

Enrico Schlesinger

Algebra lineare e geometria

ZANICHELLI

I diritti di elaborazione in qualsiasi forma o opera, di memorizzazione anche digitale su supporti di qualsiasi tipo (inclusi magnetici e ottici), di riproduzione e di adattamento totale o parziale con qualsiasi mezzo (compresi i microfilm e le copie fotostatiche), i diritti di noleggio, di prestito e di traduzione sono riservati per tutti i paesi. L'acquisto della presente copia dell'opera non implica il trasferimento dei suddetti diritti né li esaurisce.

Le fotocopie per uso personale (cioè privato e individuale, con esclusione quindi di strumenti di uso collettivo) possono essere effettuate, nei limiti del 15% di ciascun volume, dietro pagamento alla S.I.A.E del compenso previsto dall'art. 68, commi 4 e 5, della legge 22 aprile 1941 n. 633. Tali fotocopie possono essere effettuate negli esercizi commerciali convenzionati S.I.A.E. o con altre modalità indicate da S.I.A.E.

Per le riproduzioni ad uso non personale (ad esempio: professionale, economico, commerciale, strumenti di studio collettivi, come dispense e simili) l'editore potrà concedere a pagamento l'autorizzazione a riprodurre un numero di pagine non superiore al 15% delle pagine del presente volume. Le richieste per tale tipo di riproduzione vanno inoltrate a

Centro Licenze e Autorizzazioni per le Riproduzioni Editoriali (CLEARedi)
Corso di Porta Romana, n. 108
20122 Milano
e-mail autorizzazioni@clearedi.org e sito web www.clearedi.org

L'editore, per quanto di propria spettanza, considera rare le opere fuori del proprio catalogo editoriale, consultabile al sito www.zanichelli.it/f_catalog.html. La fotocopia dei soli esemplari esistenti nelle biblioteche di tali opere è consentita, oltre il limite del 15%, non essendo concorrenziale all'opera.
Non possono considerarsi rare le opere di cui esiste, nel catalogo dell'editore, una successiva edizione, le opere presenti in cataloghi di altri editori o le opere antologiche. Nei contratti di cessione è esclusa, per biblioteche, istituti di istruzione, musei ed archivi, la facoltà di cui all'art. 71 - ter legge diritto d'autore.
Maggiori informazioni sul nostro sito: www.zanichelli.it/fotocopie/

Realizzazione editoriale: CompoMat, Configni (RI)

Copertina:

- *Progetto grafico:* Miguel Sal & C., Bologna
- *Realizzazione:* CL'EM, Milano
- *Immagine di copertina:* © Isabella Nenci

Prima edizione: luglio 2011

Ristampa:

5

4



3

2

2015

2014

2013

Questo libro è stampato su carta che rispetta le foreste. www.zanichelli.it/la-casa-editrice/carta-e-ambiente/

Realizzare un libro è un'operazione complessa, che richiede numerosi controlli: sul testo, sulle immagini e sulle relazioni che si stabiliscono tra essi.

L'esperienza suggerisce che è praticamente impossibile pubblicare un libro privo di errori. Saremo quindi grati ai lettori che vorranno segnalarceli.

Per segnalazioni o suggerimenti relativi a questo libro l'indirizzo a cui rivolgersi è:

Zanichelli editore S.p.A.
Via Irnerio 34
40126 Bologna
fax 051293322
e-mail: linea_universitaria@zanichelli.it
sito web: www.zanichelli.it

Prima di effettuare una segnalazione è possibile verificare se questa sia già stata inviata in precedenza, identificando il libro interessato all'interno del nostro catalogo on line (www.zanichelli.it/f_catalog.html) e selezionando il link ERRATA CORRIGE, dove sono disponibili le eventuali correzioni in formato PDF.

Per comunicazioni di tipo commerciale: universita@zanichelli.it



Questo libro è stampato su carta che rispetta le foreste.
www.zanichelli.it/la-casa-editrice/carta-e-ambiente/

Stampa: Tipografia Babina

Via Aldo Moro 18, 40068 San Lazzaro di Savena (BO)

per conto di Zanichelli editore S.p.A.

Via Irnerio 34, 40126 Bologna

A mia moglie e ai miei figli,
che mi hanno sopportato
nel lungo periodo di stesura
di questo volume

Indice generale

Prefazione	ix
1 Spazio euclideo e vettori	1
1 Introduzione	1
2 Richiami di geometria euclidea nello spazio	4
3 Dalla geometria all'algebra dei vettori	10
4 Sistemi di riferimento e coordinate	21
4.1 Sistema di riferimento in un piano	21
4.2 Sistemi di riferimento nello spazio	26
5 Coordinate cartesiane nello spazio	30
6 Proiezioni ortogonali e prodotto scalare	33
7 Prodotto vettoriale e prodotto misto	40
8 Geometria analitica di rette e piani nello spazio	46
8.1 Equazioni parametriche di una retta	46
8.2 Equazione cartesiana di un piano	54
8.3 Equazioni cartesiane di una retta nello spazio	56
8.4 Equazioni parametriche di un piano nello spazio	58
8.5 Distanza di due punti. Equazione di una sfera	59
8.6 Distanza tra un punto e un piano	60
8.7 Distanza tra un punto e una retta nello spazio	61
8.8 Distanza tra due rette	63
2 Sistemi lineari	69
1 Introduzione	69
2 Equazioni lineari	70
3 Esempi introduttivi	72
4 Vettori colonna	77
5 Sistemi lineari e matrici	80
6 Metodo di eliminazione di Gauss	86
6.1 Operazioni elementari sulle righe	90
6.2 L'algoritmo di Gauss	91
6.3 Rango di una matrice. Teoremi di Cramer e Rouché-Capelli	96

3 Algebra delle matrici	119
1 Introduzione	119
2 Somma e prodotto per uno scalare	121
3 Il prodotto righe per colonne	122
4 Matrici invertibili	128
5 Matrice trasposta. Matrici simmetriche	133
6 L'algoritmo di Gauss-Jordan e il calcolo dell'inversa	137
7 Fattorizzazione LU	141
7.1 La fattorizzazione nel caso non siano necessari scambi di righe	141
7.2 La fattorizzazione nel caso generale	147
8 Prodotto di matrici a blocchi	150
4 Spazi vettoriali	153
1 Introduzione	153
2 Assiomi ed esempi	154
3 Sottospazi	158
4 Combinazioni lineari	163
5 Indipendenza lineare	168
6 Basi e Dimensione	173
7 Coordinate	184
8 Rango di una matrice	191
9 Equazioni cartesiane di un sottospazio	202
10 Operazioni sui sottospazi	205
5 Applicazioni lineari	217
1 Introduzione	217
2 Applicazioni lineari	219
3 Nucleo, fibre e immagine	226
4 Algebra delle applicazioni lineari	232
5 Il teorema di rappresentazione	240
6 Teorema di nullità più rango e applicazioni	255
6 Determinante	261
1 Introduzione	261
2 Determinante e mosse di Gauss	262
3 Determinante di matrici di permutazione	266
4 Formula esplicita per il determinante	271
5 Sviluppi di Laplace	280
6 Il teorema di Binet e il determinante di un'applicazione lineare	289
7 Determinante e rango	292
8 Complementi	295
7 Autovalori e autovettori	301
1 Introduzione	301
2 Autovettori e autovalori di un'applicazione lineare	301
3 Autovettori e autovalori di una matrice	307
4 Ricerca di autovalori e autovettori	311

5	Matrici simili	331
6	Il problema della forma canonica	339
8	Spazi euclidei	363
1	Introduzione	363
2	Spazi euclidei	364
3	Il teorema di Pitagora e la disuguaglianza di Schwarz	370
4	Basi ortonormali e matrici ortogonali	377
5	Proiezioni ortogonali e algoritmo di Gram-Schmidt	389
6	Equazioni normali e il metodo dei minimi quadrati	404
7	Matrici di proiezioni ortogonali	412
8	Il caso complesso	416
9	Complementi	425
9.1	Il teorema di Eulero sulle rotazioni dello spazio	425
9.2	Riflessioni ortogonali	427
9	Teoremi spettrali e forme quadratiche	431
1	Introduzione	431
2	Teorema spettrale	432
3	Forme quadratiche	441
4	La decomposizione ai valori singolari	462
5	Il caso complesso	477
6	Matrici normali reali	482
7	Quadratiche	486
	Indice analitico	497

Prefazione

Il presente volume integra i due volumi *Analisi matematica 1* e *Analisi matematica 2* di Bramanti, Pagani e Salsa, in modo da fornire un testo completo per i corsi di matematica di base delle facoltà scientifiche. L'*algebra lineare* è la branca della matematica che si occupa dello studio dei *sistemi lineari* e più in generale delle *applicazioni lineari*. Le applicazioni lineari rivestono un ruolo centrale in matematica perché sono le funzioni più semplici; lo studio di ogni altra funzione è ricondotto, in prima approssimazione, a quello di una funzione lineare: la derivata di una funzione è l'applicazione lineare che meglio approssima la funzione localmente; esattamente come la retta tangente è la retta che meglio approssima una curva in un intorno del punto di tangenza. In coordinate un'applicazione lineare si rappresenta mediante una *matrice* e, come conseguenza, l'algebra lineare dal punto di vista computazionale si riduce all'algebra delle matrici. L'aspetto computazionale sta assumendo un ruolo sempre più importante col crescere della capacità di calcolo dei computer, che già oggi rende possibile la risoluzione di sistemi lineari con centinaia di migliaia di incognite. Il risultato è che all'ingegnere e allo scienziato dei nostri giorni è richiesta una conoscenza sempre più ampia e approfondita dell'algebra lineare.

Nella trattazione non ho cercato di seguire il percorso logico più breve possibile, ho invece privilegiato un'esposizione degli argomenti che porti gradualmente dal concreto all'astratto, cercando così di ovviare a quella percezione di eccessiva astrazione che sembra essere la principale difficoltà nell'affrontare lo studio dell'algebra lineare. Specificamente, nel primo capitolo si mostra come gli assiomi degli spazi vettoriali abbiano origine da quelli della geometria euclidea e si trattano dettagliatamente le operazioni sui vettori geometrici con cui lo studente è familiare dallo studio della fisica; nel secondo capitolo si introducono i sistemi lineari e si descrive il metodo di eliminazione di Gauss, fornendo così uno strumento di calcolo importante; il terzo capitolo è sostanzialmente dedicato al prodotto di matrici. Solo nel quarto e nel quinto capitolo si sviluppa la teoria astratta degli spazi vettoriali e delle applicazioni lineari, sempre privilegiando un approccio costruttivo; per esempio, le dimostrazioni del teorema di nullità più rango e della formula di Grassmann forniscono un metodo per costruire delle basi dei sottospazi coinvolti.

Il volume è stato pensato per poter essere impiegato per due scopi differenti: come libro di testo per un corso di base e come libro di riferimento di algebra lineare per tutto il corso degli studi; sono pertanto trattati, di norma negli ultimi paragrafi di ciascun capitolo, alcuni risultati meno elementari, che non trovano spazio in un corso di base di algebra lineare e il cui studio non è comunque richiesto per la comprensione degli argomenti fondamentali.

In un corso tipico di 35 ore di lezione e 25 ore di esercitazione, il testo può essere utilizzato secondo queste linee: un'esposizione molto veloce delle operazioni sui vettori geometrici dello spazio euclideo, che includa prodotto scalare, proiezioni ortogonali, prodotto vettoriale e determinante di matrici 2×2 e 3×3 ; il metodo di eliminazione di Gauss e il teorema di Rouché-Capelli del secondo capitolo richiedono da 3 a 5 ore di lezione; nel capitolo 3, il paragrafo sulla fattorizzazione LU è scritto come un complemento e può essere omesso, così come gli ultimi due paragrafi del capitolo 4; il capitolo 5 sulle applicazioni lineari è fondamentale: si può poi scegliere se seguire l'ordine del testo (determinante e autovalori) o anticipare lo studio degli spazi euclidei: in ogni caso, al determinante non andrebbe dedicato troppo spazio e il paragrafo 2 del capitolo sul determinante è studiato per consentire una trattazione molto veloce dell'argomento; l'ultimo paragrafo del capitolo sugli autovalori è dedicato alla forma canonica di Jordan e dovrebbe essere omesso in un corso di base; una trattazione minimale del capitolo sugli spazi euclidei potrebbe limitarsi ai primi cinque paragrafi. L'ultimo capitolo del libro contiene due argomenti fondamentali, che sono il teorema spettrale e la sua applicazione allo studio delle forme quadratiche reali, insieme a molti complementi, in genere non elementari, che di nuovo sono pensati più come riferimento che non come parte del libro di testo: tra questi la fattorizzazione di Cholesky delle matrici definite positive, la legge di inerzia di Sylvester, la decomposizione ai valori singolari, la matrice pseudoinversa, la forma canonica di una matrice normale reale, la forma canonica euclidea dell'equazione di una quadrica.

Ringraziamenti

Desidero ringraziare le persone che più mi hanno aiutato nella stesura di questo testo: Sandro Salsa che mi ha proposto di scriverlo nell'ambito della revisione del volume *Matematica: calcolo infinitesimale e algebra lineare* di Bramanti, Pagani e Salsa; Marco Bramanti che mi ha fornito numerosi e preziosi suggerimenti in corso d'opera; Irene Sabadini che ha riletto criticamente le numerose versioni preliminari e che ha sperimentato l'impostazione didattica del volume nel corso di Algebra lineare e Geometria per gli allievi di Ingegneria Matematica del Politecnico di Milano; Marco Boella che ha scritto gli esercizi svolti del secondo capitolo; Cecilia Rizzi che ha scritto alcuni degli esercizi; Maurizio Vianello che mi ha aiutato con i problemi tecnici nell'utilizzo di LaTeX. Ho poi tratto profitto da numerose discussioni sul contenuto del testo coi colleghi Pietro Pirola, Cristina Cerutti, Luca Formaggia, Pierluigi Moseneder, Fabio Nobile, Roberto Notari, Irene Sabadini, Riccardo Sacco, Marco Verani. Uno speciale ringraziamento va a Isabella Nenci della casa editrice Zanichelli la cui consulenza è stata fondamentale per poter portare a compimento il volume.

Spazio euclideo e vettori

1 INTRODUZIONE

La parola geometria significa misura della terra e in effetti, la geometria è proprio quella parte della matematica che si occupa di descrivere e misurare lo spazio fisico. Per almeno due millenni si è considerato un unico modello matematico per lo spazio, la geometria euclidea e si è pensato che tale geometria descrivesse in modo accurato la realtà.

I principi fondamentali della geometria euclidea si trovano già negli *Elementi* di Euclide, probabilmente il trattato che ha avuto più successo nella storia dell'umanità. Scritti originariamente nel III secolo avanti Cristo, i tredici libri degli Elementi conservano ancora oggi una loro attualità, tant'è vero che sono ancora, direttamente o indirettamente, utilizzati a scuola e su internet ne sono disponibili ottime versioni critiche (<http://aleph0.clarku.edu/~djoyce/java/elements/elements.html> e <http://web.math.unifi.it/ssis/euclidel.htm> per esempio).

Il successo degli Elementi è in parte certamente motivato dall'importanza intrinseca della geometria, che fornisce un linguaggio di base alla matematica e strumenti utili a molte altre discipline per esempio: l'architettura, l'ingegneria e la fisica, che necessitano tutte di una descrizione quantitativa dello spazio che ci circonda. Ma il principale merito dell'opera risiede nell'utilizzo del *metodo assiomatico* per la trattazione della teoria. Una teoria assiomatica si fonda su un certo numero di *concetti primitivi* e di *assiomi* (o postulati). I concetti primitivi sono gli oggetti di base della teoria. Per esempio, nel caso della geometria euclidea, i concetti primitivi sono punti, rette e piani; nel caso dell'aritmetica, i concetti primitivi sono i numeri naturali. Per mezzo dei concetti primitivi si possono poi definire o costruire tutti gli altri oggetti a cui la teoria è interessata, per esempio, in geometria, i triangoli e i parallelepipedi. Gli assiomi sono invece le proprietà che i concetti primitivi devono soddisfare. La teoria consiste nel dedurre tutte le conseguenze logicamente necessarie degli assiomi: in matematica, le affermazioni dedotte si dicono *teoremi* e il processo di deduzione *dimostrazione*. La potenza del metodo risiede nel fatto che i teoremi rimangono validi qualunque interpretazione concreta si dia ai concetti primitivi, purché gli assiomi della teoria siano verificati. L'esempio fondamentale in questo libro è quello della teoria degli *spazi vettoriali*, i cui concetti primitivi sono detti vettori perché soddisfano gli

stessi assiomi dei vettori della fisica; ma anche le funzioni reali soddisfano gli stessi assiomi, per cui i teoremi della teoria si applicano ugualmente ai vettori della fisica e alle funzioni. Il primo a rendersi conto dell'analogia tra vettori e funzioni fu lo scienziato francese Fourier, che decompose una funzione periodica nella somma delle sue componenti armoniche, in modo perfettamente analogo a come un vettore si decomponesse nelle sue componenti lungo gli assi cartesiani.

Naturalmente, perché una teoria assiomatica sia utile, è fondamentale da un lato che ne risulti un modello astratto, sufficientemente vicino a quanto suggerito dall'esperienza, da poter fornire soluzioni a problemi concreti; dall'altro gli assiomi devono essere sufficientemente semplici da consentire uno sviluppo efficace della teoria.

Per quel che riguarda gli assiomi di Euclide, per quasi due millenni si è pensato che esprimessero delle verità evidenti, universalmente valide, come per esempio l'enunciato che per due punti passi una e una sola retta. Tra la fine del 1700 e il 1800 però, dopo secoli di infruttuose ricerche mirate a dedurre il quinto postulato di Euclide dai precedenti, si fece finalmente varco l'idea che gli assiomi fossero fondati sull'esperienza e non fossero delle verità a priori. Di qui fu breve il passo che portò a costruire *altri* modelli di geometria e a mettere in dubbio il fatto che il modello euclideo fosse quello che meglio descriveva la realtà fisica. Nacquero così le geometrie non eucleedee. Queste non ebbero altrettanto successo della geometria euclidea, ma il motivo è che la geometria si è da allora sviluppata seguendo, fondamentalmente, un'altra strada, aperta da Riemann. L'idea di Riemann è che un oggetto geometrico si possa scomporre in tanti piccoli pezzi elementari, così come la superficie della Terra è ben rappresentata da un atlante: a ogni carta dell'atlante corrisponde una piccola regione della superficie terrestre e questa piccola regione in buona approssimazione ha la forma di un rettangolo nel piano euclideo. Solo incollando opportunamente le carte dell'atlante, la superficie della Terra assume una forma sferica. Analogamente lo spazio dovrebbe essere descritto dapprima localmente e poi le varie carte locali dovrebbero essere incollate insieme in un modo che non è dato conoscere a priori. L'intuizione suggerisce che queste carte locali dovrebbero almeno in prima approssimazione essere descritte come lo spazio euclideo. Questo è ancora grosso modo lo stato delle cose: nessuno crede che la geometria dell'Universo debba obbedire globalmente agli assiomi euclidei, ma la geometria euclidea, con la sua generalizzazione a dimensione arbitraria, fornisce ancora il miglior modello per descrivere lo spazio nelle nostre immediate vicinanze.

La geometria euclidea conserva così un ruolo fondamentale nella matematica moderna. Però, a prima vista sembra scomparsa dai libri di testo e dai programmi degli insegnamenti universitari. Perché? Il motivo è che gli assiomi euclidei sono stati sostituiti da altri più efficaci e la geometria euclidea ha, sostanzialmente, cambiato nome ed è divenuta parte dell'algebra lineare. Scopo di questo capitolo è di far toccare con mano al lettore questa metamorfosi, con il fine di rendere più digeribile la teoria astratta che verrà sviluppata nei capitoli successivi.

In realtà, gli Elementi non contengono nemmeno un sistema di assiomi completo: gli assiomi presenti nel trattato di Euclide non consentono davvero di dedurre tutte le conseguenze che se ne vogliono trarre. Ma questo è un problema ormai risolto e sono, da tempo, disponibili formulazioni soddisfacenti di tutti gli assiomi necessari. La versione più conosciuta è quella dei *Fondamenti della Geometria* di Hilbert, che il lettore giustamente curioso può ritrovare, almeno in inglese, su internet (<http://www.gutenberg.org/etext/17384>). Nel libro di Hilbert gli assiomi sono orga-

nizzati in cinque gruppi (connessione, ordine, assioma delle parallele, congruenza e continuità) e sono più di venti.

Il vero problema della formulazione originaria della geometria euclidea risiede però nella difficoltà di sviluppo della teoria. Penso che tutti abbiamo almeno un po' sofferto cercando di capire le dimostrazioni dei teoremi sui triangoli. In realtà nonostante tutto il loro fascino, le dimostrazioni basate su costruzioni geometriche richiedono tipicamente molto ingegno e spesso risultano poco convincenti, magari semplicemente perché un disegno non illustra adeguatamente la complessità della situazione. Un metodo alternativo e spesso più efficace, di trattare un problema geometrico è quello di tradurlo in un problema algebrico. Questo metodo, che va sotto il nome di geometria analitica o geometria delle coordinate, fu escogitato nel 1600 da Fermat e Cartesio. L'idea di base è di fissare un sistema di riferimento che permetta di individuare un punto attraverso le sue coordinate: se si fissano nel piano un'origine e i quattro punti cardinali, ogni punto P del piano è descritto da due coordinate (x, y) : l'ascissa x esprime di quanto bisogna muoversi verso est o verso ovest (a seconda del segno di x) per ritrovarsi allineati verticalmente con P , mentre l'ordinata y indica di quanto bisogna poi muoversi verticalmente per arrivare in P . Le rette possono ora essere definite come i luoghi dei punti del piano le cui coordinate soddisfano un'equazione lineare della forma $y = mx + c$ o $x = c$. Possiamo quindi verificare algebricamente che per due punti passa un'unica retta: gli assiomi della geometria euclidea scompaiono, perché possono essere dedotti dalle regole dell'aritmetica.

Questo approccio, numericamente molto efficace, presenta tre inconvenienti. Il primo inconveniente è che, se ci liberiamo completamente dell'oggetto geometrico sostituendolo col suo analogo in coordinate, ci ritroviamo con delle coordinate fissate *a priori*. Spesso è invece più conveniente fissare il sistema di coordinate *a posteriori*, a seconda del problema che si deve risolvere. Per esempio, nello studiare un'ellisse conviene prendere il centro di simmetria dell'ellisse come origine delle coordinate e gli assi di simmetria dell'ellisse come assi cartesiani. Solo così si ottiene per l'ellisse una semplice equazione della forma

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

Il secondo inconveniente è che, se ci limitiamo a lavorare in coordinate, tutto quello che definiamo e dimostriamo dipende, in linea di principio, dal sistema di coordinate. Per esempio, se definiamo la somma di due vettori in un sistema di coordinate, dobbiamo poi controllare che il risultato non dipenda dalle coordinate che stiamo utilizzando.

Il terzo inconveniente è che l'algebra, coi suoi procedimenti veloci ma meccanici, rischia di oscurare la visualizzazione dei problemi che invece è di fondamentale aiuto al nostro ragionamento. Di regola, nel risolvere un problema geometrico, siamo guidati dalla nostra intuizione, che è fondata sull'esperienza visiva dello spazio. Questa intuizione deve continuare a essere la nostra guida nel dirigere i potenti strumenti dell'algebra.

Fu solo verso la fine del 1800 che si impose il linguaggio dei vettori; una delle motivazioni principali era la necessità di esprimere le leggi dell'elettromagnetismo in modo indipendente dalle coordinate. Vale la pena osservare che tra i creatori della teoria vanno annoverati il padre dell'elettromagnetismo Maxwell, il fisico matematico Gibbs e l'ingegnere Heaveside. Secondo lo storico della matematica Kline, Heaveside pensava all'analisi vettoriale come a "una forma abbreviata dell'ordinaria geometria

analitica cartesiana". Questo è un punto di vista che il lettore dovrebbe far proprio: basti osservare che è più semplice pensare a un vettore che non alla terna delle sue coordinate.

Il punto che mi preme sottolineare è che il linguaggio dei vettori ha in comune con la trattazione euclidea l'indipendenza dal sistema di coordinate e la visualizzazione geometrica e in comune con la geometria cartesiana l'aspetto algebrico, perché sui vettori è possibile definire delle operazioni algebriche con proprietà del tutto simili alle operazioni sui numeri. I vettori, quindi, sono un oggetto intrinseco, indipendente dalle coordinate, su cui a differenza dei punti, è possibile operare algebricamente e quindi in maniera efficace. Per essere specifici, le operazioni che introduciamo sui vettori sono la somma e il prodotto per uno scalare, il prodotto scalare e il prodotto vettoriale. Mediante queste operazioni è immediato ricavare formule per lunghezze e angoli e risolvere, per esempio, il problema di calcolare la distanza di due rette nello spazio (formula (8.20)).

In questo capitolo introduciamo il concetto di vettore e le operazioni sui vettori e ricaviamo dalla geometria euclidea le proprietà algebrica delle operazioni sui vettori. Nei capitoli successivi del libro ribalteremo il punto di vista: i vettori saranno i concetti primitivi dell'algebra lineare e formeranno insiemi detti *spazi vettoriali*: gli assiomi della teoria saranno le proprietà delle operazioni sui vettori. Questo ci permette di non essere assolutamente rigorosi perché tutti i risultati di questo capitolo saranno casi particolari di teoremi della teoria astratta.

2 RICHIAMI DI GEOMETRIA EUCLIDEA NELLO SPAZIO

Lo scopo di questo paragrafo è di esplicitare i principali assiomi della geometria euclidea che useremo nel seguito, con particolare attenzione agli aspetti della geometria dello spazio che potrebbero essere poco familiari per lo studente a cui sia stata insegnata come spesso avviene solamente la geometria del piano. Introduciamo anche il concetto di *direzione* che è cruciale nella definizione dei vettori.

Assiomi di incidenza

Nello spazio euclideo i principali assiomi di incidenza sono:

- *per due punti distinti passa una e una sola retta*: questo significa che, dati due punti distinti A e B , esiste una e una sola retta a cui A e B appartengono; tale retta si denota col simbolo AB ;
- *per tre punti non allineati passa uno e un solo piano*: dati tre punti A , B e C che non appartengano a una stessa retta, esiste uno e un solo piano \mathbf{H} a cui A , B e C appartengono;
- *se una retta e un piano si intersecano in più di un punto, allora la retta è contenuta nel piano*: se A e B sono due punti distinti del piano \mathbf{H} , la retta AB è contenuta in \mathbf{H} ;

- se due piani distinti hanno un punto in comune, allora lo loro intersezione è una retta.

Si osservi che l'ultimo assioma esprime il fatto che lo spazio euclideo è tridimensionale (in uno spazio di dimensione superiore l'intersezione di due piani può consistere di un unico punto!): utilizzando questo assioma saremo in grado di individuare ogni punto dello spazio con una terna di coordinate.

Assioma delle parallele e direzione di una retta

Analizziamo le possibili posizioni reciproche di due rette. Siccome per due punti distinti passa un'unica retta, due rette distinte possono avere in comune al massimo un punto.

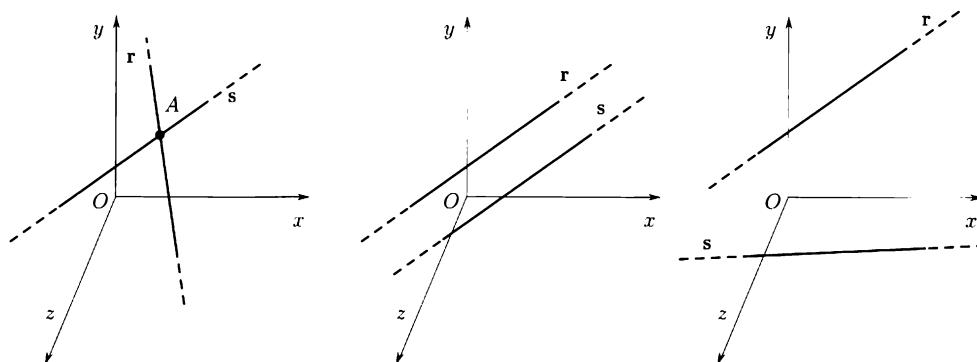


Figura 2.1.

Due rette che abbiano un punto in comune si dicono *incidenti*. Due rette incidenti **r** e **s** sono *complanari*, cioè esiste un piano che le contiene entrambe: infatti, se il punto in comune è **A** e **r** = **AB** e **s** = **AC**, allora il piano per i tre punti **A**, **B** e **C** contiene entrambe le rette. Non è detto però che due rette *complanari* siano incidenti, perché è possibile che non abbiano punti in comune. Diciamo perciò che due rette distinte sono *parallele* se sono *complanari* e non hanno punti in comune. Vi è, infine, la possibilità che le due rette non siano *complanari*, o, equivalentemente, che non siano né incidenti e né parallele: in quest'ultimo caso le due rette si dicono *sghembe*. Sarà poi necessario postulare che una retta è parallela a se stessa. Riassumendo:

Due rette distinte nello spazio possono essere:

- *incidenti* (un punto in comune, *complanari*);
- *parallele* (nessun punto in comune, ma *complanari*);
- *sghembe* (non *complanari*).

Il caso più frequente è quello che le due rette siano sghembe: infatti, quattro punti A , B , C e D presi a caso nello spazio non appartengono a un unico piano e, quindi, le due rette AB e CD sono sghembe.

Possiamo ora enunciare l'*assioma delle parallele* (in sostanza il quinto postulato di Euclide), che caratterizza la geometria euclidea e che è alla base del concetto di *vettore geometrico* perché consente di definire la *direzione* di una retta.

Assioma delle parallele

Dati una retta \mathbf{r} e un punto P , esiste un'unica retta passante per P che sia parallela a \mathbf{r} . Si dice che tale retta è ottenuta *traslando* \mathbf{r} in P .

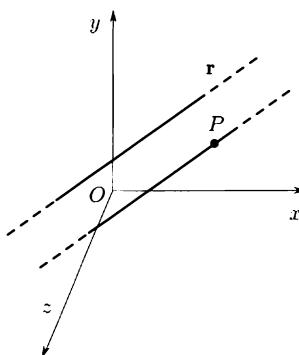


Figura 2.2. Retta per P parallela a \mathbf{r} .

Per due millenni si è cercato invano di dedurre l'assioma delle parallele dagli assiomi che lo precedono negli *Elementi*, fino a quando nel 1800 è stato formulato un modello alternativo di geometria nel quale per un punto passano infinite rette parallele a una retta data.

Utilizziamo subito l'assioma delle parallele per definire la nozione di direzione. La relazione di parallelismo è una *relazione di equivalenza*, cioè gode delle seguenti proprietà:

- *riflessiva*: una retta è parallela a se stessa;
- *simmetrica*: se \mathbf{r} è parallela a \mathbf{s} , allora \mathbf{s} è parallela a \mathbf{r} ;
- *transitiva*: se due rette \mathbf{r}_1 e \mathbf{r}_2 sono parallele a una stessa retta \mathbf{s} , allora \mathbf{r}_1 e \mathbf{r}_2 sono parallele; per dimostrare questo occorre l'assioma delle parallele (si vedano gli esercizi alla fine del paragrafo).

L'insieme delle rette dello spazio si ripartisce perciò in classi di equivalenza: due rette appartengono alla stessa classe di equivalenza se e solo se sono parallele. Una *direzione* nello spazio è per definizione una classe di equivalenza di rette parallele: in parole povere, due rette hanno la stessa direzione se sono parallele. L'assioma delle parallele si può quindi riformulare così: *per un punto passa un'unica retta con una direzione assegnata*.

Punti di una retta e numeri reali

L'intuizione ci suggerisce che i punti di una retta siano ordinati, in modo che si possa definire il segmento di estremi P e Q come l'insieme dei punti della retta che stanno *tra* P e Q . Una retta AB è l'unione di due semirette: la semiretta uscente da A passante per B consiste dei punti P per cui A non appartiene al segmento di estremi B e P ; la semiretta uscente da A che non passa per B consiste dei punti P per cui A appartiene al segmento di estremi B e P .

