psi\colon V \to \mathbb{K}^{h+1}, \qquad \Psi(v) = \begin{pmatrix} \psi(v) \\ \varphi_1(v) \\ \vdots \\ \varphi_h(v) \end{pmatrix},$$

è surgettiva ed esiste quindi $v \in V$ tale che $\Psi(v) = (1, 0, 0, \ldots, 0)^T$. Ma questo è assurdo perché un tale v appartiene all'intersezione dei nuclei di $\varphi_1, \ldots, \varphi_h$. □

Ricordiamo che per definizione un **iperpiano** in uno spazio vettoriale V è un sottospazio vettoriale che è nucleo di un funzionale lineare non nullo. Dunque ad ogni $f \in V^\vee - \{0\}$ è associato l'iperpiano $\operatorname{Ker} f$, ed ogni iperpiano è ottenuto in questo modo. Ricordiamo anche che se V ha dimensione finita $n > 0$, un sottospazio vettoriale è un iperpiano se e solo se ha dimensione $n - 1$.

Corollario 12.1.6. *Siano $\varphi, \psi \in V^\vee$ due funzionali lineari. Allora $\operatorname{Ker}\varphi \subseteq \operatorname{Ker}\psi$ se e solo se ψ è un multiplo scalare di φ. In particolare due funzionali lineari non nulli definiscono lo stesso iperpiano se e solo se ciascuno è un multiplo scalare dell'altro.*

Dimostrazione. Trattasi del Corollario 12.1.5 nel caso $r = 1$. □

Esempio 12.1.7. Tenendo presente l'Esempio 12.1.2 il teorema precedente ci dice che per ogni iperpiano $H \subseteq \mathbb{K}^n$ esiste un vettore riga non nullo $(a_1, \ldots, a_n)$ tale che

$$H = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mid a_1x_1 + \cdots + a_nx_n = 0 \right\}.$$

Abbiamo quindi ritrovato quanto visto nell'Esempio 5.4.9.

Esempio 12.1.8. L'Esempio 12.1.2 mostra che il duale dello spazio delle matrici $M_{n,1}(\mathbb{K}) = \mathbb{K}^n$ è naturalmente isomorfo allo spazio $M_{1,n}(\mathbb{K}) = \mathbb{K}^{(n)}$. Adesso generalizziamo e mostriamo che per ogni $n, m > 0$ il duale dello spazio $M_{n,m}(\mathbb{K})$ è naturalmente isomorfo a $M_{m,n}(\mathbb{K})$. Se $A = (a_{ij})$ è una matrice quadrata abbiamo già definito la sua traccia come la somma degli elementi sulla diagonale principale; si verifica immediatamente che l'applicazione traccia

$$\operatorname{Tr}\colon M_{n,n}(\mathbb{K}) \to \mathbb{K}, \qquad \operatorname{Tr}(b_{ij}) = \sum_i b_{ii},$$

è lineare. Più in generale, per ogni matrice se $A \in M_{m,n}(\mathbb{K})$ l'applicazione

$$\operatorname{Tr}_A\colon M_{n,m}(\mathbb{K}) \to \mathbb{K}, \qquad \operatorname{Tr}_A(B) = \operatorname{Tr}(AB) = \sum_i (\sum_j a_{ij}b_{ji}),$$

è lineare (notare che il prodotto AB è una matrice quadrata di ordine m). Anche l'applicazione

$$\phi\colon M_{m,n}(\mathbb{K}) \to M_{n,m}(\mathbb{K})^\vee, \qquad \phi(A) = \operatorname{Tr}_A,$$

è lineare e soddisfa la relazione

$$\langle \phi(A), B \rangle = \operatorname{Tr}(AB), \qquad A \in M_{m,n}(\mathbb{K}),\ B \in M_{n,m}(\mathbb{K}).$$

Dato che ϕ è lineare tra spazi della stessa dimensione, per mostrare che è un isomorfismo basta mostrare che è iniettiva, ossia che se $\operatorname{Tr}(AB) = 0$ per ogni $B \in M_{n,m}(\mathbb{K})$ allora $A = 0$. Se $A = (a_{ij})$, considerando la matrice E_{ij} che ha come coefficienti 1 al posto (i, j) e 0 altrove si ha che $\operatorname{Tr}(AE_{ij}) = a_{ji}$ e quindi $\operatorname{Tr}(AE_{ij}) = 0$ per ogni i, j se e solo se $A = 0$.

Esempio 12.1.9. Per ogni coppia di spazi vettoriali V, W è definita un'applicazione

$$W \times V^\vee \xrightarrow{\otimes} \operatorname{Hom}(V, W), \qquad (w, h) \mapsto w \otimes h,$$

dove

$$w \otimes h(v) = h(v)w, \qquad v \in V.$$

L'immagine dell'applicazione $w \otimes h$ è contenuta in $\mathrm{Span}(w)$; dunque rango di $w \otimes h$ è ≤ 1 ed è uguale ad 1 se e solo se h, w sono entrambi diversi da 0.

Viceversa, ogni applicazione lineare $f\colon V \to W$ di rango 1 è del tipo $w \otimes h$: infatti, preso un qualunque isomorfismo lineare $g\colon f(V) \xrightarrow{\simeq} \mathbb{K}$ si ha $f = g^{-1}(1) \otimes (g \circ f)$.

COROLLARIO 12.1.10. *Siano V, W spazi vettorial di dimensione finita. Dati $h_1, \dots, h_r \in V^\vee$ e $w_1, \dots, w_r \in W$ le seguenti condizioni sono equivalenti:*

1. *l'applicazione $f = w_1 \otimes h_1 + \cdots + w_r \otimes h_r$ ha rango r;*
2. *i vettori $w_1, \dots, w_r$ sono linearmente indipendenti in W ed i vettori $h_1, \dots, h_r$ sono linearmente indipendenti in $V^\vee$.*

DIMOSTRAZIONE. Per definizione, l'applicazione $f = w_1 \otimes h_1 + \cdots + w_r \otimes h_r$ può essere fattorizzata come

$$f\colon V \xrightarrow{\mathbf{h}} \mathbb{K}^r \xrightarrow{\mathbf{w}} W,$$

dove $\mathbf{h}$ è l'applicazione di componenti $h_1, \dots, h_r$ e w è la moltiplicazione a sinistra per il multivettore riga $(w_1, \dots, w_r)$. È quindi chiaro che il rango di f è $\leq r$ e vale l'uguaglianza se e solo se $\mathbf{h}$ e $\mathbf{w}$ hanno entrambi rango r. La conclusione segue dal Teorema 12.1.4. □

È istruttivo accennare ad una diversa dimostrazione del fatto che la prima asserzione del Corollario 12.1.10 implica la seconda, e basata sull'osservazione che valgono le relazioni

$$(x+y) \otimes h = x \otimes h + y \otimes h, \qquad x \otimes (h+k) = x \otimes h + x \otimes k, \qquad (ax) \otimes h = x \otimes (ah),$$

per ogni $x, y \in W$, $h, k \in V^\vee$ e $a \in \mathbb{K}$.

Supponiamo i vettori $w_1, \dots, w_r$ linearmente dipendenti; a meno di permutazioni degli indici possiamo supporre $w_r = \sum_{i=1}^{r-1} a_i w_i$ per opportuni $a_1, \dots, a_{r-1} \in \mathbb{K}$. Allora

$$\begin{aligned}
f &= w_1 \otimes h_1 + \cdots + w_r \otimes h_r \\
&= w_1 \otimes h_1 + \cdots + w_{r-1} \otimes h_{r-1} + \left(\sum_{i=1}^{r-1} a_i w_i \right) \otimes h_r \\
&= w_1 \otimes h_1 + \cdots + w_{r-1} \otimes h_{r-1} + \sum_{i=1}^{r-1} (a_i w_i) \otimes h_r \\
&= w_1 \otimes h_1 + \cdots + w_{r-1} \otimes h_{r-1} + \sum_{i=1}^{r-1} w_i \otimes (a_i h_r) \\
&= \sum_{i=1}^{r-1} w_i \otimes (h_i + a_i h_r)
\end{aligned}$$

e quindi f è somma di $r-1$ applicazioni di rango ≤ 1. Un ragionamento del tutto analogo mostra che se i vettori $h_1, \dots, h_r$ sono linearmente dipendenti allora f ha rango $< r$.

Esercizi.

654. Provare che i vettori

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \qquad v_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix},$$

formano una base di $\mathbb{K}^3$ e determinare i tre vettori riga che corrispondono alla base duale.

655. Sia (v_1, v_2, v_3) una base di uno spazio vettoriale V e sia $(\varphi_1, \varphi_2, \varphi_3)$ la corrispondente base duale. Per ciascuna delle seguenti basi di V, descrivere la base duale in funzione di $(\varphi_1, \varphi_2, \varphi_3)$.

1. (v_3, v_2, v_1) (risposta: $(\varphi_3, \varphi_2, \varphi_1)$),
2. $(2v_1, v_2, v_3)$,
3. $(v_1 + v_2, v_2, v_3)$,
4. $(v_1 + v_2 + v_3, v_2, v_3)$.

656. Mostrare che se $A = (a_{ij}) \in M_{n,m}(\mathbb{K})$ e $B = (b_{hk}) \in M_{m,n}(\mathbb{K})$ allora valgono le formule

$$\operatorname{Tr}(AB) = \sum_{i,j} a_{ij} b_{ji}, \qquad \operatorname{Tr}(AA^T) = \sum_{i,j} a_{ij}^2.$$

657 (♡). Siano V uno spazio vettoriale e $\phi_1, \dots, \phi_r \in V^\vee$ linearmente indipendenti. Dimostrare che esiste un sottospazio vettoriale $U \subseteq V$ di dimensione r tale che le restrizioni ad U dei funzionali ϕ_i formano una base di $U^\vee$.

12.2. Basi duali e sistemi di coordinate

Ricordiamo che un sistema di coordinate su uno spazio vettoriale V sul campo $\mathbb{K}$ è definito come una successione di applicazioni lineari $\varphi_1, \dots, \varphi_n \colon V \to \mathbb{K}$ tali che l'applicazione

$$\varphi \colon V \to \mathbb{K}^n, \qquad v \mapsto \begin{pmatrix} \varphi_1(v) \\ \vdots \\ \varphi_n(v) \end{pmatrix},$$

è un isomorfismo.

Corollario 12.2.1. *Dato uno spazio vettoriale V di dimensione finita n, una successione di applicazioni lineari $\varphi_1, \dots, \varphi_n \colon V \to \mathbb{K}$ è un sistema di coordinate su V se e soltanto se $(\varphi_1, \dots, \varphi_n)$ è una base di $V^\vee$.*

Dimostrazione. Abbiamo visto che se V ha dimensione n, allora anche $V^\vee$ ha dimensione n. Basta adesso ricordare che $\varphi_1, \dots, \varphi_n \colon V \to \mathbb{K}$ definisce un sistema di coordinate se l'applicazione

$$\varphi \colon V \to \mathbb{K}^n, \qquad v \mapsto \begin{pmatrix} \varphi_1(v) \\ \vdots \\ \varphi_n(v) \end{pmatrix},$$

è un isomorfismo di spazi vettoriali e applicare il Teorema 12.1.4. □

Data una base $(v_1, \dots, v_n)$ di uno spazio vettoriale V possiamo considerare il sistema di coordinate associato e questo ci fornisce un modo naturale per costruire una base del duale $V^\vee$: per ogni indice $i = 1, \dots, n$ consideriamo l'applicazione lineare $\varphi_i \colon V \to \mathbb{K}$ cha associa ad ogni vettore la sua i-esima coordinata nella base $(v_1, \dots, v_n)$, ossia tale che per ogni scelta di $a_1, \dots, a_n \in \mathbb{K}$ vale

$$\langle \varphi_i, a_1 v_1 + \cdots + a_n v_n \rangle = a_i.$$

Equivalentemente φ_i è l'applicazione lineare definita dalle relazioni

$$\langle \varphi_i, v_j \rangle = \delta_{ij}$$

dove δ è la funzione **delta di Kronecker** definita come

$$\delta_{ij} = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}.$$

Definizione 12.2.2. Nelle notazioni precedenti, la successione $(\varphi_1, \dots, \varphi_n)$ è una base di $V^\vee$, detta **base duale** di $(v_1, \dots, v_n)$.

Naturalmente la base duale cambia di pari passo con i cambiamenti di base. Ad esempio, in dimensione 2, se (φ_1, φ_2) è la base duale di (v_1, v_2), allora:

(1) $(\varphi_1/2, \varphi_2)$ è la base duale di $(2v_1, v_2)$;
(2) $(\varphi_1 - \varphi_2, \varphi_2)$ è la base duale di $(v_1, v_2 + v_1)$.

Proposizione 12.2.3. *Siano $(u_1, \dots, u_n)$ e $(v_1, \dots, v_n)$ due basi di uno spazio vettoriale legate dalla matrice di cambiamento*

$$A = (a_{ij}), \qquad (v_1, \dots, v_n) = (u_1, \dots, u_n)A.$$

Detta $(f_1, \dots, f_n)$ *la base duale di* $(u_1, \dots, u_n)$ *e* $(\varphi_1, \dots, \varphi_n)$ *la base duale di* $(v_1, \dots, v_n)$ *vale la formula*

$$\begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} = A \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_n \end{pmatrix}.$$

DIMOSTRAZIONE. Sia B la matrice di coefficenti $b_{ij} = \varphi_i(u_j)$; per ipotesi sappiamo che per ogni i, j si ha

$$\delta_{ij} = \varphi_i(v_j) = \varphi_i(\sum_h u_h a_{hj}) = \sum_h \varphi_i(u_h) a_{hj} = \sum_h b_{ih} a_{hj},$$

che equivale alla condizione $B = A^{-1}$. Adesso basta osservare che per ogni i, j vale

$$\left(\sum_h a_{ih} \varphi_h \right) u_j = \sum_h a_{ih} \varphi_h(u_j) = \sum_h a_{ih} b_{hj} = \delta_{ij}.$$

□

È istruttivo ridimostrare il Corollario 5.1.16 nel linguaggio degli spazi duali.

LEMMA 12.2.4. *Siano* H *un sottospazio di uno spazio vettoriale di dimensione finita* V *e* $v \in V$ *un vettore tale che* $v \notin H$*. Allora esiste* $f \in V^\vee$ *tale che*

$$H \subseteq \operatorname{Ker} f, \qquad f(v) = 1.$$

DIMOSTRAZIONE. Denotiamo $v_1 = v$, $s = \dim H + 1$ e sia $v_2, \dots, v_s$ una qualunque base di H. Siccome $v_1 \notin \operatorname{Span}(v_2, \dots, v_s)$ i vettori $v_1, v_2, \dots, v_s$ sono linearmente indipendenti e possono essere completati ad una base $v_1, \dots, v_n$ di V. Basta allora considerare come f il primo elemento della corrispondente base duale. □

TEOREMA 12.2.5. *Sia* V *uno spazio vettoriale di dimensione finita sul campo* $\mathbb{K}$ *e siano* $\varphi_1, \dots, \varphi_r \in V^\vee$*. Consideriamo l'applicazione lineare*

$$\varphi \colon V \to \mathbb{K}^r, \qquad \varphi(v) = \begin{pmatrix} \varphi_1(v) \\ \vdots \\ \varphi_r(v) \end{pmatrix},$$

che ha come componenti $\varphi_1, \dots, \varphi_r$*. Allora:*

1. φ *è surgettiva se e solo se* $\varphi_1, \dots, \varphi_r$ *sono vettori linearmente indipendenti in* $V^\vee$*;*
2. φ *è iniettiva se e solo se* $\varphi_1, \dots, \varphi_r$ *generano lo spazio duale* $V^\vee$*;*
3. φ *è bigettiva se e solo se* $\varphi_1, \dots, \varphi_r$ *sono una base del duale.*

DIMOSTRAZIONE. Il primo punto è già stato dimostrato senza ipotesi alcuna sulla dimensione di V. Per quanto riguarda il secondo punto, sia $h \leq r$ la dimensione del sottospazio vettoriale generato da $\varphi_1, \dots, \varphi_r$; a meno di permutazioni degli indici possiamo supporre $\varphi_1, \dots, \varphi_h$ linearmente indipendenti e $\varphi_i \in \operatorname{Span}(\varphi_1, \dots, \varphi_h)$ per ogni $i > h$. Per il Corollario 12.1.5 si ha

$$\operatorname{Ker} \varphi = \operatorname{Ker} \varphi_1 \cap \cdots \cap \operatorname{Ker} \varphi_r = \operatorname{Ker} \varphi_1 \cap \cdots \cap \operatorname{Ker} \varphi_h = \operatorname{Ker} \phi,$$

dove $\phi \colon V \to \mathbb{K}^h$ è l'applicazione che ha come componenti $\varphi_1, \dots, \varphi_h$. Per il punto precedente ϕ è surgettiva e quindi il suo nucleo ha dimensione uguale a $\dim V - h$. Dato che $\operatorname{Ker} \phi = \operatorname{Ker} \varphi$ otteniamo che φ è iniettiva se e solo se $\dim V = h$.

La terza proprietà segue immediatamente dalle precedenti e dal calcolo delle dimensioni. □

Esercizi.

658. Sia V spazio vettoriale di dimensione finita:
   (1) Dimostrare che ogni sottospazio vettoriale di V è intersezione di iperpiani.
   (2) Sia $v_1, \dots, v_n$ una base di V e denotiamo $v_0 = v_1 + v_2 + \cdots + v_n$. Sia $f \colon V \to V$ un'applicazione lineare tale che $f(v_i) = \lambda_i v_i$ per ogni $i = 0, \dots, n$ ed opportuni $\lambda_i \in \mathbb{K}$. Dimostrare che $\lambda_0 = \lambda_1 = \cdots = \lambda_n$.
   (3) Sia $f \colon V \to V$ lineare tale che $f(L) \subseteq L$ per ogni retta $L \subseteq V$ (retta=sottospazio di dimensione 1). Dimostrare che f è un multiplo scalare dell'identità.

(4) Sia $f\colon V \to V$ lineare tale che $f(H) \subseteq H$ per ogni iperpiano $H \subseteq V$. Dimostrare che f è un multiplo scalare dell'identità.

659. Sia V spazio vettoriale di dimensione n e siano $H_1, \dots, H_n \subseteq V$ iperpiani fissati e tali che $H_1 \cap \cdots \cap H_n = 0$. Dimostrare che esiste una base $v_1, \dots, v_n$ di V tale che $v_i \in H_j$ per ogni $i \neq j$.

12.3. Biduale e trasposta

Dato uno spazio vettoriale V sul campo $\mathbb{K}$ possiamo costruire il duale $V^\vee$ e ripetere la procedura costruendo il duale del duale, detto anche **biduale**:[1]

$$V^{\vee\vee} = \operatorname{Hom}(V^\vee, \mathbb{K}) = \operatorname{Hom}(\operatorname{Hom}(V, \mathbb{K}), \mathbb{K}).$$

Mostriamo adesso che esiste un'applicazione lineare canonica $\imath\colon V \to V^{\vee\vee}$: ad ogni vettore $v \in V$ associamo l'applicazione lineare

$$\imath(v)\colon V^\vee \to \mathbb{K}$$

definita dalla formula

$$\imath(v)(f) = f(v), \qquad f \in V^\vee.$$

Per ogni vettore $v \in V$ l'applicazione $\imath(v)$ è lineare, ciò è una diretta conseguenza della definizione di somma e prodotto per scalare in $V^\vee$. Anche l'applicazione $\imath\colon V \to V^{\vee\vee}$ è lineare, infatti per ogni $u, v \in V$ e per ogni $f \in V^\vee$ si ha

$$\imath(u+v)(f) = f(u+v) = f(u) + f(v) = \imath(u)(f) + \imath(v)(f) = (\imath(u) + \imath(v))(f)$$

e quindi $\imath(u+v) = \imath(u) + \imath(v)$ in $V^{\vee\vee}$. Similmente si prova che $\imath(\lambda v) = \lambda \imath(v)$ per $\lambda \in \mathbb{K}$.

Si noti che per definire $\imath$ non abbiamo fatto uso di basi o scelte arbitrarie e questo ci autorizza ad aggettivare *naturale* l'applicazione $\imath$.

TEOREMA 12.3.1. *Per ogni spazio vettoriale di dimensione finita V, l'applicazione naturale $\imath\colon V \to V^{\vee\vee}$ è un isomorfismo di spazi vettoriali.*

DIMOSTRAZIONE. Siccome $V, V^\vee$ e $V^{\vee\vee}$ hanno la stessa dimensione, per mostrare che $\imath$ è un isomorfismo basta mostrare che $\operatorname{Ker} \imath = 0$. Sia $v \in V$ diverso da 0, per il Lemma 12.2.4 esiste $f \in V^\vee$ tale che $f(v) \neq 0$, quindi $\imath(v)(f) = f(v) \neq 0$ e dunque $\imath(v) \neq 0$.

□

OSSERVAZIONE 12.3.2. Il Teorema 12.3.1 è completamente falso se V ha dimensione infinita. Dimostreremo infatti nel Corollario 12.5.8, che per ogni V di dimensione infinita, l'applicazione $\imath$ è iniettiva ma non è mai surgettiva.

DEFINIZIONE 12.3.3. Sia $f\colon V \to W$ una applicazione lineare. La **trasposta** di f è l'applicazione lineare $f^*\colon W^\vee \to V^\vee$ definita dalla composizione a destra con f:

$$\psi \mapsto f^*(\psi) = \psi \circ f, \qquad \begin{array}{ccc} V & \xrightarrow{\quad f \quad} & W \\ & {\scriptstyle f^*(\psi)} \searrow \quad \swarrow {\scriptstyle \psi} & \\ & \mathbb{K} & \end{array}$$

In altri termini, dato un funzionale $\psi \in W^\vee$, la sua immagine $f^*(\psi) \in V^\vee$ è determinata dalla formula

$$f^*(\psi)(v) = \psi(f(v)), \qquad \text{per ogni} \quad v \in V,$$

È immediato verificare che $f^* : W^\vee \to V^\vee$ è lineare e cioè che

$$f^*(\psi + \psi') = f^*(\psi) + f^*(\psi'), \qquad f^*(a\psi) = a\, f^*(\psi),$$

per ogni $\psi, \psi' \in W^\vee$ ed ogni $a \in \mathbb{K}$. Infatti, per ogni $v \in V$

$$f^*(\psi + \psi')(v) = (\psi + \psi')f(v) = \psi f(v) + \psi' f(v) = f^*(\psi)(v) + f^*(\psi')(v),$$

e

$$f^*(a\psi)(v) = (a\psi)(f(v)) = a(\psi f(v)) = a f^*(\psi)(v).$$

[1] A volte viene detto anche *biduale algebrico* per distinguerlo dal biduale topologico, oggetto che troverete nei corsi di analisi reale e analisi funzionale.

È molto importante notare che anche la definizione di f^* è *intrinseca*, non dipende cioè dalla scelta di basi in V e W. Date due applicazioni lineari

$$V \xrightarrow{f} W \xrightarrow{g} U$$

si ha:

$$(gf)^* = f^*g^* : U^\vee \to V^\vee.$$

Anche in questo caso le dimostrazioni sono immediate. Dobbiamo far vedere che

$$(gf)^*(\psi) = f^*g^*(\psi)\,, \quad \forall\, \psi \in U^\vee.$$

Questo vuol dire che dobbiamo mostrare che

$$(gf)^*(\psi)(u) = f^*g^*(\psi)(u)\,, \quad \forall\, \psi \in U^*, \quad \forall\, u \in U.$$

In effetti si ha:

$$(gf)^*(\psi)(u) = \psi gf(u) = (g^*(\psi))(f(u)) = (f^*g^*\psi)(u).$$

Se facciamo due volte la trasposta di un'applicazione $f\colon V \to W$ troviamo un'applicazione lineare $f^{**}\colon V^{\vee\vee} \to W^{\vee\vee}$. A questo punto non è sorprendente scoprire che, via i morfismi naturali $\imath\colon V \to V^{\vee\vee}$ e $\imath\colon W \to W^{\vee\vee}$, l'applicazione f^{**} coincide con f, o più precisamente che vale

$$f^{**} \circ \imath = \imath \circ f.$$

Sia infatti $v \in V$, per dimostrare che $f^{**}(\imath(v)) = \imath(f(v))$ come elementi di $W^{\vee\vee}$ bisogna provare che per ogni $\psi\colon W \to \mathbb{K}$, $\psi \in W^\vee$, vale

$$f^{**}(\imath(v))(\psi) = \imath(f(v))(\psi).$$

Si ha $\imath(f(v))(\psi) = \psi(f(v))$; d'altra parte

$$f^{**}(\imath(v)) = \imath(v) \circ f^*\colon W^\vee \to V^\vee \to \mathbb{K}$$

e quindi

$$f^{**}(\imath(v))(\psi) = \imath(v) \circ f^*(\psi) = \imath(v)(\psi \circ f) = \psi(f(v)).$$

Per giustificare il termine "trasposta" assegnato all'applicazione f^* mettiamoci in dimensione finita e cerchiamo di scoprire la relazione esistente tra la matrice che rappresenta f rispetto a due basi prefissate e la matrice che rappresenta f^* rispetto alle basi duali. Siano dunque $(v_1, \dots, v_n)$ una base di V, $(w_1, \dots, w_m)$ una base di W e $(\varphi_1, \dots, \varphi_n)$, $(\psi_1, \dots, \psi_m)$ le rispettive basi duali. L'applicazione f sarà allora rappresentata dall'unica matrice (a_{ij}) tale che

$$f(v_j) = \sum_i w_i a_{ij}, \qquad j = 1, \dots, n,$$

mentre f^* sarà allora rappresentata dall'unica matrice (b_{hk}) tale che

$$f^*(\psi_k) = \sum_h \varphi_h b_{hk}, \qquad k = 1, \dots, m.$$

Per ogni coppia di indici j, k abbiamo

$$f^*(\psi_k)(v_j) = \psi_k(f(v_j))$$

e sostituendo le rispettive espressioni ad entrambi i membri

$$(\sum_h \varphi_h b_{hk})(v_i) = \sum_h \varphi_h(v_i) b_{hk} = \psi_k(\sum_j w_j a_{ji}) = \sum_j \psi_k(w_j) a_{ji}.$$

L'espressione a sinistra è uguale a b_{ik} e quella a destra è uguale a a_{ki}. Dunque la matrice (b_{hk}) è uguale alla trasposta di (a_{ij}).

Esercizi.

660. Usando l'isomorfismo $M_{n,n}(\mathbb{K})^\vee \cong M_{n,n}(\mathbb{K})$ introdotto nell'Esempio 12.1.8, descrivere la trasposta f^* dell'applicazione

$$f\colon \mathbb{K} \to M_{n,n}(\mathbb{K}), \qquad f(t) = tI \quad (I = \text{matrice identità}).$$

661. Sia V uno spazio vettoriale di dimensione finita. Determinare tutte le applicazioni lineari $h\colon V \to V^\vee$ tali che $h = f^*hf$ per ogni isomorfismo lineare $f\colon V \to V$.

12.4. Dualità vettoriale

Siano V uno spazio vettoriale di dimensione finita e $V^\vee$ il suo duale. Per ogni sottoinsieme $S \subseteq V$ definiamo il suo **annullatore** $\operatorname{Ann}(S) \subseteq V^\vee$ come l'insieme di tutti i funzionali lineari che si annullano su S:

$$\operatorname{Ann}(S) = \{\phi \in V^\vee \mid \phi(s) = 0 \quad \forall s \in S\} = \{\phi \in V^\vee \mid S \subseteq \operatorname{Ker}\phi\}.$$

È facile verificare che $\operatorname{Ann}(S)$ è un sottospazio vettoriale di $V^\vee$: se $\phi, \psi \in \operatorname{Ann}(S)$ e $\lambda \in \mathbb{K}$ allora per ogni $s \in S$ vale

$$(\phi + \psi)(s) = \phi(s) + \psi(s) = 0 + 0 = 0, \qquad (\lambda\phi)(s) = \lambda\phi(s) = 0,$$

e quindi $\phi+\psi, \lambda\phi \in \operatorname{Ann}(S)$. Le seguenti proprietà seguono immediatamente dalla definizione:

(1) se $S \subseteq T$ allora $\operatorname{Ann}(T) \subseteq \operatorname{Ann}(S)$;
(2) $\operatorname{Ann}(S \cup T) = \operatorname{Ann}(S) \cap \operatorname{Ann}(T)$.

Inoltre, se $H = \operatorname{Span}(S) \subseteq V$ è il sottospazio vettoriale generato da S, allora $\operatorname{Ann}(H) = \operatorname{Ann}(S)$: siccome $S \subseteq H$, abbiamo già visto che $\operatorname{Ann}(H) \subseteq \operatorname{Ann}(S)$; viceversa se $\phi \in \operatorname{Ann}(S)$, per ogni $v \in H$ esistono $s_1, \dots, s_n \in S$ ed una combinazione lineare $v = a_1 s_1 + \cdots + a_n s_n$ e quindi $\phi(v) = \sum a_i \phi(s_i) = 0$. Questo prova che $\phi \in \operatorname{Ann}(H)$.

Lemma 12.4.1. *Siano V spazio vettoriale di dimensione finita e $H, K \subset V$ sottospazi vettoriali. Allora:*

(1) $\operatorname{Ann}(0) = V^\vee$, $\operatorname{Ann}(V) = 0$*;*
(2) *se $H \subseteq K$ allora* $\operatorname{Ann}(K) \subseteq \operatorname{Ann}(H)$*;*
(3) $\operatorname{Ann}(H + K) = \operatorname{Ann}(H) \cap \operatorname{Ann}(K)$*;*
(4) $\operatorname{Ann}(H \cap K) = \operatorname{Ann}(H) + \operatorname{Ann}(K)$.

Dimostrazione. Le prime due proprietà sono ovvie e, siccome $H + K = \operatorname{Span}(H \cup K)$, la terza segue da quanto visto poco sopra.

Siccome $H \cap K \subset H$ ne segue $\operatorname{Ann}(H) \subseteq \operatorname{Ann}(H \cap K)$; similmente $\operatorname{Ann}(K) \subseteq \operatorname{Ann}(H \cap K)$ e quindi $\operatorname{Ann}(H) + \operatorname{Ann}(K) \subseteq \operatorname{Ann}(H \cap K)$. Rimane quindi da dimostrare che se $f\colon V \to \mathbb{K}$ è lineare e $f(H \cap K) = 0$, allora esistono $f_1, f_2\colon V \to \mathbb{K}$ tali che $f = f_1 + f_2$, $f_1(H) = 0$, $f_2(K) = 0$. A tal fine supponiamo

$$\dim H \cap K = p, \qquad \dim H = q, \quad \dim K = r, \quad \dim V = n \geq r + q - p,$$

e prendiamo una base di V del tipo $(e_1, , \dots, e_q, \epsilon_1, \dots, \epsilon\epsilon_{n-q})$ dove

(1) $e_1, , \dots, e_p$ è una base di $H \cap K$;
(2) $e_1, , \dots, e_q$ è una base di H;
(3) $e_1, , \dots, e_p, \epsilon_1, \dots, \epsilon_{r-p}$ è una base di K.

Definiamo f_1, f_2 ponendo $f_1(e_i) = 0$, $f_1(\epsilon_i) = f(\epsilon_i)$ per ogni i e $f_2 = f - f_1$; allora $f_1(H)$ e $f_2(K) = 0$.

□

Se $E \subseteq V^\vee$ è un qualunque sottoinsieme si definisce il suo **luogo di zeri** $Z(E) \subseteq V$ come l'insieme dei vettori che annullanno tutti i funzionali di E:

$$Z(E) = \{v \in V \mid \phi(v) = 0 \ \forall \phi \in E\} = \bigcap_{\phi \in E} \operatorname{Ker}\phi.$$

Anche in questo caso si verifica immediatamente che $Z(E)$ è un sottospazio vettoriale.

Se ci limitiamo a considerare annullatori e luoghi di zeri di sottospazi vettoriali, abbiamo definito due applicazioni di insiemi

$$\left\{\begin{array}{c}\text{sottospazi vettoriali}\\ \text{di } V\end{array}\right\} \underset{Z}{\overset{\operatorname{Ann}}{\rightleftarrows}} \left\{\begin{array}{c}\text{sottospazi vettoriali}\\ \text{di } V^\vee\end{array}\right\}$$

che per il prossimo teorema sono una l'inversa dell'altra, ed in particolare entrambe bigettive.

Teorema 12.4.2. *Sia V uno spazio vettoriale di dimensione finita n. Allora:*

(1) *per ogni sottospazio vettoriale $W \subseteq V$ vale*

$$\dim \operatorname{Ann}(W) + \dim W = n, \qquad Z(\operatorname{Ann}(W)) = W;$$

(2) *per ogni sottospazio vettoriale* $H \subseteq V^\vee$ *vale*

$$\dim Z(H) + \dim H = n, \qquad \mathrm{Ann}(Z(H)) = H.$$

DIMOSTRAZIONE. Sia $W \subseteq V$ sottospazio vettoriale, sia r la sua dimensione e scegliamo una base $v_1, \dots, v_n$ di V tale che $v_1, \dots, v_r \in W$. Sia $\varphi_1, \dots, \varphi_n \in V^\vee$ la base duale; è immediato verificare che un generico elemento

$$a_1\varphi_1 + \cdots + a_n\varphi_n$$

del duale appartiene all'annullatore di W se e solo se $a_1 = \cdots = a_r = 0$. Dunque $\mathrm{Ann}(W)$ è generato da $\varphi_{r+1}, \dots, \varphi_n$ ed ha dunque dimensione $n - r$.

Sia $H \subseteq V^\vee$ è un sottospazio vettoriale di dimensione h e scegliamo una base $\phi_1, \dots, \phi_h$ di H. Per il Teorema 12.2.5 l'applicazione $\phi\colon V \to \mathbb{K}^h$ di componenti $\phi_1, \dots, \phi_h$ è surgettiva e quindi il sottospazio

$$Z(H) = Z(\phi_1, \dots, \phi_h) = \mathrm{Ker}\,\phi$$

ha dimensione $n - h$.

Se $S \subseteq V$ è un sottoinsieme e $s \in S$, allora $\phi(s) = 0$ per ogni $\phi \in \mathrm{Ann}(S)$ e quindi $s \in Z(\mathrm{Ann}(S))$, ossia $S \subseteq Z(\mathrm{Ann}(S))$. Similmente se E è un sottoinsieme di $V^\vee$ e $\phi \in E$ si ha $\phi(v) = 0$ per ogni $v \in Z(E)$ e quindi $\phi \in \mathrm{Ann}(Z(E))$, ossia $E \subseteq \mathrm{Ann}(Z(E))$.

Se $W \subseteq V$ è un sottospazio vettoriale, allora $W \subseteq Z(\mathrm{Ann}(W))$, i due sottospazi hanno la stessa dimensione e quindi coincidono. Se H è un sottospazio di $V^\vee$ allora $H \subseteq \mathrm{Ann}(Z(H))$ ed i due sottospazi hanno la stessa dimensione. □

COROLLARIO 12.4.3. *Siano* V *spazio vettoriale di dimensione finita e* $H, K \subset V^\vee$ *sottospazi vettoriali. Allora:*

(1) *se* $H \subseteq K$ *allora* $Z(K) \subseteq Z(H)$*;*
(2) $Z(H + K) = Z(H) \cap Z(K)$*;*
(3) $Z(H \cap K) = Z(H) + Z(K)$.

DIMOSTRAZIONE. Conseguenza quasi immediata del Lemma 12.4.1; i dettagli sono lasciati per esercizio al lettore. Ad esempio, siccome Ann è iniettivo, per dimostrare che $Z(H + K) = Z(H) \cap Z(K)$ basta dimostrare che $H + K = \mathrm{Ann}(Z(H + K)) = \mathrm{Ann}(Z(H) \cap Z(K))$. D'altra parte per il lemma si ha $\mathrm{Ann}(Z(H) \cap Z(K)) = \mathrm{Ann}(Z(H)) + \mathrm{Ann}(Z(K)) = H + K$. □

In estrema sintesi, abbiamo visto che in dimensione finita annullatori e luoghi di zeri hanno proprietà analoghe. Dimostriamo adesso che, a meno di isomorfismi canonici, annullatore e luogo di zeri sono la stessa cosa.

TEOREMA 12.4.4. *Nelle notazioni precedenti se* V *ha dimensione finita ed* $\imath\colon V \to V^{\vee\vee}$ *denota l'isomorfismo canonico, allora per* $W \subseteq V$ *e* $H \subseteq V^\vee$ *sottospazi vettoriali si ha:*

$$\imath(Z(H)) = \mathrm{Ann}(H), \qquad \imath(W) = \mathrm{Ann}(\mathrm{Ann}(W)).$$

DIMOSTRAZIONE. Sia $H \subseteq V^\vee$ un sottospazio vettoriale; il fatto che $\imath$ è un isomorfismo implica in particolare che $\imath(Z(H))$ ha la stessa dimensione di $Z(H)$ e quindi la stessa dimensione di $\mathrm{Ann}(H)$. Per mostrare l'uguaglianza $\imath(Z(H)) = \mathrm{Ann}(H)$ basta quindi mostrare che $\imath(Z(H)) \subseteq \mathrm{Ann}(H)$. Sia $v \in Z(H)$, allora per ogni $f \in H$ vale

$$\imath(v)(f) = f(v) = 0$$

e questo implica che $\imath(v) \in \mathrm{Ann}(H)$. Quando $H = \mathrm{Ann}(W)$, siccome già sappiamo che $W = Z(\mathrm{Ann}(W)) = Z(H)$ si ottiene

$$\imath(W) = \imath(Z(\mathrm{Ann}(W)) = \mathrm{Ann}(\mathrm{Ann}(W)).$$

□

Il combinato dei precedenti 4 enunciati è il nocciolo di quella che viene comunemente chiamata *dualità vettoriale*. Le sue applicazioni più rilevanti riguardano la geometria proiettiva e vanno quindi al di là degli obiettivi di queste note; rimanendo nell'ambito dell'algebra lineare possiamo usare la dualità vettoriale per ridimostrare, in maniera più astratta e concettuale, che il rango di una matrice coincide con il rango della sua trasposta.

TEOREMA 12.4.5. *Sia $f\colon V \to W$ un'applicazione lineare tra spazi di dimensione finita. Allora vale*

$$\operatorname{Ker} f^* = \operatorname{Ann}(f(V)), \qquad f^*(W^\vee) = \operatorname{Ann}(\operatorname{Ker} f),$$

ed in particolare f e f^* hanno lo stesso rango.

DIMOSTRAZIONE. Vale $\phi \in \operatorname{Ker} f^*$ se e solo se $f^*(\phi)(v) = 0$ per ogni $v \in V$ e siccome $f^*(\phi)(v) = \phi(f(v))$ questo equivale a dire che $\phi(w) = 0$ per ogni $w \in f(V)$. Abbiamo quindi provato che $\operatorname{Ker} f^* = \operatorname{Ann}(f(V))$. Per le formule sul computo delle dimensioni abbiamo che

$$\operatorname{rg}(f^*) = \dim W - \dim \operatorname{Ker} f^* = \dim W - \dim \operatorname{Ann}(f(V)) = \dim f(V) = \operatorname{rg}(f).$$

È del tutto ovvio che ogni funzionale del tipo $f^*(\phi) = \phi \circ f$ si annulla sul nucleo di f e quindi $f^*(W^\vee) \subseteq \operatorname{Ann}(\operatorname{Ker} f)$; d'altra parte l'uguaglianza dei ranghi ed il teorema del rango provano che $f^*(W^\vee)$ e $\operatorname{Ann}(\operatorname{Ker} f)$ hanno la stessa dimensione e questo basta per concludere la dimostrazione. □

Esercizi.

662. Sia V uno spazio vettoriale di dimensione finita e siano $U, W \subseteq V$ due sottospazi tali che $U \oplus W = V$. Provare che $V^\vee = \operatorname{Ann}(U) \oplus \operatorname{Ann}(W)$.

663. Mostrare che per ogni sottoinsieme finito S di uno spazio vettoriale V di dimensione finita la sua chiusura lineare è uguale a $Z(\operatorname{Ann}(S))$.

664. Siano V uno spazio vettoriale di dimensione finita e

$$f\colon V \to V^\vee, \qquad u \mapsto f_u,$$

un'applicazione lineare. Consideriamo l'applicazione

$$\phi\colon V \times V \to \mathbb{K}, \qquad \phi(u, v) = f_u(v).$$

Dimostrare che ϕ è lineare in ciascuna variabile, ossia che

$$\phi(\lambda u, v) = \phi(u, \lambda v) = \lambda\phi(u, v), \qquad \lambda \in \mathbb{K}$$

$$\phi(u_1 + u_2, v) = \phi(u_1, v) + \phi(u_2, v), \qquad \phi(u, v_1 + v_2) = \phi(u, v_1) + \phi(u, v_2),$$

e che

$$A = \{u \in V \mid \phi(u, v) = 0 \ \ \forall\, v \in V\}, \qquad B = \{v \in V \mid \phi(u, v) = 0 \ \ \forall\, u \in V\},$$

sono sottospazi vettoriali della stessa dimensione.

665. Provare che una successione di spazi vettoriali di dimensione finita è esatta se e solo se la sua dualizzata è esatta; in particolare $f\colon V \to W$ è iniettiva (risp.: surgettiva) se e solo se $f^*\colon W^\vee \to V^\vee$ è surgettiva (risp.: iniettiva). Provare che i due enunciati del Teorema 5.6.8 sono uno il dualizzato dell'altro.

666. Siano $f\colon V \to V$ un endomorfismo di uno spazio vettoriale di dimensione finita e $0 \neq \varphi \in V^\vee$. Provare che φ è un autovettore di f^* se e solo se $\operatorname{Ker}(\varphi)$ è un sottospazio f-invariante di V.

12.5. Il principio del massimo

Nei capitoli precedenti abbiamo usato più volte il fatto che se X è un insieme finito, allora ogni famiglia $\mathcal{F}$ di sottoinsiemi di X possiede elementi con un massimo numero di elementi, e quindi possiede elementi massimali, definiti nel modo seguente.

DEFINIZIONE 12.5.1. Siano X un insieme (possibilmente infinito) e $\mathcal{F} \subseteq \mathcal{P}(X)$ una famiglia (possibilmente infinita) di sottoinsiemi di X. Un elemento $M \in \mathcal{F}$ si dice **massimale** se non esistono elementi di $\mathcal{F}$ che contengono strettamente M. Equivalentemente, M è massimale se accade che per ogni $N \in \mathcal{F}$ tale che $M \subseteq N$ si ha $N = M$.

Ad esempio se $X = \{1, 2, \dots, 2017\}$ e $\mathcal{F}$ è la famiglia formata da tutti i sottoinsiemi di X con un numero pari di numeri interi, allora $\mathcal{F}$ contiene esattamente 2017 elementi massimali, ciascuno dei quali ottenuto togliendo un intero $n \in X$.

Dobbiamo prestare attenzione al fatto che, qualora X è infinito, una famiglia di sottoinsiemi di X può non avere elementi massimali. Ad esempio la famiglia di tutti i sottoinsiemi finiti di $\mathbb{N}$ non possiede elementi massimali: preso un qualsiasi sottoinsieme finito di $\mathbb{N}$ possiamo trovarne uno strettamente più grande aggiungendo un elemento.

L'esperienza matematica mostra che è molto utile trovare condizioni su di una famiglia $\mathcal{F} \subseteq \mathcal{P}(X)$ che siano sufficienti per dedurre che $\mathcal{F}$ abbia elementi massimali; il principio del massimo fornisce una di queste condizioni (non l'unica possibile).

DEFINIZIONE 12.5.2. Una famiglia $\mathcal{C} \subseteq \mathcal{P}(X)$ di sottoinsiemi di X si dice una **catena** se per ogni coppia $A, B \in \mathcal{C}$ vale $A \subseteq B$ oppure $B \subseteq A$.

Ad esempio, la famiglia di tutti gli intervalli chiusi del tipo $[-a, a]$, al variare di a tra i numeri reali positivi è una catena di sottoinsiemi di $\mathbb{R}$, mentre la famiglia di tutti gli intervalli chiusi $[a, b]$ non è una catena.

DEFINIZIONE 12.5.3. Una famiglia $\mathcal{F} \subseteq \mathcal{P}(X)$ di sottoinsiemi di X si dice **strettamente induttiva** se, per ogni catena non vuota $\mathcal{C} \subseteq \mathcal{F}$, l'unione di tutti gli elementi di $\mathcal{C}$ appartiene ad $\mathcal{F}$:

$$\bigcup_{A \in \mathcal{C}} A \ \in \mathcal{F}.$$

Ad esempio, la famiglia dei sottoinsiemi di $\mathbb{N}$ che non contengono potenze di 11 è strettamente induttiva, mentre la famiglia di tutti i sottoinsiemi finiti di $\mathbb{N}$ non è strettamente induttiva.

TEOREMA 12.5.4 (Principio del massimo di Hausdorff). *Siano X un insieme e $\mathcal{F} \subseteq \mathcal{P}(X)$ una famiglia non vuota strettamente induttiva. Allora la famiglia $\mathcal{F}$ possiede elementi massimali.*

La dimostrazione del principio del massimo, che può essere omessa ad una prima lettura, verrà data nella Sezione 13.4. Qui daremo invece un cenno di dimostrazione del Teorema 12.5.4 nel caso molto particolare $X = \mathbb{N}$, allo scopo di illustrare dove entra in gioco l'ipotesi che la famiglia sia strettamente induttiva.

Supponiamo quindi di avere una famiglia non vuota e strettamente induttiva $\mathcal{F} \subseteq \mathcal{P}(\mathbb{N})$. Fra tutti gli insiemi in $\mathcal{F}$ scegliamone uno, che indicheremo A_0, tale che $A_0 \cap \{0\}$ contiene il maggior numero di elementi. Successivamente, tra tutti gli insiemi in $\mathcal{F}$ che contengono A_0 scegliamone uno, che indicheremo A_1, tale che $A_1 \cap \{0, 1\}$ contiene il maggior numero di elementi. Poi si prosegue alla stessa maniera per ciascun intero maggiore di 1 arrivando a costruire ricorsivamente una successione di insiemi

$$A_0 \subseteq A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots,$$

in $\mathcal{F}$ dove A_n è scelto tra gli elementi di $\mathcal{F}$ che contengono A_{n-1} e con $A_n \cap \{0, 1, \dots, n\}$ più grande possibile.

Adesso la famiglia $\{A_i\}$ è chiaramente una catena in $\mathcal{F}$ e, per ipotesi, l'unione $B = \cup_{n=0}^{\infty} A_n$ appartiene ad $\mathcal{F}$; dimostriamo che B è un elemento massimale. Supponiamo per assurdo che esista $C \in \mathcal{F}$ che contiene strettamente B e indichiamo con N il minimo intero positivo dell'insieme non vuoto $C - B$. Ma allora si ha

$$A_{N-1} \subseteq A_N \subseteq B \subset C, \qquad N \in C \cap \{1, \dots, N\}, \qquad N \notin A_N \cap \{0, 1, \dots, N\},$$

in contraddizione con la scelta di A_N tra gli insiemi in $\mathcal{F}$ che contengono A_{N-1} e hanno intersezione massimale con $\{1, \dots, N\}$.

Vediamo adesso alcune applicazioni del principio del massimo all'algebra lineare.

TEOREMA 12.5.5 (semisemplicità degli spazi vettoriali). *Siano W uno spazio vettoriale e $V \subseteq W$ un sottospazio vettoriale. Allora esiste un sottospazio vettoriale $U \subseteq W$ tale che $W = V \oplus U$.*

DIMOSTRAZIONE. Consideriamo la famiglia $\mathcal{F} \subseteq \mathcal{P}(V)$ dei sottospazi vettoriali di W che hanno in comune con V il solo vettore nullo e dimostriamo che tale famiglia è non vuota e strettamente induttiva. Il sottospazio 0 appartiene a $\mathcal{F}$ per ovvi motivi. Se $\mathcal{C} \subseteq \mathcal{F}$ è una catena, consideriamo il sottoinsieme

$$M = \bigcup_{H \in \mathcal{C}} H, \qquad M \subseteq V.$$

Se $v_1, v_2 \in M$, per definizione esistono due sottospazi $H_1, H_2 \in C$ tali che $v_1 \in H_1, v_2 \in H_2$. Siccome $\mathcal{C}$ è una catena si ha $H_1 \subseteq H_2$ oppure $H_2 \subseteq H_1$. In ogni caso esiste un indice i tale che $v_1, v_2 \in H_i$ e di conseguenza $av_1 + bv_2 \in H_i \subseteq M$ per ogni $a, b \in \mathbb{K}$; questo prova che M è un sottospazio vettoriale di W. Dimostriamo adesso che $V \cap M = 0$; se $v \in V \cap M$ in particolare $v \in M$ ed esiste $H \in C$ tale che $v \in H$ e siccome $V \cap H = 0$ si ha $v = 0$.

Dunque $M \in \mathcal{F}$ e per il principio del massimo la famiglia $\mathcal{F}$ possiede un elemento massimale U; per ipotesi $V \cap U = 0$ e per provare che $V \oplus U = W$ basta quindi dimostrare che $V + U = W$.

Supponiamo per assurdo che $V + U \neq W$, scelto un vettore $w \notin V + U$ consideriamo il sottospazio $S = U \oplus \mathbb{K}w$; ogni vettore di $V \cap S$ è del tipo $v = u + aw$, con $v \in V$, $u \in U$ e $a \in \mathbb{K}$. Siccome $w \notin V + U$, dalla relazione $v - u = aw$ deduciamo che $a = 0$ e $u = v \in V \cap U$. Siccome $U \cap V = 0$ si ha $v = u = 0$ e questo prova che $V \cap S = 0$. Il sottospazio S contraddice la massimalità di U e quindi l'ipotesi $V + U \neq W$. □

COROLLARIO 12.5.6. *Sia $f\colon V \to W$ un'applicazione lineare. Allora f è iniettiva se e solo se la sua trasposta $f^*\colon W^\vee \to V^\vee$ è surgettiva. In particolare, per ogni spazio vettoriale W e per ogni vettore $w \neq 0$ esiste $\beta \in W^\vee$ tale che $\beta(w) = 1$.*

DIMOSTRAZIONE. Supponiamo f iniettiva, per il Teorema 12.5.5 esiste un sottospazio vettoriale $U \subset W$ tale che $W = f(V) \oplus U$. Dato un qualsiasi funzionale lineare $\alpha\colon V \to \mathbb{K}$ basta considerare il funzionale lineare $\beta\colon W \to \mathbb{K}$ definito ponendo

$$\beta(w) = 0 \text{ se } w \in U, \quad \beta(w) = \alpha(f^{-1}(w)) \text{ se } w \in f(V),$$

ed osservare che $f^*(\beta) = \alpha$. Se $w \in W - \{0\}$ l'applicazione $\mathbf{w}\colon \mathbb{K} \to W$, $\mathbf{w}(t) = tw$, è iniettiva e abbiamo appena dimostrato che esiste $\beta \in W^\vee$ tale che $\mathbf{w}^*\beta(1) = \beta(w) = 1$.

Supponiamo adesso f non iniettiva e sia v un vettore non nullo del nucleo di f; abbiamo appena dimostrato che esiste $\alpha \in V^\vee$ tale che $\alpha(v) = 1$. D'altra parte, per ogni $\beta \in W^\vee$ si ha $f^*\beta(v) = \beta(f(v)) = \beta(0) = 0$ e questo implica in particolare che f^* non è surgettiva. □

TEOREMA 12.5.7 (esistenza di basi non ordinate). *Dato un qualsiasi spazio vettoriale V esiste un sottoinsieme $B \subseteq V$ tale che:*

1. *ogni sottoinsieme finito e non vuoto di B è formato da vettori linearmente indipendenti;*
2. *ogni vettore di V è combinazione lineare di un numero finito di vettori di B.*

DIMOSTRAZIONE. Consideriamo la famiglia $\mathcal{F} \subseteq \mathcal{P}(V)$ definita tramite la seguente proprietà: un sottoinsieme $A \subseteq V$ appartiene ad $\mathcal{F}$ se e solo se ogni sottoinsieme finito e non vuoto di A è formato da vettori linearmente indipendenti.

La famiglia $\mathcal{F}$ non è vuota in quanto contiene il sottoinsieme vuoto. Vogliamo adesso dimostrare che se $\mathcal{C} \subseteq \mathcal{F}$ è una catena, per ogni successione finita di vettori distinti

$$v_1, \ldots, v_n \in H := \bigcup_{A \in \mathcal{C}} A$$

si ha che $v_1, \ldots, v_n$ sono linearmente indipendenti. Per definizione di H, esiste una successione $A_1, \ldots, A_n \in \mathcal{C}$ tale che $v_i \in A_i$ per ogni i. Siccome $\mathcal{C}$ è una catena, a meno di permutare gli indici possiamo supporre

$$A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n$$

e quindi $\{v_1, \ldots, v_n\}$ è un sottoinsieme finito di A_n e quindi tali vettori sono linearmente indipendenti.

Per il principio del massimo esiste un sottoinsieme massimale $B \in \mathcal{F}$. Mostriamo che per ogni vettore $v \in V$ esiste una successione finita $u_1, \ldots, u_n \in B$ ed una combinazione lineare

$$v = a_1u_1 + \cdots + a_nu_n. \tag{12.1}$$

Se $v \in B$ il fatto è ovvio; se invece $v \notin B$, per la condizione di massimalità $B \cup \{v\} \notin \mathcal{F}$ e dunque esiste un sottoinsieme finito $\{u_1, \dots, u_n\} \subseteq B$ tale che i vettori $v, u_1, \dots, u_n$ sono linearmente dipendenti, ossia esiste una combinazione lineare non nulla:

$$0 = \gamma v + b_1 u_1 + \cdots + b_n u_n.$$

Poiché $u_1, \dots, u_n$ sono linearmente indipendenti si deve avere $\gamma \neq 0$ e di conseguenza

$$v = \frac{-1}{\gamma}(b_1 u_1 + \cdots + b_n u_n).$$

□

Concludiamo la sezione dimostrando che per ogni spazio vettoriale V di dimensione infinita, il morfismo $\imath\colon V \to V^{\vee\vee}$ non è mai surgettivo.

Se A è un qualsiasi sottoinsieme di uno spazio vettoriale si definisce $\mathrm{Span}(A)$ come l'insieme di tutte le combinazioni lineari finite di elementi di A. In altri termini, $\mathrm{Span}(A)$ è l'unione dei sottospazi $\mathrm{Span}(A')$ al variare di A' tra tutti i sottoinsiemi finiti di A.

Corollario 12.5.8. *Sia V uno spazio vettoriale di dimensione infinita, allora l'applicazione naturale $\imath\colon V \to V^{\vee\vee}$ è iniettiva ma non è surgettiva.*

Dimostrazione. Diamo solamente una traccia di dimostrazione, lasciando come esercizio per il lettore il completamento di tutti i dettagli. Sia $v \in V$ un vettore non nullo, per il Corollario 12.5.6 esiste $f\colon V \to \mathbb{K}$ tale che $f(v) = \imath(v)(f) = 1$; abbiamo quindi dimostrato che il nucleo di $\imath$ contiene il solo vettore nullo.

Per il Teorema 12.5.7 esiste un sottoinsieme $B \subseteq V$ tale che:

(1) ogni sottoinsieme finito di B è formato da vettori linearmente indipendenti;
(2) ogni vettore di V è combinazione lineare di un numero finito di vettori in B.

Osserviamo adesso che ogni applicazione di insiemi $f\colon B \to \mathbb{K}$ si estende ad un'unica applicazione lineare $f\colon V \to \mathbb{K}$. Infatti ogni vettore $v \in V$ si scrive in maniera unica nella forma

$$v = a_1 u_1 + \cdots + a_n u_n, \qquad a_i \in \mathbb{K}, \quad u_1, \dots, u_n \in B \text{ distinti},$$

e basta definire $f(v) = a_1 f(u_1) + \cdots + a_n f(u_n)$. Consideriamo adesso, per ogni $u \in B$ il funzionale lineare $\tilde{u}\colon V \to \mathbb{K}$ definito dalle condizioni

$$\tilde{u}(u) = 1, \qquad \tilde{u}(w) = 0, \qquad w \in B,\ w \neq u.$$

Se indichiamo con $H \subseteq V^\vee$ il sottospazio vettoriale generato dai funzionali $\tilde{u}$, allora, poiché B è infinito si ha $H \neq V^\vee$, in quanto non appartiene ad H il funzionale $\eta\colon V \to \mathbb{K}$ definito dalla condizione $\eta(u) = 1$ per ogni $u \in B$. Per la semisemplicità dello spazio vettoriale $V^\vee$ possiamo trovare un sottospazio $K \neq 0$ tale che $H \oplus K = V^\vee$ e, siccome $K \to K^{\vee\vee}$ è iniettiva, si ha $K^\vee \neq 0$ ed esiste un funzionale non nullo $f\colon K \to \mathbb{K}$. Se $\pi\colon V^\vee = H \oplus K \to K$ è la proiezione sul secondo fattore, è immediato vedere che il funzionale non nullo $f\pi \in V^{\vee\vee}$ si annulla in H e non appartiene all'immagine di $\imath\colon V \to V^{\vee\vee}$. □

In certe situazioni è comodo avere formulazioni equivalenti del principio del massimo: in letterature se ne trovano molte ne esistono molte ed in queste note ne vedremo due: il lemma di Tukey ed il lemma di Zorn.

Definizione 12.5.9. Una famiglia $\mathcal{F} \subseteq \mathcal{P}(X)$ di sottoinsiemi di X si dice di **carattere finito** se vale la proprietà che un sottoinsieme $A \subseteq X$ appartiene a $\mathcal{F}$ se e solo se ogni sottoinsieme finito di A appartiene a $\mathcal{F}$.

Ad esempio, se $f\colon X \to \mathbb{R}$ è un'applicazione surgettiva di insiemi, la famiglia dei sottoinsiemi Y tali che la restrizione $f_{|Y}\colon Y \to \mathbb{R}$ è iniettiva è di carattere finito, mentre la famiglia dei sottoinsiemi Z tali che la restrizione $f_{|Z}\colon Z \to \mathbb{R}$ è surgettiva non è di carattere finito.

Corollario 12.5.10 (Lemma di Tukey). *Siano X un insieme e $\mathcal{F} \subseteq \mathcal{P}(X)$ una famiglia non vuota di carattere finito. Allora la famiglia $\mathcal{F}$ è strettamente induttiva e possiede elementi massimali.*

DIMOSTRAZIONE. Siano $\mathcal{F} \subseteq \mathcal{P}(X)$ una famiglia di carattere finito e $\mathcal{C} \subseteq \mathcal{F}$ una catena. Per ogni successione finita di elementi

$$x_1, \dots, x_n \in H := \bigcup_{A \in \mathcal{C}} A$$

esiste una sucecssione $A_1, \dots, A_n \in \mathcal{C}$ tale che $x_i \in A_i$ per ogni i. Siccome $\mathcal{C}$ è una catena, a meno di permutare gli indici possiamo supporre

$$A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n,$$

quindi $\{x_1, \dots, x_n\}$ è un sottoinsieme finito di A_n e, siccome $A_n \in \mathcal{F}$ ne segue che $\{x_1, \dots, x_n\} \in \mathcal{F}$. Abbiamo dunque dimostrato che ogni sottoinsieme finito di H appartiene alla famiglia di carattere finito $\mathcal{F}$ e quindi $H \in \mathcal{F}$. La dimostrazione del lemma di Tukey è quindi una diretta conseguenza del principio del massimo. □

Esercizi.

667. Dire se le seguenti famiglie di sottoinsiemi di $\mathbb{N}$ sono o meno strettamemte induttive:
(1) famiglia dei sottoinsiemi $Z \subseteq \mathbb{N}$ tali che l'applicazione $Z \mapsto \mathbb{N}$, $n \mapsto n^3 - n$, è iniettiva;
(2) famiglia dei sottoinsiemi $Z \subseteq \mathbb{N}$ tali che l'applicazione $Z \mapsto \mathbb{N}$, $n \mapsto n^3 - n$, non è surgettiva;
(3) famiglia dei sottoinsiemi $Z \subseteq \mathbb{N}$ che non contengono numeri primi;
(4) famiglia dei sottoinsiemi $Z \subseteq \mathbb{N}$ che contengono al più un numero finito di potenze di 2;
(5) famiglia dei sottoinsiemi $Z \subseteq \mathbb{N}$ chiusi per la somma, ossia tali che se $a, b \in Z$ allora $a + b \in Z$;
(6) famiglia dei sottoinsiemi $Z \subseteq \mathbb{N}$ chiusi per la somma che non contengono il numero 37.

668. Indichiamo poi con $\mathcal{P}(X)^0$ **l'insieme delle parti finite di** X, ossia $\mathcal{P}(X)^0 \subseteq \mathcal{P}(X)$ ed un sottoinsieme $A \subseteq X$ appartiene a $\mathcal{P}(X)^0$ se e solo se A contiene un numero finito di elementi.

Siano X un insieme e $\mathcal{F}^0$ un sottoinsieme di $\mathcal{P}(X)^0$ con le proprietà che $\emptyset \in \mathcal{F}^0$ e, se $A \in \mathcal{F}^0$ e $B \subset A$, allora anche $B \in \mathcal{F}^0$. Definiamo $\mathcal{F} \subseteq \mathcal{P}(X)$ nel modo seguente: per un sottoinsieme $A \subseteq X$ vale $A \in \mathcal{F}$ se e solo se ogni sottoinsieme finito di A appartiene a $\mathcal{F}^0$. Provare che $\mathcal{F}$ è una famiglia strettamente induttiva.

669. Siano dati uno spazio vettoriale V ed un sottoinsieme $A \subseteq V$ tale che ogni sottoinsieme finito di A è formato da vettori linearmente indipendenti. Provare che esiste un sottoinsieme $B \subseteq V$ con le seguenti proprietà:
(1) $A \cap B = \emptyset$;
(2) ogni sottoinsieme finito di $A \cup B$ è formato da vettori linearmente indipendenti;
(3) ogni vettore di V è combinazione lineare di un numero finito di vettori di $A \cup B$.

670. Siano V lo spazio vettoriale delle funzioni reali continue definite nell'intervallo aperto $]-1, 1[$ e sia $A \subset V$ l'insieme dalle funzioni

$$f_a \colon]-1, 1[\to \mathbb{R}, \qquad f_a(t) = \frac{1}{1 - at}, \qquad a \in [-1, 1].$$

Provare che ogni sottoinsieme finito di A è formato da vettori linearmente indipendenti.

671 (Teorema di scambio, ☕). Per semplicità notazionale, diremo che un sottoinsieme A di uno spazio vettoriale è linearmente indipendente se ogni suo sottoinsieme finito è formato da vettori linearmente indipendenti.

Sia B un insieme di generatori di uno spazio vettoriale V e sia $A \subseteq V$ un sottoinsieme non vuoto di vettori linearmente indipendenti. Si denoti con $p \colon A \times B \to A$ e $q \colon A \times B \to B$ le proiezioni sui fattori e si consideri la famiglia $\mathcal{F}$ formata da tutti i sottoinsiemi $C \subseteq A \times B$ che godono delle seguenti proprietà:
(1) le restrizioni $p \colon C \to A$ e $q \colon C \to B$ sono entrambe iniettive;
(2) $(A - C) \cup q(C)$ è un insieme linearmente indipendente.

Provare che $\mathcal{F}$ è strettamente induttiva. Applicare il principio del massimo per dimostrare che esiste un'applicazione iniettiva $f\colon A \to B$ la cui immagine $f(A)$ è ancora formata da vettori linearmente indipendenti.

672 (☕). Siano V, W due spazi vettoriali, $A = \emptyset$ e $B \subseteq V$ un sottoinsieme che soddisfa le condizioni del Teorema 12.5.7. Provare che per ogni applicazione di insiemi $h\colon B \to W$ vi è un'unica applicazione lineare $f\colon V \to W$ tale che $f(v) = h(v)$ per ogni $v \in B$. Dedurre che $\mathrm{Hom}(V, W)$ è uno spazio vettoriale di dimensione finita se e solo se V e W sono entrambi di dimensione finita.

673 (☕). Usare il risultato dell'Esercizio 257 e l'esistenza di basi non ordinate per provare che $\mathbb{R}$ ed $\mathbb{R}^2$ sono isomorfi come spazi vettoriali su $\mathbb{Q}$.

674. Dimostrare che il lemma di Tukey implica l'assioma della scelta. Se $f\colon X \to Y$ è surgettiva considerare la famiglia $\mathcal{F} \subseteq \mathcal{P}(X)$ dei sottoinsiemi $A \subseteq X$ tali che la restrizione $f\colon A \to Y$ è iniettiva.

675. Siano X, Y insiemi qualsiasi e indichiamo con $p\colon X \times Y \to X$ e $q\colon X \times Y \to Y$ le proiezioni. Sia $\mathcal{F} \subseteq \mathcal{P}(X \times Y)$ la famiglia dei sottoinsiemi $A \subseteq X \times Y$ tali che le applicazioni

$$p\colon A \to X, \qquad q\colon A \to Y,$$

sono entrambe iniettive. Provare che:

(1) La famiglia $\mathcal{F}$ è strettamente induttiva e di carattere finito.
(2) Se $M \in \mathcal{F}$ è un elemento massimale, allora almeno una delle due proiezioni

$$p\colon M \to X, \qquad q\colon M \to Y,$$

è bigettiva.
(3) O esiste un'applicazione iniettiva $X \to Y$ oppure esiste un'applicazione iniettiva $Y \to X$.

676. Sia X un insieme infinito. Dimostrare che esiste un sottoinsieme infinito $A \subseteq X$ ed un'applicazione iniettiva $A \to \mathbb{N}$.

12.6. Complementi: forme alternanti

Una delle più importanti e naturali generalizzazioni del duale di uno spazio vettoriale V sul campo $\mathbb{K}$ è data dagli spazi di forme multilineari ed alternanti. Ricordiamo dalla Definizione 8.1.7 che un'applicazione

$$V \times \cdots \times V \ \to \mathbb{K}$$

si dice multilineare se è separatamente lineare in ciascuna variabile.

Nel contesto delle applicazioni multilineari è utile introdurre la seguente notazione: se $a_1, \ldots, a_n$ è una successione di elementi in un insieme e vogliamo passare ad una sottosuccessione, si utilizza il simbolo $\widehat{\ }$ per indicare gli elementi esclusi, ossia

$$a_1, \ldots, \widehat{a_i}, \ldots, a_n = a_1, \ldots, a_{i-1}, a_{i+1}, , \ldots, a_n.$$

Si ha quindi, ad esempio:

$$a_1, \widehat{a_2}, a_3, a_4 = a_1, a_3, a_4\,, \qquad a_1, \widehat{a_2}, a_3, \widehat{a_4}, a_5, a_6 = a_1, a_3, a_5, a_6.$$

Definizione 12.6.1. Sia V uno spazio vettoriale sul campo $\mathbb{K}$, un'applicazione multilineare

$$\omega\colon \underbrace{V \times \cdots \times V}_{p \text{ fattori}} \to \mathbb{K}$$

si dice una **p-forma alternante** se $\omega(v_1, \ldots, v_p) = 0$ ogni volta che $v_i = v_{i+1}$ per qualche indice $i < p$.

Il concetto non è del tutto nuovo in quanto abbiamo ampia esperienza della funzione determinante $\det\colon M_{n,n}(\mathbb{K}) \to \mathbb{K}$, interpretata come n-forma alternante sullo spazio dei vettori colonna.

L'insieme di tutte le p-forme alternanti

$$\omega\colon \underbrace{V \times \cdots \times V}_{p \text{ fattori}} \to \mathbb{K},$$

dotato delle naturali operazioni di somma e prodotto per scalare, è uno spazio vettoriale che denoteremo $\Omega^p(V)$. In particolare si ha $\Omega^1(V) = V^\vee$ ed è utile porre per convenzione $\Omega^0(V) = \mathbb{K}$ e $\Omega^q(V) = 0$ per ogni $q < 0$.

Le stesse identiche considerazioni fatte nel Lemma 8.2.4 mostrano che se $\omega \in \Omega^p(V)$ e $v_1, \dots, v_p \in V$ allora

$$\omega(v_{\sigma(1)}, \dots, v_{\sigma(p)}) = (-1)^\sigma \omega(v_1, \dots, v_p)$$

per ogni permutazione σ e, se i vettori $v_1, \dots, v_p$ sono linearmente dipendenti, allora $\omega(v_1, \dots, v_p) = 0$.

Nello studio delle forme alternanti risulterà fondamentale l'introduzione di due prodotti:

$$\begin{aligned} \Omega^p(V) \times \Omega^q(V) &\xrightarrow{\wedge} \Omega^{p+q}(V) \quad \text{prodotto wedge o esterno,} \\ V \times \Omega^p(V) &\xrightarrow{\lrcorner} \Omega^{p-1}(V) \quad \text{prodotto di contrazione o interno.} \end{aligned}$$

Il prodotto interno è quello più facile da definire: dati infatti $v \in V$ e $\omega \in \Omega^p(V)$ si definisce $v \lrcorner \omega \in \Omega^{p-1}(V)$ mediante la formula

$$v \lrcorner \omega(u_1, \dots, u_{p-1}) = \omega(v, u_1, \dots, u_{p-1}).$$

Per $p = 1$ ritroviamo il solito accoppiamento di dualità.

LEMMA 12.6.2. *Siano V spazio vettoriale ed $\omega \in \Omega^p(V)$, $p > 0$:*

(1) *vale $\omega = 0$ se e soltanto se $v \lrcorner \omega = 0$ per ogni $v \in V$,*
(2) *$v \lrcorner (v \lrcorner \omega) = 0$ per ogni $v \in V$,*
(3) *$v_1 \lrcorner (v_2 \lrcorner \omega) = -v_2 \lrcorner (v_1 \lrcorner \omega)$ per ogni $v_1, v_2 \in V$.*

DIMOSTRAZIONE. Evidente. □

Il prodotto esterno è un po' più complicato da definire e richiede l'introduzione delle **permutazioni shuffle**: diremo che una permutazione

$$\sigma\colon \{1, \dots, n\} \to \{1, \dots, n\}$$

è uno shuffle di tipo $(p, n-p)$, con $0 \le p \le n$, se $\sigma(i) < \sigma(i+1)$ per ogni $i \neq p$, ossia se

$$\sigma(1) < \sigma(2) < \cdots < \sigma(p), \qquad \sigma(p+1) < \sigma(p+2) < \cdots < \sigma(n).$$

Indicheremo con $S(p, n-p) \subseteq \Sigma_n$ l'insieme degli shuffle di tipo $(p, n-p)$; tale insieme contiene esattamente $\binom{n}{p}$ elementi, essendo chiaro che ogni $\sigma \in S(p, n-p)$ è univocamente determinato dal sottoinsieme $\{\sigma(1), \sigma(2), \dots, \sigma(p)\} \subseteq \{1, \dots, n\}$. Si noti che $S(0, n)$ e $S(n, 0)$ contengono solamente la permutazione identità.

Il prodotto esterno di due forme $\omega \in \Omega^p(V), \eta \in \Omega^q(V)$ si definisce mediante la formula

$$\omega \wedge \eta(u_1, \dots, u_{p+q}) = \sum_{\sigma \in S(p,q)} (-1)^\sigma \omega(u_{\sigma(1)}, \dots, u_{\sigma(p)}) \eta(u_{\sigma(p+1)}, \dots, u_{\sigma(p+q)}). \tag{12.2}$$

Per $p = 0$ ritroviamo il prodotto per scalare, mentre per $\phi, \psi \in \Omega^1(V) = V^\vee$ e $u, v \in V$ si ha

$$\phi \wedge \psi(u, v) = \phi(u)\psi(v) - \phi(v)\psi(u).$$

È chiaro che il prodotto esterno $\omega \wedge \eta$ è multilineare, mentre la prova che si tratta di una forma alternante richiede qualche considerazione non banale. Supponiamo, nella situazione della Formula (12.2) che $u_i = u_{i+1}$ per un indice fissato $0 < i < p + q$. Possiamo allora considerare gli insiemi

$$\begin{aligned} A &= \{\sigma \in S(p, q) \mid i \in \sigma(\{1, \dots, p\}), \quad i+1 \in \sigma(\{p+1, \dots, p+q\})\}, \\ B &= \{\sigma \in S(p, q) \mid i \in \sigma(\{p+1, \dots, p+q\}), \quad i+1 \in \sigma(\{1, \dots, p\})\}, \end{aligned}$$

e, poiché ω ed η sono entrambe forme alternanti, si può scrivere

$$\omega\wedge\eta(u_1,\dots,u_{p+q})=\sum_{\sigma\in A\cup B}(-1)^\sigma\omega(u_{\sigma(1)},\dots,u_{\sigma(p)})\eta(u_{\sigma(p+1)},\dots,u_{\sigma(p+q)}).$$

Se τ è la trasposizione che scambia i ed $i+1$ e $\sigma\in A$, allora $\tau\sigma\in B$; viceversa, se $\sigma\in B$, allora $\tau\sigma\in A$. Possiamo quindi scrivere

$$\begin{aligned}\omega\wedge\eta(u_1,\dots,u_{p+q})=&\sum_{\sigma\in A}(-1)^\sigma\omega(u_{\sigma(1)},\dots,u_{\sigma(p)})\eta(u_{\sigma(p+1)},\dots,u_{\sigma(p+q)})\\&+\sum_{\sigma\in A}(-1)^{\tau\sigma}\omega(u_{\tau\sigma(1)},\dots,u_{\tau\sigma(p)})\eta(u_{\tau\sigma(p+1)},\dots,u_{\tau\sigma(p+q)}).\end{aligned}$$

Per ipotesi $u_i=u_{i+1}$ e quindi per ogni $\sigma\in A$ si ha $u_{\sigma(h)}=u_{\tau\sigma(h)}$ per ogni h. Basta adesso osservare che $(-1)^{\tau\sigma}=-(-1)^\sigma$ per concludere che $\omega\wedge\eta(u_1,\dots,u_{p+q})=0$.