Inoltre un punto P su una delle semirette uscenti da A è individuato dalla lunghezza \overline{AP} del segmento di estremi A e P . Tale lunghezza, fissata come unità di misura la lunghezza \overline{AB} , sarà espressa dal numero reale non negativo $\overline{AP}/\overline{AB}$. In altre parole, per ogni numero reale positivo x assumiamo che ci siano sulla retta AB esattamente due punti a distanza x dal punto A , uno dalla parte di B , l'altro dalla parte opposta; le distanze sono calcolate rispetto all'unità di misura \overline{AB} . Possiamo ora assegnare a un numero reale x un punto $P(x)$ della retta: se $x > 0$, il punto $P(x)$ è il punto della retta a distanza x da A sulla semiretta passante per B ; se $x < 0$, il punto $P(x)$ è il punto della retta a distanza $|x|$ da A sulla semiretta che non passa per B ; infine, a $x = 0$ associamo il punto A : $P(0) = A$.

In accordo con questo discorso intuitivo, assumeremo di ottenere in questo modo una corrispondenza biunivoca, cioè che ogni punto della retta sia della forma $P(x)$ e che $P(x_1) \neq P(x_2)$ se $x_1 \neq x_2$. Il punto $P = P(x)$ è individuato così dal numero reale $x = x(P)$, che si dice *coordinata* di P rispetto al *sistema di riferimento* $\{A, B\}$ della retta. Il punto A , che corrisponde a $x = 0$, è l'*origine* delle coordinate. Riassumendo

Corrispondenza tra i punti di una retta e i numeri reali

Vi è una corrispondenza biunivoca tra l'insieme \mathbb{R} dei numeri reali e i punti della retta AB . Intuitivamente, tale corrispondenza associa a un numero reale x l'unico punto $P = P(x)$ della retta AB tale che

- a) La distanza di P da A è

$$\frac{\overline{AP}}{\overline{AB}} = |x|$$

- b) P giace sulla semiretta uscente da A e passante per B se $x > 0$; P giace sulla semiretta uscente da A che non passa per B se $x < 0$.

Ortogonalità

Due rette complanari si dicono *ortogonali* o *perpendicolari* se sono incidenti in un punto O in cui formano quattro angoli uguali e quindi retti. Nello spazio due rette \mathbf{r}_1 e \mathbf{r}_2 si dicono *ortogonali* o *perpendicolari* se sono perpendicolari le rette \mathbf{s}_1 e \mathbf{s}_2 ottenute traslando \mathbf{r}_1 e \mathbf{r}_2 in uno stesso punto O (\mathbf{s}_1 e \mathbf{s}_2 sono le rette per O parallele rispettivamente a \mathbf{r}_1 e a \mathbf{r}_2). Si può mostrare che la definizione non dipende dalla scelta del punto O : la perpendicolarità è conservata dal parallelismo.

Bisogna fare attenzione che nello spazio due rette perpendicolari possono benissimo non essere incidenti: per esempio, nello spazio cartesiano ogni retta parallela all'asse

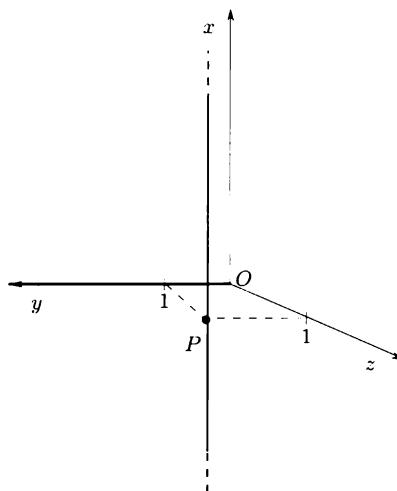


Figura 2.3. La retta di equazioni $y = z = 1$ è perpendicolare all'asse y

x è perpendicolare all'asse y , ma, se non appartiene al piano $z = 0$, non ha punti in comune con l'asse y .

Mostrare che per un punto P e una retta r non contenente P passa uno e un solo piano.
Suggerimento: si tratta del piano passante per P e due punti A e B di r .

Una retta r e un piano \mathbf{H} si dicono paralleli se non hanno punti in comune oppure se r è contenuta in \mathbf{H} . Mostrare che:

1. una retta r e un piano \mathbf{H} sono paralleli se e solo se esiste una retta s che è contenuta in \mathbf{H} e che è parallela a r ;
2. una retta r e un piano \mathbf{H} non sono paralleli se e solo se hanno esattamente un punto in comune.

Due piani \mathbf{H}_1 e \mathbf{H}_2 si dicono paralleli se non hanno punti in comune oppure coincidono. Mostrare che:

1. due piani \mathbf{H}_1 e \mathbf{H}_2 sono paralleli se e solo se l'insieme delle rette parallele a \mathbf{H}_1 coincide con l'insieme delle rette parallele a \mathbf{H}_2 : intuitivamente, due piani sono paralleli se le direzioni contenute in \mathbf{H}_1 coincidono con quelle contenute in \mathbf{H}_2 ;
2. due piani \mathbf{H}_1 e \mathbf{H}_2 non sono paralleli se e solo se l'intersezione $\mathbf{H}_1 \cap \mathbf{H}_2$ è una retta.

Si dice che una retta r e un piano \mathbf{H} sono perpendicolari se r è perpendicolare a ogni retta contenuta nel piano \mathbf{H} . Mostrare che:

1. dati una retta r e un punto P , esiste un unico piano \mathbf{H} passante per P e perpendicolare a r ; inoltre \mathbf{H} è l'unione delle infinite rette che passano per P e sono perpendicolari a r ;
2. dati un piano \mathbf{H} e un punto P , esiste un'unica retta r passante per P e perpendicolare a \mathbf{H} ;

3. mostrare che due rette perpendicolari a uno stesso piano sono parallele: le rette perpendicolari a uno stesso piano hanno quindi un'unica direzione, che si dice direzione *normale* al piano. Mostrare che due piani sono paralleli se e soltanto se hanno la stessa direzione normale;
4. mostrare che un piano \mathbf{H} e una retta \mathbf{r} sono paralleli se e solo se la direzione normale al piano è perpendicolare alla retta \mathbf{r} .

Dati una retta \mathbf{r} e un punto P , esistono infiniti piani per P paralleli a \mathbf{r} : si tratta dei piani che contengono la retta \mathbf{s} per P che è parallela a \mathbf{r} . Per parametrizzare tali piani, si fissi nel piano per P perpendicolare a \mathbf{s} la circonferenza Γ con centro in P e raggio 1. A ogni punto Q della circonferenza si associa il piano \mathbf{H}_Q che contiene Q e la retta \mathbf{s} . Si mostri che in questo modo si ottiene una corrispondenza biunivoca tra la circonferenza Γ e l'insieme dei piani per P paralleli a \mathbf{r} .

 Dati un piano \mathbf{H} e un punto P , si mostri che esistono infinite rette per P parallele a \mathbf{H} e l'unione di tali rette è l'unico piano \mathbf{H}_1 passante per P che sia parallelo a \mathbf{H} . Anche queste rette possono essere messe in corrispondenza biunivoca coi punti di una circonferenza: se Γ_1 nel piano \mathbf{H}_1 è la circonferenza di centro P e raggio 1, le rette per P parallele a \mathbf{H} sono le rette PQ che congiungono P a un punto Q della circonferenza Γ_1 .

 Dimostrare che la relazione di parallelismo per le rette dello spazio è una relazione di equivalenza.

Soluzione: il punto è mostrare la proprietà transitiva: se \mathbf{r} è parallela a \mathbf{s} ed \mathbf{s} è parallela a \mathbf{l} , allora \mathbf{r} è parallela a \mathbf{l} . Supponiamo che $\mathbf{r} \parallel \mathbf{s}$ ed $\mathbf{s} \parallel \mathbf{l}$: dobbiamo mostrare $\mathbf{r} \parallel \mathbf{l}$. Questo è ovvio se $\mathbf{r} = \mathbf{s}$. Quindi possiamo supporre che \mathbf{r} ed \mathbf{s} siano distinte, così che esiste un unico piano \mathbf{H} che contiene \mathbf{r} ed \mathbf{s} , ed $\mathbf{r} \cap \mathbf{s} = \emptyset$. Primo caso: la retta \mathbf{l} è contenuta nel piano \mathbf{H} . Allora \mathbf{r} ed \mathbf{l} sono complanari, e occorre mostrare che \mathbf{r} ed \mathbf{l} sono uguali oppure non hanno punti in comune. Se \mathbf{r} ed \mathbf{l} si intersecano in un punto P , allora \mathbf{r} ed \mathbf{l} sono due rette passanti per P e parallele a \mathbf{s} e per l'assioma delle parallele si ha $\mathbf{r} = \mathbf{l}$. Quindi \mathbf{r} ed \mathbf{l} o coincidono o non hanno alcun punto in comune, il che significa che sono parallele. Secondo caso: la retta \mathbf{l} non è contenuta nel piano \mathbf{H} , quindi esiste un punto P di \mathbf{l} che non appartiene a \mathbf{H} . Siano \mathbf{H}_r e \mathbf{H}_s passanti per P che contengono rispettivamente \mathbf{r} ed \mathbf{s} . Si noti che $\mathbf{r} = \mathbf{H}_r \cap \mathbf{H}$ perché \mathbf{H}_r e \mathbf{H} sono due piani distinti contenenti \mathbf{r} e, analogamente, $\mathbf{s} = \mathbf{H}_s \cap \mathbf{H}$. Il punto P appartiene ai due piani \mathbf{H}_r e \mathbf{H}_s , che perciò si incontrano lungo una retta $\mathbf{l}' = \mathbf{H}_r \cap \mathbf{H}_s$. La retta \mathbf{l}' non incontra il piano \mathbf{H} perché

$$\mathbf{l}' \cap \mathbf{H} = \mathbf{H}_r \cap \mathbf{H}_s \cap \mathbf{H} = (\mathbf{H}_r \cap \mathbf{H}) \cap (\mathbf{H}_s \cap \mathbf{H}) = \mathbf{r} \cap \mathbf{s} = \emptyset$$

In particolare, \mathbf{l}' non ha punti in comune con \mathbf{r} che è contenuta in \mathbf{H} . Ora \mathbf{l}' è complanare a \mathbf{r} , perché entrambe le rette giacciono in \mathbf{H}_r . Siccome \mathbf{l}' ed \mathbf{r} non hanno punti in comune e sono complanari, $\mathbf{l}' \parallel \mathbf{r}$. Lo stesso ragionamento mostra che $\mathbf{l}' \parallel \mathbf{s}$. Quindi \mathbf{l}' passa per P ed è parallela a \mathbf{s} . Ma anche la retta \mathbf{l} passa per P ed è parallela a \mathbf{s} . Per l'assioma delle parallele $\mathbf{l}' = \mathbf{l}$ e allora $\mathbf{l}' \parallel \mathbf{r}$ significa $\mathbf{l} \parallel \mathbf{r}$ come volevasi dimostrare.

 Questo esercizio descrive brevemente il modello di Poincaré per il *piano iperbolico*. In questo modello i *punti* sono i punti di un cerchio C nel piano euclideo, esclusi quelli del bordo. Le *rette* sono i diametri del cerchio e gli archi di circonferenza in C che intersecano il bordo di C perpendicolarmente. Si dimostra che, con questa definizione di punti e rette, i primi quattro assiomi euclidei sono soddisfatti (per esempio: per due punti passa una e una sola retta). Mostrare che invece l'assioma delle parallele non vale: per un punto passano infinite rette parallele a una retta data (cf. <http://math.youngzones.org/Non-Egeometry/poincare.html>).

3 DALLA GEOMETRIA ALL'ALGEBRA DEI VETTORI

Definiamo ora, nel contesto della geometria euclidea, il concetto di vettore e le operazioni di somma di vettori e prodotto di un vettore per uno scalare. Diamo largo spazio alle proprietà delle operazioni sui vettori, in primo luogo perché l'utilità dei vettori risiede proprio nel fatto che su di essi si possono effettuare delle operazioni con proprietà simili a quelle familiari dell'aritmetica: in secondo luogo, in un capitolo successivo tratteremo la teoria astratta degli spazi vettoriali, i cui assiomi sono proprio le operazioni sui vettori: questo paragrafo serve quindi anche a motivare gli assiomi della teoria astratta.

Il più semplice problema di misura che ci si può porre è quello di calcolare la distanza \overline{AB} tra due punti A e B . Per risolverlo, possiamo immaginare di avere fissato un'unità di misura per le lunghezze e di avere a disposizione uno strumento, come un righello o come il metro flessibile del sarto o semplicemente una corda, che calcoli \overline{AB} in rapporto all'unità di misura. Per misura la distanza \overline{AB} dobbiamo tendere la nostra corda tra A e B . Di qui nasce la nozione di segmento di retta: la corda tesa tra A e B . Questa nozione precede quella di retta: la retta per A e B si ottiene estendendo il segmento AB all'infinito.

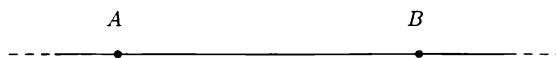


Figura 3.1. La retta AB .

Per arrivare alla nozione di vettore occorre definire il verso di una retta. Intuitivamente, un verso di una retta è uno dei due sensi in cui si può percorrere la retta. Per rendere rigoroso il concetto, si può osservare che ci sono esattamente due *ordinamenti* $<$ dei punti di una retta che rispettano la seguente proprietà: se $P < Q$ e R appartiene al segmento PQ , allora $P < R < Q$. Un verso della retta è uno di questi due possibili ordinamenti.

Un *vettore applicato* o *segmento orientato* \overrightarrow{AB} è un segmento AB insieme al *verso* della retta per A e B per cui A precede B . Assegnare un vettore applicato è la stessa cosa che assegnare una coppia ordinata di punti (A, B) , il primo estremo A e il secondo estremo B del vettore.



Figura 3.2. Il vettore \overrightarrow{AB} .

Possiamo associare a un vettore applicato \overrightarrow{AB} il suo *punto di applicazione* A , il suo *modulo* (o *lunghezza* o *norma*), la sua *direzione* e il suo *verso*. Il modulo di \overrightarrow{AB} è la lunghezza del segmento AB ; lo si denota coi simboli \overline{AB} oppure $\|\overrightarrow{AB}\|$. La direzione di \overrightarrow{AB} è la direzione della retta per A e B : due vettori \overrightarrow{AB} e \overrightarrow{CD} hanno la stessa direzione se e solo se le due rette AB e CD sono parallele. Possiamo specificare quando due

vettori \overrightarrow{AB} e \overrightarrow{CD} con la stessa direzione hanno lo stesso verso oppure verso opposto. Siccome \overrightarrow{AB} e \overrightarrow{CD} hanno la stessa direzione, le due rette AB e CD sono parallele e quindi giacciono in uno stesso piano **H**. La retta AC divide il piano in due semipiani; i vettori \overrightarrow{AB} e \overrightarrow{CD} hanno lo stesso verso se B e D appartengono allo stesso semipiano, cioè giacciono dalla stessa parte rispetto alla retta AC ; altrimenti \overrightarrow{AB} e \overrightarrow{CD} hanno verso opposto.

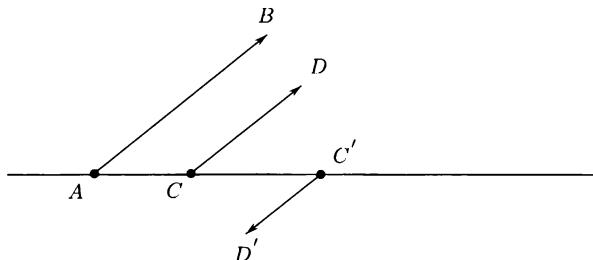


Figura 3.3. \overrightarrow{AB} e \overrightarrow{CD} hanno stesso verso, \overrightarrow{AB} e $\overrightarrow{C'D'}$ verso opposto.

L'assioma delle parallele si traduce nel fatto che *un vettore applicato è determinato da punto di applicazione, modulo, direzione e verso*: esiste uno e un solo vettore applicato con punto di applicazione, modulo, direzione e verso assegnati. Infatti supponiamo che il punto di applicazione sia A , che il modulo sia d e che la direzione sia quella della retta \mathbf{r} . Un vettore \overrightarrow{AB} ha la stessa direzione di \mathbf{r} se e solo se B giace sulla retta \mathbf{s} che passa per A ed è parallela a \mathbf{r} . Il fatto che la retta \mathbf{s} sia univocamente determinata segue dall'assioma delle parallele. Infine, sulla retta \mathbf{s} ci sono due punti B e B' la cui distanza da A è uguale al modulo assegnato e la scelta tra questi due punti è dettata dal verso assegnato.

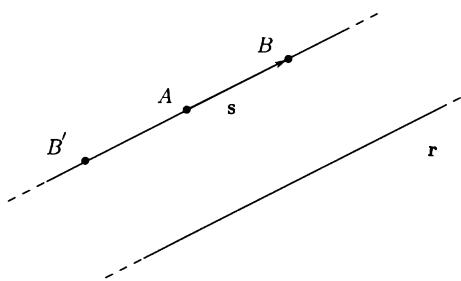


Figura 3.4. Vettore applicato in A e parallelo alla retta \mathbf{r} .

La nozione di vettore ha le sue origini in fisica, dove è usato per rappresentare, per esempio, forze e velocità. Quando un vettore rappresenta una forza meccanica che agisce su una particella, il suo punto di applicazione rappresenta la posizione della particella, il suo modulo è l'intensità della forza, la sua direzione e il suo verso sono quelli dello spostamento che la forza imprime alla particella. Quando un vettore rappresenta, invece, la velocità di una particella che percorre una certa traiettoria, il punto di applicazione indica la posizione della particella, il suo modulo è la velocità

scalare, la sua direzione è la direzione tangente alla traiettoria e il suo verso è il verso di percorrenza della traiettoria.

Un *vettore libero* è una grandezza dotata di modulo, direzione e verso. Questo significa che un vettore libero è un vettore applicato \overrightarrow{AB} di cui si dimentica il punto di applicazione. Più precisamente, diciamo che due vettori applicati \overrightarrow{AB} e \overrightarrow{CD} sono equivalenti se hanno ugual modulo, direzione e verso. In tal caso scriviamo $\overrightarrow{AB} \equiv \overrightarrow{CD}$. Identificando vettori applicati equivalenti si ottengono i vettori liberi: un *vettore libero* è una classe di equivalenza di vettori applicati, ovvero è un vettore applicato \overrightarrow{AB} insieme a tutti i vettori \overrightarrow{CD} equivalenti a \overrightarrow{AB} .

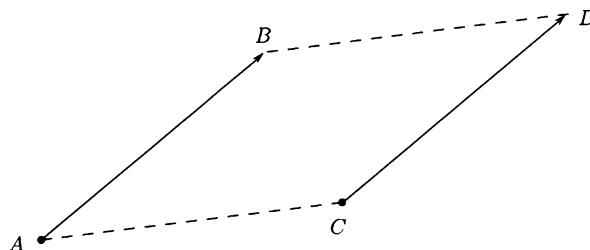


Figura 3.5. Vettori applicati equivalenti.

Non è facile pensare a un vettore come a un insieme di vettori applicati tutti equivalenti tra loro. Può aiutare un esempio a cui siamo tutti abituati dai primi anni di scuola: quello dei numeri razionali. Un numero razionale è infatti una frazione $\frac{a}{b}$ insieme a tutte le frazioni $\frac{c}{d}$ equivalenti ad $\frac{a}{b}$, nel senso che $ad - bc = 0$. Le due frazioni $\frac{1}{2}$ e $\frac{2}{4}$ rappresentano lo stesso numero razionale, così come due vettori applicati con lo stesso modulo, direzione e verso, rappresentano lo stesso vettore libero.

D'ora innanzi per *vettore* intenderemo un *vettore libero* e denoteremo i vettori con lettere minuscole in grassetto; useremo il simbolo $\|\mathbf{v}\|$ per indicare il modulo di \mathbf{v} . Designeremo l'insieme dei vettori liberi col simbolo \mathbf{V} .

Scriviamo $\mathbf{v} \equiv \overrightarrow{AB}$ per indicare che il vettore libero \mathbf{v} è rappresentato dal vettore applicato \overrightarrow{AB} : il modulo, la direzione e il verso di \mathbf{v} sono quelli di \overrightarrow{AB} . La relazione si può leggere in senso inverso per ritrovare il vettore applicato a partire dal vettore libero \mathbf{v} e dal punto di applicazione A . Si dice che B è ottenuto traslando A di \mathbf{v} e si scrive anche $\mathbf{v} = B - A$ per motivi che diventeranno chiari più avanti quando definiremo la differenza di due vettori.

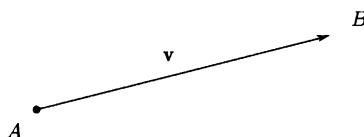


Figura 3.6. Il punto B è ottenuto traslando A del vettore \mathbf{v} .

In questo modo si associa a un vettore \mathbf{v} una trasformazione dello spazio, che si dice appunto *traslazione* del vettore \mathbf{v} e che denoteremo col simbolo $\tau_{\mathbf{v}}$: per definizione

$\tau_{\mathbf{v}}(A) = B$ significa $\mathbf{v} \equiv \overrightarrow{AB}$. Come vuole l'intuizione, una traslazione è una trasformazione dello spazio che sposta tutti i punti di una distanza fissa nella stessa direzione e verso: la distanza, direzione e verso specificati da \mathbf{v} . Si noti che un vettore \mathbf{v} è determinato dalla traslazione $\tau_{\mathbf{v}}$: basta considerare il traslato $B = \tau_{\mathbf{v}}(A)$ di un punto arbitrario A per ricostruire \mathbf{v} , perché $\mathbf{v} \equiv \overrightarrow{AB}$. Quindi possiamo identificare un vettore \mathbf{v} con la traslazione $\tau_{\mathbf{v}}$, visualizzando così immediatamente tutti i vettori applicati che rappresentano \mathbf{v} .

Una volta fissato un punto O nello spazio, detto origine, si possono identificare:

1. l'insieme dei punti P dello spazio;
2. l'insieme dei vettori \overrightarrow{OP} applicati in O ;
3. l'insieme dei vettori liberi \mathbf{v} .

Il punto P si identifica col vettore \overrightarrow{OP} , che si dice *vettore posizione* del punto P rispetto all'origine O . Siccome il punto di applicazione O è fissato, un vettore applicato \overrightarrow{OP} è individuato dal suo modulo, direzione e verso, cioè dal corrispondente vettore libero \mathbf{v} .

Assegnare un punto P è equivalente ad assegnare il suo vettore posizione \overrightarrow{OP} . Il vantaggio nell'usare i vettori è che sui vettori è possibile definire in modo naturale delle operazioni algebriche, la somma e il prodotto per scalare, per le quali valgono formalmente le principali regole di calcolo dell'aritmetica; mentre questo non è possibile sui punti.

Somma di vettori

Consideriamo due vettori applicati \overrightarrow{OP} e \overrightarrow{PQ} tali che il secondo estremo del primo vettore coincida con il punto di applicazione del secondo vettore. La loro somma è per definizione il vettore \overrightarrow{OQ} :

$$(3.1) \quad \overrightarrow{OQ} = \overrightarrow{OP} + \overrightarrow{PQ}$$

Procediamo a definire la somma di due vettori liberi qualsiasi \mathbf{v} e \mathbf{w} . Fissiamo un punto O . Sia $\overrightarrow{OP} \equiv \mathbf{v}$ il vettore \mathbf{v} applicato in O e sia $\overrightarrow{PQ} \equiv \mathbf{w}$ il vettore \mathbf{w} applicato in P . La *somma* $\mathbf{v} + \mathbf{w}$ di \mathbf{v} e \mathbf{w} è per definizione il vettore libero rappresentato da \overrightarrow{OQ} .

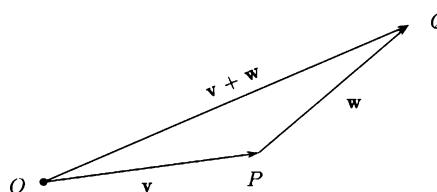


Figura 3.7. Somma di due vettori.

Naturalmente occorre verificare che la definizione non dipende dal punto di applicazione O scelto per \mathbf{v} . Se al posto di O scegliamo un altro punto A e $\mathbf{v} \equiv \overrightarrow{AB}$ e $\mathbf{w} \equiv \overrightarrow{BC}$, occorre verificare che i vettori OQ e AC abbiano lo stesso modulo, direzione e verso

e quindi definiscono lo stesso vettore libero $\mathbf{v} + \mathbf{w}$. Per questo basta osservare che i due triangoli $\triangle OPQ$ e $\triangle ABC$ sono congruenti.

Si noti che per la somma di vettori si usa lo stesso simbolo $+$ dell'aritmetica, nonostante si tratti di un'operazione sui vettori e non sui numeri. La giustificazione è che, come vedremo più avanti, la somma di vettori gode delle stesse proprietà formali della somma tra numeri.

La somma dei vettori \mathbf{v} e \mathbf{w} si può anche definire con la *regola del parallelogramma*, che permette di definire la somma di vettori applicati in uno *stesso* punto O . Dati due vettori \overrightarrow{OP} e \overrightarrow{OR} , si costruisce il parallelogramma che ha tre vertici in O , P e R . Detto Q il quarto vertice del parallelogramma, si pone

$$\overrightarrow{OP} + \overrightarrow{OR} = \overrightarrow{OQ}$$

quindi il vettore somma è la diagonale del parallelogramma uscente da O .

La regola del parallelogramma è suggerita dalla fisica: se i vettori \overrightarrow{OP} e \overrightarrow{OR} rappresentano due forze agenti simultaneamente su una particella situata nel punto O , il vettore somma $\overrightarrow{OP} + \overrightarrow{OR}$ rappresenta la forza risultante. La costruzione del parallelogramma tecnicamente richiede che i tre punti O , P e R non siano allineati; a questo problema si può rimediare facilmente, sostituendo R con un punto R' molto vicino a R ma non appartenente alla retta OP e facendo poi tendere R' a R .

Mostriamo che la regola del parallelogramma è coerente con la definizione precedente di somma di vettori liberi. Se \mathbf{v} e \mathbf{w} sono due vettori liberi, possiamo applicarli in O ottenendo $\mathbf{v} \equiv \overrightarrow{OP}$ e $\mathbf{w} \equiv \overrightarrow{OR}$. Il vettore \overrightarrow{PQ} rappresenta anch'esso \mathbf{w} perché ha uguale modulo, direzione e verso di \overrightarrow{OR} . Quindi la diagonale \overrightarrow{OQ} del parallelogramma rappresenta il vettore somma $\mathbf{v} + \mathbf{w}$ come definito precedentemente.

Somma di vettori come composizione di traslazioni

La traslazione $\tau_{\mathbf{v}+\mathbf{w}}$ definita dal vettore somma $\mathbf{v} + \mathbf{w}$ è la traslazione $\tau_{\mathbf{v}}$ seguita dalla traslazione $\tau_{\mathbf{w}}$. Questo è immediato dalla definizione, si veda la figura 3.7: il punto O viene traslato da \mathbf{v} in P e il punto P viene traslato da \mathbf{w} in Q ; ma il punto Q si ottiene anche traslando O del vettore somma $\mathbf{v} + \mathbf{w} \equiv \overrightarrow{OQ}$. In simboli:

$$(3.2) \quad \tau_{\mathbf{w}} \circ \tau_{\mathbf{v}} = \tau_{\mathbf{v}+\mathbf{w}}$$

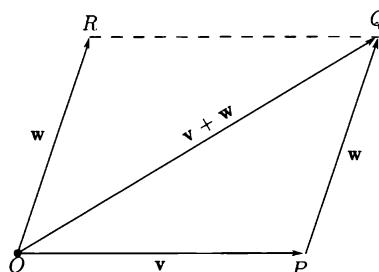


Figura 3.8. La regola del parallelogramma.

dove il simbolo \circ denota come al solito la composizione delle due funzioni τ_v e τ_w , cioè la funzione definita applicando prima τ_v e poi τ_w :

$$\tau_w \circ \tau_v(O) = \tau_w(\tau_v(O))$$

Proprietà della somma di vettori

Sull'insieme \mathbf{V} dei vettori liberi abbiamo definito l'operazione di somma. Rispetto a tale operazione, \mathbf{V} è un *gruppo abeliano*: questo significa che la somma di vettori gode delle stesse proprietà formali della somma di numeri e precisamente:

V1 Proprietà commutativa La somma di vettori è commutativa: per ogni coppia di vettori v e w vale l'uguaglianza

$$(3.3) \quad \mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$$

Questo segue dal fatto che nel parallelogramma della figura 3.8 la diagonale \overrightarrow{OQ} rappresenta sia $\mathbf{v} + \mathbf{w}$ sia $\mathbf{w} + \mathbf{v}$ perché $\mathbf{v} \equiv \overrightarrow{OP} \equiv \overrightarrow{RQ}$ e $\mathbf{w} \equiv \overrightarrow{OR} \equiv \overrightarrow{PQ}$.

Poiché $\tau_w \circ \tau_v = \tau_{v+w}$, il fatto che la somma di vettori sia commutativa significa che le traslazioni commutano: traslare prima del vettore v e poi del vettore w dà lo stesso risultato che traslare prima di w e poi del vettore v .

V2 Proprietà associativa La somma è un'operazione associativa: questo significa che

$$(3.4) \quad (\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$$

comunque si scelgano i vettori u , v e w . Possiamo perciò scrivere $\mathbf{u} + \mathbf{v} + \mathbf{w}$ al posto di $(\mathbf{u} + \mathbf{v}) + \mathbf{w}$ e $\mathbf{u} + (\mathbf{v} + \mathbf{w})$.

Per dimostrare la proprietà associativa con una costruzione geometrica, applichiamo i tre vettori in un punto O : $\mathbf{u} \equiv \overrightarrow{OP}$, $\mathbf{v} \equiv \overrightarrow{OQ}$ e $\mathbf{w} \equiv \overrightarrow{OR}$.

Quindi costruiamo come in figura il parallelepipedo che ha come spigoli uscenti dal vertice O i segmenti OP , OQ e OR . Sia S il vertice del parallelepipedo opposto a O . Si verifica che il vettore \overrightarrow{OS} rappresenta tanto $(\mathbf{u} + \mathbf{v}) + \mathbf{w}$ quanto $\mathbf{u} + (\mathbf{v} + \mathbf{w})$. A rigore, questa dimostrazione richiede che i quattro punti O , P , Q e R non stiano

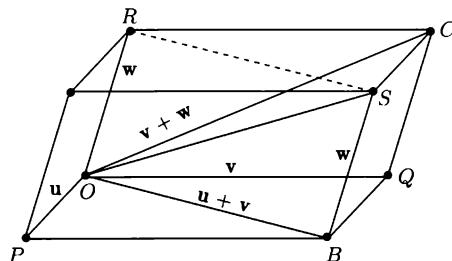


Figura 3.9. $\overrightarrow{OS} = \overrightarrow{OB} + \overrightarrow{BS} \equiv (\mathbf{u} + \mathbf{v}) + \mathbf{w}$; $\overrightarrow{OS} = \overrightarrow{OC} + \overrightarrow{CS} \equiv \mathbf{u} + (\mathbf{v} + \mathbf{w})$.

tutti su uno stesso piano, ma il caso *degenero* in cui i quattro punti appartengono a uno stesso piano si può dedurre da quello non degenero con un processo di limite.

Possiamo dare una dimostrazione più semplice e convincente, ma meno intuitiva, sfruttando l'interpretazione della somma di vettori come composizione delle corrispondenti traslazioni. Il punto è che l'operazione di composizione di funzioni è evidentemente associativa: per ogni terna di funzioni f , g , h vale l'uguaglianza $(h \circ g) \circ f = h \circ (g \circ f)$ perché le funzioni ai due membri associano entrambe a un elemento x l'elemento

$$h(g(f(x)))$$

Applicando questa uguaglianza alle traslazioni definite da \mathbf{u} , \mathbf{v} e \mathbf{w} otteniamo

$$(3.5) \quad (\tau_{\mathbf{w}} \circ \tau_{\mathbf{v}}) \circ \tau_{\mathbf{u}} = \tau_{\mathbf{w}} \circ (\tau_{\mathbf{v}} \circ \tau_{\mathbf{u}})$$

Tenendo conto che $\tau_{\mathbf{v}+\mathbf{w}} = \tau_{\mathbf{w}} \circ \tau_{\mathbf{v}}$ per (3.2), l'uguaglianza (3.5) è equivalente alla (3.4) che risulta così dimostrata.

V3 Esistenza dell'elemento neutro Il *vettore nullo*, denotato col simbolo $\mathbf{0}$, è l'unico vettore che ha modulo nullo. Se si pensa a un vettore applicato \vec{AB} come a una coppia ordinata di punti, il vettore nullo è rappresentato dal caso degenero in cui $A = B$. La direzione e il verso del vettore nullo non sono definiti, ma per convenzione e per coerenza con l'algebra che svilupperemo, diciamo che il vettore nullo è parallelo a ogni altro vettore, come se avesse ogni direzione. Se pensiamo ai vettori come traslazioni, il vettore nullo corrisponde alla traslazione che lascia ogni punto al suo posto. Da questo segue immediatamente che $\mathbf{0}$ è un *elemento neutro* per la somma di vettori:

$$(3.6) \quad \mathbf{v} + \mathbf{0} = \mathbf{v} \quad \text{per ogni } \mathbf{v} \in \mathbf{V}$$

V4 Esistenza dell'opposto di un vettore dato Per ogni vettore \mathbf{v} esiste uno e un solo vettore $-\mathbf{v}$, detto *opposto* di \mathbf{v} , tale che

$$(3.7) \quad \mathbf{v} + (-\mathbf{v}) = \mathbf{0}$$

Per la proprietà commutativa (3.7) è equivalente a $-\mathbf{v} + \mathbf{v} = \mathbf{0}$.

Per verificare l'esistenza dell'opposto, osserviamo che, se $\mathbf{v} \equiv \vec{AB}$, il vettore $-\mathbf{v} \equiv \vec{BA}$ sommato a \mathbf{v} dà come risultato il vettore nullo.

Per quanto riguarda l'unicità, osserviamo che, perché possa valere (3.6), il vettore $-\mathbf{v}$ deve avere modulo e direzione uguali a quelli di \mathbf{v} e verso opposto; quindi $-\mathbf{v}$ è univocamente determinato da \mathbf{v} .

Mostriamo di nuovo l'unicità dell'opposto come conseguenza dall'esistenza di *un* vettore opposto e delle proprietà formali V1-V2-V3. Cominciamo con l'osservare che l'esistenza di *un* vettore opposto insieme alle proprietà V1-V2-V3 implica la:

Legge di cancellazione della somma

Dall'uguaglianza

$$(3.8) \quad \mathbf{v} + \mathbf{w}_1 = \mathbf{v} + \mathbf{w}_2$$

segue $\mathbf{w}_1 = \mathbf{w}_2$.

Per dimostrare la legge di cancellazione, sommiamo $-\mathbf{v}$ a entrambi i membri di (3.8) ottenendo:

$$(3.9) \quad -\mathbf{v} + (\mathbf{v} + \mathbf{w}_1) = -\mathbf{v} + (\mathbf{v} + \mathbf{w}_2)$$

Per la proprietà associativa il primo membro di (3.9) è uguale a

$$(-\mathbf{v} + \mathbf{v}) + \mathbf{w}_1 = \mathbf{0} + \mathbf{w}_1 = \mathbf{w}_1$$

Per lo stesso motivo il membro di destra di (3.9) è uguale a \mathbf{w}_2 , per cui $\mathbf{w}_1 = \mathbf{w}_2$ e la legge di cancellazione è dimostrata.

Possiamo ora mostrare che il vettore opposto di \mathbf{v} è univocamente determinato da \mathbf{v} . Infatti, se \mathbf{w}_1 e \mathbf{w}_2 sono due vettori opposti di \mathbf{v} , allora

$$\mathbf{v} + \mathbf{w}_1 = \mathbf{0} = \mathbf{v} + \mathbf{w}_2$$

e dalla legge di cancellazione segue $\mathbf{w}_1 = \mathbf{w}_2$.

La *differenza* di due vettori è definita ponendo

$$\mathbf{v} - \mathbf{w} = \mathbf{v} + (-\mathbf{w})$$

Naturalmente $\mathbf{w} + (\mathbf{v} - \mathbf{w}) = \mathbf{v}$: analogamente al caso dell'aritmetica, la differenza di \mathbf{v} e \mathbf{w} è quel vettore che sommato a \mathbf{w} dà come risultato \mathbf{v} .

Si noti che, comunque si scelgano tre punti O , P e Q , vale l'uguaglianza

$$\overrightarrow{QP} \equiv \overrightarrow{OP} - \overrightarrow{OQ}$$

come illustrato nella figura 3.10. In altre parole, comunque si scelga l'origine O , la differenza tra il vettore posizione di P e il vettore posizione di Q ha il modulo, direzione e verso del vettore \overrightarrow{QP} . Per questo spesso si scrive $\overrightarrow{QP} = P - Q$. Mentre non è possibile definire la somma di due punti senza fissare un'origine, la differenza di due punti è un vettore ben definito.

Prodotto di un vettore per uno scalare

In aritmetica, dopo la somma di due numeri, si introduce l'altra operazione fondamentale, il prodotto. A differenza della somma, non esiste un perfetto analogo del prodotto nel caso dei vettori, ma è possibile definire un'operazione, il *prodotto di un*

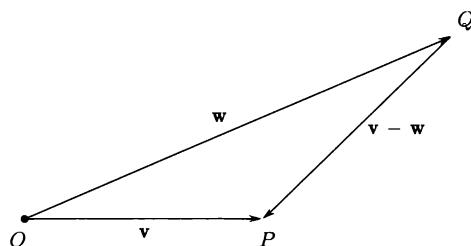


Figura 3.10. Differenza di due vettori.

vettore per uno scalare, che conserva molte delle proprietà del prodotto di due numeri. Il termine *scalare* in questo capitolo significa *numero reale*. Tradizionalmente, uno scalare denota una grandezza descritta da un solo numero, in opposizione a un vettore, che è una grandezza definita da più parametri. Dati uno scalare $t \in \mathbb{R}$ e un vettore $\mathbf{v} \in \mathbf{V}$, il prodotto del vettore \mathbf{v} per lo scalare t è il vettore $t\mathbf{v}$ definito come segue:

- i) il modulo di $t\mathbf{v}$ è il modulo di \mathbf{v} moltiplicato per $|t|$. In simboli:

$$\|t\mathbf{v}\| = |t| \|\mathbf{v}\|$$

- ii) la direzione di $t\mathbf{v}$ coincide con quella di \mathbf{v} :

- iii) il verso di $t\mathbf{v}$ è quello di \mathbf{v} se $t > 0$, è il verso opposto se $t < 0$.

Se $t = 0$, il modulo di $0\mathbf{v}$ è nullo qualunque sia \mathbf{v} , quindi $0\mathbf{v} = \mathbf{0}$ per ogni \mathbf{v} . Se $t \neq 0$, possiamo *dividere* un vettore \mathbf{v} per lo scalare t : per definizione questo significa moltiplicare \mathbf{v} per l'inverso di t . In simboli:

$$\frac{\mathbf{v}}{t} = \frac{1}{t}\mathbf{v} \quad (t \neq 0).$$

Naturalmente, è possibile definire il prodotto di un *vettore applicato* \overrightarrow{OP} per uno scalare t : il punto di applicazione di $t\overrightarrow{OP}$ è O mentre, modulo, direzione e verso sono definiti come nel caso dei vettori liberi.

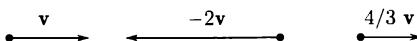


Figura 3.11. Moltiplicazione di \mathbf{v} per -2 e per $4/3$.

Le seguenti proprietà del prodotto di un vettore per uno scalare sono altrettanto fondamentali di quelle della somma di vettori: come vedremo nelle dimostrazioni, esse traducono in termini algebrici la teoria euclidea della congruenza e della proporzionalità.

V5 Proprietà distributiva del prodotto per scalare rispetto alla somma di vettori:

$$t(\mathbf{v} + \mathbf{w}) = t\mathbf{v} + t\mathbf{w} \quad \text{per ogni } t \in \mathbb{R} \text{ e ogni } \mathbf{v}, \mathbf{w} \in \mathbf{V}$$

V6 Proprietà distributiva del prodotto per uno scalare rispetto alla somma di scalari:

$$(s+t)\mathbf{v} = s\mathbf{v} + t\mathbf{v} \quad \text{per ogni } s, t \in \mathbb{R} \text{ e ogni } \mathbf{v} \in \mathbf{V}$$

V7 Proprietà associativa del prodotto per scalare:

$$s(t\mathbf{v}) = (st)\mathbf{v} \quad \text{per ogni } s, t \in \mathbb{R} \text{ e ogni } \mathbf{v} \in \mathbf{V}$$

V8 Normalizzazione del prodotto per scalare:

$$1\mathbf{v} = \mathbf{v} \quad \text{per ogni } \mathbf{v} \in V$$

Diamo ora una dimostrazione geometrica delle proprietà **V5** e **V6**.

Per dimostrare **V5**, consideriamo, come nella figura 3.12, due triangoli ABC e DEF tali che $\mathbf{v} \equiv \overrightarrow{AB}$, $\mathbf{w} \equiv \overrightarrow{BC}$, $t\mathbf{v} \equiv \overrightarrow{DE}$ e $t\mathbf{w} \equiv \overrightarrow{EF}$. Dobbiamo far vedere che il vettore $t(\mathbf{v} + \mathbf{w}) \equiv \overrightarrow{tAC}$ coincide col vettore $t\mathbf{v} + t\mathbf{w} \equiv \overrightarrow{DF}$.

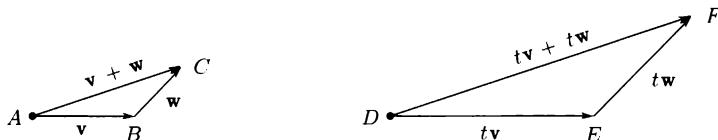


Figura 3.12. La proprietà distributiva $t(\mathbf{v} + \mathbf{w}) = t\mathbf{v} + t\mathbf{w}$.

Per costruzione le rette AB e DE sono parallele, così come le rette BC e EF . Quindi i due angoli ABC e DEF sono uguali. Inoltre

$$\overline{DE} : \overline{AB} = \overline{EF} : \overline{BC} = |t|.$$

Da Euclide Libro VI proposizione 6 segue che i due triangoli ABC e DEF sono simili. Questo implica che le due rette AC e DF sono parallele e che

$$\overline{DF} : \overline{AC} = |t|$$

Questo mostra che \overrightarrow{DF} e \overrightarrow{AC} hanno la stessa direzione, e che il rapporto dei loro moduli è uguale a $|t|$. Quindi i vettori $t\overrightarrow{AC}$ e \overrightarrow{DF} hanno lo stesso modulo e la stessa direzione, ed è evidente che hanno anche lo stesso verso. Concludiamo che $t\overrightarrow{AC} \equiv \overrightarrow{DF}$ come volevasi dimostrare.

Mostriamo ora la proprietà distributiva del prodotto per scalare rispetto alla somma di scalari:

$$(s+t)\mathbf{v} = s\mathbf{v} + t\mathbf{v} \quad \text{per ogni } s, t \in \mathbb{R} \text{ e ogni } \mathbf{v} \in V$$

Supponiamo per semplicità che s e t siano entrambi positivi. In questo caso $(s+t)\mathbf{v}$ e $s\mathbf{v} + t\mathbf{v}$ hanno la stessa direzione e verso di \mathbf{v} , per cui esistono tre punti allineati A , B e C , con B appartenente al segmento AC , tali che $s\mathbf{v} \equiv \overrightarrow{AB}$ e $t\mathbf{v} \equiv \overrightarrow{BC}$, come illustrato in figura 3.13. Ora $\overline{AC} = \overline{AB} + \overline{BC}$ perché B appartiene al segmento AC : questo è il fatto geometrico che stiamo traducendo in algebra nella forma della proprietà **V6**. Quindi

$$\overline{AC} = \overline{AB} + \overline{BC} = \|s\mathbf{v}\| + \|t\mathbf{v}\|$$



Figura 3.13. La proprietà distributiva $(s+t)\mathbf{v} = s\mathbf{v} + t\mathbf{v}$.

Ora, per definizione del prodotto per scalare, $\|sv\| = |s| \|\mathbf{v}\|$ e $\|tv\| = |t| \|\mathbf{v}\|$. Usando la proprietà distributiva dell'aritmetica e il fatto che $|s| + |t| = |s + t|$ perché $s, t > 0$, troviamo

$$\overrightarrow{AC} = |s + t| \|\mathbf{v}\| = \|(s + t)\mathbf{v}\|$$

Quindi, i vettori \overrightarrow{AC} e $(s+t)\mathbf{v}$ hanno lo stesso modulo; hanno anche la stessa direzione e verso di \mathbf{v} , per cui $\overrightarrow{AC} \equiv (s+t)\mathbf{v}$. Ma

$$\overrightarrow{AC} \equiv \overrightarrow{AB} + \overrightarrow{BC} \equiv s\mathbf{v} + t\mathbf{v}$$

per definizione di somma di vettori. Quindi i vettori $(s + t)\mathbf{v}$ e $s\mathbf{v} + t\mathbf{v}$ sono uguali perché sono entrambi rappresentati da \overrightarrow{AC} e questo conclude la dimostrazione nel caso $s, t > 0$.

Nel caso in cui $s > 0, t < 0$ e $s + t > 0$ la dimostrazione è simile, con la differenza che ora il punto C si trova tra A e B , per cui $\overrightarrow{AC} = \overrightarrow{AB} - \overrightarrow{BC}$. Ma allora

$$\overrightarrow{AC} = |s| \|\mathbf{v}\| - |t| \|\mathbf{v}\| = |s + t| \|\mathbf{v}\| = \|(s + t)\mathbf{v}\|$$

e si può poi concludere come nel caso precedente. Gli altri casi possibili seguono facilmente dai due che abbiamo esaminati.

Le due proprietà **V7** e **V8** sono conseguenza immediata della definizione del prodotto per uno scalare: la proprietà **V7** segue dal fatto che due vettori sono uguali se hanno uguale modulo, direzione e verso, tenuto conto della regola di moltiplicazione dei segni e del fatto che il modulo del prodotto di due numeri è uguale al prodotto dei moduli; la proprietà **V8** vale perché il prodotto di un vettore per lo scalare 1 lascia invariati il modulo, la direzione e il verso del vettore.

Visto che queste proprietà sono evidenti, perché ci siamo dati la pena di enunciarle? Il motivo è in realtà importante: nella teoria astratta degli spazi vettoriali i concetti primitivi di *punti*, *rette* e *piani* vengono sostituiti dai *vettori* e gli assiomi euclidei dagli assiomi **V1-V8**, insieme alla richiesta che un vettore si possa applicare in un unico modo in un punto dato. È poi possibile definire rette e piani, sostanzialmente come faremo nel prossimo paragrafo e *dedurre* dalle proprietà dei vettori gli assiomi della geometria euclidea. Nella matematica moderna la geometria euclidea viene così riformulata in termini algebrici. Il motivo è che gli assiomi nella forma algebrica **V1-V8** sono di molto più facile e ampio utilizzo rispetto agli assiomi euclidei. Si tratta di un metodo molto efficace: in sostanza si rimpiazzano complicate costruzioni geometriche con semplici equazioni. Inoltre, la teoria degli spazi vettoriali si applica con successo non solo ai vettori geometrici, ma a tutti quegli insiemi in cui è possibile definire somma e prodotto per scalare: per esempio anche all'insieme delle funzioni di una variabile reale, permettendo l'uso di idee e tecniche geometriche nella teoria delle funzioni e delle equazioni differenziali.



Vediamo qualche esempio di proprietà che si possono dedurre formalmente da **V1-V8**.

Abbiamo già osservato che $0\mathbf{v} = \mathbf{0}$ per ogni vettore \mathbf{v} , usando la definizione di prodotto per scalare. Si può dedurre questa uguaglianza formalmente dalle proprietà **V1-V8**: infatti

$$\mathbf{v} + 0\mathbf{v} = 1\mathbf{v} + 0\mathbf{v} = (1+0)\mathbf{v} = 1\mathbf{v} = \mathbf{v} = \mathbf{v} + \mathbf{0}$$

e quindi $0\mathbf{v} = \mathbf{0}$ per la legge di cancellazione della somma.

Analogamente, $t\mathbf{0} = \mathbf{0}$ per ogni scalare t : infatti per ogni vettore \mathbf{v}

$$t\mathbf{0} = t(\mathbf{0} + \mathbf{0}) = t\mathbf{0} + t\mathbf{0}$$

e sommando $-t\mathbf{0}$ otteniamo $\mathbf{0} = t\mathbf{0}$.

Vale anche la

Legge di annullamento del prodotto per scalare.

Se $t\mathbf{v} = \mathbf{0}$, allora $t = 0$ oppure $\mathbf{v} = \mathbf{0}$.

Questo segue dal fatto che $\|t\mathbf{v}\| = |t|\|\mathbf{v}\|$. Ma è anche una conseguenza delle proprietà **V1-V8**: infatti, se $t\mathbf{v} = \mathbf{0}$, ma $t \neq 0$, allora

$$\mathbf{0} = t^{-1}\mathbf{0} = t^{-1}(t\mathbf{v}) = (t^{-1}t)\mathbf{v} = 1\mathbf{v} = \mathbf{v}$$

Il vettore opposto $-\mathbf{v}$ coincide col prodotto $(-1)\mathbf{v}$ del vettore \mathbf{v} per lo scalare -1 . Questo si può vedere geometricamente, in quanto entrambi i vettori hanno lo stesso modulo e la stessa direzione di \mathbf{v} , ma verso opposto. Ma lo si può vedere anche come conseguenza formale delle proprietà **V1-V8**. Infatti

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

4 SISTEMI DI RIFERIMENTO E COORDINATE

Introduciamo sistemi di riferimento in un piano e nello spazio. Questo consente di identificare punti e vettori con terne di numeri reali e riduce le operazioni sui vettori a operazioni sui numeri reali; l'algebra dei vettori diviene così un vero strumento di calcolo.

4.1 Sistema di riferimento in un piano

Due vettori si dicono paralleli se hanno la stessa direzione; per definizione il vettore nullo è parallelo a ogni altro vettore. Poiché abbiamo assunto che i punti di una retta siano in corrispondenza biunivoca con i numeri reali, due vettori paralleli differiscono solo per il prodotto per uno scalare. Precisamente:

Sia \mathbf{v} un vettore non nullo e sia \mathbf{w} un vettore parallelo a \mathbf{v} . Allora esiste un unico scalare $t \in \mathbb{R}$ tale che $\mathbf{w} = t\mathbf{v}$.

In effetti, t è lo scalare che ha modulo uguale al rapporto dei moduli di \mathbf{w} e \mathbf{v} e il segno positivo o negativo a seconda che \mathbf{v} e \mathbf{w} abbiano verso uguale oppure opposto. In ogni caso

$$\mathbf{w} = \pm \frac{\|\mathbf{w}\|}{\|\mathbf{v}\|} \mathbf{v}$$

e $t = \pm \frac{\|\mathbf{w}\|}{\|\mathbf{v}\|}$. Lo scalare $t = 0$ corrisponde a $\mathbf{w} = \mathbf{0}$.

Consideriamo ora una retta \mathbf{r} passante per un punto O e parallela a un vettore \mathbf{v} . Se P è un punto della retta, il vettore \overrightarrow{OP} è diretto come \mathbf{r} e quindi è parallelo a \mathbf{v} .

Per quanto abbiamo appena osservato, esiste un unico scalare t tale che

$$(4.1) \quad \overrightarrow{OP} \equiv t\mathbf{v}$$

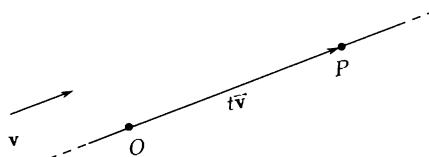


Figura 4.1. Retta per O parallela a \mathbf{v} .

Il valore assoluto di t rappresenta la lunghezza $\|\overrightarrow{OP}\|$ se come unità di misura si prende il modulo di \mathbf{v} , perché

$$|t| = \frac{\|\overrightarrow{OP}\|}{\|\mathbf{v}\|}$$

D'altra parte, il segno di t è positivo o negativo a che seconda che \overrightarrow{OP} abbia lo stesso verso di \mathbf{v} oppure verso opposto. In sostanza, t indica di quanto dobbiamo spostarci da O nella direzione di \mathbf{v} se vogliamo arrivare in P .

Possiamo ora introdurre delle coordinate nel piano \mathbf{H} passante per tre punti non allineati O , A e B . L'idea è che ogni punto P del piano si può ritrovare a partire da O mediante uno spostamento, specificato dalla coordinata x , nella direzione del vettore $\mathbf{u} \equiv \overrightarrow{OA}$, seguito da uno spostamento, specificato dalla coordinata y , nella direzione del vettore $\mathbf{v} \equiv \overrightarrow{OB}$. Si pensi a una cartina geografica in cui una località è individuata dalla sua latitudine e dalla sua longitudine.

Il punto O sarà l'origine delle coordinate, le rette OA e OB gli assi del sistema di riferimento. Per ottenere le coordinate occorre fissare un'unità di misura e un verso su ciascuno degli assi e, quindi, un vettore \mathbf{u} parallelo a OA e un vettore \mathbf{v} parallelo a OB : la scelta naturale è di prendere $\mathbf{u} \equiv \overrightarrow{OA}$, il vettore posizione di A rispetto all'origine, e $\mathbf{v} \equiv \overrightarrow{OB}$. Possiamo allora sostituire i punti A e B coi loro vettori posizione \mathbf{u} e \mathbf{v} . Tenendo conto che O , A e B non sono allineati se e solo se \mathbf{u} e \mathbf{v} non sono paralleli, diamo la seguente definizione:

Sistema di riferimento in un piano

Un *sistema di riferimento* in un piano \mathbf{H} è una terna $(O, \mathbf{u}, \mathbf{v})$ dove:

- a) O è un punto di \mathbf{H} , detto *origine* delle coordinate;
- b) \mathbf{u} e \mathbf{v} sono due vettori paralleli al piano \mathbf{H} e non paralleli tra loro. Questo significa che le rette per O dirette come \mathbf{u} e \mathbf{v} sono contenute nel piano \mathbf{H} e sono distinte: tali rette si dicono *assi del sistema di riferimento* e si dice che $\{\mathbf{u}, \mathbf{v}\}$ è una *base* del piano \mathbf{H} .

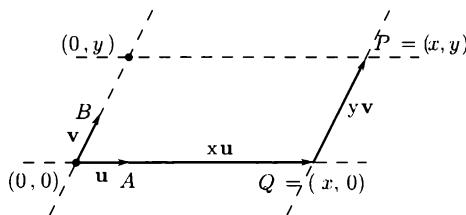


Figura 4.2. Coordinate nel piano.

Possiamo ora determinare le coordinate di un punto P del piano rispetto al sistema di riferimento $(O, \mathbf{u}, \mathbf{v})$ come illustrato dalla figura 4.2. Sia \mathbf{r} la retta per P parallela a \mathbf{v} ; siccome \mathbf{v} non è parallelo a \mathbf{u} , la retta \mathbf{r} interseca l'asse OA in un punto Q .

Siccome \overrightarrow{OQ} è parallelo a \mathbf{u} , esiste un unico scalare x tale che $\overrightarrow{OQ} \equiv x\mathbf{u}$. Analogamente, poiché \overrightarrow{QP} è parallelo a \mathbf{v} , esiste un unico scalare $y \in \mathbb{R}$ tale che

$$\overrightarrow{QP} \equiv y\mathbf{v}$$

Quindi il vettore posizione di P rispetto all'origine O è

$$(4.2) \quad \overrightarrow{OP} = \overrightarrow{OQ} + \overrightarrow{QP} \equiv x\mathbf{u} + y\mathbf{v}.$$

Poniamo $x(P) = x$ e $y(P) = y$ e diciamo che $x(P)$ e $y(P)$ sono le *coordinate* di P rispetto al sistema di riferimento. A ogni punto del piano abbiamo così associato una coppia ordinata $(x(P), y(P))$ di numeri reali. Viceversa, a una coppia ordinata (x, y) di numeri reali, possiamo associare il punto P il cui vettore posizione è $x\mathbf{u} + y\mathbf{v}$ come in (4.2): è chiaro che P appartiene a \mathbf{H} e così otteniamo una corrispondenza biunivoca tra i punti del piano e le coppie ordinate di numeri reali. Una volta fissato il sistema di riferimento, possiamo quindi identificare P con la coppia delle sue coordinate e scrivere semplicemente $P = (x, y)$. Si noti che l'ordine delle coordinate è importante: il punto $(1, 2)$ è diverso dal punto $(2, 1)$!

Componenti di un vettore

Vogliamo ora rivedere quanto appena discusso in termini di vettori. Cominciamo con l'esprimere algebricamente la condizione che un vettore \mathbf{w} sia parallelo al piano \mathbf{H} , cioè abbia la direzione di una retta contenuta in \mathbf{H} . Applichiamo il vettore in O ottenendo $\mathbf{w} \equiv \overrightarrow{OP}$. Siccome O è un punto di \mathbf{H} , il vettore \mathbf{w} è parallelo a \mathbf{H} se e solo se $P \in \mathbf{H}$, cioè per (4.2) se e solo se esistono due scalari x e y (le coordinate di P) tali che

$$(4.3) \quad \mathbf{w} = x\mathbf{u} + y\mathbf{v}$$

Intuitivamente, quest'equazione prova che le direzioni dei vettori del piano \mathbf{H} si ottengono muovendosi per un primo tratto nella direzione di \mathbf{u} e poi di un secondo tratto nella direzione di \mathbf{v} .

I vettori che si possono scrivere nella forma (4.3) si dicono *combinazioni lineari* di \mathbf{u} e \mathbf{v} . Concludiamo:

Supponiamo che $\{\mathbf{u}, \mathbf{v}\}$ sia una base del piano \mathbf{H} . Un vettore \mathbf{w} è parallelo ad \mathbf{H} se e solo se è una combinazione lineare (4.3) di \mathbf{u} e \mathbf{v} .

I coefficienti x e y della combinazione lineare (4.3) sono univocamente determinati da \mathbf{w} : si dice che x e y sono le *coordinate* o le *componenti* del vettore \mathbf{w} rispetto alla base $\{\mathbf{u}, \mathbf{v}\}$ del piano.

Verifichiamo direttamente l'unicità delle componenti: se

$$x_1\mathbf{u} + y_1\mathbf{v} = x_2\mathbf{u} + y_2\mathbf{v}$$

sottraendo a entrambi i membri dell'uguaglianza il membro di destra troviamo

$$(x_1 - x_2)\mathbf{u} + (y_1 - y_2)\mathbf{v} = \mathbf{0}$$

Se si avesse $x_1 - x_2 \neq 0$, allora si ricaverebbe

$$\mathbf{u} = -\frac{y_1 - y_2}{x_1 - x_2}\mathbf{v}$$

e \mathbf{u} sarebbe parallelo a \mathbf{v} , contraddicendo l'ipotesi che $\{\mathbf{u}, \mathbf{v}\}$ sia una base. Quindi $x_1 = x_2$. Analogamente si mostra $y_1 = y_2$.

Per distinguere punti da vettori, useremo la notazione $\mathbf{w} = \begin{bmatrix} x \\ y \end{bmatrix}$ per identificare un vettore con la coppia delle sue componenti, una volta che il sistema di riferimento sia stato fissato.

Rispetto alla base $\{\mathbf{u}, \mathbf{v}\}$, il vettore \mathbf{u} ha componenti $x = 1$ e $y = 0$ perché:

$$\mathbf{u} = 1\mathbf{u} + 0\mathbf{v}$$

Analogamente, \mathbf{v} ha componenti $x = 0$ e $y = 1$. Quindi

$$\mathbf{u} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Linearità delle componenti e \mathbb{R}^2

La somma di due vettori paralleli al piano \mathbf{H} è ancora un vettore parallelo ad \mathbf{H} e un vettore rimane parallelo ad \mathbf{H} se viene moltiplicato per uno scalare t . Il fatto fondamentale è che le componenti del vettore somma $\mathbf{w}_1 + \mathbf{w}_2$ sono la somma delle corrispondenti componenti di \mathbf{w}_1 e \mathbf{w}_2 e che le componenti di $t\mathbf{w}$ si ottengono moltiplicando per t le corrispondenti componenti di \mathbf{w} : una funzione che gode di queste due proprietà si dice *lineare*.

PROPOSIZIONE 4.1 Supponiamo fissata una base del piano \mathbf{H} . La componente x è una funzione *lineare* sull'insieme \mathbf{W} dei vettori paralleli a \mathbf{H} :

$$(4.4) \quad x(\mathbf{w}_1 + \mathbf{w}_2) = x(\mathbf{w}_1) + x(\mathbf{w}_2) \quad \text{per ogni } \mathbf{w}_1, \mathbf{w}_2 \in \mathbf{W}$$

$$(4.5) \quad x(t\mathbf{w}) = tx(\mathbf{w}) \quad \text{per ogni } t \in \mathbb{R} \text{ e ogni } \mathbf{w} \in \mathbf{W}$$

Analogamente, la componente y è una funzione lineare: (4.4) e (4.5) valgono ancora se si sostituisce x con y .

DIMOSTRAZIONE. Se $\mathbf{w}_1 = x_1 \mathbf{u} + y_1 \mathbf{v}$ e $\mathbf{w}_2 = x_2 \mathbf{u} + y_2 \mathbf{v}$, allora

$$\mathbf{w}_1 + \mathbf{w}_2 = (x_1 \mathbf{u} + y_1 \mathbf{v}) + (x_2 \mathbf{u} + y_2 \mathbf{v}) = (x_1 + x_2) \mathbf{u} + (y_1 + y_2) \mathbf{v}$$

Questo mostra che $\mathbf{w}_1 + \mathbf{w}_2$ è anch'esso parallelo a \mathbf{H} e che le sue componenti si ottengono sommando le corrispondenti componenti di \mathbf{w}_1 e \mathbf{w}_2 . Analogamente si mostra $x(t\mathbf{w}) = tx(\mathbf{w})$.

La linearità delle componenti è un fatto estremamente utile perché riconduce le operazioni sui vettori del piano a operazioni sui numeri. Per formalizzare questo punto introduciamo l'insieme \mathbb{R}^2 delle coppie ordinate di numeri reali:

$$\mathbb{R}^2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x, y \in \mathbb{R} \right\}$$

In \mathbb{R}^2 definiamo le operazioni di somma e prodotto per scalare componente per componente:

$$(4.6) \quad \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \end{bmatrix} \quad t \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} tx \\ ty \end{bmatrix}$$

Per la proposizione precedente, l'identificazione $\mathbf{w} = \begin{bmatrix} x \\ y \end{bmatrix}$ di un vettore geometrico col vettore algebrico delle sue coordinate preserva le operazioni di somma e prodotto per scalare:

$$\mathbf{w}_1 + \mathbf{w}_2 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \end{bmatrix} \quad \text{e} \quad t\mathbf{w} = t \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} tx \\ ty \end{bmatrix}$$

Le operazioni sui vettori del piano sono così ricondotte a operazioni sulle loro coordinate.

OSSERVAZIONE Consideriamo due punti A e B del piano \mathbf{H} . Le componenti del vettore \overrightarrow{AB} sono le differenze delle corrispondenti coordinate dei punti A e B :

$$\overrightarrow{AB} \equiv \begin{bmatrix} x_B - x_A \\ y_B - y_A \end{bmatrix}$$

Infatti

$$\overrightarrow{AB} \equiv \overrightarrow{OB} - \overrightarrow{OA} \equiv \begin{bmatrix} x_B \\ y_B \end{bmatrix} - \begin{bmatrix} x_A \\ y_A \end{bmatrix} = \begin{bmatrix} x_B - x_A \\ y_B - y_A \end{bmatrix}$$

4.2 Sistemi di riferimento nello spazio

Per poter assegnare delle coordinate a un punto dello spazio abbiamo bisogno di fissare un punto origine O e tre vettori con direzioni *indipendenti*, in modo che sia possibile raggiungere ogni punto dello spazio traslando O nelle tre direzioni assegnate. Precisiamo cosa voglia dire direzioni indipendenti.

DEFINIZIONE 4.2 I vettori \mathbf{u} , \mathbf{v} e \mathbf{w} si dicono *linearmente dipendenti* se esistono tre scalari a , b e c non tutti nulli tali che

$$(4.7) \quad a\mathbf{u} + b\mathbf{v} + c\mathbf{w} = \mathbf{0}$$

I vettori \mathbf{u} , \mathbf{v} e \mathbf{w} si dicono *linearmente indipendenti* se non sono *linearmente dipendenti*.