Sia il prodotto esterno che quello interno sono lineari in entrambe le variabili, e cioè valgono le relazioni:

(1) $(a\omega_1+b\omega_2)\wedge\eta=a\omega_1\wedge\eta+b\omega_2\wedge\eta$,
(2) $\omega\wedge(a\eta_1+b\eta_2)=a\omega\wedge\eta_1+b\omega\wedge\eta_2$,
(3) $(av_1+bv_2)\lrcorner\eta=av_1\lrcorner\eta+bv_2\lrcorner\eta$,
(4) $v\lrcorner(a\eta_1+b\eta_2)=av\lrcorner\eta_1+bv\lrcorner\eta_2$.

Le dimostrazioni sono del tutto elementari e lasciate per esercizio al lettore.

TEOREMA 12.6.3. *Il prodotto esterno di forme alternanti gode delle seguenti proprietà:*

(1) *(Formula di Leibniz) per ogni $\omega\in\Omega^p(V)$, $\eta\in\Omega^q(V)$ e $v\in V$ si ha*
$$v\lrcorner(\omega\wedge\eta)=(v\lrcorner\omega)\wedge\eta+(-1)^p\omega\wedge(v\lrcorner\eta);$$
(2) *(Regola dei segni di Koszul[2]) per ogni $\omega\in\Omega^p(V)$ e $\eta\in\Omega^q(V)$ si ha*
$$\omega\wedge\eta=(-1)^{pq}\eta\wedge\omega;$$
(3) *(Associatività) per ogni terna di forme alternanti ω,η,μ si ha*
$$(\omega\wedge\eta)\wedge\mu=\omega\wedge(\eta\wedge\mu);$$
(4) *per ogni $\phi_1,\dots,\phi_n\in V^\vee$ ed ogni $v_1,\dots,v_n\in V$ si ha*
$$(\phi_1\wedge\cdots\wedge\phi_n)(v_1,\dots,v_n)=\det(\phi_i(v_j)).$$

DIMOSTRAZIONE. 1) Dati $u_1,\dots,u_{p+q-1}\in V$ e due permutazioni shuffle $\sigma\in S(p-1,q)$, $\tau\in S(p,q-1)$ osserviamo che la segnatura della permutazione

$$(v,u_1,\dots,u_{p+q})\mapsto(v,u_{\sigma(1)},\dots,u_{\sigma(p+q-1)})$$

è uguale alla segnatura di σ, mentre la segnatura di

$$(v,u_1,\dots,u_{p+q})\mapsto(u_{\tau(1)},\dots,u_{\tau(p)},v,u_{\tau(p+1)},\dots,u_{\tau(p+q-1)})$$

è uguale alla segnatura di τ moltiplicata per $(-1)^p$. Abbiamo quindi

$$\begin{aligned}v\lrcorner(\omega\wedge\eta)(u_1,\dots,u_{p+q-1})=&\\=&\omega\wedge\eta(v,u_1,\dots,u_{p+q-1})\\=&\sum_{\sigma\in S(p-1,q)}(-1)^\sigma\omega(v,u_{\sigma(1)},\dots,u_{\sigma(p-1)})\eta(u_{\sigma(p)},\dots,u_{\sigma(p+q-1)})\\&+(-1)^p\sum_{\tau\in S(p,q-1)}(-1)^\tau\omega(u_{\tau(1)},\dots,u_{\tau(p)})\eta(v,u_{\tau(p+1)},\dots,u_{\tau(p+q-1)})\\=&(v\lrcorner\omega)\wedge\eta(u_1,\dots,u_{p+q-1})+(-1)^p\omega\wedge(v\lrcorner\eta)(u_1,\dots,u_{p+q-1}).\end{aligned}$$

2) È possibile dimostrare la regola dei segni di Koszul con ragionamenti sulla segnatura simili al punto precedente. Alternativamente possiamo utilizzare la formula di Leibniz assieme

[2]Jean-Louis Koszul (leggasi kózul), matematico francese nato il 3 gennaio 1921, principalmente noto per il *complesso di Koszul*, ossia per la generalizzazione all'algebra commutativa del complesso di spazi vettoriali

$$\cdots\to\Omega^4(V)\xrightarrow{v\lrcorner}\Omega^3(V)\xrightarrow{v\lrcorner}\Omega^2(V)\xrightarrow{v\lrcorner}\Omega^1(V)\xrightarrow{v\lrcorner}\Omega^0(V),\qquad v\in V.$$

ad un procedimento per induzione su $p+q$. La regola dei segni è banalmente vera quando $p+q=0$; se $p+q>0$, per dimostrare $\omega\wedge\eta=(-1)^{pq}\eta\wedge\omega$ basta provare che

$$v\lrcorner(\omega\wedge\eta)=(-1)^{pq}v\lrcorner(\eta\wedge\omega)$$

per ogni vettore $v\in V$. Per la formula di Leibniz

$$v\lrcorner(\omega\wedge\eta)=(v\lrcorner\omega)\wedge\eta+(-1)^p\omega\wedge(v\lrcorner\eta)$$

e per l'ipotesi induttiva

$$\begin{aligned}v\lrcorner(\omega\wedge\eta)&=(-1)^{(p-1)q}\eta\wedge(v\lrcorner\omega)+(-1)^{p+p(q-1)}(v\lrcorner\eta)\wedge\omega\\&=(-1)^{pq}((-1)^q\eta\wedge(v\lrcorner\omega)+(v\lrcorner\eta)\wedge\omega)\\&=(-1)^{pq}v\lrcorner(\eta\wedge\omega).\end{aligned}$$

3) Come per il punto 2, date le forme alternanti $\omega\in\Omega^n(V),\eta\in\Omega^m(V),\mu\in\Omega^p(V)$ dimostriamo la formula

$$(\omega\wedge\eta)\wedge\mu=\omega\wedge(\eta\wedge\mu)$$

per induzione su $n+m+p$ usando la formula di Leibniz. Per ogni $v\in V$ si ha

$$\begin{aligned}&v\lrcorner((\omega\wedge\eta)\wedge\mu)=(v\lrcorner(\omega\wedge\eta))\wedge\mu+(-1)^{n+m}(\omega\wedge\eta)\wedge(v\lrcorner\mu)=\\&=((v\lrcorner\omega)\wedge\eta)\wedge\mu+(-1)^n(\omega\wedge(v\lrcorner\eta))\wedge\mu+(-1)^{n+m}(\omega\wedge\eta)\wedge(v\lrcorner\mu)\end{aligned}$$

che per l'ipotesi induttiva è uguale a

$$\begin{aligned}&(v\lrcorner\omega)\wedge(\eta\wedge\mu)+(-1)^n\omega\wedge((v\lrcorner\eta)\wedge\mu)+(-1)^{n+m}\omega\wedge(\eta\wedge(v\lrcorner\mu))=\\&=(v\lrcorner\omega)\wedge(\eta\wedge\mu)+(-1)^n\omega\wedge(v\lrcorner(\eta\wedge\mu))=v\lrcorner(\omega\wedge(\eta\wedge\mu)).\end{aligned}$$

4) Il risultato è certamente vero per $n=1$. Per $n>1$, dati $\phi_1,\dots,\phi_n\in V^\vee$ e $v_1,\dots,v_n\in V$ consideriamo la matrice $A=(\phi_i(v_j))$, $i,j=1,\dots,n$. Per induzione su n si hanno le uguaglianze

$$(\phi_1\wedge\cdots\widehat{\phi_i}\cdots\wedge\phi_n)(v_2,\dots,v_n)=\det(A_{i1})$$

dove, come al solito, A_{i1} è la sottomatrice di A ottenuta cancellando la riga i e la prima colonna. Dalla formula di Leibniz si ottiene

$$\begin{aligned}(\phi_1\wedge\cdots\wedge\phi_n)(v_1,\dots,v_n)&=v_1\lrcorner(\phi_1\wedge\cdots\wedge\phi_n)(v_2,\dots,v_n)\\&=\sum_{i=1}^n(-1)^{i+1}\phi_i(v_1)(\phi_1\wedge\cdots\widehat{\phi_i}\cdots\wedge\phi_n)(v_2,\dots,v_n)\\&=\sum_{i=1}^n(-1)^{i+1}\phi_i(v_1)\det(A_{i1})\end{aligned}$$

e la conclusione segue dallo sviluppo di Laplace rispetto alla prima colonna. □

COROLLARIO 12.6.4. *Siano V uno spazio vettoriale e $v\in V$, $\phi\in V^\vee=\Omega^1(V)$ tali che $v\lrcorner\phi=\phi(v)=1$. Allora ogni forma alternante $\omega\in\Omega^p(V)$ si scrive in modo unico come*

$$\omega=\phi\wedge\omega_1+\omega_2$$

dove

$$\omega_1\in\Omega^{p-1}(V),\quad\omega_2\in\Omega^p(V),\qquad v\lrcorner\omega_1=v\lrcorner\omega_2=0.$$

Inoltre:

(1) *se $u\lrcorner\phi=u\lrcorner\omega=0$ per un dato vettore $u\in V$, allora $u\lrcorner\omega_1=u\lrcorner\omega_2=0$;*
(2) *se $v\lrcorner\psi=\psi\wedge\omega=0$ per un dato funzionale $\psi\in V^\vee$, allora $\psi\wedge\omega_1=\psi\wedge\omega_2=0$.*

DIMOSTRAZIONE. Per l'esistenza della decomposizione basta considerare le forme

$$\omega_1=v\lrcorner\omega,\qquad\omega_2=\omega-\phi\wedge\omega_1,$$

che soddisfano le condizioni richieste in quanto

$$v\lrcorner\omega_1=v\lrcorner(v\lrcorner\omega)=0,\qquad v\lrcorner(\phi\wedge\omega_1)=(v\lrcorner\phi)\wedge\omega_1=\omega_1=v\lrcorner\omega.$$

Per l'unicità, se $\phi\wedge\omega_1+\omega_2=\phi\wedge\eta_1+\eta_2$, con $v\lrcorner\omega_1=v\lrcorner\omega_2=v\lrcorner\eta_1=v\lrcorner\eta_2=0$, allora

$$v\lrcorner(\omega_1-\eta_1)=0,\qquad\omega_1-\eta_1=v\lrcorner(\phi\wedge(\omega_1-\eta_1))=v\lrcorner(\eta_2-\omega_2)=0$$

e di conseguenza

$$\eta_2-\omega_2=\phi\wedge(\omega_1-\eta_1)=0.$$

Supponiamo adesso $u \lrcorner \phi = u \lrcorner \omega = 0$, allora

$$0 = u \lrcorner \omega = \phi \wedge (-u \lrcorner \omega_1) + u \lrcorner \omega_2.$$

D'altra parte

$$v \lrcorner (-u \lrcorner \omega_1) = u \lrcorner (v \lrcorner \omega_1) = 0, \qquad v \lrcorner (u \lrcorner \omega_2) = -u \lrcorner (v \lrcorner \omega_2) = 0,$$

e per l'unicità della decomposizione si ha $u \lrcorner \omega_1 = u \lrcorner \omega_2 = 0$.

Nel caso in cui $v \lrcorner \psi = \psi \wedge \omega = 0$, si ha

$$0 = \omega \wedge \psi = \phi \wedge (\omega_1 \wedge \psi) + \omega_2 \wedge \psi$$

e dal fatto che $v \lrcorner (\omega_1 \wedge \psi) = v \lrcorner (\omega_2 \wedge \psi) = 0$ segue per unicità che $\omega_1 \wedge \psi = \omega_2 \wedge \psi = 0$ □

Il precedente corollario consente di dimostrare elegantemente che se V ha dimensione finita n, allora lo spazio $\Omega^p(V)$ ha dimensione finita $\binom{n}{p}$.

TEOREMA 12.6.5. *Sia $v_1, \dots, v_n$ una base dello spazio vettoriale V e sia $\phi_1, \dots, \phi_n \in V^\vee$ la corrispondente base duale. Allora, per ogni $p > 0$ le forme alternanti*

$$\phi_{i_1} \wedge \phi_{i_2} \wedge \cdots \wedge \phi_{i_p}, \qquad 0 < i_1 < i_2 < \cdots < i_p \leq n,$$

formano una base di $\Omega^p(V)$.

DIMOSTRAZIONE. Dati gli indici $0 < j_1 < \cdots < j_p \leq n$, dalla formula

$$\phi_{i_1} \wedge \phi_{i_2} \wedge \cdots \wedge \phi_{i_p}(v_{j_1}, \dots, v_{j_p}) = \det(\phi_{i_h}(v_{j_k})) = \begin{cases} 1 & \text{se } i_h = j_h, \ \forall\, h, \\ 0 & \text{altrimenti,} \end{cases}$$

segue che tali forme sono linearmente indipendenti. Per dimostrare che generano, per ogni $k = 0, 1, \dots, n$ consideriamo gli spazi vettoriali

$$\Omega^p_k = \{\omega \in \Omega^p(V) \mid v_i \lrcorner \omega = 0, \quad \forall\, i \leq k\}, \qquad p > 0,$$

e dimostriamo per induzione su $n - k$ che le forme

$$\phi_{i_1} \wedge \phi_{i_2} \wedge \cdots \wedge \phi_{i_p}, \qquad k < i_1 < i_2 < \cdots < i_p \leq n,$$

formano una base di Ω^p_k. Il passo iniziale $n = k$ consiste nel dimostrare che $\Omega^p_n = 0$ per ogni $p > 0$; data $\omega \in \Omega^p_n$ e $u_1, \dots, u_p \in V$ si può scrivere $u_1 = \sum a_i v_i$ e quindi

$$\begin{aligned} \omega(u_1, \dots, u_p) &= \omega(\sum a_i v_i, u_2, \dots, u_p) \\ &= \sum a_i \omega(v_i, u_2, \dots, u_p) = \sum a_i \, v_i \lrcorner \omega(u_2, \dots, u_p) = 0. \end{aligned}$$

Se $\omega \in \Omega^p_{k-1}$ con $0 < k \leq n$, siccome $v_k \lrcorner \phi_k = 1$ e $v_i \lrcorner \omega = v_i \lrcorner \phi_k = 0$ per ogni $i < k$, per il Corollario 12.6.4 si ha

$$\omega = \phi_k \wedge \omega_1 + \omega_2, \qquad \omega_1 \in \Omega^{p-1}_k, \quad \omega_2 \in \Omega^p_k,$$

e tutto segue dall'ipotesi induttiva. □

Una forma alternante $\omega \in \Omega^p(V)$ si dice **decomponibile** se esistono $\phi_1, \dots, \phi_p \in V^\vee$ tali che

$$\omega = \phi_1 \wedge \cdots \wedge \phi_p.$$

Segue dal Teorema 12.6.5 che in dimensione finita ogni forma alternante è combinazione lineare di un numero finito di forme decomponibili.

La nozione di applicazione trasposta si estende immediatamente alle forme alternanti. Se $f \colon V \to W$ è un'applicazione lineare, per ogni p si definisce l'applicazione $f^* \colon \Omega^p(W) \to \Omega^p(V)$ ponendo

$$f^*\omega(u_1, \dots, u_p) = \omega(f(u_1), \dots, f(u_p)), \qquad \omega \in \Omega^p(W), \quad u_1, \dots, u_p \in V.$$

Segue immediatamente dalle definizioni di prodotto esterno ed interno che valgono le formule

$$f^*(\omega \wedge \eta) = f^*\omega \wedge f^*\eta, \qquad v \lrcorner f^*\omega = f^*(f(v) \lrcorner \omega), \qquad \omega \in \Omega^p(W), \quad v \in V.$$

Viene talvolta detto che, in dimensione finita, il prodotto interno è il trasposto del prodotto esterno, intendendo con ciò quanto descritto nel seguente teorema.

TEOREMA 12.6.6. *Sia V uno spazio vettoriale di dimensione finita. Vi è un'unica successione di isomorfismi di spazi vettoriali $D_p\colon \Omega^p(V^\vee) \to \Omega^p(V)^\vee$, $p > 0$, tali che:*

(1) *per ogni $v \in V$ ed ogni $\eta \in V^\vee = \Omega^1(V)$ vale*
$$D_1(\imath(v))(\eta) = \eta(v) = v \lrcorner \eta,$$
dove $\imath\colon V \to V^{\vee\vee} = \Omega^1(V^\vee)$ indica l'isomorfismo canonico.

(2) *per ogni $p > 0$, ogni $\omega \in \Omega^p(V^\vee)$, $v \in V$, $\eta \in \Omega^{p+1}(V)$:*
$$D_{p+1}(\omega \wedge \imath(v))(\eta) = D_p(\omega)(v \lrcorner \eta).$$

DIMOSTRAZIONE. Siccome lo spazio vettoriale $\Omega^p(V^\vee)$ è generato dalle forme $\imath(v_1) \wedge \cdots \wedge \imath(v_p)$ al variare di $v_1, \dots, v_p \in V$, l'unicità è chiara e basta dimostrare esistenza e bigettività delle applicazioni D_p.

Scegliamo una base $e_1, \dots, e_n$ di V e indichiamo con $\phi_1, \dots, \phi_n \in V^\vee$ la corrispondente base duale. Per semplicità notazionale identifichiamo V con $V^{\vee\vee}$ tramite $\imath$, per cui $e_1, \dots, e_n$ diventa la base duale di $\phi_1, \dots, \phi_n$. Definiamo le applicazioni lineari $D_p\colon \Omega^p(V^\vee) \to \Omega^p(V)^\vee$ mediante la formula

$$D_p(\omega)(\eta) = \sum \omega(\phi_{i_1}, \dots, \phi_{i_p})\eta(e_{i_p}, \dots, e_{i_1}), \tag{12.3}$$

dove $\omega \in \Omega^p(V^\vee)$ e $\eta \in \Omega^p(V)$ e la sommatoria è fatta sull'insieme di tutte le successioni $0 < i_1 < i_2 < \cdots < i_p \le n$. Dato un vettore $v = \sum a_i e_i \in V$ ed una forma $\eta \in \Omega^1(V)$ si ha

$$D_1(\imath(v))(\eta) = \sum_{i=1}^n \imath(v)(\phi_i)\eta(e_i) = \sum_{i=1}^n \phi_i(v)\eta(e_i) = \sum_{i=1}^n a_i\eta(e_i) = \sum_{i=1}^n \eta(a_i e_i) = \eta(v).$$

Consideriamo adesso $\omega \in \Omega^{p-1}(V^\vee)$, $v = \sum a_i e_i \in V$, $\eta \in \Omega^p(V)$. Si ha:

$$\begin{aligned}
&\sum_{i_1<\cdots<i_p} (\omega \wedge \imath(v))(\phi_{i_1}, \dots, \phi_{i_p})\eta(e_{i_p}, \dots, e_{i_1}) \\
&\quad = \sum_{i_1<\cdots<i_p} \sum_{h=1}^p (-1)^{p-h}\omega(\dots, \widehat{\phi_{i_h}}, \dots)\phi_{i_h}(v)\eta(e_{i_p}, \dots, v_{e_1}) \\
&\quad = \sum_{i_1<\cdots<i_p} \sum_{h=1}^p (-1)^{p-h}\omega(\dots, \widehat{\phi_{i_h}}, \dots)a_{i_h}\eta(e_{i_p}, \dots, e_{i_1}) \\
&\quad = \sum_{i_1<\cdots<i_p} \sum_{h=1}^p \omega(\dots, \widehat{\phi_{i_h}}, \dots)a_{i_h}\eta(e_{i_h}, e_{i_p}, \dots, \widehat{e_{i_h}}, \dots, e_{i_1})
\end{aligned}$$

Dall'altro lato abbiamo

$$\begin{aligned}
&\sum_{j_1<\cdots<j_{p-1}} \omega(\phi_{j_1}, \dots, \phi_{j_{p-1}})\eta(v, e_{j_{p-1}}, \dots, e_{j_1}) \\
&\quad = \sum_{j_1<\cdots<j_{p-1}} \sum_{k=1}^n \omega(\phi_{j_1}, \dots, \phi_{j_{p-1}})a_k\eta(e_k, e_{j_{p-1}}, \dots, e_{j_1})
\end{aligned}$$

e la due espressioni coincidono poiché $\eta(e_k, e_{j_{p-1}}, \dots, e_{j_1}) = 0$ ogni volta che $k = j_s$ per qualche s. Abbiamo con ciò dimostrato l'esistenza delle applicazioni D_p, ed anche che l'espressione (12.3) è indipendente dalla scelta della base.

Sappiamo che due basi di $\Omega^p(V^\vee)$ e $\Omega^p(V)$ sono formate rispettivamente dalle forme decomponibili $e_{j_1} \wedge \cdots \wedge e_{j_p}$, $\phi_{j_p} \wedge \cdots \wedge \phi_{j_1}$ al variare di $0 < j_1 < \cdots < j_p \le n$. Si ha

$$D_p(e_{j_1} \wedge \cdots \wedge e_{j_p})(\eta) = D_{p-1}(e_{j_1} \wedge \cdots \wedge e_{j_{p-1}})(e_{j_p} \lrcorner \eta) = \cdots = \eta(e_{j_p}, \dots, e_{j_1}).$$

Ne consegue che per ogni successione $i_1 < \cdots < i_p$ si ha

$$D_p(e_{j_1} \wedge \cdots \wedge e_{j_p})(\phi_{i_p} \wedge \cdots \wedge \phi_{i_1}) = \begin{cases} 1 & \text{se } i_s = j_s\ \forall s, \\ 0 & \text{altrimenti.} \end{cases}$$

e quindi D_p trasforma basi in basi. □

Non è difficile dimostrare che se $\dim V \ge p+2 \ge 4$ esistono forme in $\Omega^p(V)$ che non sono decomponibili. Tra i criteri più noti per stabilire se una forma è decomponibile, quello delle relazioni quadratiche di Plücker è certamente il più celebre.

TEOREMA 12.6.7. *Dato uno spazio di dimensione finita V ed una forma $\omega \in \Omega^p(V)$ si considerino le due applicazioni lineari:*

$$e_\omega \colon V^\vee \to \Omega^{p+1}(V), \qquad e_\omega(\phi) = \phi \wedge \omega,$$
$$i_\omega \colon V \to \Omega^{p-1}(V), \qquad i_\omega(v) = v \lrcorner \omega.$$

Allora la forma ω è decomponibile se e solo se $e_\omega \circ i_\omega^* = 0$, viz. se e solo se si annulla l'applicazione composta

$$\Omega^{p-1}(V)^\vee \xrightarrow{i_\omega^*} V^\vee \xrightarrow{e_\omega} \Omega^{p+1}(V).$$

DIMOSTRAZIONE. Se $\omega = 0$ non c'è nulla da dimostrare. Iniziamo la dimostrazione provando che per ogni $\omega \neq 0$ vale l'inclusione

$$\operatorname{Ker}(e_\omega) \subseteq \operatorname{Ann}(\operatorname{Ker} i_\omega) = i_\omega^*(\Omega^{p-1}(V)^\vee).$$

Infatti, se $\phi \wedge \omega = 0$, allora per ogni $v \in \operatorname{Ker}(i_\omega)$ si ha $v \lrcorner \omega = 0$ e per Leibniz

$$0 = v \lrcorner (\phi \wedge \omega) = \phi(v) \wedge \omega \;\Rightarrow\; \phi \in \operatorname{Ann}(\operatorname{Ker} i_\omega).$$

D'altra parte la condizione $e_\omega \circ i_\omega^* = 0$ è del tutto equivalente a $i_\omega^*(\Omega^{p+1}(V)^\vee) \subseteq \operatorname{Ker}(e_\omega)$ e quindi il teorema si riconduce a dimostrare che $\omega \neq 0$ è decomponibile se e solo se $\operatorname{Ker}(e_\omega) = \operatorname{Ann}(\operatorname{Ker} i_\omega)$.

Se $\omega \neq 0$ è decomponibile vuol dire che esistono $\phi_1, \dots, \phi_p \in V^\vee$ linearmente indipendenti e tali che $\omega = \phi_1 \wedge \cdots \wedge \phi_p$. Estendiamo tali funzionali ad una base $\phi_1, \dots, \phi_n$ di $V^\vee$ e indichiamo con $v_1, \dots, v_n \in V$ la corrispondente base duale. Si verifica immediatamente che

$$\operatorname{Ker}(e_\omega) = \operatorname{Span}(\phi_1, \dots, \phi_p), \qquad \operatorname{Ker}(i_\omega) = \operatorname{Span}(v_{p+1}, \dots, v_n),$$

e di conseguenza che $\operatorname{Ker}(e_\omega) = \operatorname{Ann}(\operatorname{Ker} i_\omega)$.

Viceversa, se $\operatorname{Ker}(e_\omega) = \operatorname{Ann}(\operatorname{Ker} i_\omega)$ possiamo trovare una base $\phi_1, \dots, \phi_n$ di $V^\vee$, con base duale $v_1, \dots, v_n \in V$, tale che

$$\operatorname{Ker}(e_\omega) = \operatorname{Ann}(\operatorname{Ker} i_\omega) = \operatorname{Span}(\phi_1, \dots, \phi_s), \qquad \operatorname{Ker}(i_\omega) = \operatorname{Span}(v_{s+1}, \dots, v_n).$$

Per il Corollario 12.6.4, relativo alla coppia ϕ_s, v_s, possiamo scrivere

$$\omega = \phi_s \wedge \omega_1 + \omega_2,$$

in modo tale che

$$v_i \lrcorner \omega_1 = v_i \lrcorner \omega_2 = 0, \quad \text{per ogni} \quad i \geq s,$$
$$\phi_i \wedge \omega_1 = \phi_i \wedge \omega_2 = 0, \quad \text{per ogni} \quad i < s.$$

Inoltre $\phi_s \wedge \omega_2 = \phi_s \wedge \omega - \phi_s \wedge \phi_s \wedge \omega = 0$ e quindi si hanno le inclusioni:

$$\operatorname{Ann}(\operatorname{Ker} i_{\omega_1}) \subseteq \operatorname{Span}(\phi_1, \dots, \phi_{s-1}) \subseteq \operatorname{Ker}(e_{\omega_1}),$$
$$\operatorname{Ann}(\operatorname{Ker} i_{\omega_2}) \subseteq \operatorname{Span}(\phi_1, \dots, \phi_{s-1}) \subset \operatorname{Span}(\phi_1, \dots, \phi_s) \subseteq \operatorname{Ker}(e_{\omega_2}),$$

dalle quali si ricava $\omega_2 = 0$ e $\operatorname{Ann}(\operatorname{Ker} i_{\omega_1}) = \operatorname{Span}(\phi_1, \dots, \phi_{s-1}) = \operatorname{Ker}(e_{\omega_1})$. Basta adesso ragionare per induzione su p per dedurre che ω_1 e $\omega = \phi_s \wedge \omega_1$ sono decomponibili. □

COROLLARIO 12.6.8 (Relazioni quadratiche di Plücker). *Sia V uno spazio vettoriale di dimensione finita. Una forma alternante $\omega \in \Omega^p(V)$ è decomponibile se e solo se per ogni successione di $2p$ vettori $v_1, v_2, \dots, v_{2p} \in V$ si ha*

$$\sum_{i=1}^{p+1} (-1)^{i-1} \omega(v_1, \dots, \widehat{v_i}, \dots, v_{p+1}) \omega(v_i, v_{p+2}, \dots, v_{2p}) = 0. \tag{12.4}$$

DIMOSTRAZIONE. Per il Teorema 12.6.7 basta mostrare che la condizione (12.4) equivale alla condizione $e_\omega \circ i_\omega^* = 0$.

Per cominciare, osserviamo che dal Teorema 12.6.5 segue immediatamente che gli operatori di valutazione

$$[v_{p+1}, \dots, v_{2p}] \colon \Omega^{p-1}(V) \to \mathbb{K}, \quad [v_{p+1}, \dots, v_{2p}](\omega) = \omega(v_{p+1}, \dots, v_{2p}),$$

formano, al variare di $v_{p+1}, \dots, v_{2p} \in V$ un insieme di generatori dello spazio duale $\Omega^{p-1}(V)^\vee$. Poiché

$$[v_{p+1}, \dots, v_{2p}](i_\omega(v)) = v \lrcorner \omega(v_{p+1}, \dots, v_{2p}) = \omega(v, v_{p+1}, \dots, v_{2p}),$$

dalla definizione di trasposta otteniamo

$$i_\omega^*([v_{p+1}, \dots, v_{2p}]) \colon V \to \mathbb{K}, \qquad i_\omega^*([v_{p+1}, \dots, v_{2p}])(v) = \omega(v, v_{p+1}, \dots, v_{2p}).$$

Direttamente dalla definizione di prodotto esterno si ha quindi

$$\begin{aligned} & e_\omega(i^*_\omega([v_{p+1},\dots,v_{2p}]))(v_1,\dots,v_{p+1}) \\ &\quad = \sum_{i=1}^{p+1}(-1)^{i-1} i^*_\omega([v_{p+1},\dots,v_{2p}])(v_i)\omega(v_1,\dots,\widehat{v_i},\dots,v_{p+1}) \\ &\quad = \sum_{i=1}^{p+1}(-1)^{i-1}\omega(v_i,v_{p+2},\dots,v_{2p})\omega(v_1,\dots,\widehat{v_i},\dots,v_{p+1}). \end{aligned}$$

□

Esercizi.

677 (♡). Sia V spazio vettoriale di dimensione 3 e sia $\phi_1,\dots,\phi_4$ un insieme di generatori di $V^\vee$. Dimostrare che $\omega = \phi_1\wedge\phi_2 + \phi_3\wedge\phi_4 \neq 0$.

678 (☕, ♡). Sia $\omega = \phi_1\wedge\phi_2 + \cdots + \phi_{2r-1}\wedge\phi_{2r} \in \Omega^2(V)$ e sia assuma che il sottospazio vettoriale generato dai funzionali $\phi_1,\dots,\phi_{2r} \in V^\vee$ abbia dimensione strettamente maggiore di r. Dimostrare che $\omega \neq 0$.

679. Se $\omega \in \Omega^p(V)$ con p intero dispari, provare che $\omega\wedge\omega = 0$ mostrando che $v\lrcorner(\omega\wedge\omega) = 0$ per ogni vettore $v \in V$. Dare un esempio di forma $\eta \in \Omega^2(V)$ tale che $\eta\wedge\eta \neq 0$.

680. Mostrare che per ogni forma $\omega \in \Omega^p(V)$ l'insieme $\{v \in V \mid v\lrcorner\omega = 0\}$ è un sottospazio vettoriale di V. Se $\phi_1,\dots,\phi_n$ è una base di $V^\vee$ e $n \geq 4$, provare che la forma $\phi_1\wedge\phi_2 + \phi_3\wedge\phi_4$ non è decomponibile.

681. Provare che, in caratteristica diversa da 2, una forma $\eta \in \Omega^2(V)$ è decomponibile se e solo se $\eta\wedge\eta = 0$.

682. Sia $f\colon V \to W$ lineare surgettiva. Provare che

$$f^*\Omega^p(W) = \{\omega \in \Omega^p(V) \mid v\lrcorner\omega = 0 \text{ per ogni } v \in \operatorname{Ker} f\}.$$

683. Siano V uno spazio vettoriale di dimensione finita n e ω un generatore di $\Omega^n(V)$. Provare che:

(1) L'applicazione $i_\omega\colon V \to \Omega^{n-1}(V)$, $i_\omega(v) = v\lrcorner\omega$, è un isomorfismo. Dedurre che ogni forma in $\Omega^{n-1}(V)$ è decomponibile.
(2) Per ogni $\eta \in \Omega^p(V)$ vi è un'unica applicazione lineare $h_\eta\colon \Omega^{n-p}(V) \to \mathbb{K}$ tale che $\eta\wedge\mu = h_\eta(\mu)\omega$ per ogni $\mu \in \Omega^{n-p}(V)$. Dedurre che la scelta di ω determina un isomorfismo $\Omega^p(V) \simeq \Omega^{n-p}(V)^\vee$.

CAPITOLO 13

Spazi quoziente

Tipicamente, le strutture quoziente riservano allo studente uno scoglio concettuale non irrilevante, ed è proprio per questo motivo che abbiamo ritardato il più possibile la loro introduzione, nella convinzione di trovare adesso lettori più maturi dal punto di vista matematico. Gli spazi vettoriali quoziente rappresentano una nozione fondamentale ed ineliminabile in tutta la matematica, che pertanto deve essere assolutamente compresa e ben digerita. Gran parte delle difficoltà concettuali sulle strutture quoziente si ritrovano anche in tutte le costruzioni rigorose dei numeri reali, una delle quali viene proposta nella Sezione 13.3 di questo capitolo.

13.1. Relazioni di equivalenza

Sia X un insieme, diremo che un sottoinsieme $R \subseteq X \times X$ del prodotto cartesiano di X con se stesso definisce una **relazione di equivalenza** se sono soddisfatte le seguenti tre proprietà:

Riflessiva: $(x, x) \in R$ *per ogni* $x \in X$*;*
Simmetrica: *se* $(x, y) \in R$*, allora* $(y, x) \in R$*;*
Transitiva: *se* $(x, y) \in R$ *e* $(y, z) \in R$*, allora* $(x, z) \in R$*.*

Ad esempio l'intero prodotto $R = X \times X$ definisce una relazione di equivalenza detta *banale*, così come la diagonale $R = \Delta = \{(x, x) \in X \times X \mid x \in X\}$ definisce una relazione di equivalenza detta *discreta*.

È facile vedere che la precedenti proprietà sono indipendenti, ossia che due delle tre non implicano la terza. Ad esempio per $X = \mathbb{Z}$, il sottoinsieme $R = \emptyset$ soddisfa le proprietà simmetrica e transitiva ma non quella riflessiva; il sottoinsieme $R = \{(x, y) \mid x \leq y\}$ soddisfa le proprietà riflessiva e transitiva ma non quella simmetrica, mentre $R = \{(x, y) \mid |x - y| \leq 1\}$ soddisfa le proprietà riflessiva e simmetrica ma non quella transitiva.

È sempre possibile descrivere un sottoinsieme del prodotto $R \subseteq X \times X$ mediante una relazione binaria $\sim$, dove scriveremo $x \sim y$ se e solo se $(x, y) \in R$. Ad esempio, nella relazione di equivalenza banale vale $x \sim y$ per ogni $x, y \in X$, mentre nella relazione di equivalenza discreta vale $x \sim y$ se e solo se $x = y$. In termini di $\sim$ le prededenti proprietà diventano:

Riflessiva: $x \sim x$ *per ogni* $x \in X$*;*
Simmetrica: *se* $x \sim y$*, allora* $y \sim x$*;*
Transitiva: *se* $x \sim y$ *e* $y \sim z$*, allora* $x \sim z$*.*

Dunque, una relazione di equivalenza può essere vista come una relazione binaria $\sim$ che soddisfa le condizioni riflessiva, simmetrica e transitiva; in tal caso se $x \sim y$ diremo che x è equivalente ad y.

Esempio 13.1.1. Abbiamo visto che ogni insieme possiede le relazioni di equivalenza banale e discreta: tali relazioni coincidono se e solo se l'insieme contiene al più un punto.

Esempio 13.1.2. Sull'insieme degli interi $\mathbb{Z}$ definiamo la relazione $\sim$ ponendo $x \sim y$ se e solo se $x - y$ è divisibile per 2. Lasciamo al lettore la semplice verifica che si tratta di una relazione di equivalenza.

Esempio 13.1.3. La relazione di similitudine tra matrici quadrate (Definizione 9.1.2), è una relazione di equivalenza.

Definizione 13.1.4. Sia $\sim$ una relazione di equivalenza su un insieme X. La **classe di equivalenza** di un elemento $x \in X$ è il sottoinsieme

$$[x] \subseteq X, \qquad [x] = \{y \in X \mid y \sim x\}.$$

Si dimostra facilmente che se $[x] \cap [y] \neq \emptyset$ allora $x \sim y$ e $[x] = [y]$: infatti se $[x] \cap [y] \neq \emptyset$, scelto un qualsiasi elemento $z \in [x] \cap [y]$ si ha per definizione di classe di equivalenza $z \sim x$ e $z \sim y$, per simmetria $x \sim z$ e per la proprietà transitiva $x \sim y$. Adesso per ogni $u \in [x]$ si ha $u \sim x$, $x \sim y$ e per transitività $u \sim y$, ossia $u \in [y]$; abbiamo dunque provato che $[x] \subseteq [y]$, per simmetria vale anche $[y] \subseteq [x]$ e quindi $[x] = [y]$.

Detto in altri termini, le classi di equivalenza di una relazione di equivalenza sull'insieme X dànno una decomposizione di X come unione di sottoinsiemi a due a due disgiunti. Le classi di equivalenza determinano univocamente la relazione di equivalenza e le tre proprietà caratterizzanti diventano:

Riflessiva: *$x \in [x]$ per ogni $x \in X$;*
Simmetrica: *se $x \in [y]$, allora $y \in [x]$;*
Transitiva: *se $x \in [y]$ e $y \in [z]$, allora $x \in [z]$.*

Esempio 13.1.5. Data un'applicazione tra insiemi $f\colon X \to Q$, possiamo definire sull'insieme X una relazione di equivalenza $\sim$ ponendo $x \sim y$ se e solo se $f(x) = f(y)$. Le classi di equivalenza di tale relazione coincidono con le fibre di f; più precisamente, per ogni $x \in X$ vale

$$[x] = f^{-1}(\{f(x)\}) = \{y \in X \mid f(y) \in \{f(x)\}\}.$$

Ogni relazione di equivalenza può essere pensata del tipo descritto nell'Esempio 13.1.5. Data una qualsiasi relazione di equivalenza $\sim$ su un insieme X possiamo considerare l'applicazione

$$\pi\colon X \to \mathcal{P}(X), \qquad \pi(x) = [x],$$

che ad ogni elemento associa la corrispondente classe di equivalenza. Per definizione di classe di equivalenza si ha $x \sim y$ se se solo se $\pi(x) = \pi(y)$.

Definizione 13.1.6. Nelle notazioni precedenti l'immagine dell'applicazione $\pi\colon X \to \mathcal{P}(X)$ viene detta **insieme quoziente** di X per la relazione $\sim$ e viene indicato $X/\!\sim$:

$$X/\!\sim = \{[x] \mid x \in X\} \subseteq \mathcal{P}(X).$$

L'applicazione

$$\pi\colon X \to X/\!\sim, \qquad \pi(x) = [x],$$

viene detta **proiezione al quoziente**.

Nella pratica matematica, la descrizione esplicita del quoziente come sottoinsieme dell'insieme delle parti rappresenta un'inelegante zavorra; si tende quindi ad ignorare tale descrizione sostituendola nei ragionamenti con le proprietà della proiezione:

(1) la proiezione al quoziente $\pi\colon X \to X/\!\sim$ è surgettiva;
(2) dati $x, y \in X$, vale $x \sim y$ se e solo se $\pi(x) = \pi(y)$.

Data una relazione di equivalenza $\sim$ su un insieme X, lasciamo per esercizio al lettore la prova che, per un sottoinsieme $S \subseteq X$, le seguenti condizioni sono equivalenti:

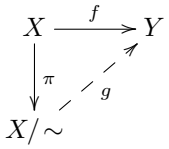
(1) per ogni $x \in X$ vi è un unico $s \in S$ tale che $s \sim s$;
(2) S interseca ogni classe di equivalenza in uno ed un solo punto;
(3) la restrizione della proiezione $\pi_{|S}\colon S \to X/\!\sim$ è bigettiva;
(4) S è l'immagine di un'applicazione $s\colon X/\!\sim \to X$ tale che $\pi(s(q)) = q$ per ogni $q \in X/\!\sim$.

Se tali condizioni sono verificate, diremo che S è un **insieme di rappresentanti** per la relazione $\sim$.

Per ogni relazione di equivalenza esiste un insieme di rappresentanti: infatti la proiezione al quoziente è surgettiva e per l'assioma della scelta (pagina 45) esiste un'applicazione $s\colon X/\!\sim \to X$ tale che $\pi s([x]) = [x]$ per ogni classe di equivalenza $[x]$.

Lemma 13.1.7. *Siano $\sim$ una relazione di equivalenza su X, $\pi\colon X \to X/\!\sim$ la proiezione al quoziente e $f\colon X \to Y$ un'applicazione. Sono fatti equivalenti:*

(1) *L'applicazione f è costante sulle classi di equivalenza, ossia $f(x) = f(y)$ ogni volta che $x \sim y$.*
(2) *Esiste un'unica applicazione $g\colon X/\!\sim \to Y$ tale che $f = g\pi$.*



DIMOSTRAZIONE. Esercizio. □

Esercizi.

684. Per le relazioni appresso definite sull'insieme degli uomini e delle donne viventi, dire quali sono di equivalenza:

(1) $a \sim b$ se a e b hanno un antenato in comune dopo l'anno 1000;
(2) $a \sim b$ se a e b abitano a meno di 200 km di distanza;
(3) $a \sim b$ se a e b hanno lo stesso padre;
(4) $a \sim b$ se a e b sono clienti della MidelPhone;
(5) $a \sim b$ se a è meno giovane di b.

685. Dire quali, tra le seguenti relazioni binarie sull'insieme $M_{n,n}(\mathbb{C})$ delle matrici $n \times n$ a coefficienti complessi, sono relazioni di equivalenza:

(1) $A \sim B$ se e solo se $A - B \in M_{n,n}(\mathbb{R})$;
(2) $A \sim B$ se e solo se $A - B$ è nilpotente;
(3) $A \sim B$ se e solo se $\det(A) = \det(B)$;
(4) $A \sim B$ se e solo se $A^2 = B^2$;
(5) $A \sim B$ se e solo se $A = \lambda B$ per qualche $\lambda \in \mathbb{C}$;
(6) $A \sim B$ se e solo se $A = \lambda B$ per qualche $\lambda \in \mathbb{C} - \{0\}$.

686. Sull'insieme $X \subseteq \mathbb{Z} \times \mathbb{Z}$ formato dalle coppie (n, m) di interi con $m > 0$ consideriamo la relazione

$$(n, m) \sim (p, q) \iff nq = mp, \qquad (n, m), (p, q) \in X.$$

Verificare che si tratta di una relazione di equivalenza e che esiste una bigezione naturale tra l'insieme quoziente $X/\sim$ e l'insieme $\mathbb{Q}$ dei numeri razionali.

687. Siano X un insieme e $\mathcal{R} \subseteq \mathcal{P}(X \times X)$ una famiglia di relazioni di equivalenza. Provare che l'intersezione

$$S = \bigcap \{R \in \mathcal{R}\} \subseteq X \times X,$$

di tutte le relazioni in $\mathcal{R}$ è ancora una relazione di equivalenza.

13.2. Spazi vettoriali quoziente

Siano V uno spazio vettoriale sul campo $\mathbb{K}$ ed U un sottospazio di V. Si consideri in V la relazione

$$w \sim v \iff w - v \in U.$$

È immediato verificare che $\sim$ è una relazione di equivalenza e che $v \sim 0$ se e soltanto se $v \in U$. Si denota l'insieme quoziente $V/\sim$ con il simbolo V/U. Dato $v \in V$ si denota con $[v]$ la classe di equivalenza di v in V/U e con

$$\pi \colon V \longrightarrow V/U, \qquad v \mapsto \pi(v) = [v],$$

la proiezione al quoziente.

Si dota V/U della struttura di spazio vettoriale ponendo, per $w, v \in V$ e $a \in \mathbb{K}$,

$$[w] + [v] = [w + v]\,, \qquad a[v] = [av]\,, \qquad 0_{V/U} = [0_V].$$

Con questa struttura, la proiezione π risulta essere una applicazione lineare surgettiva e $\operatorname{Ker} \pi = U$. Le verifiche sono immediate.

Nel trattare gli spazi quoziente torna utile usare la nozione di successione esatta corta (Definizione 5.6.5). Ricordiamo che una successione di spazi vettoriali ed applicazioni lineari

$$0 \to U \xrightarrow{f} V \xrightarrow{g} W \to 0$$

si dice **esatta corta** se f è iniettiva, g è surgettiva e $\operatorname{Ker} g = f(U)$; in particolare si ha $W = g(V)$, $f\colon U \to \operatorname{Ker} g$ è un isomorfismo e, se V ha dimensione finita, allora per il teorema del rango abbiamo

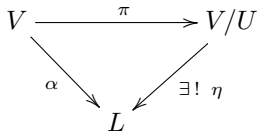
$$\dim V = \dim \operatorname{Ker} g + \dim g(V) = \dim U + \dim W.$$

Per costruzione si ha una successione esatta corta

$$0 \to U \xrightarrow{\iota} V \xrightarrow{\pi} V/U \to 0$$

dove ι è l'inclusione. Dunque, se V ha dimensione finita si ha $\dim V/U = \dim V - \dim U$.

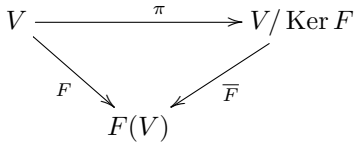
Il quoziente V/U verifica la seguente proprietà universale. Data una qualsiasi applicazione lineare $\alpha\colon V \to L$ tale che $\alpha(U) = 0$, esiste un'unica applicazione lineare $\eta\colon V/U \to L$ tale che $\eta\pi = \alpha$:



Infatti basta porre $\eta([v]) = \alpha(v)$ e verificare che la definizione è ben posta. Da questa proprietà segue che, se $F\colon V \to W$ è una qualunque applicazione lineare, allora si ha un isomorfismo lineare

$$\overline{F}\colon V/\operatorname{Ker} F \xrightarrow{\cong} F(V).$$

Infatti, per la proprietà universale si ha un diagramma



Essendo $\overline{F}$ suriettiva ed essendo la dimensione del suo dominio uguale a quella del suo codominio, si ha che $\overline{F}$ è un isomorfismo. In particolare, se F è suriettiva, si ha $V/\operatorname{Ker} F \cong W$.

Definizione 13.2.1. Data una applicazione lineare $f\colon V \to W$ si definisce il **conucleo** (detto anche cokèr, abbreviazione del termine inglese cokernel) di F come:

$$\operatorname{Coker} f = \frac{W}{f(V)}.$$

Si ha una successione esatta

$$0 \to \operatorname{Ker} f \xrightarrow{\iota} V \xrightarrow{f} W \xrightarrow{p} \operatorname{Coker} f \to 0$$

dove ι è l'inclusione e p è la proiezione. Si hanno inoltre le formule

$$\dim \operatorname{Ker} f = \dim V - \operatorname{rango} f, \qquad \dim \operatorname{Coker} f = \dim W - \operatorname{rango} f.$$

Esempio 13.2.2. Data una successione esatta corta

$$0 \to U \xrightarrow{f} V \xrightarrow{g} W \to 0$$

l'applicazione G si fattorizza ad una applicazione bigettiva

$$V/\operatorname{Ker} g \to W$$

e quindi, induce un isomorfismo

$$\operatorname{Coker} f \cong W.$$

Esempio 13.2.3. Gli spazi vettoriali quoziente, sebbene non strettamente necessari, ci aiutano a provare l'analogo della Proposizione 10.3.7 per gli endomorfismi nilpotenti. Abbiamo visto nel Teorema 10.4.7 che ogni endomorfismo nilpotente, preso singolarmente, è triangolabile ma in generale non è possibile mettere contemporaneamente in forma triangolare superiore due distinti endomorfismi nilpotenti. Infatti se $f, g\colon V \to V$ sono rappresentati in una opportuna base con matrici triangolari strettamente superiori, allora ogni combinazione

lineare $af + bg$ è ancora rappresentata, nella stessa base, da una matrice triangolare strettamente superiore ed è quindi nilpotente. È quindi evidente che, ad esempio, non è possibile triangolarizzare contemporaneamente le due matrici nilpotenti

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

poiché la loro somma non è nilpotente.

Al fine di descrivere una condizione necessaria e sufficiente alla triangolazione simultanea di endomorfismi nilpotenti, per ogni successione $f_1, f_2, \ldots \in \operatorname{Hom}(V, V)$, possibilmente con ripetizioni, definiamo i commutatori iterati $[f_1, \ldots, f_n] \in \operatorname{Hom}(V, V)$, $n > 0$, mediante le formule ricorsive:

$$[f_1] = f_1, \qquad [f_1, f_2] = f_1 f_2 - f_2 f_1, \qquad [f_1, f_2, f_3] = [[f_1, f_2], f_3],$$
$$[f_1, \ldots, f_n] = [[f_1, f_2, \ldots, f_{n-1}], f_n], \qquad n > 1.$$

Teorema 13.2.4. *Siano V spazio vettoriale di dimensione finita e $\mathcal{N} \subseteq \operatorname{Hom}(V, V)$ un sottoinsieme di endomorfismi. Allora esiste una base di V in cui ogni elemento di $\mathcal{N}$ si rappresenta con una matrice triangolare strettamente superiore se e solo se:*

1. *per ogni $n \geq 1$ e per ogni successione $f_1, \ldots, f_n \in \mathcal{N}$, il commutatore iterato $[f_1, \ldots, f_n]$ è nilpotente;*
2. *esiste un intero positivo m tale che $[f_1, \ldots, f_m] = 0$ per ogni successione $f_1, \ldots, f_m \in \mathcal{N}$.*

Nota: in entrambe le condizioni del teorema, alle successioni è consentito avere elementi ripetuti, come ad esempio nel caso in cui $\mathcal{N} \subseteq \operatorname{Hom}(\mathbb{R}^3, \mathbb{R}^3)$ è formato dalle due matrici

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Dimostrazione. Indichiamo con k la dimensione di V e con $N \subseteq M_{k,k}(\mathbb{K})$ è il sottospazio delle matrici triangolari strettamente superiori. Si osserva che se $A, B \in N$, allora anche $[A, B] \in N$ e si dimostra facilmente per induzione su r che per ogni $A_1, \ldots, A_r \in N$, se a_{ij} è un coefficiente non nullo della matrice $[A_1, \ldots, A_r]$, allora $j \geq i + r$: in particolare $[A_1, \ldots, A_k] = 0$ per ogni $A_1, \ldots, A_k \in N$.

Viceversa, supponiamo che $\mathcal{N} \subseteq \operatorname{Hom}(V, V)$ soddisfa le due condizioni del teorema e sia m il più piccolo intero tale che $[f_1, \ldots, f_m] = 0$ per ogni $f_1, \ldots, f_m \in \mathcal{N}$. Esistono quindi $f_1, \ldots, f_{m-1} \in \mathcal{N}$ tali che l'endomorfismo $g = [f_1, \ldots, f_{m-1}]$ è nilpotente ma non nullo. Per come abbiamo scelto m si ha $[g, f] = 0$ per ogni $f \in \mathcal{N}$ e quindi il sottospazio $U = \operatorname{Ker} g$ è f-invariante per ogni $f \in \mathcal{N}$. Per induzione sulla dimensione esiste una base $u_1, \ldots, u_r$ di U ed una base $w_{r+1}, \ldots w_k$ di V/U rispetto alle quali ogni f si rappresenta con una matrice triangolare strettamente superiore. Basta adesso sollevare ciascun w_i ad un vettore u_i di V e $u_1, \ldots, u_k$ è la base cercata. □

Esercizi.

688. Dati uno spazio vettoriale V e due sottospazi $U \subseteq W \subseteq V$, mostrare che esiste una successione esatta

$$0 \to \frac{W}{U} \to \frac{V}{U} \to \frac{V}{W} \to 0$$

e dedurre che esiste un isomorfismo

$$\frac{V/U}{W/U} \cong \frac{V}{W}.$$

689. Dati uno spazio vettoriale V e due sottospazi $U, W \subseteq V$, mostrare che esiste una successione esatta

$$0 \to W \cap U \to U \to \frac{U + W}{W} \to 0$$

e dedurre che esiste un isomorfismo

$$\frac{U + W}{W} \cong \frac{U}{U \cap W}.$$

690. Data una successione esatta corta

$$0 \to U \xrightarrow{f} V \xrightarrow{g} W \to 0$$

ed un'applicazione lineare $h\colon Z \to W$ denotiamo

$$V \times_W Z = \{(v, z) \in V \times Z \mid g(v) = h(z)\},$$

con $p\colon V \times_W Z \to Z$ la proiezione $p(v, z) = z$ e con $q\colon U \to V \times_W Z$ l'applicazione $q(u) = (u, 0)$. Dimostrare che

$$0 \to U \xrightarrow{q} V \times_W Z \xrightarrow{p} Z \to 0$$

è una successione esatta.

691. Data una successione esatta corta

$$0 \to U \xrightarrow{f} V \xrightarrow{g} W \to 0$$

ed un'applicazione lineare $h\colon U \to Z$ denotiamo

$$Z \amalg_U V = \frac{Z \oplus V}{K},$$

dove K è il sottospazio vettoriale di $Z \oplus V$ formato dalle coppie $(h(u), -f(u))$, $u \in U$. Dimostrare che esiste una successione esatta corta

$$0 \to Z \to Z \amalg_U V \to W \to 0.$$

692. Dato un diagramma commutativo

$$\begin{array}{ccccccccc} 0 & \longrightarrow & U & \longrightarrow & V & \longrightarrow & W & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle h} & & \\ 0 & \longrightarrow & U & \longrightarrow & V & \longrightarrow & W & \longrightarrow & 0 \end{array}$$

di spazi vettoriali di dimensione finita con le righe esatte e *identiche* tra loro, dimostrare che

$$\operatorname{Tr} g = \operatorname{Tr} f + \operatorname{Tr} h, \qquad \det(g) = \det(f)\det(h), \qquad p_g(t) = p_f(t)p_h(t).$$

693 (Ⓐ). 1) Definiamo per ricorrenza la successione di numeri interi

$$c_0 = 1, \qquad c_{n+1} = (2n+1)c_n, \qquad n \geq 0.$$

Dimostrare che il diagramma

$$0 \to \mathbb{K}[t] \xrightarrow{f} \mathbb{K}[t] \xrightarrow{g} \mathbb{K} \to 0, \qquad f(p(t)) = tp(t) - p(t)', \quad g\left(\sum a_i t^i\right) = \sum_n c_n a_{2n},$$

è una successione esatta corta di spazi vettoriali sul campo $\mathbb{K}$.

2) Siete prigionieri di un tiranno, che per un misto di sadismo e generosità vi offre la libertà qualora riusciate a calcolare l'integrale indefinito

$$\int p(x)e^{-x^2}dx$$

dove $p(x)$ è una funzione polinomiale. A voi non è consentito scegliere $p(x)$ ma potete optare per una delle seguenti 4 possibilità: la funzione $p(x)$ è pari, la funzione $p(x)$ è dispari, $p(0) = 0$, il polinomio p ha grado ≤ 2. Cosa scegliete?

3) Che legame c'è tra i precedenti punti 1) e 2)?

694. Siano V uno spazio vettoriale di dimensione finita ed $f\colon V \to V$ un endomorfismo. Provare che se esiste una filtrazione $0 = V_0 \subseteq V_1 \subseteq V_2 \subseteq \cdots \subseteq V_k = V$ di sottospazi f-invarianti tale che le fattorizzazioni ai quozienti

$$f_i\colon \frac{V_i}{V_{i-1}} \to \frac{V_i}{V_{i-1}}, \qquad f_i([v]) = [f(v)],$$

sono triangolabili, allora anche f è triangolabile.

695 (La filtrazione dei pesi, ☕). Dato un endomorfismo nilpotente f su uno spazio vettoriale di dimensione finita V, oltre alle filtrazioni dei nuclei e delle immagini, riveste una certa importanza la cosiddetta *filtrazione dei pesi*. Ponendo per convenzione f^0 uguale all'identità, per ogni intero k definiamo

$$W_k = \sum_{\substack{a,b \geq 0 \\ a-b=k+1}} (\operatorname{Ker} f^a \cap f^b(V)) \,.$$

Dimostrare che:

(1) $W_k \subseteq W_{k+1}$ per ogni $k \in \mathbb{Z}$;
(2) se $\tau \geq 0$ e $f^{\tau+1} = 0$, allora $W_\tau = V$ e $W_{-\tau-1} = 0$;
(3) $f(W_k) \subseteq W_{k-2}$ per ogni $k \in \mathbb{Z}$;
(4) se in un'opportuna base $u_1, \dots, u_n$ di V si ha $f(u_i) = u_{i+1}$ per ogni $i < n$, allora $u_i \in W_{n-2i+1}$ e $u_i \notin W_{n-2i}$ per ogni i;
(5) $f^k \colon \dfrac{W_k}{W_{k-1}} \to \dfrac{W_{-k}}{W_{-k-1}}$ è un isomorfismo per ogni $k \geq 0$.

(Suggerimento: per l'ultimo punto usare la forma canonica di Jordan.)

696. Siano $A, B, H \in M_{2,2}(\mathbb{C})$ le matrici:

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

Verificare che $[A, B] = AB - BA = H$, $[H, A] = 2A$, $[H, B] = -2B$, e dedurre che le matrici A e B non sono simultaneamente triangolabili.

697. Indichiamo con $[f_1, \dots, f_n]$ il commutatore iterato di $f_1, \dots, f_n$. Provare che se $[f_i, f_n] = 0$ per ogni $1 \leq i < n$, allora $[f_1, \dots, f_n] = 0$. (Suggerimento: induzione su n e identità di Jacobi, vedi Esercizio 320).

13.3. La costruzione dei numeri reali

La teoria degli insiemi quoziente e degli spazi vettoriali quoziente fornisce un modo concettualmente valido per definire in maniera rigorosa l'insieme dei numeri reali assieme alla sua struttura di spazio vettoriale sul campo $\mathbb{Q}$, nel quale è facile definire un prodotto ed una relazione di ordine che soddisfano le ben note proprietà. La costruzione che proponiamo è una mediazione tra la classica costruzione di Cantor, basata sulle successioni di Cauchy, e la descrizione dei numeri reali come sviluppi decimali infiniti.

Fissato un qualsiasi numero razionale $a \in \mathbb{Q}$, indichiamo con $|a| \in \mathbb{Q}$ il suo valore assoluto; osserviamo che per ogni $a, b \in \mathbb{Q}$ si ha $|ab| = |a|\,|b|$ e vale la cosiddetta *disuguaglianza triangolare* $|a + b| \leq |a| + |b|$.

Denotiamo con $\mathbb{Q}^{\mathbb{N}}$ l'insieme di tutte le applicazioni $a \colon \mathbb{N} \to \mathbb{Q}$; possiamo pensare ogni $a \in \mathbb{Q}^{\mathbb{N}}$ come una successione $a = (a_0, a_1, \dots, a_n, \dots)$, $n \in \mathbb{N}$, di numeri razionali. Come nell'Esempio 4.2.5, l'insieme $\mathbb{Q}^{\mathbb{N}}$ possiede una naturale struttura di spazio vettoriale sul campo $\mathbb{Q}$, dove le operazioni di somma e di prodotto per scalare sono definite come:

$$(a_0, a_1, \dots) + (b_0, b_1, \dots) = (a_0 + b_0, a_1 + b_1, \dots), \quad r(a_0, a_1, \dots) = (ra_0, ra_1, \dots).$$

Denotiamo con $j \colon \mathbb{Q} \to \mathbb{Q}^{\mathbb{N}}$ l'applicazione lineare iniettiva che associa ad ogni numero razionale r la successione $j(r) = (r, r, \dots)$ che vale costantemente r e con con $T \colon \mathbb{Q}^{\mathbb{N}} \to \mathbb{Q}^{\mathbb{N}}$ l'applicazione lineare surgettiva che abbassa gli indici di una unità:

$$T(a_0, a_1, a_2, a_3, \dots) = (a_1, a_2, a_3, \dots).$$

È chiaramente possibile considerare le potenze nonnegative di T:

$$T^m(a_0, a_1, a_2, a_3, \dots) = (a_m, a_{1+m}, a_{2+m}, \dots), \qquad m \geq 0.$$

Definizione 13.3.1. Denotiamo con $\mathfrak{m} \subseteq \mathbb{Q}^{\mathbb{N}}$ il sottoinsieme delle successioni $a = (a_0, a_1, a_2, \dots)$ per cui esiste una costante $M \in \mathbb{Q}$ tale che

$$|a_n| \leq \frac{M}{10^n} \quad \text{per ogni} \quad n \in \mathbb{N}.$$

Denotiamo con $R \subseteq \mathbb{Q}^{\mathbb{N}}$ il sottoinsieme delle successioni $a \in \mathbb{Q}^{\mathbb{N}}$ tali che $a - Ta \in \mathfrak{m}$.

Chiameremo gli elementi di R *serie decimali*: segue immediatamente dalla definizione che $a \in R$ se e solo se esiste una costante $M \in \mathbb{Q}$ tale che

$$|a_n - a_{n+1}| = |a_n - (Ta)_n| \leq \frac{M}{10^n} \quad \text{per ogni} \quad n \in \mathbb{N}.$$

Diremo inoltre che una serie decimale a è *unitaria* se la costante M può essere presa uguale ad 1, ossia se

$$|a_n - a_{n+1}| = |a_n - (Ta)_n| \leq \frac{1}{10^n} \quad \text{per ogni} \quad n \in \mathbb{N}.$$

È del tutto evidente che se a è una serie decimale, allora $T^m a$ è unitaria per ogni valore di m sufficientemente grande.

Esempio 13.3.2. Ogni sviluppo decimale infinito

$$m, \alpha_1\alpha_2\alpha_3 \ldots \qquad m \in \mathbb{Z}, \quad \alpha_n = 0, \ldots, 9,$$

definisce in maniera canonica una serie decimale unitaria: infatti se $a_n = m, \alpha_1 \ldots \alpha_n$ indica la frazione decimale ottenuta considerando solo le prime n cifre dopo la virgola (ossia mettendo le rimanenti uguali a 0) si ha $a_{n+1} - a_n = \pm\frac{\alpha_{n+1}}{10^{n+1}}$ e quindi

$$|a_n - a_{n+1}| \leq \frac{9}{10^{n+1}} \leq \frac{1}{10^n}.$$

Lemma 13.3.3. *I sottoinsiemi $R, \mathfrak{m} \subseteq \mathbb{Q}^{\mathbb{N}}$ sono sottospazi vettoriali e si ha:*

$$\mathfrak{m} + j(\mathbb{Q}) \subseteq R, \qquad \mathfrak{m} \cap j(\mathbb{Q}) = 0, \qquad T(R) \subseteq R, \qquad T(\mathfrak{m}) \subseteq \mathfrak{m}. \tag{13.1}$$

Dimostrazione. Mostriamo che $\mathfrak{m}$ è un sottospazio vettoriale: essendo chiaro che se $0 \in \mathfrak{m}$, basta mostrare che tale sottoinsieme è chiuso per combinazioni lineari. Se $a, b \in \mathfrak{m}$, per definizione esistono due costanti $N, M \in \mathbb{Q}$ tali che per ogni n vale

$$|a_n| \leq \frac{N}{10^n}, \qquad |b_n| \leq \frac{M}{10^n}.$$

Per la disuguaglianza triangolare, per ogni coppia di numeri razionali r, s si ha

$$|ra_n + sb_n| \leq |ra_n| + |sb_n| \leq \frac{|rN| + |sM|}{10^n}$$

e questo prova che $ra + sb \in \mathfrak{m}$. Dati $a, b \in R$ per definizione $a - Ta, b - Tb \in \mathfrak{m}$ e quindi per ogni $r, s \in \mathbb{Q}$ si ha

$$(ra + sb) - T(ra + sb) = r(a - Ta) + s(b - Tb) \in \mathfrak{m}.$$

La dimostrazione delle quattro relazioni (13.1) viene lasciata come (semplice) esercizio per il lettore. □

Definiamo lo **spazio vettoriale dei numeri reali** $\mathbb{R}$ come lo spazio vettoriale quoziente $\mathbb{R} = \dfrac{R}{\mathfrak{m}}$. Per ogni $a \in R$ denotiamo con $[a] \in \mathbb{R}$ la corrispondente classe di equivalenza: per definizione si ha $[a] = [b]$ se e solo se $a - b \in \mathfrak{m}$. Poiché $a - T^m a \in \mathfrak{m}$ per ogni $a \in R$ ed ogni intero positivo m, è sempre possibile rappresentare ogni numero da una serie decimale unitaria.

L'applicazione $\mathbb{Q} \to \mathbb{R}$, $r \mapsto [j(r)]$ è iniettiva e per semplicità notazionale identificheremo $\mathbb{Q}$ con la sua immagine dentro $\mathbb{R}$; con tale convenzione si ha $[j(r)] = r$ per ogni $r \in \mathbb{Q}$.

La struttura di $\mathbb{R}$ come spazio vettoriale su $\mathbb{Q}$ ci permette di definire le operazioni di somma di due numeri reali e di prodotto di un numero reale per un numero razionale. Il nostro prossimo obiettivo è quello di definire il prodotto di due numeri reali; a tal fine è utile la seguente caratterizzazione delle serie decimali.

Lemma 13.3.4. *Dato $a \in \mathbb{Q}^{\mathbb{N}}$, vale $a \in R$ se e solo se esiste una costante $N \in \mathbb{Q}$ tale che per $|a_n - a_{n+k}| \leq \frac{N}{10^n}$ per ogni $n, k \in \mathbb{N}$. In particolare, ogni serie decimale è una successione limitata di numeri razionali, ossia se $a \in R$ allora esiste una costante P tale che $|a_n| \leq P$ per ogni n.*

DIMOSTRAZIONE. Una implicazione è banale. Viceversa, se esiste M tale che $|a_n - a_{n+1}| < \frac{M}{10^n}$, allora per ogni n, k si ha (vedi Esempio 2.3.9)

$$|a_n - a_{n+k}| \leq \sum_{i=1}^{k} |a_{n+i} - a_{n+i-1}| \leq \frac{M}{10^n} \sum_{i=1}^{k} \frac{1}{10^{i-1}} \leq \frac{10M}{9} \frac{1}{10^n}.$$

In particolare $|a_0 - a_k| \leq \frac{10M}{9}$ per ogni k e per la disuguaglianza triangolare

$$|a_k| \leq |a_0| + \frac{10M}{9}.$$

□

Lo spazio $\mathbb{Q}^{\mathbb{N}}$ possiede un prodotto naturale $\cdot$, ottenuto dalla moltiplicazione componente per componente:

$$(a_0, a_1, \ldots) \cdot (b_0, b_1, \ldots) = (a_0 b_0, a_1 b_1, \ldots).$$

Un tale prodotto eredita dalla moltiplicazione di numeri razionali le proprietà associativa, commutativa e distributiva:

$$a, b, c \in \mathbb{Q}^{\mathbb{N}}, \qquad a \cdot (b \cdot c) = (a \cdot b) \cdot c, \quad a \cdot b = b \cdot a, \quad a \cdot (b + c) = a \cdot b + a \cdot c.$$

Inoltre, per ogni $r \in \mathbb{Q}$ ed ogni $a \in \mathbb{Q}^{\mathbb{N}}$ vale $j(r) \cdot a = ra$.

LEMMA 13.3.5. *Nelle notazioni precedenti, per ogni $a, b \in R$ ed ogni $c \in \mathfrak{m}$, si ha $a \cdot b \in R$ e $a \cdot c \in \mathfrak{m}$.*

DIMOSTRAZIONE. Per ipotesi esiste una costante M tale che $|c_n| \leq \frac{M}{10^n}$ per ogni n. Sappiamo che le serie decimali sono limitate, in particolare esiste una costante P tale che $|a_n| \leq P$ per ogni n. Ne segue che

$$|a_n c_n| \leq \frac{PM}{10^n}$$

che vuol dire $a \cdot c \in \mathfrak{m}$. Per definizione di R si ha $a - Ta, b - Tb \in \mathfrak{m}$ e quindi

$$a \cdot b - T(a \cdot b) = a \cdot b - Ta \cdot Tb = a \cdot (b - Tb) + (a - Ta) \cdot Tb$$

e abbiamo dimostrato che $a \cdot (b - Tb) \in \mathfrak{m}$ e $(a - Ta) \cdot Tb \in \mathfrak{m}$. □

Possiamo quindi definire il prodotto di due numeri reali mediante la formula

$$[a] \cdot [b] = [a \cdot b]$$

ed il Lemma 13.3.5 mostra che tale prodotto non dipende dalla scelta di a, b all'interno delle proprie classi di equivalenza.

PROPOSIZIONE 13.3.6. *Sia $t \in \mathbb{R}$, $t \neq 0$. Vi è un unico numero reale $t^{-1} \in \mathbb{R}$ tale che $t \cdot t^{-1} = t^{-1} \cdot t = 1$.*

DIMOSTRAZIONE. L'unicità è facile da dimostrare: se $h \in \mathbb{R}$ è un altro numero reale tale che $h \cdot t = 1$, per l'associatività del prodotto si ha

$$h = h \cdot 1 = h \cdot (s \cdot s^{-1}) = (h \cdot s) \cdot s^{-1} = 1 \cdot s^{-1} = s^{-1}.$$

Per dimostrare l'esistenza è sufficiente dimostrare che se $a \in R$, $a \notin \mathfrak{m}$, allora esiste un intero positivo m ed una serie decimale $b \in R$ tale che $(T^m a) \cdot b = j(1)$.

Per il Lemma 13.3.4 esiste una costante $N > 0$ tale che $|a_n - a_{n+k}| \leq \frac{N}{10^n}$ per ogni $n, k \in \mathbb{N}$. Per ipotesi $a \notin \mathfrak{m}$ e quindi esiste un intero $m \geq 0$ tale che $|a_m| \geq \frac{2N}{10^m}$; per la disuguaglianza triangolare

$$|a_{m+n}| \geq |a_m| - |a_m - a_{m+n}| \geq \frac{N}{10^m}, \quad \text{per ogni} \quad n \in \mathbb{N}.$$

Ne consegue che

$$(T^m a) \cdot b = j(1), \qquad \text{dove} \quad b_n = \frac{1}{a_{n+m}},$$

e per concludere la dimostrazione basta dimostrare che $b \in R$. Per ogni $n \in \mathbb{N}$ si ha

$$|b_n - b_{n+1}| = \left| \frac{1}{a_{n+m}} - \frac{1}{a_{n+m+1}} \right| = \left| \frac{a_{n+m+1} - a_{n+m}}{a_{n+m} a_{n+m+1}} \right| = \frac{|a_{n+m+1} - a_{n+m}|}{|a_{n+m}||a_{n+m+1}|}$$

$$\leq \frac{10^{2m}}{N^2} |a_{n+m} - a_{n+m+1}| \leq \frac{10^{2m}}{N^2} \frac{N}{10^{n+m}} = \frac{10^m}{N} \frac{1}{10^n}.$$

□

L'ordinamento usuale su $\mathbb{Q}$ induce in maniera naturale una relazione di ordine sullo spazio R delle serie decimali: date $a, b \in R$ scriveremo $a \leq b$ se $a_n \leq b_n$ per ogni $n \in \mathbb{N}$.

LEMMA 13.3.7. *Per due numeri reali $s, t \in \mathbb{R}$ le seguenti condizioni sono equivalenti:*

(1) *esistono $a, b \in R$ tali che $s = [a]$, $t = [b]$ e $a \leq b$;*
(2) *per ogni $a \in R$ tale che $s = [a]$, esiste $b \in R$ con $t = [b]$ e $a \leq b$;*
(3) *per ogni $b \in R$ tale che $t = [b]$, esiste $a \in R$ con $s = [a]$ e $a \leq b$;*
(4) *per ogni $a, b \in R$ tali che $s = [a]$ e $t = [b]$, esiste $c \in \mathfrak{m}$ tale che $a \leq b + c$.*

DIMOSTRAZIONE. Facile esercizio. □

DEFINIZIONE 13.3.8. Dati due numeri reali s, t, scriveremo $s \leq t$ se valgono le condizioni equivalenti del Lemma 13.3.7.

Notiamo che se $a, b \in R$ e $T^m a \leq T^m b$ per qualche $m > 0$, allora $[a] \leq [b]$.

PROPOSIZIONE 13.3.9. *La relazione $\leq$ su $\mathbb{R}$ della Definizione 13.3.8 estende la usuale relazione di ordine tra numeri razionali. Valgono inoltre le seguenti proprietà:*

(1) (proprietà riflessiva) *$s \leq s$ per ogni $s \in \mathbb{R}$;*
(2) (proprietà antisimmetrica) *se $s, t \in \mathbb{R}$, $s \leq t$ e $t \leq s$ allora $s = t$;*
(3) (proprietà transitiva) *se $s, t, u \in \mathbb{R}$, $s \leq t$ e $t \leq u$ allora $s \leq u$;*
(4) (totalità dell'ordinamento) *per ogni $s, t \in \mathbb{R}$ si ha $s \leq t$ oppure $t \leq s$;*
(5) (parte intera) *per ogni $s \in \mathbb{R}$ esiste un unico intero $\lfloor s \rfloor \in \mathbb{Z}$ tale che*
$$\lfloor s \rfloor \leq s < \lfloor s \rfloor + 1;$$
(6) (densità dei razionali) *siano $s, t \in \mathbb{R}$ con $s < t$. Allora esiste $r \in \mathbb{Q}$ tale che $s < r < t$;*
(7) (proprietà Archimedea) *siano $s, t \in \mathbb{R}$. Se $0 < s$ esiste $n \in \mathbb{N}$ tale che $t < ns$.*

DIMOSTRAZIONE. Siano r, s due numeri razionali, essendo del tutto evidente che $r \leq s$ se e solo se esiste $c \in \mathfrak{m}$ tale che $j(r) \leq j(s) + c$, ne consegue che la relazione $\leq$ tra numeri reali estende la usuale relazione di ordine tra i numeri razionali. La relazione $s \leq s$ è del tutto ovvia.