Il significato geometrico di questa definizione è contenuto nella seguente proposizione:

PROPOSIZIONE 4.3 Dati tre vettori $\mathbf{u} \equiv \overrightarrow{OA}$, $\mathbf{v} \equiv \overrightarrow{OB}$, $\mathbf{w} \equiv \overrightarrow{OC}$, le seguenti affermazioni sono equivalenti:

- 1) \mathbf{u} , \mathbf{v} e \mathbf{w} sono linearmente indipendenti;
- 2) il vettore \mathbf{u} è non nullo, \mathbf{v} non è parallelo a \mathbf{u} e \mathbf{w} non è combinazione lineare di \mathbf{u} e \mathbf{v} ;
- 3) i punti O , A , B e C non giacciono su uno stesso piano.

Di seguito mostriamo le varie implicazioni:

DIMOSTRAZIONE.

1) \implies 2)

Se $\mathbf{u} = \mathbf{0}$, i tre vettori sono dipendenti perché la (4.7) vale con $a = 1$ e $b = c = 0$. Se $\mathbf{u} \neq \mathbf{0}$, ma \mathbf{v} è parallelo a \mathbf{u} , allora esiste uno scalare t tale che $\mathbf{v} = t\mathbf{u}$, quindi la (4.7) vale con $a = t$ e $b = -1$ e $c = 0$. Infine se $\mathbf{w} = x\mathbf{u} + y\mathbf{v}$ è combinazione lineare di \mathbf{u} e \mathbf{v} , allora la (4.7) vale con $a = x$ e $b = y$ e $c = -1$ e i tre vettori sono dipendenti. Quindi se i tre vettori sono indipendenti deve valere anche 2).

2) \implies 1)

Supponiamo che valga la relazione (4.7): occorre mostrare che, sotto l'ipotesi 2), dev'essere $a = b = c = 0$. Infatti se fosse $c \neq 0$, allora dalla (4.7) possiamo ricavare

$$\mathbf{w} = -\frac{a}{c}\mathbf{u} - \frac{b}{c}\mathbf{v}$$

e quindi $\mathbf{w} = x\mathbf{u} + y\mathbf{v}$ è combinazione lineare di \mathbf{u} e \mathbf{v} , contraddicendo 2). Quindi $c = 0$ e

$$a\mathbf{u} + b\mathbf{v} = \mathbf{0}$$

Se b fosse diverso da zero, avremmo che $\mathbf{v} = -\frac{a}{b}\mathbf{u}$ sarebbe parallelo a \mathbf{u} , contraddicendo 2) di nuovo. Quindi $b = 0$ e la (4.7) si riduce ad $a\mathbf{u} = \mathbf{0}$. Per la legge di cancellazione del prodotto per scalare, siccome $\mathbf{u} \neq \mathbf{0}$, dobbiamo avere $a = 0$ e, quindi, $a = b = c = 0$ come volevasi dimostrare.

2) \Rightarrow 3)

Siccome \mathbf{u} e \mathbf{v} non sono paralleli, $(O, \mathbf{u}, \mathbf{v})$ è un sistema di riferimento del piano \mathbf{H} che contiene i tre punti O, A, B . Se C appartenesse a \mathbf{H} , $\mathbf{w} \equiv \overrightarrow{OC}$ sarebbe parallelo a \mathbf{H} e quindi sarebbe una combinazione lineare di \mathbf{u} e \mathbf{v} . Ma questo contraddice 2). Quindi i punti O, A, B e C non sono complanari, cioè 3) vale.

3) \Rightarrow 2)

Innanzitutto, i quattro punti in questione devono essere distinti perché tre punti sono complanari. Quindi $\mathbf{u} \neq \mathbf{0}$.

Se \mathbf{v} fosse parallelo a \mathbf{u} , allora i tre punti O, A e B sarebbero allineati e quindi il piano per OAC conterebbe anche B , contraddicendo 3). Infine se \mathbf{u} e \mathbf{v} non sono paralleli, ma $\mathbf{w} = x\mathbf{u} + y\mathbf{v}$, allora \mathbf{w} sarebbe parallelo al piano OAB e, poiché $\overrightarrow{OC} \equiv \mathbf{w}$, anche C apparterrebbe al piano per OAB , contraddicendo 3).

Sistema di riferimento nello spazio

Un *sistema di riferimento* $(O, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ dello spazio consiste di:

- un punto O detto *origine* delle coordinate;
- tre vettori linearmente indipendenti $\mathbf{u}_1, \mathbf{u}_2$ e \mathbf{u}_3 , che formano la *base* del sistema di riferimento.

Le rette per O dirette come $\mathbf{u}_1, \mathbf{u}_2$ e \mathbf{u}_3 si dicono *assi del sistema di riferimento*.

Per la proposizione 4.3, dare un sistema di riferimento è equivalente ad assegnare quattro punti non complanari insieme a un loro ordinamento.

Fissato un sistema di riferimento, possiamo associare a un punto P una terna ordinata di numeri reali (x, y, z) analogamente al caso del piano. Da P tracciamo la retta \mathbf{r} parallela a $\mathbf{u}_3 \equiv \overrightarrow{OC}$. Sia \mathbf{H} il piano per i tre punti O, A e B . Per ipotesi \mathbf{u}_3 non è combinazione lineare di \mathbf{u}_1 e \mathbf{u}_2 , il che significa geometricamente che la retta \mathbf{r} non è parallela al piano \mathbf{H} . Ne segue che \mathbf{r} taglia \mathbf{H} in un unico punto Q . Siccome \overrightarrow{QP} è parallelo a \mathbf{u}_3 , esiste un unico scalare z tale che

$$\overrightarrow{QP} \equiv z\mathbf{u}_3$$

D'altra parte, poiché Q appartiene al piano \mathbf{H} ,

$$\overrightarrow{OQ} \equiv x\mathbf{u}_1 + y\mathbf{u}_2$$

dove (x, y) sono le coordinate di Q rispetto al sistema di riferimento $(O, \mathbf{u}_1, \mathbf{u}_2)$ di \mathbf{H} .

Gli scalari x, y e z sono univocamente determinati da P e dal sistema di riferimento: si dicono le *coordinate* del punto P rispetto al sistema di riferimento $(O, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$.

Viceversa, il punto P è determinato dalla terna ordinata delle sue coordinate, perché per costruzione

$$(4.8) \quad \overrightarrow{OP} = \overrightarrow{OQ} + \overrightarrow{QP} \equiv x\mathbf{u}_1 + y\mathbf{u}_2 + z\mathbf{u}_3$$

Geometricamente, P è ottenuto traslando l'origine O prima di un vettore $x\mathbf{u}_1$, poi di un vettore $y\mathbf{u}_2$ e infine di un vettore $z\mathbf{u}_3$. Si verifica immediatamente che le coordinate

del punto così costruito sono esattamente (x, y, z) . Riassumendo: un punto P ha coordinate (x, y, z) se e solo se il suo vettore posizione è

$$(4.9) \quad \overrightarrow{OP} \equiv x\mathbf{u}_1 + y\mathbf{u}_2 + z\mathbf{u}_3$$

Fissato il sistema di riferimento, possiamo così identificare un punto P con la terna delle sue coordinate e scrivere addirittura $P = (x, y, z)$.

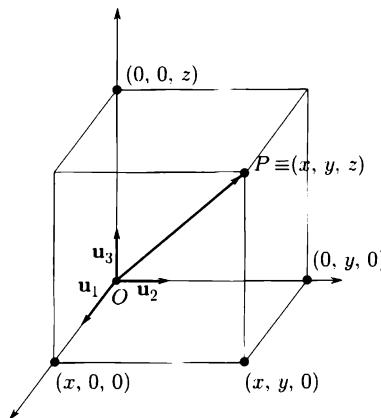


Figura 4.3. Il vettore $\overrightarrow{OP} \equiv x\mathbf{u}_1 + y\mathbf{u}_2 + z\mathbf{u}_3$.

Analogamente al caso del piano, possiamo ora introdurre le *componenti* di un arbitrario vettore \mathbf{v} . Applichiamo il vettore nell'origine, ottenendo $\mathbf{v} \equiv \overrightarrow{OP}$. Se il punto P ha coordinate (x, y, z) , allora per la (4.8) si ha

$$(4.10) \quad \mathbf{v} = x\mathbf{u}_1 + y\mathbf{u}_2 + z\mathbf{u}_3$$

Si dice che \mathbf{v} è *combinazione lineare* dei vettori $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ con coefficienti x, y e z . Poiché \mathbf{v} è univocamente determinato da P , i coefficienti x, y e z sono determinati da \mathbf{v} e si dicono *componenti* o coordinate di \mathbf{v} rispetto alla *base* $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ del sistema di riferimento.

L'unicità delle componenti si può verificare algebricamente come conseguenza dell'indipendenza lineare dei vettori della base: supponiamo che

$$x\mathbf{u}_1 + y\mathbf{u}_2 + z\mathbf{u}_3 = x'\mathbf{u}_1 + y'\mathbf{u}_2 + z'\mathbf{u}_3$$

Sottraendo il secondo membro di questa uguaglianza dal primo troviamo

$$(x - x')\mathbf{u}_1 + (y - y')\mathbf{u}_2 + (z - z')\mathbf{u}_3 = \mathbf{0}$$

Siccome i tre vettori sono linearmente indipendenti, dobbiamo avere

$$x - x' = y - y' = z - z' = 0$$

quindi $x = x'$, $y = y'$ e $z = z'$ come volevasi dimostrare.

Il fatto che *ogni* vettore si possa scrivere nella forma (4.10) è l'espressione algebrica della tridimensionalità dello spazio: nello spazio tridimensionale ci sono soltanto tre direzioni indipendenti. Geometricamente, abbiamo usato l'ipotesi di tridimensionalità sotto la forma che una retta e un piano non paralleli tra loro si incontrano in un punto. Nella teoria astratta degli spazi vettoriali la limitazione che lo spazio sia tridimensionale viene rimossa.

Rispetto alla base $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ il vettore \mathbf{u}_1 ha componenti $x = 1, y = 0$ e $z = 0$ perché:

$$\mathbf{u}_1 = 1 \mathbf{u}_1 + 0 \mathbf{u}_2 + 0 \mathbf{u}_3$$

Analogamente, \mathbf{u}_2 ha componenti $x = 0, y = 1$ e $z = 0$ e \mathbf{u}_3 ha componenti $x = 0, y = 0$ e $z = 1$.

Possiamo identificare un vettore geometrico \mathbf{v} con il vettore algebrico delle sue componenti e scrivere $\mathbf{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$. Un simbolo equivalente è

$$[x, y, z]^T = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

che è tipograficamente più comodo (l'apice T denota l'operazione di *trasposizione* che trasforma righe in colonne). Per esempio $\mathbf{u}_2 = [0, 1, 0]^T$.

Linearità delle componenti di un vettore

Come nel caso di un piano si dimostra che le componenti sono funzioni lineari:

PROPOSIZIONE 4.4 Fissata una base dello spazio, la componente

$$x = x(\mathbf{v}) : \mathbf{V} \rightarrow \mathbb{R}$$

è una funzione lineare:

$$(4.11) \quad x(\mathbf{v}_1 + \mathbf{v}_2) = x(\mathbf{v}_1) + x(\mathbf{v}_2) \quad \text{per ogni } \mathbf{v}_1, \mathbf{v}_2 \in \mathbf{V}$$

$$(4.12) \quad x(t\mathbf{v}) = tx(\mathbf{v}) \quad \text{per ogni } t \in \mathbb{R} \text{ e ogni } \mathbf{v} \in \mathbf{V}$$

Analogamente, le componenti y e z sono funzioni lineari.

Denotiamo con \mathbb{R}^3 l'insieme dei *vettori colonna con tre componenti reali*:

$$\mathbb{R}^3 = \left\{ [x, y, z]^T : x, y, z \in \mathbb{R} \right\}$$

con le operazioni di somma e prodotto per scalare definite componendo per componente:

$$(4.13) \quad \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \\ z_1 + z_2 \end{bmatrix}, \quad t \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} tx \\ ty \\ tz \end{bmatrix}$$

La definizione di somma e prodotto per scalare in \mathbb{R}^3 è data in modo che l'identificazione $\mathbf{v} = [x, y, z]^T$ sia compatibile con le operazioni sui vettori geometrici: per la proposizione 4.1:

$$\mathbf{v}_1 + \mathbf{v}_2 = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \\ z_1 + z_2 \end{bmatrix}$$

Si noti che nella formula il primo $+$ denota la somma di vettori liberi, il secondo la somma in \mathbb{R}^3 . Analogamente per il prodotto scalare:

$$t\mathbf{v} = t \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} tx \\ ty \\ tz \end{bmatrix}$$

Possiamo quindi effettuare tutti i conti direttamente sulle componenti, riconducendo le operazioni sui vettori a operazioni sui numeri.

5 COORDINATE CARTESIANE NELLO SPAZIO

Il nostro prossimo obiettivo è trovare un'espressione per il modulo di un vettore in termini delle sue coordinate. Si ottiene un'espressione semplice solo se si utilizza un sistema di riferimento $(O, \mathbf{i}, \mathbf{j}, \mathbf{k})$ in cui i vettori della base siano a due a due perpendicolari e abbiano modulo uno (questo intuitivamente significa aver fissato la medesima unità di misura sui tre assi di riferimento). Analizziamo quindi brevemente le condizioni di lunghezza 1 e perpendicolarità.

DEFINIZIONE 5.1 Un *versore* è un vettore il cui modulo è uguale a 1. In simboli: \mathbf{e} è un versore se $\|\mathbf{e}\| = 1$.

Per definizione di prodotto per scalare. vale l'uguaglianza

$$(5.1) \quad \|t\mathbf{v}\| = |t| \|\mathbf{v}\|$$

per ogni scalare t e per ogni vettore \mathbf{v} . In particolare,

- i) se \mathbf{e} è un versore, il modulo di $t\mathbf{e}$ è precisamente $|t|$.
- ii) se \mathbf{v} è un vettore non nullo, esistono esattamente due versori diretti come \mathbf{v} :

$$\mathbf{e}_1 = \frac{\mathbf{v}}{\|\mathbf{v}\|} \quad \text{ed} \quad \mathbf{e}_2 = -\frac{\mathbf{v}}{\|\mathbf{v}\|}$$

Il versore \mathbf{e}_1 è ottenuto dividendo \mathbf{v} per il suo modulo; è l'unico versore che ha la stessa direzione e lo stesso verso di \mathbf{v} . L'altro versore \mathbf{e}_2 è il vettore opposto di \mathbf{e}_1 : ha la stessa direzione di \mathbf{v} , ma verso opposto.

Due vettori non nulli $\mathbf{v} \equiv \overrightarrow{OP}$ e $\mathbf{w} \equiv \overrightarrow{OQ}$ dividono il piano per O , P e Q in due angoli: uno convesso $\alpha \in [0, \pi]$ e l'altro concavo $\beta = 2\pi - \alpha$. L'angolo convesso α si dice *angolo formato da* \mathbf{v} e \mathbf{w} ; non dipende dalla scelta del punto di applicazione O . Un angolo convesso è determinato dal suo coseno, perché il coseno è una funzione biunivoca dell'intervallo $[0, \pi]$ sull'intervallo $[-1, 1]$. Quindi per conoscere α è sufficiente

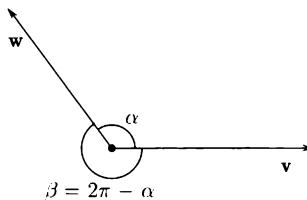


Figura 5.1. L'angolo α formato da \mathbf{v} e \mathbf{w} .

conoscere $\cos(\alpha)$. Si noti che $\cos(\beta) = \cos(2\pi - \alpha) = \cos(\alpha)$. Quando i due vettori sono paralleli, l'angolo α è zero se i due vettori hanno lo stesso verso, mentre $\alpha = \pi$ se hanno verso opposto.

DEFINIZIONE 5.2 Due vettori non nulli si dicono *ortogonali* (o *perpendicolari*) se l'angolo da essi formato è un angolo retto. Il vettore nullo è per definizione ortogonale a qualsiasi altro vettore. Il simbolo $\mathbf{v} \perp \mathbf{w}$ significa che \mathbf{v} e \mathbf{w} sono ortogonali.

OSSERVAZIONE Due rette si dicono ortogonali (o perpendicolari) se i loro vettori direzione sono ortogonali. Nello spazio due rette ortogonali *non* sono necessariamente incidenti. Per esempio, se si fissa nello spazio un riferimento cartesiano, in cui gli assi di riferimento sono perpendicolari tra loro, ogni retta del piano xy è perpendicolare all'asse z (indipendentemente dal fatto che contenga l'origine).

Il teorema fondamentale sull'ortogonalità è il Teorema di Pitagora, che tradotto nell'algebra dei vettori diventa:

TEOREMA 5.3 (Teorema di Pitagora)

Se \mathbf{v} e \mathbf{w} sono ortogonali, allora

$$(5.2) \quad \|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2$$

DIMOSTRAZIONE. Se uno dei due vettori è nullo, l'enunciato è ovvio. Se i due vettori sono non nulli, applichiamo i due vettori in un punto O : $\mathbf{v} \equiv \overrightarrow{OP}$ e $\mathbf{w} \equiv \overrightarrow{OQ}$. Per ipotesi l'angolo \widehat{POQ} è retto, quindi il parallelogramma generato da OP e OQ è un rettangolo. Il vettore somma $\mathbf{v} + \mathbf{w} \equiv \overrightarrow{OR}$ è una diagonale del rettangolo, per cui la tesi segue dall'usuale teorema di Pitagora:

$$\|\mathbf{v} + \mathbf{w}\|^2 = \overline{OR}^2 = \overline{OP}^2 + \overline{OQ}^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2$$

DEFINIZIONE 5.4 Un sistema di riferimento $(O, \mathbf{i}, \mathbf{j}, \mathbf{k})$ dello spazio si dice *cartesiano* se i vettori \mathbf{i} , \mathbf{j} e \mathbf{k}

a) sono dei versori:

$$\|\mathbf{i}\| = \|\mathbf{j}\| = \|\mathbf{k}\| = 1$$

b) sono a due a due perpendicolari:

$$\mathbf{i} \perp \mathbf{j}, \quad \mathbf{i} \perp \mathbf{k}, \quad \mathbf{j} \perp \mathbf{k}$$

c) formano una terna destrorsa: se \mathbf{i} e \mathbf{j} hanno rispettivamente il verso del pollice e dell'indice della mano destra, il versore \mathbf{k} ha il verso del dito medio.

Una terna di vettori per cui valgano a) e b) si dice *base ortonormale: normale* perché i vettori sono normalizzati in modo da avere lunghezza 1 e *orto* perché i vettori sono a due a due ortogonali.

La definizione che abbiamo dato di terna destrorsa è intuitiva, ma non rigorosa. Non ne faremo uso finché non introdurremo il prodotto vettoriale e allora saremo in grado di darne una versione rigorosa.

Le coordinate (x, y, z) di un punto P si dicono rispettivamente l'*ascissa*, l'*ordinata* e la *quota* del punto. La retta per O diretta come il versore \mathbf{i} si dice asse delle ascisse o delle x ; analogamente l'asse delle ordinate o delle y (rispettivamente delle quote o delle z) è la retta per O diretta come il versore \mathbf{j} (la retta per O diretta come il versore \mathbf{k}). Il punto P di coordinate (x, y, z) ha vettore posizione

$$\overrightarrow{OP} \equiv x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

Possiamo finalmente esprimere il modulo del vettore \overrightarrow{OP} in funzione delle coordinate:

PROPOSIZIONE 5.5 (Modulo di un vettore in coordinate cartesiane)

Se $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ formano una base ortonormale dello spazio, il modulo del vettore $\mathbf{v} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ è:

$$(5.3) \qquad \|\mathbf{v}\| = \sqrt{x^2 + y^2 + z^2}$$

DIMOSTRAZIONE. Il vettore $\mathbf{v} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ è la somma del vettore $x\mathbf{i} + y\mathbf{j}$ e del vettore $z\mathbf{k}$. Questi due vettori sono perpendicolari perché il primo è parallelo al piano xy e il secondo è parallelo all'asse z . Per il teorema di Pitagora:

$$\|\mathbf{v}\|^2 = \|x\mathbf{i} + y\mathbf{j}\|^2 + \|z\mathbf{k}\|^2$$

I vettori $x\mathbf{i}$ e $y\mathbf{j}$ sono perpendicolari perché il primo è diretto come l'asse delle x , il secondo come l'asse y . Usando ancora il teorema di Pitagora troviamo

$$(5.4) \qquad \|\mathbf{v}\|^2 = \|x\mathbf{i} + y\mathbf{j}\|^2 + \|z\mathbf{k}\|^2 = \|x\mathbf{i}\|^2 + \|y\mathbf{j}\|^2 + \|z\mathbf{k}\|^2$$

Fin qui abbiamo solo usato la perpendicolarità dei vettori della base; dal fatto che \mathbf{i} sia un versore segue $\|x\mathbf{i}\| = |x|$ e analogamente $\|y\mathbf{j}\| = |y|$ e $\|z\mathbf{k}\| = |z|$. Tenuto conto di questo, la (5.3) segue dalla (5.4).

6 PROIEZIONI ORTOGONALI E PRODOTTO SCALARE

Il nostro prossimo obiettivo è mostrare che le coordinate cartesiane di un punto P si ottengono proiettando il punto sugli assi coordinati. Un problema collegato è quello di ricavare un'espressione in coordinate per il coseno dell'angolo α formato da due vettori: questo per esempio consente di stabilire immediatamente se due vettori sono perpendicolari una volta che ne siano note le componenti. Il percorso logico parte dalla nozione di proiezione ortogonale, che è immediatamente legata al coseno di un angolo, e conduce allo strumento tecnico fondamentale che è *il prodotto scalare di due vettori*.

Cominciamo col mostrare che, nel caso di un sistema di riferimento cartesiano, le coordinate di un punto P si ottengono proiettando P ortogonalmente sugli assi coordinati. Per questo introduciamo la nozione di *proiezione ortogonale* un vettore \mathbf{v} lungo un vettore non nullo \mathbf{w} .

L'intuizione ci suggerisce di procedere così. Applichiamo \mathbf{v} in un punto O come in figura (6.1): $\mathbf{v} \equiv \overrightarrow{OP}$. Consideriamo la retta \mathbf{r} per O diretta come \mathbf{w} : è la retta su cui vogliamo proiettare \mathbf{v} . Per determinare la proiezione, nel piano per \mathbf{r} e P tracciamo la retta \mathbf{s} passante per P perpendicolare a \mathbf{r} . La retta \mathbf{s} interseca la retta \mathbf{r} in un punto Q che è la proiezione ortogonale di P su \mathbf{r} perché l'angolo $\widehat{\overrightarrow{OQ}\overrightarrow{P}}$ è retto. Diremo che il vettore \overrightarrow{OQ} è la proiezione ortogonale di \overrightarrow{OP} sulla retta \mathbf{r} .

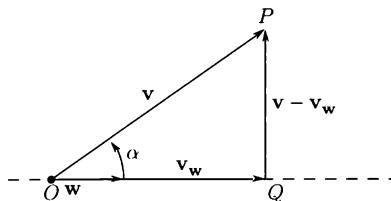


Figura 6.1. Proiezione ortogonale di \mathbf{v} lungo \mathbf{w} .

Osserviamo che \overrightarrow{OQ} è diretto come \mathbf{w} , mentre $\overrightarrow{QP} \equiv \overrightarrow{OP} - \overrightarrow{OQ}$ è perpendicolare a \mathbf{w} . Il vettore

$$\mathbf{v}_w \equiv \overrightarrow{OQ}$$

ha perciò le seguenti proprietà:

- a) \mathbf{v}_w è parallelo a \mathbf{w} ;
- b) $\mathbf{v} - \mathbf{v}_w$ è perpendicolare a \mathbf{w} .

Mostriamo che a) e b) determinano \mathbf{v}_w . Supponiamo infatti che a) e b) valgano per un vettore $\mathbf{u} \equiv \overrightarrow{OQ_1}$. Allora da a) segue che Q_1 appartiene alla retta \mathbf{r} . D'altra parte, $\mathbf{v} - \mathbf{u} \equiv \overrightarrow{Q_1P}$ per b) è perpendicolare a \mathbf{w} , quindi Q_1 appartiene alla retta \mathbf{s} (la retta per P perpendicolare a \mathbf{r} nel piano che contiene P ed \mathbf{r}). Concludiamo che Q_1 è il punto di intersezione di \mathbf{r} ed \mathbf{s} , per cui $\mathbf{u} = \mathbf{v}_w$.

Quindi, dal punto di vista algebrico, determinare la proiezione ortogonale di \mathbf{v} lungo \mathbf{w} significa decomporre \mathbf{v} nella somma di un vettore \mathbf{v}_w diretto come \mathbf{w} e di un vettore $\mathbf{v} - \mathbf{v}_w$ perpendicolare a \mathbf{w} .

DEFINIZIONE 6.1 Sia \mathbf{w} un vettore non nullo. La *proiezione ortogonale* di un vettore \mathbf{v} nella direzione di \mathbf{w} è il vettore \mathbf{v}_w che soddisfa le due proprietà:

- a) \mathbf{v}_w è parallelo a \mathbf{w} ;
- b) $\mathbf{v} - \mathbf{v}_w$ è perpendicolare a \mathbf{w} .

Osserviamo che \mathbf{v}_w dipende solo dalla direzione di \mathbf{w} perché le proprietà a) e b) dipendono solo dalla direzione di \mathbf{w} . Notiamo inoltre che la proiezione ortogonale \mathbf{v}_w è un vettore.

Se $\alpha \in [0, \pi]$ è l'angolo formato da \mathbf{v} e \mathbf{w} , il coseno dell'angolo α è (per definizione!) il rapporto (con segno) tra le lunghezze di \mathbf{v}_w e di \mathbf{v} . Precisamente:

PROPOSIZIONE 6.2 Siano \mathbf{v} e \mathbf{w} due vettori non nulli e sia α l'angolo formato da \mathbf{v} e \mathbf{w} . Per la proiezione ortogonale di \mathbf{v} lungo \mathbf{w} vale la formula:

$$(6.1) \quad \mathbf{v}_w = \cos(\alpha) \|\mathbf{v}\| \frac{\mathbf{w}}{\|\mathbf{w}\|}$$

DIMOSTRAZIONE. L'enunciato è sostanzialmente la definizione di coseno. Il punto è che in un triangolo rettangolo il rapporto tra la lunghezza di un cateto e l'ipotenusa dipende solo dall'angolo α che essi formano, tale rapporto è per definizione il coseno dell'angolo α . La formula (6.1) dice esattamente questo nel caso in cui \mathbf{v} e \mathbf{w} formino un angolo acuto (i due membri hanno segno positivo). La formula rimane valida anche quando l'angolo è ottuso per la definizione del segno del coseno.

Il nostro prossimo obiettivo è trovare un'espressione per la proiezione \mathbf{v}_w e quindi per $\cos(\alpha)$, in termini delle componenti di \mathbf{v} e \mathbf{w} . Si ottiene un risultato più semplice rendendo simmetrici i ruoli di \mathbf{v} e \mathbf{w} in (6.1). Per questo si introduce il prodotto scalare di \mathbf{v} e \mathbf{w} :

DEFINIZIONE 6.3 Siano \mathbf{v} e \mathbf{w} due vettori non nulli e sia α l'angolo da essi formato. Il *prodotto scalare* (o *prodotto interno*) di \mathbf{v} e \mathbf{w} è il *numero reale* definito dalla formula

$$(6.2) \quad \mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos(\alpha)$$

Se uno dei vettori \mathbf{v} e \mathbf{w} è il vettore nullo, si pone $\mathbf{v} \cdot \mathbf{w} = 0$.

Possiamo riscrivere la (6.1) esprimendo la proiezione ortogonale in termini del prodotto scalare:

$$(6.3) \quad \mathbf{v}_w = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{w}\|^2} \mathbf{w}$$

In particolare, se \mathbf{e} è un versore,

$$(6.4) \quad \mathbf{v}_e = (\mathbf{v} \cdot \mathbf{e}) \mathbf{e}$$

Supponiamo che \mathbf{i} , \mathbf{j} e \mathbf{k} formino una base ortonormale dello spazio, cioè siano tre versori a due a due perpendicolari. Allora

$$\mathbf{i} \cdot \mathbf{i} = \mathbf{j} \cdot \mathbf{j} = \mathbf{k} \cdot \mathbf{k} = 1 \quad \text{e} \quad \mathbf{i} \cdot \mathbf{j} = \mathbf{i} \cdot \mathbf{k} = \mathbf{j} \cdot \mathbf{k} = 0$$

Per mostrare la prima uguaglianza, osserviamo che l'angolo formato da un vettore con se stesso è nullo, quindi

$$\mathbf{i} \cdot \mathbf{i} = \|\mathbf{i}\|^2 \cos(0) = \|\mathbf{i}\|^2 = 1$$

Nell'ultimo passaggio abbiamo usato l'ipotesi che \mathbf{i} sia un versore.

L'annullarsi del prodotto scalare $\mathbf{i} \cdot \mathbf{j}$ segue dalla perpendicolarità di \mathbf{i} e \mathbf{j} :

$$\mathbf{i} \cdot \mathbf{j} = \|\mathbf{i}\| \|\mathbf{j}\| \cos\left(\frac{\pi}{2}\right) = 0$$

Si noti che *due vettori sono perpendicolari se e solo se il loro prodotto scalare è nullo* perché un angolo $\alpha \in [0, \pi]$ ha coseno nullo se e soltanto se è un angolo retto.

Mostriamo ora che in un sistema di riferimento cartesiano le coordinate si ottengono mediante proiezioni ortogonali sugli assi coordinati e quindi si possono esprimere mediante il prodotto scalare.

PROPOSIZIONE 6.4 Sia $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ una base ortonormale dello spazio. Se $\mathbf{v} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$, allora

$$(6.5) \quad x = \mathbf{v} \cdot \mathbf{i}, \quad y = \mathbf{v} \cdot \mathbf{j}, \quad z = \mathbf{v} \cdot \mathbf{k}$$

Quindi le coordinate x , y e z di un vettore \mathbf{v} sono il prodotto scalare di \mathbf{v} coi versori \mathbf{i} , \mathbf{j} e \mathbf{k} degli assi coordinati.

DIMOSTRAZIONE. Il vettore $y\mathbf{j} + z\mathbf{k}$ è parallelo al piano yz , quindi è perpendicolare a \mathbf{i} . Nella somma

$$\mathbf{v} = x\mathbf{i} + (y\mathbf{j} + z\mathbf{k})$$

il vettore $x\mathbf{i}$ è parallelo a \mathbf{i} , mentre $y\mathbf{j} + z\mathbf{k}$ è perpendicolare a \mathbf{i} , quindi $x\mathbf{i}$ è la proiezione ortogonale \mathbf{v}_i di \mathbf{v} sull'asse x . D'altra parte per la (6.4)

$$\mathbf{v}_i = (\mathbf{v} \cdot \mathbf{i})\mathbf{i}$$

Confrontando le due espressioni di \mathbf{v}_i si ottiene $x = \mathbf{v} \cdot \mathbf{i}$. Il ragionamento per le altre due coordinate è analogo.

□

Supponiamo che \mathbf{i} , \mathbf{j} e \mathbf{k} formino una base ortonormale dello spazio. Possiamo ritrovare le componenti di \mathbf{i} rispetto a tale base usando la (6.5):

$$x = \mathbf{i} \cdot \mathbf{i} = 1, \quad y = \mathbf{i} \cdot \mathbf{j} = 0, \quad z = \mathbf{i} \cdot \mathbf{k} = 0$$

Il ragionamento si può ripetere per \mathbf{j} e \mathbf{k} , ritrovando:

$$\mathbf{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Il prodotto scalare ha un'espressione in coordinate cartesiane molto semplice, anche se questo non è affatto evidente dalla definizione che ne abbiamo dato. Ricaveremo tale espressione dalle proprietà algebriche del prodotto scalare:

PROPOSIZIONE 6.5 Il prodotto scalare gode delle seguenti proprietà:

a) commutatività: $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$;

b) linearità nel primo fattore:

$$(\mathbf{v}_1 + \mathbf{v}_2) \cdot \mathbf{w} = \mathbf{v}_1 \cdot \mathbf{w} + \mathbf{v}_2 \cdot \mathbf{w}$$

e

$$(t\mathbf{v}) \cdot \mathbf{w} = t(\mathbf{v} \cdot \mathbf{w})$$

c) positività: $\mathbf{v} \cdot \mathbf{v} \geq 0$ e l'uguaglianza vale se e solo se $\mathbf{v} = \mathbf{0}$.

DIMOSTRAZIONE. La commutatività e la seconda uguaglianza in b) sono evidenti dalla definizione. Mostriamo ora che

$$(6.6) \quad (\mathbf{v}_1 + \mathbf{v}_2) \cdot \mathbf{w} = \mathbf{v}_1 \cdot \mathbf{w} + \mathbf{v}_2 \cdot \mathbf{w}$$

Se $\mathbf{w} = \mathbf{0}$, entrambi i membri sono 0. Se $\mathbf{w} \neq \mathbf{0}$, consideriamo il versore $\mathbf{i} = \frac{\mathbf{w}}{\|\mathbf{w}\|}$. Prendendo due versori \mathbf{j} e \mathbf{k} perpendicolari a \mathbf{i} , costruiamo un sistema di riferimento cartesiano $(O, \mathbf{i}, \mathbf{j}, \mathbf{k})$. Per la proposizione 6.4 il prodotto scalare $\mathbf{v} \cdot \mathbf{i}$ è la componente x di \mathbf{v} rispetto a tale riferimento. Siccome la componente x è una funzione lineare, dalla proposizione 6.4 segue

$$(\mathbf{v}_1 + \mathbf{v}_2) \cdot \mathbf{i} = x(\mathbf{v}_1 + \mathbf{v}_2) = x(\mathbf{v}_1) + x(\mathbf{v}_2) = \mathbf{v}_1 \cdot \mathbf{i} + \mathbf{v}_2 \cdot \mathbf{i}$$

Moltiplicando entrambi i membri per lo scalare $\|\mathbf{w}\|$ otteniamo (6.6).

Infine, se \mathbf{v} è un vettore non nullo, l'angolo che \mathbf{v} forma con se stesso è nullo, quindi

$$\mathbf{v} \cdot \mathbf{v} = \|\mathbf{v}\|^2 \cos(0) = \|\mathbf{v}\|^2$$

da cui segue c).

OSSERVAZIONE Naturalmente, dalla commutatività e dalla linearità nel primo fattore, segue che il prodotto scalare è lineare anche nel secondo fattore, cioè

$$\mathbf{v} \cdot (\mathbf{w}_1 + \mathbf{w}_2) = \mathbf{v} \cdot \mathbf{w}_1 + \mathbf{v} \cdot \mathbf{w}_2 \quad \text{e} \quad \mathbf{v} \cdot (t\mathbf{w}) = t(\mathbf{v} \cdot \mathbf{w})$$

Possiamo finalmente ricavare l'espressione in coordinate cartesiane del prodotto scalare:

PROPOSIZIONE 6.6 Se $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ formano una base ortonormale dello spazio, allora il prodotto scalare dei vettori $\mathbf{v}_1 = x_1 \mathbf{i} + y_1 \mathbf{j} + z_1 \mathbf{k}$ e $\mathbf{v}_2 = x_2 \mathbf{i} + y_2 \mathbf{j} + z_2 \mathbf{k}$ è:

$$(6.7) \quad \mathbf{v}_1 \cdot \mathbf{v}_2 = x_1 x_2 + y_1 y_2 + z_1 z_2$$

DIMOSTRAZIONE. Utilizzando la linearità del prodotto scalare nella prima componente troviamo

$$(6.8) \quad \mathbf{v}_1 \cdot \mathbf{v}_2 = (x_1 \mathbf{i} + y_1 \mathbf{j} + z_1 \mathbf{k}) \cdot \mathbf{v}_2 = x_1 \mathbf{i} \cdot \mathbf{v}_2 + y_1 \mathbf{j} \cdot \mathbf{v}_2 + z_1 \mathbf{k} \cdot \mathbf{v}_2$$

Per la proposizione 6.4 abbiamo $v_2 \cdot i = x_2$, $v_2 \cdot j = y_2$ e $v_2 \cdot k = z_2$. La (6.7) segue quindi dalla (6.8).

È importante notare che tramite il prodotto scalare possiamo calcolare tutto quanto abbiamo introdotto in precedenza: modulo di un vettore e quindi distanza tra due punti, coseno dell'angolo tra due vettori e proiezioni scalari. Per questo, nel contesto astratto del capitolo sugli spazi vettoriali euclidei, definiremo lunghezze e angoli a partire dal prodotto scalare. Esplicitamente:

a) Il modulo di un vettore \mathbf{v} è

$$(6.9) \quad \|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$$

b) L'angolo α formato da due vettori non nulli \mathbf{v} e \mathbf{w} è determinato da:

$$(6.10) \quad \cos(\alpha) = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|}$$

c) La proiezione ortogonale di un vettore \mathbf{v} nella direzione di un vettore non nullo \mathbf{w} è

$$(6.11) \quad \mathbf{v}_{\mathbf{w}} = \frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}} \mathbf{w}$$

Le formule (6.9) e (6.10) seguono immediatamente dalla definizione di prodotto scalare, tenendo conto che un vettore forma con se stesso un angolo nullo, $\cos(0) = 1$, mentre la formula per la proiezione ortogonale è già stata discussa nella (6.3).

Supponiamo che nello spazio sia fissato un sistema di riferimento cartesiano. Dati i vettori $\mathbf{v} = [3, 5, 3]^T$ e $\mathbf{w}_a = [4, -2, a]^T$, determiniamo gli eventuali valori del parametro a per i quali i due vettori sono perpendicolari.

Per questo occorre ricordare che $\mathbf{v} \perp \mathbf{w}_a$ se e solo se $\mathbf{v} \cdot \mathbf{w}_a = 0$. Ora

$$\mathbf{v} \cdot \mathbf{w}_a = \begin{bmatrix} 3 \\ 5 \\ -3 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ -2 \\ a \end{bmatrix} = 12 - 10 - 3a = 2 - 3a$$

quindi i due vettori sono perpendicolari se e solo se $a = \frac{2}{3}$.



Dati i tre punti $A = (1, 2, 3)$, $B = (2, 2, 3)$ e $C = (2, 3, 4)$, determiniamo l'angolo α formato dai vettori \overrightarrow{AB} e \overrightarrow{AC} .

Cominciamo col trovare le componenti di \overrightarrow{AB} : sono le differenze delle coordinate di B e di A perché:

$$\overrightarrow{AB} \equiv \overrightarrow{OB} - \overrightarrow{OA} = \begin{bmatrix} x_B \\ y_B \\ z_B \end{bmatrix} - \begin{bmatrix} x_A \\ y_A \\ z_A \end{bmatrix} = \begin{bmatrix} x_B - x_A \\ y_B - y_A \\ z_B - z_A \end{bmatrix}$$

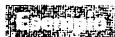
Quindi

$$\overrightarrow{AB} \equiv \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{e} \quad \overrightarrow{AC} \equiv \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Possiamo ora calcolare

$$\cos(\alpha) = \frac{\overrightarrow{AB} \cdot \overrightarrow{AC}}{\|\overrightarrow{AB}\| \|\overrightarrow{AC}\|} = \frac{1}{\sqrt{3}} = \frac{\sqrt{3}}{3}$$

Quindi $\alpha = \arccos(\frac{\sqrt{3}}{3})$ (poco meno di 55 gradi). Osserviamo che $\overrightarrow{AB} = \mathbf{i}$ e $\overrightarrow{AC} = \mathbf{i} + \mathbf{j} + \mathbf{k}$. Poiché \mathbf{i} è lo spigolo di un cubo di lato unitario e $\mathbf{i} + \mathbf{j} + \mathbf{k}$ è la diagonale del cubo, α è l'angolo formato dalla diagonale di un cubo con uno spigolo adiacente.



Determiniamo la proiezione ortogonale \mathbf{v}_r del vettore $\mathbf{v} = [1, 2, 3]^T$ sulla retta r che congiunge l'origine con il punto $A = (3, 1, -1)$.

La retta r è parallela al vettore

$$\mathbf{w} \equiv \overrightarrow{OA} \equiv [3, 1, -1]^T$$

Usando la formula (6.11) troviamo

$$(6.12) \quad \mathbf{v}_w = \frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}} \mathbf{w} = \frac{3+2-3}{9+1+1} \mathbf{w} = \left[\frac{6}{11}, \frac{2}{11}, -\frac{2}{11} \right]^T$$

Possiamo verificare il risultato controllando che la proiezione ortogonale \mathbf{v}_w sia perpendicolare al vettore $\mathbf{v} - \mathbf{v}_w$:

$$\mathbf{v} - \mathbf{v}_w = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} \frac{6}{11} \\ \frac{2}{11} \\ -\frac{2}{11} \end{bmatrix} = \begin{bmatrix} \frac{5}{11} \\ \frac{20}{11} \\ \frac{35}{11} \end{bmatrix}$$

per cui

$$\mathbf{v}_w \cdot (\mathbf{v} - \mathbf{v}_w) = \begin{bmatrix} \frac{6}{11} \\ \frac{2}{11} \\ \frac{-2}{11} \end{bmatrix} \cdot \begin{bmatrix} \frac{5}{11} \\ \frac{20}{11} \\ \frac{35}{11} \end{bmatrix} = \frac{30 + 40 - 70}{11^2} = 0$$

Dati i punti $A = (3, 2, 1)$, $B = (2, 2, 5)$, $C = (1, 2, 7)$, scrivere la proiezione ortogonale di \overrightarrow{AB} su \overrightarrow{AC} . Calcolare l'area del triangolo $\triangle ABC$.

Posto $A = (2, 1, 3)$ e $B = (1, 1, 4)$, scrivere i (due) versori che sono diretti come la retta AB . Ripetere l'esercizio per due punti generali $A = (a_1, a_2, a_3)$ e $B = (b_1, b_2, b_3)$.

Trovare un versore che forma angoli uguali coi tre semiassi positivi.

Usando le proprietà del prodotto scalare mostrare l'uguaglianza (teorema di Carnot o del coseno):

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2\mathbf{u} \cdot \mathbf{v} + \|\mathbf{v}\|^2$$

Dimostrare la *disuguaglianza di Schwarz*: $|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\| \|\mathbf{v}\|$. Dedurne (usando il teorema di Carnot) la *disuguaglianza triangolare*:

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

Quando vale l'uguaglianza in queste due disuguaglianze?

Proiezione ortogonale su un piano Sia \mathbf{n} un vettore perpendicolare al piano \mathbf{H} . Dato un vettore libero \mathbf{v} , si ponga $\mathbf{v}_H = \mathbf{v} - \mathbf{v}_n$, dove \mathbf{v}_n è la proiezione ortogonale di \mathbf{v} lungo \mathbf{n} . Si dice che \mathbf{v}_H è la proiezione ortogonale di \mathbf{v} su \mathbf{H} , perché \mathbf{v}_H è parallelo a \mathbf{H} e $\mathbf{v} - \mathbf{v}_H$ è perpendicolare a \mathbf{H} .

- Supponiamo che $O \in \mathbf{H}$, $\mathbf{v} \equiv \overrightarrow{OP}$ e $\mathbf{v}_H \equiv \overrightarrow{OQ}$. Mostrare che la retta QP è ortogonale a \mathbf{H} e che Q è il punto di \mathbf{H} più vicino a P .
- Scrivere una formula per le componenti di \mathbf{v}_H quando $\mathbf{n} = [3, 2, 1]^T$, e per un generico vettore $\mathbf{n} = [a, b, c]^T$.
- Se \mathbf{H} è il piano xy , mostrare che la proiezione ortogonale di $[x, y, z]^T$ è $[x, y, 0]^T$.
- Supponiamo che \mathbf{u} e \mathbf{w} siano due vettori non nulli, paralleli a \mathbf{H} e perpendicolari tra loro. Mostrare che:

$$\mathbf{v}_H = \mathbf{v}_u + \mathbf{v}_w$$

Osservare che l'ipotesi che \mathbf{u} e \mathbf{w} siano perpendicolari tra loro è necessaria affinché questa formula sia vera.

Sia \mathbf{H} il piano per l'origine perpendicolare al vettore $[1, 3, 2]^T$. Scrivere il vettore $\mathbf{v} = [4, 1, 5]^T$ come somma di un vettore parallelo ad \mathbf{H} e di un vettore perpendicolare ad \mathbf{H} .

7 PRODOTTO VETTORIALE E PRODOTTO MISTO

Introduciamo ora un'operazione, il *prodotto vettoriale*, che a due vettori dati associa un vettore ad essi perpendicolare:

DEFINIZIONE 7.1 (Prodotto vettoriale)

Il *prodotto vettoriale* di due vettori \mathbf{v} e \mathbf{w} è il vettore $\mathbf{v} \times \mathbf{w}$ (indicato anche col simbolo $\mathbf{v} \wedge \mathbf{w}$) così definito:

- a) se \mathbf{v} e \mathbf{w} non sono paralleli,
 - il modulo di $\mathbf{v} \times \mathbf{w}$ è uguale all'area del parallelogramma di spigoli \mathbf{v} e \mathbf{w} ; quindi
$$\|\mathbf{v} \times \mathbf{w}\| = \|\mathbf{v}\| \|\mathbf{w}\| |\sin(\alpha)|$$

dove α è l'angolo formato da \mathbf{v} e \mathbf{w} ;

 - la direzione di $\mathbf{v} \times \mathbf{w}$ è quella perpendicolare al piano generato da \mathbf{v} e da \mathbf{w} ; quindi $\mathbf{v} \times \mathbf{w}$ è perpendicolare sia a \mathbf{v} sia a \mathbf{w} ;
 - il verso di $\mathbf{v} \times \mathbf{w}$ è tale che la terna di vettori \mathbf{v} , \mathbf{w} , $\mathbf{v} \times \mathbf{w}$ sia destrorsa;
- b) se \mathbf{v} e \mathbf{w} sono paralleli (incluso il caso in cui uno dei due vettori sia nullo), il prodotto vettoriale di \mathbf{v} e \mathbf{w} è il vettore nullo.

Il *parallelogramma di spigoli* \mathbf{v} e \mathbf{w} si costruisce applicando i vettori \mathbf{v} e \mathbf{w} in un punto A : posto $\mathbf{v} \equiv \overrightarrow{AB}$, $\mathbf{w} \equiv \overrightarrow{AC}$ e $\mathbf{v} + \mathbf{w} \equiv \overrightarrow{AD}$, i vertici del parallelogramma sono A , B , C e D . L'angolo α è l'angolo tra \overrightarrow{AB} e \overrightarrow{AC} . Quando α tende a 0 o π , cioè \mathbf{w} diviene parallelo a \mathbf{v} , il parallelogramma collassa sulla retta AC e la sua area tende a zero. Quindi il caso in cui \mathbf{v} e \mathbf{w} sono paralleli è ottenuto come limite del caso generale.

Dalla definizione segue immediatamente:

Prodotto vettoriale dei versori cartesiani

Se $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ è una terna ortonormale destrorsa,

$$\mathbf{i} \times \mathbf{j} = \mathbf{k} = -\mathbf{j} \times \mathbf{i} \quad \mathbf{j} \times \mathbf{k} = \mathbf{i} = -\mathbf{k} \times \mathbf{j} \quad \mathbf{k} \times \mathbf{i} = \mathbf{j} = -\mathbf{i} \times \mathbf{k}$$

$$\mathbf{i} \times \mathbf{i} = \mathbf{j} \times \mathbf{j} = \mathbf{k} \times \mathbf{k} = \mathbf{0}$$

Analogamente al caso del prodotto scalare, ricaveremo l'espressione in coordinate del prodotto vettoriale da queste uguaglianze e dalle seguenti fondamentali proprietà:

Proprietà del prodotto vettoriale

- a) Anticommutatività: $\mathbf{v} \times \mathbf{w} = -\mathbf{w} \times \mathbf{v}$.

b) Bilinearità: per ogni $t_1, t_2 \in \mathbb{R}$

$$(t_1\mathbf{v}_1 + t_2\mathbf{v}_2) \times \mathbf{w} = t_1(\mathbf{v}_1 \times \mathbf{w}) + t_2(\mathbf{v}_2 \times \mathbf{w})$$

$$\mathbf{v} \times (t_1\mathbf{w}_1 + t_2\mathbf{w}_2) = t_1(\mathbf{v} \times \mathbf{w}_1) + t_2(\mathbf{v} \times \mathbf{w}_2)$$

c) Annullamento: $\mathbf{v} \times \mathbf{w} = \mathbf{0}$ se e solo se \mathbf{v} e \mathbf{w} sono paralleli.

In particolare, $\mathbf{v} \times \mathbf{v} = \mathbf{0}$.

Mentre le proprietà a) e c) seguono immediatamente dalla definizione, la bilinearità del prodotto vettoriale è più difficile da verificare (cf. l'esercizio 26). Dalle proprietà del prodotto vettoriale se ne deduce l'espressione in coordinate cartesiane:

Espressione in coordinate cartesiane del prodotto vettoriale

Supponiamo che $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ siano una terna ortonormale destrorsa. Il prodotto vettoriale di

$$\mathbf{v} = x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k} \quad \text{e} \quad \mathbf{w} = x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k}$$

è il vettore

$$(7.1) \quad \mathbf{v} \times \mathbf{w} = (y_1z_2 - y_2z_1)\mathbf{i} - (x_1z_2 - x_2z_1)\mathbf{j} + (x_1y_2 - x_2y_1)\mathbf{k}$$

DIMOSTRAZIONE. Dalla bilinearità del prodotto vettoriale segue:

$$\begin{aligned} & (x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k}) \times (x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k}) = \\ &= x_1\mathbf{i} \times (x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k}) + y_1\mathbf{j} \times (x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k}) + z_1\mathbf{k} \times (x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k}) = \\ &= x_1y_2\mathbf{i} \times \mathbf{j} + x_1z_2\mathbf{i} \times \mathbf{k} + y_1x_2\mathbf{j} \times \mathbf{i} + y_1z_2\mathbf{j} \times \mathbf{k} + z_1x_2\mathbf{k} \times \mathbf{i} + z_1y_2\mathbf{k} \times \mathbf{j} = \\ &= (y_1z_2 - y_2z_1)\mathbf{i} - (x_1z_2 - x_2z_1)\mathbf{j} + (x_1y_2 - x_2y_1)\mathbf{k} \end{aligned}$$

Per ricordare l'espressione in coordinate del prodotto vettoriale, poniamo

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

$$\begin{vmatrix} A & B & C \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{vmatrix} = \begin{vmatrix} y_1 & z_1 \\ y_2 & z_2 \end{vmatrix} A - \begin{vmatrix} x_1 & z_1 \\ x_2 & z_2 \end{vmatrix} B + \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} C$$

dove A, B e C possono essere numeri o vettori. Queste espressioni definiscono il *determinante* di una matrice 2×2 e 3×3 , rispettivamente. L'importanza dei determinanti va ben oltre il legame con il prodotto vettoriale.

Possiamo ora riscrivere l'espressione in coordinate (7.1) del prodotto vettoriale come segue:

$$(7.2) \quad \mathbf{v} \times \mathbf{w} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{vmatrix} = \begin{vmatrix} y_1 & z_1 \\ y_2 & z_2 \end{vmatrix} \mathbf{i} - \begin{vmatrix} x_1 & z_1 \\ x_2 & z_2 \end{vmatrix} \mathbf{j} + \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} \mathbf{k}$$

Un primo motivo per cui i determinanti sono importanti è che forniscono uno strumento per calcolare aree e volumi:

PROPOSIZIONE 7.2 Nel piano cartesiano, la misura dell'area del parallelogramma di lati $\overrightarrow{OP} \equiv [a, b]^T$ e $\overrightarrow{OQ} \equiv [c, d]^T$ è il valore assoluto di $ad - bc$.

DIMOSTRAZIONE. Pensiamo il piano xy come al piano di equazione $z = 0$ nello spazio cartesiano. La misura dell'area in questione è il modulo del vettore $\mathbf{v} \times \mathbf{w}$ per definizione di prodotto vettoriale. Calcoliamo tale prodotto vettoriale:

$$\mathbf{v} \times \mathbf{w} = \begin{bmatrix} a \\ b \\ 0 \end{bmatrix} \times \begin{bmatrix} c \\ d \\ 0 \end{bmatrix} = (ad - bc)\mathbf{k}$$

Infine $(ad - bc)\mathbf{k}$ ha modulo uguale al valore assoluto di $ad - bc$ perché \mathbf{k} è un versore.

Si può ricavare la stessa formula con un calcolo più diretto, sfruttando la

Identità di Lagrange

$$\|\mathbf{v} \times \mathbf{w}\|^2 = \|\mathbf{v}\|^2 \|\mathbf{w}\|^2 - (\mathbf{v} \cdot \mathbf{w})^2$$

Dimostriamo l'identità di Lagrange: se α è l'angolo formato da \mathbf{v} e \mathbf{w} , il modulo del prodotto vettoriale è $\|\mathbf{v}\| \|\mathbf{w}\| |\sin(\alpha)|$, mentre

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos(\alpha)$$

Quindi l'identità di Lagrange segue dal fatto che $\sin^2(\alpha) = 1 - \cos^2(\alpha)$.

Possiamo infine ricavare la formula per l'area del parallelogramma a partire dal fatto che $ab + cd = \mathbf{v} \cdot \mathbf{w}$:

$$\begin{aligned} \|\mathbf{v} \times \mathbf{w}\|^2 &= \|\mathbf{v}\|^2 \|\mathbf{w}\|^2 - (\mathbf{v} \cdot \mathbf{w})^2 = (a^2 + b^2)(c^2 + d^2) - (ac + bd)^2 \\ &= a^2d^2 + b^2c^2 - 2acbd = (ad - bc)^2 \end{aligned}$$

Passiamo ora al calcolo del volume di un parallelepipedo. A questo scopo introduciamo il *prodotto misto o triplo* di tre vettori, che in coordinate è il determinante di una matrice 3×3 .

DEFINIZIONE 7.3 (Prodotto misto di tre vettori)

Il prodotto misto di tre vettori \mathbf{u} , \mathbf{v} e \mathbf{w} è lo scalare $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$.

In coordinate cartesiane, se $\mathbf{u} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} d \\ e \\ f \end{bmatrix}$ e $\mathbf{w} = \begin{bmatrix} g \\ h \\ l \end{bmatrix}$, allora

$$(7.3) \quad \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & l \end{vmatrix}$$

L'espressione in coordinate cartesiane si ricava così

$$\begin{aligned} \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) &= \begin{bmatrix} a \\ b \\ c \end{bmatrix} \cdot \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ d & e & f \\ g & h & l \end{vmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \cdot \left(\begin{vmatrix} e & f \\ h & l \end{vmatrix} \mathbf{i} - \begin{vmatrix} d & f \\ g & l \end{vmatrix} \mathbf{j} + \begin{vmatrix} d & e \\ g & h \end{vmatrix} \mathbf{k} \right) = \\ &= \begin{vmatrix} e & f \\ h & l \end{vmatrix} a - \begin{vmatrix} d & f \\ g & l \end{vmatrix} b + \begin{vmatrix} d & e \\ g & h \end{vmatrix} c = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & l \end{vmatrix} \end{aligned}$$

Il prodotto misto calcola un volume, più precisamente:

PROPOSIZIONE 7.4 Consideriamo nello spazio tre vettori $\mathbf{u} = \overrightarrow{OA}$, $\mathbf{v} = \overrightarrow{OB}$ e $\mathbf{w} = \overrightarrow{OC}$. Allora:

- a) il valore assoluto del prodotto misto $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$ è uguale al volume del parallelepipedo di spigoli \overrightarrow{OA} , \overrightarrow{OB} e \overrightarrow{OC} ;
- b) se il prodotto misto è non nullo, il suo segno è positivo se la terna $\mathbf{u}, \mathbf{v}, \mathbf{w}$ è destrorsa, negativo se la terna è sinistrorsa.

DIMOSTRAZIONE. Il valore assoluto del prodotto misto è

$$||\mathbf{v}|| ||\mathbf{w}|| |\sin(\alpha)| ||\mathbf{u}|| |\cos(\beta)|$$

dove α è l'angolo tra \mathbf{v} e \mathbf{w} e β è l'angolo tra \mathbf{u} e $\mathbf{v} \times \mathbf{w}$. Ora $||\mathbf{v}|| ||\mathbf{w}|| |\sin(\alpha)|$ è l'area di base del parallelepipedo; mentre $||\mathbf{u}|| |\cos(\beta)|$ è la lunghezza della proiezione di \mathbf{u} nella direzione perpendicolare alla base ed è quindi l'altezza del parallelepipedo. Il volume del parallelepipedo è uguale al prodotto base per altezza, quindi, al valore assoluto del prodotto misto dei tre vettori. Questo dimostra il primo punto.

Per determinare il segno del prodotto misto, ragioniamo così. Posto $\mathbf{n} = \mathbf{v} \times \mathbf{w}$, la terna $\mathbf{v}, \mathbf{w}, \mathbf{n}$ è destrorsa per definizione di prodotto vettoriale. Quindi, tale è la terna $\mathbf{n}, \mathbf{v}, \mathbf{w}$ (ruotare il dito medio al posto del pollice). Il prodotto misto è per definizione il prodotto scalare $\mathbf{u} \cdot \mathbf{n}$, quindi ha segno positivo se la proiezione ortogonale di \mathbf{u} lungo \mathbf{n} ha lo stesso verso di \mathbf{n} , cioè se $\mathbf{u}, \mathbf{v}, \mathbf{w}$ è una terna destrorsa.

La proposizione precedente specifica il valore assoluto e il segno del prodotto misto e quindi lo determina completamente: il prodotto misto è quel numero il cui valore assoluto è il volume del parallelepipedo generato da \mathbf{u} , \mathbf{v} e \mathbf{w} e il cui segno è positivo o negativo a seconda che \mathbf{u} , \mathbf{v} e \mathbf{w} formino una terna destrorsa o sinistrorsa.

Come conseguenza importante, ricaviamo un criterio per stabilire, nella terminologia del paragrafo 4, quando \mathbf{u} , \mathbf{v} e \mathbf{w} formano una base di \mathbb{R}^3 :

PROPOSIZIONE 7.5 Consideriamo tre vettori liberi \mathbf{u} , \mathbf{v} e \mathbf{w} . Allora:

- a) i tre vettori sono linearmente indipendenti se e solo se il loro prodotto misto $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$ è diverso da zero;
- b) se $\mathbf{v} \times \mathbf{w} \neq \mathbf{0}$, ma $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = 0$, allora \mathbf{u} è combinazione lineare di \mathbf{v} e \mathbf{w} .

DIMOSTRAZIONE. Applichiamo i tre vettori nell'origine: $\mathbf{u} \equiv \overrightarrow{OA}$, $\mathbf{v} \equiv \overrightarrow{OB}$ e $\mathbf{w} \equiv \overrightarrow{OC}$. Per la proposizione 4.3 i tre vettori sono indipendenti se e solo se i punti O , A , B e C non sono complanari e questo succede se e solo se il parallelepido da essi generato ha volume diverso da zero. Il punto a) segue dal fatto che tale volume è il modulo del prodotto misto $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$.

Supponiamo ora che $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = 0$. Per il punto a) i tre vettori sono linearmente dipendenti, quindi esistono tre scalari non tutti nulli tali che

$$a\mathbf{u} + b\mathbf{v} + c\mathbf{w} = \mathbf{0}$$

Se $a = 0$, allora $b\mathbf{v} + c\mathbf{w} = \mathbf{0}$ con b e c non entrambi nulli. Ma allora \mathbf{v} e \mathbf{w} sono paralleli, contraddicendo l'ipotesi $\mathbf{v} \times \mathbf{w} \neq \mathbf{0}$. Quindi dev'essere $a \neq 0$ e

$$\mathbf{u} = -\frac{b}{a}\mathbf{v} - \frac{c}{a}\mathbf{w}$$

è combinazione lineare di \mathbf{v} e \mathbf{w} .

Possiamo infine definire rigorosamente la nozione di terna destrorsa, in funzione di una base fissata $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ (intuitivamente, il pollice, l'indice e il medio della mano destra). Data una terna di vettori $\mathbf{u}, \mathbf{v}, \mathbf{w}$, definiamo il loro prodotto misto mediante la formula (7.3) in termini delle coordinate rispetto alla base $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$. Diciamo che la terna $\mathbf{u}, \mathbf{v}, \mathbf{w}$ è destrorsa se il prodotto misto ha segno positivo, sinistrorsa se il segno è negativo.

Determinare un vettore \mathbf{w} perpendicolare ai due vettori $\mathbf{u} = [1, 2, 3]^T$ e $\mathbf{v} = [3, 2, 1]^T$. Verificare il risultato controllando che $\mathbf{u} \cdot \mathbf{w} = \mathbf{v} \cdot \mathbf{w} = 0$.

Determinare l'area del parallelogramma che ha tre vertici in $A = (0, 1, 2)$, $B = (0, 2, 3)$ e $C = (1, 2, 3)$ calcolando il modulo del prodotto vettore $\overrightarrow{AB} \times \overrightarrow{AC}$. Controllare che si ottiene lo stesso risultato calcolando il modulo di $\overrightarrow{AC} \times \overrightarrow{BC}$. Posto $D = (2, 2, 2)$, calcolare il volume del parallelepipedo di spigoli \overrightarrow{AB} , \overrightarrow{AC} , \overrightarrow{AD}

Nel piano cartesiano siano $A = (1, 2)$, $B = (2, 4)$, $C = (-2, 7)$. Calcolare l'area del triangolo ABC .

Suggerimento: l'area di un triangolo è la metà dell'area di un parallelogramma.

Calcolare il volume del tetracdro di vertici $A = (1, 2, 0)$, $B = (2, 4, 0)$, $C = (-2, 7, 0)$, $D = (1, 1, 1)$.

Suggerimento: il volume di un tetraedro è un terzo del prodotto dell'area di base per l'altezza; quindi è un sesto del volume di un parallelepipedo.

Per quali valori del parametro $a \in \mathbb{R}$ i tre vettori $\mathbf{u} = [1, 2, a]^T$, $\mathbf{v} = [2, 4, 7]^T$ e $\mathbf{w} = [0, 0, 1]^T$ formano una base di \mathbb{R}^3 ?

Mostrare che il prodotto misto ha le seguenti simmetrie:

$$\begin{aligned}\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) &= -\mathbf{v} \cdot (\mathbf{u} \times \mathbf{w}) \\ \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) &= \mathbf{w} \cdot (\mathbf{u} \times \mathbf{v}) = \mathbf{v} \cdot (\mathbf{u} \times \mathbf{w})\end{aligned}$$

Il prodotto vettoriale non è un'operazione associativa: trovare tre vettori \mathbf{u} , \mathbf{v} e \mathbf{w} tali che

$$(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} \neq \mathbf{u} \times (\mathbf{v} \times \mathbf{w})$$

Mostrare l'uguaglianza

$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\mathbf{w}$$

Suggerimento: ridursi al caso $\mathbf{v} = \mathbf{i}$ e in cui $\mathbf{w} = ai + bj$ appartenga al piano xy .

Siano \mathbf{v} e \mathbf{w} due vettori non paralleli. Mostrare che $\mathbf{v} \times \mathbf{w}$ è perpendicolare a tutte le combinazioni lineari $x\mathbf{v} + y\mathbf{w}$ (geometricamente, se $\mathbf{v} \equiv \overrightarrow{OA}$ e $\mathbf{w} \equiv \overrightarrow{OB}$, $\mathbf{v} \times \mathbf{w}$ è normale al piano OAB).

Supponiamo che due piani \mathbf{H}_1 e \mathbf{H}_2 si intersechino nella retta \mathbf{r} . Mostrare che il prodotto vettoriale dei vettori normali di \mathbf{H}_1 e \mathbf{H}_2 è parallelo a \mathbf{r} .

Dimostrare che il prodotto vettoriale è bilineare completando la seguente traccia

- a) usando la definizione di prodotto vettoriale si mostri che per ogni scalare t

$$t(\mathbf{v} \times \mathbf{w}) = (t\mathbf{v}) \times \mathbf{w} = \mathbf{v} \times (t\mathbf{w})$$

e se ne deduca che per controllare la bilinearità del prodotto vettoriale basta verificare che

$$(\mathbf{u} + \mathbf{v}) \times \mathbf{k} = \mathbf{u} \times \mathbf{k} + \mathbf{v} \times \mathbf{k}$$

sotto l'ipotesi aggiuntiva che \mathbf{k} sia un versore.