(Antisimmetria) Supponiamo $s \leq t$ e $t \leq s$. Siccome $s \leq t$, possiamo allora trovare $a, b \in A$ tali che $s = [a]$, $t = [b]$ e $a \leq b$, mentre da $t \leq s$ segue che esiste $c \in \mathfrak{m}$ tale che $b \leq a + c$. Dunque $0 \leq b - a \leq c$ e questo implica $b - a \in \mathfrak{m}$.

(Transitività) Se $s \leq t$ e $t \leq u$ per ogni $a \in A$ tale che $s = [a]$ possiamo trovare $b \in A$ tale che $t = [b]$ e $a \leq b$, e poi possiamo trovare $c \in A$ tale che $[c] = u$ e $b \leq c$. Dunque $a \leq c$ e questo implica $s \leq t$.

(Totalità) Scegliamo due rappresentanti per s, t, diciamo $s = [a]$, $t = [b]$, $a, b \in R$. Sappiamo che esiste una costante N tale che
$$|a_n - a_{n+k}| \leq \frac{N}{10^n}, \qquad |b_n - b_{n+k}| \leq \frac{N}{10^n}$$
per ogni n, k. Supponiamo che $s \leq t$ sia falso, allora per ogni $c \in \mathfrak{m}$ esiste un intero m tale che $a_m \geq b_m + c_m$. Prendendo $c_n = \frac{2N}{10^n}$ per la disuguaglianza triangolare $a_n \geq b_n$ per ogni $n \geq m$, ossia $T^m b \leq T^m a$, che implica $t = [T^m b] \leq [T^m a] = s$.

(Parte intera) Sia $a \in R$ tale che $[a] = s$. Per il Lemma 13.3.4 esiste $M \in \mathbb{Q}$ tale che $|a_n - a_0| \leq M$ per ogni n e quindi $j(a_0 - M) \leq a \leq j(a_0 + M)$, che implica $a_0 - M \leq s \leq a_0 + M$. Si deduce che l'insieme S degli interi $m \leq s$ è non vuoto (contiene ogni intero minore di $a_0 - M$) ed è limitato superiormente da $a_0 + M$. Basta quindi definire $\lfloor s \rfloor$ come il massimo di S.

(Densità) Siano $a, b \in R$ tali che $s = [a]$, $t = [b]$ e $a \leq b$. Sappiamo che esiste una costante N tale che
$$|a_n - a_{n+k}| \leq \frac{N}{10^n}, \qquad |b_n - b_{n+k}| \leq \frac{N}{10^n},$$
per ogni n, k e, dato che $b - a \notin \mathfrak{m}$ esiste un indice m tale che $b_m > a_m + \frac{2N}{10^m}$. Dunque
$$a_n \leq a_m + \frac{N}{10^m}, \qquad b_m - \frac{N}{10^m} \leq b_n \quad \text{per ogni } n \geq m,$$

$$T^m a \le j(a_m + \frac{N}{10^m}), \qquad j(b_m - \frac{N}{10^m}) \le T^m b,$$

che implicano le disuguaglianze

$$s \le a_m + \frac{N}{10^m}, \qquad b_m - \frac{N}{10^m} \le t.$$

È adesso del tutto chiaro che il numero razionale $r = \frac{1}{2}(a_m + b_m)$ ha le proprietà richieste: infatti basta osservare che $a_m + \frac{N}{10^m} < r < b_m - \frac{N}{10^m}$.

(Proprietà Archimedea) Per i punti precedenti esistono $r, q \in \mathbb{Q}$ tali che $0 < r < s$ e $t \le q$ e basta prendere un qualsiasi intero positivo n tale che $nr \ge q$. □

Esempio 13.3.10. Mostriamo che l'applicazione

$$f\colon \mathbb{R} \to \mathcal{P}(\mathbb{N}), \qquad f(x) = \{n \in \mathbb{N} \mid \lfloor 10^{n+1} x \rfloor > 10 \lfloor 10^n x \rfloor \},$$

è surgettiva. Sia $S \subseteq \mathbb{N}$ un qualunque sottoinsieme e denotiamo, per ogni $n \in \mathbb{N}$, con

$$S_n = \{s \in S \mid s < n\}, \qquad a_n = \sum_{s \in S_n} \frac{1}{10^{s+1}}.$$

È chiaro che per ogni n vale

$$a_n \le a_{n+1} \le a_n + \frac{1}{10^{n+1}}$$

e quindi che la successione $a = (a_0, a_1, \ldots)$ è una serie decimale unitaria, che definisce pertanto un numero reale $x = [a]$. Siccome

$$S = \{n \mid a_{n+1} > a_n\} = \{n \mid 10^{n+1} a_{n+1} > 10(10^n a_n)\}$$

basta provare che $\lfloor 10^n x \rfloor = 10^n a_n$ per ogni $n \ge 0$. Si ha

$$10^n a_n \le 10^n a_{n+k} \le 10^n a_n + \sum_{i=1}^{k} \frac{1}{10^i} \le 10^n a_n + \frac{1}{9} < 10^n a_n + 1$$

per ogni $k \ge 0$ e questo implica che $10^n a_n \le 10^n x < 10^n a_n + 1$.

Lemma 13.3.11. *Siano n un intero positivo e $A, B \subseteq \mathbb{Q}$ due sottoinsiemi non vuoti tali che $a \le b + 1/n$ per ogni $a \in A$ ed ogni $b \in B$. Allora esiste un numero razionale $r \in \mathbb{Q}$ tale che*

$$a - \frac{1}{n} \le r \le b + \frac{1}{n}$$

per ogni $a \in A$ ed ogni $b \in B$.

Dimostrazione. Scegliamo $a_0 \in A$, $b_0 \in B$ e consideriamo l'insieme

$$S = \{m \in \mathbb{Z} \mid m \le nb \quad \forall\, b \in B\}.$$

Tale insieme è non vuoto in quanto contiene la parte intera di $na_0 - 1$ ed è limitato superiormente in quanto ogni $m \in S$ è minore di nb_0. Detto $N = \max(S)$, il numero $r = \frac{N+1}{n}$ è quello cercato. Per costruzione $r \le b + \frac{1}{n}$ per ogni $b \in B$ ed esiste un $b_1 \in B$ tale che $r = \frac{N+1}{n} > b_1$. Se per assurdo esiste un $a \in A$ tale che $a > r + \frac{1}{n} = \frac{N+2}{n}$, allora si avrebbe $a > b_1 + \frac{1}{n}$ in contrasto con le ipotesi. □

Teorema 13.3.12 (principio di completezza). *Siano $S, T \subseteq \mathbb{R}$ due sottoinsiemi non vuoti tali che $s \le t$ per ogni $s \in S$ ed ogni $t \in T$. Allora esiste un numero reale $\xi \in \mathbb{R}$ tale che*

$$s \le \xi \le t$$

per ogni $s \in S$ ed ogni $t \in T$.

Dimostrazione. Poiché ogni numero reale è rappresentato da una serie decimale unitaria, possiamo scegliere due sottoinsiemi $A, B \subseteq R$ tali che

$$S = \{[a] \mid a \in A\}, \qquad |a_n - a_{n+1}| \le \frac{1}{10^n} \quad \text{per ogni } \ a \in A,$$

$$T = \{[b] \mid b \in B\}, \qquad |b_n - b_{n+1}| \le \frac{1}{10^n} \quad \text{per ogni } \ b \in B.$$

Osserviamo adesso che per ogni intero n ed ogni $a \in A$, $b \in B$ si ha:

$$a_n \le b_n + \frac{3}{10^n}.$$

Infatti, se per assurdo esistesse un indice m tale che $a_m > b_m + \frac{3}{10^m}$, allora per ogni $l > m$ si avrebbe

$$a_l \geq a_m - \frac{10}{9}\frac{1}{10^m}, \qquad b_l < b_m + \frac{10}{9}\frac{1}{10^m}, \qquad a_l > b_l + \frac{7}{9}\frac{1}{10^m}$$

in contraddizione con il fatto che $[a] \leq [b]$. Per il Lemma 13.3.11 possiamo trovare una successione r_n di numeri razionali tali che

$$a_n - \frac{3}{10^n} \leq r_n \leq b_n + \frac{3}{10^n} \qquad \text{per ogni } \ a \in A, b \in B;$$

Adesso costruiamo in maniera ricorsiva un'altra successione c_n di numeri razionali tale che per ogni intero positivo n si abbiano le diseguaglianze:

$$|c_n - c_{n+1}| \leq \frac{4}{10^n}, \qquad a_n - \frac{3}{10^n} \leq c_n \leq b_n + \frac{3}{10^n} \qquad \text{per ogni } \ a \in A, b \in B.$$

Per iniziare poniamo $c_1 = r_1$ e supponiamo di aver definito $c_1, \ldots, c_n$. Per ogni $a \in A$ ed ogni $b \in B$ si ha

$$c_n + \frac{4}{10^n} \geq a_n + \frac{1}{10^n} \geq a_{n+1}$$
$$b_{n+1} \geq b_n - \frac{1}{10^n} \geq c_n - \frac{4}{10^n}.$$

Possiamo quindi definire

$$c_{n+1} = \min\left(r_{n+1}\,,\, c_n + \frac{4}{10^n}\right) \qquad \text{se} \quad r_{n+1} \geq c_n,$$
$$c_{n+1} = \max\left(r_{n+1}\,,\, c_n - \frac{4}{10^n}\right) \qquad \text{se} \quad r_{n+1} \leq c_n.$$

In definitiva $c = (c_1, c_2, \ldots) \in R$ e $\xi = [c]$ è il numero cercato. □

Per concludere, mostriamo che è possibile rappresentare ogni numero reale s mediante uno sviluppo decimale e che, se s non è una frazione decimale, allora un tale sviluppo è unico. Per definizione, una frazione decimale è un numero razionale della forma $\frac{m}{10^n}$, con n, m numeri interi, che possiede due sviluppi decimali, uno finito ed un altro infinito periodico con periodo 9. Supponiamo quindi che $s \in \mathbb{R}$ non sia una frazione decimale e, a meno di moltiplicazione per -1, possiamo supporre $s > 0$. In tali ipotesi esiste una unica successione di numeri naturali $m_0, m_1, \ldots \in \mathbb{N}$ tali che

$$\frac{m_n}{10^n} < s < \frac{m_n + 1}{10^n} \quad \text{per ogni} \quad n \in \mathbb{N}.$$

Ne consegue che

$$10m_{n-1} \leq m_n < 10(m_{n-1} + 1) \iff 0 \leq \alpha_n = m_n - 10m_{n-1} \leq 9 \quad \text{per ogni } n > 0,$$

e che s è rappresentato dallo sviluppo decimale $m_0{,}\alpha_1\alpha_2 \ldots$. Lasciamo per esercizio al lettore che tale sviluppo è unico e non è periodico di periodo 9.

Esercizi.

698. Siano $s, t \in \mathbb{R}$, $s, t \geq 0$. Provare che $st \geq 0$.

699. Per ogni $a = (a_0, a_1, \ldots) \in \mathbb{Q}^{\mathbb{N}}$ definiamo $|a| = (|a_0|, |a_1|, \ldots)$. Provare che per ogni $a \in R$ si ha $[|a|]$ è uguale al valore assoluto del numero reale $[a]$.

700. Date due successioni $a, b \in \mathbb{Q}^N$ ed un numero razionale $0 < c < 1$, scriveremo:

(1) $a \preceq_c b$ se esiste una costante $M \in \mathbb{Q}$ tale che $a_n \leq b_n + Mc^n$ per ogni n;
(2) $a \sim_c b$ se esiste una costante $M \in \mathbb{Q}$ tale che $|a_n - b_n| \leq Mc^n$ per ogni n;

Definiamo poi

$$R_c = \{a \in \mathbb{Q}^N \mid a \sim_c Ta\}, \qquad \mathfrak{m}_c = \{a \in \mathbb{Q}^N \mid a \sim_c 0\}.$$

Dimostrare che:

(1) $R_c, \mathfrak{m}_c$ sono sottospazi vettoriali di $\mathbb{Q}^{\mathbb{N}}$ e $\mathfrak{m}_c \subseteq R_c$;
(2) se $0 < c \leq d < 1$, le inclusioni naturali $R_c \subseteq R_d$ e $\mathfrak{m}_c \subseteq \mathfrak{m}_d$ inducono un isomorfismo di spazi vettoriali quoziente

$$f\colon \frac{R_c}{\mathfrak{m}_c} \to \frac{R_d}{\mathfrak{m}_d}.$$

701 (☕, ♡). Provare che non esistono quattro polinomi $p_1(x), p_2(x), p_3(x), p_4(x)$ a coefficienti reali tali che:

(1) $p_1(a) < p_2(a) < p_3(a) < p_4(a)$ per piccoli valori di $a > 0$, $a \in \mathbb{R}$,
(2) $p_3(a) < p_1(a) < p_4(a) < p_2(a)$ per piccoli valori di $a < 0$, $a \in \mathbb{R}$.

13.4. Complementi: insiemi ordinati e lemma di Zorn

Come preannunciato, diamo una dimostrazione del Teorema 12.5.4 come conseguenza dell'assioma della scelta; diciamo subito che occorre fare molta attenzione alla notazione, ed in particolare al fatto che la scrittura $A \subset B$ indica che A è un sottoinsieme proprio di B, ossia:

$$A \subset B \iff A \subseteq B \text{ e } A \neq B.$$

Ricordiamo che una famiglia $\mathcal{F}$ di sottoinsiemi di X si dice strettamente induttiva se, per ogni catena non vuota $\mathcal{C} \subseteq \mathcal{F}$, l'unione di tutti gli elementi di $\mathcal{C}$ appartiene ad $\mathcal{F}$.

TEOREMA 13.4.1. *Siano X un insieme, $\mathcal{F}$ una famiglia strettamente induttiva di sottoinsiemi di X e sia $f\colon \mathcal{F} \to \mathcal{F}$ un'applicazione tale che $A \subseteq f(A)$ per ogni $A \in \mathcal{F}$. Allora, per ogni $Q \in \mathcal{F}$ esiste $P \in \mathcal{F}$ tale che $Q \subseteq P$ e $f(P) = P$.*

DIMOSTRAZIONE. Basta trovare un elemento $P \in \mathcal{F}$ tale che $Q \subseteq P$ e $f(P) \subseteq P$. Chiameremo *torre* una qualunque sottofamiglia $\mathcal{A} \subseteq \mathcal{F}$ che soddisfa le seguenti condizioni S1, S2 ed S3:

S1: $Q \in \mathcal{A}$.
S2: $f(\mathcal{A}) \subseteq \mathcal{A}$.
S3: Se $\mathcal{A}$ contiene una catena non vuota $\mathcal{C}$, allora $\bigcup_{A \in \mathcal{C}} A \in \mathcal{A}$.

Indichiamo con $\mathbf{T}$ la collezione delle torri in $\mathcal{F}$; la famiglia $\mathbf{T}$ non è vuota perché contiene $\mathcal{F}$. Osserviamo che

$$\mathcal{M} = \bigcap_{\mathcal{A} \in \mathbf{T}} \mathcal{A}$$

soddisfa le tre condizioni precedenti e pertanto $\mathcal{M} \in \mathbf{T}$. Notiamo anche che la famiglia $\{B \in \mathcal{F} \mid Q \subseteq B\}$ appartiene ad $\mathbf{T}$ e quindi $Q \subseteq B$ per ogni $B \in \mathcal{M}$.

Diremo, provvisoriamente, che un sottoinsieme T di X è *buono* se $T \in \mathcal{M}$ e se per ogni $A \in \mathcal{M}$, $A \subset T$, vale $f(A) \subseteq T$; ad esempio Q è un sottoinsieme buono.

LEMMA 13.4.2. *Per ogni sottoinsieme buono $T \in \mathcal{M}$ e per ogni $A \in \mathcal{M}$ si ha $A \subseteq T$ oppure $f(T) \subseteq A$.*

DIMOSTRAZIONE. Sia $T \in \mathcal{M}$ buono e consideriamo la famiglia

$$\mathcal{N} = \{A \in \mathcal{M} \mid A \subseteq T \text{ oppure } f(T) \subseteq A\}.$$

Siccome $\mathcal{N} \subseteq \mathcal{M}$, per dimostrare che $\mathcal{N} = \mathcal{M}$ basta mostrare che $\mathcal{N}$ soddisfa le proprietà S1, S2 ed S3. Per quanto riguarda S1 abbiamo già osservato che $Q \subseteq B$ per ogni $B \in \mathcal{M}$; in particolare $Q \subseteq T$ e quindi $Q \in \mathcal{N}$.

Per dimostrare S2, ossia che $f(A) \in \mathcal{N}$ per ogni $A \in \mathcal{N}$ consideriamo la seguente casistica:

(1) se $A \subset T$ allora, dato che T è buono si ha $f(A) \subseteq T$ e quindi $f(A) \in \mathcal{N}$.
(2) se $A = T$, allora $f(T) \subseteq f(T)$ e quindi $f(T) \in \mathcal{N}$.
(3) se $f(T) \subseteq A$, allora $f(T) \subseteq A \subseteq f(A)$ e quindi $f(A) \in \mathcal{N}$.

Per dimostrare S3, sia $\mathcal{C}$ una catena non vuota contenuta in $\mathcal{N}$ e sia $H = \bigcup_{A \in \mathcal{C}} A \in \mathcal{M}$. Se $H \subseteq T$ allora $H \in \mathcal{N}$, altrimenti esiste $A \in \mathcal{C}$ che non è contenuto in T e quindi deve valere $f(T) \subseteq A$; a maggior ragione $f(T) \subseteq H$ e quindi $H \in \mathcal{N}$. □

LEMMA 13.4.3. *Ogni elemento di $\mathcal{M}$ è un sottoinsieme buono.*

DIMOSTRAZIONE. Indichiamo con $\mathcal{T}$ la famiglia dei sottoinsiemi buoni di $\mathcal{M}$:

$$\mathcal{T} = \{T \in \mathcal{M} \mid f(A) \subseteq T \text{ per ogni } A \in \mathcal{M},\ A \subset T\}.$$

Come nel precedente lemma, per dimostrare che $\mathcal{T} = \mathcal{M}$ basta mostrare che $\mathcal{T}$ soddisfa le condizioni S1,S2 ed S3; abbiamo già osservato che Q è buono e quindi che $\mathcal{T}$ soddisfa S1. Per quanto riguarda S2 occorre dimostrare che se T è buono, allora anche $f(T)$ è buono, ossia

che $f(A) \subseteq f(T)$ per ogni $A \in \mathcal{M}$, $A \subset f(T)$. Per il Lemma 13.4.2 le condizioni $A \in \mathcal{M}$, $A \subset f(T)$, implicano che $A \subseteq T$ e quindi basta applicare f per ottenere $f(A) \subseteq f(T)$.

Mostriamo adesso che $\mathcal{T}$ soddisfa la condizione S3. Siano $\mathcal{C} \subseteq \mathcal{T}$ una catena e $H = \bigcup_{A\in\mathcal{C}} A \in \mathcal{M}$. Dobbiamo provare che H è buono, e cioè che se $A \in \mathcal{M}$ e $A \subset H$, allora $f(A) \subseteq H$. Siccome H non è contenuto in A esiste $T \in \mathcal{C}$ tale che $T \not\subseteq A$ e quindi, a maggior ragione $f(T) \not\subseteq A$. Per il Lemma 13.4.2 si ha $A \subseteq T$, che assieme alla condizione $T \not\subseteq A$ implica $A \subset T$. Dato che T è buono si ha $f(A) \subseteq T \subseteq H$.

□

Tornando alla dimostrazione del Teorema 13.4.1, osserviamo che è sufficiente dimostrare che $\mathcal{M}$ è una catena di X. Infatti, in tal caso per le proprietà S3 ed S2 si ha

$$P := \bigcup_{A\in\mathcal{M}} A \in \mathcal{M}, \qquad f(P) \in \mathcal{M},$$

e quindi $f(P) \subseteq P$.

Siano dunque $S, T \in \mathcal{M}$ e supponiamo che $S \not\subseteq T$. Per il Lemma 13.4.3 T è buono e dunque, per il Lemma 13.4.2, si ha $f(T) \subseteq S$; siccome $T \subseteq f(T)$ a maggior ragione si ha $T \subseteq S$. □

Siamo adesso in grado di dimostrare il Teorema 12.5.4 come conseguenza del Teorema 13.4.1 e dell'assioma della scelta. Consideriamo la famiglia $\mathcal{R} \subseteq \mathcal{F} \times \mathcal{F}$ formata dalle coppie (A, B) tali che $A \subset B$. Se, per assurdo, $\mathcal{F}$ non contiene elementi massimali, allora la proiezione sul primo fattore $\mathcal{R} \to \mathcal{F}$ è surgettiva e quindi, per l'assioma della scelta, possiamo trovare un'applicazione $f\colon \mathcal{F} \to \mathcal{F}$ tale che

$$A \subset f(A), \qquad A \neq f(A),$$

per ogni $A \in \mathcal{F}$, in contraddizione con il Teorema 13.4.1.

Vediamo adesso un'altra formulazione equivalente del principio del massimo, estremamente utile e nota ai matematici con il nome di Lemma di Zorn.

Un **ordinamento** in un insieme X è una relazione binaria $\leq$ che soddisfa le tre proprietà:

Riflessiva: *$x \leq x$ per ogni $x \in X$.*
Antisimmetrica: *se $x \leq y$ e $y \leq x$, allora $x = y$.*
Transitiva: *se $x \leq y$ e $y \leq z$, allora $x \leq z$.*

Un ordinamento viene anche detto una **relazione d'ordine**. Se $\leq$ è un ordinamento si definisce $x < y$ se $x \leq y$ e $x \neq y$.

Esempio 13.4.4. Sia Y un insieme; dati due sottoinsiemi $A, B \subseteq Y$ definiamo

$$A \leq B \quad \text{se} \quad A \subseteq B.$$

La relazione $\leq$ è un ordinamento su $\mathcal{P}(Y)$ detto di *inclusione*.

Un ordinamento su X si dice **totale** se per ogni $x, y \in X$ vale $x \leq y$ oppure $y \leq x$. Un **insieme ordinato** è un insieme dotato di un ordinamento; un **insieme totalmente ordinato** è un insieme dotato di un ordinamento totale.[1] Ad esempio, i numeri reali, con la consueta relazione di ordine è un insieme totalmente ordinato.

Ogni sottoinsieme di un insieme ordinato è a sua volta un insieme ordinato, con la relazione di ordine indotta.

Definizione 13.4.5. Sia $(X, \leq)$ un insieme ordinato: Un sottoinsieme $C \subseteq X$ si dice una **catena** se per ogni $x, y \in C$ vale $x \leq y$ oppure $x \geq y$. In altri termini $C \subseteq X$ è una catena se e solo se C è un insieme totalmente ordinato per la relazione di ordine indotta.

Corollario 13.4.6. *Sia $(X, \leq)$ un insieme ordinato. La famiglia $\mathcal{F} \subseteq \mathcal{P}(X)$ di tutte le catene è strettamente induttiva e quindi possiede elementi massimali rispetto all'inclusione.*

[1] Questa definizione non è universalmente accettata: alcuni chiamano ordinamenti gli ordinamenti totali e ordinamenti parziali gli ordinamenti. Altri usano la parola **poset** (dall'inglese Partially Ordered SET) per indicare un insieme ordinato.

DIMOSTRAZIONE. Siccome $\mathcal{F}$ non è vuota (contiene la catena vuota), basta provare che è strettamente induttiva. Sia $\mathcal{C} \subseteq \mathcal{F}$ una catena e consideriamo due elementi

$$x_1, x_2 \in H := \bigcup_{A \in \mathcal{C}} A.$$

Dunque esistono $A_1, A_2 \in \mathcal{C}$ tali che $x_i \in A_i$ per ogni i. Siccome $\mathcal{C}$ è una catena, a meno di permutare gli indici possiamo supporre $A_1 \subseteq A_2$ quindi $x_1, x_2 \in A_2$; ma per ipotesi A_2 è una catena in X e quindi si ha $x_1 \leq x_2$ oppure $x_2 \leq x_1$. Questo prova che anche H è una catena in X, ossia che $H \in \mathcal{F}$.

□

DEFINIZIONE 13.4.7. Sia $(X, \leq)$ un insieme ordinato:

(1) Sia $C \subset X$ un sottoinsieme e $x \in X$. Diremo che x è un **maggiorante** di C se $x \geq y$ per ogni $y \in C$.
(2) Diremo che $m \in X$ è un **elemento massimale** di X se è l'unico maggiorante di se stesso, cioè se $\{x \in X \mid m \leq x\} = \{m\}$.

TEOREMA 13.4.8 (Lemma di Zorn). *Sia $(X, \leq)$ un insieme ordinato non vuoto. Se ogni catena in X possiede almeno un maggiorante, allora X possiede elementi massimali.*

DIMOSTRAZIONE. Per il Corollario 13.4.6 esiste una catena massimale $C \subseteq X$. Sia $m \in X$ un maggiorante per C e dimostriamo che m è un elemento massimale. Per assurdo, se m non è massimale esiste $x \in X$ tale che $m < x$. Allora il sottoinsieme $C \cup \{x\}$ è una catena che contiene strettamente C. □

Dunque il lemma di Zorn è una conseguenza del principio del massimo (Teorema 12.5.4 e quindi dell'assioma della scelta. Viceversa, è facile dimostrare che sia l'assioma della scelta che il principio del massimo sono conseguenze del lemma di Zorn: ad esempio, se X è un insieme e $\mathcal{F} \subseteq \mathcal{P}(X)$ una famiglia non vuota strettamente induttiva, allora $\mathcal{F}$ è un insieme ordinato rispetto all'inclusione. Per ogni catena $\mathcal{C} \subseteq \mathcal{F}$ l'insieme dei maggioranti è non vuoto ed ha come minimo l'unione di tutti gli elementi di $\mathcal{C}$. Per il Lemma di Zorn la famiglia $\mathcal{F}$ possiede elementi massimali e questo prova il principio del massimo.

Mostriamo adesso che il lemma di Zorn implica l'assioma della scelta: sia $g\colon Y \to X$ un'applicazione surgettiva di insiemi non vuoti e mostriamo che il lemma di Zorn implica l'esistenza di un'applicazione $f\colon X \to Y$ tale che $g(f(x)) = x$ per ogni $x \in X$. Introduciamo l'insieme $\mathcal{S}$ i cui elementi sono le coppie (E, f) tali che:

(1) $E \subset X$ è un sottoinsieme.
(2) $f\colon E \to Y$ è un'applicazione tale che $gf(x) = x$ per ogni $x \in E$.

L'insieme $\mathcal{S}$ non è vuoto, esso contiene infatti la coppia $(\emptyset, \emptyset \hookrightarrow Y)$. Su $\mathcal{S}$ è possibile ordinare gli elementi per *estensione*, definiamo cioè $(E, h) \leq (F, k)$ se k estende h: in altri termini $(E, h) \leq (F, k)$ se e solo se $E \subset F$ e $h(x) = k(x)$ per ogni $x \in E$. Mostriamo adesso che ogni catena in $\mathcal{S}$ possiede maggioranti. Sia $\mathcal{C} \subset \mathcal{S}$ una catena e consideriamo l'insieme

$$A = \bigcup_{(E,h) \in \mathcal{C}} E.$$

Definiamo poi $a\colon A \to Y$ nel modo seguente: se $x \in A$ allora esiste $(E, h) \in \mathcal{C}$ tale che $x \in E$, e si pone $a(x) = h(x)$. Si tratta di una buona definizione, infatti se $(F, k) \in \mathcal{C}$ e $x \in F$ si ha, poiché $\mathcal{C}$ è una catena $(E, h) \leq (F, k)$ oppure $(E, h) \geq (F, k)$. In entrambi i casi $x \in E \cap F$ e $h(x) = k(x)$. È chiaro che $(A, a) \in \mathcal{S}$ è un maggiorante di $\mathcal{C}$.

Per il lemma di Zorn esiste un elemento massimale $(U, f) \in \mathcal{S}$ e basta dimostrare che $U = X$. Se così non fosse esisterebbe $y \in Y$ tale che $g(y) \notin U$ e la coppia $(U \cup \{g(y)\}, f')$, che estende (U, f) e tale che $f'(g(y)) = y$, appartiene a $\mathcal{S}$ e contraddice la massimalità di (U, f).

Il lemma di Zorn ha moltissime applicazioni in matematica e rappresenta quindi uno strumento indispensabile del matematico. A titolo di esempio, useremo il lemma di Zorn per dimostrare il prossimo teorema di importanza fondamentale.

TEOREMA 13.4.9. *Ogni insieme possiede ordinamenti totali.*

DIMOSTRAZIONE. Ai soli fini di questa dimostrazione conviene interpretare un ordinamento totale su un insieme I come un sottoinsieme $A \subseteq I \times I$ che gode delle seguenti proprietà:

(1) $(x, x) \in A$ per ogni $x \in I$,
(2) se $(x, y) \in A$ e $(y, x) \in A$, allora $x = y$,
(3) se $(x, y) \in A$ e $(y, z) \in A$, allora $(x, z) \in A$,
(4) per ogni $x, y \in X$ vale $(x, y) \in A$ oppure $(y, x) \in A$.

È infatti chiaro che la relazione $x \leq y \iff (x, y) \in A$ è un ordinamento totale se e solo se il sottoinsieme $A \subseteq I \times I$ soddisfa le precedenti 4 condizioni.

Sia adesso H un qualunque insieme e indichiamo con X la famiglia di tutte le coppie (I, A) con $I \subseteq H$ sottoinsieme e $A \subseteq I \times I$ ordinamento totale. È immediato osservare X non è vuoto, infatti contiene la coppia $(\emptyset, \emptyset)$, e che la relazione di estensione

$$(I, A) \leq (J, B) \iff I \subseteq J \text{ e } A = B \cap (I \times I),$$

definisce un ordinamento su X. Volendo applicare il lemma di Zorn occorre verificare che ogni catena in X possiede maggioranti: questo è molto facile, infatti detta $C \subseteq X$ una qualunque catena un suo maggiorante è dato dalla coppia (J, B), dove

$$J = \cup_{(I,A)\in C} I, \qquad B = \cup_{(I,A)\in C} A.$$

Lasciamo per esercizio al lettore la semplice verifica che (J, B) è un elemento di X che maggiora la catena C.

Dunque X soddisfa le ipotesi del Lemma di Zorn, esiste quindi un elemento massimale (M, S) e per concludere la dimostrazione basta provare che $M = X$. Se esiste $x \in X - M$, definendo

$$N = M \cup \{x\}, \qquad R = S \cup (\{x\} \times N),$$

si ha $(N, R) \in X$ e $(M, S) \leq (N, R)$, in contraddizione con la massimalità di (M, R). □

Esercizi.

702. Sia X un insieme ordinato con la proprietà che ogni suo sottoinsieme non vuoto possiede massimo e minimo. Mostrare che X è finito e totalmente ordinato.

703. Sia $\prec$ una relazione su di un insieme X tale che:

(1) Per ogni $x, y \in X$, almeno una delle due relazioni $x \prec y$, $y \prec x$ è falsa.
(2) Se vale $x \prec y$ e $y \prec z$, allora $x \prec z$.

Dimostrare che la relazione

$$x \leq y \quad \iff \quad x \prec y \text{ oppure } x = y$$

è una relazione di ordine.

704. Sia V uno spazio vettoriale (non necessariamente di dimensione finita) e consideriamo l'insieme $G(V)$ dei suoi sottospazi vettoriali, ordinato per inclusione (Esempio 13.4.4). Dimostrare che per ogni sottoinsieme $X \subseteq V$ che contiene 0 la famiglia

$$\{F \in G(V) \mid F \subseteq X\}$$

possiede elementi massimali.

CAPITOLO 14

Fattorizzazione di polinomi e forma canonica razionale

Finora siamo stati in grado di trovare forme speciali e semplici di endomorfismi triangolabili, ossia di endomorfismi il cui polinomio caratteristico si decompone in un prodotto di polinomi di primo grado. Per il teorema fondamentale dell'algebra questo è sempre vero su campo dei numeri complessi, ma già su $\mathbb{R}$ è facile trovare esempi di matrici 2×2 che non sono triangolabili.

In questo capitolo mostreremo che è sempre possibile mettere un endomorfismo nella cosidetta **forma canonica razionale**, ossia si può trovare una base dello spazio vettoriale in cui l'endomorfismo si rappresenta con una matrice diagonale a blocchi, in cui ogni blocco è una matrice compagna di tipo particolare. Come conseguenza dimostreremo in particolare che il polinomio caratteristico divide una potenza del polinomio minimo.

14.1. Il massimo comune divisore di polinomi

Le stesse considerazioni usate nella Sezione 10.1 per introdurre la nozione di polinomio minimo possono essere utilmente utilizzate per introdurre il concetto di massimo comune divisore di polinomi.

Siano $p_1, \dots, p_n \in \mathbb{K}[t]$ dei polinomi, indichiamo con $(p_1, p_2, \dots, p_n) \subseteq \mathbb{K}[t]$ l'insieme di tutti i polinomi che si possono scrivere nella forma

$$f_1 p_1 + f_2 p_2 + \cdots + f_n p_n, \qquad \text{con} \quad f_1, \dots, f_n \in \mathbb{K}[t]. \tag{14.1}$$

Se i polinomi p_i non sono tutti nulli, allora $(p_1, p_2, \dots, p_n)$ contiene certamente dei polinomi monici: infatti, se ad esempio $p_1 \neq 0$ e c è il suo coefficiente direttivo, $p_1 = cx^a + \cdots$, allora ponendo $f_1 = 1/c$, $f_2 = \cdots = f_n = 0$, il polinomio

$$f_1 p_1 + f_2 p_2 + \cdots + f_n p_n = \frac{p_1}{c}$$

è monico.

Tra tutti i polinomi monici (e quindi non nulli) di $(p_1, p_2, \dots, p_n)$ scegliamone uno di grado minimo. Un tale polinomio è unico perché se ce ne fossero due

$$p = f_1 p_1 + \cdots + f_n p_n = t^h + a_1 t^{h-1} + \cdots + a_h,$$
$$q = g_1 p_1 + \cdots + g_n p_n = t^h + b_1 t^{h-1} + \cdots + b_h,$$

posto

$$r = (f_1 - g_1) p_1 + \cdots + (f_n - g_n) p_n = ct^s + \cdots, \qquad c \neq 0,$$

risulterebbe $s < h$ e il polinomio r/c sarebbe un polinomio monico di grado strettamente inferiore al grado di p. Ciò è assurdo a meno che non sia $q = p$.

Denoteremo con $\mathrm{MCD}(p_1, \dots, p_n)$ il polinomio monico di grado minimo che appartiene a $(p_1, \dots, p_n)$ e lo chiameremo **massimo comune divisore** di $p_1, \dots, p_n$ in $\mathbb{K}[t]$. Tale nome è giustificato dal seguente teorema:

Teorema 14.1.1. *Il massimo comune divisore p dei polinomi $p_1, \dots, p_n$ in $\mathbb{K}[t]$ è l'unico polinomio monico che soddisfa le seguenti condizioni:*

1. *p divide tutti i polinomi $p_1, \dots, p_n$;*
2. *se $g \in \mathbb{K}[t]$ divide tutti i polinomi $p_1, \dots, p_n$, allora divide anche p.*

Dimostrazione. Proviamo prima che $p = \mathrm{MCD}(p_1, \dots, p_n)$ soddisfa le condizioni. Indichiamo con $f_1, \dots, f_n \in \mathbb{K}[t]$ dei polinomi tali che $p = f_1 p_1 + \cdots + f_n p_n$.

(1) A meno di scambio di indici basta mostrare che p divide p_1. Per l'algoritmo di divisione si ha $p_1 = qp + r$, con $\deg(r) < \deg(p)$; se per assurdo fosse $r = ct^h + \cdots \neq 0$, allora

$$\frac{r}{c} = \frac{p_1 - qp}{c} = \frac{p_1 - qf_1}{c}p_1 + \frac{-qf_2}{c}p_2 + \cdots + \frac{-qf_n}{c}p_n$$

sarebbe un polinomio monico di grado inferiore scrivibile come in (14.1).

(2) Se g divide tutti i polinomi p_i, allora si ha $p_1 = gq_1$, $p_2 = gq_2$,..., $p_n = gq_n$ e quindi

$$p = f_1p_1 + \cdots + f_np_n = f_1q_1g + \cdots + f_nq_ng = (f_1q_1 + \cdots + f_nq_n)g.$$

Sia q un polinomio monico. Se q divide tutti i polinomi $p_1, \ldots, p_n$, allora dalla seconda condizione segue che q divide p. Se ogni polinomio che divide $p_1, \ldots, p_n$ divide anche q, questo vale in particolare per p. Abbiamo quindi dimostrato che se un polinomio monico q soddisfa entrambe le precedenti condizioni, allora p e q si dividono l'uno con l'altro e quindi devono coincidere. □

È utile osservare che dal Teorema 14.1.1 segue immediatamente che il massimo comune divisore di $p_1, \ldots, p_n$ è uguale al massimo comune divisore di p_1 e $\mathrm{MCD}(p_2, \ldots, p_n)$.

Esempio 14.1.2. Il Teorema 10.7.3 afferma che per ogni matrice quadrata A, il rapporto tra il suo polinomio caratteristico ed il suo polinomio minimo è uguale, a meno del segno, al massimo comune divisore dei coefficienti della matrice aggiunta di $A - tI$.

Definizione 14.1.3. Due polinomi si dicono **relativamente primi** se il loro massimo comune divisore è uguale a 1.

Per il calcolo effettivo del massimo comune divisore usiamo le stesse idee utilizzate nella riduzione di Gauss, ossia:

(1) Il massimo comune divisore è invariante per permutazione degli indici, ossia
$$\mathrm{MCD}(p_1, \ldots, p_n) = \mathrm{MCD}(p_{\sigma(1)}, \ldots, p_{\sigma(n)})$$
per ogni permutazione σ.

(2) Il massimo comune divisore è invariante per moltiplicazione dei polinomi per scalari diversi da 0, ossia
$$\mathrm{MCD}(p_1, \ldots, p_n) = \mathrm{MCD}(a_1p_1, \ldots, a_np_n), \qquad a_1, \ldots, a_n \in \mathbb{K} - \{0\}.$$

(3) Se $p_1 \neq 0$ e $p_i = 0$ per ogni $i > 1$, allora $(p_1, \ldots, p_n) = \{fp_1 \mid f \in \mathbb{K}[t]\}$ e quindi $\mathrm{MCD}(p_1, \ldots, p_n)$ è il polinomio monico associato a p_1.

(4) Se $p_1 \neq 0$ e $p_2 = h_1p_1 + r_1$, con $\deg(r_1) < \deg(p_1)$, allora
$$\mathrm{MCD}(p_1, p_2, p_3, \ldots, p_n) = \mathrm{MCD}(p_1, r_2, p_3, \ldots, p_n).$$
A tal fine basta dimostrare che $(p_1, p_2, p_3, \ldots, p_n) = (p_1, r_2, p_3, \ldots, p_n)$ come sottoinsiemi di $\mathbb{K}[t]$ e questo segue dalla relazione
$$f_1p_1 + f_2p_2 + \cdots + f_np_n = (f_1 + hf_2)p_1 + f_2r_2 + \cdots + f_np_n$$

A questo punto è chiaro come funziona l'algoritmo (detto di Euclide): al primo passo si ordinano i polinomi $p_1, \ldots, p_n$ in modo tale che
$$\deg(p_1) \leq \cdots \leq \deg(p_k), \qquad p_{k+1} = \cdots = p_n = 0.$$
Al secondo passo si sostituiscono i polinomi $p_2, \ldots, p_k$ con i loro resti per la divisione per p_1. Ciò fatto si ritorna al primo passo; siccome la somma dei gradi dei polinomi non nulli diminuisce ad ogni secondo passaggio, ad un certo punto la procedura termina con $p_2 = \cdots = p_n = 0$ ed in tal caso il polinomio monico associato a p_1 è il massimo comune divisore.

Esempio 14.1.4.
$$\begin{aligned}
\mathrm{MCD}(x^4 - 1, x^6 - 1) &= MCD(x^4 - 1, x^2 - 1) && [x^6 - 1 = x^2(x^4 - 1) + (x^2 - 1)] \\
&= \mathrm{MCD}(x^2 - 1, x^4 - 1) \\
&= \mathrm{MCD}(x^2 - 1, 0) && [x^4 - 1 = (x^2 + 1)(x^2 - 1)] \\
&= x^2 - 1.
\end{aligned}$$

L'algoritmo di Euclide dimostra anche che **il massimo comune divisore non dipende dalla scelta del campo** dove si considerano i coefficienti dei polinomi coinvolti.

TEOREMA 14.1.5. *Sia $\mathbb{K} \subset \mathbb{C}$ un campo di numeri e siano $f, g \in \mathbb{K}[t]$ due polinomi non nulli. Allora* $\mathrm{MCD}(f,g) = 1$ *se e solo se f e g non hanno radici comuni su $\mathbb{C}$.*

DIMOSTRAZIONE. Se $\mathrm{MCD}(f,g) = 1$ vuol dire che si può scrivere $1 = af + bg$ e quindi per ogni $\alpha \in \mathbb{C}$ si ha $a(\alpha)f(\alpha) + b(\alpha)g(\alpha) = 1$; in particolare $f(\alpha)$ e $g(\alpha)$ non possono essere entrambi uguali a 0. Viceversa se $\mathrm{MCD}(f,g) = h$ ha grado positivo, allora h divide sia f che g e quindi ogni radice complessa di h è una radice comune di f e g. □

Esercizi.

705. Calcolare il massimo comune divisore per ciascuna coppia dei seguenti polinomi a coefficienti razionali:

$$x^2 + 1, \qquad x^6 + x^3 + x + 1, \qquad x^4 + 1, \qquad x^7.$$

706. Siano $n, m > 1$ due interi positivi senza fattori comuni. Usare l'algoritmo di Euclide per dimostrare che

$$\mathrm{MCD}(1 + x + x^2 + \cdots + x^{n-1}, 1 + x + x^2 + \cdots + x^{m-1}) = 1.$$

707. Sia d il massimo comune divisore di due interi positivi $n, m > 1$. Provare che

$$\mathrm{MCD}(x^n - 1, x^m - 1) = x^d - 1.$$

708 (Divisori elementari). Per ogni matrice $A \in M_{n,n}(\mathbb{K})$ ed ogni $0 \le k \le n$ definiamo il polinomio $D_k(A,t) \in \mathbb{K}[t]$, detto *$k$-esimo divisore elementare* di A, come il massimo comune divisore dei determinanti dei minori di ordine k della matrice $A - tI$. Ad esempio: $D_n(A,t) = (-1)^n p_A(t)$; $D_k(\lambda I, t) = (t - \lambda)^k$; $D_1(A,t) = 1$ eccetto il caso in cui $A = \lambda I$; se A è una matrice compagna allora $D_k(A,t) = 1$ per ogni $k < n$. Segue immediatamente dagli sviluppi di Laplace che $D_k(A,t)$ divide $D_{k+1}(A,t)$ per ogni $0 \le k < n$.

Usare il teorema di Binet generalizzato (Esercizio 469) per provare che matrici simili hanno gli stessi divisori elementari.

709 (☕, ♡). Siano $p_1, \ldots, p_n \in \mathbb{K}[t]$ polinomi monici. Provare che esistono $n(n-1)$ polinomi $a_{ij} \in \mathbb{K}[t]$, con $i = 2, \ldots, n$, $j = 1, \ldots, n$ tali che

$$\mathrm{MCD}(p_1, \ldots, p_n) = \det \begin{pmatrix} p_1 & p_2 & \cdots & p_n \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

14.2. Polinomi irriducibili e fattorizzazione unica

Un polinomio $p(t) \in \mathbb{K}[t]$ si dice **invertibile** se esiste un altro polinomio $q(t)$ tale che $p(t)q(t) = 1$.

Se $p(t) = a_0 \neq 0$ ha grado 0, allora $p(t)$ è invertibile (basta considerare $q(t) = a_0^{-1}$); viceversa se $p(t)$ è invertibile allora

$$0 = \deg(1) = \deg(p(t)q(t)) = \deg(p(t)) + \deg(q(t))$$

da cui segue che $\deg(p(t)) = 0$.

Un polinomio $p(t) \in \mathbb{K}[t]$ si dice **riducibile** se può essere scritto come prodotto di due polinomi non invertibili.

Un polinomio $p(t) \in \mathbb{K}[t]$ si dice **irriducibile** se non è né invertibile né riducibile. In altri termini $p(t) \in \mathbb{K}[t]$ è irriducibile se ha grado ≥ 1 se se ogni volta che si scrive come prodotto $p(t) = a(t)b(t)$ ne consegue che $a(t)$ è invertibile oppure che $b(t)$ è invertibile.

Ad esempio ogni polinomio di grado 1 è irriducibile; per il teorema fondamentale dell'algebra, un polinomio in $\mathbb{C}[t]$ di grado almeno 2 è riducibile.

È chiaro che il concetto di irriducibilità dipende dal campo in cui consideriamo i coefficienti dei polinomi: ad esempio il polinomio $x^2 - 2$ è irriducibile in $\mathbb{Q}[t]$ ma diventa riducibile in $\mathbb{R}[t]$ dato che $x^2 - 2 = (x + \sqrt{2})(x - \sqrt{2})$. Similmente $x^2 + 1$ è irriducibile in $\mathbb{R}[t]$ ma diventa riducibile su $\mathbb{C}[t]$.

LEMMA 14.2.1. *Siano $p(t), f(t), g(t) \in \mathbb{K}[t]$ tre polinomi non nulli, con $p(t)$ irriducibile.*

(1) *se* $p(t)$ *non divide* $f(t)$*, allora* $\mathrm{MCD}(p(t), f(t)) = 1$*;*
(2) *se* $p(t)$ *divide il prodotto* $f(t)g(t)$*, allora* $p(t)$ *divide almeno uno dei due fattori* $f(t)$ *e* $g(t)$*.*

DIMOSTRAZIONE. Sia $h(t) = \mathrm{MCD}(p(t), f(t))$, siccome $h(t)$ divide il polinomio irriducibile $p(t)$, si ha $h(t) = 1$ oppure $h(t) = cp(t)$, per un opportuno $c \in \mathbb{K}$. Nel secondo caso, siccome $h(t)$ divide $f(t)$, ne segue che anche $p(t)$ divide $f(t)$.

Supponiamo adesso che $p(t)$ divide $f(t)g(t)$ e supponiamo per assurdo che $\mathrm{MCD}(p(t), f(t)) = 1$ e $\mathrm{MCD}(p(t), g(t)) = 1$. Esistono allora dei polinomi $a(t), b(t), c(t), d(t)$ tali che

$$a(t)p(t) + b(t)f(t) = 1, \qquad c(t)p(t) + d(t)g(t) = 1\ .$$

Moltiplicando membro a membro si ottiene

$$\begin{aligned} 1 &= (a(t)p(t) + b(t)f(t))(c(t)p(t) + d(t)g(t)) \\ &= (a(t)c(t)p(t) + a(t)d(t)g(t) + b(t)f(t)c(t))p(t) + (b(t)d(t))f(t)g(t) \end{aligned}$$

e quindi $1 = \mathrm{MCD}(p(t), f(t)g(t))$. □

TEOREMA 14.2.2 (Fattorizzazione unica). *Ogni polinomio di grado positivo si fattorizza in maniera essenzialmente unica come un prodotto di polinomi irriducibili. Unicità essenziale significa che se*

$$f = p_1 \cdots p_n = q_1 \cdots q_m$$

con i polinomi p_i, q_j *irriducibili, allora* $n = m$ *e, a meno di riordinare gli indici, per ogni* h *i polinomi* p_h *e* q_h *differiscono per una costante moltiplicativa.*

DIMOSTRAZIONE. Dimostrare l'esistenza è facile: se f è di grado positivo allora non è invertibile, quidi è irriducibile oppure riducibile. Nel primo caso abbiamo finito, nel secondo si ha $f = f_1 f_2$ con i polinomi f_1, f_2 entrambi di grado minore a quello di f. Adesso ripetiamo il ragionamento per f_1, f_2 e per i loro eventuali fattori. Siccome il grado non può diminuire indefinitamente, dopo un numero finito di passi troviamo una fattorizzazione

$$f = p_1 \cdots p_n$$

con i polinomi p_i irriducibili.

Mostriamo adesso l'unicità. Se

$$f = p_1 \cdots p_n = aq_1 \cdots q_m$$

con a una costante non nulla ed i polinomi p_i, q_j irriducibili, allora p_n divide il prodotto $q_1 \cdots q_m$ e quindi divide almeno uno dei fattori; a meno di riordinare gli indici possiamo supporre che p_n divide q_m, $q_m = p_n h$. Per ipotesi q_m è irriducibile e quindi h è una costante non nulla. Dividendo per p_n si ottiene quindi

$$p_1 \cdots p_{n-1} = (ah)q_1 \cdots q_{m-1}$$

e si ripete il ragionamento per p_{n-1} fino all'esaurimento di tutti i fattori. □

Una prima conseguenza del Teorema 14.2.2 è che, come avviene per i numeri interi, ogni polinomio $f \in \mathbb{K}[t]$ si può scrivere in maniera essenzialmente unica come

$$f = cp_1^{a_1} \cdots p_n^{a_n}$$

con $c \in \mathbb{K}$, $c \neq 0$, $a_1, \ldots, a_n > 0$ ed i polinomi p_i monici, irriducibili e distinti tra loro. Una tale scomposizione viene detta **fattorizzazione normalizzata**.

Esercizi.

710 (☕, Teorema di Sylvester). Siano $A \in M_{n,n}(\mathbb{K})$ e $B \in M_{m,m}(\mathbb{K})$ tali che i rispettivi polinomi caratteristici $p_A(t)$ e $p_B(t)$ siano senza fattori comuni di grado positivo. Dimostrare che il sistema lineare omogeneo

$$AX = XB, \qquad X \in M_{n,m}(\mathbb{K})\ ,$$

possiede solo la soluzione banale $X = 0$.

(Suggerimento: sia $V = \{X \in M_{n,m}(\mathbb{K}) \mid AX = XB\}$. Calcolare, in funzione di $p_A(t)$ e $p_B(t)$ i polinomi caratteristici degli endomorfismi

$$L_A, R_B \colon M_{n,m}(\mathbb{K}) \to M_{n,m}(\mathbb{K}), \qquad L_A(X) = AX, \quad R_B(X) = XB,$$

e mostrare che $L_A(V) \subseteq V$, $R_B(V) \subseteq V$).

711 (☕). Siano $\mathbb{K}$ un campo di caratteristica 0 e $p(t) \in \mathbb{K}[t]$ un polinomio di grado d.

1) Provare che esistono dei polinomi $p_1(t), \dots, p_n(t) \in \mathbb{K}[t]$ di gradi $\deg p_i(t) = d - i$ e tali che per ogni polinomio $h(t) \in \mathbb{K}[t]$ si ha

$$p(t + h(t)) = p(t) + p_1(t)h(t) + p_2(t)h(t)^2 + \cdots + p_d(t)h(t)^d.$$

2) Provare che se $p(t)$ non ha fattori multipli, allora $p(t)$ e $p_1(t)$ non hanno fattori comuni ed esiste una successione di polinomi $b_1(t), b_2(t), \dots$ tale che

$$p(t)^i \text{ divide } 1 - p_1(t)b_i(t)$$

per ogni intero positivo i.

3) Provare che se $p(t)$ non ha fattori multipli, allora esiste una successione di polinomi $a_1(t), a_2(t), \dots$ tale che

$$p(t)^{n+1} \text{ divide } p\left(t + a_1(t)p(t) + a_2(t)p(t)^2 + \cdots + a_n(t)p(t)^n\right)$$

per ogni intero positivo n.

712. Sia f un endomorfismo di uno spazio vettoriale di dimensione finita.

(1) Dedurre dal Teorema 10.2.4 che se un polinomio irriducibile divide il polinomio caratteristico di f, allora divide anche il polinomio minimo.

(2) Sia $h(t)$ un polinomio monico irriducibile di grado d che divide il polinomio caratteristico di f. Provare che $\operatorname{Ker}(h(f)) \neq 0$ e che per ogni $0 \neq v \in \operatorname{Ker}(h(f))$ i vettori $v, f(v), \dots, f^{d-1}(v)$ sono linearmente indipendenti. Dedurre che in una opportuna base, l'endomorfismo f si rappresenta con una matrice a blocchi

$$\begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$$

dove A è la matrice compagna di $h(t)$.

(3) Dedurre dai punti precedenti che se due endomorfismi f, g hanno lo stesso polinomio caratteristico e $p(t) \in \mathbb{K}[t]$ è un qualunque polinomio, allora anche gli endomorfismi $p(f), p(g)$ hanno lo stesso polinomio caratteristico.

14.3. Decomposizione primaria ed endomorfismi semisemplici

In tutta la sezione, il simbolo V denoterà uno spazio vettoriale di dimensione finita su di un campo $\mathbb{K}$.

LEMMA 14.3.1. *Siano $f\colon V \to V$ un endomorfismo e $h(t), k(t) \in \mathbb{K}[t]$ polinomi senza fattori comuni. Allora*

$$\operatorname{Ker}(h(f)) \cap \operatorname{Ker}(k(f)) = 0.$$

In particolare, se $h(t)$ non ha fattori in comune con il polinomio minimo $q_f(t)$, allora l'endomorfismo $h(f)$ è invertibile.

DIMOSTRAZIONE. Il sottospazio $H = \operatorname{Ker}(h(f)) \cap \operatorname{Ker}(k(f))$ è f-invariante e dunque il polinomio minimo di $f_{|H}\colon H \to H$ divide $h(t)$ e $k(t)$. Se $H \neq 0$, allora il polinomio minimo di $f_{|H}$ ha grado positivo. Se $h(t)$ e $q_f(t)$ non hanno fattori comuni si ha

$$\operatorname{Ker}(h(f)) = \operatorname{Ker}(h(f)) \cap \operatorname{Ker}(q_f(f)) = 0.$$

□

TEOREMA 14.3.2. *Sia $p(t) = q_1(t)^{a_1} \cdots q_s(t)^{a_s}$ la fattorizzazione normalizzata di un polinomio monico che annulla un endomorfismo $f\colon V \to V$. Per ogni sottospazio f-invariante $U \subseteq V$ esiste una decomposizione in somma diretta di sottospazi f-invarianti.*

$$U = (\operatorname{Ker}(q_1(f)^{a_1}) \cap U) \oplus \cdots \oplus (\operatorname{Ker}(q_s(f)^{a_s}) \cap U).$$

DIMOSTRAZIONE. Ponendo $h(t) = q_1(t)^{a_1}$ e $k(t) = p(t)/h(t) = q_2(t)^{a_2} \cdots q_s(t)^{a_s}$, i polinomi $h(t)$ e $k(t)$ non hanno fattori comuni. Per il Lemma 14.3.1 $\operatorname{Ker}(h(f)) \cap \operatorname{Ker}(k(f)) = 0$ ed a maggior ragione $(\operatorname{Ker}(h(f)) \cap U) \cap (\operatorname{Ker}(k(f)) \cap U) = 0$.

D'altra parte, $h(f)k(f) = 0$ e quindi è ben definita l'applicazione lineare $h(f)\colon U \to \operatorname{Ker}(k(f)) \cap U$ il cui nucleo è $\operatorname{Ker}(h(f)) \cap U$.

Per il teorema del rango si ha $\dim U \leq \dim(\operatorname{Ker}(k(f)) \cap U) + \dim(\operatorname{Ker}(h(f)) \cap U)$, e per la formula di Grassmmann $(\operatorname{Ker}(h(f)) \cap U) \oplus (\operatorname{Ker}(k(f)) \cap U) = U$. Adesso si procede per induzione su s ripetendo il ragionamento per il sottospazio $\operatorname{Ker}(k(f)) \cap U$ invariante per l'endomorfismo $f\colon \operatorname{Ker}(k(f)) \to \operatorname{Ker}(k(f))$ annullato dal polinomio $q_2(t)^{a_2} \cdots q_s(t)^{a_s}$. □

COROLLARIO 14.3.3 (Decomposizione primaria). *Sia $q_f(t) = q_1(t)^{a_1} \cdots q_s(t)^{a_s}$ la fattorizzazione normalizzata del polinomio minimo di endomorfismo $f\colon V \to V$. Allora si ha una decomposizione in somma diretta di sottospazi f-invarianti.*

$$V = \operatorname{Ker}(q_1(f)^{a_1}) \oplus \cdots \oplus \operatorname{Ker}(q_s(f)^{a_s}).$$

DIMOSTRAZIONE. Ovvia conseguenza del Teorema 14.3.3. □

Nelle notazioni del Corollario 14.3.3, i sottospazi $\operatorname{Ker}(q_i(f)^{a_i}) \subseteq V$ vengono detti **autospazi generalizzati** di f.

DEFINIZIONE 14.3.4. Un endomorfismo $f\colon V \to V$ si dice **semisemplice** se per ogni sottospazio f-invariante $U \subseteq V$ esiste un sottospazio f-invariante $W \subseteq V$ tale che $V = U \oplus W$.

TEOREMA 14.3.5. *Sia $q_f(t) = q_1(t)^{a_1} \cdots q_s(t)^{a_s}$ la fattorizzazione normalizzata del polinomio minimo di un endomorfismo $f\colon V \to V$. Allora f è semisemplice se e solo se $a_i = 1$ per ogni i.*

DIMOSTRAZIONE. Dimostriamo prima che tale condizione è necessaria. Supponiamo che $a_1 > 1$ e proviamo che il sottospazio proprio f-invariante

$$U = \operatorname{Ker}(q_1(f)^{a_1 - 1} \cdots q_s(f)^{a_s})$$

non possiede complementari f-invarianti. Se per assurdo $V = U \oplus W$ con $f(W) \subseteq W$, allora $q_1(f)(W) \subset U \cap W = 0$ e quindi

$$W \subseteq \operatorname{Ker}(q_1(f)) \subseteq \operatorname{Ker}(q_1(f)^{a_1 - 1} \cdots q_s(f)^{a_s}) = U.$$

Supponiamo adesso $a_i = 1$ per ogni i e sia $U \subseteq V$ un sottospazio f-invariante. Consideriamo prima il caso $s = 1$, ossia il caso in cui $q_f(t)$ è un polinomio irriducibile. Denotiamo con d il grado di $q_f(t)$, allora per ogni $v \in V$ il vettore $f^d(v)$ è combinazione lineare di $v, f(v), \ldots, f^{d-1}(v)$ e quindi il sottospazio

$$H(v) = \operatorname{Span}(v, f(v), \ldots, f^{d-1}(v))$$

è f-invariante di dimensione d. Sia $U \subseteq V$ un sottospazio f-invariante proprio e prendiamo un qualsiasi vettore $v_1 \in V - U$. Allora $U \cap H(v_1)$ è un sottospazio f-invariante di dimensione $< d$ e di conseguenza il polinomio minimo della restrizione di f a $U \cap H(v_1)$ ha grado $< d$ e divide $q_f(t)$; ma questo è possibile solo se $U \cap H(v_1) = 0$. Se $U \oplus H(v_1) = V$ abbiamo finito, altrimenti si ripete il ragionamento con un $v_2 \in V - U \oplus H(v_1)$ fino a quando arriviamo ad una decomposizione $V = U \oplus H(v_1) \oplus \cdots \oplus H(v_n)$.

Se $s > 1$, per il Teorema 14.3.2 si hanno due decomposizioni in somma diretta

$$U = (\operatorname{Ker}(q_1(f)) \cap U) \oplus \cdots \oplus (\operatorname{Ker}(q_s(f)) \cap U), \qquad V = \operatorname{Ker}(q_1(f)) \oplus \cdots \oplus \operatorname{Ker}(q_s(f)).$$

Quindi è sufficiente dimostrare che per ogni i il sottospazio, $\operatorname{Ker}(q_i(f)) \cap U$ possiede un complementare f-invariante in $\operatorname{Ker}(q_i(f))$ e basta osservare che il polinomio minimo della restrizione $f\colon \operatorname{Ker}(q_i(f)) \to \operatorname{Ker}(q_i(f))$ è esattamente $q_i(t)$. □

Esercizi.

713. Provare che un endomorfismo è diagonalizzabile se e solo se è triangolabile e semisemplice.

714 (Proiezioni primarie). Sia $q_f(t) = q_1(t)^{a_1} \cdots q_s(t)^{a_s}$ la fattorizzazione normalizzata del polinomio minimo di endomorfismo $f\colon V \to V$. Si assuma $s > 1$, siano $a(t), b(t) \in \mathbb{K}[t]$ polinomi tali che

$$a(t)q_1(t)^{a_1} + b(t)(q_2(t)^{a_2} \cdots q_s(t)^{a_s}) = 1.$$

e si consideri l'endomorfismo $p = b(f)q_2(f)^{a_2} \cdots q_s(f)^{a_s} \colon V \to V$. Provare che il nucleo di p è $\operatorname{Ker}(q_2(f)^{a_2}) \oplus \cdots \oplus \operatorname{Ker}(q_s(f)^{a_s})$ e che $p(v) = v$ per ogni $v \in \operatorname{Ker}(q_1(f)^{a_1})$.

715 (☕). Usare il risultato dell'Esercizio 711 per dimostrare che se il campo $\mathbb{K}$ è di caratteristica 0, allora per ogni endomorfismo $f\colon V \to V$ esiste un polinomio $h(t) \in \mathbb{K}[t]$ tale che $h(f)$ è semisemplice e $h(f) - f$ è nilpotente.

14.4. Spazi ciclici, irriducibili e indecomponibili

Per decomporre ulteriormente ed in maniera f-invariante gli addendi della decomposizione primaria di un endomorfismo abbiamo bisogno di introdurre alcuni nuovi concetti.

Definizione 14.4.1. Sia f un endomorfismo di uno spazio vettoriale V di dimensione n. Diremo che V è **f-ciclico** se esiste un vettore $v \in V$ tale che

$$(v, f(v), \dots, f^{n-1}(v)), \qquad n = \dim V,$$

è una base di V.

Lemma 14.4.2. *Per un endomorfismo $f\colon V \to V$ le seguenti condizioni sono equivalenti:*

1. *V è f-ciclico;*
2. *in una opportuna base, l'endomorfismo f è rappresentato da una matrice compagna;*
3. *esiste un vettore $v \in V$ tale che*

$$\operatorname{Span}(v, f(v), f^2(v), \dots) = V.$$

Dimostrazione. Un endomorfismo si rappresenta con una matrice compagna (Definizione 9.3.1) nella base $(v_1, \dots, v_n)$ se e solo se $v_{i+1} = f^i(v_1)$ per ogni $i = 1, \dots, n-1$; questo prova l'equivalenza tra le prime due condizioni.

Se la successione di vettori $v, f(v), \dots, f^m(v), \dots$ è un insieme di generatori, indichiamo con N il più piccolo intero tale che i vettori

$$v, f(v), \dots, f^N(v),$$

siano linearmente dipendenti e denotiamo $W = \operatorname{Span}(v, f(v), \dots, f^{N-1}(v))$; chiaramente i vettori $v, f(v), \dots, f^{N-1}(v)$ sono linearmente indipendenti. Dimostriamo per induzione su $r \geq 0$ che $f^{N+r}(v) \in W$; questo implicherà che $W = V$. Se $a_0 v + a_1 f(v) + \cdots + a_N f^N(v) = 0$ con gli a_i non tutti nulli, si deve avere $a_N \neq 0$ e quindi $f^N(v) \in W$. Per ogni $r > 0$ si ha

$$0 = f^r(0) = a_0 f^r(v) + a_1 f^{r+1}(v) + \cdots + a_N f^{N+r}(v)$$

e quindi $f^{N+r}(v) \in \operatorname{Span}(f^r(v), \dots, f^{N+r-1}(v))$. Per l'ipotesi induttiva $f^i(v) \in W$ per ogni $i < N + r$ e quindi

$$f^{N+r}(v) \in \operatorname{Span}(f^r(v), \dots, f^{N+r-1}(v)) \subset W.$$

□

Definizione 14.4.3. Sia $f\colon V \to V$ un endomorfismo di uno spazio vettoriale di dimensione finita sul campo $\mathbb{K}$. Diremo che V è **f-irriducibile** se gli unici sottospazi f-invarianti $W \subseteq V$ sono $W = 0$ e $W = V$.

Diremo che V è **f-indecomponibile** se non è possibile scrivere $V = U \oplus W$ con U e W sottospazi f-invarianti diversi da 0.

Se V è f-irriducibile allora è anche f-indecomponibile. Il viceversa è falso, in quanto possono esistere sottospazi f-invarianti che non ammettono complementari f-invarianti (vedi Esercizio 716).

L'obbiettivo di questa sezione è dimostrare il seguenti risultati.

TEOREMA 14.4.4. *Sia $f\colon V \to V$ un endomorfismo di uno spazio vettoriale di dimensione finita n sul campo $\mathbb{K}$. Sono condizioni equivalenti:*

(1) *V è f-indecomponibile,*
(2) *V è f-ciclico ed il polinomio minimo $q_f(t)$ è una potenza di un polinomio irriducibile,*
(3) *il polinomio minimo $q_f(t)$ ha grado $n = \dim V$ ed è una potenza di un polinomio irriducibile.*

Se le precedenti condizioni sono verificate e $q_f(t) = k(t)^m$, con $k(t) \in \mathbb{K}[t]$ irriducibile di grado $d = \dfrac{n}{m}$, allora

$$\dim \operatorname{Ker} k(f)^j = jd, \qquad \text{per ogni} \quad j = 1, \dots, m\,.$$

COROLLARIO 14.4.5. *Sia $f\colon V \to V$ un endomorfismo di uno spazio vettoriale di dimensione finita n sul campo $\mathbb{K}$. Sono condizioni equivalenti:*

(1) *V è f-irriducibile,*
(2) *V è f-ciclico ed il polinomio minimo $q_f(t)$ è irriducibile,*
(3) *il polinomio minimo $q_f(t)$ è irriducibile di grado $n = \dim V$,*
(4) *il polinomio caratteristico $p_f(t)$ è irriducibile.*

Le dimostrazioni sono abbastanza lunghe e non banali; per chiarezza espositiva le divideremo in una serie di risultati parziali.

LEMMA 14.4.6. *Se il polinomio caratteristico $p_f(t)$ di un endomorfismo $f\colon V \to V$ è irriducibile, allora V è f-irriducibile.*

DIMOSTRAZIONE. Se V non è irriducibile, esiste un sottospazio f-invariante $0 \neq A \subsetneq V$ ed il polinomio caratteristico della restrizione $f_{|A}\colon A \to A$ divide il polinomio caratteristico $p_f(t)$. □

LEMMA 14.4.7. *Se V è f-indecomponibile, allora il polinomio minimo $q_f(t)$ è una potenza di un polinomio irriducibile.*

DIMOSTRAZIONE. Se $q_f(t)$ non è una potenza di un polinomio irriducibile, allora possiamo scrivere $q_f(t) = h(t)k(t)$, dove $h(t)$ e $k(t)$ sono polinomi di grado positivo senza fattori comuni. L'endomorfismo $h(f)$ non è invertibile, altrimenti si avrebbe $k(f) = 0$ e per lo stesso motivo anche $k(f)$ non è invertibile. Dunque

$$\operatorname{Ker}(h(f)) \neq 0, \qquad \operatorname{Ker}(k(f)) \neq 0, \qquad \operatorname{Ker}(h(f)) \cap \operatorname{Ker}(k(f)) = 0.$$

D'altra parte esistono due polinomi $a(t), b(t)$ tali che $a(t)h(t) + b(t)k(t) = 1$ e quindi per ogni vettore $v \in V$ si ha

$$v = I(v) = (a(f)h(f) + b(f)k(f))(v) = a(f)h(f)(v) + b(f)k(f)(v)$$

e siccome $a(f)h(f)(v) \in \operatorname{Ker}(k(f))$ e $b(f)k(f)(v) \in \operatorname{Ker}(h(f))$ abbiamo provato che

$$V = \operatorname{Ker}(h(f)) \oplus \operatorname{Ker}(k(f))$$

in contraddizione con la f-irriducibilità di V. □

LEMMA 14.4.8. *Se V è f-indecomponibile, allora V è f-ciclico. In particolare il polinomio minimo di f coincide, a meno del segno, con il polinomio caratteristico.*

DIMOSTRAZIONE. Per il Lemma 14.4.7 sappiamo che esiste un polinomio irriducibile $k(t)$ ed un intero positivo m tale che $q_f(t) = k(t)^m$. Per ogni vettore $v \in V$ denotiamo con $\alpha(v)$ il più piccolo intero non negativo tale che $k(f)^{\alpha(v)}(v) = 0$:

$$\alpha(v) = \min\{a \mid k(f)^a(v) = 0\}.$$

Chiaramente $\alpha(v) \leq m$ per ogni vettore v e vale $\alpha(v) = 0$ se e solo se $v = 0$. Indichiamo inoltre con $A(v) \subseteq V$ il sottospazio vettoriale generato da $v, f(v), f^2(v), \dots$:

$$A(v) = \operatorname{Span}(v, f(v), f^2(v), \dots) = \{h(f)v \mid h(t) \in \mathbb{K}[t]\}.$$

Chiaramente $A(v)$ è un sottospazio f-invariante per ogni $v \in V$; dimostrare che V è f-ciclico equivale a provare che esiste un vettore $v \in V$ tale che $A(v) = V$.

Sia $n > 0$ il più piccolo intero tale che esistono n vettori $v_1 \dots, v_n \in V$ con la proprietà che

$$A(v_1) + A(v_2) + \cdots + A(v_n) = V \ . \tag{14.2}$$

Fra tutte le n-uple $v_1, \dots, v_n$ che soddisfano (14.2) scegliamone una $(v_1, \dots, v_n)$ tale che $\sum_{i=1}^{n} \alpha(v_i)$ sia minima. Dimostriamo che in tal caso vale

$$V = A(v_1) \oplus \cdots \oplus A(v_n).$$

A tal fine basta dimostrare che se $h_1(t), \dots, h_n(t) \in \mathbb{K}[t]$ e

$$h_1(f)v_1 + \cdots + h_n(f)v_n = 0$$

allora $h_i(f)v_i = 0$ per ogni i. Scriviamo $h_i(t) = k_i(t)k(t)^{b_i}$, con $b_i \geq 0$ e $k_i(t)$ non divisibile per $k(t)$; a meno di permutazioni degli indici possiamo supporre che $b_i < \alpha(v_i)$ se $i \leq r$ e $b_i \geq \alpha(v_i)$ se $i > r$; possiamo inoltre supporre $b_1 \leq b_2 \leq \cdots \leq b_r$. Siccome $h_i(f)v_i = 0$ per ogni $i > r$ basta dimostrare che $r = 0$; supponiamo per assurdo $r > 0$, allora si ha:

$$h_1(f)v_1 + \cdots + h_r(f)v_r = 0.$$

Poniamo

$$w_1 = k_1(f)v_1 + \sum_{i=2}^{r} k_i(f)k(f)^{b_i - b_1} v_i, \qquad W = A(w_1) + A(v_2) + \cdots + A(v_n).$$

Per il Lemma 14.3.1 l'endomorfismo $k_1(f)\colon V \to V$ è invertibile e W è un sottospazio f-invariante; in particolare $k_1(f)^{-1}(W) \subset W$, da ciò segue che $v_1 \in W$ e quindi $W = V$. Se $w_1 = 0$ contraddiciamo la minimalità di n, mentre se $w_1 \neq 0$ si ha $k(f)^{b_1}(w_1) = 0$ contraddicendo la minimalità di $\sum_{i=1}^{n} \alpha(v_i)$.

Adesso, per come abbiamo scelto n ogni $A(v_i)$ è un sottospazio f-invariante diverso da 0 e l'ipotesi di indecomponibilità implica $n = 1$.

Il fatto che polinomio minimo e caratteristico coincidono a meno di moltiplicazione per ± 1 segue dal Lemma 10.2.1.

□

Lemma 14.4.9. *Sia $f\colon V \to V$ un endomorfismo di uno spazio vettoriale di dimensione finita n sul campo $\mathbb{K}$. Se esiste un polinomio monico irriducibile $k(t) \in \mathbb{K}[t]$ tale che*

$$q_f(t) = (-1)^n p_f(t) = k(t)^m$$

per qualche $m > 0$, allora V è f-indecomponibile.