- b) sia \mathbf{H} il piano perpendicolare al versore \mathbf{k} (si può assumere che \mathbf{k} sia il versore dell'asse z e quindi che \mathbf{H} sia il piano xy); si mostri che per ogni vettore \mathbf{u} parallelo ad \mathbf{H} , il prodotto vettoriale $\mathbf{u} \times \mathbf{k}$ si ottiene ruotando \mathbf{u} di $\pi/2$ (in senso orario rispetto a \mathbf{k}) nel piano \mathbf{H} . Se ne deduca che

$$(\mathbf{u} + \mathbf{v}) \times \mathbf{k} = \mathbf{u} \times \mathbf{k} + \mathbf{v} \times \mathbf{k}$$

se \mathbf{u} e \mathbf{v} sono paralleli ad \mathbf{H} .

- c) Sia $\mathbf{u}_{\mathbf{H}}$ la proiezione ortogonale del vettore \mathbf{u} sul piano \mathbf{H} : se \mathbf{H} è il piano xy , $(x, y, z)_{\mathbf{H}} = (x, y, 0)$. Si mostri

$$\mathbf{u} \times \mathbf{k} = \mathbf{u}_{\mathbf{H}} \times \mathbf{k}$$

- d) Si conclude ora che

$$\begin{aligned}(\mathbf{u} + \mathbf{v}) \times \mathbf{k} &\stackrel{c)}{=} (\mathbf{u} + \mathbf{v})_{\mathbf{H}} \times \mathbf{k} = (\mathbf{u}_{\mathbf{H}} + \mathbf{v}_{\mathbf{H}}) \times \mathbf{k} = \\ &\stackrel{b)}{=} \mathbf{u}_{\mathbf{H}} \times \mathbf{k} + \mathbf{v}_{\mathbf{H}} \times \mathbf{k} = \mathbf{u} \times \mathbf{k} + \mathbf{v} \times \mathbf{k}\end{aligned}$$

8 GEOMETRIA ANALITICA DI RETTE E PIANI NELLO SPAZIO

In questo paragrafo supponiamo di aver fissato un sistema di riferimento nello spazio con origine nel punto O . Un punto P è determinato dalle sue coordinate (x, y, z) . Non assumiamo a priori che il sistema di riferimento sia cartesiano.

8.1 Equazioni parametriche di una retta

Consideriamo la retta \mathbf{r} passante per il punto A e diretta come il vettore \mathbf{v} . Un punto $P = (x, y, z)$ appartiene alla retta \mathbf{r} se e solo se \overrightarrow{AP} è parallelo a \mathbf{v} , cioè se esiste uno scalare $t \in \mathbb{R}$ tale che $\overrightarrow{AP} \equiv t\mathbf{v}$.

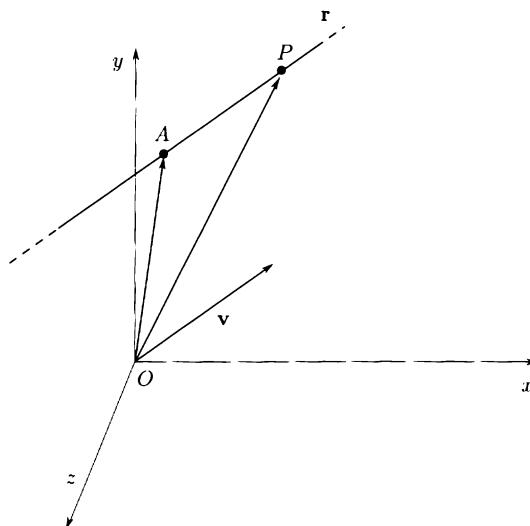


Figura 8.1. $\overrightarrow{OP} = \overrightarrow{OA} + t\mathbf{v}$.

Sommmando a entrambi i membri di questa uguaglianza il vettore \overrightarrow{OA} , otteniamo

$$(8.1) \quad \overrightarrow{OP} \equiv \overrightarrow{OA} + t\mathbf{v}$$

Scrivendo (8.1) in coordinate troviamo le *equazioni parametriche* della retta:

Equazioni parametriche di una retta nello spazio

Un punto $P = (x, y, z)$ appartiene alla retta \mathbf{r} passante per il punto $A = (x_A, y_A, z_A)$ e diretta come il vettore $\mathbf{v} = [a, b, c]^T$ se e soltanto se esiste uno scalare t tale che

$$(8.2) \quad \begin{cases} x = x_A + at \\ y = y_A + bt \\ z = z_A + ct \end{cases}$$

In particolare, P appartiene alla retta AB se e solo se esiste $t \in \mathbb{R}$ tale che:

$$(8.3) \quad \begin{cases} x = x_A + (x_B - x_A)t \\ y = y_A + (y_B - y_A)t \\ z = z_A + (z_B - z_A)t \end{cases}$$

I coefficienti a , b e c con cui compare il parametro t nelle equazioni (8.12) sono le componenti del vettore direzione \mathbf{v} : si dicono *parametri direttori* della retta. Il vettore direzione non è univocamente determinato dalla retta \mathbf{r} : ogni altro vettore parallelo a $[a, b, c]^T$, cioè un vettore di coordinate $[\lambda a, \lambda b, \lambda c]^T$ con $\lambda \neq 0$, è un vettore direzione di \mathbf{r} . Anche x_A , y_A e z_A possono essere sostituiti con le coordinate di un altro punto della retta \mathbf{r} . Le equazioni parametriche sono perciò tutt'altro che uniche.

Per la retta \mathbf{r} che congiunge i punti $A = (1, 2, 0)$ e $B = (0, 0, -3)$ le equazioni (8.3) sono:

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

Le coordinate di A e di B si ritrovano ponendo $t = 0$ e $t = 1$ (così si può controllare di aver scritto correttamente le equazioni). Altre equazioni parametriche per la stessa retta sono

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

che si ottengono pensando a \mathbf{r} come alla retta che passa per B con vettore direzione \overrightarrow{BA} . Questa volta B corrisponde a $u = 0$, A a $u = 1$.

OSSERVAZIONE È tradizionale scrivere le equazioni nella forma (8.2), ma è chiaro che x, y, z sono funzioni $x(t), y(t), z(t)$ della variabile t . Le equazioni parametriche definiscono la funzione $P(t) : \mathbb{R} \rightarrow \mathbb{R}^3$ che al numero $t \in \mathbb{R}$ associa il punto dello spazio

$$P(t) = (x_A + at, y_A + bt, z_A + ct)$$

La retta \mathbf{r} è l'immagine di tale funzione in \mathbb{R}^3 , cioè l'insieme dei punti $P(t)$ al variare di t in \mathbb{R} . Il vettore derivato $P'(t)$ di tale funzione è costante e coincide con il vettore direzione \mathbf{v} . Se immaginiamo che la retta sia percorsa da una particella che al tempo t si trova nel punto $P(t)$, il vettore velocità del moto è costante nel tempo e uguale al vettore direzione \mathbf{v} .

Coseni direttori di una retta

Si hanno dei vantaggi se si sceglie un vettore direzione \mathbf{v} che sia un *versore*, cioè che abbia modulo 1:

- a) se \mathbf{v} è un versore, la distanza di due punti $P(t_1)$ e $P(t_2)$ della retta è esattamente $|t_2 - t_1|$ perché:

$$\|\overrightarrow{P(t_1)P(t_2)}\| = \|\overrightarrow{OP(t_2)} - \overrightarrow{OP(t_1)}\| = \|t_2\mathbf{v} - t_1\mathbf{v}\| = |t_2 - t_1| \|\mathbf{v}\| = |t_2 - t_1|$$

L'interpretazione fisica è chiara: il punto $P(t)$ percorre la retta con velocità scalare $\|\mathbf{v}\|$. Se tale velocità è 1, la distanza percorsa in un intervallo di tempo Δt è $|\Delta t|$:

- b) se le coordinate sono cartesiane e $\mathbf{v} = [c_1, c_2, c_3]^T$ è un versore, l'ascissa c_1 di \mathbf{v} è il coseno dell'angolo α che \mathbf{v} forma col semiasse positivo delle ascisse. Infatti

$$c_1 = x(\mathbf{v}) = \mathbf{v} \cdot \mathbf{i} = \|\mathbf{v}\| \|\mathbf{i}\| \cos(\alpha) = \cos(\alpha)$$

Per un motivo analogo $c_2 = y(\mathbf{v})$ e $c_3 = z(\mathbf{v})$ sono i coseni degli angoli che \mathbf{v} forma coi semiassi positivi delle ordinate e delle quote. Si dice che c_1 , c_2 e c_3 sono i *coseni direttori* della retta. I coseni direttori sono univocamente determinati a meno del segno (perché \mathbf{v} e $-\mathbf{v}$ sono gli unici versori paralleli alla retta).

Il prezzo da pagare è che per trasformare un vettore $\mathbf{w} = [a, b, c]^T$ in un versore occorre dividere \mathbf{w} per lo scalare $\|\mathbf{w}\| = \sqrt{a^2 + b^2 + c^2}$.

Posizione reciproca di due rette nello spazio

Come abbiamo visto nel paragrafo 2, due rette distinte nello spazio possono esserci:

- incidenti (un punto in comune, complanari);
- parallele (nessun punto in comune, ma complanari);
- sghembe (non complanari).

È semplice stabilire se due rette siano parallele quando ne siano note delle equazioni parametriche: infatti due rette sono parallele se e solo se i loro vettori direzione sono paralleli. Esplicitamente, due rette con vettori direzione $\mathbf{v} = [a, b, c]^T$ e $\mathbf{w} = [a', b', c']^T$ sono parallele se e solo se esiste uno scalare t tale che $a = ta'$, $b = tb'$ e $c = tc'$. Sappiamo anche che questo è equivalente all'annullarsi del prodotto vettoriale $\mathbf{v} \times \mathbf{w}$.

D'altra parte, non è immediato riconoscere dalle equazioni parametriche se due siano sghembe o incidenti, come mostra l'esempio seguente.

Consideriamo i punti $A = (0, 1, 2)$, $B = (1, 3, 5)$, $C = (a, 0, 3)$ e $D = (a+6, 2, 1)$, dove a è un parametro reale. Sia \mathbf{r}_1 la retta per A e B e sia \mathbf{r}_2 la retta per C e D . Vogliamo determinare gli eventuali valori di a per cui le due rette sono incidenti. A questo scopo scriviamo le equazioni parametriche delle due rette, con l'accortezza di denotare i parametri con lettere diverse, diciamo s per \mathbf{r}_1 e t per \mathbf{r}_2 :

$$\mathbf{r}_1 : \begin{cases} x = s \\ y = 1 + 2s \\ z = 2 + 3s \end{cases} \quad \mathbf{r}_2 : \begin{cases} x = a + 6t \\ y = 2t \\ z = 3 - 2t \end{cases}$$

Ora un punto $P = (x, y, z)$ appartiene a entrambe le rette se e solo se le sue coordinate si possono esprimere in funzione sia del parametro s sia del parametro t ; quindi, se e solo se esistono s e t tali che

$$\begin{cases} x(P) = s = a + 6t \\ y(P) = 1 + 2s = 2t \\ z(P) = 2 + 3s = 3 - 2t \end{cases}$$

Otteniamo così un sistema lineare nelle incognite s e t , che possiamo riscrivere portando a primo membro i termini con le variabili e a secondo membro i termini noti:

$$\begin{cases} s - 6t = a \\ 2s - 2t = -1 \\ 3s + 2t = 1 \end{cases}$$

Sommmando le ultime due equazioni troviamo $s = 0$ e sostituendo nel sistema troviamo $t = \frac{1}{2}$ e $s = -\frac{a}{6}$: concludiamo che il sistema ha la soluzione $(s, t) = \left(0, \frac{1}{2}\right)$ se $a = -3$ e non ha soluzioni se $a \neq -3$.

Quindi, se $a = -3$, le due rette sono incidenti e si incontrano nel punto $A = (0, 1, 2)$ che corrisponde al valore $s = 0$ nelle equazioni di \mathbf{r}_1 e a $t = \frac{1}{2}$ nelle equazioni di \mathbf{r}_2 . Le due rette non coincidono perché i vettori direzione delle due rette sono $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ e $\begin{bmatrix} 6 \\ 2 \\ 2 \end{bmatrix}$ che non sono paralleli.

Se invece $a \neq -3$, le due rette non sono incidenti e non sono nemmeno parallele perché i vettori direzione non sono paralleli. Quindi, se $a \neq -3$, le due rette sono sghembe.

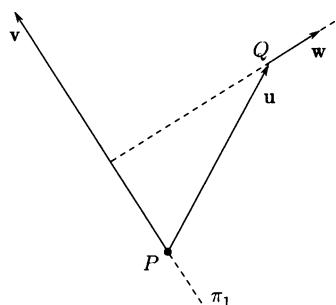


Figura 8.2. Figura di due rette sghembe e del vettore \mathbf{u} : la prima retta passa per P ed è diretta come \mathbf{v} , la seconda passa per Q ed è diretta come \mathbf{w} ; infine $\mathbf{u} = \overrightarrow{PQ}$.

Utilizzando prodotto scalare e prodotto vettoriale, possiamo determinare la posizione reciproca di due rette senza dover risolvere alcun sistema lineare. Infatti supponiamo che \mathbf{r}_1 sia la retta per P con vettore direzione \mathbf{v} e che \mathbf{r}_2 sia la retta per Q con vettore direzione \mathbf{w} . Sia $\mathbf{u} \equiv \overrightarrow{PQ}$ il vettore che congiunge il punto P della prima retta col punto Q della seconda retta. Le due rette sono complanari se e solo se il parallelepipedo generato da \mathbf{u} , \mathbf{v} e \mathbf{w} ha volume nullo. Quindi le due rette sono sghembe se e solo se il prodotto misto $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$ è diverso da zero.

Riassumendo

Posizione reciproca di due rette nello spazio

Siano \mathbf{r}_1 la retta per P con vettore direzione \mathbf{v} ed \mathbf{r}_2 la retta per Q con vettore direzione \mathbf{w} . Sia $\mathbf{u} \equiv \overrightarrow{PQ}$. Allora

- \mathbf{r}_1 e \mathbf{r}_2 sono parallele se $\mathbf{v} \times \mathbf{w} = \mathbf{0}$;
- \mathbf{r}_1 e \mathbf{r}_2 sono incidenti in un punto se $\mathbf{v} \times \mathbf{w} \neq \mathbf{0}$, ma $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = 0$;
- \mathbf{r}_1 e \mathbf{r}_2 sono sghembe se $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) \neq 0$.

Nell'esempio precedente,

$$\mathbf{u} \equiv \overrightarrow{AC} = \begin{bmatrix} a \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} 6 \\ 2 \\ -2 \end{bmatrix}$$

per cui

$$\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \begin{vmatrix} a & 1 & 6 \\ -1 & 2 & 2 \\ 1 & 3 & -2 \end{vmatrix} = -10a - 0 - 30 = -10(a + 3)$$

Concludiamo che le due rette sono sghembe se e solo se $a \neq -3$, in accordo con quanto stabilito precedentemente.

Rette nel piano cartesiano

In modo analogo si trovano le equazioni parametriche di una retta in un piano: la differenza è che ci sono due coordinate anzichè tre. La retta per il punto (x_A, y_A) con vettore direzione $\mathbf{v} = [p, q]^T$ ha equazioni parametriche $P(t) = (x_A + pt, y_A + qt)$.

Tutte le rette che non sono parallele all'asse delle y possono essere parametrizzate usando come parametro l'ascissa x . Infatti una retta è parallela all'asse y se e solo se è nulla la prima coordinata p del suo vettore direzione \mathbf{v} . Se $p \neq 0$, dall'equazione $x = x_A + pt$ possiamo ricavare $t = \frac{x-x_A}{p}$. Sostituendo nell'espressione di y e ponendo $m = \frac{q}{p}$, otteniamo le equazioni parametriche $P(x) = (x, y_A + m(x - x_A))$. Si noti che queste equazioni si ottengono pensando \mathbf{r} come la retta per il punto $(0, y_A - mx_A)$ diretta come il vettore $[1, m]^T$. In conclusione:

Il luogo dei punti $P = (x, y)$ del piano le cui coordinate soddisfano l'equazione

$$(8.4) \quad y = mx + c$$

è la retta per $(0, c)$ con vettore direzione $[1, m]^T$

Anche una retta \mathbf{r} parallela all'asse y ha un'equazione cartesiana: se la retta taglia l'asse x nel punto $(c, 0)$, l'equazione di \mathbf{r} è $x = c$. Non è in realtà necessario trattare questo caso separatamente. Per le rette non parallele all'asse y l'equazione (8.4) si può riscrivere nella forma $-mx + y = c$ e, quindi, *ogni* retta del piano ha un'equazione cartesiana *lineare*, cioè della forma

$$ax + by = d$$

con coefficienti a e b non entrambi nulli: $a = 1$ e $b = 0$ nel caso dell'equazione $x = c$, $a = -m$ e $b = 1$ nel caso dell'equazione $-mx + y = c$. Viceversa:

Se a e b sono due numeri reali non entrambi nulli, il luogo dei punti $P = (x, y)$ del piano le cui coordinate soddisfano l'equazione

$$(8.5) \quad ax + by = d$$

è una retta con vettore direzione $[b, -a]^T$.

Infatti, se $a \neq 0$, l'equazione (8.5) è equivalente a $y = mx + c$ con $m = -\frac{a}{b}$ e $c = \frac{d}{b}$. Il vettore direzione della retta è allora

$$\begin{bmatrix} 1 \\ m \end{bmatrix} = \frac{1}{b} \begin{bmatrix} b \\ -a \end{bmatrix}$$

che è parallelo a $[b, -a]^T$. Se invece $b = 0$, l'equazione è equivalente a $x = \frac{d}{a}$, che è l'equazione di un retta parallela all'asse y e, quindi, con vettore direzione

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = -\frac{1}{a} \begin{bmatrix} b \\ -a \end{bmatrix}.$$

OSSERVAZIONE L'equazione cartesiana (8.5) di una retta non è univocamente determinata perché l'insieme delle soluzioni non cambia se si moltiplica l'equazione per uno scalare non nullo. Ma questa è la sola ambiguità: due rette coincidono se e solo se le loro equazioni cartesiane differiscono per la moltiplicazione per uno scalare non nullo.

OSSERVAZIONE Se le coordinate sono cartesiane (cioè i vettori \mathbf{i} e \mathbf{j} della base del sistema di riferimento sono perpendicolari e hanno modulo 1), allora $m = \operatorname{tg}(\alpha)$ dove α è l'angolo formato da \mathbf{i} , cioè dal semiasse positivo delle ascisse, col vettore direzione della retta: m si dice *coefficiente angolare* della retta. Inoltre:

in coordinate cartesiane, la retta di equazione $ax + by = d$ è perpendicolare al vettore $\mathbf{n} = [a, b]^T$ dei coefficienti dell'equazione.

Questo segue dal fatto che il vettore direzione della retta è $\mathbf{v} = [b, -a]^T$ e il prodotto scalare di questi due vettori è nullo:

$$\mathbf{n} \cdot \mathbf{v} = ab + b(-a) = 0$$

Il modo forse più veloce di scrivere l'equazione cartesiana della retta per due punti A e B è il seguente. Siccome la retta è diretta come \overrightarrow{AB} , l'equazione ha la forma

$$(y_B - y_A)x - (x_B - x_A)y = d$$

Occorre determinare ancora d e per questo basta imporre che (x_A, y_A) risolva l'equazione.

Per esempio, se $A = (1, 0)$ e $B = (3, 3)$, allora la retta ha equazione della forma

$$3x - 2y = d$$

e imponendo il passaggio per A ricaviamo

$$3 = d$$

Quindi la retta ha equazione $3x - 2y = 3$, o anche

$$y = \frac{3}{2}x - \frac{3}{2}$$

Parametrizzazione di un segmento

Un punto P appartiene alla retta AB se $\overrightarrow{AP} = t\overrightarrow{AB}$; i punti del segmento AB sono quelli che hanno coordinata t compresa tra 0 e 1.

In termini dei vettori posizione rispetto all'origine l'uguaglianza precedente si riscrive

$$\overrightarrow{OP} = \overrightarrow{OA} + t\overrightarrow{AB} = \overrightarrow{OA} + t(\overrightarrow{OB} - \overrightarrow{OA}) = (1-t)\overrightarrow{OA} + t\overrightarrow{OB}$$

Quindi possiamo riscrivere le equazioni parametriche della retta in un modo in cui i ruoli di A e B sono più simmetrici:

Sia P un punto della retta per A e B . Allora esiste uno e un solo scalare $t \in \mathbb{R}$ tale che

$$(8.6) \quad \overrightarrow{OP} = (1-t)\overrightarrow{OA} + t\overrightarrow{OB}$$

I punti del segmento AB corrispondono ai valori di t compresi tra 0 e 1.

Fissiamo $t \in (0, 1)$ e consideriamo il corrispondente punto P sul segmento AB : allora $\overrightarrow{AP} = t\overrightarrow{AB}$ e

$$\overrightarrow{PB} = \overrightarrow{AB} - \overrightarrow{AP} = (1-t)\overrightarrow{AB}$$

Quindi

$$\overrightarrow{AP} : \overrightarrow{PB} = t : (1-t)$$

ovvero il punto P divide il segmento AB in due segmenti il rapporto delle cui lunghezze è uguale a $\frac{t}{1-t}$. In particolare, per $t = \frac{1}{2}$ otteniamo il punto medio del segmento. Sostituendo in (8.6) vediamo che il vettore posizione del punto medio di AB è la media aritmetica dei vettori posizione di A e di B :

Il punto medio M del segmento AB ha vettore posizione

$$(8.7) \quad \overrightarrow{OM} = \frac{1}{2}(\overrightarrow{OA} + \overrightarrow{OB})$$

Baricentro di un triangolo

Consideriamo un triangolo ABC . Vogliamo mostrare che le tre mediane di un triangolo si incontrano in un punto. Una mediana del triangolo è il segmento che congiunge un vertice col punto medio del lato opposto. Poniamo $\mathbf{v} \equiv \overrightarrow{AB}$ e $\mathbf{w} \equiv \overrightarrow{AC}$. Detti M ed N i punti di medi dei segmenti AB e AC , allora

$$\overrightarrow{AM} = \frac{1}{2} \overrightarrow{AB} \equiv \frac{1}{2} \mathbf{v} \quad \text{e} \quad \overrightarrow{AN} = \frac{1}{2} \overrightarrow{AC} \equiv \frac{1}{2} \mathbf{w}$$

Sia P il punto di intersezione delle mediane CM e BN . Siccome P appartiene a CM , per (8.6) il suo vettore posizione rispetto ad A ha la forma

$$\overrightarrow{AP} = (1-t) \overrightarrow{AC} + t \overrightarrow{AM} \equiv (1-t)\mathbf{w} + \frac{t}{2}\mathbf{v}$$

per un unico scalare t . Poiché il punto P appartiene anche alla mediana BN , possiamo scrivere:

$$\overrightarrow{AP} = (1-s) \overrightarrow{AB} + s \overrightarrow{AN} \equiv (1-s)\mathbf{v} + \frac{s}{2}\mathbf{w}$$

per un unico scalare s . Uguagliando le due espressioni di \overrightarrow{AP} concludiamo

$$(1-t)\mathbf{w} + \frac{t}{2}\mathbf{v} = (1-s)\mathbf{v} + \frac{s}{2}\mathbf{w}$$

ovvero

$$\left(1-t-\frac{s}{2}\right)\mathbf{w} = \left(1-s-\frac{t}{2}\right)\mathbf{v}$$

Se $1-t-\frac{s}{2}$ fosse diverso da zero, dividendo l'uguaglianza precedente per $1-t-\frac{s}{2}$ potremmo scrivere \mathbf{w} come multiplo di \mathbf{v} e, quindi, \mathbf{w} e \mathbf{v} sarebbero paralleli. Questo è assurdo perché \mathbf{w} e \mathbf{v} sono diretti come i lati AC e AB del triangolo. Quindi lo scalare $1-t-\frac{s}{2}$ è nullo. Per un motivo analogo $1-s-\frac{t}{2}=0$. Perciò s e t soddisfano il sistema lineare

$$\begin{cases} s + \frac{t}{2} = 1 \\ t + \frac{s}{2} = 1 \end{cases}$$

Tale sistema ammette l'unica soluzione $s = t = \frac{2}{3}$. Quindi il punto P di intersezione delle due mediane divide entrambi i segmenti CM e BN nella proporzione $2 : 1$. Per simmetria anche il punto di intersezione della mediana CM con la terza mediana AQ del triangolo deve dividere il segmento CM nella proporzione $2 : 1$ e, quindi, è ancora il punto P . Questo mostra che le tre mediane si incontrano in uno stesso punto P , che è detto *baricentro* del triangolo.

Il vettore posizione del baricentro ha un'espressione simmetrica rispetto ai vertici del triangolo: per ricavarla, sostituiamo $t = \frac{2}{3}$ nell'espressione precedente di \overrightarrow{AP} trovando

$$\overrightarrow{AP} = \frac{1}{3} \overrightarrow{AC} + \frac{2}{3} \overrightarrow{AM} = \frac{1}{3} \overrightarrow{AC} + \frac{2}{3} \frac{1}{2} \overrightarrow{AB} = \frac{1}{3} (\overrightarrow{AB} + \overrightarrow{AC}).$$

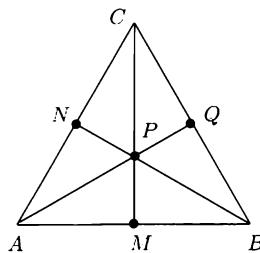


Figura 8.3. Le tre mediane di un triangolo si incontrano nel baricentro.

Sommando a entrambi i membri \overrightarrow{OA} troviamo l'espressione desiderata:

$$(8.8) \quad \overrightarrow{OP} = \frac{1}{3} (\overrightarrow{OA} + \overrightarrow{OB} + \overrightarrow{OC})$$

8.2 Equazione cartesiana di un piano

Vogliamo ora determinare l'equazione di un piano nello spazio cartesiano. Un piano \mathbf{H} nello spazio è determinato se conosciamo

- 1) tre punti non allineati A, B, C di \mathbf{H} , oppure
- 2) un punto $A \in \mathbf{H}$ e una base $\{\mathbf{v}, \mathbf{w}\}$ di \mathbf{H} , oppure
- 3) un punto $A \in \mathbf{H}$ e un vettore \mathbf{n} normale ad \mathbf{H} .

Ricordiamo che una base di \mathbf{H} è una coppia di vettori \mathbf{v} e \mathbf{w} paralleli ad \mathbf{H} , ma non paralleli tra loro. L'equivalenza dei dati in 1) e 2) si ottiene ponendo $\mathbf{v} \equiv \overrightarrow{AB}$ e $\mathbf{w} \equiv \overrightarrow{AC}$, come spiegato nel paragrafo 4. L'equivalenza di 2) e 3) segue dal fatto che un vettore è parallelo ad \mathbf{H} , cioè è combinazione lineare di \mathbf{v} e \mathbf{w} , se e solo se è perpendicolare a \mathbf{n} , cioè a $\mathbf{v} \times \mathbf{w}$ (cf. la proposizione 7.5).

Dati un punto $A \in \mathbf{H}$ e un vettore normale ad \mathbf{H} , è immediato scrivere un'equazione cartesiana di \mathbf{H} :

Equazione cartesiana di un piano

Supponiamo siano fissate nello spazio delle coordinate cartesiane.

- a) Il punto $P = (x, y, z)$ appartiene al piano \mathbf{H} passante per $A = (x_A, y_A, z_A)$ e perpendicolare a $\mathbf{n} = [a, b, c]^T$ se e solo se:

$$a(x - x_A) + b(y - y_A) + c(z - z_A) = 0$$

- b) Fissati quattro numeri reali a, b, c e d , con $(a, b, c) \neq (0, 0, 0)$, il luogo dei punti $P = (x, y, z)$ le cui coordinate soddisfano l'equazione lineare

$$ax + by + cz = d$$

è un piano con vettore normale $[a, b, c]^T$.

DIMOSTRAZIONE.

a). La condizione che $P = (x, y, z)$ appartenga a \mathbf{H} è che il vettore \overrightarrow{AP} sia contenuto in \mathbf{H} , il che equivale a dire che $\overrightarrow{AP} \cdot \mathbf{n} = 0$. In coordinate questa uguaglianza si scrive

$$\begin{bmatrix} x - x_A \\ y - y_A \\ z - z_A \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = a(x - x_A) + b(y - y_A) + c(z - z_A) = 0$$

b). Sia \mathbf{H} il luogo dei punti $P = (x, y, z)$ che soddisfano l'equazione: $ax + by + cz = d$. Cerchiamo innanzitutto un punto A di \mathbf{H} . Se $a \neq 0$, il punto $A = (\frac{d}{a}, 0, 0)$ soddisfa l'equazione e quindi appartiene a \mathbf{H} . Se $a = 0$, allora $b \neq 0$ oppure $c \neq 0$ e possiamo prendere per A il punto $(0, \frac{d}{b}, 0)$ oppure il punto $(0, 0, \frac{d}{c})$. Scriviamo $A = (x_A, y_A, z_A)$.

Abbiamo determinato A in modo che $ax_A + by_A + cz_A = d$. Sostituendo nell'equazione di \mathbf{H} troviamo $ax + by + cz = ax_A + by_A + cz_A$ e quindi

$$a(x - x_A) + b(y - y_A) + c(z - z_A) = 0$$

Dal punto a) segue che \mathbf{H} è il piano per A con vettore normale $[a \ b \ c]^T$.

OSSERVAZIONE Si ricordi che nel piano cartesiano vale un risultato analogo: il luogo dei punti le cui coordinate soddisfano l'equazione $ax + by = d$ è una retta perpendicolare al vettore $[a \ b]^T$.

OSSERVAZIONE L'equazione cartesiana di un piano è determinata a meno di una costante non nulla: se $ax + by + cz = d$ e $a'x + b'y + c'z = d'$ sono due equazioni dello stesso piano, allora esiste un numero reale $t \neq 0$ tale che $a' = ta$, $b' = tb$, $c' = tc$, $d' = td$. Brevemente, due equazioni individuano lo stesso piano se e solo se l'una è un multiplo dell'altra.

Infatti, l'esistenza di $t \neq 0$ tale che $a' = ta$, $b' = tb$, $c' = tc$ segue dal fatto che due vettori normali al piano sono paralleli. Se poi A è un punto del piano, le sue coordinate soddisfano entrambe le equazioni, quindi

$$td = t(ax_A + by_A + cz_A) = a'x_A + b'y_A + c'z_A = d'$$

Il piano \mathbf{H} passante per tre punti non allineati A, B e C è il piano per A con vettore normale $\mathbf{n} \equiv \overrightarrow{AB} \times \overrightarrow{AC}$; la condizione che un punto P appartenga ad \mathbf{H} è

$$(8.9) \quad \overrightarrow{AP} \cdot (\overrightarrow{AB} \times \overrightarrow{AC}) = 0$$

Questa equazione si può interpretare così il membro di sinistra è il prodotto misto dei tre vettori \overrightarrow{AP} , \overrightarrow{AB} e \overrightarrow{AC} ; il punto P appartiene al piano per A, B e C se e solo se il parallelepipedo con spigoli \overrightarrow{AP} , \overrightarrow{AB} e \overrightarrow{AC} ha volume nullo.

Scrivendo la (8.9) in coordinate si ottiene la formula esplicita (8.10) per l'equazione del piano \mathbf{H} . Si tenga però presente che nello svolgere gli esercizi può esserci più semplice calcolare le componenti del vettore normale $\mathbf{n} = \overrightarrow{AB} \times \overrightarrow{AC}$ e poi scrivere l'equazione del piano perpendicolare a \mathbf{n} e passante per A .

Equazione cartesiana di un piano per tre punti

Un punto $P = (x, y, z)$ appartiene al piano per i tre punti non allineati A, B e C se e solo se

$$(8.10) \quad \begin{vmatrix} x - x_A & y - y_A & z - z_A \\ x_B - x_A & y_B - y_A & z_B - z_A \\ x_C - x_A & y_C - y_A & z_C - z_A \end{vmatrix} = 0$$

L'equazione del piano \mathbf{H} per i tre punti $A = (1, 0, -1)$, $B = (2, 1, 2)$ e $C = (-1, 2, -1)$ è

$$\begin{aligned} 0 &= \begin{vmatrix} x - 1 & y & z + 1 \\ 1 & 1 & 3 \\ -2 & 2 & 0 \end{vmatrix} = (x - 1) \begin{vmatrix} 1 & 3 \\ 2 & 0 \end{vmatrix} - y \begin{vmatrix} 1 & 3 \\ -2 & 0 \end{vmatrix} + (z + 1) \begin{vmatrix} 1 & 1 \\ -2 & 2 \end{vmatrix} = \\ &= -6(x - 1) - 6y + 4(z + 1) = -6x - 6y + 4z + 10 \end{aligned}$$

ovvero:

$$3x + 3y - 2z = 5$$

Negli esercizi si può verificare velocemente di non avere fatto errori sostituendo nell'equazione trovata le coordinate dei punti: se non si sono sbagliati i conti, le coordinate di A, B e C soddisfano l'equazione.

Supponiamo ora di dover determinare l'intersezione del piano \mathbf{H} con la retta di equazioni parametriche $P(t) = (1 + t, 2 + t, 3 + t)$. Per questo basta trovare il valore di t per cui le coordinate di $P(t)$ soddisfano l'equazione del piano. Ora

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

per cui $P(t)$ appartiene ad \mathbf{H} se e solo se $t = \frac{1}{2}$: la retta e il piano si incontrano nel punto $P(\frac{1}{2}) = (\frac{3}{2}, \frac{5}{2}, \frac{7}{2})$.

OSSERVAZIONE Se usando la formula (8.10) si trova un'equazione con tutti i coefficienti delle variabili nulli, allora il prodotto vettore $\overrightarrow{AB} \times \overrightarrow{AC}$ è nullo e quindi i tre punti A, B, C sono allineati. Esistono in questo caso infiniti piani passanti per A, B, C : quelli contenenti la retta per A, B e C .

8.3 Equazioni cartesiane di una retta nello spazio

Due piani \mathbf{H}_1 e \mathbf{H}_2 sono paralleli se e solo se hanno la stessa direzione normale. Possiamo quindi prendere lo stesso vettore normale per entrambi i piani e scrivere equazioni cartesiane $ax + by + cz = d_1$ e $ax + by + cz = d_2$ per i due piani. Se $d_1 = d_2$, i due piani coincidono, mentre se $d_1 \neq d_2$ le due equazioni sono incompatibili e i

due piani non hanno punti in comune. Se invece i due piani non sono paralleli, si intersecano in una retta:

Equazioni cartesiane di una retta nello spazio

Supponiamo che i vettori $\mathbf{n}_1 = [a_1, b_1, c_1]^T$ e $\mathbf{n}_2 = [a_2, b_2, c_2]^T$ non siano paralleli. Allora il luogo dei punti $P = (x, y, z)$ dello spazio le cui coordinate risolvono il sistema lineare

$$(8.11) \quad \begin{cases} a_1x + b_1y + c_1z = d_1 \\ a_2x + b_2y + c_2z = d_2 \end{cases}$$

è una retta con vettore direzione $\mathbf{n}_1 \times \mathbf{n}_2$.

Viceversa, ogni retta è il luogo dei punti le cui coordinate soddisfano due equazioni della forma (8.11).

DIMOSTRAZIONE. Siano \mathbf{H}_1 e \mathbf{H}_2 i piani di equazione cartesiana $a_1x + b_1y + c_1z = d_1$ e $a_2x + b_2y + c_2z = d_2$. Il luogo dei punti le cui coordinate soddisfano il sistema lineare (8.11) è l'intersezione dei due piani; per ipotesi i due piani non sono paralleli, quindi si intersecano lungo una retta. Tale retta è perpendicolare a \mathbf{n}_1 perché è contenuta in \mathbf{H}_1 , ed è perpendicolare a \mathbf{n}_2 perché è contenuta in \mathbf{H}_2 . Quindi $\mathbf{n}_1 \times \mathbf{n}_2$ è diretto come la retta intersezione dei due piani.

Viceversa, data una retta \mathbf{r} , esiste un punto P dello spazio non appartenente a \mathbf{r} . Sia \mathbf{H}_1 il piano per P contenente \mathbf{r} . Esiste un punto Q dello spazio che non appartiene ad \mathbf{H}_1 . Sia \mathbf{H}_2 il piano per Q contenente \mathbf{r} . I due piani non sono paralleli (perché contengono la retta \mathbf{r} e non coincidono), quindi la loro intersezione è una retta che contiene \mathbf{r} e pertanto coincide con \mathbf{r} .

Risolvere il sistema lineare (8.11) è equivalente a trovare le equazioni parametriche della retta \mathbf{r} . Viceversa, eliminando il parametro t dalle tre equazioni parametriche di una retta se ne possono ricavare le (due) equazioni cartesiane. Illustriamo questo punto con un esempio specifico, rimandando considerazioni più generali e dettagliate al capitolo sui sistemi lineari.

I piani di equazione $x+2y+3z=0$ e $x-y+4z=0$ si intersecano in una retta \mathbf{r} perché i vettori normali $[1, 2, 3]^T$ e $[1, -1, 4]^T$ non sono paralleli. Per ricavare le equazioni parametriche della retta \mathbf{r} risolviamo il sistema lineare:

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

Sottraendo la prima equazione dalla seconda, otteniamo il sistema equivalente

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

Per ogni valore t assegnato arbitrariamente a y , la seconda equazione permette di ricavare $z = 3t$. Sostituendo nella prima equazione troviamo $x = 2 - 11t$. La retta ha quindi equa-

zioni parametriche $(x(t), y(t), z(t)) = (2 - 11t, t, 3t)$. Si può controllare l'esattezza dei conti accertandosi che $(x(t), y(t), z(t))$ verifichi le equazioni dei due piani.

Supponiamo invece di voler ricavare delle equazioni cartesiane per la retta di equazioni parametriche $(x(t), y(t), z(t)) = (1 + t, 2 + t, 3 + t)$. Possiamo ricavare $t = x - 1$ dalla prima equazione e sostituendo nelle altre due equazioni, troviamo

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

Quindi la retta ha equazioni cartesiane $x - y + 1 = x - z + 2 = 0$.

8.4 Equazioni parametriche di un piano nello spazio

Abbiamo visto che i punti di un piano si possono esprimere in funzione di due parametri: se \mathbf{H} è il piano per il punto A generato da \mathbf{v} e \mathbf{w} , un punto P appartiene a \mathbf{H} se e solo se esistono due scalari t_1 e t_2 tali che

$$\overrightarrow{AP} \equiv t_1 \mathbf{v} + t_2 \mathbf{w}$$

Sommendo \overrightarrow{OA} otteniamo l'espressione del vettore posizione di P rispetto all'origine O del sistema di riferimento dello spazio:

$$\overrightarrow{OP} \equiv \overrightarrow{OA} + t_1 \mathbf{v} + t_2 \mathbf{w}$$

Riscrivendo questa uguaglianza componente per componente si ottengono le equazioni parametriche del piano:

Equazioni parametriche di un piano

Un punto $P = (x, y, z)$ appartiene al piano \mathbf{H} per il punto $A = (x_A, y_A, z_A)$ e generato dai vettori $\mathbf{v} = [a_1, b_1, c_1]^T$ e $\mathbf{w} = [a_2, b_2, c_2]^T$ se e soltanto se esiste uno scalare t tale che

$$(8.12) \quad \begin{cases} x = x_A + a_1 t_1 + a_2 t_2 \\ y = y_A + b_1 t_1 + b_2 t_2 \\ z = z_A + c_1 t_1 + c_2 t_2 \end{cases}$$

Si noti che il punto A del piano corrisponde ai valori $t_1 = 0$ e $t_2 = 0$ dei parametri.

È sempre possibile eliminare t_1 e t_2 dalle equazioni parametriche di un piano per ottenerne l'equazione cartesiana

$$(8.13) \quad ax + by + cz = d$$

Naturalmente il vettore $[a, b, c]^T$ è a meno di un multiplo il *prodotto vettoriale* dei due vettori \mathbf{v} e \mathbf{w} che generano il piano.

Viceversa, dall'equazione cartesiana di un piano se ne possono ricavare delle equazioni parametriche. Per questo basta risolvere l'equazione (8.13). C'è una difficoltà: l'equazione ha infinite soluzioni, in corrispondenza biunivoca coi punti del piano. Per esprimere le soluzioni dell'equazione abbiamo perciò bisogno di due parametri. Se per

esempio il coefficiente a di x nell'equazione è diverso da zero, possiamo ricavare x in funzione di y e z ed esprimere le soluzioni di (8.13) in funzione dei parametri $t_1 = y$ e $t_2 = z$. Il risultato è

$$(8.14) \quad \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{d}{a} - \frac{b}{a}t_1 - \frac{c}{a}t_2 \\ t_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} \frac{d}{a} \\ 0 \\ 0 \end{bmatrix} + t_1 \begin{bmatrix} -\frac{b}{a} \\ 1 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} -\frac{c}{a} \\ 0 \\ 1 \end{bmatrix} = \overrightarrow{OA} + t_1 \mathbf{v} + t_2 \mathbf{w}$$

dove $A = (\frac{d}{a}, 0, 0)$, $\mathbf{v} = [-\frac{b}{a}, 1, 0]^T$ e $\mathbf{w} = [-\frac{c}{a}, 0, 1]^T$. Queste sono le equazioni parametriche per il piano di equazione cartesiana (8.5).

Risolvere la (8.13) significa, sostanzialmente, trovare un sistema di riferimento $(A, \mathbf{v}, \mathbf{w})$. Osserviamo che la base $\{\mathbf{v}, \mathbf{w}\}$ così ottenuta *non* è in generale ortonormale: si è così portati, in modo naturale, a considerare sistemi di riferimento non cartesiani.



Consideriamo, come nell'esempio di p. 56, il piano **H** di equazione

$$3x + 3y - 2z = 5$$

Possiamo ricavare x in funzione di y e z e scrivere come sopra le coordinate del punto generico del piano nella forma

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{5}{3} - t_1 + \frac{2}{3}t_2 \\ t_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} \frac{5}{3} \\ 0 \\ 0 \end{bmatrix} + t_1 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} \frac{2}{3} \\ 0 \\ 1 \end{bmatrix}$$

Un sistema di riferimento del piano **H** è $(A, \mathbf{v}, \mathbf{w})$ dove $A = (\frac{5}{3}, 0, 0)$, $\mathbf{v} = [-1, 1, 0]^T$ e $\mathbf{w} = [\frac{2}{3}, 0, 1]^T$.

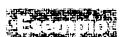
8.5 Distanza di due punti. Equazione di una sfera

Il modulo del vettore $\mathbf{v} = [x, y, z]^T$ è:

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{x^2 + y^2 + z^2}.$$

La distanza $d(A, B)$ tra due punti A e B dello spazio è il modulo del vettore $\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA}$, quindi

$$(8.15) \quad d(A, B) = \|\overrightarrow{AB}\| = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2 + (z_B - z_A)^2}.$$



Equazione di una sfera

La *sfera* di centro C e raggio $R > 0$ è l'insieme dei punti che hanno distanza da C pari a R . Dalla (8.15) segue che un punto di coordinate (x, y, z) appartiene a tale sfera se e solo se

$$(8.16) \quad (x - x_C)^2 + (y - y_C)^2 + (z - z_C)^2 = R^2$$

Sviluppando i quadrati si può riscrivere l'equazione nella forma

$$(8.17) \quad x^2 + y^2 + z^2 + ax + by + cz + d = 0$$

dove $a = -2x_C$, $b = -2y_C$, $c = -2z_C$ e $d = x_C^2 + y_C^2 + z_C^2 - R^2$.

Viceversa, se $4d < a^2 + b^2 + c^2$, il luogo dei punti le cui coordinate soddisfano (8.17) è la sfera di centro $C = (-\frac{a}{2}, -\frac{b}{2}, -\frac{c}{2})$ e raggio $R = \frac{1}{2}\sqrt{a^2 + b^2 + c^2 - 4d}$. Piuttosto che ricordare questa formula, per trovare centro e raggio conviene completare i quadrati e riportare la (8.17) alla forma della (8.16).

Per esempio, per trovare il centro e il raggio della sfera di equazione $x^2 + x + y^2 + z^2 = 0$, completiamo i quadrati e riscriviamo l'equazione nella forma:

$$\left(x + \frac{1}{2}\right)^2 + y^2 + z^2 = \frac{1}{4}$$

Si tratta perciò della sfera di centro $C = (-\frac{1}{2}, 0, 0)$ e raggio $R = \frac{1}{2}$. Per controllare i conti, notiamo che $(0, 0, 0)$ soddisfa l'equazione $x^2 + x + y^2 + z^2 = 0$ e, in effetti, l'origine dista $\frac{1}{2} = R$ dal centro C e quindi appartiene alla sfera.

8.6 Distanza tra un punto e un piano

La distanza $d(P, H)$ tra un punto P e un piano H nello spazio è per definizione il minimo delle distanze $d(P, A)$ di P da un punto A del piano. Tale minimo esiste ed è uguale a $d(P, B)$ dove B è l'intersezione del piano H con la retta per P perpendicolare al piano H : infatti, per ogni altro punto A di H , BP è perpendicolare a AB e, quindi, il triangolo ABP è rettangolo con ipotenusa AP ;

$$d(P, A) = \overline{AP} > \overline{BP} = d(P, B)$$

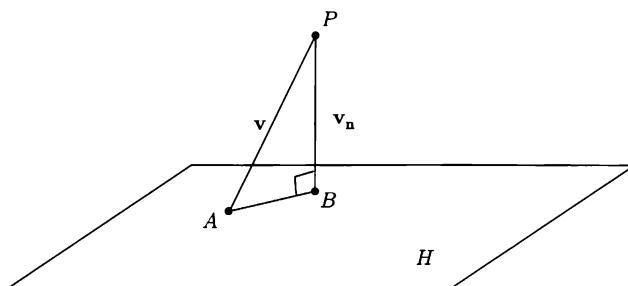


Figura 8.4. Distanza di un punto da un piano.

In termini di vettori,

$$\overrightarrow{AP} = \overrightarrow{AB} + \overrightarrow{BP}$$

con \overrightarrow{BP} e \overrightarrow{AB} perpendicolari tra loro. Quindi \overrightarrow{BP} è la proiezione ortogonale v_n del vettore $v = \overrightarrow{AP}$ nella direzione normale al piano. La distanza $d(P, H)$ è perciò uguale al modulo di tale proiezione ortogonale. Questo consente di ricavare una formula esplicita per la distanza:

Distanza di un punto da un piano

La distanza di un punto $P = (x_P, y_P, z_P)$ dal piano \mathbf{H} di equazione $ax + by + cz = d$ è:

$$d(P, \mathbf{H}) = \frac{|ax_P + by_P + cz_P - d|}{\sqrt{a^2 + b^2 + c^2}}$$

DIMOSTRAZIONE. Come già detto in precedenza, fissiamo un punto A del piano. Il vettore $\mathbf{n} = [a, b, c]^T$ è perpendicolare al piano. La distanza cercata è la norma della proiezione ortogonale \mathbf{v}_n del vettore $\mathbf{v} \equiv \overrightarrow{AP}$ nella direzione normale al piano.

La proiezione ortogonale è il vettore

$$\mathbf{v}_n = \frac{\mathbf{v} \cdot \mathbf{n}}{\mathbf{n} \cdot \mathbf{n}} \mathbf{n}$$

la cui norma è

$$\|\mathbf{v}_n\| = \frac{|\mathbf{v} \cdot \mathbf{n}|}{\mathbf{n} \cdot \mathbf{n}} \|\mathbf{n}\| = \frac{|\mathbf{v} \cdot \mathbf{n}|}{\|\mathbf{n}\|}$$

Ora

$$\mathbf{v} \cdot \mathbf{n} = \begin{bmatrix} x_P - x_A \\ y_P - y_A \\ z_P - z_A \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = a(x_P - x_A) + b(y_P - y_A) + c(z_P - z_A).$$

Siccome A appartiene al piano, $ax_A + by_A + cz_A = d$, quindi

$$\mathbf{v} \cdot \mathbf{n} = ax_P + by_P + cz_P - d$$

per cui

$$d(P, \mathbf{H}) = \|\mathbf{v}_n\| = \frac{|ax_P + by_P + cz_P - d|}{\sqrt{a^2 + b^2 + c^2}}$$

OSSERVAZIONE Nel piano, vale una formula analoga per la distanza di un punto $P = (x, y)$ dalla retta \mathbf{r} di equazione $ax + by = c$:

$$d(P, \mathbf{r}) = \frac{|ax + by - c|}{\sqrt{a^2 + b^2}}$$

8.7 Distanza tra un punto e una retta nello spazio

La distanza $d(P, \mathbf{r})$ tra un punto P e una retta \mathbf{r} nello spazio è per definizione il minimo delle distanze $d(P, Q)$ di P da un punto Q della retta. Tale minimo esiste: se C è il punto di intersezione di \mathbf{r} con il piano \mathbf{H} che passa per P ed è perpendicolare a \mathbf{r} , C è il punto di \mathbf{r} più vicino a P . Infatti, per costruzione, per ogni punto Q di \mathbf{r} diverso da C , il triangolo QCP è rettangolo, QP ne è l'ipotenusa, CP un cateto. Quindi $d(P, Q) > d(P, C)$ e perciò $d(P, \mathbf{r}) = d(P, C)$. Dati un punto A della retta e un suo vettore direzione \mathbf{w} possiamo calcolare tale distanza in due modi:

Distanza di un punto da una retta nello spazio

Sia \mathbf{r} la retta per A diretta come il vettore \mathbf{w} . La distanza $d(P, \mathbf{r})$ tra un punto P e la retta \mathbf{r} è:

$$(8.18) \quad d(P, \mathbf{r}) = \frac{\|\mathbf{v} \times \mathbf{w}\|}{\|\mathbf{w}\|} = \left\| \mathbf{v} - \frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}} \mathbf{w} \right\|$$

dove $\mathbf{v} \equiv \overrightarrow{AP}$.

DIMOSTRAZIONE. Sia C il punto di intersezione di \mathbf{r} con il piano \mathbf{H} che passa per P ed è perpendicolare a \mathbf{r} . Se θ è l'angolo \widehat{ACP} , allora:

$$d(P, \mathbf{r}) = \|\overrightarrow{CP}\| = \|\overrightarrow{AP}\| \sin(\theta) = \frac{\|\overrightarrow{AP} \times \mathbf{w}\|}{\|\mathbf{w}\|}$$

Per ottenere la seconda formula per la distanza, osserviamo che

$$\overrightarrow{AP} = \overrightarrow{AC} + \overrightarrow{CP}$$

e che \overrightarrow{AC} è parallelo a \mathbf{r} e \overrightarrow{CP} è perpendicolare ad essa. Quindi \overrightarrow{AC} è la proiezione ortogonale $\overrightarrow{AP}_{\mathbf{w}}$ di \overrightarrow{AP} su \mathbf{r} , per cui:

$$d(P, \mathbf{r}) = \|\overrightarrow{CP}\| = \|\mathbf{v} - \mathbf{v}_w\| = \left\| \mathbf{v} - \frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}} \mathbf{w} \right\|$$

Esempio

Consideriamo $P = (1, 0, 1)$ e la retta \mathbf{r} per $A = (1, 1, 0)$ e $B = (0, 1, 1)$. Allora $\mathbf{w} = \overrightarrow{AB} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ e $\overrightarrow{AP} = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$, quindi $\overrightarrow{AP} \times \overrightarrow{AB} = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$ e

$$d(P, \mathbf{r}) = \frac{\|\overrightarrow{AP} \times \overrightarrow{AB}\|}{\|\overrightarrow{AB}\|} = \frac{\sqrt{3}}{\sqrt{2}}$$

Possiamo anche usare la seconda formula per la distanza:

$$\mathbf{v} - \frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}} \mathbf{w} = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ -1 \\ \frac{1}{2} \end{bmatrix}$$

Ritroviamo così

$$d(P, \mathbf{r}) = \sqrt{\frac{1}{4} + 1 + \frac{1}{4}} = \sqrt{\frac{3}{2}}$$

OSSERVAZIONE Si può anche procedere determinando C esplicitamente. Questo ha l'unico vantaggio di non dover ricordare alcuna formula: è più elementare, ma più lungo:

- a) si determinano delle equazioni parametriche di \mathbf{r} ; si indicano con $R(t)$ le coordinate del punto di \mathbf{r} corrispondente al valore t del parametro;

- b) si scrive l'equazione del piano \mathbf{H} per P perpendicolare a \mathbf{r} : se $P = (x_P, y_P, z_P)$ e $[a, b, c]^T$ è un vettore direzione per \mathbf{r} , l'equazione di \mathbf{H} è

$$a(x - x_P) + b(y - y_P) + c(z - z_P) = 0;$$

- c) si determina il punto C in cui la retta \mathbf{r} interseca il piano \mathbf{H} : in pratica, si trova l'unico valore t_0 del parametro t per cui le coordinate del punto $R(t)$ soddisfano l'equazione del piano. Quindi $C = R(t_0)$;
- d) la distanza tra P e \mathbf{r} è uguale alla distanza tra P e C .

Nell'esempio precedente $P = (1, 0, 1)$ e \mathbf{r} è la retta per $A = (1, 1, 0)$ e $B = (0, 1, 1)$, che ha vettore direzione $[-1, 0, 1]^T$. Quindi $R(t) = (1 - t, 1, t)$ sono le equazioni parametriche della retta.

Il piano per P perpendicolare a \mathbf{r} ha equazione $-(x - 1) + (z - 1) = 0$ ovvero

$$x - z = 0$$

Cerchiamo il punto C in cui la retta \mathbf{r} interseca il piano \mathbf{H} : poiché C appartiene alla retta, $C = R(t) = (1 - t, 1, t)$ per un valore $t = t(C)$ del parametro. Ma C appartiene anche a \mathbf{H} per cui le sue coordinate devono soddisfare l'equazione del piano:

$$(1 - t) - t = 0$$

Troviamo $t(C) = 1/2$, per cui $C = (1/2, 1, 1/2)$ e

$$d(P, \mathbf{r}) = d(P, C) = \sqrt{(1 - 1/2)^2 + (0 - 1)^2 + (1 - 1/2)^2} = \sqrt{3/2}$$

8.8 Distanza tra due rette

Questo esempio è molto più complesso dei precedenti; capirlo richiede tempo ed energia, che saranno ripagati dall'acquisizione di una migliore intuizione per la geometria dello spazio. Si consiglia vivamente di ragionare sulla figura seguente

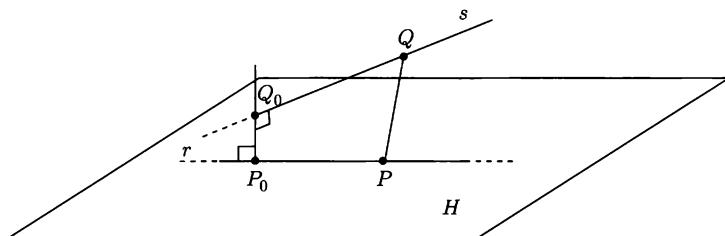


Figura 8.5. Distanza di due rette sghembe.

La distanza $d(\mathbf{r}, \mathbf{s})$ tra due rette nello spazio è per definizione il minimo della funzione $d(P, Q)$ al variare del punto P sulla retta \mathbf{r} e del punto Q sulla retta \mathbf{s} . Se le due rette sono parallele, tale minimo coincide con la distanza di un punto arbitrario P

della prima retta dalla seconda retta e abbiamo già visto come calcolare tale distanza. Supponiamo quindi che le due rette non siano parallele e siano \mathbf{v} e \mathbf{w} i vettori direzione delle due rette. Allora il vettore $\mathbf{n} = \mathbf{v} \times \mathbf{w}$ è non nullo e perpendicolare a entrambe le rette: in particolare, i tre vettori sono indipendenti e nessuno dei tre è combinazione lineare degli altri due. Sia \mathbf{H} il piano che contiene \mathbf{r} ed è parallelo a \mathbf{n} . La retta \mathbf{s} non è parallela a \mathbf{H} , perché \mathbf{w} non è combinazione lineare di \mathbf{v} ed \mathbf{n} e, quindi, interseca \mathbf{H} in un unico punto Q_0 . Nel piano \mathbf{H} la retta per Q_0 perpendicolare a \mathbf{r} ha direzione \mathbf{n} e taglia \mathbf{r} in un punto P_0 . La retta P_0Q_0 per costruzione è diretta come \mathbf{n} , quindi è perpendicolare sia a \mathbf{r} sia a \mathbf{s} ; si verificare facilmente che $d(P_0, Q_0) \leq d(P, Q)$ per ogni P in \mathbf{r} e ogni Q in \mathbf{s} . Quindi $d(\mathbf{r}, \mathbf{s}) = d(P_0, Q_0)$.

Per calcolare tale distanza non è necessario trovare esplicitamente i due punti P_0 e Q_0 : infatti, scelti arbitrariamente un punto P su \mathbf{r} e un punto Q su \mathbf{s} , il vettore $\overrightarrow{P_0Q_0}$ è la proiezione ortogonale $(\overrightarrow{PQ})_{\mathbf{n}}$ di \overrightarrow{PQ} lungo \mathbf{n} . Per me questo non è evidente, ma possiamo ragionare così. La proiezione ortogonale della somma di due vettori è la somma delle proiezioni ortogonali:

$$(\mathbf{a} + \mathbf{b})_{\mathbf{n}} = \mathbf{a}_{\mathbf{n}} + \mathbf{b}_{\mathbf{n}}$$

Questo segue per esempio dalla formula esplicita $\mathbf{u}_{\mathbf{n}} = \frac{\mathbf{u} \cdot \mathbf{n}}{\mathbf{n} \cdot \mathbf{n}} \mathbf{n}$ tenuto conto della bilinearità del prodotto scalare.

Poiché

$$\overrightarrow{PQ} = \overrightarrow{PP_0} + \overrightarrow{P_0Q_0} + \overrightarrow{Q_0Q}$$

la proiezione ortogonale $(\overrightarrow{PQ})_{\mathbf{n}}$ si decompone come

$$(\overrightarrow{PQ})_{\mathbf{n}} = (\overrightarrow{PP_0})_{\mathbf{n}} + (\overrightarrow{P_0Q_0})_{\mathbf{n}} + (\overrightarrow{Q_0Q})_{\mathbf{n}}$$

Il primo e il terzo addendo sono nulli perché $\overrightarrow{PP_0}$ e $\overrightarrow{Q_0Q}$ sono perpendicolari a \mathbf{n} , mentre il secondo addendo $(\overrightarrow{P_0Q_0})_{\mathbf{n}}$ coincide con $\overrightarrow{P_0Q_0}$ perché P_0 e Q_0 giacciono su una retta parallela a \mathbf{n} .

Concludiamo che $\overrightarrow{P_0Q_0}$ è la proiezione ortogonale di \overrightarrow{PQ} lungo \mathbf{n} . Possiamo quindi dedurre la formula per la distanza delle due rette:

$$(8.19) \quad d(\mathbf{r}, \mathbf{s}) = ||\overrightarrow{P_0Q_0}|| = ||(\overrightarrow{PQ})_{\mathbf{n}}|| = \frac{|\overrightarrow{PQ} \cdot \mathbf{n}|}{\sqrt{\mathbf{n} \cdot \mathbf{n}}}$$

Riassumendo:

Distanza di due rette non parallele

Sia \mathbf{r} la retta per P diretta come \mathbf{v} e sia \mathbf{s} la retta per Q diretta come \mathbf{w} . Se $\mathbf{v} \times \mathbf{w} \neq 0$, allora la distanza tra le due rette è:

$$(8.20) \quad d(\mathbf{r}, \mathbf{s}) = \frac{|\overrightarrow{PQ} \cdot (\mathbf{v} \times \mathbf{w})|}{\|\mathbf{v} \times \mathbf{w}\|}$$

Si osservi che l'ipotesi $\mathbf{v} \times \mathbf{w} \neq 0$ significa che le due rette non sono parallele. Se le rette sono parallele, la distanza tra le rette è uguale alla distanza della prima retta

da un punto qualsiasi della seconda. Se le due rette sono incidenti, il vettore \overrightarrow{PQ} è parallelo al piano che contiene le due rette, per cui $\overrightarrow{PQ} \cdot (\mathbf{v} \times \mathbf{w}) = 0$: la distanza di due rette incidenti è zero!



Nel piano cartesiano sia \mathbf{r} la retta per $A = (1, 0)$ diretta come $\mathbf{v} = [2, 3]^T$. Verificare che \mathbf{r} coincide con la retta di equazioni parametriche

$$\begin{cases} x = u \\ y = -\frac{3}{2} + \frac{3}{2}u \end{cases}$$

Si trovi l'equazione cartesiana del piano di \mathbb{R}^3 passante per i punti $P_1 = (0, 1, 2)$, $P_2 = (1, 2, 3)$, $P_3 = (1, 3, 5)$. Si determinino le componenti dei due versori normali al piano.

Trovare l'equazione del piano di \mathbb{R}^3 passante per il punto $P = (0, 1, 2)$ e perpendicolare alla retta di equazioni $x + y + z = x - 2y + 3z = 0$.



Si considerino i piani \mathbf{H}_1 di equazione $x - y + 3z - 5 = 0$ e \mathbf{H}_2 di equazione $2x + 2y - 4z = 0$.

- trovare una retta contenuta in \mathbf{H}_1 e parallela a \mathbf{H}_2 ;
- mostrare che non esiste una retta contenuta in \mathbf{H}_1 e perpendicolare a \mathbf{H}_2 ;
- trovare un piano perpendicolare a \mathbf{H}_1 e \mathbf{H}_2 .



Si considerino i piani \mathbf{H}_1 di equazione $x - y + 2z - 4 = 0$ e \mathbf{H}_2 di equazione $2x + 5y + z - 9 = 0$:

- trovare una retta parallela ad entrambi i piani;
- mostrare che non esiste una retta perpendicolare ad entrambi i piani.

Si determinino le equazioni cartesiane della retta \mathbf{r} di \mathbb{R}^3 perpendicolare al piano di equazione $x - 2y + z = 35245$ e passante per il punto $P = (1, 2, 3)$. Sia \mathbf{s} la retta di \mathbb{R}^3 passante per i punti $Q = (1, 2, 3)$ e $R = (3, 2, a)$ dove a è un parametro reale. Si stabilisca se \mathbf{r} e \mathbf{s} sono perpendicolari o parallele per qualche valore di a .



Trovare delle equazioni parametriche e delle equazioni cartesiane della retta di \mathbb{R}^3 passante per $P = (1, 2, 3)$ e perpendicolare al piano di equazione $x + 3y + z = 300$. Calcolare la distanza del punto P dal piano $x + y + z = 0$.

Calcolare la distanza del punto $P = (2, 3, 5)$ dalla retta \mathbf{r} di equazioni cartesiane $x + y - 3 = x - y + z - 6 = 0$.



Trovare le equazioni della retta passante per $P = (1, 1, 1)$ e perpendicolare alle rette \mathbf{r} , di equazioni $x = y + 2 = z - 3$ ed \mathbf{s} , di equazioni $x = 2y + 1 = 3z - 4$.



Mostrare che le diagonali di un parallelogramma si bisecano.

37 Determinare l'equazione del luogo dei punti $X = (x, y, z)$ equidistanti da $P = (2, 3, 1)$ e $Q = (-1, 2, 5)$ in due modi distinti: 1) imponendo che $\|PX\|^2 = \|QX\|^2$; 2) osservando che si tratta del piano perpendicolare a PQ che passa per il punto medio del segmento PQ .

38 Trovare centro e raggio della sfera di equazione $x^2 + y^2 + z^2 - 4x + 6y + 6 = 0$.

39 Trovare:

- l'equazione della sfera di centro $P = (1, 2, 4)$ e raggio 7. Fissato poi un parametro reale a , si consideri la retta \mathbf{r} di equazioni parametriche $(a+t, 2+t, 3+t)$. Si determinino i punti (se esistono) in cui la retta taglia la sfera;
- la distanza della retta \mathbf{r} dal centro della sfera. Confrontando tale distanza con il raggio, si stabilisca se \mathbf{r} taglia la sfera in due punti distinti, in un unico punto o in nessun punto. Si verifichi che il risultato ottenuto è in accordo col primo punto dell'esercizio.