Dimostrazione. Se V non è indecomponibile, allora esiste una decomposizione $V = A \oplus B$ con A e B sottospazi invarianti di dimensione positiva. Siano $a(t)$ e $b(t)$ i polinomi caratteristici delle restrizioni di f ad A e B rispettivamente. Allora $p_f(t) = a(t)b(t)$ e per la fattorizzazione unica dei polinomi ne segue che $a(t) = \pm k(t)^{m_1}$, $b(t) = \pm k(t)^{m_2}$, con $m_1 + m_2 = m$. Per il teorema di Cayley–Hamilton, applicato alle restrizioni di f ai sottospazi A, B, il polinomio $k(t)^{\max(m_1, m_2)}$ annulla f e di conseguenza è divisibile per il polinomio minimo. Questo implica che $m_1 + m_2 \leq \max(m_1, m_2)$ e quindi che $A = 0$ oppure $B = 0$. □

Lemma 14.4.10. *Sia V uno spazio vettoriale di dimensione n e sia $g\colon V \to V$ un endomorfismo tale che $g^m = 0$, con $m > 0$. Allora:*

$$\dim \operatorname{Ker}(g^h) \geq \frac{hn}{m} \qquad \textit{per ogni} \quad h = 1, \dots, m.$$

Inoltre, se $\dim \operatorname{Ker}(g^{m-1}) \leq \dfrac{(m-1)n}{m}$, *allora* $\dim \operatorname{Ker}(g^h) = \dfrac{hn}{m}$ *per ogni* $h = 1, \dots, m$.

Dimostrazione. Siccome $g^m = 0$ la disuguaglianza è vera per $h = m$. Ponendo $\alpha_i = \dim \operatorname{Ker}(g^i) - \dim \operatorname{Ker}(g^{i-1})$, per il Lemma 11.1.1 si ha:

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_m \geq \alpha_{m+1} = 0, \qquad \sum_{i=1}^{m} \alpha_i = n.$$

Siano h un intero fissato $0 < h < m$, e $x = \dim \operatorname{Ker}(g^h)$; si ha $x = \alpha_1 + \cdots + \alpha_h$, $n - x = \alpha_{h+1} + \cdots + \alpha_m$ e quindi

$$\frac{x}{h} \geq \frac{n-x}{m-h} \quad \Rightarrow \quad x \geq \frac{hn}{m} \ .$$

Se $\alpha_1 > \alpha_m$, lo stesso ragionamento mostra che $x > \dfrac{hn}{m}$ per ogni $h = 1, \dots, m-1$. In particolare se $\dim \operatorname{Ker}(g^{m-1}) = \dfrac{(m-1)n}{m}$ ne consegue che $\alpha_1 = \cdots = \alpha_m = \dfrac{n}{m}$. □

LEMMA 14.4.11. *Se V è f-indecomponibile e $q_f(t) = k(t)^m$ con $k(t)$ irriducibile di grado d, allora $\dim \operatorname{Ker} k(f)^u = ud$ per ogni $u = 1, \dots, m$.*

DIMOSTRAZIONE. Sia n la dimensione di V, allora $dm = n$ e quindi, per il Lemma 14.4.10 applicato all'endomorfismo $g = k(f)$ è sufficiente dimostrare che

$$\dim \operatorname{Ker} k(f)^{m-1} \leq (m-1)d,$$

o, equivalentemente che il sottospazio f-invariante $H = k(f)^{m-1}(V) \neq 0$ ha dimensione $\geq d$. Adesso basta osservare che il polinomio caratteristico di $f_{|H} \colon H \to H$ divide il polinomio caratteristico di f e quindi è, a meno del segno, uguale ad una potenza di $k(t)$. □

Esercizi.

716. Provare che gli unici sottospazi f-invarianti dell'endomorfismo

$$f \colon \mathbb{K}^2 \to \mathbb{K}^2, \qquad f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ x \end{pmatrix},$$

sono $0, \mathbb{K}^2$ e $\{x = 0\}$.

717. Sia $f \colon V \to V$ un endomorfismo nilpotente. Dimostrare che V è f-indecomponibile se e solo se il nucleo di f ha dimensione 1; dimostrare che V è f-irriducibile se e solo se V ha dimensione 1.

718. Siano $f, g \colon V \to V$ endomorfismi tali che $fg = gf$. Dimostrare che se V è f-ciclico, allora $g = p(f)$ per qualche polinomio $p(t) \in \mathbb{K}[t]$.

719 (☕). Sia $f \colon V \to V$ un endomorfismo di uno spazio vettoriale di dimensione n e sia m il grado del polinomio minimo $q_f(t)$. Dimostrare:

(1) sia $q_f(t) = q_1(t)^{a_1} \cdots q_s(t)^{a_s}$, $a_i > 0$, la fattorizzazione normalizzata del polinomio minimo, allora per ogni $i = 1, \dots, s$ esiste un vettore $v_i \in \operatorname{Ker}(q_1(f)^{a_i}) - \operatorname{Ker}(q_1(f)^{a_i - 1})$;
(2) siano $v_1, \dots, v_s$ come al punto precedente e poniamo $v = v_1 + \cdots + v_s$. Allora i vettori $v, f(v), \dots, f^{m-1}(v)$ sono linearmente indipendenti;
(3) V è f-ciclico se e solo se $m = n$.

14.5. La forma canonica razionale

Il teorema di fattorizzazione unica per i polinomi ha un suo analogo per gli endomorfismi di uno spazio di dimensione finita. Per la precisione, in una opportuna base ogni endomorfismo si rappresenta con una matrice diagonale a blocchi, in cui ogni blocco è la matrice compagna di una potenza di un polinomio monico irriducibile; inoltre tali blocchi sono unici a meno dell'ordine.

LEMMA 14.5.1. *Sia $f \colon V \to V$ un endomorfismo di uno spazio vettoriale di dimensione finita sul campo $\mathbb{K}$. Esiste allora una decomposizione in somma diretta*

$$V = V_1 \oplus \cdots \oplus V_s$$

dove per ogni i si ha $V_i \neq 0$, $f(V_i) \subseteq V_i$ e V_i è f-indecomponibile.

DIMOSTRAZIONE. Si consideri la famiglia di tutte le decomposizioni in somma diretta $V = V_1 \oplus \cdots \oplus V_s$ con $V_i \neq 0$ e $f(V_i) \subseteq V_i$ per ogni i; si noti che per una tale decomposizione si ha $\sum_i \dim V_i = \dim V$ e quindi $s \leq \dim V$. Inoltre la famiglia non è vuota perché contiene la decomposizione $V = V_1$. Adesso scegliamo una decomposizione con s massimo e mostriamo che in tal caso ogni V_i è f-indecomponibile. Se V_1 non è f-indecomponibile allora si può scrivere $V_1 = U \oplus W$ con U, W f-invarianti e non nulli; dunque $V = U \oplus W \oplus V_2 \oplus \cdots \oplus V_s$, in contraddizione con la massimalità di s.

□

TEOREMA 14.5.2 (Forma canonica razionale). *Sia $f\colon V \to V$ un endomorfismo di uno spazio vettoriale di dimensione finita sul campo $\mathbb{K}$. Esiste allora una base nella quale f si rappresenta con una matrice diagonale a blocchi*

$$\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \vdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_s \end{pmatrix}$$

dove ogni A_i è una matrice compagna il cui polinomio minimo è una potenza di un polinomio irriducibile. Inoltre i blocchi $A_1, \dots, A_s$ sono unici a meno di permutazioni degli indici.

DIMOSTRAZIONE. L'esistenza segue immediatamente dal Lemma 14.5.1. Infatti si ha $V = V_1 \oplus \cdots \oplus V_s$ con $f(V_i) \subset V_i$ ed ogni V_i f-indecomponibile. Per il Teorema 14.4.4 ogni V_i possiede una base rispetto alla quale l'endomorfismo $f\colon V_i \to V_i$ si rappresenta con una matrice compagna A_i il cui polinomio caratteristico è, a meno del segno, una potenza di un polinomio irriducibile. Basta adesso considerare l'unione, per $i = 1, \dots, s$, di tali basi per ottenere la matrice diagonale a blocchi richiesta.

La prova dell'unicità richiede invece alcune considerazioni aggiuntive. Siano $k_1(t), \dots, k_r(t)$ i fattori monici irriducibili del polinomio caratteristico di f, siccome

$$p_f(t) = \prod p_{A_i}(t)$$

per ogni $i = 1, \dots, s$ esistono due indici j, l tali che $p_{A_i}(t) = \pm k_j(t)^l$. Siccome la matrice compagna A_i è univocamente determinata dalla coppia (j, l), per dimostrare l'unicità della decomposizione basta dimostrare che il numero di volte, contate con molteplicità, in cui una data coppia (j, l) compare dipende solo da f. Per semplicità notazionale trattiamo il caso $j = 1$, per gli altri indici $j = 2, \dots, r$ il ragionamento è del tutto simile. Indicando con d il grado di $k_1(t)$, la successione di numeri razionali

$$a_u = \frac{1}{d}(\dim \operatorname{Ker} k_1(f)^u - \dim \operatorname{Ker} k_1(f)^{u-1}), \qquad u > 0,$$

dipende solo dall'endomorfismo f. Restringendo l'attenzione ai sottospazi V_i si hanno tre possibilità:

(1) $\operatorname{Ker} k_1(f)^u \cap V_i = 0$ se $k_1(t)$ non divide $p_{A_i}(t)$;
(2) $\operatorname{Ker} k_1(f)^u \cap V_i = V_i$ se $p_{A_i}(t) = \pm k_1(t)^l$ con $u \geq l$;
(3) $\dim(\operatorname{Ker} k_1(f)^u \cap V_i) = ud$ se $p_{A_i}(t) = \pm k_1(t)^l$ con $u \leq l$.

È allora chiaro che a_u coincide con il numero di coppie $(1, l)$ tali che $u \geq l$ e quindi che $a_u - a_{u+1}$ è esattamente il numero di volte in cui compare la coppia $(1, u)$.

□

La dimostrazione dell'unicità che abbiamo appena ci dice anche che la forma canonica razionale di un endomorfismo f dipende solo dalle dimensioni dei nuclei degli endomorfismi $h(f)$ al variare di $h(t)$ tra i divisori monici del polinomio caratteristico. Essendo due endomorfismi con la stessa forma canonica razionale evidentemente coniugati abbiamo contemporaneamente dimostrato anche il seguente risultato.

TEOREMA 14.5.3. *Due endomorfismi $f, g\colon V \to V$ sono simili se e solo se hanno lo stesso polinomio caratteristico $p(t) = p_f(t) = p_g(t)$ e se gli endomorfismi $h(f), h(g)$ hanno lo stesso rango per ogni polinomio $h(t)$ che divide $p(t)$.*

OSSERVAZIONE 14.5.4. Rileggendo il Teorema 10.2.4 alla luce della fattorizzazione unica dei polinomi, abbiamo che ogni fattore irriducibile del polinomio caratteristico divide il polinomio minimo.

Questo fatto può essere provato anche come semplice corollario della forma canonica razionale; se $k(t)$ è monico irriducibile e divide il polinomio caratteristico di un endomorfismo $f\colon V \to V$, allora esiste un sottospazio $V_i \subset V$ che è f-invariante e tale che il polinomio minimo della restrizione $f\colon V_i \to V_i$ è una potenza di $k(t)$. Se il polinomio minimo $q_f(t)$ non fosse divisibile per $k(t)$, per il Lemma 14.3.1 l'applicazione lineare $q_f(f)\colon V_i \to V_i$ sarebbe invertibile.

COROLLARIO 14.5.5. *Siano $f\colon V \to V$ un endomorfismo e $q \in \mathbb{K}[t]$ polinomio. Allora* $\det(q(f)) = 0$ *se e solo se q ha fattori in comune con il polinomio caratteristico di f.*

DIMOSTRAZIONE. Il polinomio $q(t)$ ha fattori in comune con il polinomio caratteristico se e solo se ha fattori in comune con il polinomio minimo. □

Esercizi.

720. Sia V uno spazio vettoriale di dimensione finita sul campo $\mathbb{Q}$ dei numeri razionali e sia $f\colon V \to V$ un endomorfismo tale che $f^5 = Id$. Si assuma che f non abbia punti fissi non banali, ossia che $f(v) \neq v$ per ogni $v \neq 0$; dimostrare che la dimensione di V è un multiplo intero di 4.

721 (☕). Sia $f\colon V \to V$ un endomorfismo. Dimostrare che V è f-ciclico se e solo se ogni endomorfismo di V che commuta con f è del tipo $p(f)$ per qualche polinomio $p(t) \in \mathbb{K}[t]$.

14.6. Complementi: il risultante di due polinomi

Sia $\mathbb{K}$ un campo e siano $f(x), g(x) \in \mathbb{K}[x]$ polinomi di gradi n e m rispettivamente, diciamo

$$f(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n, \qquad g(x) = b_0x^m + b_1x^{m-1} + \cdots + b_m, \qquad a_0, b_0 \neq 0\,.$$

La **matrice di Sylvester** della coppia f, g è la matrice quadrata di ordine $n + m$ definita come

$$S(f(x), g(x)) = \begin{pmatrix} a_0 & a_1 & \cdot & a_n & & & \\ & a_0 & \cdot & \cdot & a_n & & \\ & & \ddots & \ddots & & \ddots & \\ & & & a_0 & a_1 & \cdot & a_n \\ b_0 & b_1 & \cdot & \cdot & b_m & & \\ & \ddots & \ddots & & & \ddots & \\ & & b_0 & b_1 & \cdot & \cdot & b_m \end{pmatrix}, \tag{14.3}$$

dove i coefficienti $a_0, \ldots, a_n$ sono posizionati obliquamente sulle prime m righe ed i coefficienti $b_0, \ldots, b_m$ sono posizionati obliquamente sulle ultime n righe. Ad esempio la matrice di Sylvester dei polinomi $a_0x^2 + a_1x + a_2$ e $b_0x^2 + b_1x + b_2$ è

$$\begin{pmatrix} a_0 & a_1 & a_2 & 0 \\ 0 & a_0 & a_1 & a_2 \\ b_0 & b_1 & b_2 & 0 \\ 0 & b_0 & b_1 & b_2 \end{pmatrix}.$$

Notiamo che le prime m righe di $S(f(x), g(x))$ rappresentano le coordinate dei polinomi

$$x^{m-1}f(x), \ldots, xf(x), f(x) \in \mathbb{K}[x]_{<n+m}$$

rispetto alla base $x^{n+m+1}, \ldots, x, 1$. Similmente le ultime n righe rappresentano le coordinate dei polinomi

$$x^{n-1}g(x), \ldots, xg(x), g(x) \in \mathbb{K}[x]_{<n+m}$$

nella medesima base.

DEFINIZIONE 14.6.1. Il **risultante** di due polinomi è il determinante della loro matrice di Sylvester:

$$R(f(x), g(x)) = |S(f(x), g(x))|.$$

Ad esempio, il risultante dei polinomi $x^2 - 2$ e $2x^2 - x$ è uguale a:

$$R(x^2 - 2, 2x^2 - x) = \begin{vmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 0 & -2 \\ 2 & -1 & 0 & 0 \\ 0 & 2 & -1 & 0 \end{vmatrix} = 14.$$

Abbiamo incontrato alcuni esempi di matrici di Sylvester negli Esercizi 423 e 446. In particolare dalla formula dell'Esercizio 446,

$$\begin{vmatrix} 1 & -\lambda & 0 & \dots & 0 \\ 0 & 1 & -\lambda & \dots & 0 \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & 0 & \dots & -\lambda \\ a_0 & a_1 & a_2 & \dots & a_n \end{vmatrix} = a_0\lambda^n + a_1\lambda^{n-1} + \cdots + a_n, \tag{14.4}$$

segue immediatamente che per ogni polinomio $g(x) \in \mathbb{K}[x]$ ed ogni $\lambda \in \mathbb{K}$ si ha

$$R(x - \lambda, g(x)) = g(\lambda). \tag{14.5}$$

TEOREMA 14.6.2. *Due polinomi non nulli $f(x), g(x) \in \mathbb{K}[x]$ hanno fattori comuni di grado positivo se e solo se il loro risultante si annulla. Equivalentemente, si ha* $\mathrm{MCD}(f(x), g(x)) = 1$ *se e solo se $R(f(x), g(x)) \neq 0$.*

DIMOSTRAZIONE. Siano n, m i gradi di $f(x), g(x)$. Abbiamo già osservato che $R(f(x), g(x)) \neq 0$ se e solo se gli $n + m$ polinomi

$$x^{m-1}f(x), \dots, xf(x), f(x), x^{n-1}g(x), \dots, xg(x), g(x)$$

formano una base dello spazio vettoriale $\mathbb{K}[x]_{<n+m}$ dei polinomi di grado $< n + m$. Dunque, se $R(f(x), g(x)) \neq 0$, possiamo trovare dei coefficienti $c_1, \dots, c_m, d_1, \dots, d_n \in \mathbb{K}$ tali che

$$c_1x^{m-1}f(x) + \cdots + c_mf(x) + d_1x^{n-1}g(x) + \cdots + d_ng(x) = 1.$$

Possiamo riscrivere la precedente equazione nella forma

$$G(x)f(x) + F(x)g(x) = 1,$$

dove

$$G(x) = c_1x^{m-1} + \cdots + c_m, \quad F(x) = d_1x^{n-1} + \cdots + d_n,$$

che dà immediatamente $\mathrm{MCD}(f(x), g(x)) = 1$.

Viceversa se $R(f(x), g(x)) = 0$, possiamo trovare dei coefficienti $c_1, \dots, c_m, d_1, \dots, d_n \in \mathbb{K}$ non tutti nulli e tali che

$$c_1x^{m-1}f(x) + \cdots + c_mf(x) + d_1x^{n-1}g(x) + \cdots + d_ng(x) = 0.$$

Come prima possiamo riscrivere la precedente equazione nella forma

$$G(x)f(x) + F(x)g(x) = 0,$$

dove

$$G(x) = c_1x^{m-1} + \cdots + c_m, \qquad F(x) = d_1x^{n-1} + \cdots + d_n.$$

Supponiamo per fissare le idee $F(x) \neq 0$, allora $\deg F(x) < \deg f(x)$ ed a maggior ragione $f(x)$ non divide $F(x)$. Siccome $f(x)$ divide il prodotto $F(x)g(x)$ deve esistere un fattore irriducibile di $f(x)$ che divide $g(x)$. □

È del tutto evidente che il risultante $R(f(x), g(x))$ è una funzione polinomiale dei coefficienti di $f(x)$ e $g(x)$. La dimostrazione del Teorema 14.6.2 mostra che è sempre possibile trovare due polinomi $F(x), G(x)$ tali che $G(x)f(x) + F(x)g(x) = R(f(x), g(x))$; lo sviluppo di Laplace del determinante mostra che anche i coefficienti di tali polinomi possono essere scelti come funzioni polinomiali dei coefficienti di $f(x)$ e $g(x)$. A tal fine, il trucco è considerare la matrice di Sylvester (14.3) a coefficienti in $\mathbb{K}[x]$. Adesso se all'ultima colonna aggiungiamo la penultima moltiplicata per x, la terzultima moltiplicata per x^2 e così via fino ad aggiungere la prima colonna moltiplicata per x^{n+m-1}, otteniamo la colonna

$$(x^{m-1}f(x), \dots, f(x), x^{n-1}g(x), \dots, g(x))^T.$$

Basta adesso effettuare lo sviluppo di Laplace rispetto all'ultima colonna per trovare la formula

$$R(f(x), g(x)) = G(x)f(x) + F(x)g(x),$$

dove

$$\begin{aligned} F(x) &= d_1x^{n-1} + \cdots + d_n, & d_i &= (-1)^{n+i}|S(f(x), g(x))_{i+m,n+m}|, \\ G(x) &= c_1x^{m-1} + \cdots + c_m, & c_i &= (-1)^{n+m+i}|S(f(x), g(x))_{i,n+m}|. \end{aligned}$$

TEOREMA 14.6.3 (Relazioni di bilinearità). *Dati tre polinomi non nulli* $f(x), g(x), h(x) \in \mathbb{K}[x]$ *si hanno le formule*

$$R(f(x)h(x), g(x)) = R(f(x), g(x))R(h(x), g(x)),$$
$$R(f(x), h(x)g(x)) = R(f(x), h(x))R(f(x), g(x)).$$

DIMOSTRAZIONE. Segue dalle proprietà del determinante che se $f(x)$ ha grado n e $g(x)$ ha grado m, allora $R(f(x), g(x)) = (-1)^{nm}R(g(x), f(x))$ e $R(af(x), g(x)) = a^m R(f(x), g(x))$ per ogni $a \in \mathbb{K}$, $a \neq 0$. È quindi sufficiente dimostrare la relazione $R(f(x), h(x)g(x)) = R(f(x), h(x))R(f(x), g(x))$ con l'ulteriore ipotesi che $f(x)$ sia un polinomio monico.

Supponiamo quindi il polinomio $f(x)$ monico di grado n, dato un qualsiasi polinomio $g(x)$, indichiamo con $C = (c_{ij})$ la matrice quadrata di ordine n a coefficienti in $\mathbb{K}$ tale che per ogni $i = 1, \dots, n$ vale

$$x^{n-i}g(x) = h_i(x)f(x) + \sum_{j=1}^{n} c_{ij}x^{n-j}, \qquad \text{con } \ h_i(x) \in \mathbb{K}[x].$$

Vogliamo dimostrare che $R(f(x), g(x)) = \det(c_{ij})$; a tal fine basta osservare che, se m è il grado di $g(x)$, allora $h_i(x)$ ha grado $< m$ ed ogni polinomio $h_i(x)f(x)$ è una combinazione lineare a coefficienti in $\mathbb{K}$ di $f(x), xf(x), \dots, x^{m-1}f(x)$. È dunque possibile sommare ad ognuna delle ultime n righe della matrice $S(f(x), g(x))$ dei multipli delle prime m righe in modo tale che diventi una matrice triangolare a blocchi della forma

$$\begin{pmatrix} T & * \\ 0 & C \end{pmatrix},$$

dove T è una matrice triangolare superiore di ordine m con i coefficienti della diagonale tutti uguali a 1.

Indichiamo con $g\colon \mathbb{K}[x] \to \mathbb{K}[x]$ l'applicazione lineare data dalla moltiplicazione per $g(x)$ e con $(f(x)) \subseteq \mathbb{K}[x]$ lo spazio vettoriale di tutti i polinomi divisibili per $f(x)$. Siccome $g((f(x))) \subseteq (f(x))$ si ha una fattorizzazione al quoziente

$$g\colon \frac{\mathbb{K}[x]}{(f(x))} \to \frac{\mathbb{K}[x]}{(f(x))}$$

che è rappresentata dalla matrice C nella base $x^{n-1}, \dots, x, 1$. Abbiamo quindi dimostrato che $R(f(x), g(x)) = \det(g)$ e per il teorema di Binet

$$R(f(x), h(x)g(x)) = \det(hg) = \det(h)\det(g) = R(f(x), h(x))R(f(x), g(x)).$$

□

Esercizi.

722 (Invarianza per traslazione). Dimostrare che, per ogni $f(x), g(x) \in \mathbb{K}[x]$ e per ogni $a \in \mathbb{K}$ vale

$$R(f(x-a), g(x-a)) = R(f(x), g(x)).$$

723. Usare le relazioni di bilinearità e la Formula (14.4) per mostrare che, se $f(x) = a_0 \prod_{i=1}^{n}(x - \alpha_i)$, allora

$$R(f(x), g(x)) = a_0^m \prod_{i=1}^{n} g(\alpha_i)$$

e quindi che, se $g = b_0 \prod_{i=1}^{m}(x - \beta_i)$, allora vale

$$R(f(x), g(x)) = a_0^m b_0^n \prod_{i=1}^{n}\prod_{j=1}^{m}(\alpha_i - \beta_j).$$

CAPITOLO 15

Forme bilineari e quadratiche

In questo capitolo tratteremo la teoria delle forma bilineari e quadratiche con gli argomenti presentati in ordine di generalità decrescente del campo base e degli spazi vettoriali coinvolti: inizieremo con spazi vettoriali definiti su campi arbitrari, per poi passare a campi di caratteristica $\neq 2$ e finire con i campi dei numeri reali e complessi. Similmente inizieremo con spazi vettoriali qualunque, per poi passare a quelli di dimensione finita e terminare con quelli numerici.

15.1. Nozioni base

In questa sezione daremo la definizione e alcuni semplici esempi di forme bilineari e quadratiche.

DEFINIZIONE 15.1.1. Sia V uno spazio vettoriale su di un campo $\mathbb{K}$ fissato. Un'applicazione

$$\varphi\colon V \times V \to \mathbb{K}$$

si dice una **forma bilineare** se per ogni vettore $v \in V$ le due applicazioni

$$\varphi(-, v)\colon V \to \mathbb{K}, \qquad w \mapsto \varphi(w, v),$$
$$\varphi(v, -)\colon V \to \mathbb{K}, \qquad w \mapsto \varphi(v, w),$$

sono lineari.

In altri termini φ è bilineare se per ogni terna di vettori $v, w, z \in V$ e per ogni coppia di scalari $a, b \in \mathbb{K}$ vale

$$\varphi(v, aw + bz) = a\varphi(v, w) + b\varphi(v, z), \qquad \varphi(aw + bz, v) = a\varphi(w, v) + b\varphi(z, v).$$

ESEMPIO 15.1.2. Se $f, g\colon V \to \mathbb{K}$ sono due applicazioni lineari, allora le tre applicazioni

$$\varphi_1,\ \varphi_2,\ \varphi_3\colon V \times V \to \mathbb{K},$$
$$\varphi_1(x, y) = f(x)g(y), \quad \varphi_2(x, y) = f(x)g(y) + g(x)f(y), \quad \varphi_3(x, y) = f(x)g(y) - g(x)f(y),$$

sono bilineari. Più in generale per ogni intero $n > 0$ ed ogni successione di $2n$ applicazioni lineari $f_1, \ldots, f_n, g_1, \ldots, g_n\colon V \to \mathbb{K}$, l'applicazione

$$\varphi\colon V \times V \to \mathbb{K}, \qquad \varphi(x, y) = \sum_{i=1}^{n} f_i(x)g_i(y),$$

è bilineare.

DEFINIZIONE 15.1.3. Una forma bilineare $\varphi\colon V \times V \to \mathbb{K}$ si dice:

(1) **Simmetrica** se $\varphi(x, y) = \varphi(y, x)$ per ogni $x, y \in V$.
(2) **Antisimmetrica** se $\varphi(x, y) = -\varphi(y, x)$ per ogni $x, y \in V$.
(3) **Alternante** se $\varphi(x, x) = 0$ per ogni $x \in V$.

La forma φ_2 dell'Esempio 15.1.8 è simmetrica, mentre la forma φ_3 è alternante.

È facile osservare che ogni forma alternante è anche antisimmetrica, infatti se φ è alternante allora per ogni $x, y \in V$ si ha:

$$0 = \varphi(x + y, x + y) = \varphi(x, x) + \varphi(x, y) + \varphi(y, x) + \varphi(y, y) = \varphi(x, y) + \varphi(y, x)\,.$$

Viceversa se φ è antisimmetrica ed il campo $\mathbb{K}$ ha caratteristica diversa da 2 (cioè $1{+}1 \neq 0$) allora φ è anche alternante: infatti, ponendo $x = y$ vale $\varphi(x, x) = -\varphi(x, x) = 0$ e quindi $2\varphi(x, x) = 0$.

Abbiamo già osservato che il determinante di una matrice 2×2 è una forma bilineare alternante sullo spazio dei vettori colonna $\mathbb{K}^2$.

Osservazione 15.1.4. Talvolta vengono usate notazioni alternative per denotare una forma bilineare $V \times V \to \mathbb{K}$, come ad esempio:

(1) $V \times V \ni (x, y) \mapsto x \cdot y$ (notazione prodotto), od anche $x * y$ ecc.;
(2) $V \times V \ni (x, y) \mapsto [x, y]$ (notazione parentesi o bracket), od anche $\langle x, y \rangle$ ecc.;
(3) $V \times V \ni (x, y) \mapsto \langle x \mid y \rangle$ (notazione bra-ket).

Come per le applicazioni lineari, anche le forme bilineari possono essere sommate e moltiplicate per scalari. Più precisamente, date $\varphi, \psi \colon V \times V \to \mathbb{K}$ bilineari e $a, b \in \mathbb{K}$ si ha

$$a\varphi + b\psi \colon V \times V \to \mathbb{K}, \qquad (a\varphi + b\psi)(x, y) = a\varphi(x, y) + b\psi(x, y).$$

È immediato osservare che se φ, ψ sono simmetriche (risp.: antisimmetriche, alternanti), allora anche $a\varphi + b\psi$ è simmetrica (risp.: antisimmetrica, alternante). In caratteristica $\neq 2$ ogni forma bilineare φ si scrive in modo unico come somma di una forma simmetrica e di una forma alternante, vedi Esercizio 724:

$$\varphi(x, y) = \frac{\varphi(x, y) + \varphi(y, x)}{2} + \frac{\varphi(x, y) - \varphi(y, x)}{2}. \tag{15.1}$$

Definizione 15.1.5. Sia V uno spazio vettoriale su di un campo $\mathbb{K}$. Un'applicazione

$$q \colon V \to \mathbb{K}$$

si dice una **forma quadratica** se soddisfa le seguenti due condizioni:

(1) $q(tx) = t^2 q(x)$ per ogni $x \in V$ ed ogni $t \in \mathbb{K}$;
(2) l'applicazione

$$\varphi \colon V \times V \to \mathbb{K}, \qquad \varphi(x, y) = q(x + y) - q(x) - q(y),$$

è una forma bilineare.

Esempio 15.1.6. Se $f, g \colon V \to \mathbb{K}$ sono due applicazioni lineari, allora l'applicazione

$$q \colon V \to \mathbb{K}, \qquad q(x) = f(x)g(x),$$

è quadratica. Infatti $q(tx) = f(tx)g(tx) = t^2 f(x)g(x) = t^2 q(x)$, mentre per quanto visto nell'Esempio 15.1.2 l'applicazione

$$(x, y) \mapsto q(x + y) - q(x) - q(x) = f(x)g(y) + f(y)g(x)$$

è una forma bilineare simmetrica.

Esempio 15.1.7. L'applicazione

$$\Phi \colon \mathbb{K}^3 \to \mathbb{K}, \qquad \Phi \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1 x_2 + x_2 x_3,$$

è quadratica. Possiamo scrivere $\Phi(x) = \pi_1(x)\pi_2(x) + \pi_2(x)\pi_3(x)$ dove $\pi_i \colon \mathbb{K}^3 \to \mathbb{K}$ è la i-esima coordinata. Adesso, per ogni intero $n > 0$ ed ogni successione di $2n$ applicazioni lineari $f_1, \ldots, f_n, g_1, \ldots, g_n \colon V \to \mathbb{K}$, l'applicazione

$$q \colon V \times V \to \mathbb{K}, \qquad q(x) = \sum_{i=1}^{n} f_i(x) g_i(x),$$

è quadratica.

Alla stessa maniera delle forme bilineari, anche le forme quadratiche possono essere sommate e moltiplicate per scalare.

Esempio 15.1.8. Sia $b \colon V \times V \to \mathbb{K}$ una forma bilineare, allora l'applicazione

$$q \colon V \to \mathbb{K}, \qquad q(x) = b(x, x),$$

è una forma quadratica. Infatti $q(tx) = b(tx, tx) = tb(x, tx) = t^2 b(x, x) = t^2 q(x)$ e

$$q(x + y) - q(x) - q(y) = b(x + y, x + y) - b(x, x) - b(y, y) = b(x, y) + b(y, x)$$

è bilineare. È chiaro che due forme bilineari inducono la stessa forma quadratica se e solo se la loro differenza è alternante.

Osservazione 15.1.9. È possibile dimostrare che ogni forma quadratica si ottiene da una forma bilineare secondo la regola descritta nell'Esempio 15.1.8. In caratteristica $\neq 2$ la dimostrazione è molto semplice e viene riportata nel prossimo Lemma 15.1.10, mentre la dimostrazione per campi arbitrari è decisamente più ostica e viene rimandata agli Esercizi 738 e 739.

Da ciò ne consegue in particolare che lo spazio vettoriale delle forme quadratiche è naturalmente isomorfo al quoziente dello spazio delle forme bilineari per il sottospazio delle forme alternanti.

Lemma 15.1.10. *Sia V uno spazio vettoriale su di un campo $\mathbb{K}$ di caratteristica $\neq 2$. Un'applicazione $q\colon V \to \mathbb{K}$ è una forma quadratica se e solo se esiste una forma bilineare $b\colon V \times V \to \mathbb{K}$ simmetrica e tale che*

$$q(x) = b(x,x) \quad \textit{per ogni} \quad x \in V.$$

*In tale situazione la forma b è unica ed è detta la **polare** di q.*

Dimostrazione. Una implicazione è già stata vista nell'Esempio 15.1.8. Viceversa, se $q\colon V \to \mathbb{K}$ è quadratica allora $\varphi(x,y) = q(x+y) - q(x) - q(y)$ è una forma bilineare simmetrica e basta dimostrare che $q(x) = \dfrac{1}{2}\varphi(x,x)$ per ogni $x \in V$. A tal fine basta osservare che

$$2q(x) = 4q(x) - q(x) - q(x) = q(2x) - q(x) - q(x) = \varphi(x,x)$$

e quindi che $q(x) = b(x,x)$, dove $b = \varphi/2$. Per dimostrare l'unicità della polare occorre dimostrare che se $q(x) = b(x,x)$ con $b\colon V \times V \to \mathbb{K}$ bilineare simmetrica allora $2b = \varphi$. Usando la simmetria di b si ha

$$\varphi(x,y) = b(x+y, x+y) - b(x,x) - b(y,y) = b(x,y) + b(y,x) = 2b(x,y)$$

per ogni $x, y \in V$.

□

Esercizi.

724. Provare che, in caratteristica $\neq 2$, la Formula 15.1 descrive l'unico modo possibile di scrivere una forma bilineare come somma di una forma bilineare simmetrica e di una alternante.

725. Siano $\varphi\colon V \times V \to \mathbb{K}$ una forma bilineare e $q\colon V \to \mathbb{K}$ una forma quadratica. Sia $f\colon W \to V$ un'applicazione lineare; provare che l'applicazione

$$f^*\varphi\colon W \times W \to \mathbb{K}, \qquad f^*\varphi(x,y) = \varphi(f(x), f(y))$$

è una forma bilineare e che l'applicazione

$$f^*q\colon W \to \mathbb{K}, \qquad f^*q(x) = q(f(x))$$

è una forma quadratica.

726 (Ⓐ). Sia V lo spazio vettoriale delle funzioni continue sull'intervallo $[0,1]$ dimostrare che l'applicazione

$$\varphi\colon V \times V \to \mathbb{R}, \qquad \varphi(f,g) = \int_0^1 f(t)g(t)dt,$$

è una forma bilineare.

727. Sia S un insieme di generatori di uno spazio vettoriale V e sia $\varphi\colon V \times V \to \mathbb{K}$ una forma bilineare. Provare che φ è alternante se e solo se è antisimmetrica e $\varphi(v,v) = 0$ per ogni $v \in S$.

728 (Vero o falso?). Dimostrare oppure confutare le seguenti affermazioni:

(1) sia $\varphi\colon V \times V \to \mathbb{K}$ una forma bilineare. Allora il sottoinsieme

$$U = \{v \in V \mid \varphi(v,u) = 0 \text{ per ogni } u \in V\}$$

è un sottospazio vettoriale;

(2) siano $\varphi\colon V \times V \to \mathbb{K}$ una forma bilineare e $u_1, u_1 \in V$ vettori linearmente indipendenti. Allora il sottoinsieme
$$W = \{v \in V \mid \varphi(v, u_1)\varphi(v, u_1) = 0\}$$
è un sottospazio vettoriale;
(3) sia $q\colon V \to \mathbb{K}$ una forma quadratica. Allora il sottoinsieme
$$Z = \{v \in V \mid q(v) = 0\}$$
è un sottospazio vettoriale.

729. Siano $\mathbb{K}$ un campo qualsiasi e $q\colon V \to \mathbb{K}$ lineare surgettiva. Provare che q soddisfa le condizioni (1) e (2) della Definizione 15.1.5 se e solo se $\mathbb{K} = \mathbb{F}_2 = \{0, 1\}$. (Suggerimento: se $\mathbb{K} \neq \mathbb{F}_2$, $x \in V$ è un qualunque vettore tale che $f(x) \neq 0$ e $t \in \mathbb{K}$ è un qualunque scalare $\neq 0, 1$; per linearità $q(tx) = tq(x)$ e siccome $t^2 \neq t$ si ha $q(tx) \neq t^2 q(x)$.)

730 (☕). Siano V uno spazio vettoriale e $f\colon V \to \mathbb{K}$ lineare fissata. Quali sono le applicazioni lineari $g\colon V \to \mathbb{K}$ tali che la forma bilineare
$$\varphi\colon V \times V \to \mathbb{K}, \qquad \varphi(x, y) = f(x)g(y),$$
è simmetrica. (Per semplificare il problema e togliere il ☕ assumere che V abbia dimensione finita e che $\mathbb{K}$ sia infinito.)

731 (☕). Siano V uno spazio vettoriale di dimensione finita e φ una forma bilineare su V. dimostrare che le due applicazioni:
$$f\colon V \to V^\vee, \qquad f(v)(h) = \varphi(v, h),$$
$$g\colon V \to V^\vee, \qquad g(v)(h) = \varphi(v, h),$$
sono lineari ed hanno lo stesso rango. (Suggerimento: dualità vettoriale).

15.2. Rango e congruenza di forme bilineari

A partire da questo momento e fino alla fine del capitolo, tutti gli spazi vettoriali sono assunti, salvo avviso contrario, di dimensione finita.

Siano V uno spazio vettoriale di dimensione n e $\varphi\colon V \times V \to \mathbb{K}$ una forma bilineare; ogni base $e_1, \dots, e_n$ di V permette di associare alla forma φ la matrice
$$B \in M_{n,n}(\mathbb{K}), \qquad B = (b_{ij}), \qquad b_{ij} = \varphi(e_i, e_j),$$
i cui coefficienti sono i valori di φ calcolati nelle n^2 coppie di elementi della base.

La precedente costruzione è reversibile e la matrice B determina univocamente la forma φ, nel senso descritto dal seguente lemma.

LEMMA 15.2.1. *Sia $e_1, \dots, e_n$ una base fissata di uno spazio vettoriale V. Per ogni matrice $B = (b_{ij}) \in M_{n,n}(\mathbb{K})$ esiste un'unica forma bilineare $\varphi\colon V \times V \to \mathbb{K}$ tale che $\varphi(e_i, e_j) = b_{ij}$ per ogni i, j. Inoltre la forma φ è simmetrica (risp.: alternante) se e solo se la matrice B è simmetrica (risp.: alternante).*

DIMOSTRAZIONE. *Unicità.* Se φ è una forma bilineare su V, data una qualunque coppia di vettori $u = \sum x_i e_i$ e $v = \sum y_i e_i$, $x_i, y_i \in \mathbb{K}$, per bilinearità si ha
$$\varphi(u, v) = \varphi(\sum x_i e_i, \sum y_j e_j) = \sum_{i,j} x_i y_j \, \varphi(e_i, e_j) = \sum_{i,j} x_i y_j b_{ij}$$
ossia $\varphi(u, v) = x^T B y$, dove $x, y \in \mathbb{K}^n$ sono i vettori delle coordinate di u, v rispetto alla base $e_1, \dots, e_n$ e $x^T B y$ denota l'usuale prodotto righe×colonne. In particolare φ è univocamente definita dagli scalari $b_{ij} = \varphi(e_i, e_j)$.

Esistenza. Basta definire φ usando la formula
$$\varphi(u, v) = \sum_{i,j} b_{ij} x_i y_j = x^T B y\,, \qquad u = \sum x_i e_i,\ v = \sum y_i e_i,$$
trovata nella dimostrazione dell'unicità.

Se φ è simmetrica, allora $b_{ij} = \varphi(e_i, e_j) = \varphi(e_j, e_i) = b_{ji}$ e la matrice B è simmetrica. Viceversa se la matrice B è simmetrica, per ogni $u = \sum x_i e_i$ e $v = \sum y_i e_i$ si ha

$$\varphi(u, v) = x^T B y = (x^T B y)^T = y^T B^T x = y^T B x = \varphi(v, u).$$

Analogo ragionamento nel caso alternante. □

Il Lemma 15.2.1 applicato agli spazi vettoriali numerici ed alle loro basi canoniche ci dice in particolare che ogni applicazione bilineare $\varphi\colon \mathbb{K}^n \times \mathbb{K}^n \to \mathbb{K}$ si scrive in maniera unica come

$$\varphi(x, y) = x^T B y, \qquad B \in M_{n,n}(\mathbb{K}), \quad x, y \in \mathbb{K}^n.$$

Ovviamente cambiando base anche la matrice B cambia; uno dei principali obiettivi di questa sezione sarà analizzare come cambia B al variare della base e dimostrare in particolare che il rango di B è invariante per cambi di base: potremo quindi definire il **rango** di φ come il rango di B.

Sia $\varepsilon_1, \dots, \varepsilon_n$ un'altra base di V, e sia $A = (a_{ij}) \in M_{n,n}(\mathbb{K})$ la corrispondente matrice di cambiamento di base:

$$(\varepsilon_1, \dots, \varepsilon_n) = (e_1, \dots, e_n)A \iff \varepsilon_i = \sum_j a_{ji} e_j.$$

Sia $C = (c_{ij})$ la matrice dei valori di φ nella base ε_i, allora

$$c_{ij} = \varphi(\varepsilon_i, \varepsilon_j) = \varphi(\sum_h a_{hi} e_h, \sum_k a_{kj} e_k) = \sum_{h,k} a_{hi} b_{hk} a_{kj}$$

e quindi $C = A^T B A$. Lo stesso conto dimostra anche che se partiamo da una qualunque matrice invertibile $U \in GL_n(\mathbb{K})$, allora la matrice $U^T B U$ rappresenta la forma blineare φ nella base $(e_1, \dots, e_n)U$.

Definizione 15.2.2. Due matrici $B, C \in M_{n,n}(\mathbb{K})$ si dicono **congruenti** se esiste una matrice invertibile $A \in GL_n(\mathbb{K})$ tale che $C = A^T B A$.

Abbiamo dimostrato che due matrici sono congruenti se e solo se rappresentano la medesima forma bilineare in due basi diverse.

Esempio 15.2.3. Anche se a prima vista la nozione di congruenza è simile alla nozione di similitudine, le due relazioni si comportano molto diversamente. Ad esempio, sul campo dei numeri reali le due matrici

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}^T \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

sono congruenti ma non sono simili (esercizio: perché?). Analogamente le due matrici

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

sono simili ma non congruenti (esercizio: perché?).

Due matrici congruenti B, C hanno lo stesso rango: infatti se $C = A^T B A$ per qualche matrice invertibile A, allora anche A^T è invertibile e sappiamo che il rango è invariante per moltiplicazione, sia a destra che a sinistra, con una matrice invertibile. Questo prova in particolare che la prossima definizione è ben posta.

Definizione 15.2.4. Sia $\varphi\colon V \times V \to \mathbb{K}$ una forma bilineare su uno spazio vettoriale di dimensione finita. Il **rango** di φ è il rango della matrice di coefficienti $\varphi(e_i, e_j)$, dove $e_1, \dots, e_n$ è una (qualunque) base di V.

Poiché una matrice e la sua trasposta hanno lo stesso rango, segue immediatamente che una forma bilineare φ ha il medesimo rango della forma bilineare φ^T ottenuta scambiando l'ordine delle variabili: $\varphi^T(u, v) = \varphi(v, u)$. Esistono altre possibili definizioni equivalenti di rango: le principali sono elencate nei seguenti due teoremi.

Teorema 15.2.5. *Sia $\varphi\colon V \times V \to \mathbb{K}$ una forma bilineare su uno spazio vettoriale di dimensione finita n. Per un numero intero r le seguenti condizioni sono equivalenti:*

(1) *φ ha rango r;*

(2) *il sottospazio vettoriale*

$$\mathrm{RKer}(\varphi) = \{v \in V \mid \varphi(u,v) = 0 \text{ per ogni } u \in V\}$$

ha dimensione $n - r$.

(3) *il sottospazio vettoriale*

$$\mathrm{LKer}(\varphi) = \{u \in V \mid \varphi(u,v) = 0 \text{ per ogni } v \in V\}$$

ha dimensione $n - r$*;*

DIMOSTRAZIONE. Siccome $\mathrm{LKer}(\varphi) = \mathrm{RKer}(\varphi^T)$ basta dimostrare l'equivalenza delle prime due condizioni. Fissiamo una base di V, ossia un isomorfismo di spazi vettoriali $f\colon \mathbb{K}^n \to V$ e sia B matrice dei valori di φ in tale base. Per costruzione $\varphi(f(x), f(y)) = x^T B y$ e

$$f^{-1}(\mathrm{RKer}(\varphi)) = \{y \in \mathbb{K}^n \mid x^T B y = 0 \text{ per ogni } x \in \mathbb{K}^n\}.$$

Se x è lo i-esimo vettore della base canonica, allora $x^T B y$ coincide con la i-esima coordinata di By e quindi si ha

$$f^{-1}(\mathrm{RKer}(\varphi)) = \{y \in \mathbb{K}^n \mid By = 0\} = \operatorname{Ker} B.$$

Basta adesso applicare il teorema del rango per concludere la dimostrazione. □

Il prossimo teorema ripete sostanzialmente il precedente nel formalismo degli spazi vettoriali duali.

TEOREMA 15.2.6. *Sia* $\varphi\colon V \times V \to \mathbb{K}$ *una forma bilineare su uno spazio vettoriale di dimensione finita* n*. Per un numero intero* r *le seguenti condizioni sono equivalenti:*

(1) φ *ha rango* r*;*

(2) *l'applicazione lineare*

$$h\colon V \to V^\vee, \qquad h(u)(v) = \varphi(u,v),$$

ha rango r*;*

(3) *esistono due succesioni* $f_1, \dots, f_r \in V^\vee$ *e* $g_1, \dots, g_r \in V^\vee$*, ciascuna formata da* r *funzionali lineari linearmente indipendenti tali che*

$$\varphi(u,v) = \sum_{i=1}^{r} f_i(u) g_i(v);$$

(4) r *è il minimo intero per cui esistono due succesioni* $f_1, \dots, f_r \in V^\vee$ *e* $g_1, \dots, g_r \in V^\vee$ *tali che*

$$\varphi(u,v) = \sum_{i=1}^{r} f_i(u) g_i(v);$$

DIMOSTRAZIONE. Mostriamo prima l'equivalenza tra (1) e (2). Prendiamo una base qualsiasi $e_1, \dots, e_n$ di V e sia $f_1, \dots, f_n$ la corrispondente base duale: $f_i(e_j) = \delta_{ij}$. I coefficienti della matrice B che rappresenta φ nella base $e_1, \dots, e_n$ sono $b_{ij} = \varphi(e_i, e_j)$. Denotiamo con $A = (a_{ij})$ la matrice che rappresenta h nelle basi $e_1, \dots, e_n$ e $f_1, \dots, f_n$: per definizione $h(e_i) = \sum_j a_{ji} f_j$ per ogni i e quindi

$$a_{ji} = h(e_i)(e_j) = \varphi(e_i, e_j) = b_{ij}.$$

dunque $A = B^T$ e le due matrici A, B hanno lo stesso rango. Notiamo incidentalmente che

$$\varphi(u,v) = \sum_{i,j} b_{ij} f_i(u) f_j(v) = \sum_{i=1}^{n} f_i(u) g_i(v), \quad \text{dove} \quad g_i = \sum_j b_{ij} f_j,$$

per ogni $u, v \in V$, e quindi che la condizione (4) è ben posta.

Proviamo adesso che (3) implica (2). A tal fine basta estendere $f_1, \dots, f_r$ ad una base $f_1, \dots, f_n$ di $V^\vee$ e considerare la corrispondente base duale $e_1, \dots, e_n$ di $V = V^{\vee\vee}$. Similmente estendiamo $g_1, \dots, g_r$ ad una base $g_1, \dots, g_n$ di $V^\vee$ a cui corrisponde una base duale $\varepsilon_1, \dots, \varepsilon_n$. Denotiano con $A = (a_{ij})$ la matrice che rappresenta h nelle basi e_i, g_j, ossia $h(e_i) = \sum_j a_{ji} g_j$. Allora

$$a_{ji} = h(e_i)(\varepsilon_j) = \varphi(e_i, \varepsilon_j) = \sum_{k=1}^{r} f_k(e_i) g_k(\varepsilon_j) = \begin{cases} 1 & \text{se } i = j \le r, \\ 0 & \text{altrimenti,} \end{cases}$$

da cui segue immediatamente che A ha rango r.

Dimostriamo adesso che (4) implica (3), ossia che se r è il minimo tale che $\varphi = \sum_{i=1}^{r} f_i g_i$, allora le due successioni $f_1, \dots, f_r$ e $g_1, \dots, g_r$ sono linearmente indipendenti. Supponiamo per assurdo che $f_1, \dots, f_r$ siano linearmente dipendenti, allora a meno di permutare gli indici si avrebbe una relazione del tipo

$$f_r = a_1 f_1 + \cdots + a_{r-1} f_{r-1}, \qquad a_i \in \mathbb{K},$$

da cui segue

$$\begin{aligned}\varphi(u,v) &= \sum_{i=1}^{r-1} f_i(u) g_i(v) + \sum_{i=1}^{r-1} a_i f_i(u) g_r(v) \\ &= \sum_{i=1}^{r-1} f_i(u)(g_i + a_i g_r)(v) = \sum_{i=1}^{r-1} f_i(u) l_i(v),\end{aligned}$$

dove $l_i = g_i + a_i g_r \in V^\vee$ per ogni $i = 1, \dots, r-1$, in contraddizione con la minimalità di r. Quindi i funzionali $f_1, \dots, f_r$ sono linearmente indipendenti e lo stesso ragionamento si applica anche ai funzionali $g_1, \dots, g_r$.

Per concludere la dimostrazione del teorema, abbiamo dimostrato che si può sempre scrivere $\varphi(u,v) = \sum_{i=1}^{n} f_i g_i$ e che quindi l'insieme delle espressioni $\varphi = \sum_{\text{finita}} f_i g_i$ è non vuoto. Se tra queste espressioni ne prendiamo una con il numero minimo di addendi, abbiamo dimostrato che tale numero è uguale ad r. □

Definizione 15.2.7. Una forma bilineare si dice **non degenere** se ha rango massimo.

Segue immediatamente dal Teorema 15.2.5 che per una forma bilineare $\varphi\colon V \times V \to \mathbb{K}$ le seguenti condizioni sono equivalenti:

(1) φ è non degenere;
(2) $\varphi(u,v) = 0$ per ogni v, se e solo se $u = 0$;
(3) $\varphi(u,v) = 0$ per ogni u se e solo se $v = 0$.

Ad esempio la forma $x^T B y$ su $\mathbb{K}^n$ è non degenere se e solo se la matrice B è invertibile.

In letteratura si usa spesso le seguente terminologia:

(1) un **prodotto interno** è una forma bilineare simmetrica non degenere;
(2) una **forma simplettica** è una forma bilineare alternante non degenere.[1]

Definizione 15.2.8. Il **prodotto interno canonico** sugli spazi vettoriali numerici $\mathbb{K}^n$ è definito come l'applicazione bilineare simmetrica

$$\mathbb{K}^n \times \mathbb{K}^n \xrightarrow{\cdot} \mathbb{K}, \qquad x \cdot y = x^T y = \sum_{i=1}^{n} x_i y_i.$$

Abbiamo già incontrato il prodotto scalare sugli spazi $\mathbb{R}^n$, che coincide con il prodotto interno canonico.

Esercizi.

732. Sia φ una forma bilineare su $\mathbb{K}^n$. Dimostrare che esiste, ed è unica, una matrice $B \in M_{n,n}(\mathbb{K})$ tale che $\varphi(x,y) = x \cdot By = B^T x \cdot y$.

733. Siano V uno spazio vettoriale di dimensione finita, $\varphi\colon V \times V \to \mathbb{R}$ una forma bilineare e $S \subseteq V$ un insieme di generatori. Mostrare che

$$\operatorname{LKer}(\varphi) = \{u \in V \mid \varphi(u,v) = 0 \text{ per ogni } v \in S\}.$$

734 (☕, ♡). Siano V uno spazio vettoriale di dimensione finita e $\varphi\colon V \times V \to \mathbb{K}$ una forma bilineare con la proprietà che

$$\varphi(x,y) = 0 \quad \text{se e solo se} \quad \varphi(y,x) = 0.$$

Dimostrare che φ è simmetrica oppure alternante.

[1] L'aggettivo simplèttico (o simplèctico), dal greco symplektikós (*συμπλεκτικός*) "relativo all'intreccio", è stato introdotto in matematica da Hermann Weyl nel 1939 e non ha nulla a che vedere con il medesimo termine usato in zoologia e/o mineralogia.

735 (♡). Siano V uno spazio vettoriale di dimensione finita, $\varphi\colon V \times V \to \mathbb{R}$ una forma bilineare e $f\colon V \to V$ lineare tale che $\varphi(f(x), f(y)) = \varphi(x, y)$ per ogni $x, y \in V$. Dimostrare che:

(1) Se φ è non degenere allora f è iniettiva.
(2) Il rango di f è maggiore od uguale al rango di φ.

736. Siano φ, ψ forme bilineari sul medesimo spazio vettoriale V di dimensione finita. Dimostrare che se φ è non degenere esiste $f\colon V \to V$ lineare tale che $\psi(x, y) = \varphi(x, f(y))$. (Sugg.: se B è una matrice simmetrica invertibile, $x^T Ay = x^T BB^{-1}Ay$.)

737. Sia $B = (b_{ij})$ una matrice $n \times n$ simmetrica a coefficienti in un campo $\mathbb{K}$ di caratteristica $\neq 2$. Definiamo un'applicazione $\varphi\colon \mathbb{K}^n \times \mathbb{K}^n \to \mathbb{K}$ mediante la formula

$$\varphi(x, y) = \det \begin{pmatrix} b_{11} & \dots & b_{1n} & y_1 \\ \vdots & \ddots & \vdots & \vdots \\ b_{n1} & \dots & b_{nn} & y_n \\ x_1 & \dots & x_n & 0 \end{pmatrix}.$$

Dimostrare:

(1) φ è bilineare simmetrica. Interpretare, in funzione di B, i coefficienti della matrice A tale che $\varphi(x, y) = x^T Ay$.
(2) Determinare, in funzione del rango di B, il rango di φ.

738 (♡). Sia $e_1, \dots, e_n$ una base fissata di uno spazio vettoriale V. Dimostrare che per ogni forma quadratica $q\colon V \to \mathbb{K}$ vi è un'unica forma bilineare $\varphi\colon V \times V \to \mathbb{K}$ tale che $q(x) = \varphi(x, x)$ per ogni x e $\varphi(e_i, e_j) = 0$ per ogni $i > j$.

739 (☕). Siano V uno spazio vettoriale di dimensione infinita e $q\colon V \to \mathbb{K}$ una forma quadratica. Si consideri l'insieme $\mathcal{F}$ formato dalle coppie (W, φ) in cui $W \subseteq V$ è un sottospazio vettoriale e φ è una forma bilineare su W tale che $q(x) = \varphi(x, x)$ per ogni $x \in W$. Usare il lemma di Zorn per dimostrare che $\mathcal{F}$ contiene (almeno) un elemento massimale (M, b) e ragionare, mutatis mutandis, come nell'Esercizio 738 per provare che $M = V$.

15.3. Forme bilineari simmetriche

A partire da questo momento e fino alla fine del capitolo assumeremo, salvo avviso contrario, che tutti gli spazi vettoriali abbiano dimensione finita e tutti i campi abbiano caratteristica diversa da 2 (e quindi $2, 4, 8, 16, \dots \neq 0$).

Se $\varphi\colon V \times V \to \mathbb{K}$ è una forma bilineare simmetrica, definiamo la forma quadratica associata a φ come l'applicazione (vedi Esempio 15.1.8):

$$\Phi\colon V \to \mathbb{K}, \qquad \Phi(v) = \varphi(v, v).$$

Notiamo che per ogni coppia di vettori $x, y \in V$ vale la **formula di polarizzazione**

$$\varphi(x, y) = \frac{1}{2}(\Phi(x + y) - \Phi(x) - \Phi(y))$$

e quindi ogni forma bilineare simmetrica è univocamente determinata dalla forma quadratica associata; in particolare la restrizione di Φ ad un sottospazio vettoriale $W \subseteq V$ è identicamente nulla se e solo se la restrizione di φ a $W \times W$ è identicamente nulla.

Viceversa data una forma quadratica $\Phi\colon V \to \mathbb{K}$, l'applicazione

$$\varphi\colon V \times V \to \mathbb{K}, \qquad \varphi(x, y) = \frac{1}{2}(\Phi(x + y) - \Phi(x) - \Phi(y))$$

è una forma bilineare simmetrica detta **forma polare** di Φ.

Abbiamo visto che in un sistema di coordinate $x_1, \dots, x_n$, ogni forma bilineare simmetrica φ si scrive come

$$\varphi(x, y) = \sum_{ij} b_{ij} x_i y_j$$

con la matrice (b_{ij}) simmetrica ed univocamente determinata. Come conseguenza, ogni forma quadratica Φ si esprime come un polinomio omogeneo di secondo grado nelle variabili $x_1, \dots, x_n$. Lasciamo per esercizio al lettore la semplice verifica del fatto che se

$$\Phi(x) = \sum_{i \le j} a_{ij} x_i x_j$$

allora la forma polare è uguale a

$$\varphi(x, y) = \sum_i a_{ii} x_i y_i + \frac{1}{2} \sum_{i<j} a_{ij}(x_i y_j + x_j y_i).$$

Ad esempio la polare della forma quadratica $\Phi(x) = x_1^2 - x_1 x_2 - 2x_2^2$ è la forma bilineare

$$\varphi(x, y) = x_1 y_1 - \frac{1}{2}(x_1 y_2 + x_2 y_1) - 2x_2 y_2.$$

Chi conosce le regole base di calcolo delle derivate parziali può, se lo desidera, calcolare la polare φ di una forma quadratica Φ usando la **Formula di Eulero**

$$\varphi(x, y) = \frac{1}{2} \sum_{i=1}^{n} y_i \frac{\partial \Phi}{\partial x_i}(x).$$

ESEMPIO 15.3.1.

(1) Il prodotto interno canonico su $\mathbb{K}^n$ è una forma bilineare simmetrica.
(2) Il **piano iperbolico** su $\mathbb{K}$ si definisce come lo spazio vettoriale $\mathbb{K}^2$ dotato della forma bilineare simmetrica
$$U(x, y) = x_1 y_2 + x_2 y_1 = x^T \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} y.$$
(3) Abbiamo visto che su $\mathbb{K}^n$ ogni forma bilineare prende la forma $x^T B y$ per un'opportuna matrice B. Tale forma è simmetrica se e solo se la matrice B è simmetrica.

DEFINIZIONE 15.3.2. Il **nucleo** di una forma bilineare simmetrica φ è il sottospazio vettoriale

$$\operatorname{Ker} \varphi = \{v \in V \mid \varphi(v, w) = 0 \ \forall w \in V\} = \{v \in V \mid \varphi(w, v) = 0 \ \forall w \in V\}$$

e l'**indice di nullità** di φ è la dimensione di $\operatorname{Ker} \varphi$.

Per quanto visto nelle sezioni precedenti il rango della forma φ può essere calcolato mediate la formula $\operatorname{rg}(\varphi) = \dim V - \dim \operatorname{Ker} \varphi$ ed una forma bilineare simmetrica è un prodotto interno se e solo se il suo nucleo è il sottospazio nullo 0. Equivalentemente, *φ è non degenere se e solo se per ogni $v \neq 0$ esiste $w \in V$ tale che $\varphi(v, w) \neq 0$.*

LEMMA 15.3.3. *Una forma bilineare simmetrica $\varphi \colon V \times V \to \mathbb{K}$ è non degenere se e solo se per ogni successione $\mathbf{u} = \{u_1, \dots, u_m\}$ di vettori linearmente indipendenti di V, l'applicazione*

$$\varphi_{\mathbf{u}} \colon V \to \mathbb{K}^m, \qquad v \mapsto \begin{pmatrix} \varphi(u_1, v) \\ \vdots \\ \varphi(u_m, v) \end{pmatrix},$$

è surgettiva.

DIMOSTRAZIONE. Se $u_1, \dots, u_n$ è una base di V allora per definizione il nucleo di $\operatorname{Ker} \varphi$ coincide con il nucleo di $\varphi_{\mathbf{u}}$ e questo prova l'implicazione del "se". Viceversa, supponiamo φ non degenere e siano $u_1, \dots, u_m$ linearmente indipendenti; basta allora completare tale successione ad una base $\mathbf{u} = \{u_1, \dots, u_m, \dots, u_n\}$ e considerare la composizione dell'isomorfismo lineare $\varphi_{\mathbf{u}} \colon V \to \mathbb{K}^n$ con la proiezione sulle prime m coordinate.

□

OSSERVAZIONE 15.3.4. Bisogna fare attenzione alla reale possibilità che la restrizione di una forma non degenere ad un sottospazio vettoriale proprio possa essere degenere. Ad esempio il prodotto interno canonico $x^T \cdot y$ è non degenere su $\mathbb{C}^2$, ma la sua restrizione al sottospazio $\{x \in \mathbb{C}^2 \mid x_2 = i x_1\}$ è identicamente nulla e quindi degenere.

DEFINIZIONE 15.3.5. Un vettore $x \in V$ si dice **isotropo** rispetto ad una forma quadratica $\Phi \colon V \to \mathbb{K}$ se $\Phi(x) = 0$.

In generale, i vettori isotropi non formano un sottospazio vettoriale. Ad esempio se $\Phi(x) = f(x)g(x)$ con $f, g\colon V \to \mathbb{K}$ lineari, allora i vettori isotropi di Φ sono dati dall'unione del nucleo di f e del nucleo di g. Ogni vettore appartenente al nucleo di una forma quadratica è isotropo, ma il viveversa è generalmente falso.

DEFINIZIONE 15.3.6. Sia φ una forma bilineare simmetrica sullo spazio vettoriale V; due vettori $x, y \in V$ si dicono φ-ortogonali se $\varphi(x, y) = 0$. Una base $e_1, \dots, e_n$ di V si dice φ-ortogonale se $\varphi(e_i, e_j) = 0$ per ogni $i \neq j$.

Quando la forma bilineare è chiara dal contesto parleremo più semplicemente di ortogonalità anziche di φ-ortogonalità.

Se $x_1, \dots, x_n$ sono le coordinate associate ad una base φ-ortogonale $e_1, \dots, e_n$ si ha

$$\varphi(x, y) = \sum_{i,j} \varphi(e_i, e_j) x_i y_j = \sum_{i=1}^{n} \lambda_i \, x_i y_i, \qquad \lambda_i \in \mathbb{K},$$

e la forma quadratica associata diventa

$$\Phi(x) = \sum_i \lambda_i x_i^2, \qquad \lambda_i = \Phi(e_i), \quad x = \sum x_i e_i \, .$$

Dato che la matrice $\varphi(e_i, e_j)$ è diagonale abbiamo dimostrato il seguente utile criterio.

PROPOSIZIONE 15.3.7. *Il rango di una forma bilineare simmetrica φ è uguale al numero di indici i tali che $\varphi(e_i, e_i) \neq 0$, calcolati rispetto ad una qualunque base φ-ortogonale $e_1, \dots, e_n$.*

Diamo adesso due distinte dimostrazioni del seguente importante risultato.

TEOREMA 15.3.8. *Ogni forma bilineare simmetrica φ su di uno spazio vettoriale di dimensione finita ammette basi φ-ortogonali.*

DIMOSTRAZIONE. Delle due dimostrazioni proposte, la prima, breve ed elegante, mostrerà solamente l'esistenza mentre la seconda darà anche un metodo di costruzione della base che potrà essere usato negli esercizi.

Prima dimostrazione. Dimostriamo il teorema per induzione su $n = \dim V$; se $n = 1$ qualsiasi vettore non nullo è una base φ-ortogonale. Se $n > 1$ si possono avere due casi, nel primo φ è identicamente nulla e quindi ogni base è φ-ortogonale. Se invece φ non è identicamente nulla, per la formula di polarizzazione esiste un vettore $v_1 \in V$ tale che $\varphi(v_1, v_1) \neq 0$. Denotiamo $W = \{w \in V \mid \varphi(v_1, w) = 0\}$ e poiché W è il nucleo dell'applicazione lineare $\varphi(v_1, -)\colon V \to \mathbb{K}$ si ha $\dim W \geq n - 1$; siccome $v_1 \notin W$ si ha $\dim W = n - 1$ e $V = \mathbb{K}v_1 \oplus W$. Per induzione esiste una base $v_2, \dots, v_n \in W$ che è φ-ortogonale. È allora chiaro che $v_1, v_2, \dots, v_n$ è una base φ-ortogonale di V.

Seconda dimostrazione. Diremo che una base $e_1, \dots, e_n$ è *k-ortogonale* se $\varphi(e_i, e_j) = 0$ per ogni $i < k$ e per ogni $j \neq i$. Notiamo che per stabilire se una base $e_1, \dots, e_n$ è k-ortogonale basta verificare che $\varphi(e_i, e_j) = 0$ per ogni $i < k$ e per ogni $j > i$: infatti se $j < i$ allora $j < k$, $i > j$ e quindi $\varphi(e_i, e_j) = \varphi(e_j, e_i) = 0$.

Diremo che una base k-ortogonale $e_1, \dots, e_n$ è *correttamente ordinata* se $\varphi(e_k, e_k) \neq 0$ oppure se $\varphi(e_k, e_j) = 0$ per ogni $j > k$ (nel secondo caso la base risulta $(k+1)$-ortogonale, e viceversa).

Partiamo da una base qualsiasi, che per vacuità di condizioni risulta essere 1-ortogonale. Illustriamo adesso un algoritmo che permette, a partire da una base k-ortogonale, con $1 \leq k \leq n$, di costruire esplicitamente, prima una base k-ortogonale correttamente ordinata, e poi un'altra $(k+1)$-ortogonale.

1) Sia $e_1, \dots, e_n$ una base di k-ortogonale, se $\varphi(e_k, e_j) = 0$ per ogni $j > k$ allora la base è già $(k+1)$-ortogonale andate al punto 4). Altrimenti sia $l \leq n$ il minimo indice tale che $\varphi(e_k, e_l) \neq 0$: per la k-ortogonalità della base si ha $l \geq k$. Se $k = l$ andate al punto 3). Se $l > k$ andate punto 2).

2) Se siete arrivati qui è perché $\varphi(e_k, e_k) = 0$ e $\varphi(e_k, e_l) \neq 0$ per qualche $l > k$. dalla formula $\varphi(e_k + e_l, e_k + e_l) - \varphi(e_k - e_l, e_k - e_l) = 4\varphi(e_k, e_l) \neq 0$ segue che i due addendi al primo membro non possono essere entrambi nulli e possiamo certamente trovare $a \in \{1, -1\} \subset \mathbb{K}$ in modo che $\varphi(e_k + ae_l, e_k + ae_l) \neq 0$. Sostituite e_k con $e_k + ae_l$, lasciate invariate gli altri elementi della base e andate al punto 3).

3) Se siete arrivate qui è perché avete una base $e_1, \dots, e_n$ k-ortogonale correttamente ordinata che non è $(k+1)$-ortogonale; questo implica in particolare che $\varphi(e_k, e_k) \neq 0$.

Poniamo $v_i = e_i$ per ogni $i \leq k$, mentre se $i > k$ poniamo

$$v_i = e_i - \frac{\varphi(e_i, e_k)}{\varphi(e_k, e_k)} e_k.$$

Si verifica facilmente che $v_1, \dots, v_n$ è una base $(k+1)$-ortogonale: adesso andate al punto 4).

4) Se $k+1 \leq n$ si aumenta k di una unità e si torna al punto 1). Se $k+1 = n+1$ abbiamo trovato una base n-ortogonale che, per definizione, è φ-ortogonale. □

È importante osservare che nel teorema di esistenza delle basi ortogonali sono fondamentali le ipotesi che la forma sia simmetrica e che il campo abbia caratteristica $\neq 2$. Infatti la matrice che rappresenta una forma bilineare in una base ortogonale è diagonale, quindi simmetrica. D'altra parte, su un qualunque campo di caratteristica 2 il piano iperbolico non possiede basi ortogonali (Esercizio 742).

Definizione 15.3.9. Diremo che due forme bilineari simmetriche $\varphi, \psi \colon V \times V \to \mathbb{K}$ sono **congruenti** se esiste un'applicazione lineare invertibile $A \colon V \to V$ tale che

$$\varphi(x, y) = \psi(Ax, Ay), \qquad \text{per ogni } x, y \in V.$$

Diremo che due forme quadratiche $\Phi, \Psi \colon V \to \mathbb{K}$ sono **congruenti** se esiste un'applicazione lineare invertibile $A \colon V \to V$ tale che $\Phi(x) = \Psi(Ax)$ per ogni $x \in V$.

La relazione di congruenza di forme bilineari è strettamente correlata alla nozione di cogruenza di matrici. Le forme bilineari simmetriche su $\mathbb{K}^n$ sono in bigezione naturale con le matrici simmetriche $B \in M_{n,n}(\mathbb{K})$: ad ogni matrice simmetrica B corrisponde la forma bilineare $x^T B y$. Due forme bilineari simmetriche $b(x, y) = x^T B y$ e $c(x, y) = x^T C y$ sono per definizione congruenti se esiste un isomorfismo lineare di $\mathbb{K}^n$, ossia una matrice inveribile A tale che

$$x^T B y = b(x, y) = c(Ax, Ay) = (Ax)^T C (Ay) = x^T (A^T C A) y, \qquad \text{per ogni} x, y \in \mathbb{K}^n.$$

Dato che la precedente condizione è soddisfatta se e solo se $B = A^T C A$ ne consegue che b e c sono congruenti cone forma bilineari se e solo B e C sono congruenti come matrici.

Corollario 15.3.10. *Su di un qualunque campo di caratteristica $\neq 2$ ogni matrice simmetrica è congruente ad una matrice diagonale.*

Dimostrazione. Sia B una matrice simmetrica, per il Teorema 15.3.8 esiste una base $v_1, \dots, v_n$ di $\mathbb{K}^n$ che è ortogonale per la forma bilineare $x^T B y$. Se A è la matrice (invertibile) le cui colonne sono $v_1, \dots, v_n$, i coefficienti della matrice $A^T B A$ coincidono con i valori $v_i^T B v_j$, che si annullano per $i \neq j$. □

Lemma 15.3.11. (1) *Due forme bilineari simmetriche sono congruenti se e solo se le forme quadratiche associate sono congruenti.*

(2) *La congruenza è una relazione di equivalenza sull'insieme delle forme bilineari simmetriche e/o quadratiche.*

Dimostrazione. Siano Φ, Ψ le forme quadratiche associate a due forme bilineari simmetriche φ, ψ. Se φ, ψ sono congruenti esiste $A \colon V \to V$ lineare invertibile tale che $\varphi(x, y) = \psi(Ax, Ay)$ per ogni x, y. In particolare, per ogni $x \in V$ vale $\Phi(v) = \varphi(v, v) = \psi(Ax, Ax) = \Psi(Ax)$ e quindi anche Φ, Ψ sono congruenti.

Viceversa se Φ, Ψ sono congruenti, diciamo $\Phi(x) = \Psi(Ax)$, allora per ogni $x, y \in V$ vale

$$\begin{aligned}\varphi(x, y) &= \frac{1}{2}(\Phi(x+y) - \Phi(x) - \Phi(y)) = \frac{1}{2}(\Psi(A(x+y)) - \Psi(Ax) - \Psi(Ay)) \\ &= \frac{1}{2}(\Psi(Ax + Ay) - \Psi(Ax) - \Psi(Ay)) = \psi(Ax, Ay)\end{aligned}$$

e quindi φ, ψ sono congruenti.

Per dimostrare che la congruenza è una relazione di equivalenza, per ogni forma bilineare φ sullo spazio vettoriale V ed ogni isomorfismo lineare $A \colon V \to V$ conviene introdurre la notazione $A^* \varphi$ per indicare la forma bilineare

$$A^* \varphi(x, y) = \varphi(Ax, Ay).$$

Se $A, B\colon V \to V$ sono due applicazioni lineari invertibili allora

$$(AB)^*\varphi(x,y) = \varphi(ABx, ABy) = A^*\varphi(Bx, By) = B^*A^*\varphi(x,y)$$

da cui segue che $(AB)^*\varphi = B^*A^*\varphi$.

Denotiamo con $\sim$ la relazione di congruenza, cioè $\varphi \sim \psi$ se e soltanto se esiste $A\colon V \to V$ lineare invertibile tale che $\varphi = A^*\psi$. Tale relazione è:

(1) *Riflessiva* poiché $I^*\varphi = \varphi$, essendo I l'identità su V.
(2) *Simmetrica* poiché se $\varphi = A^*\psi$ allora $(A^{-1})^*\varphi = (A^{-1})^*A^*\psi = (AA^{-1})^*\psi = \psi$.
(3) *Transitiva* poiché se $\varphi = A^*\psi$ e $\psi = B^*\eta$ allora $\varphi = A^*B^*\eta = (BA)^*\eta$.

È utile tenere presente che, se due forme bilineari simmetriche φ, ψ sono congruenti, diciamo $\varphi = A^*\psi$ e $e_1, \dots, e_n$ è una base di V, allora $\epsilon_1 = Ae_1, \dots, \epsilon_n = Ae_n$ è ancora una base di V e vale $\varphi(e_i, e_j) = \psi(\epsilon_i, \epsilon_j)$ per ogni i, j. Viceversa se esistono due basi $e_1, \dots, e_n$ e $\epsilon_1, \dots, \epsilon_n$ di V tali che $\varphi(e_i, e_j) = \psi(\epsilon_i, \epsilon_j)$ per ogni i, j, allora, detta A l'applicazione lineare tale che $Ae_i = \epsilon_i$, allora $\varphi = A^*\psi$. Questo è sostanzialmente la stessa cosa del fatto che due forme bilineari sono congruenti se e solo se sono sono rappresentate dalla stessa matrice simmetrica in basi opportune. □

In maniera del tutto simile, se $\Phi\colon V \to \mathbb{K}$ è una forma quadratica e $A\colon W \to V$ è un'applicazione lineare si definisce la forma quadratica $A^*\Phi\colon W \to \mathbb{K}$ mediante la formula

$$A^*\Phi(w) = \Phi(Aw), \qquad w \in W.$$

In particolare, due forme quadratiche $\Phi, \Psi\colon V \to \mathbb{K}$ sono congruenti se e solo se esiste $A\colon V \to V$ lineare invertibile tale che $A^*\Phi = \Psi$.

DEFINIZIONE 15.3.12. Sia $\varphi\colon V \times V \to \mathbb{K}$ una forma bilineare simmetrica. Per ogni sottospazio vettoriale $W \subseteq V$ definiamo il suo **φ-ortogonale** come

$$W^\perp = \{v \in V \mid \varphi(v, w) = 0 \quad \text{per ogni} \quad w \in W\}.$$

In particolare $W \cap W^\perp$ coincide con il nucleo delle restrizione di φ a W e $W \subseteq (W^\perp)^\perp$. Un sottospazio vettoriale $W \subseteq V$ si dice:

(1) **anisotropo** se $W \cap W^\perp = 0$, ossia se la restrizione di φ a W è non degenere;
(2) **isotropo** se $W \cap W^\perp \neq 0$, ossia se la restrizione di φ a W è degenere;
(3) **totalmente isotropo** se $W \subseteq W^\perp$, ossia se $\varphi(x, y) = 0$ per ogni $x, y \in W$.

LEMMA 15.3.13. *Sia $\varphi\colon V \times V \to \mathbb{K}$ una forma bilineare simmetrica non degenere. Per ogni sottospazio vettoriale $W \subseteq V$ valgono le formule*

$$\dim W^\perp + \dim W = \dim V, \qquad (W^\perp)^\perp = W.$$

DIMOSTRAZIONE. Sia $w_1, \dots, w_m$ una base di W, allora $v \in W^\perp$ se e solo se $\varphi(w_i, v) = 0$ per ogni $i = 1, \dots, m$: il "solo se" è chiaro, mentre se $\varphi(w_i, v) = 0$ per ogni i allora per ogni $w = \sum a_i w_i \in W$ si ha $\varphi(v, w) = \sum a_i \varphi(w_i, v) = 0$. Dunque $W^\perp$ coincide con il nucleo dell'applicazione lineare

$$V \to \mathbb{K}^m, \qquad v \mapsto \begin{pmatrix} \varphi(w_1, v) \\ \vdots \\ \varphi(w_m, v) \end{pmatrix},$$

e questo prova che $\dim W^\perp \geq \dim V - m = \dim V - \dim W$. Sia H un complementare di W in V, ossia $H \oplus W = V$. Abbiamo visto che $\dim H^\perp + \dim H \geq \dim V$ e quindi che $\dim W^\perp + \dim H^\perp \geq \dim V$.

Ogni vettore $x \in W^\perp \cap H^\perp$ appartiene al nucleo di φ: infatti per ogni $v \in V$ possiamo scrivere $v = w + h$ con $w \in W$ e $h \in H$ ed allora $\varphi(x, v) = \varphi(x, w) + \varphi(x, h) = 0$. Siccome φ è non degenere si ha $W^\perp \cap H^\perp = 0$ $\dim W^\perp + \dim W = \dim V$ per la formula di Grassmann.

La medesima formula applicata a $W^\perp$ ci dice che $\dim(W^\perp)^\perp = \dim W$ e siccome $W \subseteq (W^\perp)^\perp$ per ovvii motivi si ha $W = (W^\perp)^\perp$

□

Esercizi.

740. Mostrare che il prodotto interno canonico ed il piano iperbolico sono non-degeneri.

741. Determinare tutti i vettori isotropi del piano iperbolico e del prodotto interno canonico su $\mathbb{C}^2$.

742. Provare che su un qualunque campo di caratteristica 2 il piano iperbolico non possiede basi ortogonali.

743. Sia φ la forma polare di una forma quadratica $\Phi\colon V \to \mathbb{K}$. Dimostrare che

$$\operatorname{Ker}\varphi = \{x \in V \mid \Phi(x+v) = \Phi(v) \quad \forall\, v \in V\}.$$

744. Sia $W \subseteq V$ un sottospazio totalmente isotropo rispetto ad una forma bilineare simmetrica non degenere. Provare che $2\dim W \leq \dim V$.

15.4. Applicazioni ortogonali e riflessioni

Come al solito lavoriamo in caratteristica diversa da 2 e su spazi vettoriali di dimensione finita. In tutta la sezione denoteremo con V uno spazio vettoriale di dimensione finita fissato dotato di un prodotto interno, ossia di una forma bilineare simmetrica non degenere $\varphi\colon V \times V \to \mathbb{K}$.

Definizione 15.4.1. Un'applicazione lineare $f\colon V \to V$ si dice **ortogonale** rispetto ad una forma bilineare simmetrica non degenere $\varphi\colon V \times V \to \mathbb{K}$, o più semplicemente φ-ortogonale, se

$$\varphi(fx, fy) = \varphi(x, y) \quad \text{per ogni} \quad x, y \in V.$$

Denoteremo con $O(V, \varphi) \subseteq \operatorname{Hom}_{\mathbb{K}}(V, V)$ l'insieme delle applicazioni φ-ortogonali.

Dall'ipotesi che φ è non degenere segue che ogni applicazione φ-ortogonale è invertibile: infatti se $fx = 0$, allora per ogni y vale $\varphi(x, y) = \varphi(fx, fy) = 0$ e quindi $x \in \operatorname{Ker}\varphi = 0$. È immediato osservare che $O(V, \varphi)$ è non vuoto (contiene l'identità) e che se $f, g \in O(V, \varphi)$, allora anche $f^{-1}, fg \in O(V, \varphi)$.

Esempio 15.4.2. Sia $W \subseteq V$ un sottospazio anisotropo, ossia tale che $W \cap W^\perp = 0$. Allora $V = W \oplus W^\perp$ ed esiste un'unica applicazione lineare

$$R_W\colon V \to V \quad \text{tale che} \quad \begin{cases} R_W(v) = v & \text{se } v \in W, \\ R_W(v) = -v & \text{se } v \in W^\perp, \end{cases}$$

chiamata **riflessione** ortogonale rispetto a W. È facile dimostrare che R_W è ortogonale rispetto a φ: dati $x_1, x_2 \in V$ possiamo scrivere $x_i = w_i + v_i$ con $w_i \in W$ e $v_i \in W^\perp$ e quindi, siccome $\varphi(w_i, v_j) = 0$ per ogni i, j si ha

$$\begin{aligned} \varphi(R_W x_1, R_W x_2) &= \varphi(w_1 - v_1, w_2 - v_2) = \varphi(w_1, w_2) + \varphi(v_1, v_2) \\ &= \varphi(w_1 + v_1, w_2 + v_2) = \varphi(x_1, x_2). \end{aligned}$$

Definizione 15.4.3. Sia $\varphi\colon V \times V \to \mathbb{K}$ bilineare simmetrica. Per ogni vettore $v \in V$ tale che $\varphi(v, v) \neq 0$ si definisce l'applicazione lineare

$$S_v\colon V \to V, \qquad S_v(x) = x - 2\,\frac{\varphi(x, v)}{\varphi(v, v)}\,v.$$

Se φ è non degenere, allora S_v coincide con la riflessione R_W rispetto all'iperpiano

$$W = \{w \in W \mid \varphi(v, w) = 0\}$$

ed è quindi φ-ortogonale. Infatti $v \notin W$, $v \in W^\perp$ e siccome $\dim W^\perp = 1$ si ha $W^\perp = \operatorname{Span}(v)$. Per ogni $w \in W$ ed ogni $a \in \mathbb{K}$ si ha

$$S_v w = w - 2\frac{\varphi(w, v)}{\varphi(v, v)}v = w = R_W w, \qquad S_v(av) = av - 2\frac{\varphi(av, v)}{\varphi(v, v)}v = -av = R_W(av).$$

OSSERVAZIONE 15.4.4. È importante osservare che se $\varphi(x,x)=\varphi(y,y)$ e $\varphi(x-y,x-y)\neq 0$, allora $\varphi(x-y,x+y)=0$, $S_{x-y}x=y$ e $S_{x-y}x=y$. Infatti

$$\varphi(x,x-y)=\varphi(x,x)-\varphi(x,y)=\varphi(y,y)-\varphi(y,x)=-\varphi(y,x-y),$$
$$\varphi(x-y,x-y)=2\varphi(x,x-y),$$

e quindi

$$S_{x-y}x=x-2\frac{\varphi(x,x-y)}{\varphi(x-y,x-y)}(x-y)=x-(x-y)=y.$$

TEOREMA 15.4.5 (Teorema di cancellazione di Witt). *Siano $\varphi\colon V\times V\to\mathbb{K}$ una forma bilineare simmetrica non degenere, $W\subseteq V$ un sottospazio vettoriale e $g\colon W\to V$ un'applicazione lineare* iniettiva *tale che*

$$\varphi(gx,gy)=\varphi(x,y)\qquad \text{per ogni}\quad x,y\in W.$$

Allora esiste un'applicazione lineare $f\colon V\to V$ tale che:

1. *f estende g, ossia $fx=gx$ per ogni $x\in W$;*
2. *f è composizione di un numero finito di riflessioni del tipo S_v e quindi $\varphi(fx,fy)=\varphi(x,y)$ per ogni $x,y\in V$.*

DIMOSTRAZIONE. La dimostrazione del teorema si ottiene mettendo assieme i seguenti due lemmi.

LEMMA 15.4.6. *Nelle ipotesi del Teorema 15.4.5 esiste un sottospazio anisotropo $U\subseteq V$ ed un'applicazione lineare iniettiva $f\colon U\to V$ tali che $W\subseteq U$, $f_{|W}=g$ e $\varphi(fx,fy)=\varphi(x,y)$ per ogni $x,y\in U$.*

DIMOSTRAZIONE. Dimostriamo il teorema per induzione sulla dimensione di $W\cap W^\perp$. Se $W\cap W^\perp=0$ basta porre $U=W$ e $f=g$. Supponiamo $W\cap W^\perp\neq 0$ e sia $e_1,\dots,e_m\in W$ una base φ-ortogonale: siccome $V^\perp=0$ si ha necessariamente $W\neq V$, ed a meno di permutazioni degli indici possiamo supporre $\varphi(e_m,e_m)=0$. Sappiamo che l'applicazione

$$V\to\mathbb{K}^m,\qquad x\mapsto(\varphi(x,e_1),\dots,\varphi(x,e_m))^T,$$

è surgettiva e quindi possiamo trovare un vettore $e_{m+1}\in V$ tale che $\varphi(e_m,e_{m+1})=1$ e $\varphi(e_i,e_{m+1})=0$ per ogni $i<m$; siccome $e_m\in W\cap W^\perp$ si ha $e_{m+1}\notin W$, per ogni $s\in\mathbb{K}$ si ha

$$\varphi(e_{m+1}+se_m,e_{m+1}+se_m)=\varphi(e_{m+1},e_{m+1})+2s,$$

e quindi, a meno di aggiungere ad e_{m+1} un opportuno multiplo scalare di e_m, possiamo anche supporre $\varphi(e_{m+1},e_{m+1})=0$.

Siccome $g(e_1),\dots,g(e_m)$ è una base φ-ortogonale di $g(W)$, in maniera del tutto simile possiamo trovare $v_{m+1}-V-g(W)$ tale che $\varphi(g(e_m),v_{m+1})=1$, $\varphi(g(e_i),v_{m+1})=0$ per ogni $i<m$ e $\varphi(v_{m+1},v_{m+1})=0$. Se denotiamo con U il sottospazio generato da $e_1,\dots,e_{m+1}$, allora possiamo estendere g ad un'applicazione iniettiva $f\colon U\to V$ ponendo $f(e_{m+1})=v_{m+1}$ e $f(e_i)=g(e_i)$ per $i\leq m$.

Per concludere resta da dimostrare che la dimensione di $U\cap U^\perp$ è strettamente minore di quella di $W\cap W^\perp$. Si ha $U\cap U^\perp\subseteq W$: infatti se $x=x_1e_1+\cdots+x_{m+1}e_{m+1}\in U\cap U^\perp$, allora $\varphi(x,e_m)=x_{m+1}=0$ e quindi $x\in W$. Siccome $W\subseteq U$ si ha $U^\perp\subseteq W^\perp$; quindi

$$U\cap U^\perp\subseteq W\cap W^\perp.$$

Ma $\varphi(e_m,e_{m+1})=1\neq 0$, il vettore e_m non appartiene a $U^\perp$ e quindi

$$U\cap U^\perp\neq W\cap W^\perp.$$

□

LEMMA 15.4.7. *Se, in aggiunta alle ipotesi del Teorema 15.4.5, il sottospazio W è anisotropo, allora esiste $f\colon V\to V$ che estende g e che è composizione di al più $2\dim W$ riflessioni del tipo S_v.*

DIMOSTRAZIONE. Siano $m=\dim W$ e $e_1,\dots,e_m\in W$ una base φ-ortogonale. Per ipotesi la restrizione $\varphi\colon W\times W\to\mathbb{K}$ è non degenere e quindi $\varphi(e_i,e_i)\neq 0$ per ogni $i=1,\dots,m$.

Denotiamo $v_i=g(e_i)$ e dimostriamo per induzione su $k=0,\dots,m$ che esiste un'applicazione $f_k\in O(V,\varphi)$, composizione si al più $2k$ riflessioni S_v e tale che $f(e_i)=\pm v_i$ per ogni $i\leq k$; per $k=0$ il risultato è vero per vacuità di condizioni.

Sia adesso $0 < k \leq m$ e supponiamo che esista $h\colon V \to V$, composizione di al più $2(k-1)$ riflessioni di tipo S_v e tale che $h(e_i) = v_i$ per ogni $i < k$. In particolare, ponendo $u_k = h(e_k)$, per le ipotesi fatte su g e h si ha

$$\Phi(u_k) = \Phi(v_k) = \Phi(e_k) \neq 0, \qquad \varphi(v_i, u_k) = \varphi(v_i, v_k) = \varphi(e_i, e_k) = 0 \quad \forall\, i < k.$$

Dalla formula

$$\Phi(u_k + v_k) + \Phi(u_k - v_k) = 2(\Phi(u_k) + \Phi(v_k)) = 4\Phi(e_k) \neq 0$$

segue che $\Phi(u_k+v_k)$ e $\Phi(u_k-v_k)$ non sono entrambi nulli. Se $\Phi(u_k-v_k) \neq 0$ allora $S_{u_k-v_k}(v_i) = v_i$ per ogni $i < k$ e per l'Ossevazione 15.4.4

$$S_{u_k-v_k}(u_k) = u_k - \frac{2\varphi(u_k, u_k - v_k)}{\varphi(u_k - v_k, u_k - v_k)}(u_k - v_k) = v_k.$$

Per dimostrare il passo induttivo basta quindi considerare $f = S_{u_k-v_k}h$. Se invece $\Phi(u_k+v_k) \neq 0$ allora $S_{v_k}S_{u_k+v_k}(v_i) = v_i$ per ogni $i < k$, $S_{v_k}S_{u_k+v_k}(u_k) = v_k$ e per dimostrare il passo induttivo basta quindi considerare $f = S_{v_k}S_{u_k+v_k}h$. □

L'utilizzo in sequenza dei due precedenti lemmi dimostra il teorema di cancellazione di Witt. □

Corollario 15.4.8. *Siano U, W due sottospazi totalmente isotropi rispetto ad una forma bilineare simmetrica non degenere $\varphi\colon V \times V \to \mathbb{K}$. Se $\dim U \leq \dim W$, allora esiste $f \in O(V, \varphi)$ tale che $f(U) \subseteq W$.*

Dimostrazione. Basta applicare il teorema di cancellazione di Witt a qualunque applicazione lineare iniettiva $g\colon U \to W$. □

Esercizi.

745. Determinare tutte le applicazioni ortogonali del piano iperbolico in sé.

746. Provare che l'applicazione lineare indotta da una matrice $A \in M_{n,n}(\mathbb{K})$ è ortogonale rispetto al prodotto interno canonico su $\mathbb{K}^n$ se e solo se $A^T = A^{-1}$.

747. Sia $B = (b_{ij})$ una matrice $n \times n$ simmetrica a coefficienti in un campo $\mathbb{K}$ di caratteristica $\neq 2$ e sia $H \subset \mathbb{K}^n$ un iperpiano di equazione $\sum_{i=1}^n a_i x_i = 0$. Provare che:

(1) H è isotropo per la forma $x^T By$ se e soltanto se esiste $x \in H$ tale che $Bx = (a_1, \ldots, a_n)^T$.
(2) H è isotropo per la forma $x^T By$ se e soltanto se

$$\det \begin{pmatrix} b_{11} & \ldots & b_{1n} & a_1 \\ \vdots & \ddots & \vdots & \vdots \\ b_{n1} & \ldots & b_{nn} & a_n \\ a_1 & \ldots & a_n & 0 \end{pmatrix} = 0.$$

Dedurre che ogni iperpiano è isotropo se e soltanto se il rango di B è $\leq n-2$.

15.5. Forme quadratiche reali e complesse

In un campo arbitrario, capire quando due forme quadratiche sono congruenti è in generale arduo. Fortunatamente ciò non vale nei campi dei numeri reali e complessi, dove le classi di congruenza sono univocamente determinate da pochi semplici invarianti numerici. Abbiamo già dimostrato che avere lo stesso rango è condizione necessaria affinché due forme quadratiche siano congruenti. Sui numeri complessi tale condizione è anche sufficiente.

Teorema 15.5.1. *Sia V uno spazio vettoriale di dimensione finita su $\mathbb{C}$. Due forme quadratiche $\Phi, \Psi\colon V \to \mathbb{C}$ sono congruenti se e solo se hanno lo stesso rango.*

DIMOSTRAZIONE. Sia $z_1, \dots, z_n$ un sistema di coordinate su V. Mostriamo che ogni forma quadratica Φ di rango r è congruente alla *forma standard*

$$I_r(z) = \sum_{i=1}^{r} z_i^2.$$

Sia φ la forma bilineare associata a Φ e consideriamo una base φ-ortogonale $\epsilon_1, \dots, \epsilon_n$. Abbiamo visto che il rango di φ (=rango di Φ) è uguale a numero di indici i tali che $\Phi(\epsilon_i) \neq 0$; a meno di permutazioni di indici possiamo supporre $\Phi(\epsilon_i) = 0$ se $i > r$ e $\Phi(\epsilon_i) = \lambda_i \neq 0$ se $i \leq r$. Scegliamo per ogni $i \leq r$ una radice quadrata μ_i di λ_i e consideriamo la nuova base

$$v_i = \epsilon_i \text{ se } i > r, \qquad v_i = \frac{\epsilon_i}{\mu_i} \text{ se } i \leq r.$$

Per costruzione vale $\Phi(v_i) = 0$ se $i > r$ e $\Phi(v_i) = 1$ se $i \leq r$ e quindi Φ è congruente alla forma I_r. □

Sui numeri reali due forme quadratiche possono avere lo stesso rango senza tuttavia essere congruenti; è quindi necessario introdurre un ulteriore invariante.

DEFINIZIONE 15.5.2. Una forma quadratica $\Phi\colon V \to \mathbb{R}$ su di uno spazio vettoriale reale V si dice:

(1) **definita positiva** (e talvolta scriveremo $\Phi > 0$) se $\Phi(x) > 0$ per ogni $x \in V$, $x \neq 0$.
(2) **definita negativa** ($\Phi < 0$) se $\Phi(x) < 0$ per ogni $x \in V$, $x \neq 0$, o equivalentemente se $-\Phi$ è definita positiva.
(3) **semidefinita positiva** ($\Phi \geq 0$) se $\Phi(x) \geq 0$ per ogni $x \in V$.
(4) **semidefinita negativa** ($\Phi \leq 0$) se $\Phi(x) \leq 0$ per ogni $x \in V$, o equivalentemente se $-\Phi$ è semidefinita positiva.
(5) **indefinita** in tutti gli altri casi, ossia se esistono $v, w \in V$ tali che $\Phi(v) > 0$ e $\Phi(w) < 0$.

Le stesse denominazioni si applicano alle forme bilineari considerando le forme quadratiche associate.

Ad esempio la forma quadratica associata al prodotto interno canonico su $\mathbb{R}^n$ è definita positiva. Una forma bilineare simmetrica reale si dice un **prodotto scalare** se la forma quadratica associata è definita positiva. Spesso il prodotto interno canonico sugli spazi $\mathbb{R}^n$ viene detto **prodotto scalare canonico**.

Sia φ è una forma bilineare simmetrica; in un sistema di coordinate $x_1, \dots, x_n$ corrispondente ad una base φ-ortogonale, la forma quadratica associata si scrive

$$\Phi(x) = \sum_{i=1}^{n} \lambda_i x_i^2, \qquad \lambda_i \in \mathbb{R}.$$

È immediato osservare che $\Phi > 0$ (risp.: $\Phi \geq 0$, $\Phi < 0$, $\Phi \leq 0$) se e solo se $\lambda_i > 0$ (risp.: $\lambda_i \geq 0$, $\lambda_i < 0$, $\lambda_i \leq 0$) per ogni $i = 1, \dots, n$.

LEMMA 15.5.3. *Sia $\Phi\colon V \to \mathbb{R}$ una forma quadratica, $A\colon V \to V$ lineare invertibile e $W \subseteq V$ un sottospazio vettoriale. Allora la restrizione di $A^*\Phi$ a W è definita positiva se e solo se lo è anche la restrizione di Φ a $A(W)$.*

Identico risultato si ottiene sostituendo il termine definita con semidefinita e/o positiva con negativa.

DIMOSTRAZIONE. Basta ricordarsi che per ogni $w \in W$ vale $A^*\Phi(w) = \Phi(Aw)$. □

Dal precedente lemma segue immediatamente che, per ogni forma quadratica Φ, i due numeri:

(1) **indice di positività** $\Phi_+ =$ massima dimensione di un sottospazio $W \subseteq V$ tale che la restrizione di Φ a W è definita positiva.
(2) **indice di negatività** $\Phi_- = (-\Phi)_+ =$ massima dimensione di un sottospazio $W \subseteq V$ tale che la restrizione di Φ a W è definita negativa.

sono invarianti per congruenza.

DEFINIZIONE 15.5.4. La coppia (Φ_+, Φ_-) e la differenza $\Phi_+ - \Phi_-$ sono dette rispettivamente **segnatura** e **indice** della forma quadratica Φ. [2]

TEOREMA 15.5.5 (Teorema di Sylvester). *Data una forma quadratica reale in forma diagonale (ossia in un sistema di coordinate $x_1, \dots, x_n$ relativo ad una base ortogonale)*

$$\Phi(x) = \sum_{i=1}^{n} \lambda_i x_i^2, \qquad \lambda_i \in \mathbb{R},$$

si considerino i numeri:

(1) $p =$ *numero di indici i tali che $\lambda_i > 0$.*
(2) $q =$ *numero di indici i tali che $\lambda_i < 0$.*

Allora vale $\Phi_+ = p$ e $\Phi_- = q$. In particolare il rango di φ è uguale alla somma $\Phi_+ + \Phi_-$, gli interi p e q non dipendono dalla particolare base ortogonale scelta e sono invarianti per congruenza.

DIMOSTRAZIONE. È sufficiente dimostrare che $p = \Phi_+$: infatti considerando la forma quadratica opposta $-\Phi$ i coefficienti λ_i cambiano tutti di segno e $(-\Phi)_+ = \Phi_-$. A meno di permutazione di indici possiamo supporre $\lambda_1, \dots, \lambda_p > 0$, $\lambda_{p+1}, \dots, \lambda_{p+q} < 0$ e $\lambda_{p+q+1}, \dots, \lambda_n = 0$. Sia $L \subseteq V$ il sottospazio definito dalle equazioni $x_{p+1} = \cdots = x_n = 0$ e denotiamo con $\pi\colon V \to L$ la proiezione sulle prime p coordinate $\pi(x_1, \dots, x_n) = (x_1, \dots, x_p, 0, \dots, 0)$. La restrizione di Φ al sottospazio L è definita positiva e quindi, per definizione di Φ_+, si ha $p \le \Phi_+$.

Viceversa, sempre per definizione di Φ_+, esiste un sottospazio vettoriale $W \subseteq V$ di dimensione Φ_+ tale che $\Phi(x) > 0$ per ogni $x \in W$, $x \neq 0$. Dimostriamo che l'applicazione lineare $\pi\colon W \to L$ è iniettiva; da questo seguirà che $\Phi_+ = \dim W \le \dim L = p$.

Sia dunque $x = (x_1, \dots, x_n) \in W$ tale che $\pi(x) = 0$, ciò significa che $x_1 = \dots = x_p = 0$ e quindi che $\Phi(x) = \sum_{i>p} \lambda_i x_i^2 \le 0$. D'altra parte Φ è definita positiva su W e la condizione $\Phi(x) \le 0$ implica necessariamente $x = 0$. □

È immediato osservare che la segnatura è univocamente determinata da rango e indice.

COROLLARIO 15.5.6. *Due forme quadratiche definite su di uno spazio vettoriale reale V sono congruenti se e solo se hanno la stessa segnatura.*

DIMOSTRAZIONE. Fissiamo un isomorfismo $V \cong \mathbb{R}^n$ (cioè fissiamo una base); consideriamo poi, per ogni coppia di interi positivi p, q tali che $p + q \le n$, la forma quadratica

$$I_{p,q}(x) = x_1^2 + x_2^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_{p+q}^2.$$

e dimostriamo che ogni forma quadratica è congruente ad $I_{p,q}$ per opportuni p, q. Notiamo poi che, per il teorema di Sylvester la coppia (p, q) è uguale alla segnatura.

Sia dunque Φ una forma quadratica; sappiamo che esiste un sistema di coordinate lineari $y_1, \dots, y_n$ su $\mathbb{R}^n$ rispetto al quale

$$\Phi(y) = \sum_{i=1}^{n} \lambda_i y_i^2$$

A meno di permutazioni di indici possiamo assumere:

(1) $\lambda_i > 0$ se $i \le p$;
(2) $\lambda_i < 0$ se $p < i \le p + q$;
(3) $\lambda_i = 0$ se $i > p + q$.

Nel sistema di coordinate

$$\begin{cases} z_i = \sqrt{\lambda_i}\, y_i & \text{se } i \le p \\ z_i = \sqrt{-\lambda_i}\, y_i & \text{se } p < i \le p + q \\ z_i = y_i & \text{se } p + q < i \end{cases}$$

[2] Alcuni autori chiamano **segnatura** la differenza $\Phi_+ - \Phi_-$; altri autori chiamano segnatura la terna (Φ_0, Φ_+, Φ_-) dove Φ_0, detto indice di nullità, è la dimensione del nucleo della forma bilineare simmetrica associata.

la forma quadratica diventa

$$\Phi(z) = \sum_{i=1}^{p} z_i^2 - \sum_{i=p+1}^{p+q} z_i^2$$

che è congruente a $I_{p,q}$. □

ESEMPIO 15.5.7. Calcoliamo, al variare del parametro $\lambda \in \mathbb{R}$, il rango e la segnatura della forma quadratica:

$$\Phi\colon \mathbb{R}^2 \to \mathbb{R}, \qquad \Phi(x) = x_1^2 + 2\lambda x_1 x_2 - x_2^2.$$

Nella base canonica e_1, e_2 la matrice simmetrica corrispondente è

$$B_\lambda = \begin{pmatrix} 1 & \lambda \\ \lambda & -1 \end{pmatrix}$$

Osserviamo che $\Phi(e_1) > 0$, da cui deduciamo $\Phi_+ \geq 1$, e che $\Phi(e_2) < 0$ da cui deduciamo $\Phi_- \geq 1$. Siccome $\Phi_+ + \Phi_- \leq 2$ dovrà necessariamente essere $\Phi_+ = \Phi_- = 1$. Dunque Φ ha rango 2 e segnatura $(1,1)$, indipendentemente dal valore di λ, e le matrici B_λ sono tutte congruenti tra loro.

Il segno del determinante di una matrice simmetrica è invariante per congruenza. Infatti se B, C sono matrici reali simmetriche congruenti, allora esiste una matrice invertibile A tale che $A^T B A = C$ e quindi

$$\det C = \det(A^T)\det(B)\det(A) = \det(B)\det(A)^2.$$

Siccome ogni matrice simmetrica è congruente ad una matrice diagonale (Corollario 15.3.10), segue immediatamente dal teorema di Sylvester che, per una forma quadratica

$$\Phi\colon \mathbb{R}^n \to \mathbb{R}, \qquad \Phi(x) = x^T B x,$$

vale

(1) $\det(B) = 0$ se e solo se $\operatorname{rg}(\Phi) < n$.
(2) $\det(B) > 0$ se e solo se $\operatorname{rg}(\Phi) = n$ e Φ_- pari.
(3) $\det(B) < 0$ se e solo se $\operatorname{rg}(\Phi) = n$ e Φ_- dispari.

Riprendendo le notazioni della Sezione 10.6, per ogni matrice $A \in M_{n,n}(\mathbb{R})$ ed ogni $k \leq n$ denotiamo con $A[k]$ la sottomatrice formata dalle prime k righe e k colonne.

COROLLARIO 15.5.8 (Criterio di Sylvester). *Sia $A \in M_{n,n}(\mathbb{R})$ una matrice simmetrica tale che $\det(A[k]) \neq 0$ per ogni $k = 1, \ldots, n$. Allora l'indice di negatività delle forma quadratica $x \mapsto x^T A x$ è uguale al numero di cambiamenti di segno della successione $1, \det(A[1]), \ldots, \det(A[n])$.*

DIMOSTRAZIONE. Ragioniamo per induzione su n. Denotiamo con ϕ la forma bilineare simmetrica $\phi(x,y) = x^T A y$, con Φ la forma quadratica associata e con $\Phi_\pm$ i suoi indici di posività e nagatività. Siccome $\det(A) \neq 0$, per il teorema di Sylvester vale $\Phi_+ + \Phi_- = n$.

Osserviamo che l'ipotesi $\det(A[k]) \neq 0$ è del tutto equivalente a dire che la restrizione di ϕ al sottospazio generato dai primi k vettori della base canonica è nondegenere.

Denotiamo con $\mathbb{R}^{n-1} \subset \mathbb{R}^n$ il sottospazio generato dai primi $n-1$ vettori della base canonica e con a, b gli indici di positività e negatività della restrizione di Φ a $\mathbb{R}^{n-1}$. Per induzione possiamo supporre che b è uguale al numero di cambiamenti di segno della successione $1, \det(A[1]), \ldots, \det(A[n-1])$

Chiaramente $a \leq \Phi_+$, $b \leq \Phi_-$ e siccome anche $\det(A[n-1]) \neq 0$ si ha $a + b = n - 1$. Basta quindi dimostrare che: se i determinanti di $A[n-1]$ e A hanno lo stesso segno allora $\Phi_+ = a + 1$; se i determinanti di $A[n-1]$ e A hanno segni opposti allora $\Phi_- = b + 1$.

Denotando come al solito con $e_1, \ldots, e_n$ la base canonica, sia $v_n = A^{-1}(e_n)$; allora $w^T A[n-1] v_n = w^T e_n = 0$ per ogni $w \in \mathbb{R}^{n-1}$ e, poiché la restrizione di ϕ a $\mathbb{R}^{n-1}$ è non degenere, si ha $v_n \notin \mathbb{R}^{n-1}$.

Sia adesso $v_1, \ldots, v_{n-1}$ una base ϕ-ortogonale di $\mathbb{R}^{n-1}$, allora $v_1, \ldots, v_n$ è una base ϕ-ortogonale di $\mathbb{R}^{n-1}$. Siccome il segno del determinante è invariate per congruenza si ha che il segno di $\det(A[n])$ è ugual al segno del prodotto $\lambda \det(A[n-1])$. □

ESEMPIO 15.5.9. Calcoliamo, in funzione di $\lambda \in \mathbb{R}$, rango e segnatura della matrice simmetrica

$$B_\lambda = \begin{pmatrix} 1 & 1 \\ 1 & \lambda \end{pmatrix},$$

ossia rango e segnatura della forma quadratica associata. Sia ha $\Phi(e_1) = 1$, dove e_1, e_2 denota la base canonica e Φ la forma quadratica associata alla matrice: dunque $\Phi_+ \geq 1$ e $\Phi_- \leq 1$ Il determinante è uguale a $\lambda - 1$ e quindi

(1) per $\lambda = 1$ il rango è 1 e la segnatura è (1,0).
(2) per $t > 1$ il rango è 2, $|B_\lambda| > 0$ e quindi Φ_- è pari. Necessariamente $\Phi_- = 0$ e la segnatura è $(2, 0)$.
(3) per $\lambda < 1$ il rango è 2, $|B_\lambda| < 0$ e quindi Φ_- è dispari. Necessariamente $\Phi_- = 1$ e la segnatura è $(1, 1)$.

Attenzione: per $0 < \lambda < 1$ tutti i coefficienti di B_λ sono positivi e tuttavia la forma quadratica $x^T B_\lambda x$ non è definita positiva.

ESEMPIO 15.5.10. Calcoliamo il rango e la segnatura della forma quadratica Φ associata alla matrice simmetrica

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

La diagonale contiene sia valori positivi che negativi e quindi $\Phi_+ \geq 1$, $\Phi_- \geq 1$. Il determinante della matrice è $-5 < 0$, quindi Φ_- è dispari e di conseguenza $\Phi_+ = 2$, $\Phi_- = 1$.

ESEMPIO 15.5.11. Calcoliamo gli indici Φ_+, Φ_- della forma quadratica Φ associata alla matrice simmetrica

$$\begin{pmatrix} 1 & 1 & 350! \\ 1 & 2 & \pi \\ 350! & \pi & -3 \end{pmatrix}$$

La diagonale contiene sia valori positivi che negativi e quindi $\Phi_+ \geq 1$, $\Phi_- \geq 1$. La restrizione della forma quadratica al sottospazio U generato dai primi due vettori della base canonica è rappresentata dalla matrice

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

e quindi è definita positiva per quanto visto nell'Esempio 15.5.9. Quindi $\Phi_+ \geq 2$ e dunque $\Phi_+ = 2$ e $\Phi_- = 1$.

ESEMPIO 15.5.12. Calcoliamo rango e segnatura della forma quadratica

$$\Phi \colon \mathbb{R}^3 \to \mathbb{R}, \quad \Phi(x) = x_1^2 - 2x_1x_2 - 2x_1x_3.$$

Denotiamo con φ la polare di Φ, ossia laforma bilineare simmetrica associata. Nella base canonica e_1, e_2, e_3 la matrice simmetrica corrispondente $B = (\varphi(e_i, e_j))$ è

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

Applichiamo l'algoritmo di ortogonalizzazione illustrato nella seconda dimostrazione del Teorema 15.3.8; poiché $\varphi(b_1, b_1) \neq 0$, nel primo passaggio si considera la nuova base

- $c_1 = b_1 = (1, 0, 0)$,
- $c_2 = b_2 - \dfrac{\varphi(b_2, b_1)}{\varphi(b_1, b_1)} b_1 = b_2 + b_1 = (1, 1, 0)$,
- $c_3 = b_3 - \dfrac{\varphi(b_3, b_1)}{\varphi(b_1, b_1)} b_1 = b_3 + b_1 = (1, 0, 1)$.

La matrice $C = (\varphi(c_i, c_j))$ è

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

Siccome $\varphi(c_2, c_2) \neq 0$ si ripete l'algoritmo costruendo la nuova base:

- $d_1 = c_1 = (1, 0, 0)$,

- $d_2 = c_2 = (1, 1, 0)$,
- $d_3 = c_3 - \dfrac{\varphi(c_3, c_2)}{\varphi(c_2, c_2)} c_2 = c_3 - c_2 = (0, -1, 1)$.

La matrice $D = (\varphi(d_i, d_j))$ è

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

e quindi, per il teorema di Sylvester, la forma Φ ha rango 2 e segnatura $(1, 1)$.

TEOREMA 15.5.13 (Disuguaglianze di Cauchy–Schwarz). *Siano φ una forma bilineare su uno spazio vettoriale reale V e $u, v \in V$. Si denoti con Φ la forma quadratica associata a φ e con $U \subseteq V$ il sottospazio vettoriale generato da u, v. Allora:*

(1) *se Φ è definita (positiva o negativa) su U allora $\varphi(u, v)^2 \leq \Phi(u)\Phi(v)$, ed il segno $=$ vale se e solo se u, v sono linearmente dipendenti;*

(2) *se Φ è indefinita e non degenere su U allora $\varphi(u, v)^2 \geq \Phi(u)\Phi(v)$ ed il segno $=$ vale se e solo se u, v sono linearmente dipendenti;*

(3) *se Φ è degenere su U allora $\varphi(u, v)^2 = \Phi(u)\Phi(v)$.*

DIMOSTRAZIONE. Quando diciamo che Φ è definita positiva su U si intende che la restrizione $\Phi\colon U \to \mathbb{K}$ è definita positiva. Se $u = 0$ oppure $v = 0$ il risultato è chiaro. Se $u, v \neq 0$ sono linearmente dipendenti si ha $v = su$ per qualche $s \in \mathbb{R}$, $s \neq 0$, e tutto segue dalle formule

$$\phi(u, v) = s\Phi(u), \qquad \Phi(v) = s^2\Phi(u).$$

Non è quindi restrittivo supporre u, v vettori linearmente indipendenti, ossia $\dim U = 2$. Sia $x, y \in U$ una base φ-ortogonale e scriviamo

$$u = ax + by, \quad v = cx + dy, \quad ad - cb \neq 0, \quad (ad - cb)^2 > 0.$$

Siccome

$$\begin{aligned} \varphi(u, v) &= \varphi(ax + by, cx + dy) = ac\Phi(x) + bd\Phi(y) \\ \Phi(u) &= \varphi(ax + by, ax + by) = a^2\Phi(x) + b^2\Phi(y) \\ \Phi(v) &= \varphi(cx + dy, cx + dy) = c^2\Phi(x) + d^2\Phi(y) \end{aligned}$$

si ha

$$\Phi(u)\Phi(v) - \varphi(u, v)^2 = (ad - bc)^2\Phi(x)\Phi(y)$$

e per concludere basta osservare che $\Phi(x)\Phi(y) > 0$ se la restrizione di Φ ad U è definita, $\Phi(x)\Phi(y) < 0$ se la restrizione di Φ ad U è indefinita e $\Phi(x)\Phi(y) = 0$ se la restrizione di Φ ad U è degenere. □

ESEMPIO 15.5.14. Nei testi di relatività ristretta si introduce il cosiddetto prodotto scalare di Minkowski e si classificano i vettori dello spazio-tempo come di tipo tempo, di tipo spazio e di tipo luce. La prossima proposizione studia alcune proprietà matematiche dell'insieme, denotato T, dei vettori di tipo tempo.

PROPOSIZIONE 15.5.15. *Sia V uno spazio vettoriale reale di dimensione $n+1$ e sia $\langle -, - \rangle$ una forma bilineare simmetrica su V di segnatura $(1, n)$. Si considerino i due sottoinsiemi*

$$T = \{v \in V \mid \langle v, v \rangle > 0\}, \qquad \overline{T} = \{v \in V \mid v \neq 0,\ \langle v, v \rangle \geq 0\}.$$

Allora:

(1) *sia $W \subseteq V$ un sottospazio vettoriale di dimensione $m + 1$. Allora la restrizione di $\langle -, - \rangle$ a W ha segnatura $(1, m)$ se e solo se $W \cap T \neq \emptyset$;*
(2) *per ogni $v \in T$ ed ogni $w \in \overline{T}$ vale $\langle v, w \rangle \neq 0$;*
(3) *per ogni $u \in T$, $v, w \in \overline{T}$ tali che $\langle u, v \rangle > 0$ e $\langle u, w \rangle > 0$ si ha $\langle v, w \rangle \geq 0$.*

In particolare, la relazione $\sim$ sull'insieme T definita come $u \sim v$ se $\langle u, v \rangle > 0$, è una relazione di equivalenza.

DIMOSTRAZIONE. Dato che la segnatura è $(1, n)$ si ha $T \neq \emptyset$ ed esiste un iperpiano $N \subseteq V$ in cui la forma bilineare è definita negativa.

(*1*) Dato W di dimensione $m+1$, se la restrizione di $\langle -, - \rangle$ ha segnatura $(1, m)$ allora W contiene almeno un vettore v con $\langle v, v \rangle > 0$ e quindi $T \cap W \neq \emptyset$. Viceversa, se esiste $v \in T \cap W$ allora $\mathrm{Span}(v)$ è un sottospazio di dimensione 1 in cui la forma è definita positiva, mentre $N \cap W$ è un sottospazio di dimensione $\geq m$ in cui la forma è definita negativa.

(*2*) Se fosse per assurdo $\langle v, w \rangle = 0$, con $v \in T$ e $w \in \overline{T}$ allora v, w devono necessariamente essere linearmente indipendenti, e la forma bilineare ristretta a $W = \mathrm{Span}(v, w)$ sarebbe semidefinita positiva, in contraddizione con il fatto che $W \cap N \neq 0$.

(*3*) Se per assurdo esistessero $u \in T$, $v, w \in \overline{T}$ tali che $\langle u, v \rangle > 0$, $\langle u, w \rangle > 0$ e $\langle v, w \rangle < 0$, allora per ogni numero reale non negativo $t \geq 0$ si ha $u + tw \in T$ e tuttavia $\langle v, u + tw \rangle = \langle v, u \rangle + t\langle v, w \rangle$ si annullerebbe per $t = -\langle v, u \rangle / \langle v, w \rangle$ in contraddizione con il punto precedente. □

La relazione $\sim$ introdotta nella precedente proposizione possiede esattamemte due classi di equivalenza. Infatti, sia $v \in T$ un vettore fissato e siano $u, w \in T$ non equivalenti a v, ossia tali che $\langle v, u \rangle \leq 0$ e $\langle v, w \rangle \leq 0$. Per il punto (*2*) si ha $\langle v, u \rangle < 0$ e $\langle v, w \rangle < 0$, quindi $v \sim -u$, $v \sim -w$ e per la proprietà transitiva $-u \sim -w$, che implica immediatamente $u \sim w$.

Si ha dunque una partizione di T in classi di equivalenza:

$$T = T_+ \cup T_-, \quad \text{dove} \quad T_+ = \{u \in T \mid \langle v, u \rangle > 0\}, \quad T_- = \{u \in T \mid \langle -v, u \rangle > 0\}.$$

Se $u, w \in T_+$ allora $tu + (1-t)w \in T_+$ per ogni $0 \leq t \leq 1$: infatti siccome $\langle v, u \rangle > 0$ e $\langle v, w \rangle > 0$ si ha $\langle v, tu + (1-t)w \rangle = t\langle v, u \rangle + (1-t)\langle v, w \rangle > 0$. Abbiamo quindi dimostrato che ogni classe di equivalenza è un sottoinsieme convesso di V.

Inoltre, dato $v \in T$, il punto (*2*) implica

$$\{w \in \overline{T} \mid \langle v, w \rangle \geq 0\} = \{w \in \overline{T} \mid \langle v, w \rangle > 0\}, \tag{15.2}$$

ed il punto (*3*) implica che l'insieme (15.2) dipende solo dalla classe di equivalenza di v.

Esercizi.

748. Sia φ una forma bilineare simmetrica reale di rango r e indice σ. Provare:
   (1) $r - \sigma$ è pari.
   (2) φ è semidefinita positiva se e solo se $r = \sigma$.
   (3) φ è semidefinita negativa se e solo se $r = -\sigma$.
   (4) φ è semidefinita se e solo se ogni vettore isotropo appartiene al nucleo di φ.
   (5) Usare i punti 1, 2 e 4 per trovare, senza bisogno di far calcoli, rango e indice della forma quadratica dell'Esempio 15.5.12.

749. Sia $t > 1$ numero reale. Dimostrare che la matrice simmetrica A di coefficienti $a_{ij} = t^{(i-1)(j-1)}$, $i, j = 1, \ldots, n$, è definita positiva. Mostrare inoltre che se $0 < t < 1$ allora l'indice di A dipende solo dalla classe di resto modulo 2 di n.

750. Provare che una forma quadratica reale è definita se e solo se 0 è l'unico vettore isotropo. Si noti che l'unica verifica non banale è mostrare che se una forma quadratica è indefinita allora esiste un vettore isotropo non nullo.

751. Calcolare rango e segnatura delle forme quadratiche reali

$$q, s \colon M_{2,2}(\mathbb{R}) \to \mathbb{R}, \qquad q(A) = \det(A), \quad s(A) = \mathrm{Tr}(A^2).$$

752. Sia $B \in M_{n,n}(\mathbb{R})$ una matrice simmetrica non nulla con tutti gli elementi sulla diagonale uguali a 0. Dimostrare che la forma quadratica reale associata è indefinita. (Suggerimento: considerare la restrizione ai sottospazi generati da due vettori della base canonica.)

753. Sia $H \subset M_{3,3}(\mathbb{R})$ il sottospazio vettoriale delle matrici simmetriche a traccia nulla. Calcolare rango e segnatura della forma quadratica $\Phi \colon H \to \mathbb{R}$, $\Phi(A) = \mathrm{Tr}(A^2)$.

754. Per ogni $n \geq 3$ sia $B_n \in M_{n,n}(\mathbb{R})$ la matrice simmetrica di coefficienti

$$b_{ij} = \begin{cases} 0 & \text{se } 2 \leq i \leq n-1 \text{ e } 2 \leq j \leq n-1 \\ 1 & \text{altrimenti} \end{cases}$$

Determinare rango e segnatura di B_3, B_4.

755. Determinare rango e segnatura di B_n introdotta nell'Esercizio 754 per ogni $n > 4$.

756. Siano $f, g\colon V \to \mathbb{R}$ lineari e linearmente indipendenti come vettori di $\mathrm{Hom}_{\mathbb{R}}(V, \mathbb{R})$. Mostrare che $\Phi(x) = f(x)g(x)$ è una forma quadratica di rango 2 e segnatura $(1, 1)$. Cosa si può dire se f e g sono linearmente dipendenti?

757. Calcolare rango e segnatura della forma quadratica

$$\Phi\colon \mathbb{R}^{350} \to \mathbb{R}, \qquad \Phi(x) = \sum_{i=2}^{350} (2x_1x_i + x_i^2).$$

758. Calcolare rango e segnatura della forma quadratica

$$\Phi\colon \mathbb{R}^{350} \to \mathbb{R}, \qquad \Phi(x) = (x_1 + 2x_2 + 3x_3 + \cdots + 350x_{350})^2 - x_1^2.$$

759. Sia $A = (A_{ij})$ una matrice simmetrica reale $n \times n$ di rango 1 semidefinita positiva. Dimostrare che esistono $a_1, \ldots, a_n \in \mathbb{R}$ tali che $A_{ij} = a_i a_j$ per ogni i, j.

760. Sia $(a_{ij}) \in M_{n,n}(\mathbb{R})$ simmetrica e semidefinita positiva. Provare che

$$\max_{i,j} |a_{ij}| = \max_i a_{ii}.$$

761 (☕, ♡, Lemma di Zariski). Sia $A = (a_{ij})$ una matrice simmetrica reale $n \times n$ tale che $2a_{ii} \geq \sum_{j=1}^{n} |a_{ij}|$ per ogni indice i. Provare che A è semidefinita positiva.

762. Determinare rango e segnatura della matrice

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

763. Siano $A, B \in M_{2,2}(\mathbb{R})$ simmetriche e definite positive. È vero o falso che la matrice simmetrica $AB + BA$ è definita positiva?

764. Sia V lo spazio vettoriale su $\mathbb{R}$ una cui base è data da $\cos t$, $\sin t$. Si definisca

$$\langle f,\, g\rangle := \int_{-\pi}^{\pi} f(x)g(x)dx.$$

Verificare che si tratta di un prodotto scalare . Trovare la matrice associata rispetto alla base data.

765. Sia V lo spazio vettoriale dei polinomi reali in una variabile x di grado minore o uguale a 2. Dimostrare che ponendo, per $p, q \in V$

$$\langle p, q\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$$

si definisce un prodotto scalare in V. Trovare la matrice dell'operatore $T\colon V \to V$ definito da

$$\langle Tp, q\rangle = \langle p, \frac{dq}{dx}\rangle$$

nella base $1, x, \frac{3}{2}x^2 - \frac{1}{2}$.

766. Determinare rango e segnatura della forma quadratica $\Phi\colon \mathbb{R}^4 \to \mathbb{R}$

$$\Phi(x) = x_1x_2 + x_2x_3 + x_3x_4.$$

767. Determinare tutti i vettori isotropi della forma quadratica $\Phi\colon \mathbb{R}^4 \to \mathbb{R}$ associata alla matrice

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

768. Si consideri la forma quadratica in $\mathbb{R}^3$ definita da

$$\Phi(x) = 2x_1x_2 + 6x_2x_3$$

a) Se ne calcoli rango e segnatura.
b) Si trovi una base di $\mathbb{R}^3$ in cui Φ è scritta in forma canonica

769. Sia $B = (b_{ij})$ una matrice reale $n \times n$ simmetrica Definiamo un'applicazione $\varphi\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ mediante la formula

$$\varphi(x, y) = \det \begin{pmatrix} b_{11} & \dots & b_{1n} & y_1 \\ \vdots & \ddots & \vdots & \vdots \\ b_{n1} & \dots & b_{nn} & y_n \\ x_1 & \dots & x_n & 0 \end{pmatrix}.$$

Dimostrare:

(1) φ è bilineare simmetrica. Interpretare, in funzione di B, i coefficienti della matrice A tale che $\varphi(x, y) = x^T Ay$.
(2) Determinare, in funzione del rango di B, il rango di φ.
(3) Se $\det B \neq 0$ determinare, in funzione della segnatura di B, la segnatura di φ.

770 (☕). Siano $A, B \in M_{n,n}(\mathbb{R})$ simmetriche con $A = (a_{ij})$ definita positiva e $B = (b_{ij})$ semidefinita positiva. Dimostrare che:

(1) dati $x_1, \dots, x_n \in \mathbb{R}$, la matrice di coefficienti $c_{ij} = x_i a_{ij} x_j$ è semidefinita positiva di rango uguale alla quantità di x_i diversi da 0;
(2) $\operatorname{Tr}(AB) \geq 0$ e vale $\operatorname{Tr}(AB) = 0$ se e solo se $B = 0$;
(3) se B è definita positiva allora la matrice di coefficienti $c_{ij} = a_{ij}b_{ij}$ è definita positiva.

15.6. Eliminazione di Gauss simmetrica

Abbiamo visto che due matrici simmetriche sono congruenti se e solo se rappresentano la stessa forma bilineare $\varphi\colon V \times V \to \mathbb{K}$ rispetto a due basi dello spazio vettoriale V.

Sia $e_1, \dots, e_n$ una base di V e consideriamo la matrice $B = (b_{ij}) = (\varphi(e_i, e_j))$. Vediamo come si trasforma la matrice B quando agiamo sulla base mediante una delle seguenti operazioni elementari già viste in relazione all'eliminazione di Gauss:

(1) moltiplicare un vettore della base per uno scalare non nullo;
(2) aggiungere ad un vettore della base un multiplo scalare di un altro vettore della base;
(3) scambiare tra di loro due vettori della base.

Analizziamo come cambia la matrice B in ciascuno dei tre casi.

(1) se moltiplichiamo e_h per $\lambda \neq 0$, ossia se consideriamo la nuova base

$$\varepsilon_h = \lambda e_h, \qquad \varepsilon_i = e_i \quad \text{per } i \neq h,$$

allora la matrice $C = (c_{ij}) = (\varphi(\varepsilon_i, \varepsilon_j))$ che rappresenta ϕ nella nuova base soddisfa le uguaglianze

$$\begin{cases} c_{hh} = \lambda^2 b_{hh} & \\ c_{hi} = \lambda b_{hi}, \quad \lambda c_{ih} = \lambda b_{ih} & \text{per } i \neq h, \\ c_{ij} = b_{ij} & \text{per } i, j \neq h. \end{cases}$$

Equivalentemente C si ottiene da B moltiplicando prima la h-esima riga per λ e poi moltiplicando per λ la colonna h-esima della matrice così ottenuta.

(2) cambiamo adesso la base aggiungendo a e_h un multiplo scalare di e_k, con $h \neq k$, ossia consideriamo una nuova base $\varepsilon_1, \dots, \varepsilon_n$ dove

$$\varepsilon_h = e_h + \lambda e_k, \qquad \varepsilon_i = e_i \quad \text{per } i \neq h.$$

La matrice $C = (c_{ij}) = (\varphi(\varepsilon_i, \varepsilon_j))$ che rappresenta ϕ nella nuova base soddisfa le uguaglianze

$$\begin{cases} c_{hh} = b_{hh} + \lambda b_{hk} + \lambda b_{kh} + \lambda^2 b_{kk} & \\ c_{hi} = b_{hi} + \lambda b_{ki}, \quad c_{ih} = b_{ih} + \lambda b_{ik} & \text{per } i \neq h, \\ c_{ij} = b_{ij} & \text{per } i, j \neq h. \end{cases}$$

Dunque la matrice C si ottiene dalla matrice B mediante *due operazioni elementari simmetriche*: nella prima si somma alla riga h la riga k moltiplicata per λ; nella seconda si somma alla colonna h la colonna k **della nuova matrice** moltiplicata per λ.

(3) se scambiamo e_h con e_k, la nuova matrice si ottiene prima scambiando prima le righe h e k di B e poi scambiando le colonne h e k della matrice così ottenuta.

Vediamo un esempio numerico in cui le precedenti tre operazioni elementari simmetriche sono eseguite sulla matrice

$$B = (\varphi(e_i, e_j)) = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

Se moltiplichiamo il secondo vettore della base per 2 otteniamo

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \xrightarrow{R_2 \mapsto 2R_2} \begin{pmatrix} 1 & 1 & 0 \\ 2 & 2 & 2 \\ 0 & 1 & 1 \end{pmatrix} \xrightarrow{C_2 \mapsto 2C_2} \begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 2 \\ 0 & 2 & 1 \end{pmatrix}.$$

Se aggiungiamo al primo vettore della base il triplo del secondo otteniamo

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \xrightarrow{R_1 \mapsto R_1 + 3R_2} \begin{pmatrix} 4 & 4 & 3 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \xrightarrow{C_1 \mapsto C_1 + 3C_2} \begin{pmatrix} 16 & 4 & 3 \\ 4 & 1 & 1 \\ 3 & 1 & 1 \end{pmatrix}.$$

Se scambiamo tra di loro il primo ed il secondo vettore della base otteniamo

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \xrightarrow{C_1 \leftrightarrow C_2} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Con il termine di **eliminazione di Gauss simmetrica** intenderemo la trasformazione in forma diagonale di una matrice simmetrica mediante una successione finita di operazioni elementari sulle rigonne (righe-colonne) descritte sopra ai punti (1),(2), (3).

Esempio 15.6.1. Vediamo un esempio dove la seconda rigonna è moltiplicata per 2:

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 3 \end{pmatrix} \to \begin{pmatrix} 0 & 1 & 1 \\ 2 & 2 & 4 \\ 1 & 2 & 3 \end{pmatrix} \to \begin{pmatrix} 0 & 2 & 1 \\ 2 & 4 & 4 \\ 1 & 4 & 3 \end{pmatrix}.$$

Esempio 15.6.2. Cerchiamo una matrice diagonale congruente a

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

Possiamo applicare l'eliminazione di Gauss simmetrica con l'obiettivo di annullare tutti i coefficienti esterni alla diagonale principale. Come primo passo annulliamo il coefficiente $1,2$ (prima riga, seconda colonna) sottraendo alla seconda rigonna (riga-colonna) la prima rigonna:

$$B = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & -3 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 1 & 1 & -3 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & -3 \end{pmatrix}.$$

Come secondo passo sottraiamo alla terza rigonna la prima rigonna:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & -3 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -4 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -4 \end{pmatrix}.$$

Possiamo quindi dire che B ha rango 3 e, nel caso reale, segnatura -1.

Esempio 15.6.3. Eseguiamo l'eliminazione di Gauss simmetrica della matrice

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

Iniziamo scambiando tra loro prima e seconda rigonna:

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 3 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 3 \end{pmatrix}.$$

Poi togliamo alla seconda rigonna la prima ed alla terza il doppio della prima:

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

Per finire togliamo la seconda rigonna alla terza:

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

Dunque la matrice B ha rango 2 e, sui numeri reali, segnatura 0.

Esercizi.

771. Determinare rango e segnatura della forma quadratica $\Phi\colon \mathbb{R}^4 \to \mathbb{R}$

$$\Phi(x) = x_1^2 - x_2^2 - x_3^2 + 2x_2x_3 + 2x_4x_1.$$

772. Calcolare rango e segnatura della forma quadratica

$$\Phi\colon \mathbb{R}^4 \to \mathbb{R}, \qquad \Phi(x) = x_1^2 - x_3^2 - x_4^2 + 4x_1x_2 - 2x_1x_3 + 4x_2x_3 + 2x_2x_4 - 6x_3x_4.$$

773 (♡). Determinare rango e segnatura della forma quadratica $\Phi\colon \mathbb{R}^4 \to \mathbb{R}$

$$\Phi(x) = x_1x_2 + x_2^2 + 2x_2x_4 - x_3^2.$$

774. Determinare rango e segnatura della forma quadratica $\Phi\colon \mathbb{R}^4 \to \mathbb{R}$

$$\Phi(x) = x_1x_2 + x_2x_3 + x_3x_4.$$

775. Determinare tutti i vettori isotropi della forma quadratica $\Phi\colon \mathbb{R}^4 \to \mathbb{R}$ associata alla matrice

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

776. Determinare rango e segnatura della forma quadratica $\Phi\colon \mathbb{R}^4 \to \mathbb{R}$

$$\Phi(x) = x_1^2 - x_2^2 - x_3^2 + 2x_2x_3 + 2x_4x_1.$$

777. Calcolare rango e segnatura della forma quadratica

$$\Phi\colon \mathbb{R}^5 \to \mathbb{R}, \qquad \Phi(x) = x_1^2 + 2x_2^2 - x_3^2 - x_5^2 + 2x_1x_2 + 4x_1x_3 + x_2x_3 - 2x_5x_4.$$

15.7. Il teorema spettrale

Abbiamo già dimostrato che ogni matrice simmetrica reale è diagonalizzabile. In questa sezione diamo una diversa dimostrazione, ed al tempo stesso un rafforzamento, di questo fatto noto come teorema spettrale reale.

LEMMA 15.7.1. *Siano V spazio vettoriale di dimensione finita su $\mathbb{R}$ e $f\colon V \to V$ un endomorfismo lineare. Esistono allora due vettori $u, v \in V$ non entrambi nulli, e due numeri reali a, b tali che*

$$f(u) = au - bv, \qquad f(v) = bu + av.$$

DIMOSTRAZIONE. Non è restrittivo supporre $V = \mathbb{R}^n$ e di conseguenza $f = L_A$ l'applicazione lineare associata ad una matrice $A \in M_{n,n}(\mathbb{R})$. Possiamo pensare A come una matrice a coefficienti complessi e quindi estendere nel modo canonico L_A ad un'applicazione lineare $L_A\colon \mathbb{C}^n \to \mathbb{C}^n$. Se scriviamo ogni vettore di $\mathbb{C}^n$ nella forma $u + iv$, con $u, v \in \mathbb{R}^n$ ed i unità immaginaria, allora si ha

$$L_A(u + iv) = L_A(u) + iL_A(v).$$

Sia $\lambda = a+ib \in \mathbb{C}$ un autovalore complesso di A e sia $u+iv \in \mathbb{C}^n$ il corrispondente autovettore. Si ha

$$L_A(u) + iL_A(v) = (a+ib)(u+iv) = (au - bv) + i(av + bu)$$

che equivale a

$$L_A(u) = au - bv, \qquad L_A(v) = bu + av.$$

□

Riprendendo le notazioni introdotte nella Sezione 7.6, indichiamo con $x \cdot y = \sum x_i y_i$ il prodotto scalare canonico in $\mathbb{R}^n$, con $\|x\| = \sqrt{x \cdot x}$ la norma del vettore x e con $O_n(\mathbb{R})$ il corrispondente sottogruppo di matrici ortogonali:

$$O_n(\mathbb{R}) := O(\mathbb{R}^n, \cdot) = \{E \in M_{n,n}(\mathbb{R}) \mid E^T = E^{-1}\}.$$

Siccome $\det(E^T) = \det(E)$ e $\det(E^{-1}) = \det(E)^{-1}$ segue immediatamente che per ogni matrice ortogonale E vale $\det(E)^2 = 1$.

Per ogni matrice $A \in M_{n,n}(\mathbb{R})$ e per ogni coppia di vettori $x, y \in \mathbb{R}^n$ vale

$$Ax \cdot y = (Ax)^T y = x^T A^T y = x \cdot A^T y$$

ed in particolare:

(1) Se A è simmetrica, allora $Ax \cdot y = x \cdot Ay$ per ogni x, y.
(2) Se A è antisimmetrica, allora $Ax \cdot y = -x \cdot Ay$ per ogni x, y.
(3) Se E è ortogonale, allora $Ex \cdot Ey = x \cdot E^T Ey = x \cdot y$ per ogni x, y.

LEMMA 15.7.2. *Siano $A \in M_{n,n}(\mathbb{R})$ una matrice simmetrica e $V \subseteq \mathbb{R}^n$ un sottospazio di dimensione positiva tale che $AV \subseteq V$. Allora V contiene un autovettore non banale per A.*

DIMOSTRAZIONE. Abbiamo visto che esistono due vettori $u, v \in V$ non entrambi nulli e tali che

$$Au = au - bv, \qquad Av = bu + av,$$

per qualche coppia di numeri reali a, b. Siccome A è simmetrica vale $Au \cdot v = u \cdot Av$ e quindi

$$Au \cdot v = a(u \cdot v) - b(v \cdot v) = u \cdot Av = a(u \cdot v) + b(u \cdot u).$$

Semplificando si ottiene $b(\|u\|^2 + \|v\|^2) = 0$ e siccome $\|u\|^2 + \|v\|^2 > 0$ si ha $b = 0$. Di conseguenza $f(u) = au$, $f(v) = av$ e qualunque tra u o v non sia nullo risulta essere un autovettore di A. □

DEFINIZIONE 15.7.3. Una base $v_1, \ldots, v_n$ di $\mathbb{R}^n$ si dice **ortonormale** se $v_i \cdot v_j = \delta_{ij}$ per ogni i, j, o equivalentemente se la matrice $E = (v_1, \ldots, v_n)$ è ortogonale.

TEOREMA 15.7.4 (Teorema spettrale). *Sia $A \in M_{n,n}(\mathbb{R})$ una matrice simmetrica. Esiste allora una base di $\mathbb{R}^n$ ortonormale e di autovettori per A.*

DIMOSTRAZIONE. Sia $V \subseteq \mathbb{R}^n$ un sottospazio vettoriale tale che $AV \subseteq V$ e dimostriamo per induzione sulla dimensione di V che esiste una base ortonormale di V fatta da autovettori per A. Se $V = 0$ non c'è nulla da dimostrare. Sia $u_1 \in V$ un autovettore per A, diciamo $Au_1 = \lambda_1 u_1$; a meno di dividere u_1 per la sua norma $\|u_1\|$ non è restrittivo supporre $\|u_1\| = 1$. Si consideri adesso il sottospazio $W = \{v \in V \mid u_1 \cdot v = 0\}$. Se $w \in W$, allora $Aw \in W$, infatti $Aw \in V$ e vale

$$u_1 \cdot Aw = Au_1 \cdot w = \lambda_1 u_1 \cdot w = \lambda_1 u_1 \cdot w = 0.$$

Per l'ipotesi induttiva esiste una base ortonormale $u_2, \ldots, u_s$ di W di autovettori per A e quindi $u_1, u_2, \ldots, u_s$ è una base ortonormale di V di autovettori per A. □

COROLLARIO 15.7.5. *Sia $A \in M_{n,n}(\mathbb{R})$ una matrice simmetrica. Esiste allora una matrice ortogonale E tale che $E^{-1}AE = E^T AE$ è diagonale.*

DIMOSTRAZIONE. Sia $e_1, \ldots, e_n$ la base canonica di $\mathbb{R}^n$ e $u_1, \ldots, u_n$ una base ortonormale di autovettori per A, diciamo $Au_i = \lambda_i u_i$. Denotiamo con E la matrice ortogonale del cambio di base, ossia $Ee_i = u_i$. Allora

$$E^{-1}AEe_i = E^{-1}Au_i = E^{-1}\lambda_i u_i = \lambda_i e_i$$

e quindi $E^{-1}AE$ è diagonale. □

COROLLARIO 15.7.6. *Una matrice simmetrica $A \in M_{n,n}(\mathbb{R})$ è definita (risp. semidefinita) positiva se e solo se i suoi autovalori sono tutti positivi (risp.: non negativi).*

DIMOSTRAZIONE. Se A è diagonale il risultato è chiaro. Altrimenti prendiamo una matrice ortogonale E tale che $E^{-1}AE = D$ sia diagonale. Allora A ha gli stessi autovalori di D e per ogni $x \in \mathbb{R}^n$ vale

$$x \cdot Dx = x \cdot E^{-1}AEx = Ex \cdot AEx.$$

Siccome E è invertibile, ne segue che $Ex \cdot AEx > 0$ per ogni $x \neq 0$ se e solo se $y \cdot Ay > 0$ per ogni $y \neq 0$. □

COROLLARIO 15.7.7. *Una matrice simmetrica $A \in M_{n,n}(\mathbb{R})$ è definita positiva se e solo se è semidefinita positiva e invertibile.*

DIMOSTRAZIONE. Gli autovalori sono tutti positivi se e solo se sono tutti non negativi ed il determinante è diverso da 0. □

Già sappiamo che esistono matrici complesse non nulle che sono simmetriche e nilpotenti, e quindi non diagonalizzabili. Però, se anziché considerare le matrici simmetriche si considerano quelle Hermitiane, allora esiste la versione complessa del teorema spettrale. Prima di poterlo enunciare occorre introdurre il concetto di base unitaria.

DEFINIZIONE 15.7.8. Il **prodotto Hermitiano canonico** su $\mathbb{C}^n$ è definito come

$$\langle v, w\rangle = v^T\overline{w} = \sum_i v_i\overline{w_i} \in \mathbb{C}, \qquad v, w \in \mathbb{C}^n.$$

Si noti che per ogni $v, w \in \mathbb{C}^n$ si ha $\langle v, w\rangle = \overline{\langle w, v\rangle}$; in particolare per ogni $v \in \mathbb{C}^n$ il prodotto $\langle v, v\rangle = \sum_i |v_i|^2$ è un numero reale non negativo ed ha quindi senso definire la **norma** di v come

$$\|v\| = \sqrt{\langle v, v\rangle} = \sqrt{\sum_i |v_i|^2}\,.$$

Chiaramente $v = 0$ se e solo se $\|v\| = 0$. Il prodotto Hermitiamo canonico non è una forma bilineare perché non è $\mathbb{C}$-lineare nella seconda variabile; valgono infatti le relazioni

(1) $\langle av_1 + bv_2, w\rangle = a\langle v_1, w\rangle + b\langle v_2, w\rangle$,
(2) $\langle v, aw_1 + bw_2\rangle = \overline{a}\langle v, w_1\rangle + \overline{b}\langle v, w_2\rangle$.

Il prodotto Hermitiano può essere a buon diritto considerato non degenere in ogni sottospazio, in virtù del seguente risultato.

LEMMA 15.7.9. *Per ogni sottospazio vettoriale $V \subset \mathbb{C}^n$ sia*

$$V^\perp = \{w \in \mathbb{C}^n \mid \langle w, v\rangle = 0 \quad \forall\, v \in V\}.$$

Allora $V^\perp$ è un sottospazio vettoriale e $V \oplus V^\perp = \mathbb{C}^n$.

DIMOSTRAZIONE. Siccome $\langle v, v\rangle = 0$ se e solo se $v = 0$ ne consegue che $V \cap V^\perp = 0$. La dimostrazione che $\dim V + \dim V^\perp \geq n$ è del tutto simile a quella del Lemma 15.3.13 ed è lasciata per esercizio. □

Prima di enunciare il teorema spettrale nel caso Hermitiano, ricordiamo che una matrice $H \in M_{n,n}(\mathbb{C})$ si dice Hermitiana se $H^T = \overline{H}$. Equivalentemente una matrice H è Hermitiana se e solo se per ogni $v, w \in \mathbb{C}^n$ vale la formula $\langle Hv, w\rangle = \langle v, Hw\rangle$. Infatti, se $H^T = \overline{H}$ si ha

$$\langle Hv, w\rangle = (Hv)^T\overline{w} = v^TH^T\overline{w} = v^T\overline{H}\overline{w} = v^T\overline{Hw} = \langle v, Hw\rangle.$$

Viceversa, se $\langle Hv, w\rangle = \langle v, Hw\rangle$ per ogni $v, w \in \mathbb{C}$ il conto appena fatto mostra che $v^T(H^T - \overline{H})\overline{w} = 0$ e questo basta per dedurre che $H^T - \overline{H} = 0$ (è sufficiente sviluppare il prodotto $v^T(H^T - \overline{H})\overline{w} = 0$ per v, w elementi della base canonica).

LEMMA 15.7.10. *Gli autovalori delle matrici Hermitiane sono numeri reali.*

DIMOSTRAZIONE. Supponiamo $Hv = \lambda v$, con $H \in M_{n,n}(\mathbb{C})$ Hermitiana, $v \in \mathbb{C}^n$ e $\lambda \in \mathbb{C}$. Allora

$$\lambda\|v\|^2 = \langle \lambda v, v\rangle = \langle Hv, v\rangle = \langle v, Hv\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\|v\|^2$$

e siccome $\|v\| \neq 0$ ne consegue che $\lambda = \overline{\lambda}$. □

DEFINIZIONE 15.7.11. Sia $V \subset \mathbb{C}^n$ un sottospazio vettoriale di dimensione m. Una base $u_1, \ldots, u_m \in V$ si dice **unitaria** se $\|u_i\| = 1$ per ogni i e $\langle u_i, u_j \rangle = 0$ per ogni $i \neq j$.

TEOREMA 15.7.12. *Sia $V \subset \mathbb{C}^n$ un sottospazio invariante per una matrice Hermitiana $H \in M_{n,n}(\mathbb{C})$. Allora V possiede una base unitaria di autovettori di H.*

DIMOSTRAZIONE. La dimostrazione è praticamente identica a quella del Teorema 15.7.4 e diamo solamente alcuni brevi cenni. Si ragiona per induzione sulla dimensione di V; se $\dim V = 0$ non c'è nulla da dimostrare.

Se $V \neq 0$, siccome H è triangolabile il sottospazio V contiene un autovettore per H: $v \in V$, $v \neq 0$, $Hv = \lambda v$. Il vettore $u_1 = v/\|v\|$ è unitario e si applica l'ipotesi induttiva al sottospazio $W = V \cap \mathrm{Span}(u_1)^\perp$, beninteso dopo aver usato la formula $\langle Hv, w \rangle = \langle v, Hw \rangle$ per dimostrare che anche W è invariante per H. □

Esercizi.

778. Sia $A = (a_{ij}) \in M_{n,n}(\mathbb{R})$ e si assuma che esistano $b_1, \ldots, b_n > 0$ tali che $b_i a_{ij} = b_j a_{ji}$ per ogni i, j. Dimostrare che A è diagonalizzabile.

779. Usare il criterio di Sylvester (Corollario 15.5.8) ed il teorema spettrale 15.7.4 per dimostrare quanto affermato nell'Osservazione 10.6.5.

780. Sia $A \in M_{n,n}(\mathbb{R})$ una matrice simmetrica semidefinita positiva. Provare che $\det(I + A) \geq 1$.

781 (☕). Siano $A, B \in M_{n,n}(\mathbb{R})$ matrici simmetriche semidefinite positive. Provare che $\det(I + AB) \neq 0$.

782. Sia $A = (a_{ij})$ una matrice simmetrica reale $n \times n$ e siano $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ i suoi autovalori, contati con molteplicità. Dimostrare che per ogni $i = 1, \ldots, n$ vale

$$a_{ii} \geq \min\{\lambda_1, \ldots, \lambda_n\}.$$

(Suggerimento: per quali valori di $t \in \mathbb{R}$ la matrice $A + tI$ è definita positiva?).

783. Sia $A \in M_{n,n}(\mathbb{R})$ antisimmetrica. Dimostrare che:
(1) A^2 è simmetrica e semidefinita negativa.
(2) $A - I$ è invertibile.
(3) $(I + A)(I - A)^{-1} = (I - A)^{-1}(I + A)$.
(4) La matrice $E = (I - A)^{-1}(I + A)$ è ortogonale.
(5) Il determinante di $E = (I - A)^{-1}(I + A)$ è uguale a 1.

784. Siano $u_1, \ldots, u_n$ e $v_1, \ldots, v_n$ due basi ortonormali di $\mathbb{R}^n$. Dimostrare che la matrice E di coefficienti $e_{ij} = u_i \cdot v_j$ è ortogonale.

785. Dimostrare che per una matrice $A \in M_{n,n}(\mathbb{R})$ simmetrica e ortogonale la segnatura è uguale alla traccia.

786. Sia $A \in M_{n,n}(\mathbb{R})$ una matrice simmetrica semidefinita positiva. Mostrare che se A^2 è un multiplo scalare dell'identità, allora anche A è un multiplo scalare dell'identità.

787 (♡). Per ogni $n > 0$ sia $B_{n,n} \in M_n(\mathbb{R})$ la matrice simmetrica di coefficienti

$$b_{ij} = i + j - 2, \qquad i, j = 1, \ldots, n.$$

Determinare rango e segnatura di B_1, B_2 e B_3.

788 (☕, ♡). Determinare rango e segnatura della matrice B_n introdotta nell'Esercizio 787 per ogni $n \geq 4$.

789. Sia $A \in M_{n,n}(\mathbb{R})$ simmetrica e definita positiva. Si dimostri che per ogni ε di valore assoluto sufficientemente piccolo $A + \varepsilon I$ è definita positiva.

790. Trovare una base unitaria del sottospazio $H \subset \mathbb{C}^3$ di equazione

$$x - iy + z = 0, \qquad i = \sqrt{-1}.$$

791. Trovare una base unitaria di autovettori per la matrice Hermitiana

$$\begin{pmatrix} 1 & 6i \\ -6i & 4 \end{pmatrix}.$$

792. Dimostrare che ogni matrice $L \in M_{n,n}(\mathbb{C})$ si decompone in modo unico nella forma $L = H + iK$ con H, K Hermitiane.

(☕) Si supponga adesso che nella decomposizione precedente tutti gli autovalori di H siano positivi. Si dimostri che $|\det L| \geq \det H$. Quando vale l'uguaglianza?

793. Provare che per ogni coppia di numeri complessi a, b di modulo 1 ed ogni numero reale t, la matrice

$$\begin{pmatrix} \cos(t)a & \sin(t)b \\ -\sin(t)\overline{b} & \cos(t)\overline{a} \end{pmatrix}$$

è unitaria con determinante uguale a 1. Provare inoltre che ogni matrice unitaria di ordine 2 e determinante 1 si scrive, in maniera non unica, in tale forma.

15.8. I gruppi ortogonali

In questa sezione studieremo alcune proprietà dei **gruppi ortogonali**

$$O_n(\mathbb{R}) = \{A \in M_{n,n}(\mathbb{R}) \mid A\,A^T = I\},$$

ossia dei gruppi delle applicazioni ortogonali di $\mathbb{R}^n$ rispetto al prodotto scalare canonico. Abbiamo visto nella dimostrazione del teorema di cancellazione di Witt che ogni applicazione ortogonale è composizione al più $2n$ riflessioni del tipo S_u, con $u \neq 0$, che ricordiamo essere definite dalla formula

$$S_u(x) = x - 2\frac{x \cdot u}{\|u\|^2}u.$$

Si noti che $S_u = S_{\lambda u}$ per ogni numero reale $\lambda \neq 0$; a meno di sostituire u con $u/\|u\|$ otteniamo che ogni riflessione è del tipo S_u con $\|u\| = 1$.

Uno degli obiettivi di questa sezione sarà quello di dimostrare che per ogni $A \in O_n(\mathbb{R})$, il rango di $A - I$ coincide con il minimo intero k tale che A può essere scritto come composizione di k riflessioni del tipo S_u.

Iniziamo trattando i casi particolari $O_1(\mathbb{R})$, $O_2(\mathbb{R})$ ed $O_3(\mathbb{R})$ per poi passare al caso generale. È immediato osservare che una omotetia è ortogonale se e solo se è uguale a $\pm\,\mathrm{Id}$; in particolare $O_1(\mathbb{R}) = \{\pm\,\mathrm{Id}\} = \{\mathrm{Id}, S_1\}$ è un gruppo formato da due elementi.

Sul piano $\mathbb{R}^2$ esistono molti esempi non banali di applicazioni ortogonali:

(1) **Rotazioni**: per ogni numero reale α, la rotazione attorno all'origine in $\mathbb{R}^2$ di angolo α è data dalla matrice ortogonale

$$R_\alpha = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}.$$

(2) **Riflessioni**: dato $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} \cos(\beta) \\ \sin(\beta) \end{pmatrix}$ = vettore unitario di $\mathbb{R}^2$ (unitario significa $\|u\|^2 = u_1^2 + u_2^2 = 1$), la riflessione associata è data dalla matrice

$$S_u = \begin{pmatrix} u_2^2 - u_1^2 & -2u_1u_2 \\ -2u_1u_2 & u_1^2 - u_2^2 \end{pmatrix} = \begin{pmatrix} -\cos(2\beta) & -\sin(2\beta) \\ -\sin(2\beta) & \cos(2\beta) \end{pmatrix}.$$

PROPOSIZIONE 15.8.1. *Ogni applicazione ortogonale di $\mathbb{R}^2$ è una rotazione oppure una riflessione: il primo caso accade se e solo se il determinante è $+1$, il secondo se e solo se il determinante è -1.*

DIMOSTRAZIONE. È immediato osservare che $\det(R_\alpha) = 1$ e $\det(S_u) = -1$. Viceversa, sia $A \in O_2(\mathbb{R})$ una matrice ortogonale, ossia $AA^T = I$, siccome $|A| = |A^T|$, per il teorema di Binet si ha $|A|^2 = 1$ e quindi $|A| = \pm 1$.

Se $|A| = 1$, poiché i vettori colonna di A sono una base ortonormale si può scrivere

$$\begin{pmatrix} \cos(\alpha) & \sin(\beta) \\ \sin(\alpha) & \cos(\beta) \end{pmatrix},$$

dove $\alpha, \beta \in \mathbb{R}$ devono soddisfare le condizioni

$$1 = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta) = \cos(\alpha+\beta), \quad 0 = \cos(\alpha)\sin(\beta) + \sin(\alpha)\cos(\beta) = \sin(\alpha+\beta),$$

da cui segue $\alpha + \beta = 2k\pi$, e quindi $\cos(\beta) = \cos(\alpha)$, $\sin(\beta) = -\sin(\alpha)$.

Se $|A| = -1$ la matrice possiede due autovalori distinti, uno negativo λ ed uno positivo μ. Dato che A è ortogonale ne segue che $\lambda^2 = \mu^2 = 1$, $\lambda = -1$, $\mu = +1$. È chiaro che se u è un autovettore unitario per λ, allora $A = S_u$. □

Come conseguenza della Proposizione 15.8.1 abbiamo che il determinante di una matrice $A \in O_2(\mathbb{R})$ è uguale a $|A| = (-1)^{\mathrm{rg}(A-I)}$, mentre dimostreremo nel prossimo Teorema 15.8.2 che la medesima formula vale per ogni intero positivo n ed ogni $A \in O_n(\mathbb{R})$.

Appare dunque naturale classificare le applicazioni ortogonali $A \in O_3(\mathbb{R})$ di $\mathbb{R}^3$ in base al rango della matrice $A - I$. Chiaramente $\mathrm{rg}(A - I) = 0$ se e solo se $A = I$ è l'identità.

(1) se $\mathrm{rg}(A - I) = 1$, allora esiste un piano $V \subseteq \mathbb{R}^3$ di punti fissi ed un vettore $0 \neq u \in V^\perp$. Dato che il prodotto scalare è definito positivo si ha $\|u\| \neq 0$ e quindi $V = \mathrm{Span}(u)^\perp$. Da ciò segue necessariamente che $A = S_u$.

(2) se $\mathrm{rg}(A - I) = 2$ esiste un vettore non nullo $u \in \mathbb{R}^3$, unico a meno di moltiplicazione per scalare, tale che $Au = u$, ed una decomposizione in somma diretta

$$\mathbb{R}^3 = \mathrm{Span}(u) \oplus H, \qquad H = \mathrm{Span}(u)^\perp.$$

Siccome $Au = u$, si ha $A(H) \subset H$ e A non possiede punti fissi in H. Dalla classificazione delle isometrie di $\mathbb{R}^2$ ne consegue che la restrizione di A ad H è una rotazione diversa dall'identità. In tal caso chiameremo A **rotazione** di asse $\mathrm{Span}(u)$. Può essere utile notare che in una base ortonormale e_1, e_2, e_3, con $e_1 \in \mathrm{Span}(u)$, $e_2, e_3 \in \mathrm{Span}(u)^\perp$, la rotazione A si rappresenta con una matrice del tipo

$$R_\alpha = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{pmatrix}.$$

(3) se $\mathrm{rg}(A - I) = 3$, siccome ogni polinomio di terzo grado possiede radici reali, esiste almeno un autovettore $u \in \mathbb{R}^3$, diciamo $Au = \lambda u$. Abbiamo già osservato che $\lambda^2 = 1$, mentre per le ipotesi sul rango di $A - I$ deve essere $\lambda \neq 1$. Non rimane quindi che $Au = -u$. Se denotiamo con $B = S_u A$ si ottiene $Bu = S_u(-u) = u$, e quindi $\mathrm{rg}(B - I) \leq 2$. Dato che ogni vettore di $\mathrm{Ker}(B - I) \cap \mathrm{Span}(u)^\perp$ è un punto fisso di A, per la formula di Grassmann deve necessariamente essere $\mathrm{rg}(B - I) \leq 2$. In conclusione, se $\mathrm{rg}(A - I) = 3$ allora A è la composizione di una rotazione di asse $\mathrm{Span}(u)$ e della riflessione S_u. In una base ortonormale e_1, e_2, e_3, con $e_1 \in \mathrm{Span}(u)$, $e_2, e_3 \in \mathrm{Span}(u)^\perp$, l'applicazione ortogonale A si rappresenta con una matrice del tipo

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{pmatrix}.$$

Passiamo adesso a studiare il caso generale e consideriamo lo spazio $\mathbb{R}^n$ dotato del prodotto scalare canonico.

Teorema 15.8.2. *Sia $E \in O_n(\mathbb{R})$, allora esiste una base $u_1, \ldots, u_k$ di $\mathrm{Ker}(E - I)^\perp$ (in particolare $k = \mathrm{rg}(E - I)$) tali che*

$$E = S_{u_1} \cdots S_{u_k}.$$

Viceversa se $E = S_{v_1} \cdots S_{v_h}$ con $v_i \neq 0$ per ogni i, allora $h = \mathrm{rg}(E - I) + 2m$ per qualche intero $m \geq 0$.

La dimostrazione è una conseguenza immediata dei seguenti tre lemmi.

Lemma 15.8.3. *Ogni applicazione ortogonale $E \colon \mathbb{R}^n \to \mathbb{R}^n$ può essere scritta, in modo non unico, come composizione di al più $\mathrm{rg}(E - I)$ riflessioni del tipo S_u con $u \in \mathrm{Ker}(E - I)^\perp$.*

Dimostrazione. Induzione su $k = \mathrm{rg}(E - I)$: se $k = 0$ allora $E = I$ ed il risultato è chiaro. Supponiamo quindi $k > 0$, allora il sottospazio $V = \mathrm{Ker}(E - I)^\perp$ ha dimensione k ed è un autospazio per E. Infatti se $v \in \mathrm{Ker}(E - I)^\perp$ allora per ogni $w \in \mathrm{Ker}(E - I)$ si ha $Ew = w$ e quindi

$$Ev \cdot w = Ev \cdot Ew = v \cdot w = 0$$

da cui segue $Ev \in \mathrm{Ker}(E-I)^\perp$. Siccome $k>0$ esiste $v \in \mathrm{Ker}(E-I)^\perp$ tale che $Ev \neq v$ e si consideri il vettore $u = Ev - v \neq 0$. Allora $u \in \mathrm{Ker}(E-I)^\perp$, $S_u(x) = x$ per ogni $x \in \mathrm{Ker}(E-I)$,

$$(Ev-v)\cdot(Ev-v) = Ev\cdot Ev - v\cdot Ev - Ev\cdot v + v\cdot v = 2(Ev\cdot Ev - v\cdot Ev),$$

$$S_u(Ev) = Ev - 2\frac{(Ev-v)\cdot Ev}{(Ev-v)\cdot(Ev-v)}(Ev-v) = Ev-(Ev-v) = v$$

da cui segue che $\mathrm{Ker}(E-I)+\mathrm{Span}(v) \subseteq \mathrm{Ker}(S_uE-I)$. Dunque $\mathrm{rg}(S_uE-I) < \mathrm{rg}(E-I)$, $\mathrm{Ker}(S_uE-I)^\perp \subset \mathrm{Ker}(E-I)^\perp$ e per l'ipotesi induttiva si può scrivere $S_uE = S_{u_1}\cdots S_{u_h}$ con $h < \mathrm{rg}(E-I)$, $u_i \in \mathrm{Ker}(S_uE-I)^\perp$ e quindi $E = S_uS_{u_1}\cdots S_{u_h}$ con $u, u_i \in \mathrm{Ker}(E-I)^\perp$. □

LEMMA 15.8.4. *Se $u_1,\dots,u_k \in \mathbb{R}^n$ sono linearmente indipendenti, allora*

$$\mathrm{Ker}(S_{u_1}S_{u_2}\cdots S_{u_k} - I) = \mathrm{Span}(u_1,\dots,u_k)^\perp.$$

DIMOSTRAZIONE. L'inclusione $\supseteq$ è chiara poiché se $v \in \mathrm{Span}(u_1,\dots,u_k)^\perp$ allora $S_{u_i}v = v$ per ogni i. Dimostriamo l'inclusione inversa per induzione su k; a meno di dividere ciascun u_i per la sua norma possiamo supporre $\|u_i\| = 1$ per ogni i e quindi

$$S_{u_i}x = x - 2(x\cdot u_i)u_i.$$

Per $k=0$ il risultato è vero in quanto entrambi gli spazi sono tutto $\mathbb{R}^n$. Supponiamo quindi $k>0$ e per assurdo che $\dim \mathrm{Ker}(S_{u_1}S_{u_2}\cdots S_{u_k}-I) > n-k$. Per la formula di Grassmann esistono $x,y \in \mathrm{Span}(u_1,\dots,u_k)$, $x,y\neq 0$, tali che $S_{u_1}S_{u_2}\cdots S_{u_k}(x) = x$ e $y \in \mathrm{Span}(u_2,\dots,u_k)^\perp$. Denotiamo $E = S_{u_2}\cdots S_{u_k}$, allora per l'ipotesi induttiva

$$\mathrm{Ker}(E-I) = \mathrm{Span}(u_2,\dots,u_k)^\perp.$$

Poiché $y \notin \mathrm{Span}(u_2,\dots,u_k)$ possiamo scrivere in maniera unica $x = ay+k$, $u_1 = by+h$, con $a,b \in \mathbb{R}$, e $h,k \in \mathrm{Span}(u_2,\dots,u_k)$ e $b\neq 0$. Si noti che $y\cdot h = y\cdot k = 0$ e quindi $x\cdot y = a\|y\|^2$, $u_1\cdot y = b\|y\|^2$, $Ey = y$, $Ex = ay + Ek$ e quindi

$$x = S_{u_1}Ex = ay + Ek - 2(u_1\cdot(ay+Ek))(by+h).$$

Dato che $b=0$ deve essere $u_1\cdot(ay+Ek) = 0$ da cui $x = ay+Ek$, che implica $Ek = k$, ossia $k \in \mathrm{Ker}(E-I) = \mathrm{Span}(u_2,\dots,u_k)^\perp$. Ma questo è possibile solo se $k=0$ da cui segue $x = ay =$ e $x = ay - 2(u_1\cdot(ay))(by+h)$. Siccome $b=0$ la seconda uguaglianza implica $u_1\cdot(ay) = ab\|y\|^2 = 0$ da cui $a=0$. □

LEMMA 15.8.5. *Siano $u_1,\dots,u_k \in \mathbb{R}^n$ non nulli e linearmente dipendenti. Se $k=2$ allora $S_{u_1}S_{u_2} = I$, mentre se $k\geq 3$ possiamo trovare $v_1,\dots,v_{k-2} \in \mathrm{Span}(u_1,\dots,u_k)$ non nulli e tali che*

$$S_{u_1}S_{u_2}\cdots S_{u_k} = S_{v_1}S_{v_2}\cdots S_{v_{k-2}}.$$

DIMOSTRAZIONE. Abbiamo già visto che se u_1,u_2 sono proporzionali, allora $S_{u_1} = S_{u_2}$ e quindi $S_{u_1}S_{u_2} = I$. Se $k=3$ allora $E = S_{u_1}S_{u_2}S_{u_3}$ è l'identità sul sottospazio $\mathrm{Span}(u_1,u_2,u_3)^\perp$ che ha dimensione $\geq n-2$. Per il Lemma 15.8.3 si ha che $S_{u_1}S_{u_2}S_{u_3}$ è composizione di $k\leq 2$ riflessioni del tipo S_u con $u \in \mathrm{Ker}(E-I)^\perp \subseteq \mathrm{Span}(u_1,u_2,u_3)$. Dato che il determinate di S_u è -1 per il teorema di Binet deve necessariamente essere $k=1$.

Dimostriamo adesso il caso $k>3$ per induzione su k. Se u_{k-1} e u_k sono linearmente dipendenti, allora $S_{u_{k-1}}S_{u_k} = I$; similmente se $u_1,\dots,u_{k-1}$ sono linearmente dipendenti, per induzione possiamo scrivere $S_{u_1}S_{u_2}\cdots S_{u_{k-1}}$ come un prodotto di $k-3$ riflessioni. In entrambi i casi $S_{u_1}S_{u_2}\cdots S_{u_k}$ risulta uguale al prodotto di $k-2$ riflessioni.

Non è quindi restrittivo supporre $\dim \mathrm{Span}(u_{k-1},u_k) = 2$ e $\dim \mathrm{Span}(u_1,\dots,u_{k-2}) = k-2$; per la formula di Grassmann esiste un vettore non nullo $v \in \mathrm{Span}(u_1,\dots,u_{k-2}) \cap \mathrm{Span}(u_{k-1},u_k)$.

Allora vale

$$S_{u_1}\cdots S_{u_{k-2}}S_{u_{k-1}}S_{u_k} = S_{u_1}\cdots S_{u_{k-2}}S_vS_vS_{u_{k-1}}S_{u_k}.$$

Per induzione $S_{u_1}\cdots S_{u_{k-2}}S_v$ è uguale al prodotto di $k-3$ riflessioni e $S_vS_{u_{k-1}}S_{u_k}$ è una riflessione.

□

Esercizi.

794. Sia $E \in O_n(\mathbb{R})$ con n dispari. Dimostrare che esiste $v \neq 0 \in V$ tale che $E^2 v = v$.

795. Per ogni vettore non nullo $u \in \mathbb{R}^n$ indichiamo con S_u la riflessione ortogonale rispetto all'iperpiano perpendicolare a u. Dimostrare che:

(1) Se u, v, w sono linearmente indipendenti, allora $S_v S_w$ non è l'identità e la composizione $S_u S_v S_w$ non è una riflessione rispetto ad un iperpiano;
(2) Sia $E\colon \mathbb{R}^n \to R^n$ ortogonale, allora vale $E S_u E^{-1} = S_{Eu}$;
(3) Siano $u, v, w, z \in \mathbb{R}^3$. Allora esistono $a, b \in \mathbb{R}^3$ tali che $S_u S_v S_w S_z = S_a S_b$.

796. Scrivere la matrice rispetto alla base canonica della proiezione ortogonale di $\mathbb{R}^3$ sul piano di equazione $2x - y + 3z = 0$.

797. Sia $U \subset \mathbb{R}^4$ il piano generato dai vettori $(1, 1, 0, 0)$ e $(1, 2, 3, 4)$. Trovare un piano $V \subset \mathbb{R}^4$ tale che

$$P_U P_V = P_V P_U = 0,$$

dove P_U e P_V sono le proiezioni ortogonali su U e V rispettivamente.

798. Sia $A \in M_{n,n}(\mathbb{R})$. Dimostrare che
a) A è una proiezione ortogonale $\iff A^2 = A = A^T \iff A^T$ è una proiezione
b) A è una riflessione $\iff A^2 = I = AA^T \iff A^T$ è una riflessione.

799. Determinare la matrice della proiezione ortogonale $P\colon \mathbb{R}^3 \to \mathbb{R}^3$ sul piano $x - 2y + z = 0$ rispetto alla base canonica. Scrivere $P((1, 1, 1)^T)$.

800. Scrivere le matrici rispetto alla base canonica delle proiezioni ortogonali di $\mathbb{R}^3$ sui piani di equazione

$$2x + y - 3z = 0, \qquad x + 2y - z = 0, \qquad x + y + z = 0.$$

801. Trovare una base v_1, v_2 per il sottospazio $U \subset \mathbb{R}^3$ definito dall'equazione $x - y + 3z = 0$ tale che $v_1^T v_2 = 0$ e $(0, 0, 1)v_1 = 1$.

802. Determinare l'immagine del vettore $v = (-2, 0, 1)^T$ nella riflessione ortogonale

$$f : \mathbb{R}^3 \to \mathbb{R}^3$$

rispetto alla retta generata dal vettore $u = (1, 2, 1)^T$.

803. Sia A una matrice simmetrica definita positiva. Si dimostri che vi è una matrice triangolare superiore B tale che $A = B^T B$

804. Sia $f = R_V \circ R_W\colon \mathbb{R}^3 \to \mathbb{R}^3$ la composizione delle riflessioni ortogonali rispetto ai piani $V = \{x + y = 0\}$ e $W = \{y + z = 0\}$. Trovare la matrice che rappresenta f nella base canonica.

805. Calcolare rango e segnatura della forma quadratica reale

$$x^2 + y^2 + z^2 + t^2 + 2xt - 2xy - 2yt - 2zt.$$

806. Calcolare rango e segnatura della forma quadratica

$$\Phi\colon \mathbb{R}^{11} \to \mathbb{R}, \qquad \Phi(x) = (x_1 + x_2 + x_3 + \cdots + x_{10})^2 - x_3^2.$$

807. Calcolare rango e segnatura della forma quadratica

$$\Phi\colon \mathbb{R}^4 \to \mathbb{R}, \qquad \Phi(x) = x_1^2 - x_3^2 - x_4^2 + 4x_1x_2 - 2x_1x_3 + 4x_2x_3 + 2x_2x_4 - 6x_3x_4.$$

808. Scrivere la matrice rispetto alla base canonica della proiezione ortogonale di $\mathbb{R}^3$ sul piano di equazione $2x + y - 3z = 0$.

809. Calcolare rango e segnatura della forma quadratica

$$\Phi\colon \mathbb{R}^4 \to \mathbb{R}, \qquad \Phi(x) = x_1^2 - x_3^2 - x_4^2 + 4x_1x_2 + 2x_1x_3 - 4x_2x_3 + 2x_2x_4 - 6x_3x_4.$$

810. Calcolare rango e segnatura della forma quadratica

$$\Phi\colon \mathbb{R}^3 \to \mathbb{R}, \qquad \Phi(x) = x_1^2 + x_2^2 - x_3^2 + 12x_1x_3 + 78x_2x_3.$$

811. Sia $\Phi\colon V \to \mathbb{R}$ una forma quadratica di rango massimo e indefinita (ossia tale che $\Phi(v) > 0$ e $\Phi(w) < 0$ per qualche $v, w \in V$). Dimostrare che in una opportuna base la forma Φ si rappresenta con una matrice simmetrica con coefficienti nulli sulla diagonale.

812. Calcolare rango e segnatura della forma quadratica

$$\Phi\colon \mathbb{R}^4 \to \mathbb{R}, \qquad \Phi(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + 4(x_1x_2 + x_1x_3 + x_2x_3 + x_2x_4 + x_3x_4).$$

813. Trovare una base ortonormale del piano in $\mathbb{R}^3$ di equazione $x_1 + x_2 - x_3 = 0$.

814. Trovare una base ortonormale dello spazio delle matrici reali 2×2 a traccia nulla rispetto al prodotto scalare

$$\langle A, B\rangle = \mathrm{Tr}(A^T B).$$

15.9. Complementi: Radici quadrate e scomposizione polare

Sia A una matrice quadrata reale, allora A^TA è simmetrica e semidefinita positiva, infatti per ogni $x \in \mathbb{R}^n$ vale

$$x \cdot A^TAx = x^TA^TAx = (Ax)^T(Ax) = \|Ax\|^2 \geq 0.$$

Se inoltre $\det(A) \neq 0$, allora A^TA è invertibile e definita positiva.

In particolare, per ogni matrice simmetrica P, la matrice $P^2 = P^TP$ è simmetrica e semidefinita positiva.

Teorema 15.9.1. *Sia $A \in M_{n,n}(\mathbb{R})$ simmetrica e semidefinita positiva. Allora esiste un'unica matrice P simmetrica e semidefinita positiva tale che $P^2 = A$.*

Dimostrazione. Per il teorema spettrale non è restrittivo supporre A diagonale. Gli elementi sulla diagonale di A sono non negativi e dunque esiste una unica matrice diagonale P semidefinita positiva tale che $P^2 = A$. Resta da dimostrare che se R è una matrice simmetrica semidefinita positiva e R^2 è diagonale, allora anche R è diagonale, ossia che ogni elemento della base canonica è un autovettore per R. Siano $\lambda_1 > \lambda_2 > \cdots > \lambda_s \geq 0$ gli autovalori di R; siccome R è diagonalizzabile si ha

$$\mathbb{R}^n = \bigoplus_{i=1}^{s} \{v \in \mathbb{R}^n \mid Rv = \lambda_i v\}.$$

D'altra parte per ogni $i = 1, \ldots, s$ si ha

$$\{v \in \mathbb{R}^n \mid Rv = \lambda_i v\} \subseteq \{v \in \mathbb{R}^n \mid R^2 v = \lambda_i^2 v\}$$

e siccome $\lambda_i^2 \neq \lambda_j^2$ per ogni $i \neq j$ si ha

$$\mathbb{R}^n = \bigoplus_{i=1}^{s} \{v \in \mathbb{R}^n \mid R^2 v = \lambda_i^2 v\}.$$

In particolare R ed R^2 hanno gli stessi autovettori. □

Definizione 15.9.2. Data una matrice simmetrica e semidefinita positiva A, la sua radice quadrata è l'unica matrice $\sqrt{A}$ simmetrica e semidefinita positiva tale che $(\sqrt{A})^2 = A$.

Osservazione 15.9.3. È istruttivo dare una dimostrazione leggermente più costruttiva dell'unicità della radice quadrata per matrici simmetriche definite positive A.

Vediamo prima il caso delle matrici 2×2 e poi passiamo al caso generale. Sia $B \in M_{2,2}(\mathbb{R})$ una radice quadrata di A e siano $\lambda_1, \lambda_2 > 0$ i suoi autovalori. Per Cayley-Hamilton si ha

$$0 = (B - \lambda_1 I)(B - \lambda_2 I) = B^2 - (\lambda_1 + \lambda_2)B + \lambda_1\lambda_2 I,$$

da cui segue

$$B = \frac{B^2 + \lambda_1\lambda_2 I}{\lambda_1 + \lambda_2}.$$

Se $A = B^2$, allora $\det(A) = (\lambda_1\lambda_2)^2$ e $\mathrm{Tr}(A) = \lambda_1^2 + \lambda_2^2 = (\lambda_1 + \lambda_2)^2 - 2\lambda_1\lambda_2$ otteniamo

$$B = \frac{A + \sqrt{\det(A)}I}{\sqrt{\mathrm{Tr}(A) + 2\sqrt{\det(A)}}}.$$

Sia adesso $B \in M_{n,n}(\mathbb{R})$ definita positiva e tale che $A = B^2$; osserviamo che gli autovalori di B e le rispettive molteplicità sono univocamente determinate da A e quindi pure il polinomio caratteristico

$$\det(B - tI) = r(t^2) - tq(t^2), \qquad r, q \in \mathbb{R}[t],$$

dipende da A. Per il teorema di Cartesio 10.6.2 gli $n+1$ coefficienti di $\det(B - tI)$ cambiano segno n volte e quindi $r, q \in \mathbb{R}[t]$ sono polinomi non nulli con coefficienti positivi. Per Cayley-Hamilton

$$Bq(B^2) = r(B^2), \qquad Bq(A) = r(A).$$

Adesso, le matrici $I, A, A^2, A^3, \dots$ sono simmetriche definite positive e $q(A)$ è una combinazione lineare di $I, A, A^2, A^3, \dots$ con coefficienti positivi. Dunque anche $q(A)$ è simmetrica definita positiva, quindi invertibile e di conseguenza

$$B = r(A)q(A)^{-1}.$$

COROLLARIO 15.9.4 (Decomposizione polare). *Sia $A \in M_{n,n}(\mathbb{R})$ invertibile. Allora esistono, e sono uniche, una matrice ortogonale E ed una matrice simmetrica definita positiva P tali che $EP = A$.*

DIMOSTRAZIONE. Vediamo prima l'unicità. Se $A = EP = FR$, con E, F ortogonali e P, R simmetriche definite positive, allora

$$A^T A = (EP)^T(EP) = P^T E^T EP = P^T P = P^2.$$

Similmente $A^T A = R^2$ e per l'unicità della radice quadrata $P = R$; di conseguenza $F = AR^{-1} = AP^{-1} = E$.

Mostriamo adesso l'esistenza: siccome A è invertibile, la matrice $A^T A$ è simmetrica e definita positiva. Esiste dunque P come nell'enunciato tale che $P^2 = P^T P = A^T A$. Resta da dimostrare che $E = AP^{-1}$ è ortogonale.

$$E^T E = (P^{-1})^T A^T AP^{-1} = (P^{-1})^T P^T PP^{-1} = (PP^{-1})^T PP^{-1} = I.$$

□

Esercizi.

815. Verificare direttamente che le seguenti 6 matrici simmetriche

$$\begin{pmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{pmatrix}, \qquad \pm \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

hanno lo stesso quadrato ma solo una è semidefinita positiva.

816. Calcolare, se esiste, la radice quadrata di ciascuna delle seguenti matrici simmetriche reali:

$$\begin{pmatrix} 2 & 3 \\ 3 & 12 \end{pmatrix}, \qquad \begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix}$$

817. Siano $L_1, L_2 \in GL(n, \mathbb{R})$ e $L_i = S_i P_i$, $i = 1, 2$, le rispettive decomposizioni polari. Dimostrare che P_1 e P_2 sono coniugate se e solo se esistono E_1, E_2 ortogonali tali che $L_1 = E_1 L_2 E_2$.

CAPITOLO 16

Spazi e trasformazioni affini

Strettamente legata all'algebra lineare è la geometria affine, ossia quella parte della geometria che si occupa di punti, rette e piani nello spazio e delle relazioni di allineamento, incidenza e parallelismo tra di loro. I concetti chiave sono quelli di spazio affine e trasformazione affine, che possono essere dati a vari livelli di astrazione e generalità: in ogni caso si ha che gli spazi vettoriali sono anche spazi affini, ed ogni spazio affine è isomorfo (in maniera non canonica ed in un senso ancora da definire) ad uno spazio vettoriale. È pertanto utile dal punto di vista didattico ed espositivo trattare preliminarmente la geometria affine negli spazi vettoriali, per passare successivamente alle nozioni più astratte di spazio affine.

16.1. Combinazioni baricentriche, spazi e sottospazi affini

In tutta la sezione indicheremo con V uno spazio vettoriale su di un campo $\mathbb{K}$.

DEFINIZIONE 16.1.1. Una combinazione lineare $a_0v_0 + \cdots + a_nv_n$ di vettori $v_i \in V$ si dice una **combinazione baricentrica** se $\sum a_i = 1$.

DEFINIZIONE 16.1.2. Un sottoinsieme di V si dice un **sottospazio affine** se è chiuso per combinazioni baricentriche. Più precisamente, un sottoinsieme $H \subseteq V$ è un sottospazio affine se per ogni insieme finito $v_1, \ldots, v_n \in H$ e per ogni $a_1, \ldots, a_n \in \mathbb{K}$ tali che $\sum a_i = 1$ vale $a_1v_1 + \cdots + a_nv_n \in H$.

ESEMPIO 16.1.3.
(1) Il sottoinsieme vuoto è un sottospazio affine.
(2) Ogni sottoinsieme formato da un unico vettore è un sottospazio affine.
(3) Ogni sottospazio vettoriale è anche un sottospazio affine: il viceversa è generalmente falso.
(4) L'intersezione di una famiglia arbitraria di sottospazi affini è un sottospazio affine.
(5) Il sottoinsieme di $\mathbb{K}^n$ formato dalle soluzioni $(x_1, \ldots, x_n)^T$ di un sistema lineare

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

è un sottospazio affine.
(6) Se $H \subseteq V$ e $K \subseteq W$ sono sottospazi affini, il loro prodotto cartesiano $H \times K$ è un sottospazio affine di $V \times W$.

ESEMPIO 16.1.4. Il simplesso standard di dimensione n sul campo $\mathbb{K}$ è per definizione

$$\Delta^n = \{(a_0, \ldots, a_n) \in \mathbb{K}^{n+1} \mid a_0 + \cdots + a_n = 1\}.$$

Si tratta del più piccolo sottospazio affine di $\mathbb{K}^{n+1}$ contenente la base canonica.

DEFINIZIONE 16.1.5. Siano V, W spazi vettoriali e $H \subseteq V$, $K \subseteq W$ due sottospazi affini. Un'applicazione $f\colon H \to K$ si dice **affine** se commuta con le combinazioni baricentriche, cioè se per ogni $v_0, \ldots, v_n \in H$ e per ogni $a_0, \ldots, a_n \in \mathbb{K}$ tali che $\sum a_i = 1$ vale $f(\sum a_iv_i) = \sum a_if(v_i)$. Un **isomorfismo affine** è un'applicazione affine bigettiva.

La composizione di applicazioni affini è ancora affine. L'inverso di un isomorfismo affine è ancora un isomorfismo affine.

ESEMPIO 16.1.6. Ogni applicazione lineare è anche affine.

ESEMPIO 16.1.7. Sia V uno spazio vettoriale; per ogni vettore $h \in V$ definiamo l'applicazione **traslazione per** h come

$$T_h \colon V \to V, \qquad T_h(v) = h + v.$$

Le traslazioni sono applicazioni affini. Infatti se $v_0, \dots, v_n \in V$ e $a_0, \dots, a_n \in \mathbb{K}$ sono tali che $\sum a_i = 1$, allora

$$\sum a_i T_h(v_i) = \sum a_i(h + v_i) = \sum a_i h + \sum a_i v_i = h + \sum a_i v_i = T_h(\sum a_i v_i).$$

Si noti che le traslazioni sono applicazioni bigettive e che l'inversa di T_h è T_{-h}.

LEMMA 16.1.8. *Siano H un sottospazio affine non vuoto di uno spazio vettoriale V e $p \in H$ un suo elemento. Allora l'insieme*

$$U = \{v \in V \mid p + v \in H\} = \{q - p \mid q \in H\} = T_{-p}(H)$$

è un sottospazio vettoriale ed esiste un isomorfismo affine $f \colon U \to H$ tale che $f(0) = p$.

DIMOSTRAZIONE. Mostriamo che il sottoinsieme

$$U = \{v \in V \mid p + v \in H\}$$

è un sottospazio vettoriale. Esso contiene lo 0 poiché $p = p + 0 \in H$ per ipotesi. Se $a, b \in \mathbb{K}$ e $v, w \in U$ si ha

$$p + av + bw = (1 - a - b)p + a(p + v) + b(p + w) \in H$$

e quindi $av + bw \in U$. L'applicazione

$$f \colon U \to H, \qquad f(v) = p + v,$$

è chiaramente bigettiva e per ogni scelta di $v_1, \dots, v_n \in U$ e $a_1, \dots, a_n \in \mathbb{K}$ tali che $\sum a_i = 1$ si ha

$$f(\sum a_i v_i) = p + \sum a_i v_i = \sum a_i(p + v_i).$$

Osserviamo che f è la restrizione ad U della traslazione $T_p \colon V \to V$. □

PROPOSIZIONE 16.1.9. *Sia $f \colon V \to W$ un'applicazione affine tra due spazi vettoriali. Allora f è lineare se e solo se $f(0) = 0$.*

DIMOSTRAZIONE. Una implicazione è chiara: se f è lineare allora $f(0) = 0$. Viceversa, supponiamo che f sia affine e $f(0)$ e proviamo che f è lineare. Siano $v_1, v_2 \in V$ e $a_1, a_2 \in \mathbb{K}$; se indichiamo con $a_0 = 1 - a_1 - a_2$, allora la combinazione lineare $a_1 v_1 + a_2 v_2$ è uguale alla combinazione baricentrica $a_0 0 + a_1 v_1 + a_2 v_2$ e quindi

$$f(a_1 v_1 + a_2 v_2) = f(a_0 0 + a_1 v_1 + a_2 v_2) = a_0 f(0) + a_1 f(v_1) + a_2 f(v_2) = a_1 f(v_1) + a_2 f(v_2).$$

□

COROLLARIO 16.1.10. *Sia $f \colon V \to W$ un'applicazione affine tra due spazi vettoriali. Allora l'applicazione $T_{-f(0)} \circ f$ è lineare. In particolare ogni applicazione affine tra spazi vettoriali è la composizione di un'applicazione lineare e di una traslazione.*

DIMOSTRAZIONE. Siccome la composizione di applicazioni affini è ancora affine e vale $T_{-f(0)} \circ f(0) = 0$, segue dalla proposizione precedente che $T_{-f(0)} \circ f$ è lineare. □

Dunque, ogni applicazione affine $f \colon \mathbb{K}^n \to \mathbb{K}^m$ è univocamente determinata dalla traslazione per $f(0)$ in $\mathbb{K}^m$ e dall'applicazione lineare $T_{-f(0)} \circ f$. Se, nelle coordinate canoniche, $f(0) = (b_1, \dots, b_m)$ e $T_{-f(0)} \circ f$ è rappresentata dalla matrice (a_{ij}), allora si ha

$$f \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n + b_1 \\ \vdots \quad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n + b_m \end{pmatrix}$$

COROLLARIO 16.1.11. *Siano V, W due spazi vettoriali. Se esiste un isomorfismo affine tra di loro, allora esiste anche un isomorfismo lineare e quindi V e W hanno la stessa dimensione.*

DIMOSTRAZIONE. Se $f \colon V \to W$ un isomorfismo affine tra due spazi vettoriali, allora l'applicazione $g = T_{-f(0)} \circ f$ è lineare e siccome le traslazioni sono invertibili, anche g risulta invertibile. □

La geometria affine studia le proprietà delle figure geometriche che sono invarianti per isomorfismi affini. Per **spazio affine** su di un campo $\mathbb{K}$ intenderemo un sottospazio affine di uno spazio vettoriale su $\mathbb{K}$; per distinguere anche a livello lessicale gli ambiti vettoriale ed affine, chiameremo **punti** gli elementi di uno spazio affine. Indicheremo inoltre con $\mathbb{A}^n_{\mathbb{K}}$, o più semplicemente con $\mathbb{A}^n$ se non vi sono ambiguità sul campo, lo spazio affine corrispondente allo spazio vettoriale $\mathbb{K}^n$: ogni punto di $\mathbb{A}^n$ è quindi una n-upla di elementi del campo. Tale variazione linguistica e notazionale è fatta allo scopo di indicare il cambio di punto di vista, più legato alle proprietà geometriche e meno a quelle algebriche.

Più avanti, nella sezione complementi, daremo due ulteriori definizioni di spazio affine che si trovano in letteratura, più precisamente introdurremo i concetti di spazio affine astratto e spazio affine modellato. Mostreremo tuttavia che gli spazi affini astratti e modellati sono isomorfi a sottospazi affini di spazi vettoriali e quindi la precedente definizione di spazio affine è perfettamente compatibile con le altre due.

Il fatto da tenere presente è che in uno spazio affine sono comunque ben definite le combinazioni baricentriche. Esiste quindi una ovvia generalizzazione delle Definizioni 16.1.2 e 16.1.5: un sottoinsieme L di uno spazio affine H si dice un **sottospazio affine** se è chiuso per combinazioni baricentriche, ossia se per ogni insieme finito $p_1, \dots, p_n \in L$ e per ogni $a_1, \dots, a_n \in \mathbb{K}$ tali che $\sum a_i = 1$ vale $a_1 p_1 + \cdots + a_n p_n \in L$. Un'applicazione $f\colon H \to K$ tra spazi affini si dice **affine** se commuta con le combinazioni baricentriche.

Definizione 16.1.12. Una **affinità** di un spazio affine H è un'applicazione affine e bigettiva di H in sé.

Definizione 16.1.13 (Traslazioni). Dati due punti p, q in uno spazio affine H definiamo l'applicazione

$$T_{\overrightarrow{pq}}\colon H \to H, \qquad T_{\overrightarrow{pq}}(r) = r + q - p.$$

Notiamo che se H è un sottospazio affine di uno spazio vettoriale V, allora $T_{\overrightarrow{pq}}$ coincide con la restrizione ad H della traslazione T_{q-p} e quindi l'applicazione $T_{\overrightarrow{pq}}$, che continueremo a chiamare traslazione, è un'affinità di H con inversa $T_{\overrightarrow{qp}}$. Occorre fare attenzione che due traslazioni $T_{\overrightarrow{ab}}$ e $T_{\overrightarrow{cd}}$ possono coincidere come affinità anche se $a \neq c$ e $b \neq d$.

Lemma 16.1.14. *Per una quaterna ordinata di punti a, b, c, d in uno spazio affine le seguenti condizioni sono equivalenti:*

$$T_{\overrightarrow{ab}}(c) = d, \qquad T_{\overrightarrow{ac}}(b) = d, \qquad T_{\overrightarrow{cd}}(a) = b, \qquad T_{\overrightarrow{bd}}(a) = c,$$

$$T_{\overrightarrow{ab}} = T_{\overrightarrow{cd}}, \qquad T_{\overrightarrow{ac}} = T_{\overrightarrow{bd}}.$$

Dimostrazione. Pensando lo spazio affine contenuto in uno spazio vettoriale, le sei condizioni sono tutte equivalenti a $a + d = b + c$. □

Chiameremo **parallelogramma** in uno spazio affine una quaterna ordinata di punti a, b, c, d che soddisfa le condizioni del Lemma 16.1.14. Una prima conseguenza di tale lemma è che per due punti distinti a, b in uno spazio affine H vi è un'unica traslazione $f\colon H \to H$ tale che $f(a) = b$; infatti $T_{\overrightarrow{cd}}(a) = b$ se e solo se $T_{\overrightarrow{cd}} = T_{\overrightarrow{ab}}$.

Proposizione 16.1.15. *Sia $f\colon H \to K$ un morfismo affine. Allora per ogni $a, b \in H$ vale*

$$f \circ T_{\overrightarrow{ab}} = T_{\overrightarrow{f(a)f(b)}} \circ f.$$

In particolare, ogni morfismo affine trasforma parallelogrammi in parallelogrammi.

Dimostrazione. Per ogni $p \in H$ vale

$$f(T_{\overrightarrow{ab}}(p)) = f(p - a + b) = f(p) - f(a) + f(b) = T_{\overrightarrow{f(a)f(b)}}(f(p)).$$

Se a, b, c, d è un parallelogramma in H, allora

$$T_{\overrightarrow{f(a)f(b)}}(f(c)) = f(T_{\overrightarrow{ab}}(c)) = f(d).$$

□

Esercizi.

818. Siano L un sottospazio affine di uno spazio affine H e a, b punti distinti di L. Mostrare che la traslazione $T_{\overrightarrow{ab}}\colon H \to H$ preserva L, ossia che $T_{\overrightarrow{ab}}(L) = L$.

819. Siano $p_0, \dots, p_m \in \mathbb{K}^n$ elementi fissati e siano $q_0, \dots, q_s$ combinazioni baricentriche di $p_0, \dots, p_m$. Mostrare che ogni combinazione baricentrica di $q_0, \dots, q_s$ è anche combinazione baricentrica di $p_0, \dots, p_m$.

820. Siano $v_1, \dots, v_n$, $n \geq 2$, vettori in uno spazio vettoriale su di un campo $\mathbb{K}$ di caratteristica 0. Dimostrare per ogni combinazione baricentrica x di $v_1, \dots, v_n$ esistono un indice $i = 1, \dots, n$ ed uno scalare $t \in \mathbb{K}$ tali che si può scrivere $x = tv_i + (1-t)y$ dove y è una combinazione baricentrica degli $n-1$ vettori v_j, $j \neq i$.

821. Siano $v_1, \dots, v_n$ vettori in uno spazio vettoriale sul campo $\mathbb{K}$. Siano $i = 1, \dots, n$ un indice fissato e $t \in \mathbb{K}$ uno scalare diverso da 0 e 1 (in particolare il campo $\mathbb{K}$ contiene almeno tre elementi). Dimostrare che ogni combinazione baricentrica di $v_1, \dots, v_n$ si può scrivere nella forma $tx+(1-t)y$ dove x è una combinazione baricentrica di $v_1, \dots, v_i$ e y è una combinazione baricentrica di $v_i, \dots, v_n$.

822. Sia E un sottoinsieme di uno spazio vettoriale su di un campo diverso da $\mathbb{F}_2 = \{0, 1\}$. Provare che E è un sottospazio affine se e solo se per ogni $u, v \in E$ e per ogni $a \in \mathbb{K}$ vale $au + (1-a)v \in E$.

823. Siano $p_0, p_1, p_2 \in \mathbb{R}^2$ non allineati. Dimostrare che ogni punto di $\mathbb{R}^2$ si può scrivere in modo unico come combinazione baricentrica di p_0, p_1, p_2.

824. Dato un qualsiasi sottoinsieme $S \subset \mathbb{K}^n$, sia $\langle S \rangle$ l'insieme di tutte le combinazioni baricentriche di elementi di S. Provare che $\langle S \rangle$ è un sottospazio affine. In particolare per ogni $p, q \in \mathbb{K}^n$, la retta

$$\overline{pq} = \{p + t(q-p) \mid t \in \mathbb{K}\} = \{(1-t)p + tq \mid t \in \mathbb{K}\}$$

è un sottospazio affine.

825. Sia $f\colon V \to W$ una applicazione affine. Dimostrare che:

(1) Se $E \subseteq V$ è un sottospazio affine, allora $f(E)$ è un sottospazio affine.
(2) Se $H \subseteq W$ è un sottospazio affine, allora $f^{-1}(H) = \{x \in V \mid f(x) \in H\}$ è un sottospazio affine.

826. Siano $f\colon \mathbb{K}^n \to \mathbb{K}^m$ una applicazione affine e $H \subseteq \mathbb{K}^n$ un sottospazio affine. Dimostrare che esiste un sottospazio vettoriale $K \subseteq \mathbb{K}^n$ tale che $f^{-1}(f(v)) \cap H = v + K$ per ogni $v \in H$.

827 (Teorema dei quadrilateri di Varignon). Siano a, b, c, d quattro punti distinti del piano affine reale $\mathbb{A}^2_{\mathbb{R}} = \mathbb{R}^2$. Provare che i 4 punti mediani dei segmenti $\overline{ab}, \overline{bc}, \overline{cd}$ e $\overline{da}$ sono i vertici di un parallelogramma.

828. Siano a, b, c, d quattro punti distinti del piano affine reale $\mathbb{A}^2_{\mathbb{R}} = \mathbb{R}^2$. Provare che esiste un'affinità f del piano affine in sé tale che $f(a) = b$, $f(b) = c$, $f(c) = d$ e $f(d) = a$ se e solo se a, b, c, d sono i vertici di un parallelogramma.

829. Sia $f\colon \mathbb{K}^n \to \mathbb{K}^m$ un'applicazione affine e siano

$$\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} = f(0), \qquad \begin{pmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{pmatrix} = f(e_i) - f(0),$$

dove $e_1, \dots, e_n$ indica la base canonica di $\mathbb{K}^n$. Provare che f manda il punto $(x_1, \dots, x_n)^T$ nel punto $(y_1, \dots, y_m)^T$ che soddisfa la relazione

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \\ 1 \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} & b_m \\ 0 & \dots & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \\ 1 \end{pmatrix}.$$

Caratterizzare inoltre le matrici $(n+1) \times (n+1)$ corrispondenti alle traslazioni in $\mathbb{K}^n$.

830 (♡). Si considerino i sei punti di $\mathbb{R}^2$:

$$p_1 = \begin{pmatrix}1\\2\end{pmatrix}, \quad p_2 = \begin{pmatrix}3\\1\end{pmatrix}, \quad p_3 = \begin{pmatrix}3\\3\end{pmatrix}, \quad q_1 = \begin{pmatrix}1\\8\end{pmatrix}, \quad q_2 = \begin{pmatrix}0\\7\end{pmatrix}, \quad q_3 = \begin{pmatrix}7\\3\end{pmatrix}.$$

Si determini l'affinità che trasforma p_i in q_i per $i = 1, 2, 3$.

831. Siano A e B spazi affini su di un campo $\mathbb{K}$ con almeno 3 elementi. Provare che un'applicazione $f\colon A \to B$ è affine se e solo se per ogni $p, q \in A$ e per ogni $a \in \mathbb{K}$ vale $f(ap + (1-a)q) = af(p) + (1-a)f(q)$.

832 (☕). Sia E un sottospazio fissato di uno spazio vettoriale V. Denotiamo con $\mathcal{A}$ l'insieme dei sottospazi vettoriali $H \subseteq V$ tali che $V = H \oplus E$. Fissato un elemento $K \in \mathcal{A}$ esiste una bigezione

$$\phi_K\colon \operatorname{Hom}(K, E) \to \mathcal{A}$$

che ad ogni applicazione lineare $f\colon K \to E$ associa il suo grafico, e cioè:

$$\phi_K(f) = \{x + f(x) \in V \mid x \in K\}.$$

Dimostrare che per ogni $K, H \in \mathcal{A}$ la composizione

$$\operatorname{Hom}(K, E) \xrightarrow{\phi_K} \mathcal{A} \xrightarrow{\phi_H^{-1}} \operatorname{Hom}(H, E)$$

è un isomorfismo affine di spazi vettoriali. (Suggerimento: mostrare che esiste un isomorfismo lineare $\alpha\colon H \to K$ tale che $x - \alpha(x) \in E$ per ogni $x \in H$.)

833 (☕). Sia $H \subseteq \mathbb{K}^n$ un sottospazio affine non contenente 0 e $f\colon H \to \mathbb{K}^m$ un'applicazione affine. Dimostrare che f è la restrizione ad H di un'applicazione lineare $g\colon \mathbb{K}^n \to \mathbb{K}^m$.

16.2. Il rapporto semplice

Dati due punti distinti p, q in uno spazio affine H, chiameremo **retta** (affine) passante per p e q l'insieme

$$pq = \{tp + (1-t)q \mid t \in \mathbb{K}\} \subset H.$$

Si noti che se $pq = qp$ (basta scambiare t con $1-t$) e se $p = q$ allora l'insieme pq coincide con il medesimo punto. Infine, se $r \in pq$ e $r \neq q$ allora $pq = rq$; infatti, se $r = \alpha p + (1-\alpha)q$ si ha $\alpha \neq 0$ e vale

$$sr + (1-s)q = \alpha t p + (1 - \alpha t)q \quad \text{per ogni} \quad t \in \mathbb{K}.$$

Si noti che pq è un sottospazio affine; infatti per ogni $a, b \in pq$ si ha

$$a = tp + (1-t)q, \qquad b = sp + (1-s)q,$$

per qualche $t, s \in \mathbb{K}$ e quindi, per ogni $\gamma \in \mathbb{K}$ vale

$$\gamma a + (1-\gamma)b = \delta p + (1-\delta)q, \qquad \delta = t\alpha + s(1-\alpha).$$

Diremo che tre punti p, q, r sono **allineati** se sono contenuti in una medesima retta; equivalentemente p, q, r sono allineati se $p = q$ oppure se $r \in pq$.

Lemma 16.2.1. *Sia H un sottospazio affine di uno spazio vettoriale V e siano $p, q, r \in H$. Allora p, q, r risultano allineati se e solo se i due vettori $q-p, r-p$ sono linearmente dipendenti in V. Se inoltre H non è un sottospazio vettoriale, ossia se $0 \notin H$, allora p, q, r risultano allineati in H se e solo se i tre vettori p, q, r sono linearmente dipendenti in V.*

Dimostrazione. Se $p = q$ il risultato è banale, possiamo quindi supporre $q - p \neq 0$. In tal caso $r \in pq$ se e solo se esiste t tale che $r = tp + (1-t)q$, ossia se e solo se $r - p = (1-t)(q-p)$.

È ovvio che se $q-p, r-p$ sono linearmente dipendenti, allora anche p, q, r sono linearmente dipendenti. Viceversa se $0 \notin H$ e $ap + bq + cr = 0$ per qualche $a, b, c \in \mathbb{K}$ non tutti nulli, allora $a + b + c = 0$: infatti, se fosse $a + b + c = d \neq 0$ si avrebbe

$$\frac{a}{d}p + \frac{b}{d}q + \frac{c}{d}r = 0, \qquad \frac{a}{d} + \frac{b}{d} + \frac{c}{d} = 1,$$

in contraddizione con l'ipotesi che H non è un sottospazio vettoriale. Dunque si ha $a = -b - c$ e

$$b(q-p) + c(r-p) = ap + bq + cr = 0.$$

□

LEMMA 16.2.2. *Siano dati tre punti a, b, c non allineati in uno spazio affine H e si considerino tre combinazioni baricentriche (Figura 16.1)*

$$c' = ta + (1-t)b, \quad a' = sb + (1-s)c, \quad b' = rc + (1-r)a.$$

Allora i tre punti a', b', c' sono allineati se e solo se

$$\begin{vmatrix} t & 1-t & 0 \\ 0 & s & 1-s \\ 1-r & 0 & r \end{vmatrix} = tsr - (t-1)(s-1)(r-1) = 0.$$

DIMOSTRAZIONE. Non è restrittivo supporre H sottospazio affine proprio di uno spazio vettoriale V; a meno di agire con una traslazioni in V possiamo supporre $0 \notin H$. In tale situazione i vettori a, b, c sono linearmente indipendenti in V ed i vettori a', b', c' sono allineati in H se e solo se sono linearmente dipendenti in V, ossia se e solo se la matrice di passaggio è singolare. □

DEFINIZIONE 16.2.3 (Rapporto semplice). In uno spazio affine, dati tre punti allineati a, b, c, con $b \neq c$, si definisce il **rapporto semplice** $[a, b; c] \in \mathbb{K}$ come l'unico scalare t tale che

$$a = tb + (1-t)c.$$

È chiaro che i rapporti semplici sono invarianti per affinità. Talvolta il rapporto semplice $[a, b; c]$ viene denotato $\dfrac{\overrightarrow{ac}}{\overrightarrow{bc}}$; tale scrittura è motivata dal fatto che, pensando i punti a, b, c contenuti in uno spazio vettoriale, si ha

$$a = tb + (1-t)c \iff c - a = t(c-b) \iff \overrightarrow{ac} = t\,\overrightarrow{bc}\,.$$

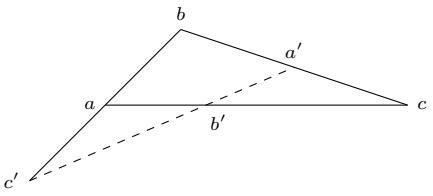


FIGURA 16.1

TEOREMA 16.2.4 (Menelao). *Siano a, b, c tre punti non allineati, ossia vertici di un triangolo non degenere, e siano $a' \in cb$, $b' \in ca$, $c' \in ab$ tre punti contenuti nei tre lati del triangolo e distinti dai vertici (Figura 16.1). Allora i tre punti a', b', c' sono allineati se e solo se*

$$[a, b; c'][b, c; a'][c, a; b'] = \frac{\overrightarrow{ac'}}{\overrightarrow{bc'}} \frac{\overrightarrow{ba'}}{\overrightarrow{ca'}} \frac{\overrightarrow{cb'}}{\overrightarrow{ab'}} = 1.$$

DIMOSTRAZIONE. Nelle stesse notazioni del Lemma 16.2.2 si ha

$$[a, b; c'] = \frac{t-1}{t}, \quad [b, c; a'] = \frac{s-1}{s}, \quad [c, a; b'] = \frac{r-1}{r}.$$

□

DEFINIZIONE 16.2.5. Due rette L, M in uno spazio affine H si dicono:

- **collineari** se hanno almeno un punto in comune;
- **parallele** se esiste una traslazione $f\colon H \to H$ tale che $f(L) = M$;
- **sghembe** se non sono né collineari né parallele.

Dunque ogni retta e al tempo stesso collineare e parallela a sé stessa. Se due rette L, M sono parallele, allora per ogni $a \in L$ e $b \in M$ si ha $T_{\overrightarrow{ab}}(L) = M$. Sia infatti f una traslazione tale che $f(L) = M$ e $c = f(a)$; allora $f = T_{\overrightarrow{ac}} = T_{\overrightarrow{bc}} \circ T_{\overrightarrow{ab}}$ e basta osservare che $T_{\overrightarrow{bc}}(M) = M$.

Le nozioni di collineazione e parallelismo si estendono ad un numero arbitrario di rette in uno spazio affine.

TEOREMA 16.2.6 (Ceva). *Siano a, b, c tre punti non allineati, ossia vertici di un triangolo non degenere, e siano $a' \in cb$, $b' \in ca$, $c' \in ab$ tre punti contenuti nei tre lati del triangolo e distinti dai vertici (Figura 16.2). Allora le tre rette aa', bb' e cc' sono collineari o parallele se e solo se*

$$[a, b; c'][b, c; a'][c, a; b'] = -1.$$

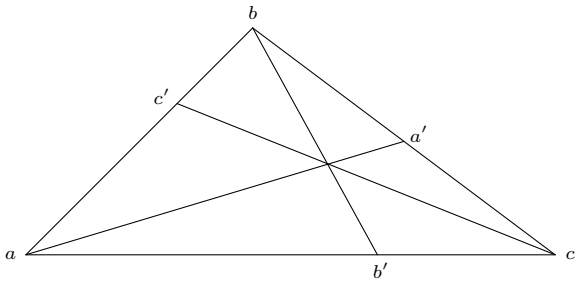


FIGURA 16.2. Il teorema di Ceva.

DIMOSTRAZIONE. Dimostriamo il teorema nel caso in cui le rette bb' e cc' si intersecano in un punto $p = ta + sb + rc$, $t + s + r = 1$. Il caso in cui tali rette sono parallele richiede una diversa dimostrazione ed è rimandato agli esercizi. È facile vedere che la retta ap interseca la retta bc se e solo se $s + r \neq 0$ o equivalentemente $t \neq 1$. Possiamo quindi supporre $t, s, r \neq 1$ e indichiamo con q il punto in cui la retta ap interseca la retta bc. Si ha

$$q = (1 - h)a + hp, \qquad \text{con} \qquad 1 - h + ht = 0,$$

e quindi

$$h = \frac{1}{1-t}, \qquad q = \frac{s}{1-t}b + \frac{r}{1-t}c, \qquad [b, c; q] = -\frac{r}{s}.$$

Si ha per simmetria

$$[a, b; c'] = -\frac{s}{t}, \qquad [c, a; b'] = -\frac{t}{r}$$

e dunque $q = a'$ se e solo se

$$[b, c; a'] = -\frac{r}{s} = -\frac{1}{[a, b; c'][c, a; b']}.$$

□

Esercizi.

834. Siano t, s, a numeri reali positivi. Si determini la funzione $f(t, s, a)$ per la quale i tre punti

$$\begin{pmatrix} t \\ at \end{pmatrix}, \qquad \begin{pmatrix} s \\ -as \end{pmatrix}, \qquad \begin{pmatrix} f(t,s,a) \\ 0 \end{pmatrix},$$

risultano allineati in $\mathbb{A}^2_{\mathbb{R}}$. In quale misura la funzione f dipende da a?

835. Dati tre punti distini ed allineati a, b, c mostrare che

$$[b, a; c] = \frac{1}{[a, b; c]}, \qquad [a, c; b] = 1 - [a, b; c].$$

836. Siano dati $a, b \in \mathbb{K}$; denotiamo con $L_1 \subset \mathbb{A}^2$ la retta passante per i punti di coordinate $(a, 0)$, $(1, b)$ e con $L_2 \subset \mathbb{A}^2$ la retta passante per i punti di coordinate $(a, 1)$, $(0, b)$. Provare che il punto di intersezione di L_1 ed L_2 appartiene alla diagonale $x = y$. (Sugg.: per semplificare i conti scrivere le equazioni delle due rette nella forma $\alpha x + \beta y = ab$ per opportuni $\alpha, \beta \in \mathbb{K}$.)

837. Sia $\mathbb{K}$ un campo con almeno tre elementi. Provare che un'applicazione bigettiva $f\colon \mathbb{K}^n \to \mathbb{K}^n$ è un'affinità se e soltanto se trasforma rette affini in rette affini (cioè se conserva gli allineamenti) e preserva i rapporti semplici.

838. Sia V uno spazio vettoriale su di un campo F. Provare che se F possiede almeno $n+1$ elementi, allora V non può essere unione di n sottospazi affini propri. In particolare uno spazio vettoriale su di un campo infinito non può essere unione finita di sottospazi affini propri. (Sugg.: induzione su n; sia per assurdo $V = \cup_{i=1}^{n} V_i$, allora a meno di traslazioni possiamo supporre $0 \in V_n$. Se $V_n \subset V_i$ per qualche $i < n$ abbiamo finito, altrimenti scegliamo $v \in V_n - \cup_{i=1}^{n-1}(V_n \cap V_i)$, $h \in V - V_n$ e consideriamo la retta affine $L = \{tv + (1-t)h \mid t \in F\}$. Esiste allora un indice i tale che L interseca V_i in almeno due punti.)

839 (☕). Trovare, se esiste, un'affinità $f\colon \mathbb{C}^2 \to \mathbb{C}^2$ tale che $f(L) \neq L$ per ogni retta affine $L \subset \mathbb{C}^2$.

16.3. Inviluppo affine, dimensione e formula di Grassmann

Se S è un sottoinsieme di uno spazio affine, indicheremo con $\langle S \rangle$ il suo **inviluppo affine**, e cioè l'intersezione di tutti i sottospazi affini che contengono S. Se H e K sono due sottospazi affini non vuoti, indicheremo talvolta con $\langle H, K \rangle$ l'inviluppo affine di $H \cup K$.

LEMMA 16.3.1. *Sia S un sottoinsieme non vuoto di uno spazio affine, allora $\langle S \rangle$ coincide con l'insieme di tutte le combinazioni baricentriche di elementi di S.*

DIMOSTRAZIONE. Indichiamo con H l'insieme di tutte le combinazioni baricentriche di elementi di S; siccome H è contenuto in ogni sottospazio affine contenente S, basta dimostrare che H è un sottospazio affine.

Siano $v_1, \dots, v_n \in H$ e $a_1, \dots, a_n \in \mathbb{K}$ tali che $\sum a_i = 1$. Ogni v_i è una combinazione baricentrica di elementi di S, diciamo

$$v_i = \sum_j b_{ij} u_{ij}, \qquad u_{ij} \in S, \ \sum_j b_{ij} = 1.$$

Ma allora

$$\sum_i a_i v_i = \sum_{i,j} a_i b_{ij} u_{ij}$$

e basta osservare che

$$\sum_{i,j} a_i b_{ij} = \sum_i a_i (\sum_j b_{ij}) = \sum_i a_i = 1.$$

□

Ricordiamo quanto visto nel Lemma 16.1.8 e nel Corollario 16.1.11: ogni spazio affine è isomorfo ad uno spazio vettoriale (pensato come spazio affine) e se due spazi vettoriali sono isomorfi come spazi affini allora sono isomorfi come spazi vettoriali. Ne segue che la seguente definizione è ben posta.

DEFINIZIONE 16.3.2. Sia K uno spazio affine non vuoto e sia $f\colon K \to W$ un isomorfismo affine con W spazio vettoriale. Si definisce la **dimensione di** K come la dimensione di W. La dimensione dello spazio affine vuoto è definita uguale a -1 per convenzione.

I punti sono tutti e soli i sottospazi affini di dimensione 0, mentre i sottospazi affini di dimensione 1 e 2 sono detti rispettivamente rette e piani. Un sottospazio affine di dimensione $n-1$ in uno spazio affine di dimensione n viene detto **iperpiano affine**.

DEFINIZIONE 16.3.3. Due sottospazi affini si dicono **paralleli** se uno è il traslato dell'altro.

Per definizione, sottospazi paralleli sono isomorfi e quindi hanno la stessa dimensione. Siccome le traslazioni formano un sottogruppo del gruppo delle affinità, la relazione di parallelismo è una relazione di equivalenza. La Proposizione 16.1.15 implica che ogni morfismo affine preserva la relazione di parallelismo.

PROPOSIZIONE 16.3.4. *Siano H, K due sottospazi paralleli di uno spazio affine E. Allora $H = K$ oppure $H \cap K = \emptyset$.*

DIMOSTRAZIONE. Siano $a, b \in E$ tali che $T_{\overrightarrow{ab}}(H) = K$ e supponiamo che esista $d \in H \cap K$. Allora esiste $c \in H$ tale che $T_{\overrightarrow{ab}}(c) = d$ e per il Lemma 16.1.14 vale $T_{\overrightarrow{ab}} = T_{\overrightarrow{cd}}$. Dunque per ogni $h \in H$ vale $T_{\overrightarrow{ab}}(h) = T_{\overrightarrow{cd}}(h) = h + d - c \in H$ e quindi $H = T_{\overrightarrow{ab}}(H) = K$. □

PROPOSIZIONE 16.3.5 (Formula di Grassmann). *Siano H, K due sottospazi di uno spazio affine E. Se $H \cap K \neq \emptyset$ allora*

$$\dim\langle H, K\rangle + \dim H \cap K = \dim H + \dim K.$$

Se $H \cap K = \emptyset$ allora $\dim\langle H, K\rangle \leq \dim H + \dim K + 1$ e l'uguaglianza vale se e solo se ogni retta in H è sghemba con ogni retta in K.

DIMOSTRAZIONE. Non è restrittivo supporre E spazio vettoriale. Se esiste $v \in H \cap K$ allora $T_{-v}H, T_{-v}K$ sono sottospazi vettoriali e si applica la formula di Grassmann dell'algebra lineare.

Supponiamo adesso che $K = \{v\}$ sia un punto esterno ad H e dimostriamo che $\dim\langle H, K\rangle = \dim H + 1$: a meno di traslazioni possiamo supporre che H sia un sottospazio vettoriale, quindi $0 \in \langle H, K\rangle$ e dunque $\langle H, K\rangle$ coincide con il sottospazio vettoriale generato da H e v.

Supponiamo adesso $H \cap K = \emptyset$; a meno di traslazioni possiamo supporre che H sia un sottospazio vettoriale. Scegliamo un vettore $v \in K$ e denotiamo $L = \langle H, v\rangle$. Allora vale $\langle H, K\rangle = \langle L, K\rangle$; quindi

$$\dim\langle H, K\rangle \leq \dim L + \dim K = \dim H + \dim K + 1$$

e l'uguaglianza vale se e solo se $\dim L \cap K = 0$. Per terminare la dimostrazione basta dimostrare che $T_{-v}(L \cap K) \subset H$ e quindi che $L \cap K$ è parallelo ad un sottospazio di H. Sia $w \in T_{-v}(L \cap K)$, allora esistono $k \in K$, $h \in H$ e $t \in \mathbb{K}$ tali che

$$w = k - v = (tv + h) - v.$$

Se $t = 1$ allora $w \in H$ come volevasi dimostrare; se invece $t \neq 1$ allora $k - tv \in H$ e quindi

$$\frac{1}{1-t}k - \frac{t}{1-t}v \in K \cap H.$$

□

Esercizi.

840. Nelle stesse notazioni del teorema di Ceva, se le rette bb' e cc' sono parallele, sia q il punto di intersezione della retta a loro parallela passante per a con la retta bc Mostrare che

$$[b, c; q] = [b', c; a] = [b, c'; a]$$

ed utilizzare l'Esercizio 835 per completare la dimostrazione del teorema di Ceva.

841. Sia $f\colon V \to W$ una applicazione affine tra spazi vettoriali e $K \subseteq V$ un sottospazio affine. Dimostrare che esiste un sottospazio vettoriale $W \subseteq V$ tale che $f^{-1}(f(v)) \cap K = v + W$ per ogni $v \in K$.

842. Siano H, K sottospazi affini disgiunti. Dimostrare che sono contenuti in iperpiani paralleli.

843. Siano a, b, c, d quattro punti in un piano affine su di un campo di caratteristica diversa da 2. Dimostrare che vale la seguente proprietà, nota come **assioma di Fano**: *almeno una coppia di diagonali opposte al quadrilatero di vertici a, b, c, d ha intersezione non vuota.*

Mostrare inoltre che l'assioma di Fano non vale in caratteristica 2.

844. Sia A uno spazio affine; $n+1$ punti $p_0, \ldots, p_n \in A$ si dicono linearmente indipendenti se l'applicazione

$$\Delta^n \to A, \qquad (t_0, \ldots, t_n) \mapsto t_0 p_0 + \cdots + t_n p_n$$

è iniettiva. In caso contrario si dicono linearmente dipendenti. Provare che la dimensione di uno spazio affine non vuoto A è uguale a

$$\dim A = \sup\{n \geq 0 \mid \text{ esistono } p_0, \ldots, p_n \in A \text{ linearmente indipendenti}\}.$$

845. Sia A uno spazio affine su $\mathbb{K}$ di dimensione finita n e siano $p_0, \ldots, p_n \in A$ dei punti linearmente indipendenti. Dimostrare che l'applicazione

$$\Delta^n \to A, \qquad (t_0, \ldots, t_n) \mapsto t_0 p_0 + \cdots + t_n p_n$$

è un isomorfismo affine.

846. Se A è uno spazio affine su $\mathbb{K}$, un'applicazione $f\colon A \to \mathbb{K}$ si dice **polinomiale** se per ogni scelta di $p_0, \dots, p_n \in A$ fissati, si ha che $f(t_0p_0 + \cdots + t_np_n)$ è un polinomio nelle variabili $t_0, \dots, t_n$. Provare che:

(1) Se $f\colon A \to B$ è un'applicazione affine e $g\colon B \to \mathbb{K}$ è polinomiale allora anche $gf\colon A \to \mathbb{K}$ è polinomiale.
(2) Le funzioni polinomiali sullo spazio affine $\mathbb{A}^n_{\mathbb{K}} \cong \mathbb{K}^n$ sono tutte e sole quelle rappresentate da polinomi nelle coordinate di $\mathbb{K}^n$.

16.4. Polinomi di Bernstein e curve di Bézier

In alcuni algoritmi usati in grafica computerizzata giocano un ruolo fondamentale i **polinomi di Bernstein** $B^n_i(t)$, $n \geq 0$, $i \in \mathbb{Z}$, definiti dalla formula:

$$B^n_i(t) = \binom{n}{i} t^i (1-t)^{n-i}\,, \qquad \text{per} \quad 0 \leq i \leq n, \tag{16.1}$$

e $B^n_i(t) = 0$ per $i < 0$ e $i > n$. Per semplicità espositiva consideriamo i polinomi di Bernstein a coefficienti reali, sebbene gran parte delle considerazioni che seguiranno sono valide su qualsiasi campo di caratteristica 0.

Sia $V_n \subset \mathbb{R}[t]$ il sottospazio vettoriale dei polinomi di grado $\leq n$; una base naturale di V_n è data dagli $n+1$ monomi $t^0, t^1, \dots, t^n$. Siccome

$$B^n_{n-i}(t) = \binom{n}{i} t^{n-i}(1-t)^i = \sum_{h=0}^{i} (-1)^h \binom{i}{h}\binom{n}{i} t^{n-i}t^h = \sum_{j=0}^{i} (-1)^{i-j}\binom{i}{j}\binom{n}{i} t^{n-j}$$

è immediato osservare che $B^n_n(t), \dots, B^n_0(t)$ sono linearmente indipendenti e quindi sono una base di V_n. La matrice di cambio di base è uguale a

$$(B^n_n(t), \dots, B^n_0(t)) = (t^n, \dots, t^0)(m_{ji}), \qquad m_{ji} = (-1)^{i-j}\binom{i}{j}\binom{n}{i}.$$

Dunque la matrice di cambio base (m_{ji}) è triangolare superiore con elementi sulla diagonale uguali a $m_{jj} = \binom{n}{j}$. Ad esempio per $n = 2$ e $n = 3$ si ha:

$$(B^2_2(t), B^2_1(t), B^2_0(t)) = (t^2, 2t - 2t^2, 1 - 2t + t^2) = (t^2, t, 1)\begin{pmatrix} 1 & -2 & 1 \\ 0 & 2 & -2 \\ 0 & 0 & 1 \end{pmatrix},$$

$$(B^3_3(t), \dots, B^3_0(t)) = (t^3, \dots, t^0)\begin{pmatrix} 1 & -3 & 3 & -1 \\ 0 & 3 & -6 & 3 \\ 0 & 0 & 3 & -3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Vediamo adesso alcune proprietà dei polinomi di Bernstein (le dimostrazioni sono lasciate per esercizio):

(1) (Simmetria) Vale $B^n_i(t) = B^n_{n-i}(1-t)$ per ogni n, i.
(2) (Relazioni ricorsive) Si ha $B^0_0 = 1$ e $B^0_i = 0$ per $i \neq 0$. Per ogni $n > 0$ ed ogni i si ha
$$B^n_i(t) = tB^{n-1}_{i-1}(t) + (1-t)B^{n-1}_i(t).$$
(3) (Partizione dell'unità, vedi Figura 16.3) Per ogni $n \geq 0$ vale
$$\sum_{i=0}^{n} B^n_i(t) = (t + (1-t))^n = 1.$$
(4) (Derivate) Le derivate dei polinomi di Bernstein soddisfano la formula:
$$B^n_i(t)' = n(\, B^{n-1}_{i-1}(t) - B^{n-1}_i(t)\,).$$

La proprietà $\sum_{i=0}^{n} B^n_i(t) = 1$ permette di usare i polinomi di Bernstein per definire curve parametriche nello spazio affine.

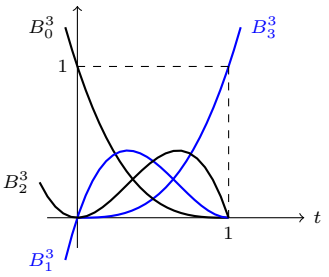


FIGURA 16.3. Grafici dei polinomi di Bernstein $B_i^3(t)$ per $0 \le t \le 1$ e $0 \le i \le 3$.

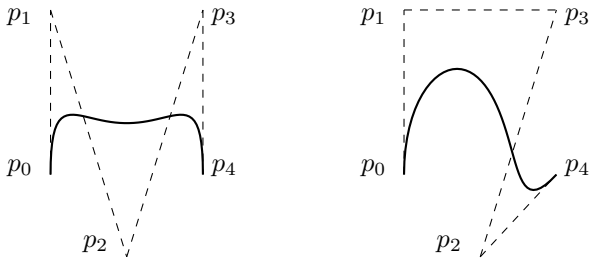


FIGURA 16.4. Due curve di Bézier del quarto grado; la prima controllata da p_0, p_1, p_2, p_3, p_4, la seconda da p_0, p_1, p_3, p_2, p_4.

DEFINIZIONE 16.4.1. Sia $p_0, p_1, \dots, p_n$ una successione di n punti in uno spazio affine H sul campo $\mathbb{R}$. Chiameremo **curva di Bézier** controllata da $p_0, p_1, \dots, p_n$ l'applicazione

$$\mathbf{b}\colon [0,1] \to H, \qquad \mathbf{b}(t) = \sum_{i=0}^{n} B_i^n(t) p_i.$$

Notiamo subito che gli estremi della curva di Bézier sono

$$\mathbf{b}(0) = p_0, \qquad \mathbf{b}(1) = p_n.$$

La costruzione delle curve di Bézier commuta con le applicazioni affini: nella situazione della Definizione 16.4.1, per ogni morfismo affine $f\colon H \to K$, la curva $f \circ \mathbf{b}$ è la curva di Bézier controllata da $f(p_0), f(p_1), \dots, f(p_n)$.

La proprietà di simmetria dei polinomi di Bernstein implica che se $\mathbf{b}(t)$ è la curva di Bézier controllata da $p_0, p_1, \dots, p_n$, allora $\mathbf{b}(1-t)$ è la curva di Bézier controllata da $p_n, p_{n-1}, \dots, p_0$. Appare invece chiaro che l'insieme $\{\mathbf{b}(t) \mid t \in [0,1]\}$ non è invariante per permutazioni dei punti p_i (vedi Figura 16.4).

Nella sostanza, la curva di Bézier controllata da una successione di punti $p_0, \dots, p_n$ è una approssimazione algebrica della poligonale di vertici $p_0, \dots, p_n$. Spesso, ma non sempre, la curva di Bézier assume un aspetto "morbido".

PROPOSIZIONE 16.4.2 (Algoritmo di de Casteljau). *Sia $p_0, p_1, \dots, p_n$ una successione di n punti in uno spazio affine H sul campo $\mathbb{R}$. Definiamo in maniera ricorsiva delle applicazioni*

$$\mathbf{b}_i^r\colon [0,1] \to H, \qquad 0 \le i, r; \quad i + r \le n;$$

ponendo $\mathbf{b}_i^0(t) = p_i$ per ogni t e

$$\mathbf{b}_i^r(t) = (1-t)\mathbf{b}_i^{r-1}(t) + t\mathbf{b}_{i+1}^{r-1}(t).$$

Allora $\mathbf{b}_0^n = \mathbf{b}$ è la curva di Bézier controllata da $p_0, p_1, \dots, p_n$.

DIMOSTRAZIONE. Segue dalla formula $B_i^r(t) = tB_{i-1}^{r-1}(t) + (1-t)B_i^{r-1}(t)$ e da una semplice induzione che per ogni coppia i, r tale che $i + r \le n$, si ha

$$\mathbf{b}_i^r(t) = \sum_{j=0}^{r} B_j^r(t) p_{i+j}.$$

□

Osservazione 16.4.3. Il primo utilizzo delle curve di Bézier è avvenuto nell'industria automobilistica attorno[1] al 1960. Infatti sia P. Bézier che P. de Casteljau lavoravano al reparto carrozzeria della Renault e della Citröen, rispettivamente.

Esercizi.

847. Dimostrare le seguenti proprietà dei polinomi di Bernstein:

$$\sum_{j=0}^{n} \frac{j}{n} B_j^n(t) = t, \qquad B_i^n(st) = \sum_{j=i}^{n} B_i^j(s) B_j^n(t),$$

$$t^i = \sum_{j=i}^{n} \frac{\binom{j}{i}}{\binom{n}{j}} B_j^n(t), \qquad B_i^n(t) B_j^m(t) = \frac{\binom{n}{i}\binom{m}{j}}{\binom{n+m}{i+j}} B_{i+j}^{n+m}(t),$$

$$\int_0^x B_i^n(t)dt = \frac{1}{n+1} \sum_{j=i+1}^{n+1} B_j^{n+1}(x),$$

$$tB_i^n(t) = \frac{i+1}{n+1} B_{i+1}^{n+1}(t), \quad (1-t)B_i^n(t) = \frac{n-i+1}{n+1} B_i^{n+1}(t),$$

$$B_i^n(t) = \frac{i+1}{n+1} B_{i+1}^{n+1}(t) + \frac{n-i+1}{n+1} B_i^{n+1}(t).$$

848. Calcolare il massimo assoluto delle funzioni di variabile reale $B_i^n \colon [0,1] \to \mathbb{R}$.

849. Sia $\mathbf{b}$ la curva di Bézier controllata da $p_0, p_1, \dots, p_n \in \mathbb{R}^k$ e provare che

$$\mathbf{b}(t) = p_0 + tn(p_1 - p_0) + t^2(\cdots)$$

Mostrare inoltre che se gli n vettori $p_i - p_0$, $i = 1, \dots, n$ sono linearmente indipendenti allora la derivata di $\mathbf{b}(t)$ è sempre diversa da 0.

850. Sia Sia $\mathbf{b} \colon [0,1] \to \mathbb{R}^2$ la curva di Bézier di terzo grado controllata dalla poligonale $(1,0), (-1,1), (1,1), (-1,0)$. Mostrare che tale curva possiede una cuspide semplice per $t = 1/2$, ossia che

$$\mathbf{b}\left(\frac{1}{2} + s\right) = \mathbf{b}\left(\frac{1}{2}\right) + As^2 + Bs^3,$$

con A, B vettori linearmente indipendenti.

16.5. Complementi: spazi affini astratti e modellati

In prima approssimazione, uno spazio affine astratto è un insieme sul quale ha senso parlare di combinazioni baricentriche.

Abbiamo già definito il simplesso n-dimensionale standard sul campo $\mathbb{K}$ come l'insieme

$$\Delta^n = \{(t_0, \dots, t_n) \in \mathbb{K}^{n+1} \mid \sum t_i = 1\}.$$

Osserviamo che Δ^n coincide con l'insieme di tutte le combinazioni baricentriche della base canonica $e_0 = (1, 0, \dots, 0)$, $e_1 = (0, 1, 0, \dots, 0)$ ecc. di $\mathbb{K}^{n+1}$.

Definizione 16.5.1. Un prespazio affine su di un campo $\mathbb{K}$ è il dato di un insieme A e di una successione di applicazioni

$$b_n \colon A^{n+1} \times \Delta^n \to A, \qquad \text{per} \quad n \geq 1.$$

È conveniente definire b_0 come l'identità su A e, per semplicità notazionale, scrivere

$$b_n(p_0, \dots, p_n, t_0, \dots, t_n) = t_0 p_0 + \cdots + t_n p_n.$$

Chiameremo le applicazioni b_n, $n \geq 0$, **combinazioni baricentriche**.

[1] Il segreto industriale che per anni ha coperto tali tecniche di progettazione non consente di dare una datazione precisa.

DEFINIZIONE 16.5.2. Un prespazio affine A su di un campo $\mathbb{K}$ si dice uno **spazio affine astratto** se:

(1) Per ogni $p_0, \dots, p_n \in A$ e per ogni $i = 0, \dots, n$ vale
$$0p_0 + \cdots + 1p_i + \cdots + 0p_n = p_i.$$
(2) Per ogni $p_0, \dots, p_n, q_0, \dots, q_m \in A$ l'insieme
$$\{(t_0, \dots, t_n, s_0, \dots, s_m) \in \Delta^n \times \Delta^m \mid t_0p_0 + \cdots + t_np_n = s_0q_0 + \cdots + s_mq_m\}$$
è un sottospazio affine di $\Delta^n \times \Delta^m \subset \mathbb{K}^{n+m+2}$.

ESEMPIO 16.5.3. Ogni spazio affine ha una struttura naturale di spazio affine astratto, dove b_n è la combinazione baricentrica usuale.

ESEMPIO 16.5.4. Sia H un iperpiano in uno spazio vettoriale V e sia A l'insieme delle rette (sottospazi lineari di dimensione 1) di V che non sono contenute in H. L'insieme A possiede una naturale struttura di spazio affine astratto, con le combinazioni baricentriche definite nel modo seguente: date $n+1$ rette $L_0, \dots, L_n \in A$ scegliamo $n+1$ vettori $v_0, \dots, v_n$ tali che $v_i \in L_i - \{0\}$ e $v_i - v_0 \in H$ per ogni $i = 0, \dots, n$. Definiamo quindi $t_0L_0 + \cdots + t_nL_n$ come la retta generata dal vettore $\sum t_iv_i$. Lasciamo per esercizio la semplice verifica che tali combinazioni baricentriche sono ben definite e definiscono una struttura di spazio affine su A.

LEMMA 16.5.5. *In ogni spazio affine astratto le combinazioni baricentriche sono simmetriche e commutano con le combinazioni baricentriche sui simplessi standard.*

Prima di passare alla dimostrazione precisiamo meglio il senso dell'enunciato. Sia A uno spazio affine astratto, dire che le combinazioni baricentriche sono simmetriche significa che per ogni $p_0, \dots, p_n \in A$, ogni $(t_0, \dots, t_n) \in \Delta^n$ ed ogni permutazione $\sigma \in \Sigma_{n+1}$ vale
$$t_0p_0 + \cdots + t_np_n = t_{\sigma(0)}p_{\sigma(0)} + \cdots + t_{\sigma(n)}p_{\sigma(n)}$$
e quindi acquista significato l'espressione $\sum_i t_ip_i$. Dire che le combinazioni baricentriche commutano con le combinazioni baricentriche sui simplessi standard significa che per ogni $p_0, \dots, p_n \in A$, ogni $(m+1)$-upla di vettori $t^0, \dots, t^m \in \Delta^n$ ed ogni $s \in \Delta^m$ vale
$$\sum_{j=0}^{m} s_j \left(\sum_{i=0}^{n} t_i^j p_i \right) = \sum_{i=0}^{n} \left(\sum_{j=0}^{m} s_j t_i^j \right) p_i.$$

DIMOSTRAZIONE. Dimostriamo prima che le combinazioni baricentriche sullo spazio affine astratto commutano con quelle sui simplessi standard. Denotiamo $q_j = t_0^j p_0 + \cdots + t_m^j p_m$; per la Definizione 16.5.2 abbiamo che
$$H = \{(v, w) \in \Delta^n \times \Delta^m \mid v_0p_0 + \cdots + v_np_n = w_0q_0 + \cdots + w_mq_m\}$$
è un sottospazio affine che contiene i vettori (t^j, e_j), per $j = 0, \dots, m$. Quindi per ogni $s = (s_0, \dots, s_m) \in \Delta^m$ si ha che $(\sum_j s_j t^j, s) \in H$ e perciò
$$(\sum_j s_j t_0^j)p_0 + \cdots + (\sum_j s_j t_n^j)p_n = s_0q_0 + \cdots + s_mq_m.$$
Per mostrare la simmetria, per ogni permutazione σ basta applicare il punto precedente ai vettori $t^j = e_{\sigma(j)}$. □

LEMMA 16.5.6. *In ogni spazio affine astratto le combinazioni baricentriche soddisfano le proprietà distributive, ossia vale*
$$0p_0 + t_1p_1 + \cdots + t_np_n = t_1p_1 + \cdots + t_np_n.$$
e, se $p_0 = p_1$, allora
$$t_0p_0 + t_1p_1 + \cdots + t_np_n = (t_0 + t_1)p_1 + \cdots + t_np_n.$$

DIMOSTRAZIONE. Per il Lemma 16.5.5 possiamo scrivere
$$t_1p_1 + \cdots + t_np_n = \sum_{j=1}^{n} t_j (\sum_{i=0}^{n} \delta_i^j p_i) = \sum_{i=0}^{n} (\sum_{j=1}^{n} t_j \delta_i^j) p_i = 0p_0 + t_1p_1 + \cdots + t_np_n.$$

Supponiamo adesso $p_0 = p_1 = p$; allora

$$H = \left\{ (v,w) \in \mathbb{K}^3 \times \mathbb{K}^2 \;\middle|\; \begin{array}{c} v_0 p + v_1 p + v_2 p + \sum_{i\geq 2} t_i p_i = w_0 p + w_1 p + \sum_{i\geq 2} t_i p_i, \\ v_0 + v_1 + v_2 = w_0 + w_1 = t_0 + t_1. \end{array} \right\}$$

è un sottospazio affine che contiene i vettori

$$a = ((t_0, 0, t_1), (t_0, t_1)), \quad b = ((0, t_1, t_0), (t_1, t_0)), \quad c = ((0, 0, t_0 + t_1), (0, t_0 + t_1))$$

e quindi contiene anche la combinazione baricentrica

$$a + b - c = ((t_0, t_1, 0), (t_0 + t_1, 0)).$$

□

Dunque le combinazioni baricentriche in uno spazio affine hanno tutte le buone proprietà che è lecito aspettarsi dalla notazione adottata.

PROPOSIZIONE 16.5.7. *Sia A uno spazio affine astratto su $\mathbb{K}$. Se $\mathbb{K}$ possiede almeno 3 elementi, allora le combinazioni baricentriche sono univocamente determinata dalle combinazioni baricentriche a due termini, ossia b_1 determina b_n per ogni n.*

DIMOSTRAZIONE. Siano $p_0, \dots, p_n \in A$, con $n > 1$ e $v = (v_0, \dots, v_n) \in \Delta^n$. Scegliamo un elemento $a \in \mathbb{K}$ diverso da 0 e da v_0 e consideriamo i vettori

$$u = (a, 1-a, 0, \dots, 0), \qquad w = \frac{a}{a - v_0}\left(v - \frac{v_0}{a} u\right) = (0, w_1, \dots, w_n).$$

Siccome $w_0 = 0$, per induzione su n i punti $q_0 = w_0 p_0 + \cdots + w_n p_n$ e $q_1 = a p_0 + (1-a) p_1$ sono determinati da b_1 e quindi lo è anche

$$v_0 p_0 + \cdots + v_n p_n = \left(1 - \frac{v_0}{a}\right) q_0 + \frac{v_0}{a} q_1.$$

□

Siamo adesso in grado di esplicitare le ovvie generalizzazioni di sottospazio, traslazione e applicazione affine.

DEFINIZIONE 16.5.8. Un sottoinsieme E di uno spazio affine astratto A si dice un sottospazio affine se è chiuso per combinazioni baricentriche.

Intersezione di sottospazi affini è ancora un sottospazio affine e per ogni sottoinsieme $S \subseteq A$ il sottospazio affine generato $\langle S \rangle$ è l'intersezione di tutti i sottospazi affini contenenti S. Equivalentemente $\langle S \rangle$ è uguale all'insieme di tutte le combinazioni baricentriche finite di elementi di S.

DEFINIZIONE 16.5.9. Un'applicazione affine è un'applicazione che commuta con le combinazioni baricentriche. Un isomorfismo affine è un'applicazione affine e bigettiva. Un isomorfismo affine di uno spazio affine in sé si dice una affinità.

ESEMPIO 16.5.10. Sia E un sottospazio di uno spazio vettoriale V. Vogliamo mostrare che l'insieme $\mathcal{A}$, dei sottospazi vettoriali $H \subset V$ tali che $V = H \oplus E$ possiede una struttura naturale di spazio affine astratto. Fissato un elemento $K \in \mathcal{A}$ esiste una bigezione

$$\phi_K \colon \operatorname{Hom}(K, E) \to \mathcal{A}$$

che ad ogni applicazione lineare $f \colon K \to E$ associa il suo grafico

$$\phi_K(f) = \{x + f(x) \in V \mid x \in K\}.$$

Per il risultato dell'Esercizio 832, per ogni $K, H \in \mathcal{A}$ la composizione

$$\operatorname{Hom}(K, E) \xrightarrow{\phi_K} \mathcal{A} \xrightarrow{\phi_H^{-1}} \operatorname{Hom}(H, E)$$

è un isomorfismo affine di spazi vettoriali e quindi vi è un'unica struttura di spazio affine astratto su $\mathcal{A}$ rispetto alla quale le applicazioni ϕ_K sono isomorfismi affini.

DEFINIZIONE 16.5.11. Chiameremo **traslazione** in uno spazio affine astratto A qualunque applicazione del tipo

$$T_{\overrightarrow{pq}} \colon A \to A, \qquad T_{\overrightarrow{pq}}(x) = x + q - p.$$

per qualche coppia di punti $p, q \in A$.

LEMMA 16.5.12. *Per ogni coppia di punti p, q in uno spazio affine astratto la traslazione $T_{\overrightarrow{pq}}$ è un'applicazione affine invertibile con inversa $T_{\overrightarrow{qp}}$.*

DIMOSTRAZIONE. Dalle proprietà distributive delle combinazioni baricentriche (Lemma 16.5.6) segue che se $\sum a_i = 1$ allora

$$T_{\overrightarrow{pq}}(a_1p_1 + \cdots + a_np_n) = a_1p_1 + \cdots + a_np_n + q - p = a_1(p_1 + q - p) + \cdots + a_n(p_n + q - p)$$

e dunque $T_{\overrightarrow{pq}}$ è un morfismo affine. Inoltre per ogni punto r si ha

$$T_{\overrightarrow{qp}} \circ T_{\overrightarrow{pq}}(r) = T_{\overrightarrow{qp}}(r + q - p) = r + q - p + p - q = r.$$

□

Continua a valere il lemma del parallelogramma 16.1.14, anche se nel caso astratto è necessaria una diversa dimostrazione.

LEMMA 16.5.13. *Per una quaterna ordinata di punti a, b, c, d in uno spazio affine astratto le seguenti condizioni sono equivalenti:*

$$T_{\overrightarrow{ab}}(c) = d, \qquad T_{\overrightarrow{ac}}(b) = d, \qquad T_{\overrightarrow{ab}} = T_{\overrightarrow{cd}}.$$

In particolare per ogni coppia di punti vi è un'unica traslazione che trasforma l'uno nell'altro.

DIMOSTRAZIONE. L'equivalenza $T_{\overrightarrow{ab}}(c) = d \iff T_{\overrightarrow{ac}}(b) = d$ segue immediatamente dalla simmetria delle combinazioni baricentriche. Se $T_{\overrightarrow{ab}} = T_{\overrightarrow{cd}}$, allora $T_{\overrightarrow{ab}}(c) = T_{\overrightarrow{cd}}(c) = c - c + d = d$. Viceversa, se $T_{\overrightarrow{ab}}(c) = d$ allora $c + b - a = d$ e quindi per ogni punto r

$$T_{\overrightarrow{cd}}(r) = r + d - c = r + (c + b - a) - c = r + b - a = T_{\overrightarrow{ab}}(r).$$

□

Dato uno spazio affine astratto A, indichiamo con $\mathcal{T}(A)$ l'insieme di tutte le traslazioni di A. Abbiamo visto che per ogni punto $p \in A$ l'applicazione $e_p \colon \mathcal{T}(A) \to A$, $f \mapsto f(p)$. Vogliamo adesso mostrare che esiste una struttura canonica di spazio vettoriale su $\mathcal{T}(A)$ che rende le applicazioni e_p degli isomorfismi affini. A tal fine dobbiamo definire un'operazione di somma $\mathcal{T}(A) \times \mathcal{T}(A) \xrightarrow{+} \mathcal{T}(A)$ ed un'operazione di prodotto per scalare $\mathbb{K} \times \mathcal{T}(A) \xrightarrow{+} \mathcal{T}(A)$ che soddisfano le condizioni elencate nella Sezione 4.2.

- Date due traslazioni f, g, scegliamo un punto $p \in A$ e poniamo $q = f(p)$, $r = g(q)$. Allora per ogni x vale

  $$g(f(x)) = g(x + q - p) = x + q - p + r - q = x + r - p = T_{\overrightarrow{pr}}(x),$$

  $$f(g(x)) = f(x + r - q) = x + r - q + q - p = x + r - p = T_{\overrightarrow{pr}}(x).$$

  Quindi $f \circ g = g \circ f \in \mathcal{T}(A)$. Definiamo la somma di due traslazioni come $f + g = f \circ g = g \circ f$. Ne segue che l'identità è l'elemento neutro per la somma e cha la somma è associativa e commutativa.
- Data una traslazione f ed uno scalare $a \in \mathbb{K}$. Scegliamo un punto $p \in A$ e definiamo il prodotto af mediante la formula

  $$af = T_{\overrightarrow{pr}}, \qquad r = (1 - a)p + af(p).$$

  Occorre mostrare che la definizione non dipende dalla scelta di p: bisogna mostrare che per ogni $q \in A$ vale

  $$T_{\overrightarrow{pr}} = T_{\overrightarrow{qs}}, \qquad r = (1 - a)p + af(p), \qquad s = (1 - a)q + af(q).$$

  Indichiamo con $z = f(q)$, dato che f è una traslazione si ha $f(x) = x + z - q$ per ogni x e quindi

  $$\begin{aligned} T_{\overrightarrow{pr}}(q) &= q + (1 - a)p + af(p) - p = q - ap + af(p) \\ &= q - ap + a(p + z - q) = (1 - a)q + af(q) = s, \end{aligned}$$

  e per il lemma sul paralleogramma vale $T_{\overrightarrow{pr}} = T_{\overrightarrow{qs}}$.

Delle 7 proprietà assiomatiche che definiscono la struttura di spazio vettoriale elencate all'inizio della Sezione 4.2 le prime 5 sono ovvie; la verifica delle ultime due viene lasciata per esercizio al lettore.

Mostriamo adesso che per ogni $p \in A$ l'applicazione bigettiva $e_p \colon \mathcal{T}(A) \to A$, $f \mapsto f(p)$, è affine. Siano $f_1, \dots, f_n \in \mathcal{T}(A)$ e $a_1, \dots, a_n \in \mathbb{K}$ tali che $\sum a_i = 1$. Dobbiamo mostrare che

$$(\sum a_i f_i)(p) = \sum a_i f_i(p).$$

Indichamo con $q_i = f_i(p)$, allora per ogni punto x vale

$$a_i f_i(x) = x + a_i q_i - a_i p,$$

e quindi

$$(\sum a_i f_i)(x) = x + \sum a_i q_i - \sum a_i p = \sum a_i (x + q_i - p),$$

che per $x = p$ diventa

$$(\sum a_i f_i)(p) = \sum a_i q_i = \sum a_i f_i(p).$$

Dunque, abbiamo dimostrato che a meno di isomorfismi affini, ogni spazio affine astratto è isomorfo ad uno spazio vettoriale e che quindi *la definizione di spazio affine data nella Sezione 16.1 non è affatto restrittiva.*

Prima di passare agli esercizi, richiamiamo la definizione di spazio affine modellato, una delle definizioni di spazio affine maggiormente presenti in letteratura.

Definizione 16.5.14. Sia V uno spazio vettoriale su di un campo $\mathbb{K}$. Uno **spazio affine modellato** su V è una terna $(A, V, \overset{\rightarrow}{})$, dove A è un insieme e $A \times A \xrightarrow{\rightarrow} V$ è un'applicazione che soddisfa gli assiomi:

(1) Per ogni $p \in A$ vale $\overrightarrow{pp} = 0$.
(2) *(Relazione di Chasles)* Per ogni $p, q, r \in A$ vale $\overrightarrow{pq} + \overrightarrow{qr} + \overrightarrow{rp} = 0$.
(3) Per ogni $p \in A$ e per ogni $v \in V$ esiste un unico punto $q \in A$ tale che $\overrightarrow{pq} = v$.

L'unico punto q descritto al punto 3 si dice **traslato** di p mediante v e si indica con $q = v + p$.

Ogni spazio vettoriale V è in modo naturale uno spazio affine modellato su sé stesso ed ogni sottospazio affine E di V è uno spazio affine modellato su $K = \{u - v \mid u, v \in E\}$.

Ogni spazio affine modellato $(A, V, \overset{\rightarrow}{})$ ha una struttura naturale di spazio affine astratto. Se $p_0, \dots, p_n \in A$ e $\sum t_i = 1$, si definisce

$$t_0 p_0 + \cdots + t_n p_n = (\sum_i t_i \overrightarrow{p_0 p_i}) + p_0.$$

In particolare per ogni $p, q, x \in A$ vale $\overrightarrow{pq} + x = T_{\overrightarrow{pq}}(x)$: infatti dalla relazione di Chasles

$$\overrightarrow{pq} = \overrightarrow{xx} - \overrightarrow{xp} + \overrightarrow{xq}$$

e quindi

$$\overrightarrow{pq} + x = (\overrightarrow{xx} - \overrightarrow{xp} + \overrightarrow{xq}) + x = x - p + q = T_{\overrightarrow{pq}}(x).$$

In conclusione, le nostre nozioni di spazio affine, spazio affine astratto e spazio affine modellato sono tra loro equivalenti.

Esercizi.

851. Siano A e B spazi affini. Provare che:
(1) Il prodotto cartesiano $A \times B$ è uno spazio affine.
(2) L'insieme di tutte le applicazioni affini $f \colon A \to B$ è uno spazio affine.

852. Mostrare che un'applicazione $f \colon A \to B$ tra spazi affini non vuoti è affine se e solo se esiste un'applicazione lineare $g \colon \mathcal{T}(A) \to \mathcal{T}(B)$ tale che $g(T_{pq}) = T_{f(p)f(q)}$ per ogni $p, q \in A$.

853. Siano $(A, V, \overset{\rightarrow}{})$ e $(B, W, \overset{\rightarrow}{})$ spazi affini modellati e non vuoti. Provare che un'applicazione $f \colon A \to B$ è affine, ossia commuta con le combinazioni baricentriche, se e solo se esiste un'applicazione lineare $g \colon V \to W$ tale che $\overrightarrow{f(p)f(q)} = g(\overrightarrow{pq})$ per ogni $p, q \in A$.

854. Sia $(A, V, \overset{\rightarrow}{})$ uno spazio affine modellato. Mostrare che per ogni $a \in A$ l'applicazione

$$V \to A, \quad v \mapsto v + a,$$

è un isomorfismo affine.

855 (Ⓐ). Per questo esercizio sono richieste alcune nozioni di teoria dei gruppi. Dimostrare che le traslazioni formano un sottogruppo normale abeliano del gruppo delle affinità di uno spazio affine in sé.

CAPITOLO 17

Complementi: le trascendenze famose

La teoria dei numeri trascendenti ha avuto inizio con una memoria di Liouville del 1844, nella quale l'autore dimostrò che esisteva una classe, piuttosto ampia, di numeri che non erano radici di equazioni algebriche a coefficienti interi. I problemi legati all'irrazionalità erano invece già ampiamente studiati e risale infatti al 1744 la dimostrazione di Eulero dell'irrazionalità di e ed al 1761 la dimostrazione di Lambert dell'irrazionalità di π.

Nel 1874 Cantor introdusse il concetto di infinito numerabile e questo portò immediatamente alla constatazione che "quasi tutti" i numeri sono trascendenti.

La dimostrazione della trascendenza di e compare in una memoria di Hermite del 1873 che è stata fonte di ispirazione per molti anni a seguire. Nel 1882 Lindemann riuscì ad estendere il lavoro di Hermite per dimostrare la trascendenza di π e, di conseguenza, a chiudere definitivamente l'antico problema della quadratura del cerchio con riga e compasso.

In questo capitolo daremo una dimostrazione dei Teoremi 4.8.2, 4.8.3 e 4.8.4, ossia dimostreremo la trascendenza dei numeri e, l e π. A differenza dei precedenti capitoli, dove le tecniche usate sono state quasi esclusivamente algebriche, avremo bisogno di usare alcuni risultati di analisi matematica solitamente insegnati al primo anno dei corsi universitari: useremo in particolare il teorema del valor medio, gli sviluppi di Taylor delle funzioni esponenziali e trigonometriche più alcune proprietà delle serie numeriche, sia reali che complesse. Useremo inoltre più volte il teorema fondamentale dell'algebra, nella cui dimostrazione abbiamo fatto uso del principio di completezza dei numeri reali, sotto forma del teorema di esistenza degli zeri.

Ricordiamo che un numero complesso si dice algebrico se è radice di un polinomio non nullo a coefficienti interi $f(x) \in \mathbb{Z}[x]$, mentre un numero complesso che non è algebrico viene detto trascendente.

LEMMA 17.0.1. *Per un numero complesso α le seguenti condizioni sono equivalenti:*

(1) *α è algebrico;*
(2) *α è radice di un polinomio monico a coefficienti razionali;*
(3) *esiste una successione finita $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{C}$ tale che*

$$\alpha = \alpha_1, \qquad \prod_{i=1}^{n}(x - \alpha_i) \in \mathbb{Q}[x].$$

DIMOSTRAZIONE. Se $\alpha = \alpha_1$ ed il polinomio monico $\prod_{i=1}^{n}(x-\alpha_i)$ ha coefficienti razionali, moltiplicando per un denominatore comune $a \in \mathbb{Z}$, $a > 0$, abbiamo che il polinomio $a\prod_{i=1}^{n}(x-\alpha_i)$ ha coefficienti interi ed annulla α che pertanto è un numero algebrico. Viceversa, se $f(\alpha) = 0$ con $f(x) \in \mathbb{Z}[x]$ di grado positivo, dividendo per il coefficiente direttivo otteniamo un polinomio monico $g(x) \in \mathbb{Q}[x]$ tale che $g(\alpha) = 0$. Per il teorema fondamentale dell'algebra il polinomio $g(x)$ si scrive come un prodotto di polinomi di primo grado

$$g(x) = \prod_{i=1}^{n}(x - \alpha_i), \qquad \alpha_i \in \mathbb{C},$$

e siccome $g(\alpha) = \prod_{i=1}^{n}(\alpha - \alpha_i) = 0$ deve essere $\alpha = \alpha_i$ per qualche i, viz. $\alpha = \alpha_1$ a meno di permutazione degli indici.

□

Abbiamo già dimostrato che somme, prodotti, opposti ed inversi di numeri algebrici sono ancora algebrici e quindi che l'insieme $\overline{\mathbb{Q}} \subset \mathbb{C}$ di tutti i numeri algebrici è un sottocampo di $\mathbb{C}$; in questo capitolo daremo una diversa dimostrazione di questo fatto usando il teorema delle funzioni simmetriche.

17.1. Irrazionalità di e ed l

Prima di dimostrare i risultati di trascendenza, occupiamoci del problema più semplice di dimostrare l'irrazionalità dei tre numeri e, l, π. In questa sezione ci occuperemo dei primi due, mentre l'irrazionalità di π sarà trattata più avanti.

Il numero di Nepero può essere definito come il valore della serie numerica

$$e = \sum_{n=0}^{+\infty} \frac{1}{n!}\,.$$

Supponiamo per assurdo che e sia razionale, esiste allora un intero positivo q tale che $qe \in \mathbb{Z}$. A maggior ragione anche il prodotto di e per $q!$ è intero e quindi

$$\sum_{n=0}^{+\infty} \frac{q!}{n!} \in \mathbb{Z}\,.$$

I primi $q+1$ termini della precedente serie sono interi e quindi

$$\sum_{n=q+1}^{+\infty} \frac{q!}{n!} = \sum_{n=0}^{+\infty} \frac{q!}{n!} - \sum_{n=0}^{q} \frac{q!}{n!} \in \mathbb{Z}\,.$$

D'altra parte, per ogni $n > q$ si hanno le disuguaglianze

$$0 < \frac{q!}{n!} = \frac{1}{(q+1)\cdots n} < \frac{1}{(q+1)(q+2)^{n-q-1}}$$

dalle quali deduciamo

$$0 < \sum_{n=q+1}^{+\infty} \frac{q!}{n!} < \sum_{n=q+1}^{+\infty} \frac{1}{(q+1)(q+2)^{n-q-1}} = \frac{1}{q+1} \sum_{m=0}^{+\infty} \frac{1}{(q+2)^m} = \frac{q+2}{(q+1)^2} < 1$$

il che è assurdo in quanto nessun intero è strettamente compreso tra 0 ed 1.

L'irrazionalità di l si dimostra in modo del tutto simile. Supponiamo per assurdo che ql sia intero per qualche intero positivo q e scegliamo un intero $a \geq 2$ sufficientemente grande tale che $q \leq 10^{a!}$. Moltiplicando ql per $10^{a!}$ otteniamo un intero positivo che si scrive come

$$lq10^{a!} = \sum_{n=1}^{a} q\frac{10^{a!}}{10^{n!}} + \sum_{n=a+1}^{+\infty} q\frac{10^{a!}}{10^{n!}}.$$

Il primo addendo del termine a destra dell'uguaglianza è intero e di conseguenza anche la somma della serie

$$\sum_{n=a+1}^{+\infty} q\frac{10^{a!}}{10^{n!}} = \sum_{i=1}^{+\infty} q\frac{10^{a!}}{10^{(a+i)!}}$$

è un numero intero. Ma questo è assurdo perché per ogni $a \geq 2$, $i > 0$ vale $(a+i)! > a! + a! + i$ e quindi

$$0 < \sum_{i=1}^{+\infty} q\frac{10^{a!}}{10^{(a+i)!}} \leq \sum_{i=1}^{+\infty} \frac{q}{10^{a!}}\,\frac{1}{10^i} \leq \sum_{i=1}^{+\infty} \frac{1}{10^i} = \frac{1}{9} < 1.$$

L'argomento che abbiamo usato funziona in generale ogniqualvolta si voglia dimostrare l'irrazionalià della somma di una serie di numeri razionali che tendono a zero in maniera molto veloce. Il lettore può facilmente verificare che tale strategia non funziona se applicata alle serie

$$e^2 = \sum_{n=0}^{+\infty} \frac{2^n}{n!}, \qquad \frac{\pi}{4} = \sum_{n=0}^{+\infty} \frac{(-1)^n}{2n+1}.$$

Esercizi.

856. Usare gli sviluppi in serie

$$e = \sum_{n\geq 0} \frac{1}{n!}\,, \qquad \frac{1}{e} = \sum_{n\geq 0} \frac{(-1)^n}{n!}\,,$$

per dimostrare che $1, e, e^{-1}$ sono linearmente indipendenti su $\mathbb{Q}$ o, equivalentemente, che e non è radice di un polinomio di secondo grado a coefficienti interi.

857. Dimostrare che $e^{\sqrt{2}}$ è irrazionale (sugg.: sviluppo in serie di $e^{\sqrt{2}} + e^{-\sqrt{2}}$).

858. Dimostrare l'irrazionalità di $\sin(1)$ e $\cos(1)$ usando gli sviluppi in serie

$$\sin(1) = \sum_{n\geq 0} \frac{(-1)^n}{(2n+1)!}, \qquad \cos(1) = \sum_{n\geq 0} \frac{(-1)^n}{(2n)!}.$$

859. Completare con i dettagli mancanti il seguente ragionamento: il calcolo dei valori approssimati di π può essere fatto mediante lo sviluppo in serie dell'arcotangente

$$\arctan(x) = \int_0^x \frac{1}{1+t^2}dt = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots$$

da cui segue

$$\frac{\pi}{4} = \arctan(1) = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1}.$$

Purtroppo tale serie converge abbastanza lentamente. Per avere stime più rapide si possono considerare i polinomi

$$g_n(x) = \frac{x^{4n}(1-x)^{4n}}{4^n} = x^{4n}\left(\frac{1-x}{\sqrt{2}}\right)^{4n}, \qquad n > 0.$$

Siccome $g_n(i) = g_n(-i) = (-1)^n$ ne segue che $g_n(x) - (-1)^n$ è divisibile per $1 + x^2$ in $\mathbb{Q}[x]$ e vale

$$\frac{\pi}{4} = \arctan(1) = \int_0^1 \frac{1-(-1)^n g_n(x)}{1+x^2}dx + (-1)^n \int_0^1 \frac{x^{4n}(1-x)^{4n}}{4^n(1+x^2)}dx.$$

In conclusione, $\int_0^1 \frac{1-(-1)^n g_n(x)}{1+x^2}dx$ è un numero razionale che approssima $\pi/4$ con un errore inferiore a 4^{-5n}.

17.2. L'operatore di derivazione

Dato un qualsiasi polinomio

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n \in \mathbb{C}[x]$$

definiamo il suo derivato $f(x)' \in \mathbb{C}[x]$ mediante la formula

$$f(x)' = (a_0 + a_1 x + \cdots + a_n x^n)' = a_1 + 2a_2 x + \cdots + n a_n x^{n-1}\,.$$

Qualora $f(x)$ abbia coefficienti reali, si vede immediatamente che $f'(x)$ coincide con la derivata di $f(x)$ intesa come limite del rapporto incrementale. L'operatore $f(x) \mapsto f'(x)$ è lineare su $\mathbb{C}$ e soddisfa la regola di Leibniz; ciò significa che:

(1) $(af(x))' = af(x)'$, $a \in \mathbb{C}$;
(2) $(f(x) + g(x))' = f(x)' + g(x)'$;
(3) $fg(x)' = f(x)'g(x) + f(x)g(x)'$.

Le prime due proprietà sono di immediata verifica. La terza proprietà, detta regola di Leibniz, si può dimostrare direttamente dalla definizione in maniera puramente algebrica: infatti, se

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n, \qquad g(x) = b_0 + b_1 x + \cdots + b_m x^m,$$

si ha

$$\begin{aligned} f(x)'g(x)+f(x)g(x)' &= (\sum_{i=0}^{n} a_i x^i)'(\sum_{j=0}^{m} b_j x^j)+(\sum_{i=0}^{n} a_i x^i)(\sum_{j=0}^{m} b_j x^j)' \\ &= (\sum_{i=0}^{n} i a_i x^{i-1})(\sum_{j=0}^{m} b_j x^j)+(\sum_{i=0}^{n} a_i x^i)(\sum_{j=0}^{m} j b_j x^{j-1}) \\ &= \sum_{k=0}^{n+m}\sum_{i+j=k} i a_i b_j x^{k-1}+\sum_{k=0}^{n+m}\sum_{i+j=k} j a_i b_j x^{k-1} \\ &= \sum_{k=0}^{n+m}\sum_{i+j=k} k a_i b_j x^{k-1} = (\sum_{k=0}^{n+m}\sum_{i+j=k} a_i b_j x^k)' = f g(x)'\,. \end{aligned}$$

Per uso futuro, notiamo che una prima conseguenza della regola di Leibniz è la formula

$$((x-a)^n)' = n(x-a)^{n-1}, \qquad a\in\mathbb{C},$$

che si dimostra per induzione su $n\geq 0$. Dato un qualsiasi polinomio $f(x)\in\mathbb{C}[x]$ possiamo farne la derivata prima, seconda, terza e più in generale la derivata h-esima $f^{(h)}(x)$:

$$f^{(0)}(x)=f(x), \quad f^{(1)}(x)=f(x)', \quad f^{(2)}(x)=f(x)'', \ \dots \quad .$$

È chiaro che se $f(x)$ ha grado d allora $f^{(h)}(x)$ ha grado $d-h$ se $h\leq d$, mentre $f^{(h)}(x)$ si annulla per $h>d$.

PROPOSIZIONE 17.2.1. *La molteplicità di un numero complesso α come radice di un polinomio $f(x)\in\mathbb{C}[x]$ è uguale al più piccolo intero h tale che $f^{(h)}(\alpha)\neq 0$.*

DIMOSTRAZIONE. Esercizio. □

Se $f(x)=\sum_{i=0}^{d} a_i x^i$, allora per ogni $h\geq 0$ si ha

$$f^{(h)}(x)=\sum_{i=h}^{d}\frac{i!}{(i-h)!}a_i x^{i-h}=h!\sum_{i=h}^{d}\binom{i}{h}a_i x^{i-h}, \qquad f^{(h)}(0)=h!\,a_h. \tag{17.1}$$

In particolare per ogni $f(x)\in\mathbb{C}[x]$ vale lo sviluppo di Taylor

$$f(x)=\sum_{h=0}^{+\infty} f^{(h)}(0)\frac{x^h}{h!},$$

dove la sommatoria ha solo un numero finito di addendi diversi da 0. Un'altra conseguenza delle formule (17.1) che useremo spesso è che se $f(x)=\sum_{i=0}^{d} a_i x^i\in\mathbb{Z}[x]$ è un polinomio a coefficienti interi, allora per ogni $h\geq 0$ ed ogni $m\in\mathbb{Z}$ il numero $f^{(h)}(m)$ è un intero divisibile per $h!$. Infatti, se $f(x)=\sum_{i=0}^{d} a_i x^i$ con $a_0,\dots,a_d\in\mathbb{Z}$, si può scrivere

$$f^{(h)}(m)=h!\sum_{i=h}^{d}\binom{i}{h}a_i m^{i-h}.$$

LEMMA 17.2.2. *Siano dati $a,n\in\mathbb{Z}$ con $n\geq 0$, $g(x)\in\mathbb{Z}[x]$ e consideriamo il polinomio*

$$f(x)=(x-a)^n g(x)\in\mathbb{Z}[x].$$

Allora $f^{(h)}(a)$ è un intero divisibile per $n!$, per ogni $h\geq 0$.

DIMOSTRAZIONE. Siccome $f^{(h)}(a)$ è divisibile per $h!$, basta dimostrare che $f^{(h)}(a)=0$ per ogni $h<n$. Ciò è senz'altro vero per $h=0$ e $n>0$; se $0<h<n$ si ha

$$f^{'}(x)=(x-a)^{n-1}(ng(x)+(x-a)g'(x))$$

e la conclusione segue per induzione su n e dal fatto che $f^{(h)}=(f')^{(h-1)}$. □

Esercizi.

860 (formule di Newton–Girard). Dati n numeri complessi $a_1, \dots, a_n$ si considerino le quantità

$$s_k = \sum_{i=0}^{n} a_i^k, \qquad k = 1, \dots, n,$$

ed i numeri complessi $p_0, \dots, p_n$ determinati dall'identità polinomiale

$$p(t) = \prod_{i=1}^{n}(1 + a_i x) = \sum_{i=0}^{n} p_i x^i;$$

$$p_0 = 1, \quad p_1 = \sum_i a_i, \quad p_2 = \sum_{i<j} a_i a_j, \quad \text{eccetera.}$$

I seguenti punti, svolti nell'ordine, permetteranno di dimostrare le *formule di Newton–Girard*, note anche come *identità di Newton*:

$$k p_k = \sum_{i=1}^{k} (-1)^{i-1} s_i \, p_{k-i}, \qquad k = 1, \dots, n. \tag{17.2}$$

(1) provare che

$$p'(x) = \sum_{i=1}^{n} a_i \frac{p(x)}{1 + a_i x};$$

(2) dati due polinomi $p(x), q(x) \in \mathbb{C}[x]$, scriviamo $p(x) \equiv q(x)$ se x^n divide la differenza $p(x) - q(x)$, ossia se i coefficienti di $p(x)$ e $q(x)$ coincidono fino al grado $n-1$. Provare che

$$a_i \frac{p(x)}{1 + a_i x} \equiv a_i p(x) \sum_{j=0}^{n-1} (-1)^j a_i^j x^j, \qquad p'(x) \equiv p(x) \left(\sum_{j=0}^{n-1} (-1)^j s_{j+1} x^j \right);$$

(3) dimostrare le formule (17.2).

17.3. Irrazionalità di π

Il modo più semplice attualmente noto per dimostrare l'irrazionalità di π fa intervenire la successione dei polinomi

$$f_n(x) = \frac{x^n (1-x)^n}{n!}, \qquad n \geq 0\,.$$

LEMMA 17.3.1. *Nelle notazioni precedenti, per ogni coppia di interi non negativi h, n ed ogni polinomio $g(x) \in \mathbb{Z}[x]$ a coefficienti interi, la derivata h-esima del polinomio $f_n(x)g(x)$ assume valori interi per $x = 0$ ed $x = 1$.*

DIMOSTRAZIONE. Il risultato è del tutto ovvio se $n = 0$ oppure se $h = 0$. Dimostriamo i rimanenti casi per induzione su h. Supponiamo quindi $n, h > 0$; siccome

$$(f_n(x)g(x))' = f_{n-1}(x)(1-2x)g(x) + f_n(x)g(x)'$$

per l'ipotesi induttiva le derivate $(h-1)$-esime di $f_{n-1}(x)(1-2x)g(x)$ e $f_n(x)g(x)'$ assumono valori interi per $x = 0$ e $x = 1$. □

Siamo adesso in grado di dimostrare non solo l'irrazionalità di π ma anche quella di π^2. Supponiamo per assurdo che $\pi^2 = \dfrac{a}{b}$, ossia $\dfrac{a}{\pi} = b\pi$, con a, b interi positivi. Per ogni intero positivo n consideriamo l'integrale

$$\int_0^1 \pi a^n f_n(x) \sin(\pi x) dx.$$

Siccome $0 \leq x^n (1-x)^n \sin(\pi x) \leq 1$ per ogni $x \in [0, 1]$ si hanno le disuguaglianze

$$0 < \int_0^1 \pi a^n f_n(x) \sin(\pi x) dx \leq \int_0^1 \pi \frac{a^n}{n!}$$

e quindi l'integrale non può essere un numero intero per ogni n sufficientemente grande tale che $n! > \pi a^n$. La contraddizione è adesso una conseguenza immediata del seguente lemma.

LEMMA 17.3.2. *Si assuma, per assurdo, che* $\pi^2 = a/b$ *con* a, b *interi non nulli e sia* $g(x) \in \mathbb{R}[x]$ *un polinomio di grado* $2n$ *tale che* $g(x)$ *e tutte le sue derivate* $g'(x)$, $g''(x)$,..., $g^{(h)}(x)$,... *assumono valori interi per* $x = 0, 1$, *allora*

$$\int_0^1 \pi a^n g(x) \sin(\pi x) dx \in \mathbb{Z}.$$

DIMOSTRAZIONE. Induzione su n. Per $n = 0$ il polinomio $g(x) = g(0)$ è una costante intera e dunque

$$\int_0^1 \pi g(0) \sin(\pi x) dx = [-g(0) \cos(\pi x)]_0^1 = 2g(0) \in \mathbb{Z}.$$

Quando $n > 0$, per la formula di integrazione per parti si ottiene

$$\begin{aligned}\int_0^1 \pi a^n g(x) \sin(\pi x) dx &= [-a^n g(x) \cos(\pi x)]_0^1 + \int_0^1 a^n g'(x) \cos(\pi x) dx \\ &= \left[-a^n g(x) \cos(\pi x) + \frac{a^n}{\pi} g'(x) \sin(\pi x)\right]_0^1 - \int_0^1 \frac{a^n}{\pi} g''(x) \sin(\pi x) dx \\ &= a^n (g(0) + g(1)) - b \int_0^1 \pi a^{n-1} g''(x) \sin(\pi x) dx.\end{aligned}$$

Basta adesso osservare che $g''(x)$ ha grado $\leq 2(n-1)$ e si può applicare l'ipotesi induttiva. □

Esercizi.

861. Siano a, b interi non negativi. Dimostrare per induzione su b che vale la formula

$$\int_0^1 x^a (1-x)^b dx = \frac{a!\, b!}{(a+b+1)!}.$$

862. Sia r un intero positivo e supponiamo per assurdo che $e^r = a/b$, con a, b interi positivi. Dedurre da ciò una contraddizione considerando, per n sufficientemente grande, l'integrale

$$\int_0^1 b r^{2n+1} \frac{x^n (1-x)^n}{n!} e^{rx} dx.$$

Dedurre dall'irrazionalità delle potenze intere di e che anche i logaritmi naturali degli interi $n > 1$ sono numeri irrazionali.

17.4. La trascendenza di l

Storicamente il primo numero reale del quale è stata dimostrata la trascendenza è il numero di Liouville

$$l = \sum_{n=1}^{+\infty} \frac{1}{10^{n!}} = 0,110001000000000000000010000000\ldots .$$

Prima di dimostrare la trascendenza di l abbiamo bisogno di un risultato di natura generale riguardante le approssimazioni razionali di numeri algebrici irrazionali. Ricordiamo dalla Definizione 4.8.6 che il grado di un numero algebrico $\alpha \in \mathbb{C}$ è il più piccolo intero positivo d tale che x è radice di un polinomio di grado d a coefficienti interi, e che un numero è razionale se e solo se è algebrico di grado 1.

LEMMA 17.4.1. *Sia* $x \in \mathbb{R}$ *un numero algebrico di grado* $d > 1$. *Allora esiste un numero reale positivo* M *tale che per ogni coppia di numeri interi* p, q, *con* $q > M$ *vale*

$$\left| x - \frac{p}{q} \right| > \frac{1}{q^{d+1}}.$$

DIMOSTRAZIONE. Per ipotesi il numero x è irrazionale. Sia $h(t)$ un polinomio di grado d a coefficienti interi tale che $h(x) = 0$ e sia $\delta > 0$ sufficientemente piccolo e tale che $h(t)$ non abbia radici reali nell'intervallo $[x-\delta, x+\delta]$ diverse da x. Indichiamo con m il massimo della funzione continua $|h'(t)|$ nell'intervallo $[x-\delta, x+\delta]$ e definiamo

$$M = \max\left(m, \sqrt[d+1]{\frac{1}{\delta}}\right).$$

Consideriamo adesso due numeri interi p, q, con $q > M$; in particolare

$$q^{d+1} > M^{d+1} \geq \frac{1}{\delta}, \qquad \frac{1}{q^{d+1}} < \delta,$$

e quindi se $|x - \frac{p}{q}| > \delta$ il lemma è banalmente verificato. Se $|x - \frac{p}{q}| \leq \delta$, allora per come abbiamo scelto δ si ha $h(\frac{p}{q}) \neq 0$ e per il teorema del valor medio esiste $\xi \in [x-\delta, x+\delta]$ tale che

$$0 \neq h\left(\frac{p}{q}\right) = h\left(\frac{p}{q}\right) - h(x) = h'(\xi)\left(\frac{p}{q} - x\right).$$

Moltiplicando per q^d e prendendo il valore assoluto

$$\left|q^d h\left(\frac{p}{q}\right)\right| = q^d |h'(\xi)| \left|\frac{p}{q} - x\right| \leq q^d m \left|\frac{p}{q} - x\right|.$$

D'altra parte $q^d h\left(\frac{p}{q}\right)$ è un intero non nullo e quindi il suo valore assoluto è ≥ 1, e quindi

$$1 \leq \left|q^d h\left(\frac{p}{q}\right)\right| \leq q^d m \left|\frac{p}{q} - x\right|$$

da cui deduciamo

$$\left|\frac{p}{q} - x\right| \geq \frac{1}{q^d m} \geq \frac{1}{q^d M} \geq \frac{1}{q^{d+1}}.$$

□

TEOREMA 17.4.2. *Il numero di Liouville* $l = \sum_n \frac{1}{10^{n!}}$ *è trascendente.*

DIMOSTRAZIONE. Abbiamo già dimostrato che l è irrazionale. Supponiamo per assurdo che sia algebrico di grado $d > 1$. Per il Lemma 17.4.1 esiste una costante $M > 0$ tale che per ogni coppia di numeri interi p, q, con $q > M$ vale

$$\left|l - \frac{p}{q}\right| > \frac{1}{q^{d+1}}.$$

Sia $N > d$ un intero tale che $10^{N!} > M$ e consideriamo i numeri interi $q = 10^{N!}$ e p definito mediante la formula

$$\frac{p}{10^{N!}} = \sum_{n=1}^{N} \frac{1}{10^{n!}}.$$

Abbiamo dunque

$$\left|l - \frac{p}{10^{N!}}\right| = \sum_{n=N+1}^{+\infty} \frac{1}{10^{n!}} = \frac{1}{10^{(N!)N}} \sum_{n=N+1}^{+\infty} \frac{1}{10^{n!-(N!)N}}$$

e siccome $n! - (N!)N \geq n - N$ per ogni $n > N$ otteniamo

$$\left|l - \frac{p}{10^{N!}}\right| \leq \frac{1}{10^{(N!)N}} \sum_{n=N+1}^{+\infty} \frac{1}{10^{n-N}} = \frac{1}{9} \frac{1}{(10^{N!})^N} < \frac{1}{(10^{N!})^N} \leq \frac{1}{q^{d+1}}$$

in contraddizione con il Lemma 17.4.1. □

Esercizi.

863. Mostrare che i numeri $\sqrt{2}+\sqrt{3}$ e $\sqrt{2}+\sqrt{3}+\sqrt{5}$ sono algebrici e calcolarne i rispettivi gradi.

17.5. La trascendenza di e

La dimostrazione della trascendenza di e utilizza la stessa idea usata nella dimostrazione dell'irrazionalità di π, ossia la stima numerica di una opportuna successione di integrali definiti. Per chiarezza espositiva conviene dedurre le stime necessarie da un opportuno lemma che enunceremo subito dopo aver richiamato la definizione di esponenziale complesso.

Ricordiamo che una serie di numeri complessi $\sum_{n=0}^{\infty} a_n$ si dice assolutamente convergente se $\sum_{n=0}^{\infty} |a_n| < +\infty$. Chiaramente se $a_n = b_n + ic_n$, con $b_n, c_n \in \mathbb{R}$, e se $\sum_{n=0}^{\infty} a_n$ è assolutamente convergente, allora pure le serie $\sum_{n=0}^{\infty} b_n$ e $\sum_{n=0}^{\infty} c_n$ sono assolutamente convergenti e si può definire

$$\sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty} b_n + i \sum_{n=0}^{\infty} c_n \in \mathbb{C} .$$

Ad esempio, la serie esponenziale

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

è assolutamente convergente per ogni $z \in \mathbb{C}$. Come nel caso delle serie reali si possono fare somme e prodotti di Cauchy di serie assolutamente convergenti senza rischio alcuno.

Allo stesso modo del caso reale si dimostra che $e^{z+w} = e^z e^w$ per ogni $z, w \in \mathbb{C}$ ed in particolare, per ogni $r, \theta \in \mathbb{R}$, si ha $e^{r+i\theta} = e^r e^{i\theta}$. Inoltre,

$$e^{i\theta} = \sum_{n=0}^{\infty} i^n \frac{\theta^n}{n!} = \sum_{m=0}^{\infty} (-1)^m \frac{\theta^{2m}}{(2m)!} + i \sum_{m=0}^{\infty} (-1)^m \frac{\theta^{2m+1}}{(2m+1)!}$$

e confrontando con i ben noti sviluppi in serie di seno e coseno ricaviamo la formula

$$e^{i\theta} = \cos\theta + i \sin\theta .$$

Ad ogni polinomio $f(x) \in \mathbb{C}[x]$ associamo due successioni di funzioni $H^n_{f(x)}, I^n_{f(x)} \colon \mathbb{C} \to \mathbb{C}$, $n = 1, 2, \dots$ definite nel modo seguente:

$$\begin{aligned}
H^1_{f(x)}(z) &= \sum_{h=0}^{\infty} f^{(h)}(z), \qquad z \in \mathbb{C}, \\
H^n_{f(x)}(z) &= H^1_{x^{n-1} f(x)^n}(z), \qquad n \geq 2,\ z \in \mathbb{C}, \\
I^1_{f(x)}(z) &= e^z \sum_{h=0}^{\infty} f^{(h)}(0) - \sum_{h=0}^{\infty} f^{(h)}(z) = e^z H^1_{f(x)}(0) - H^1_{f(x)}(z), \qquad z \in \mathbb{C}, \\
I^n_{f(x)}(z) &= I^1_{x^{n-1} f(x)^n}(z) = e^z H^n_{f(x)}(0) - H^n_{f(x)}(z), \qquad n \geq 2,\ z \in \mathbb{C}.
\end{aligned}$$

Notiamo che nella definizione di $H^1_{f(x)}, I^1_{f(x)}$ le sommatorie sono di fatto finite in quanto $f^{(h)}(x) = 0$ non appena h supera il grado di f. Ad esempio si ha

$$I^1_1(z) = e^z - 1, \quad I^1_x(z) = I^2_1(z) = e^z - z - 1, \quad I^1_{x^2}(z) = I^3_1(z) = 2e^z - z^2 - 2z - 2, \quad \dots$$

Osservazione 17.5.1. Se $f(x) \in \mathbb{R}[x]$, allora la restrizione di $I^1_{f(x)}$ all'asse reale è uguale alla soluzione del problema di Cauchy

$$\gamma'(z) = \gamma(z) + f(z), \qquad \gamma(0) = 0,$$

che si risolve nel modo standard:

$$\gamma(z) = \int_0^z f(x) e^{z-x} dx. \tag{17.3}$$

Chi studierà la teoria delle funzioni olomorfe scoprirà che la formula integrale (17.3) ha perfettamente senso anche per $z \in \mathbb{C}$, dove l'integrale è fatto lungo un qualsiasi cammino che congiunge 0 e z nel piano di Gauss.

Lemma 17.5.2. *Nelle notazioni precedenti, se $f(x) \in \mathbb{C}[x]$ allora esiste una costante positiva C, dipendente solo da $f(x)$, tale che, per ogni radice α di $f(x)$ e per ogni $n > 0$ si ha*

$$|I^n_{f(x)}(\alpha)| \leq e^{|\alpha|} C^n.$$

DIMOSTRAZIONE. Ad ogni polinomio $f(x) \in \mathbb{C}[x]$, $f(x) = \sum a_i x^i$, associamo il suo "tildato" $\tilde{f}(x) = \sum |a_i| x^i \in \mathbb{R}[x]$. Si noti che le operazioni di derivazione e tildatura commutano tra loro e che

$$\tilde{f}^{(p)}(0) = p!|a_p| = |f^{(p)}(0)| \, .$$

La disuguaglianza triangolare $|z+w| \leq |z| + |w|$, $z, w \in \mathbb{C}$, produce analoghe disuguaglianze sui polinomi tildati; in particolare, se $f, g \in \mathbb{C}[x]$, allora per ogni numero reale $a \geq 0$ vale

$$|f(a)| \leq \tilde{f}(a), \qquad 0 \leq \widetilde{f+g}(a) \leq \tilde{f}(a) + \tilde{g}(a), \qquad 0 \leq \widetilde{fg}(a) \leq \tilde{f}(a)\, \tilde{g}(a).$$

Le dimostrazioni sono lasciate per esercizio.

Mostriamo adesso che per ogni $f(x) \in \mathbb{C}[x]$ e per ogni $z \in \mathbb{C}$ vale la disuguaglianza

$$|I^1_{f(x)}(z)| \leq |z| e^{|z|} \tilde{f}(|z|) \, .$$

Sviluppando l'esponenziale in serie ed usando gli sviluppi di Taylor

$$f^{(p)}(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} f^{(p+n)}(0)$$

otteniamo

$$I^1_{f(x)}(z) = \sum_{n,p=0}^{\infty} \frac{z^n}{n!} f^{(p)}(0) - \sum_{n,p=0}^{\infty} \frac{z^n}{n!} f^{(p+n)}(0) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \sum_{p=0}^{n-1} f^{(p)}(0).$$

Ponendo $q = n - 1 - p$ si ottiene

$$I^1_{f(x)}(z) = z \sum_{p,q=0}^{\infty} \frac{z^p z^q}{(p+q+1)!} f^{(p)}(0)$$

e per la disuguaglianza triangolare

$$|I^1_{f(x)}(z)| \leq |z| \sum_{p,q=0}^{\infty} \frac{|z^q||z^p|}{(p+q+1)!} |f^{(p)}(0)| \leq |z| \sum_{p,q=0}^{\infty} \frac{|z^q|}{q!} \frac{|z|^p}{p!} \tilde{f}^{(p)}(0) = |z| e^{|z|} \tilde{f}(|z|).$$

Dopo queste stime preliminari siamo in grado di dimostrare il lemma. Siano $f(x) \in \mathbb{C}[x]$ e indichiamo con $\alpha_1, \ldots, \alpha_m \in \mathbb{C}$ le sue radici e consideriamo come costante

$$C = \max(|\alpha_1| \tilde{f}(|\alpha_1|), |\alpha_2| \tilde{f}(|\alpha_2|), \ldots, |\alpha_m| \tilde{f}(|\alpha_m|)).$$

Sia adesso n un qualunque intero positivo e consideriamo il polinomio $g(x) = x^{n-1} f(x)^n$. Abbiamo dimostrato che per ogni $z \in \mathbb{C}$ vale

$$|I^n_{f(x)}(z)| = |I^1_{g(x)}(z)| \leq |z| e^{|z|} \tilde{g}(|z|),$$

e siccome

$$\tilde{g}(|z|) \leq |z|^{n-1} \tilde{f}(|z|)^n, \qquad |z| e^{|z|} \tilde{g}(|z|) \leq |z|^n e^{|z|} \tilde{f}(|z|)^n,$$

per ogni radice α_i si ha

$$|I^n_{f(x)}(\alpha_i)| \leq |\alpha_i|^n \tilde{f}(|\alpha_i|)^n e^{|\alpha_i|} \leq e^{|\alpha_i|} C^n.$$

□

LEMMA 17.5.3. *Siano $f(x) \in \mathbb{Z}[x]$ con $f(0) \neq 0$ e p un numero primo:*

(1) *se $p > |f(0)|$, allora $H^p_{f(x)}(0)$ è un intero divisibile per $(p-1)!$ ma non per $p!$;*
(2) *se $m \in \mathbb{Z}$ e $f(m) = 0$, allora $H^p_{f(x)}(m)$ è un intero divisibile per $p!$.*

DIMOSTRAZIONE. Basta osservare che, scrivendo $g(x) = x^{p-1} f(x)^p$ si ha

$$g(x) = f(0)^p x^{p-1} + a_p x^p + a_{p+1} x^{p+1} + \cdots \qquad a_i \in \mathbb{Z}$$

e quindi

$$\sum_{h \geq 0} g^{(h)}(0) = f(0)^p (p-1)! + \sum_{h \geq p} h! a_h.$$

Se $m \in \mathbb{Z}$ e $f(m) = 0$ il polinomio $(x-m)$ divide $f(x)$, dunque $(x-m)^p$ divide $g(x)$ e $g^{(h)}(m) = 0$ per ogni $h < p$. Basta adesso osservare che, siccome $g(x) \in \mathbb{Z}[x]$, il polinomio $g^{(h)}(x)$ è divisibile per $h!$, per ogni $h > 0$.

□

TEOREMA 17.5.4. *Il numero e è trascendente.*

DIMOSTRAZIONE. Supponiamo per assurdo che e sia algebrico di grado m, allora si ha una relazione

$$a_1 e + a_2 e^2 + \cdots + a_m e^m = q, \qquad q, a_1, \ldots, a_m \in \mathbb{Z}, \quad q \neq 0.$$

Il numero q è diverso da 0 perché altrimenti, dividendo per e troveremo che e è algebrico di grado $< m$. Consideriamo il polinomio $f(x) = (x-1)(x-2)\cdots(x-m)$ e la successione di numeri complessi

$$J^n = a_1 I^n_{f(x)}(1) + \cdots + a_m I^n_{f(x)}(m), \qquad n > 0.$$

Siccome $|f(0)| = m!$, $f(1) = f(2) = \cdots = f(m) = 0$, per il Lemma 17.5.2 esiste una costante C tale che

$$|J^n| \leq |a_1| e^1 C^n + \cdots + |a_m| e^m C^n = (|a_1| e^1 + \cdots + |a_m| e^m) C^n.$$

Consideriamo adesso un numero primo $p > \max(q, m!)$, si ha:

$$\begin{aligned} J^p &= a_1 I^p_{f(x)}(1) + \cdots + a_m I^p_{f(x)}(m) \\ &= \sum_{i=1}^m a_i e^i H^p_{f(x)}(0) - \sum_{i=1}^m a_i H^p_{f(x)}(i) \\ &= q H^p_{f(x)}(0) - \sum_{i=1}^m a_i H^p_{f(x)}(i). \end{aligned}$$

Il Lemma 17.5.3 implica che $qH^p_{f(x)}(0)$ è un numero intero divisibile per $(p-1)!$ ma non per $p!$, mentre $H^p_{f(x)}(i)$ è divisibile per $p!$ per ogni $i = 1, \ldots, m$. Se ne deduce che J^p è un intero non nullo divisibile per $(p-1)!$ ed in particolare $|J^p| \geq (p-1)!$, in contraddizione con la stima

$$|J^p| \leq (|a_1| e^1 + \cdots + |a_m| e^m) C^p$$

per p molto grande. □

17.6. Polinomi simmetrici

Indichiamo con $\mathbb{Z}[x_1, \ldots, x_n]$ l'anello dei polinomi in $x_1, \ldots, x_n$ a coefficienti interi. Per definizione ogni polinomio è una somma finita di monomi, ossia di espressioni del tipo

$$a x_1^{i_1} \cdots x_n^{i_n}, \qquad a \in \mathbb{Z},\ a \neq 0, \quad i_1, \ldots, i_n \geq 0.$$

Di un tale monomio chiameremo **grado** il numero $i_1 + \cdots + i_n$, e chiameremo **peso** il numero $i_1 + 2i_2 + \cdots + n i_n$. Il grado di un polinomio è il massimo grado dei suoi monomi; similmente, il peso di un polinomio è il massimo peso dei suoi monomi. A titolo di esempio, il polinomio $x_1 x_2^4 + x_3^4 \in \mathbb{Z}[x_1, x_2, x_3]$ ha grado 5 e peso 12.

Un polinomio in cui tutti i monomi hanno lo steso grado si dice **omogeneo**; un polinomio in cui tutti i monomi hanno lo steso peso si dice **isobaro**.

È del tutto chiaro che ogni polinomio si scrive i maniera unica come come somma di polinomi isobari, che il peso della somma è minore od uguale al massimo dei pesi, mentre il peso del prodotto è la somma dei pesi.

DEFINIZIONE 17.6.1. Un polinomio $p \in \mathbb{Z}[x_1, \ldots, x_n]$ si dice **simmetrico** se è invariante per permutazione degli indici, ossia se

$$p(x_1, \ldots, x_n) = p(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$$

per ogni permutazione $\sigma \colon \{1, \ldots, n\} \to \{1, \ldots, n\}$.

Ad esempio, in $\mathbb{Z}[x_1, x_2]$ sono simmetrici i polinomi

$$x_1 + x_2, \quad x_1 x_2, \quad x_1^2 + x_2^2 = (x_1 + x_2)^2 - 2x_1 x_2, \quad x_1^2 x_2^2, \quad x_1^3 x_2 + x_1 x_2^3,$$

mentre **non sono** simmetrici i polinomi

$$x_1, \quad x_2, \quad x_1 + 2x_2, \quad x_1^2 - x_2^2, \quad x_1 x_2^2\ .$$

È chiaro che un polinomio è simmetrico se e solo se tutte le sue componenti omogenee sono simmetriche.

Osservazione 17.6.2. Dato che somme e prodotti di polinomi simmetrici sono ancora simmetrici, se $p_1, \dots, p_m \in \mathbb{Z}[x_1, \dots, x_n]$ sono simmetrici, allora per ogni $q \in \mathbb{Z}[y_1, \dots, y_m]$ ne consegue che

$$q(p_1, \dots, p_m) \in \mathbb{Z}[x_1, \dots, x_n]$$

è simmetrico.

Definizione 17.6.3. Per ogni intero positivo n, le **funzioni simmetriche elementari** $\sigma_1, \dots, \sigma_n \in \mathbb{Z}[x_1, \dots, x_n]$ sono i polinomi simmetrici definiti dalla relazione

$$t^n + \sigma_1(x_1, \dots, x_n)t^{n-1} + \cdots + \sigma_n(x_1, \dots, x_n) = \prod_{i=1}^{n}(t + x_i).$$

In altre parole, i valori delle funzioni simmetriche elementari calcolate su di una n-upla di numeri complessi $(a_1, \dots, a_n)$ sono i coefficienti del polinomio monico di grado n che ha come radici $-a_1, \dots, -a_n$. Per $n = 2$ le funzioni simmetriche elementari sono $\sigma_1 = x_1 + x_2$ e $\sigma_2 = x_1x_2$, mentre per $n = 3$ si ha

$$\sigma_1 = x_1 + x_2 + x_3, \quad \sigma_2 = x_1x_2 + x_2x_3 + x_1x_3, \quad \sigma_3 = x_1x_2x_3.$$

In generale, per la funzione $\sigma_k \in \mathbb{Z}[x_1, \dots, x_n]$ si ha

$$\sigma_k = \sum_{1 \le i_1 < \cdots < i_k \le n} x_{i_1} \cdots x_{i_k}.$$

Notiamo in particolare che σ_k è somma di $\binom{n}{k}$ monomi, che è omogeneo di grado k e che $x_{n-k+1} \cdots x_{n-1}x_n$ è il monomio di peso più alto.

Conviene definire le funzioni simmetriche σ_k per ogni $k > 0$, ponendo per convenzione $\sigma_k(x_1, \dots, x_n) = 0$ per ogni $k > n$.

Teorema 17.6.4. *Ogni polinomio simmetrico a coefficienti interi si può esprimere come un polinomio a coefficienti interi nelle funzioni simmetriche elementari.*

In altri termini, un polinomio $p \in \mathbb{Z}[x_1, \dots, x_n]$ è simmetrico se e solo se esiste un polinomio $q \in \mathbb{Z}[y_1, \dots, y_n]$ tale che

$$p(x_1, \dots, x_n) = q(\sigma_1, \dots, \sigma_n).$$

Dimostrazione. Dimostriamo per induzione sul peso che ogni polinomio simmetrico $p \in \mathbb{Z}[x_1, \dots, x_n]$ si può scrivere come somma di addendi del tipo $a\sigma_1^{\alpha_1} \cdots \sigma_n^{\alpha_n}$, con $a \in \mathbb{Z}$ e $\alpha_1, \dots, \alpha_n \ge 0$ (con la convenzione che $\sigma_i^0 = 1$). Sia m il peso di p, se $m = 0$ allora p contiene solo il termine costante ed il risultato è banalmente verificato. Supponiamo $m > 0$ e sia $p = p_0 + \cdots + p_m$ la decomposizione in componenti isobare, con p_i di peso i e $p_m \neq 0$.

Dalla simmetria di p segue che ogni monomio di p_m è del tipo $ax_1^{i_1}x_2^{i_2} \cdots x_n^{i_n}$ con $a \in \mathbb{Z}$ e $i_1 \le i_2 \le \cdots \le i_n$: se per assurdo p_m contenesse un monomio $ax_1^{i_1}x_2^{i_2} \cdots x_n^{i_n}$ con $i_k > i_{k+1}$ per qualche indice k, allora la simmetria imporrebbe che p contiene anche il monomio $ax_1^{i_1} \cdots x_k^{i_{k+1}}x_{k+1}^{i_k} \cdots x_n^{i_n}$ che però ha peso strettamente maggiore di m, in contraddizione con il fatto che p_m è la componente isobara di peso massimo.

Quindi, se poniamo

$$y_i = x_{n-i+1}x_{n-i+2} \cdots x_n, \qquad i = 1, \dots, n,$$

si ha

$$ax_1^{i_1}x_2^{i_2} \cdots x_n^{i_n} = ay_n^{i_1}x_2^{i_2-i_1}x_3^{i_3-i_1} \cdots x_n^{i_n-i_1} = ay_n^{i_1}y_{n-1}^{i_2-i_1}x_3^{i_3-i_2} \cdots x_n^{i_n-i_2} = \cdots$$

che alla fine diventa

$$ax_1^{i_1}x_2^{i_2} \cdots x_n^{i_n} = ay_n^{b_n} \cdots y_1^{b_1}, \quad b_n = i_1,\ b_{n-1} = i_2 - i_1,\ \dots$$

Abbiamo quindi dimostrato che esiste un polinomio $q \in \mathbb{Z}[y_1, \dots, y_n]$ tale che $p_m(x_1, \dots, x_n) = q(y_1, \dots, y_n)$. Basta adesso osservare che $q(y_1, \dots, y_n)$ è la componente isobara di peso più alto di $q(\sigma_1, \dots, \sigma_n)$; dunque il polinomio $p(x_1, \dots, x_n) - q(\sigma_1, \dots, \sigma_n)$ è simmetrico di peso $< m$ e si può applicare l'ipotesi induttiva. □

Ad esempio, se indichiamo con $\psi_k \in \mathbb{Z}[x_1, \dots, x_n]$ la somma delle potenze k-esime

$$\psi_k(x_1, \dots, x_n) = x_1^k + \cdots + x_n^k\,,$$

il teorema ci dice che ogni ψ_k è un polinomio nelle funzioni simmetriche elementari:

$$\psi_0 = 1, \quad \psi_1 = \sigma_1, \quad \psi_2 = \sigma_1^2 - 2\sigma_2, \quad \psi_3 = \sigma_1^3 - 3\sigma_1\sigma_2 + 3\sigma_3,$$
$$\psi_4 = \sigma_1^4 - 4\sigma_1^2\sigma_2 + 4\sigma_1\sigma_3 + 2\sigma_2^2 - 4\sigma_4 \quad \text{eccetera.}$$

Come prima applicazione del teorema delle funzioni simmetriche diamo una diversa dimostrazione del Teorema 4.8.10, ossia che i numeri algebrici formano un sottocampo di $\mathbb{C}$. A tal fine basta dimostrare che somme e prodotti di numeri algebrici sono ancora numeri algebrici in quanto se $\alpha \in \mathbb{C}$ soddisfa una relazione

$$a_0 + a_1\alpha + \cdots + a_n\alpha^n = 0, \qquad a_i \in \mathbb{Z},$$

allora $a_0 - a_1(-\alpha) + \cdots + (-1)^n a_n(-\alpha)^n = 0$, mentre se $\alpha \neq 0$ si ha

$$a_0(\alpha^{-1})^n + a_1(\alpha^{-1})^{n-1} + \cdots + a_n = 0,$$

e questo prova che opposti ed inversi di numeri algebrici sono ancora algebrici.

Teorema 17.6.5. *Somme e prodotti di numeri algebrici sono ancora numeri algebrici.*

Dimostrazione. Siano β, γ due numeri algebrici, per il Lemma 17.0.1 possiamo estendere β e γ a due successioni $\beta = \alpha_1, \alpha_2, \dots, \alpha_s$, $\gamma = \alpha_{s+1}, \dots, \alpha_n$, tali che

$$\prod_{i=1}^{s}(t - \alpha_i) \in \mathbb{Q}[t], \qquad \prod_{i=s+1}^{n}(t - \alpha_i) \in \mathbb{Q}[t].$$

In particolare $n \geq 2$,

$$f(t) = \prod_{i=1}^{n}(t - \alpha_i) \in \mathbb{Q}[t],$$

e le funzioni simmetriche elementari $\sigma_i(\alpha_1, \dots, \alpha_n)$ sono numeri razionali in quanto coincidono, a meno del segno, con i coefficienti del polinomio $f(t)$. Ogni coefficiente dei polinomi

$$g(t) = \prod_{i \neq j}(t - \alpha_i - \alpha_j), \qquad h(t) = \prod_{i \neq j}(t - \alpha_i\alpha_j)$$

è un polinomio simmetrico a coefficienti interi in $\alpha_1, \dots, \alpha_n$ ed è quindi esprimibile come un polinomio a coefficienti interi nelle funzioni $\sigma_i(\alpha_1, \dots, \alpha_n)$. In particolare i coefficienti di $g(t)$ e $h(t)$ sono numeri razionali. È poi chiaro che $\beta + \gamma = \alpha_1 + \alpha_{s+1}$ è una radice di $g(t)$ e $\beta\gamma = \alpha_1\alpha_{s+1}$ è una radice di $h(t)$. □

Esercizi.

864. Sia $A \in M_{n,n}(\mathbb{C})$ il cui polinomio caratteristico ha coefficienti interi. Dimostrare che lo stesso vale per tutte le potenze A^p, $p > 0$.

865. Indichiamo come al solito con $\sigma_1, \dots, \sigma_n \in \mathbb{Z}[x_1, \dots, x_n]$ le funzioni simmetriche elementari. Provare che se $q \in \mathbb{Z}[y_1, \dots, y_n]$ e $q(\sigma_1, \dots, \sigma_n) \equiv 0$, allora anche il polinomio q è identicamente nullo.

866 (♡). Sia $R \in \mathbb{Z}[a_0, \dots, a_n, b_0, \dots, b_m]$ il risultante universale, ossia il determinante della matrice di Sylvester (14.3), nella quale i coefficienti a_i, b_j sono pensati come indeterminate. Se le variabili a_i, b_i hanno grado 1 e peso i, per ogni i, provare che R è omogeneo di grado $n + m$ ed isobaro di peso nm.

867 (☕). Fissati due interi $0 < k \leq n$ indichiamo con I l'insieme di tutte le applicazioni iniettive $\sigma\colon \{1, \dots, k\} \to \{1, \dots, n\}$. Dimostrare che i coefficienti delle potenze di t del polinomio

$$p(t, x_1, \dots, x_n) = \prod_{\sigma \in I}(t^k + x_{\sigma(1)}t^{k-1} + \cdots + x_{\sigma(k)})$$

sono polinomi simmetrici in $x_1, \dots, x_n$. Usare questo fatto per dimostrare che:

(1) ogni polinomio monico in $\overline{\mathbb{Q}}[t]$ divide un polinomio monico in $\mathbb{Q}[t]$;
(2) il campo $\overline{\mathbb{Q}}$ è algebricamente chiuso.

17.7. La trascendenza di π

La dimostrazione della trascendenza di π è molto simile a quella della trascendenza di e; prima abbiamo però bisogno di estendere, in maniera opportuna, il risultato del Lemma 17.5.3 a polinomi con radici non necessariamente intere.

LEMMA 17.7.1. *Siano $\alpha_1, \dots, \alpha_m \in \mathbb{C}$ ed $a \neq 0$ un intero tali che il polinomio*

$$g(x) = a(x-\alpha_1)\cdots(x-\alpha_m) \in \mathbb{Z}[x]$$

abbia coefficienti interi. Allora, per ogni polinomio $f(x) \in \mathbb{Z}[x]$ di grado $\leq n$ e per ogni intero $p \geq 0$ si ha:

$$\frac{a^{n-p}}{p!}\sum_{i=1}^{m} f^{(p)}(\alpha_i) \in \mathbb{Z}.$$

DIMOSTRAZIONE. I numeri complessi $a\alpha_1, \dots, a\alpha_n$ sono le radici, contate con molteplicità, del polinomio monico a coefficienti interi

$$(x-a\alpha_1)\cdots(x-a\alpha_m) = a^{m-1}g\left(\frac{x}{a}\right).$$

Dunque $\sigma_i(a\alpha_1, \dots, a\alpha_n) \in \mathbb{Z}$ per ogni i e, per il Teorema 17.6.4, per ogni $k \geq 0$ la somma

$$\sum_{i=1}^{m}(a\alpha_i)^k = \psi_k(a\alpha_1, \dots, a\alpha_m)$$

si può esprimere come un polinomio a coefficienti interi nelle funzioni simmetriche elementari $\sigma_i(a\alpha_1, \dots, a\alpha_n)$ e di conseguenza

$$\sum_{i=1}^{m}(a\alpha_i)^k \in \mathbb{Z}.$$

Per dimostrare il lemma basta considerare il caso $f(x) = x^q$ con $q \leq n$. Se $p > q$ vale $f^{(p)} = 0$, mentre se $p \leq q$ si ha

$$\frac{a^{n-p}}{p!}\sum_{i=1}^{m} f^{(p)}(\alpha_i) = a^{n-q}\binom{q}{p}\sum_{i=1}^{m}(a\alpha_i)^{q-p} \in \mathbb{Z}.$$

□

LEMMA 17.7.2. *Siano $\alpha_1, \dots, \alpha_m \in \mathbb{C}$ ed $a \in \mathbb{Z}$ tali che il polinomio*

$$f(x) = a(x-\alpha_1)\cdots(x-\alpha_m) \in \mathbb{Z}[x]$$

abbia coefficienti interi. Sia p un numero primo:

(1) *se $p > |a^m f(0)| > 0$, allora $H^p_{a^m f(x)}(0)$ è un intero divisibile per $(p-1)!$ ma non per $p!$;*
(2) *la sommatoria $\sum_{i=1}^m H^p_{a^m f(x)}(\alpha_i)$ è un intero divisibile per $p!$.*

DIMOSTRAZIONE. Il primo punto è gia stato dimostrato nel Lemma 17.5.3. Indicando con

$$g(x) = x^{p-1}(a^m f(x))^p = a^{pm}q(x), \qquad q(x) = a^p x^{p-1}(x-\alpha_1)^p\cdots(x-\alpha_m)^p,$$

si ha per definizione

$$\sum_{i=1}^{m} H^p_{a^m f(x)}(\alpha_i) = \sum_{i=1}^{m}\sum_{h=0}^{\infty} g^{(h)}(\alpha_i) = \sum_{h=0}^{\infty} a^{pm}\sum_{i=1}^{m} q^{(h)}(\alpha_i).$$

Basta adesso osservare che ogni α_i è radice di $q(x)$ di molteplicità almeno p e quindi $q^{(h)}(\alpha_i) = 0$ per ogni $h < p$. Se $h \geq p$, allora il grado di $q(x)$ è minore di $pm + h$ e per il Lemma 17.7.1 la sommatoria $a^{pm}\sum_{i=1}^m q^{(h)}(\alpha_i)$ è un intero divisibile per $h!$. □

TEOREMA 17.7.3. *Il numero π è trascendente.*

DIMOSTRAZIONE. Abbiamo già osservato che l'unità immaginaria i è un numero algebrico. Supponiamo per assurdo che π sia algebrico, allora anche $i\pi$ è algebrico e possiamo estenderlo ad una successione $i\pi = \theta_1, \dots, \theta_d$ di numeri complessi tali che

$$\prod_{i=1}^{d}(x-\theta_i) \in \mathbb{Q}[t]$$

o, equivalentemente, tali che $\sigma_i(\theta_1,\dots,\theta_d)\in\mathbb{Q}$ per ogni $i=1,\dots,d$. Consideriamo adesso i 2^d numeri complessi, contati con molteplicità,

$$a_1\theta_1+\cdots+a_d\theta_d,\qquad a_i\in\{0,1\},$$

che a loro volta sono radici del polinomio di grado 2^d

$$q(x)=\prod_{a_1,\dots,a_d=0,1}(x-a_1\theta_1+\cdots+a_d\theta_d).$$

Per il teorema delle funzioni simmetriche, ogni coefficiente di $q(x)$ è un polinomio a coefficienti interi nelle funzioni simmetriche $\sigma_i(\theta_1,\dots,\theta_d)$; di conseguenza $q(x)$ ha coefficienti razionali e fissiamo un intero $a>0$ tale che $aq(x)\in\mathbb{Z}[x]$.

Indichiamo con $\alpha_1,\dots,\alpha_n$, $1\le n\le 2^d$, i numeri $a_1\theta_1+\cdots+a_d\theta_d$ diversi da 0 e $q=2^d-n$; la combinazione con $a_1=\cdots=a_d=0$ ci mostra che $q\ge 1$. Siccome $e^0+e^{\theta_1}=1+e^{i\pi}=0$, la relazione

$$(e^0+e^{\theta_1})\cdots(e^0+e^{\theta_d})=0$$

diventa

$$q+e^{\alpha_1}+\cdots+e^{\alpha_n}=0.$$

Consideriamo adesso il polinomio

$$f(x)=a\frac{q(x)}{x^q}=a(x-\alpha_1)\cdots(x-\alpha_n)\in\mathbb{Z}[x].$$

e, per ogni primo p sufficientemente grande consideriamo il numero complesso

$$J^p=I^p_{a^nf(x)}(\alpha_1)+\cdots+I^p_{a^nf(x)}(\alpha_n).$$

Per il Lemma 17.5.2 esiste un numero reale $C>0$, indipendente da p, tale che

$$|J^p|\le(e^{|\alpha_1|}+\cdots+e^{|\alpha_n|})C^p. \tag{17.4}$$

D'altra parte possiamo scrivere

$$-J^p=-\sum_{i=1}^n e^{\alpha_i}H^p_{a^nf(x)}(0)+\sum_{i=1}^n H^p_{a^nf(x)}(\alpha_i)=qH^p_{a^nf(x)}(0)+\sum_{i=1}^n H^p_{a^nf(x)}(\alpha_i).$$

Per il Lemma 17.7.2 il numero $\sum_{i=1}^n H^p_{a^nf(x)}(\alpha_i)$ è un intero divisibile per $p!$, mentre per $p>>0$ il numero $qH^p_{a^nf(x)}(0)$ è un intero divisibile per $(p-1)!$ ma non per $p!$. In conclusione, per ogni primo p sufficientemente grande, il numero J^p è un intero divisibile per $(p-1)!$ ma non per $p!$; in particolare $|J^p|\ge(p-1)!$ in contraddizione con la stima (17.4). □

Osservazione 17.7.4. La trascendenza di e e la trascendenza di π sono casi particolari del seguente teorema la cui dimostrazione, sebbene simile alle precedenti (ma decisamente meno elementare), è omessa.

Teorema 17.7.5 (Lindemann–Weierstrass 1885). *Siano $\alpha_1,\dots,\alpha_n$ numeri algebrici distinti. Allora i numeri $e^{\alpha_1},\dots,e^{\alpha_n}$ sono linearmente indipendenti su $\overline{\mathbb{Q}}$.*

Come immediata conseguenza si ha la trascendenza di $\log(n+1),\cos(n),\sin(n)$ per ogni intero $n>0$.

Esercizi.

868. Dimostrare che $\log_{10}(\pi)$ è irrazionale.

869. Dedurre dal teorema di Lindemann–Weierstrass che per ogni successione periodica c_n di numeri algebrici (ossia esiste $k>0$ tale $c_{n+k}=c_n$) e non tutti nulli, il numero

$$\sum_{n=0}^\infty\frac{c_n}{n!}$$

è trascendente (sugg.: matrice di Vandermonde delle radici k-esime di 1).

CAPITOLO 18

Note, commenti, curiosità e riferimenti bibliografici

Il test preliminare di autovalutazione proposto a pagina ii faceva parte, assieme ad altri questionari, di una ricerca condotta su un campione di 40.000 studenti di 9 paesi anglofoni, da un team di ricercatori inglesi e australiani, sulle condizioni socio-economiche dei *bullshiters*, ossia degli individui che dichiarano la propria competenza in settori dove invece ne hanno ben poca. I concetti di numero proprio, cambio di scala congiuntivo e frazione dichiarativa non esistono. La divisione del campione tra bullshiters e non, è stata fatta in base al punteggio attribuito a questi tre concetti farlocchi.

Capitolo 1.

:-) Il problema 1 è stato considerato tra le cause di ricorso nei confronti di un concorso a dirigente pubblico del comune di Torino. Tra i motivi dei ricorsi presentati figurava anche il fatto che la domanda delle mucche e delle galline viene considerata errata, non solo dai ricorrenti e relativi avvocati, ma anche da qualche giornalista: in un articolo pubblicato su "La Stampa" l'11 agosto 2020 il/la giornalista afferma in proposito: "Un quesito che sembra non dare gli elementi necessari per rispondere. Impossibile, quindi, non sbagliare".

In effetti, nel quesito non viene detto che le mucche hanno 4 zampe e le galline 2, elementi necessari per risolvere il problema. D'altra parte il quesito era a risposta multipla con le tre possibili risposte (32,28), (26,34) e (30,30). Quindi per rispondere correttamente bastava sapere che mucche e galline hanno un numero intero di zampe non inferiore a 2.

:-) Funtori (semplici e derivati), Hom, Tor ed Ext non sono nomi di fantasia ma enti matematici realmente esistenti ed ampiamente studiati: fanno tutti parte di quella parte della matematica chiamata teoria delle categorie. Giova osservare che in matematica, con il termine *teoria* si intende un insieme di risultati tra loro omogenei per argomenti, metodi di indagine e presentazione: oltre alla suddetta teoria delle categorie sentirete sicuramente parlare di teoria di Galois, di teoria dei numeri, di teoria degli invarianti, di teoria delle rappresentazioni, di teoria dei giochi eccetera.

Capitolo 2.

:-) il termine "Algebretta" è ripreso dall'omonimo testo di Benedetto Scimemi, molto in voga nei corsi di laurea in matematica nel periodo dal 1980 al 2000.

:-) L'uso delle maiuscole a doppio strato $\mathbb{A}, \mathbb{B}, \mathbb{C}, \mathbb{D}, \mathbb{E}, \mathbb{F}, \mathbb{G}, \mathbb{H}, \mathbb{K}, \mathbb{L}, \mathbb{M}, \mathbb{N}, \mathbb{O}, \mathbb{P}, \mathbb{Q}, \mathbb{R}, \mathbb{S}, \mathbb{T}, \mathbb{U}, \mathbb{V}, \mathbb{Z}$ ha preso piede nei lavori a stampa negli anni intorno al 1965. Precedentemente, per indicare i sistemi numerici veniva usato il grassetto maiuscolo ed il doppio strato era confinato, nella versione semplificata $\mathbb{A}, \mathbb{B}, \mathbb{C}, \mathbb{D}, \mathbb{E}, \mathbb{F}, \mathbb{G}, \mathbb{H}, \mathbb{K}, \mathbb{L}, \mathbb{M}, \mathbb{N}, \mathbb{O}, \mathbb{P}, \mathbb{Q}, \mathbb{R}, \mathbb{S}, \mathbb{T}, \mathbb{U}, \mathbb{V}, \mathbb{Z}$, alle situazioni dove risultava difficile differenziare il grassetto dal testo normale, come ad esempio nella scrittura a gesso su lavagna; in inglese il doppio strato viene chiamato blackboard.

:-) Sull'assioma della scelta possono sorgere alcune perplessità di tipo logico e fondazionale: a rigore, fare *infinite scelte* potrebbe non essere consentito. Fortunatamente si può dimostrare che tale assioma non è contraddittorio con gli altri assiomi base della matematica e quindi non si incorre in alcun errore logico nell'assumerlo come valido. Riconosco di essere piuttosto vago in questo passaggio, ma una descrizione precisa ed accurata richiederebbe una trattazione tutt'altro che banale di logica matematica e teoria assiomatica degli insiemi, argomenti che vanno molto al di là dei nostri obiettivi. Il lettore interessato può approfondire l'argomento sul libro di G. Tourlakis, *Lectures in logic and set theory, Volume 2*, Cambridge University Press (2003).

:-) L'Esercizio 102 che abbiamo sbrigativamente definito esercizio di astrazione, un filosofo del linguaggio lo avrebbe catalogato come pura *competenza inferenziale*. Noi possediamo

(normalmente) competenze inferenziali (relative alle relazioni tra oggetti) e referenziali (relative al significato degli oggetti nel mondo reale) in misura variabile e le due competenze sono cooperanti ma indipendenti. Esistono casi documentati di pazienti cerebrolesi che possedevano un solo tipo di competenza: in uno di questi, il soggetto in questione era in grado di descrivere perfettamente a parole cosa è un'anatra ma non era assolutamente in grado di riconoscerla se posto di fronte al simpatico pennuto. Per maggiori informazioni si può leggere il testo *La competenza lessicale* di Diego Marconi. È possibile che molti matematici possiedano un rapporto tra competenze inferenziali e referenziali più alto rispetto alla media e questo potrebbe spiegare perché sono spesso visti come gente strana.

Capitolo 6.

:-) L'Esercizio 338 deve essere considerato un problema esclusivamente fine a se stesso. L'autore ignora se esistono matrici invertibili in senso MTS: già per le matrici 3×3 a coefficienti complessi il problema si presenta tutt'altro che banale. Eventuali matrici invertibili in senso MTS (ammesso che esistano) non avrebbero alcuna utilità e non sono minimamente considerate in letteratura.

:-) L'Esercizio 341 mi è stato suggerito da Manfred Lehn, che ne fa largo uso nel corso di laurea in ingegneria all'Università di Magonza: a quanto racconta egli stesso, solo una piccola percentuale di studenti fornisce la risposta corretta.

Capitolo 9.

:-) I termini *autovalore* ed *autovettore* sono traduzioni, non del tutto precise dal punto di vista linguistico, dei corrispettivi tedeschi *eigenwert* ed *eigenvektor*: in tedesco "eigen" significa "proprio, caratteristico, peculiare", e non ha il significato che in italiano ha il termine "auto". I francofoni parlano di valori e vettori propri; gli anglofoni parlano di valori e vettori caratteristici, spesso chiamati direttamente eigenvalues e eigenvectors.

:-) La dimostrazione del teorema fondamentale dell'algebra è basata sulle idee dell'articolo di H. Derksen, *The fundamental theorem of algebra and linear algebra*, pubblicato sull'American Mathematical Monthly nel 2003.

Capitolo 13.

:-) Il Teorema 13.4.9 può essere visto come un banale corollario del *teorema di Zermelo*, secondo il quale ogni insieme possiede buoni ordinamenti. Un ordinamento di dice buono se ogni sottoinsieme non vuoto possiede un minimo elemento: ad esempio, l'ordinamento usuale su $\mathbb{N}$ è un buon ordinamento. Ogni buon ordinamento è anche un ordinamento totale, infatti per ogni coppia x, y di elementi si ha $\min(x, y) = x$ oppure $\min(x, y) = y$: nel primo caso $x \leq y$ e nel secondo caso $y \leq x$. Gli ordinamenti usuali sui numeri razionali e sui numeri reali sono totali ma non sono buoni ordinamenti. Per la dimostrazione del teorema di Zermelo come conseguenza del lemma di Zorn rimandiamo il lettore interessato al libro di M. Manetti, *Topologia*.

Capitolo 17.

:-) Negli ultimi tempi, a seguito degli eccessi delle politiche "publish or perish" nonché del desiderio di protagonismo di scienziati dilettanti, sono nate molte sedicenti riviste dove chiunque, pagando cifre variabili mediamente dai 100 ai 600 dollari, può pubblicare le proprie ricerche per poi fregiarsi verso il mondo (non accademico) di tale presunto marchio di qualità.

Tra queste merita sicuramente una menzione lo IOSR journal of mathematics, rivista particolarmente attiva nello spamming, che nel 2012 ha pubblicato un articolo del signor Laxman S. Gogawale in cui si "dimostra" che $\pi = 17 - 8\sqrt{3}$. Per verificare l'assurdità di tale risultato non serve nemmeno evocare la trascendenza di π: basta calcolare il valore approssimato alla terza cifra decimale di $17 - 8\sqrt{3} = 3,1435\ldots$, ancora più distante da π di $22/7 = 3,1428\ldots$.

Gogawale giustifica tale discrepanza dicendo di aver calcolato il *nuovo* valore di π, da contrapporre al "vecchio". Per fortuna, in matematica il desiderio di cambiamento, almeno per il momento, non supera le barriere della decenza anche se, in un mondo dominato dai social network dove la volontà delle masse soverchia il parere delle comunità di esperti, non mi sento di escludere a priori che si possano verificare in futuro episodi simili al *progetto di legge dell'Indiana sul pi greco* (cercare su Wikipedia).

:-) Chi è interessato ad approfondire lo studio dei numeri trascendenti segnaliamo il libro di A. Baker, *Trascendental number theory*, ed il libro di M. Ram Murty e Perusottam Rath, *Trascendental numbers*, nei quali si può trovare, tra le altre cose, la dimostrazione del Teorema 17.7.5.

CAPITOLO 19

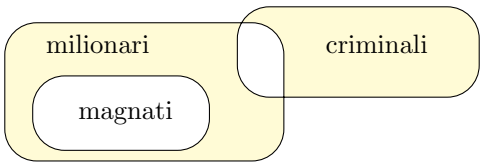
Soluzioni e suggerimenti di alcuni esercizi

In questo capitolo diamo alcune tracce, idee, suggerimenti ed aiuti utili allo svolgimento degli esercizi contrassegnati con il simbolo ♡. In qualche caso vengono fornite soluzioni complete.

6. La probabilità che si verifichino due eventi A e B è sempre minore od uguale alla probabilità che si verifichi uno solo di essi. Dunque la risposta meno probabile è certamente la terza.

18. Ad esempio: $(-1,1)$ pollo sconigliato, $(-1,0)$ anticoniglio, $(0,2)$ bipollo, $(1,-2)$ coniglio bispollato, $(3,0)$ triconiglio, $(5,-6)$ pentaconiglio esaspollato.

22. La conclusione è falsa.



23.

$$|A\cap B| = |A| - |A-B|, \qquad |A\cup B| = |A| + |B| - |A\cap B|,$$
$$|A\cup B\cup C| = |A| + |B| + |C| - |A\cap B| - |A\cap C| - |B\cap C| + |A\cap B\cap C|\,.$$

26. Sottraendo a ciascuna delle prima tre equazioni l'equazione successiva il sistema diventa

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

che si risolve rapidamente con il metodo di sostituzione.

30. Risposta: molto vicino alla madre.

31. Il test delle 4 carte è stato inventato dallo psicologo Peter Wilson nel 1966 e viene generalmente sbagliato dalla stragrande maggioranza dei soggetti a cui viene sottoposto, che oltre alla prima, indicano la terza carta anziché la quarta.

51. Scegliamo due interi positivi n, q sufficientemente grandi e tali che

$$n \geq \frac{3a}{a-c}, \qquad q^2 a \geq n^2.$$

Sia poi p l'unico intero positivo tale che

$$p^2 \leq q^2 a < (p+1)^2$$

e definiamo $b = p/q$. Chiaramente $p \geq n \geq 1$, $b^2 \leq a$ e

$$a - b^2 = \frac{q^2 a - p^2}{q^2} < \frac{(p+1)^2 - p^2}{q^2} = \frac{2p+1}{q^2} \leq \frac{3p}{q^2}.$$

D'altra parte, dividendo $p^2 \leq q^2 a$ per pq^2 si ottiene

$$\frac{3p}{q^2} \leq \frac{3a}{p} \leq \frac{3a}{n} \leq a - c \quad \Rightarrow \quad a - b^2 < a - c.$$

72. Suggerimento: induzione su n, distinguendo due casi. Se ogni sottoinsieme di k ragazzi, con $1 \le k < n$, conosce cumulativamente almeno $k+1$ ragazze, fate scegliere la fidanzata al primo ragazzo e usate l'ipotesi induttiva per far fidanzare i rimanenti $n-1$.

Se esiste un sottoinsieme di k ragazzi, con $1 \le k < n$, che conosce cumulativamente esattamente k, applicare l'ipotesi induttiva a tale sottoinsieme. Successivamente applicare l'ipotesi induttiva all'insieme dei rimanenti $n-k$ ragazzi con le scelte ristrette alle ragazze non ancora fidanzate.

80. Morto un Papa se ne fa un altro? Tale regola non è applicabile se muoiono anche tutti i cardinali, ad esempio se la terra esplode o viene inghiottita dal sole.

114. Per induzione su n possiamo supporre che se $n > 1$ allora esiste $a \in \mathbb{N}$ tale che

$$n-1 \le a^2 \le n-1+2\sqrt{n-2} < n+2\sqrt{n-1}\,.$$

Se $a^2 \ge n$ abbiamo finito, mentre se $a^2 = n-1$ allora

$$(a+1)^2 = a^2+2a+1 = n+2a = n+2\sqrt{n-1}\,.$$

Nota: se $n^2 < x < n^2+1$ allora non esistono quadrati di interi compresi tra x e $x+2\sqrt{x-1}$. Tuttavia, dato un qualsiasi numero reale $x \ge 1/2$ esiste un intero a tale che

$$x - \frac{\sqrt{4x-1}}{2} \le a^2 \le x + \frac{\sqrt{4x-1}}{2}\,.$$

Indichiamo con b la parte intera di $\sqrt{x}$. Si ha $b^2 \le x \le (b+1)^2$ e se

$$b^2 \le x - \frac{\sqrt{4x-1}}{2} = \left(\frac{\sqrt{4x-1}}{2} - \frac{1}{2}\right)^2,$$

allora $b \le \frac{\sqrt{4x-1}}{2} - \frac{1}{2}$ e quindi

$$(b+1)^2 = b^2+2b+1 \le x - \frac{\sqrt{4x-1}}{2} + 2\left(\frac{\sqrt{4x-1}}{2} - \frac{1}{2}\right) + 1 = x + \frac{\sqrt{4x-1}}{2}\,.$$

In definitiva, almeno uno tra gli interi $a = b$ e $a = b+1$ ha le proprietà richieste.

120. Siccome $\alpha, \beta > 0$, supponendo per fissare le idee che $\alpha \le \beta$, si deve avere necessariamente $1 < \alpha < 2 < \beta$. In particolare, per ogni intero n vale

$$\lfloor n\alpha \rfloor < \lfloor (n+1)\alpha \rfloor, \qquad \lfloor n\beta \rfloor < \lfloor (n+1)\beta \rfloor.$$

Per ogni intero $N > 0$ indichiamo con $a(N)$ (risp.: $b(n)$) il numero di interi positivi n tali che $1 \le \lfloor n\alpha \rfloor \le N$ (risp.: $1 \le \lfloor n\beta \rfloor \le N$). I numeri $a(N)$ e $b(N)$ si calcolano facilmente, infatti

$$1 \le \lfloor n\alpha \rfloor \le N \iff 1 \le n\alpha < N+1 \iff 1 \le n < \frac{N+1}{\alpha} \iff 1 \le n \le \left\lfloor \frac{N+1}{\alpha} \right\rfloor$$

e quindi $a(N) = \left\lfloor \dfrac{N+1}{\alpha} \right\rfloor$; alla stessa maniera si prova che $b(N) = \left\lfloor \dfrac{N+1}{\beta} \right\rfloor$. Per ipotesi α, β sono irrazionali e si hanno le disuguaglianze

$$\frac{N+1}{\alpha} - 1 < a(N) < \frac{N+1}{\alpha}, \qquad \frac{N+1}{\beta} - 1 < b(N) < \frac{N+1}{\beta},$$

$$N-1 = \frac{N+1}{\alpha} + \frac{N+1}{\beta} - 2 < a(N) + b(N) < \frac{N+1}{\alpha} + \frac{N+1}{\beta} = N+1,$$

dalle quali si deduce $a(N) + b(N) = N$. Supponiamo per assurdo che $N = \lfloor n\alpha \rfloor \ne \lfloor m\beta \rfloor$ per opportuni n, m. Allora si avrebbe $a(N) > a(N-1)$, $b(N) > b(N-1)$ in contraddizione con il fatto che $a(N) + b(N) = N$, $a(N-1) + b(N-1) = N-1$. Similmente, se $N \ne \lfloor n\alpha \rfloor$ e $N \ne \lfloor m\beta \rfloor$ per ogni n, m si avrebbe $a(N) = a(N-1)$ e $b(N) = b(N-1)$.

123. Bisogna dire se esistono o meno due numeri razionali a, b tali che

$$(a+b\sqrt{2})^2 = (a^2+2b^2) + 2ab\sqrt{2} = 3 + 2\sqrt{2},$$

e la risposta è chiaramente affermativa: $a = b = 1$.

135. Vogliamo determinare tutti i numeri complessi $z \in \mathbb{C}$ tali che $\overline{z}^2 + z = 0$. Calcoliamo prima le soluzioni reali e poi quelle con parte immaginaria diversa da 0. Se $z = a \in \mathbb{R}$ è un numero reale, allora vale $\overline{z}^2 + z = a^2 + a = 0$ se e solo se $a = 0, -1$.

Supponiamo adesso $z = a + ib$ con $b \neq 0$, allora si ha

$$\overline{z}^2 + z = (a - ib)^2 + (a + ib) = (a^2 - b^2 + a) + i(b - 2ab) = 0$$

che equivale al sistema di due equazioni

$$a^2 - b^2 + a = 0, \qquad b - 2ab = 0.$$

Siccome $b \neq 0$ dalla seconda equazione si ricava $a = 1/2$ e dalla prima $b^2 = 3/4$. In conclusione esistono esattamente 4 numeri complessi z tali che $\overline{z}^2 + z = 0$, e tali numeri sono: 0, -1, $\frac{1+i\sqrt{3}}{2}$, $\frac{1-i\sqrt{3}}{2}$.

149. Supponiamo per assurdo $|z| \geq 2\sqrt[k]{|a_k|}$ per ogni k. Allora $|z^n| = |z|^k |z|^{n-k} \geq 2^k |a_k| |z|^{n-k} = 2^k |a_k z^{n-k}|$. Dunque:

$$|z^n| > \sum_{k=1}^{n} 2^{-k} |z^n| \geq \sum_{k=1}^{n} |a_k z^{n-k}| \geq |a_1 z^{n-1} + a_2 z^{n-2} + \cdots + a_n|.$$

176. Traccia: supponiamo per assurdo $n > 0$ e che i numeri a_i non siano tutti nulli, allora per induzione su n possiamo supporre che $b = a_1 \cdots a_n \neq 0$. Sia $0 \leq h \leq k$ il più piccolo intero tale che $\sum_{i=1}^{n} a_i^{h+j} = 0$ per ogni $j = 1, \ldots, n$ e consideriamo il polinomio

$$p(t) = t^h (t - a_1) \cdots (t - a_n) \in \mathbb{C}[t]\,.$$

Allora

$$0 = \sum_{i=1}^{n} p(a_i) = b \sum_{i=1}^{n} a_i^h\,.$$

179. Suggerimento: non è restrittivo supporre $0 < h < n$; si consideri il polinomio, di grado $< n$,

$$p(x) = n(x^n - 1) + \sum_{i=0}^{n-1} (1 - \xi_i)^n - \sum_{i=0}^{n-1} (x - \xi_i)^n$$

e si dimostri, usando l'Esercizio 157, che $p(\xi_j) = 0$ per ogni $j = 0, \ldots, n-1$.

180. Il problema è chiaramente malposto in quanto non viene indicato il campo su cui si effettua la divisione. Ad esempio 1 diviso 5, ossia la soluzione dell'equazione $5x = 1$, è uguale a 2 in $\mathbb{F}_3$, è uguale a 3 in $\mathbb{F}_7$ ed è uguale ad 8 in $\mathbb{F}_{13}$.

184. Suggerimento: sia $\mathbb{K}$ un campo di caratteristica p, allora vale $(1 + x)^{pa} = (1 + x^p)^a$ in $\mathbb{K}[x]$.

207. Per ogni $a \in \mathbb{K}$ definiamo $H_a = \{x \in \mathbb{K}^n \mid x_1 = a x_2\}$, $K_a = \{x \in \mathbb{K}^n \mid x_2 = a x_1\}$. Allora si ha

$$\mathbb{K}^n = H_0 \cup_{a \in \mathbb{K}} K_a = K_0 \cup_{a \in \mathbb{K}} H_a\,.$$

208. Il problema è del tutto equivalente a chiedersi se l'equazione $xu + yv + zw = e_1$ ammette soluzioni $x, y, z \in \mathbb{R}$, ossia se il sistema lineare

$$\begin{cases} x + 3y + 5z = 1 \\ 4y + 8z = 0 \\ x + 2y + 3z = 0 \\ 2x + y = 0 \end{cases}$$

ammette soluzioni reali. Sottraendo la terza equazione alla prima e dividendo la seconda per 4 troviamo il sistema

$$\begin{cases} y + 2z = 1 \\ y + 2z = 0 \\ x + 2y + 3z = 0 \\ 2x + y = 0 \end{cases}$$

che è chiaramente inconsistente. Quindi e_1 non è combinazione lineare di u, v, w.

221. Traccia di soluzione: mostriamo che (1) implica (2). Siano $a_1, \dots, a_p \in \mathbb{K}$ non tutti nulli e tali che $\sum_{j=1}^{p} a_j(v_j - v_0) = 0$. Se poniamo $a_0 = -\sum_{j=1}^{p} a_j$ allora si ha $\sum_{j=0}^{p} a_j f(v_j) = 0$.

Mostriamo adesso che (2) implica (1). Siano $a_0, \dots, a_p \in \mathbb{K}$ non tutti nulli e tali che $\sum a_j f(v_j) = 0$. Siccome

$$\sum a_j f(v_j) = \begin{pmatrix} \sum a_j \\ \sum a_j v_j \end{pmatrix}$$

si ha che $\sum a_j = 0$ e quindi

$$a_0 = -\sum_{j=1}^{p} a_j$$

e di conseguenza $a_j \neq 0$ per qualche $j > 0$. Ma allora

$$\sum_{j \neq i} a_j(v_j - v_i) = \sum_{j \neq i} a_j v_j - (\sum_{j \neq i} a_j) v_i = \sum_{j=0}^{p} a_j v_j = 0.$$

222. Basta considerare un qualsiasi vettore $w \notin \bigcup_{i,j} \mathrm{Span}(v_i, v_j)$ e prendere $u = w - v_1$.

223. Suggerimento: induzione su n; a meno di permutazione degli indici possiamo supporre $\lim_{x \to +\infty}(p_i(x) - p_j(x)) = -\infty$ per ogni $i < j$.

230. Suggerimento: siano $a_1, \dots, a_m \in \mathbb{K}$ tali che $\sum_i a_i v_i = 0$, pensiamo $\mathbb{K}$ come uno spazio vettoriale su F e sia $W \subset \mathbb{K}$ il sottospazio vettoriale generato da $a_1, \dots, a_m$, quindi di dimensione finita su F. Si prenda una base $b_1, \dots, b_r$ di W ....

260. Traccia: se $g \circ f \neq 0$ allora $f, g \neq 0$ ed in particolare $\mathrm{Ker}(f) \neq V$, $g(W) \neq 0$; se inoltre $f \circ g = 0$ allora $g(W) \subset \mathrm{Ker}(f)$. Se $0 \neq \mathrm{Ker}(f) \neq V$ si consideri una qualsiasi applicazione lineare $g\colon W \to \mathrm{Ker}(f)$ la cui restrizione a $f(V)$ è non nulla.

263. La restrizione di f al sottospazio $y = 0$ è chiaramente surgettiva. A maggior ragione f è surgettiva ed il suo rango è uguale a 2.

284. Una implicazione è chiara. Viceversa, sia $H \subset V$ un complementare di $\mathrm{Ker}(f) = \mathrm{Ker}(g)$ in V e sia $v_1, \dots, v_n$ una base di H. Basta dimostrare che esiste uno scalare $b \in \mathbb{K}$ tale che $bf(v_i) = g(v_i)$ per ogni $i = 1, \dots, n$. Per la condizione (2) esistono $n+1$ scalari $b_0, \dots, b_n$ tali che

$$g(v_1 + \cdots + v_n) = b_0 f(v_1 + \cdots + v_n), \qquad g(v_i) = b_i f(v_i) \quad \text{per ogni } i.$$

dal fatto che $f(v_1), \dots, f(v_n)$ sono linearmente indipendenti segue che $b_0 = b_i$ per ogni i.

300. Per definizione V_{ij} è il nucleo dell'applicazione lineare surgettiva

$$M_{4,4}(\mathbb{K}) \to \mathbb{K}, \qquad (a_{ij}) \mapsto \sum_{k=1}^{4} a_{ik} - \sum_{h=1}^{4} a_{hj},$$

e per il teorema del rango V_{ij} ha dimensione $16 - 1 = 15$. Indichiamo con V l'intersezione dei sottospazi V_{ij}. Per determinare la dimensione di V, osserviamo che ogni elemento di V dipende linearmente e biunivocamente (esercizio: perché?) dai 10 coefficienti a_{11}, a_{12}, a_{13}, a_{14}, a_{21}, a_{22} ,a_{23}, a_{31}, a_{32}, a_{33}. Più precisamente l'applicazione

$$V \to \mathbb{K}^{10}, \qquad (a_{ij}) \mapsto (a_{11}, a_{12}, a_{13}, a_{14}, a_{21}, a_{22}, a_{23}, a_{31}, a_{32}, a_{33})^T,$$

è lineare e bigettiva e quindi V ha dimensione 10.

301. La dimostrazione che U_{ij} è un sottospazio vettoriale di dimensione 15 è del tutto simile al caso V_{ij} trattato nell'Esercizio 300. Indichiamo con U l'intersezione dei sottospazi U_{ij} e osserviamo che ogni elemento di U dipende linearmente e biunivocamente dai 9 coefficienti a_{11}, a_{12}, a_{13}, a_{21}, a_{22} ,a_{23}, a_{31}, a_{32}, a_{33}, e quindi U ha dimensione 9. Infatti, se $d = \sum_j a_{1j}$ è la somma dei coefficienti della prima riga, allora per ogni i, j si ha

$$d = 2\sum_{h=1}^{4} a_{hj} = \sum_{k=1}^{4} a_{ik}\,.$$

Se $2 = 0$ in $\mathbb{K}$ allora $d = 0$; mentre se $2 \neq 0$, allora anche $4 = 2^2 \neq 0$, e quindi

$$4d = \sum_i \sum_{k=1}^{4} a_{ik} = \sum_j 2 \sum_{h=1}^{4} a_{hj} = 8d$$

da cui segue $8d - 4d = 0$ da cui $d = 0$, a differenza dell'Esercizio 300 dove d può assumere qualsiasi valore.

325. Per $n = 1$ non c'è nulla da dimostrare. Se $n > 1$ possiamo supporre per induzione che il risultato sia vero per A^{n-1}, ossia

$$A^{n-1} = \frac{1}{3^{n-1}} \begin{pmatrix} c & * \\ d\sqrt{2} & * \end{pmatrix},$$

con c, d interi non divisibili per 3 e $3|(c+d)$. Dalla relazione $A^n = AA^{n-1}$ segue pertanto che il primo vettore colonna di A^n è uguale a

$$\frac{1}{3^n} \begin{pmatrix} 1 & -2\sqrt{2} \\ 2\sqrt{2} & 1 \end{pmatrix} \begin{pmatrix} c \\ d\sqrt{2} \end{pmatrix} = \frac{1}{3^n} \begin{pmatrix} a \\ b\sqrt{2} \end{pmatrix}, \qquad \text{dove} \quad a = c - 4d,\ b = 2c + d.$$

La somma $a + b = 3(c - d)$ è divisibile per 3 e quindi $3|a$ se e solo se $3|b$. Supponiamo per assurdo che $3|b$, allora anche $d = 2b - 3c - (c + d)$ sarebbe divisibile per 3, contrariamente all'ipotesi induttiva.

328. Suggerimento: $(A + B)A^{-1} = B(A^{-1} + B^{-1})$.

337. Detti $A^1, \dots, A^n \in \mathbb{K}^n$ i vettori colonna di A, per ipotesi esiste un vettore non nullo $B^1 = (b_1, \dots, b_n)^T \in \mathbb{K}^n$ ed n coefficienti $c_1, \dots, c_n$ tali che $A^i = c_i B^1$ per ogni $i = 1, \dots, n$. Basta allora prendere B come la matrice che ha come unica colonna B^1 e $C = (c_1, \dots, c_n)$ affinché $A = BC$. In particolare $a_{ii} = b_i c_i$ e quindi

$$M_{1,1}(\mathbb{K}) \ni CB = c_1 b_1 + \cdots + c_n b_n = a_{11} + \cdots + a_{nn} = \operatorname{Tr}(A).$$

Per la proprietà associativa del prodotto si ha

$$A^2 = (BC)(BC) = B(CB)C = \operatorname{Tr}(A)BC = \operatorname{Tr}(A)A\,.$$

Per finire, osserviamo che per ogni $t \in \mathbb{K}$ vale

$$(A - tI)(A - (\operatorname{Tr}(A) - t)I) = A^2 - \operatorname{Tr}(A)A + t(\operatorname{Tr}(A) - t)I = t(\operatorname{Tr}(A) - t)I$$

da cui segue che se $t \neq 0, \operatorname{Tr}(A)$ allora $A - tI$ è invertibile con inversa $\dfrac{A - (\operatorname{Tr}(A) - t)I}{t(\operatorname{Tr}(A) - t)}$.

340. Traccia: se H contiene tutte le applicazioni di rango 1 il risultato è ovvio. Altrimenti possiamo trovare una decomposizione $W = U \oplus W'$ con $\dim U = 1$ ed un'applicazione $f \notin H$ tale che $f(V) = U$. Per Grassmann esiste $g \in H$ tale che $g(V) = U$. Sia $V' = \operatorname{Ker} g$, allora il nucleo S della naturale applicazione lineare $\Phi\colon \operatorname{Hom}(V, W) \to \operatorname{Hom}(V', W')$ ha dimensione $2n - 1$, contiene f, e quindi $\dim S \cap H \leq 2n - 2$. L'immagine $\Phi(H)$ ha dimensione $\geq (n^2 - n + 1) - (2n - 2) = (n - 1)^2 - (n - 1) + 1$ e per induzione esiste $h \in H$ tale che $\Phi(h)$ è un isomorfismo. Per concludere provare che esiste un unico scalare $t \in \mathbb{K}$ tale che $h + tg$ non è un isomorfismo.

347. Suggerimento: siano a, b i massimi della riga e della colonna contenenti 1. Allora $a \geq n$, $b \geq m$, $\max(a, b) \geq n + m - 1$ e quindi $ab > nm$.

350. Supponiamo per assurdo che A non sia invertibile e siano $x_1, \dots, x_n \in \mathbb{C}$ le coordinate di un vettore non nullo del nucleo di A. Se i è un indice tale che $|x_i| \geq |x_j|$ per ogni j, si ha $|x_i| > 0$ e $\sum_j a_{ij} x_j = 0$. Dunque

$$|a_{ii}||x_i| = |a_{ii} x_i| = \left| -\sum_{j \neq i} a_{ij} x_j \right| \leq \sum_{j \neq i} |a_{ij}||x_j| \leq \sum_{j \neq i} |a_{ij}||x_i|,$$

da cui segue $|a_{ii}| \leq \sum_{j \neq i} |a_{ij}|$ in contraddizione con le ipotesi. Il vettore con tutte le coordinate uguali ad 1 appartiene al nucleo di B.

375. Si tratta di Michael Atiyah (cercate su Wikipedia) nei suoi *Advice to a Young Mathematician*. La frase è estratta dai seguenti paragrafi iniziali della sezione *Motivation*:

> A research mathematician, like a creative artist, has to be passionately interested in the subject and fully dedicated to it. Without strong internal motivation you cannot succeed, but if you enjoy mathematics the satisfaction you can get from solving hard problems is immense.
>
> The first year or two of research is the most difficult. There is so much to learn. One struggles unsuccessfully with small problems and one has serious doubts about one's ability to prove anything interesting. I went through such a period in my second year of research, and Jean-Pierre Serre, perhaps the outstanding mathematician of my generation, told me that he too had contemplated giving up at one stage.
>
> Only the mediocre are supremely confident of their ability. The better you are, the higher the standards you set yourself.

400. Per calcolare il rango della terza matrice si possono dare due possibili suggerimenti alternativi:

1) se $\begin{pmatrix} I & A \\ 0 & C \end{pmatrix}$ è equivalente per righe a $\begin{pmatrix} I & A \\ B & 0 \end{pmatrix}$, provare che $\mathrm{Ker}(L_C) = \mathrm{Ker}(L_{BA})$;

2) provare che le matrici $\begin{pmatrix} I & A \\ B & 0 \end{pmatrix}$ e $\begin{pmatrix} I & A \\ 0 & -BA \end{pmatrix}$ sono equivalenti per righe.

408. Sappiamo che se una matrice simmetrica $B \in M_{n,n}(\mathbb{R})$ è del tipo $A^T A$, allora per ogni $x \in \mathbb{R}^n$ vale $x^T Bx = (Ax)^T (Ax) \geq 0$. Nella fattispecie abbiamo

$$(2,-1,0,0) \begin{pmatrix} 1 & 2 & 0 & 0 \\ 2 & 3 & 2 & 0 \\ 0 & 2 & 3 & 2 \\ 0 & 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \\ 0 \\ 0 \end{pmatrix} = -1$$

e di conseguenza il problema ha risposta negativa.

416. Supponiamo $A \neq 0$, denotiamo con d il massimo comune divisore dei coefficienti di A e poniamo $n = pd$ e $B = A/d$. I coefficienti non nulli di B non hanno fattori comuni e dalla relazione

$$0 = \frac{(I + nB)^q - I}{n} = qB + \sum_{i=2}^{q} \binom{q}{i} n^{i-1} B^i$$

ne segue che $n = pd$ divide il numero primo q e quindi $p = n = q$, $d = 1$. Se $p = q = 2$ allora $4A + 4A^2 = 0$, e.g. se A è diagonale con tutti i coefficienti uguali a $0, -1$. Se invece $q = n \geq 3$, allora n divide $\binom{q}{2}$ e di conseguenza n^2 divide q, che è assurdo.

422. Volendo, si potrebbe inizialmente calcolare il determinante $p(x)$ della matrice a coefficienti polinomi

$$A(x) = \begin{pmatrix} x & x^2 & 2x & 3x \\ 1 & x^2 & 4 & x^3 \\ 1 & x^3 & 4x & 5 \\ 1 & x^4 & 16 & x^9 \end{pmatrix}$$

ed in un secondo momento sostituire alla x i valori $0, 1, 2$. Tuttavia il conto risulta molto più semplice se si invertono le operazioni, ossia se prima si sostituisce la x con i suddetti valori e poi si calcolano i determinanti. Otteniamo le tre matrici

$$A(0) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 4 & 0 \\ 1 & 0 & 0 & 5 \\ 1 & 0 & 16 & 0 \end{pmatrix}, \qquad A(1) = \begin{pmatrix} 1 & 1 & 2 & 3 \\ 1 & 1 & 4 & 1 \\ 1 & 1 & 4 & 5 \\ 1 & 1 & 16 & 1 \end{pmatrix}, \qquad A(2) = \begin{pmatrix} 2 & 4 & 4 & 6 \\ 1 & 4 & 4 & 8 \\ 1 & 8 & 8 & 5 \\ 1 & 16 & 16 & 2^9 \end{pmatrix},$$

che hanno tutte determinate uguale a 0: la prima perché ha una riga nulla, le altre perché hanno due colonne adiacenti uguali. Quindi $p(0) = p(1) = p(2) = 0$.

426. Trattare separatamente i casi $z = 0$ e $z \neq 0$ e considerare matrici del tipo

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & x & y \end{pmatrix}, \qquad \begin{pmatrix} x & 0 & -z \\ u & 1 & 1 \end{pmatrix}.$$

427. Per induzione su n possiamo trovare $\epsilon_2, \dots, \epsilon_n = \pm 1$ tali che il determinante $a = |A_{11} + \operatorname{diag}(\epsilon_2 d_2, \dots, \epsilon_n d_n)|$ è diverso da 0. Basta adesso osservare che

$$|A + \operatorname{diag}(d_1, \epsilon_2 d_2, \dots, \epsilon_n d_n)| - |A + \operatorname{diag}(-d_1, \epsilon_2 d_2, \dots, \epsilon_n d_n)| = 2ad_1 \neq 0\,.$$

Se $A = \operatorname{diag}(d_1, \dots, d_n)$ l'unica scelta è $\epsilon_i = 1$ per ogni i.

435. Suggerimento: siano $a_{ij} \in \{\pm 1\}$ i coefficienti di A; a meno di scambiare l'ordine delle colonne possiamo assumere

$$\begin{pmatrix} a_{22} \\ a_{32} \end{pmatrix} = \pm \begin{pmatrix} a_{23} \\ a_{33} \end{pmatrix}.$$

438. Suggerimento: per ogni $i \leq n/2$ scambiare la i-esima colonna con la $n - i + 1$-esima.

447. Suggerimento senza parole:

$$\prod (t - x_i) = t^n - \left(\sum_{i=0}^{n} x_i \right) t^{n-1} + \cdots\,.$$

454. Si consiglia di eseguire nell'ordine: sostituire alla prima riga la somma di tutte le righe; dividere la prima riga per 45 (risp.: per nm); sostituire alla prima colonna la somma di tutte le colonne; dividere la prima colonna per 9 (risp.: per n).

455. Suggerimento: sottrarre la prima riga alle altre e poi aggiungere alla prima colonna la n-esima colonna divisa per n, per ogni $n > 1$.

469. Suggerimento: Sia X la matrice diagonale $m \times m$ i cui coefficienti sulla diagonale principale sono m indeterminate distinte $x_1, \dots, x_m$. Allora ogni coefficiente della matrice AXB è un polinomio omogeneo di grado 1 nelle variabili $x_1, \dots, x_m$.

Il lettore si ponga i seguenti interrogativi sul polinomio $p(x_1, \dots, x_m) = \det(AXB)$. Qual è il suo grado? Quanto vale quando poniamo uguali a 0 almeno $m - n + 1$ variabili distinte x_i? Quanto vale quando poniamo uguali a 0 esattamente $m - n$ variabili distinte x_i? Chi sono i suoi coefficienti? Quanto vale $p(1, \dots, 1)$?

472. L'unico punto non banale è il secondo. Sia $x \in \mathbb{K}^r$ tale che $HAH^T x = 0$; siccome A e HA hanno lo stesso rango $HAx = 0$ se e solo se $Ax = 0$ e quindi vale $AH^T x = 0$. Dato che A è antisimmetrica si ha quindi $x^T HA = 0$, ma, essendo l'applicazione lineare associata ad HA surgettiva, ne consegue che il funzionale lineare $y \mapsto x^T y$ è nullo, ossia $x = 0$.

502. Suggerimento: sia $v \in \mathbb{K}^n$ tale che $Av \neq 0$; esiste una matrice B di rango 1 tale che $B(Av) = v$? In caso di risposta affermativa, quanto vale $(BA)^2 v$?

519. Come primo passo supponiamo $pfp = pf$ e dimostriamo che $\operatorname{Ker} p$ è f-invariante, ossia che se $v \in V$, $p(v) = 0$, allora $pf(v) = 0$. Questo è immediato poiché $pf(v) = pfp(v) = pf(0) = 0$.

Viceversa, supponiamo che $\operatorname{Ker} p$ sia f-invariante e dimostriamo $pf(v) = pfp(v)$ per ogni $v \in V$. Siccome $p^2 = p$ si ha $p(v - p(v)) = p(v) - p^2(v) = 0$ e quindi $v - p(v) \in \operatorname{Ker} p$, $f(v - p(v)) \in \operatorname{Ker} p$ e di conseguenza

$$pf(v) = pf(p(v) + (v - p(v)) = pfp(v) + pf(v - p(v)) = pfp(v).$$

528. Il polinomio caratteristico di A è

$$p_A(t) = \begin{vmatrix} -t & 0 & 0 & 3 \\ 1 & -t & -1 & 0 \\ 0 & -3 & -t & -1 \\ 1 & 0 & 1 & -t \end{vmatrix} = t^4 - 5t^2 + 18.$$

Essendo il discriminante di $x^2 - 5x + 18$ negativo, il polinomio $t^4 - 5t^2 + 18$ non ha radici reali e possiede 4 radici complesse distinte. Quindi la matrice non è diagonalizzabile su $\mathbb{R}$ ed è invece diagonalizzabile su $\mathbb{C}$.

533. Suggerimento: basta memorizzare il quadrato magico di Dürer e considerare la somma di matrici

$$\begin{pmatrix} 16 & 3 & 2 & 13 \\ 5 & 10 & 11 & 8 \\ 9 & 6 & 7 & 12 \\ 4 & 15 & 14 & 1 \end{pmatrix} + (n-34) \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Si noti che per ogni $n \geq 48$ la matrice ha tutti i coefficienti diversi tra loro.

556. Suggerimento: detto r il grado di $q_f(t)$, considerare i resti delle divisioni dei polinomi $1, p(t), p(t)^2, \dots, p(t)^r$ per $q_f(t)$.

615. Se f è nilpotente allora $f - \lambda I$ è invertibile e quindi surgettiva per ogni $\lambda \neq 0$. Se f non è nilpotente, allora possiede un autovalore $\lambda \neq 0$ e quindi esiste un intero positivo k tale che

$$\operatorname{Ker}(f-\lambda I)^{k-1} \neq \operatorname{Ker}(f-\lambda I)^k = \operatorname{Ker}(f-\lambda I)^{k+1}.$$

Prendiamo un vettore $v \in \operatorname{Ker}(f-\lambda I)^k - \operatorname{Ker}(f-\lambda I)^{k-1}$, siccome

$$0 = (f-\lambda I)^k v = (-\lambda)^k v + f(\cdots)$$

ne consegue che v appartiene all'immagine di f. Se per assurdo $v = (f-\lambda I)w$ si avrebbe

$$(f-\lambda I)^k w = (f-\lambda I)^{k-1} v \neq 0, \quad (f-\lambda I)^{k+1} w = (f-\lambda I)^k v = 0$$

in contraddizione con la scelta di k.

622. Suggerimento per il punto 4: costruire per induzione su i dei sottospazi vettoriali $U_1, \dots, U_s$ tali che

$$U_i \subseteq U_{i+1}, \qquad f(U_{i+1}) \subseteq U_i, \qquad f^{s-i}(V) = U_i \oplus \operatorname{Span}(f^{s-i}(v), \dots, f^{s-1}(v)).$$

630. Suggerimento: considerare A come una matrice a coefficienti complessi e determinare una relazione tra la traccia di A e le molteplicità geometriche degli autovalori.

640. La prima matrice è nilpotente di rango tre, quindi possiede un unico blocco di Jordan $J_4(0)$. La seconda matrice è studiata sulla pagina di Wikipedia dedicata alla forma canonica di Jordan: it.wikipedia.org/wiki/Forma_canonica_di_Jordan. La terza matrice è studiata su Wikibooks all'indirizzo en.wikibooks.org/wiki/Linear_Algebra/Jordan_Canonical_Form.

657. Per il Teorema 12.1.4 l'applicazione lineare

$$\phi\colon V \to \mathbb{K}^r, \qquad \phi(v) = (\phi_1(v), \dots, \phi_r(v))^T,$$

è surgettiva. Basta adesso considerare il sottospazio generato da una qualunque successione $v_1, \dots, v_r \in V$ tale che $\phi(v_1), \dots, \phi(v_r)$ è la base canonica di $\mathbb{K}^r$.

677. A meno di scambiare ϕ_1 con ϕ_3 e ϕ_2 con ϕ_4 possiamo supporre ϕ_1, ϕ_2 linearmente indipendenti. A meno di scambiare ϕ_3 con ϕ_4 e ϕ_4 con $-\phi_3$ possiamo supporre ϕ_1, ϕ_2, ϕ_3 una base di $V^\vee$. Basta adesso considerare la corrispondente base duale $v_1, v_2, v_3 \in V$ e osservare che $\omega(v_1, v_2) = 1$.

678. Per l'Esercizio 657, a meno di restringersi ad un sottospazio di V non è restrittivo supporre che $\phi_1, \dots, \phi_{2r}$ sia un insieme di generatori di $V^\vee$. Dunque $r < \dim V \leq 2r$. Se $\dim V = 2r$ abbiamo già visto che le r forme decomponibili $\phi_i \wedge \phi_{i+1}$ sono linearmente indipedenti. Se $\dim V < 2r$ allora $r > 1$ e non è restrittivo supporre che ϕ_{2r} sia combinazione lineare dei rimanenti. Denotiamo $U = \operatorname{Ker} \phi_{2r-1}$ e dimostriamo che la restrizione di ω ad U è diversa da 0. Se indichiamo con $\psi_i \in U^\vee$ la restrizione di ϕ_i ad U, essendo l'applicazione $V^\vee \to U^\vee$ surgettiva e ψ_{2r} combinazione lineare degli altri funzionali, si ha che $\psi_1, \dots, \psi_{2r-2}$ generano $U^\vee$ e siccome $\dim U \geq \dim V - 1$ la conclusione segue per induzione su r.

701. Suggerimento: per assurdo, partendo dai coefficienti di grado più basso, mostrare che le due condizioni implicano che i quattro polinomi hanno gli stessi coefficienti.

Equivalentemente, passare al limite per $x \to 0$ per dedurre (vedi Esercizio 46) che $p_1(0) = p_2(0) = p_3(0) = p_4(0)$. Considerare quindi i polinomi $q_i(x) = (p_i(x) - p_i(0))/x$ e ragionare per induzione sul massimo grado dei polinomi.

709. Induzione su n, con il caso $n = 1$ evidente. Supponiamo $n > 1$, indichiamo con $p = \mathrm{MCD}(p_1, \dots, p_n)$, $q = \mathrm{MCD}(p_2, \dots, p_n)$ e scegliamo $a, b \in \mathbb{K}[t]$ tali che $p = ap_1 - bq$.

Ponendo $g_i = p_i/q$ per $i > 1$ si ha $\mathrm{MCD}(g_2, \dots, g_n) = 1$ e per l'ipotesi induttiva esistono polinomi $a_{ij} \in \mathbb{K}[t]$, con $i = 3, \dots, n$, $j = 2, \dots, n$ tali che

$$\det \begin{pmatrix} g_2 & g_3 & \cdots & g_n \\ a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix} = 1.$$

Lasciamo al lettore la semplice verifica che

$$p = ap_1 - bq = \det \begin{pmatrix} p_1 & p_2 & \cdots & \cdot & p_n \\ b & ag_2 & \cdots & \cdot & ag_n \\ 0 & a_{32} & \cdots & \cdot & a_{3n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & a_{n2} & \cdots & \cdot & a_{nn} \end{pmatrix}.$$

734. Diamo due distinte tracce di dimostrazione:

Traccia 1: Se $\mathbb{K} = \mathbb{F}_2 = \{0, 1\}$ il risultato è ovvio, possiamo quindi supporre che $\mathbb{K}$ contenga almeno tre elementi. Sia $n = \dim V$, per risolvere l'esercizio basta provare che φ è alternante oppure che esiste una base $e_1, \dots, e_n$ di V tale che $\varphi(e_i, e_j) = \varphi(e_j, e_i)$. Se φ non è alternante, allora esiste un vettore e_1 tale che $\varphi(e_1, e_1) \neq 0$. Possiamo completare e_1 ad una base $e_1, \dots, e_n$: a meno di aggiungere a ciascun e_i un multiplo scalare di e_1 possiamo supporre $\varphi(e_i, e_i) \neq 0$: infatti

$$\varphi(e_i + te_1, e_i + te_1) = \varphi(e_i, e_i) + t(\varphi(e_1, e_i) + \varphi(e_i, e_1)) + t^2\varphi(e_1, e_1)$$

e siccome il campo $\mathbb{K}$ ha almeno tre elementi la precedente espressione non si annulla identicamente.

Per ogni coppia di indici i, j si ha

$$\varphi\left(e_i, e_j - \frac{\varphi(e_i, e_j)}{\varphi(e_i, e_i)} e_i\right) = \varphi(e_i, e_j) - \frac{\varphi(e_i, e_j)}{\varphi(e_i, e_i)} \varphi(e_i, e_i) = 0$$

e quindi

$$0 = \varphi\left(e_j - \frac{\varphi(e_i, e_j)}{\varphi(e_i, e_i)} e_i, e_i\right) = \varphi(e_j, e_i) - \varphi(e_i, e_j).$$

Traccia 2: se $\varphi = 0$ non c'è nulla da dimoastrare. Se $\varphi \neq 0$ allora, nelle notazioni dell'Esercizio 731, le applicazioni f, g sono non nulle e per ogni $v \in V$ i due funzionali lineari $f(v)$ e $g(v)$ sono linearmente dipendenti. Usare l'Esercizio 284 per provare che $f = \pm g$.

735. Osserviamo che, se $x \in \operatorname{Ker} f$, allora $\varphi(x, y) = \varphi(f(x), f(y)) = \varphi(0, f(y)) = 0$ per ogni $y \in V$. Dunque $x \in \mathrm{LKer}(\varphi)$. Vale a dire $\operatorname{Ker} f \subseteq \mathrm{LKer}(\varphi)$. In particolare, se φ è non degenere, allora $\mathrm{LKer}(\varphi) = 0$ e dunque anche $\operatorname{Ker} f = 0$, ovvero f è iniettiva, il che risolve il punto (1). Per quanto riguarda il punto (2), si ha

$$\mathrm{rango}(f) = \dim V - \dim \operatorname{Ker} f \geq \dim V - \dim \mathrm{LKer}(\varphi) = \mathrm{rango}(\varphi).$$

738. Suggerimento: sia $\varphi\colon V \times V \to \mathbb{K}$ l'unica forma bilineare tale che $\varphi(e_i, e_j) = 0$ per $i > j$, $\varphi(e_i, e_i) = q(e_i)$ e $\varphi(e_i, e_j) = q(e_i + e_j) - q(e_i) - q(e_j)$ per $i < j$. Dimostrare per induzione su k che per ogni $x \in \operatorname{Span}(e_1, \dots, e_k)$ vale $q(x) = \varphi(x, x)$.

761. Ricordiamo che per ogni numero reale a si ha $|a| = s(a) \cdot a$, dove $s\colon \mathbb{R} \to \{-1, 0, 1\}$ è la funzione segno:

$$s(a) = \begin{cases} 1 & \text{se } a > 0 \\ 0 & \text{se } a = 0 \\ -1 & \text{se } a < 0. \end{cases}$$

Per evitare ambiguità notazionali, scriveremo $\sum_{(i,j):i\neq j}$ per indicare la somma su tutte le $n^2 - n$ coppie (i, j) con $i \neq j$, mentre scriveremo $\sum_{j:j\neq i}$ per indicare la somma sugli $n - 1$

indici j diversi da i. Ad esempio,

$$\sum_{(i,j):i\neq j} a_{ij} = \sum_{i=1}^{n} \sum_{j:j\neq i} a_{ij}.$$

La condizione $2a_{ii} \geq \sum_{j=1}^{n} |a_{ij}|$ implica in particolare che $a_{ii} \geq 0$ e che i numeri

$$b_i = a_{ii} - \sum_{j:j\neq i} |a_{ij}| = a_{ii} - \sum_{j:j\neq i} |a_{ji}|, \qquad i = 1, \ldots, n,$$

sono nonnegativi. Per un vettore $x \in \mathbb{R}^n$ di coordinate $x_1, \ldots, x_n$ si ha

$$\begin{aligned} x^T Ax &= \sum_i \left(a_{ii}x_i^2 + \sum_{j:j\neq i} a_{ij}x_i x_j \right) = \sum_i \left(a_{ii}x_i^2 + \sum_{j:j\neq i} |a_{ij}|(s(a_{ij})x_i x_j) \right) \\ &= \sum_i b_i x_i^2 + \sum_{(i,j):i\neq j} |a_{ij}|(x_i^2 + s(a_{ij})x_i x_j) \\ x^T Ax &= \sum_j \left(a_{jj}x_j^2 + \sum_{i:i\neq j} a_{ij}x_i x_j \right) = \sum_j \left(a_{jj}x_j^2 + \sum_{j:j\neq i} |a_{ij}|(s(a_{ij})x_i x_j) \right) \\ &= \sum_j b_j x_j^2 + \sum_{(i,j):i\neq j} |a_{ij}|(x_j^2 + s(a_{ij})x_i x_j). \end{aligned}$$

Sommando le due espressioni si ottiene

$$\begin{aligned} 2x^T Ax &= \sum_i b_i x_i^2 + \sum_j b_j x_j^2 + \sum_{(i,j):i\neq j} |a_{ij}|(x_i^2 + x_j^2 + 2s(a_{ij})x_i x_j) \\ &= 2\sum_i b_i x_i^2 + \sum_{(i,j):i\neq j} |a_{ij}|(x_i + s(a_{ij})x_j)^2 \geq 0. \end{aligned}$$

Notiamo che la dimostrazione prova in aggiunta che se $2a_{ii} > \sum_{j=1}^{n} |a_{ij}|$ per ogni indice i, allora A è definita positiva.

773. La matrice associata alla forma quadratica Φ è

$$Q = \begin{pmatrix} 0 & 1/2 & 0 & 0 \\ 1/2 & 1 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

ed il polinomio caratteristico di Q è

$$p_Q(t) = \begin{pmatrix} -t & 1/2 & 0 & 0 \\ 1/2 & 1-t & 0 & 1 \\ 0 & 0 & -1-t & 0 \\ 0 & 1 & 0 & -t \end{pmatrix} = t^4 - \frac{9}{4}t^2 - \frac{5}{4}t = t\left(t^3 - \frac{9}{4}t - \frac{5}{4}\right).$$

Dunque $t = 0$ è una radice del polinomio caratteristico di Q di molteplicità 1. Ne segue $\text{rango}(\Phi) = 4 - 1 = 3$; inoltre la successione dei segni dei coefficienti di $p_Q(t)$ è $(+, -, -)$. C'è una sola variazione, quindi $p_Q(t)$ ha esattamente una radice positiva. Dunque $\text{segnatura}(\Phi) = (1, 2)$.

787. Si ha $B_1 = (0)$ e dunque, evidentemente $\text{rango}(B_1) = 0$ e $\text{segnatura}(B_1) = (0, 0)$. Per quanto riguarda B_2, si ha

$$B_2 = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}$$

Il polinomio caratteristico di B_2 è

$$p_{B_2}(t) = \det \begin{pmatrix} -t & 1 \\ 1 & 2-t \end{pmatrix} = t^2 - 2t - 1.$$

Il termine noto di questo polinomio è diverso da zero, dunque $\text{rango}(B_2) = 2$; inoltre la successione dei segni dei coefficienti di $p_{B_2}(t)$ è $(+, -, -)$. C'è una sola variazione, quindi

$p_{B_2}(t)$ ha esattamente una radice positiva. Dunque segnatura$(B_2) = (1,1)$. Infine,

$$B_3 = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 4 \end{pmatrix}$$

Il polinomio caratteristico di B_3 è

$$p_{B_2}(t) = \det \begin{pmatrix} -t & 1 & 2 \\ 1 & 2-t & 3 \\ 2 & 3 & 4-t \end{pmatrix} = -t^3 + 6t^2 + 6t = t\,(-t^2 + 6t + 6).$$

Dunque $t = 0$ è una radice del polinomio caratteristico di B_3 di molteplicità 1. Ne segue rango$(B_3) = 3 - 1 = 2$; inoltre la successione dei segni dei coefficienti di $p_{B_3}(t)$ è $(-,+,+)$. C'è una sola variazione, quindi $p_{B_3}(t)$ ha esattamente una radice positiva. Dunque segnatura$(B_3) = (1,1)$.

788. La matrice B_n ha la forma

$$B_n = \begin{pmatrix} 0 & 1 & 2 & \cdots & n-1 \\ 1 & 2 & 3 & \cdots & n \\ 2 & 3 & 4 & \cdots & n+1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n-1 & n & n+1 & \cdots & 2n-2 \end{pmatrix}.$$

Sottraiamo all'ultima colonna la penultima, alla penultima la terzultima e così via (quest'operazione lascia invariato il rango della matrice). Troviamo la matrice

$$\begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ 2 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n-1 & 1 & 1 & \cdots & 1 \end{pmatrix}$$

che ha evidentemente rango 2 (ci sono solamente due colonne linearmente indipendenti). Abbiamo così dimostrato

$$\text{rango}(B_n) = 2, \qquad \forall n \geq 4.$$

Per quanto riguarda la segnatura, osserviamo che il minore 2×2 in alto a sinistra ha rango 2 e dunque la segnatura di B_n coincide con la segnatura di questo minore. Ma il minore 2×2 in alto a sinistra è proprio B_2 e abbiamo già calcolato, nell'Esercizio 787, che la sua segnatura è (1,1). Ne segue

$$\text{segnatura}(B_n) = (1,1), \qquad \forall n \geq 4.$$

830. Poniamo $v_1 = p_2 - p_1$, $v_2 = p_3 - p_1$, $w_1 = q_2 - q_1$ e $w_2 = q_3 - q_1$. L'affinità cercata è quindi

$$x \mapsto A(x - p_1) + q_1\,,$$

dove A è la trasformazione lineare che manda v_i in w_i per $i = 1, 2$. In termini di matrici,

$$v_1 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}; \qquad v_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix};$$

$$w_1 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}; \qquad w_2 = \begin{pmatrix} 6 \\ -5 \end{pmatrix};$$

Dunque la matrice A è determinata dall'equazione

$$A \begin{pmatrix} 2 & 2 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 6 \\ -1 & -5 \end{pmatrix}$$

ovvero

$$A = \begin{pmatrix} -1 & 6 \\ -1 & -5 \end{pmatrix} \begin{pmatrix} 2 & 2 \\ -1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 5/4 & 7/2 \\ -3/2 & -2 \end{pmatrix}$$

In conclusione, l'affinità cercata è

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 5/4 & 7/2 \\ -3/2 & -2 \end{pmatrix} \begin{pmatrix} x-1 \\ y-2 \end{pmatrix} + \begin{pmatrix} 1 \\ 8 \end{pmatrix} = \begin{pmatrix} \frac{5}{4}x + \frac{7}{2}y - \frac{29}{4} \\ \\ -\frac{3}{2}x - 2y + \frac{27}{2} \end{pmatrix}$$

866. Sia S la matrice di Sylvester (14.3). Dato un generico numero intero t, moltiplicare a_i e b_i per t^i equivale a eseguire nell'ordine le seguenti operazioni:

(1) moltiplicare per t^i la i-esima colonna di S,
(2) dividere per t^i la i-esima riga di $S_{n,m}$ se $i \leq m$,
(3) dividere per t^{i-m} la i-esima riga di $S_{n,m}$ se $i > m$.

Alla fine il determinante di S risulta moltiplicato per t^e, dove

$$e = \sum_{i=1}^{n+m} i - \sum_{i=1}^{m} i - \sum_{i=1}^{n} i = nm.$$