40 Si considerino le rette \mathbf{r} ed \mathbf{s} di equazioni parametriche

$$\begin{cases} x = 3+2t \\ y = -2+t \\ z = 4+5t \end{cases} \quad \text{e} \quad \begin{cases} x = -3+2u \\ y = 2-u \\ z = 7-3u \end{cases}$$

Si calcoli la distanza di \mathbf{r} e \mathbf{s} e si stabilisca se sono incidenti, parallele o sghembe.

41 Trovare le equazioni parametriche ed equazioni cartesiane della retta \mathbf{r} di \mathbb{R}^3 passante per $P = (1, 2, 3)$ e perpendicolare al piano di equazione $x + y + 2z = 352$. Sia \mathbf{s} la retta di equazioni cartesiane $x + y - 3 = x - y + z - 6 = 0$. Si calcoli la distanza tra le due rette e si stabilisca se \mathbf{r} e \mathbf{s} sono sghembe (risp. incidenti, risp. parallele, risp. ortogonali).

42 Fascio di piani Sia \mathbf{r} la retta di equazioni cartesiane

$$\begin{cases} a_1x + b_1y + c_1z = d_1 \\ a_2x + b_2y + c_2z = d_2 \end{cases}$$

Il *fascio di piani per \mathbf{r}* è l'insieme dei piani che contengono \mathbf{r} . Si mostri che:

- Un piano \mathbf{H} contiene la retta \mathbf{r} se e solo la sua equazione è combinazione lineare delle due equazioni della retta, cioè ha la forma

$$(8.21) \quad \lambda(a_1x + b_1y + c_1z - d_1) + \mu(a_2x + b_2y + c_2z - d_2) = 0$$

dove λ e μ sono due numeri reali non entrambi nulli. La (8.21) si dice *equazione del fascio di piani per \mathbf{r}* .

Suggerimento: per mostrare che ogni piano \mathbf{H} del fascio per $\mathbf{r} = \mathbf{H}_1 \cap \mathbf{H}_2$ ha un'equazione della forma desiderata, considerare i vettori normali \mathbf{n} , \mathbf{n}_1 ed \mathbf{n}_2 dei tre piani; siccome \mathbf{H} contiene \mathbf{r} , il suo vettore normale \mathbf{n} è perpendicolare al vettore direzione $\mathbf{n}_1 \times \mathbf{n}_2$ della retta. Per la proposizione 7.5 $\mathbf{n} = \lambda\mathbf{n}_1 + \mu\mathbf{n}_2$; questo dà la relazione desiderata tra i coefficienti delle variabili nell'equazione. Infine testare l'equazione su un punto di \mathbf{r} per ottenerne la stessa relazione $d = \lambda d_1 + \mu d_2$ sui termini noti.

- Tutti i piani del fascio per \mathbf{r} , tranne il piano $a_2x + b_2y + c_2z = d_2$, hanno un'equazione della forma

$$(8.22) \quad a_1x + b_1y + c_1z - d_1 + t(a_2x + b_2y + c_2z - d_2) = 0$$

La (8.22) si dice *equazione affine del fascio di piani per \mathbf{r}* .

- c) I piani del fascio sono parametrizzati dalla circonferenza di raggio uno nel piano (λ, μ)

Suggerimento: dividere l'equazione del fascio per $\sqrt{\lambda^2 + \mu^2}$.

Trovare l'equazione cartesiana del piano per il punto $(3, 1, -7)$ contenente la retta di equazioni $x + y + 1 = 2x - z - 3 = 0$

Suggerimento: scrivere l'equazione affine del fascio di piani per la retta, quindi determinare t imponendo il passaggio per il punto.

Si trovi l'equazione del piano di \mathbb{R}^3 contenente i punti $P = (1, 3, 1)$ e $Q = (0, 4, 1)$ e perpendicolare al piano di equazione $x + y + z = 37$.

Suggerimento: scrivere l'equazione del fascio di piani per la retta PQ , quindi imporre la perpendicolarità dei vettori normali.

Per quali valori del parametro reale k i piani di \mathbb{R}^3 di equazioni $2x + ky + 4z = 4$, $3x + y + kz = k$ e $x - ky - 4z = k - 6$, si intersecano in una retta?

 Un altro metodo per calcolare la distanza tra due rette \mathbf{r} e \mathbf{s} è il seguente:

- si determina l'equazione cartesiana del piano \mathbf{K} che contiene \mathbf{r} e che è parallelo a \mathbf{s} (per esempio usando il fascio di piani per \mathbf{r} , se si conoscono le equazioni cartesiane di \mathbf{r});
 - si sceglie un punto qualsiasi Q della retta \mathbf{s} e si usa il fatto che $d(\mathbf{r}, \mathbf{s}) = d(K, Q)$ (perché la distanza tra K e Q è realizzata da un vettore parallelo a $\overrightarrow{P_0Q_0}$: fare il disegno).
 - si calcola la distanza $d(K, Q)$ usando la formula per la distanza tra un punto e un piano.
- Usando questo metodo ripetere gli esercizi sul calcolo della distanza di due rette.

Sistemi lineari

1 INTRODUZIONE

Un'equazione lineare è un'equazione algebrica di primo grado nelle incognite: nel caso di una incognita x ha la forma $ax = b$, nel caso di due incognite x,y la forma $ax + by = c$. Si tratta delle equazioni più semplici della matematica. L'equazione è una condizione lineare che devono soddisfare le incognite. Quando le incognite sono un numero n maggiore di 1, una sola condizione non basta a determinarle e per ottenere una soluzione unica occorre considerare un insieme di n equazioni. Questo conduce allo studio dei *sistemi lineari*, che consistono di più equazioni che le incognite devono soddisfare simultaneamente. Per esempio, per determinare l'età di due fratelli occorrono due condizioni, quali la somma e la differenza delle due età e questo si traduce in un sistema di due equazioni lineari in due incognite.

L'importanza dei sistemi lineari risiede nel fatto che esistono algoritmi in grado di risolverli (a differenza di quasi ogni altra equazione). Negli ultimi anni l'esplosione della capacità computazionale dei computers e dell'efficienza degli algoritmi ha reso possibile la soluzione in tempi accettabili di sistemi lineari con un numero di incognite sempre più grande (dell'ordine almeno delle centinaia di migliaia). Questo consente di approssimare sempre meglio la soluzione di problemi più complessi, spesso di natura applicativa, con la soluzione di un sistema lineare. Per questo lo studio dei sistemi lineari è tornato a giocare un ruolo centrale e molto attuale in matematica.

Non è però scopo di questo testo quello di illustrare metodi efficaci per la soluzione dei sistemi lineari, argomento che appartiene alla matematica numerica e comunque a corsi più avanzati. Il nostro interesse è di tipo teorico e le domande che ci devono guidare in questo capitolo sono: sotto quali condizioni il sistema ammette delle soluzioni? quando la soluzione è unica? e, se la soluzione non è unica, come si possono rappresentare tutte le soluzioni? Nel rispondere a queste domande è utile disporre dell'intuizione di un modello geometrico. Per questo ambienteremo la teoria in un analogo n dimensionale dello spazio cartesiano: rappresenteremo le incognite del sistema come un vettore $[x_1, \dots, x_n]^T$ con n componenti e denoteremo col simbolo \mathbb{R}^n lo *spazio* di tali vettori. Come nel caso di due (rispettivamente tre) variabili l'insieme delle soluzioni di un'equazione lineare $ax + by = c$ definisce una retta (rispettivamente un piano), così un'equazione lineare in x_1, \dots, x_n definisce un iperpiano in \mathbb{R}^n . L'in-

sieme delle soluzioni di un sistema di equazioni lineari in n incognite si rappresenta così come l'intersezione di iperpiani in \mathbb{R}^n e come tale ha una struttura geometrica semplice. Tale struttura si descrive facilmente una volta definite, anche per i vettori di \mathbb{R}^n , le operazioni di somma e prodotto per scalare (Teorema di Rouché-Capelli 6.8).

In questo capitolo illustriamo il *metodo di eliminazione di Gauss* che è un algoritmo per la soluzione di un arbitrario sistema lineare. L'algoritmo si fonda anch'esso sulle operazioni di somma di vettori e di prodotto di un vettore per uno scalare (i vettori in questo caso hanno come componenti i coefficienti delle equazioni del sistema). Il capitolo ha molti esempi che dovrebbero consentire allo studente di imparare a usare l'algoritmo e abituarlo a risolvere sistemi lineari di dimensioni piccole seguendo uno schema logico ordinato e non a caso come purtroppo spesso succede. In questo modo lo studente dovrebbe anche familiarizzarsi con l'algebra dei vettori, così da poter affrontare successivamente lo studio degli spazi vettoriali astratti con un'esperienza sufficiente del modello concreto di \mathbb{R}^n .

Usiamo l'eliminazione di Gauss per definire uno dei concetti più importanti dell'algebra lineare, il *rango*, che è il numero di equazioni che realmente contano nel sistema: da un sistema di m equazioni che abbia rango r si possono estrarre r equazioni indipendenti, le altre $m - r$ si possono ignorare perché sono conseguenza delle r indipendenti. La definizione che diamo di rango è operativa e consente di calcolare a mano, in modo efficiente, il rango di piccoli sistemi, ma è insoddisfacente dal punto di vista teorico: per esempio, non è ovvio dalla nostra definizione che cambiando l'ordine delle equazioni il rango rimanga invariato. A questo rimedieremo nel capitolo sugli spazi vettoriali quando daremo una definizione più concettuale di rango (come massimo numero di righe linearmente indipendenti di una matrice).

Una volta introdotto il rango r , il capitolo si conclude con i risultati fondamentali sui sistemi lineari: il teorema di Cramer 6.7, secondo cui un sistema di n equazioni in n incognite ammette un'unica soluzione se e solo se il rango è n e il teorema di Rouché-Capelli 6.8 che descrive la struttura dell'insieme delle soluzioni di un sistema. Le dimostrazioni sono costruttive e consistono, sostanzialmente, dell'algoritmo di soluzione di un sistema basato sull'eliminazione di Gauss. Dal punto di vista qualitativo, vedremo che i sistemi lineari possono essere determinati (un'unica soluzione), oppure sottodeterminati (infinte soluzioni, che dipendono da un numero di parametri pari al numero di incognite meno il rango), oppure sovradeterminati (nessuna soluzione, equazioni tra di loro incompatibili). Anche i sistemi sottodeterminati e sovradeterminati occorrono naturalmente in molte applicazioni e non sono, perciò, casi patologici che non meritino attenzione. I sistemi sottodeterminati compaiono per esempio, nella ricerca degli *autovettori*, che giocano un ruolo centrale nell'algebra lineare e nelle sue applicazioni. I sistemi sovradeterminati sorgono frequentemente quando gli errori di misura precludono l'esistenza di una soluzione esatta. In questo caso è sempre possibile determinare una soluzione approssimata, il più possibile coerente con i dati sperimentali.

2 EQUAZIONI LINEARI

Un'*equazione lineare* in n incognite x_1, x_2, \dots, x_n è un'equazione della forma

$$(2.1) \quad a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

In questa espressione a_1, \dots, a_n, b sono dei “numeri” fissati, gli a_i si dicono *coefficienti dell’equazione* e b si dice *termine noto*. L’equazione si dice *omogenea* se $b = 0$. L’equazione

$$(2.2) \quad a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0$$

si dice *equazione omogenea associata* a (2.1).

Una soluzione dell’equazione (2.1) è una n -upla ordinata $\bar{x}_1, \dots, \bar{x}_n$ di numeri per cui (2.1) è verificata:

$$a_1\bar{x}_1 + a_2\bar{x}_2 + \cdots + a_n\bar{x}_n = b$$

Per esempio la terna ordinata $x_1, x_2, x_3 = 1, 2, 3$ è una soluzione dell’equazione

$$3x_1 + 2x_2 + x_3 = 10$$

e $x_1, x_2, x_3 = 1, 2, -7$ è una soluzione dell’equazione omogenea associata

$$3x_1 + 2x_2 + x_3 = 0$$

Sarà conveniente scrivere le soluzioni come *vettori colonna*: per rappresentare una n -pla ordinata a_1, a_2, \dots, a_n useremo il simbolo

$$[a_1, a_2, \dots, a_n]^T = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

La n -upla $[0, \dots, 0]^T$ formata da n zeri si dice *vettore nullo* od *origine*; il vettore nullo è soluzione di ogni sistema omogeneo in x_1, x_2, \dots, x_n .

Insiemi numerici

È importante discutere in quale insieme ci proponiamo di risolvere un’equazione. Evidentemente un’equazione lineare $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$ ha senso in ogni insieme numerico in cui siano definiti somma e prodotto, per esempio, nell’insieme dei numeri interi \mathbb{Z} , in quello dei numeri razionali \mathbb{Q} , in quello dei numeri reali \mathbb{R} e in quello dei numeri complessi \mathbb{C} . Ma per risolvere l’equazione sono necessarie anche le operazioni inverse: sottrazione e divisione. Questo è già evidente in una sola variabile: per poter risolvere un’equazione di primo grado $ax + c = 0$ (con $a \neq 0$) nella variabile x occorre poter sottrarre $-c$ a entrambi i membri per poter riscrivere l’equazione nella forma lineare $ax = b$ e, quindi, dividere per a per trovare la soluzione:

$$x = -\frac{c}{a}$$

La risoluzione di un’equazione di primo grado richiede un insieme numerico in cui siano definite anche sottrazione e divisione. Non troveremmo nessuna soluzione dell’equazione $x + 3 = 0$ se cercassimo solo tra i numeri interi positivi, così come l’equazione $2x + 1 = 0$ non ha soluzione tra i numeri interi.

È quindi naturale studiare le equazioni lineari in un insieme in cui siano definite le quattro operazioni con le usuali proprietà dell’aritmetica: in algebra astratta un

insieme siffatto si dice . Esempi di campi sono: l'insieme dei numeri razionali \mathbb{Q} , che consiste di quei numeri che si possono rappresentare come frazioni di numeri interi, o, equivalentemente, che ammettono un'espansione decimale finita o periodica; l'insieme dei numeri reali \mathbb{R} , che consiste di tutti i numeri decimali; l'insieme \mathbb{C} dei numeri complessi; l'insieme delle funzioni razionali (funzioni che sono quoziente di due polinomi).

Per semplicità lavoreremo esclusivamente coi numeri reali e coi numeri complessi, anche se i principali risultati dell'algebra lineare, in particolare tutti quelli di questo capitolo, valgono per un campo qualsiasi. Per non dover trattare separatamente i casi reale e complesso, useremo il simbolo \mathbb{K} per denotare indifferentemente il campo dei numeri reali \mathbb{R} o il campo dei numeri complessi \mathbb{C} . Quindi supporremo che i coefficienti a_i e il termine noto b di un'equazione lineare appartengano a \mathbb{K} , siano cioè numeri reali oppure complessi e cercheremo le soluzioni $[\bar{x}_1, \dots, \bar{x}_n]^T$ tra le n -uple di numeri appartenenti a \mathbb{K} .

Negli esempi quasi sempre i coefficienti delle equazioni saranno numeri interi, per cui troveremo soluzioni che hanno per componenti numeri razionali, perché i numeri razionali formano il più piccolo campo che contiene gli interi: l'algoritmo di soluzione a partire da coefficienti interi produce soluzioni razionali. Ma nel momento in cui avremo bisogno di considerare equazioni di secondo grado sarà necessario avere a disposizione i numeri reali e i numeri complessi. Geometricamente, se ci limitassimo ai punti del piano cartesiano con coordinate razionali, la bisettrice $y = x$ del primo e terzo quadrante non intersecherebbe la circonferenza $x^2 + y^2 = 4$. In altre circostanze, per esempio quando cercheremo gli autovalori di una matrice, sarà importante poter supporre che ogni equazione algebrica abbia soluzione e per questo saranno indispensabili i numeri complessi.

3 ESEMPI INTRODUTTIVI

Prima di trattare la teoria in generale, consideriamo esempi di sistemi di equazioni lineari nelle tre incognite x, y, z . Supporremo di aver fissato una volta per tutte un sistema di riferimento cartesiano per lo spazio e identificheremo un punto P con la terna delle sue coordinate (x, y, z) . Non faremo più distinzione tra un punto $P = (x, y, z)$ e il suo vettore posizione $\overrightarrow{OP} = [x, y, z]^T$. Saremo così liberi di rappresentare una soluzione di un sistema lineare tanto come un punto quanto come un vettore. Cominciamo col discutere un sistema costituito da un'unica equazione lineare:

$$(3.1) \quad a_1x + a_2y + a_3z = b$$

Le soluzioni di (3.1) sono vettori di \mathbb{R}^3 e nel capitolo 1 abbiamo visto che l'insieme di tutte le soluzioni è un piano di \mathbb{R}^3 perpendicolare al vettore $\mathbf{n} = [a_1, a_2, a_3]^T$ (a patto che i coefficienti a_i non siano tutti nulli).



L'equazione $x + 2y + 3z = 6$ ha tra le sue soluzioni: $[6, 0, 0]^T$, $[0, 3, 0]^T$ e $[1, 1, 1]^T$. Ma come facciamo a descrivere esplicitamente tutte le soluzioni dell'equazione? Possiamo ricavare x

in funzione delle altre incognite: $x = -2y - 3z + 6$. Adesso possiamo assegnare a y un valore arbitrario t_1 e a z un valore arbitrario t_2 e ricavare la soluzione

$$(3.2) \quad \begin{cases} x = -2t_1 - 3t_2 + 6 \\ y = t_1 \\ z = t_2 \end{cases}$$

Le equazioni (3.2) sono equazioni parametriche del piano $x + 2y + 3z = 6$: al variare dei parametri t_1 e t_2 in \mathbb{R} le equazioni (3.2) descrivono tutti i punti del piano, o equivalentemente tutte le soluzioni dell'equazione; per esempio la soluzione $[6, 0, 0]^T$ si ottiene scegliendo $t_1 = t_2 = 0$, la soluzione $(0, 3, 0)$ prendendo $t_1 = 1$ e $t_2 = 0$, la soluzione $[1, 1, 1]^T$ prendendo $t_1 = t_2 = 1$. Prendendo $t_1 = 1/2$ e $t_2 = -1/3$ troviamo la soluzione $[6, 1/2, -1/3]^T$.

Possiamo riscrivere la (3.2) in forma vettoriale:

$$(3.3) \quad \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -2t_1 - 3t_2 + 6 \\ t_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} + t_1 \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$$

I vettori che compaiono al membro di destra si ottengono nel modo seguente: $\mathbf{v}_0 = [6, 0, 0]^T$ è la soluzione dell'equazione corrispondente a $t_1 = t_2 = 0$; le componenti di $\mathbf{v}_1 = [-2, 1, 0]^T$ e $\mathbf{v}_2 = [-3, 0, 1]^T$ sono, rispettivamente, i coefficienti di t_1 e t_2 nelle equazioni parametriche del piano.

Si noti che \mathbf{v}_0 è una soluzione particolare dell'equazione data, mentre \mathbf{v}_1 e \mathbf{v}_2 soddisfano l'equazione omogenea associata $x + 2y + 3z = 0$, che ha gli stessi coefficienti ma termine noto uguale a zero. Le equazioni vettoriali (3.3) mostrano che le soluzioni dell'equazione completa $x + 2y + 3z = 6$ si trovano sommando alla soluzione particolare \mathbf{v}_0 le soluzioni dell'equazione omogenea associata $x + 2y + 3z = 0$. Vedremo in seguito (proposizione 5.5) che questo vale per ogni sistema lineare.

Per risolvere l'equazione generale (3.1), procediamo come nell'esempio: se $a_1 \neq 0$, dall'equazione ricaviamo x in funzione di y e z e poi per ogni valore di y e z troviamo una e una sola soluzione del sistema (geometricamente, stiamo usando le variabili y e z come coordinate del piano di equazione (3.1)). Se invece $a_1 = 0$ ma $a_2 \neq 0$, l'equazione non pone alcun vincolo su x e possiamo ricavare y in funzione di z . Questa volta possiamo usare x e z come coordinate del piano. Per esempio, risolviamo $y + 2z = 3$. Poniamo $x = t_1$ e $z = t_2$: allora $y = 3 - 2z = 3 - 2t_2$ e le soluzioni dell'equazione sono

$$(3.4) \quad \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} t_1 \\ 3 - 2t_2 \\ t_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix} + t_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix}$$

Infine, se $a_1 = a_2 = 0$, ma a_3 non è nullo, l'equazione è equivalente all'equazione $z = b$ che ha come soluzioni i vettori della forma

$$\begin{bmatrix} t_1 \\ t_2 \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ b \end{bmatrix} + t_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

al variare di t_1 e t_2 in \mathbb{R} .



Consideriamo ora un esempio di un sistema di due equazioni nelle incognite x , y e z

$$(3.5) \quad \begin{cases} x + 2y + 3z = 6 \\ 2x - y + z = 2 \end{cases}$$

Risoluzione algebrica

L'idea è: ricavare x dalla prima equazione, eliminarla poi dalla seconda equazione, ricavare quindi dalla seconda equazione y in funzione di z , infine usare z come parametro per descrivere tutte le soluzioni. Conviene procedere così: alla seconda equazione sottraiamo la prima equazione moltiplicata per due in modo da eliminare la x dalla seconda equazione:

$$(3.6) \quad \begin{cases} x + 2y + 3z = 6 \\ -5y - 5z = -10 \end{cases}$$

Questo sistema si risolve per *sostituzione all'indietro*: poniamo $z = t$, poi ricaviamo $y = 2 - z$ dalla seconda equazione e, infine, sostituendo nella prima equazione troviamo $x = 6 - 2y - 3z = 2 - t$. Quindi le soluzioni del sistema sono i vettori

$$(3.7) \quad \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 - t \\ 2 - t \\ t \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

Poniamo $\mathbf{v}_0 = [2, 2, 0]^T$ e $\mathbf{v} = [-1, -1, 1]^T$. Il vettore \mathbf{v}_0 è una soluzione particolare del sistema (3.5), mentre \mathbf{v} , come ogni suo multiplo scalare $t\mathbf{v}$, soddisfa il sistema omogeneo associato: $\begin{cases} x + 2y + 3z = 0 \\ 2x - y + z = 0 \end{cases}$. Dalla (3.7) segue che ogni soluzione del sistema completo (3.5) si ottiene sommando alla soluzione particolare \mathbf{v}_0 una soluzione $t\mathbf{v}$ del sistema omogeneo associato.

Interpretazione geometrica

Le due equazioni del sistema sono le equazioni cartesiane di una retta nello spazio: l'insieme delle soluzioni del sistema è la retta \mathbf{r} intersezione dei due piani definiti dalle due equazioni del sistema.

La retta \mathbf{r} è perpendicolare ai vettori normali di entrambi i piani ed è quindi diretta come il prodotto vettore dei vettori normali ai due piani:

$$\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \times \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \\ -5 \end{bmatrix}$$

L'insieme delle soluzioni del sistema omogeneo associato è ancora una retta \mathbf{r}_0 diretta come \mathbf{v} . Siccome l'origine è una soluzione del sistema omogeneo, la retta \mathbf{r}_0 è la retta per l'origine parallela a \mathbf{r} .

Consideriamo ora un esempio di un sistema di tre equazioni nelle incognite x , y e z :

$$(3.8) \quad \begin{cases} x + 2y + 3z = 6 \\ 2x - y + z = 2 \\ 3x + 8y + 10z = 20 \end{cases}$$

Risoluzione algebrica

Procediamo ancora col metodo di eliminazione: l'idea è ancora di eliminare la x dalle ultime due equazioni sottraendo opportuni multipli della prima equazione e, poi, eliminando la y dall'ultima equazione sottraendo un multiplo della seconda equazione. A questo punto l'ultima equazione determina il valore di z e sostituendo all'indietro troviamo i valori di y e quindi di x .

Per eliminare x dalle ultime due equazioni, sottraiamo dalla seconda equazione due volte la prima equazione e dalla terza equazione tre volte la prima equazione:

$$(3.9) \quad \begin{cases} x + 2y + 3z = 6 \\ -5y - 5z = -10 \\ 2y + z = 2 \end{cases}$$

Ora eliminiamo la y dall'ultima equazione, sommando all'ultima equazione i due quinti della seconda:

$$(3.10) \quad \begin{cases} x + 2y + 3z = 6 \\ -5y - 5z = -10 \\ -z = -2 \end{cases}$$

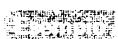
L'ultima equazione implica $z = 2$, sostituendo questo valore di z nella seconda equazione troviamo $y = 0$ e, infine, dalla prima equazione ricaviamo $x = 0$. Il sistema ha quindi un'unica soluzione $(x, y, z) = [0, 0, 2]^T$.

Interpretazione geometrica

L'insieme delle soluzioni è costituito dai punti comuni ai tre piani le cui equazioni compaiono in (3.8). I primi due piani si intersecano nella retta \mathbf{r} dell'esempio precedente, che ha vettore direzione $\mathbf{v} = [-1, -1, 1]^T$. Il terzo piano \mathbf{H} ha vettore normale $\mathbf{n} = [3, 8, 10]^T$. Siccome \mathbf{v} e \mathbf{n} non sono perpendicolari:

$$[-1, -1, 1]^T \cdot [3, 8, 10]^T = -3 - 8 + 10 = -1,$$

la retta \mathbf{r} e il piano \mathbf{H} non sono paralleli e quindi si incontrano in un unico punto $P = [0, 0, 2]^T$. Pertanto i tre piani corrispondenti alle tre equazioni del sistema hanno in comune solo il punto P .



Consideriamo nuovamente un sistema di tre equazioni in tre incognite:

$$(3.11) \quad \begin{cases} x + y - 4z = 1 \\ 2x + 3y - 10z = 2 \\ 5x - 3y - 4z = 5 \end{cases}$$

Risoluzione algebrica

Risolviamo l'uso della stessa strategia dell'esempio precedente: eliminiamo x dalla seconda e dalla terza equazione, sottraendo alla seconda il doppio della prima e alla terza il quintuplo della prima:

$$(3.12) \quad \begin{cases} x + y - 4z = 1 \\ y - 2z = 0 \\ -8y + 16z = 0 \end{cases}$$

seguendo la falsariga dell'esempio precedente, a questo punto è necessario eliminare la y nella terza equazione, in questo caso sommando alla terza equazione la seconda moltiplicata per otto, ci accorgiamo che con questa operazione *tutta la terza equazione* scompare, trasformandosi nell'identità $0 = 0$:

$$(3.13) \quad \begin{cases} x + y - 4z = 1 \\ y - 2z = 0 \\ 0 = 0 \end{cases}$$

Il sistema ha perciò solo due equazioni significative e l'insieme delle soluzioni è una retta nello spazio. Per esprimere le soluzioni, poniamo $z = t$, dalla seconda equazione ricaviamo $y = 2z = 2t$, sostituendo questo risultato nella prima ricaviamo $x = 1 + 4z - y = 1 + 2z = 1 + 2t$, le soluzioni del sistema sono i vettori

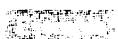
$$(3.14) \quad \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 + 2t \\ 2t \\ t \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$$

Posto $\mathbf{v}_0 = [1, 0, 0]^T$ e $\mathbf{v} = [2, 2, 1]^T$, osserviamo che anche in questo caso \mathbf{v}_0 è una soluzione particolare del sistema di partenza, mentre \mathbf{v} è una soluzione del sistema omogeneo associato:

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

Interpretazione geometrica

L'insieme delle soluzioni è costituito dai punti comuni ai tre piani indicati, che in questo caso si intersecano nella retta di equazioni parametriche (3.14). I tre piani contengono la stessa retta come le pagine aperte di un libro hanno in comune la costola rilegata.



Consideriamo un ultimo sistema di tre equazioni in tre incognite:

$$(3.15) \quad \begin{cases} x - y - z = 3 \\ 3x - 2y - 4z = 3 \\ 4x + y - 9z = 7 \end{cases}$$

Risoluzione algebrica

Adottiamo ancora la stessa strategia degli esempi precedenti: eliminiamo x dalla seconda e dalla terza equazione, sottraendo alla seconda il triplo della prima e alla terza il quadruplo della prima:

$$(3.16) \quad \begin{cases} x - y - z = 3 \\ y - z = -6 \\ 5y - 5z = -5 \end{cases}$$

proseguendo, ora eliminiamo la y dalla terza equazione, sottraendo alla terza equazione il quintuplo della seconda, il primo membro della terza equazione si annulla, il secondo no, si tratta di un'equazione impossibile:

$$(3.17) \quad \begin{cases} x - y - z = 3 \\ y - z = -6 \\ 0 = 25 \end{cases}$$

Questo sistema non ha soluzione.

Interpretazione geometrica

I tre piani, pur non essendo paralleli, non hanno alcun punto comune: i primi due si intersecano lungo una retta, che è parallela al terzo.

4 VETTORI COLONNA

Per *risolvere* i sistemi del paragrafo precedente abbiamo compiuto semplici operazioni sulle equazioni del sistema: abbiamo sommato (o sottratto) due equazioni e abbiamo moltiplicato (o diviso) un'equazione per un numero. Analogamente, per rappresentare le soluzioni di un dato sistema abbiamo impiegato le operazioni di *somma di due vettori* e di *prodotto di un vettore per uno scalare*. La trattazione dei sistemi lineari in generale richiede la naturale estensione di queste operazioni al caso di vettori con un numero arbitrario di componenti.

Denotiamo col simbolo \mathbb{K} il campo degli scalari, che può essere \mathbb{R} o \mathbb{C} e con \mathbb{K}^n l'insieme delle n -uple ordinate di elementi di \mathbb{K} . Diciamo che un elemento $\mathbf{v} \in \mathbb{K}^n$ è un *vettore colonna* con n -componenti e lo rappresentiamo coi simboli

$$\mathbf{v} = [x_1, \dots, x_n]^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Diciamo anche che x_i è l' i -esima componente di \mathbf{v} . Il motivo per cui è utile scrivere gli elementi di \mathbb{K}^n come vettori colonna diverrà chiaro solo quando avremo definito il prodotto di una matrice per un vettore. Questo non toglie che tipograficamente sia più semplice scrivere un vettore su una sola riga e usare la notazione $[x_1, x_2, \dots, x_n]^T$.

In \mathbb{K}^n , analogamente al caso di \mathbb{R}^2 e \mathbb{R}^3 , si introduce una struttura algebrica che consiste di due operazioni: la *somma di vettori*, un'operazione interna che a due vettori associa un vettore somma, e il *prodotto di un vettore per uno scalare*, un'operazione esterna che a un vettore e a uno scalare associa un vettore.

DEFINIZIONE 4.1 La *somma di due vettori* di \mathbb{K}^n è il vettore definito dalla formula:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

In altri termini, la somma di due vettori \mathbf{v} e \mathbf{w} è il vettore $\mathbf{v} + \mathbf{w}$ che ha per i -esima componente la somma delle i -esime componenti di \mathbf{v} e di \mathbf{w} .

Il *prodotto di un vettore di \mathbb{K}^n per uno scalare* $t \in \mathbb{K}$ è definito dalla formula:

$$t \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} tx_1 \\ \vdots \\ tx_n \end{bmatrix}$$

In altri termini, l' i -esima componente di $t\mathbf{v}$ è il prodotto di t per l' i -esima componente di \mathbf{v} .

Vettore nullo

Il vettore che ha tutte le componenti nulle si dice vettore nullo e si indica col simbolo $\mathbf{0}$. Ha la proprietà che $\mathbf{v} + \mathbf{0} = \mathbf{v}$ per ogni vettore \mathbf{v} con lo stesso numero di componenti.

Opposto di un vettore

Dato un vettore \mathbf{v} , l'opposto $-\mathbf{v}$ di \mathbf{v} è il vettore che ha per i -esima componente l' i -esima componente di \mathbf{v} cambiata di segno. Si noti che $-\mathbf{v}$ si ottiene anche come prodotto dello scalare -1 per il vettore \mathbf{v} :

$$(-1)\mathbf{v} = -\mathbf{v}$$

Si definisce quindi la differenza di due vettori:

$$\mathbf{v} - \mathbf{w} = \mathbf{v} + (-\mathbf{w}) = \begin{bmatrix} x_1 - y_1 \\ x_2 - y_2 \\ \vdots \\ x_n - y_n \end{bmatrix}$$

OSSERVAZIONE Si verifica (senza pena ma con un po' di noia) che le operazioni di somma e di prodotto per scalare godono delle proprietà **V1-V8** del capitolo 1. Un insieme \mathbf{V} con operazioni di somma e prodotto per uno scalare che soddisfano **V1-V8** si dice *spazio vettoriale sul campo \mathbb{K}* .

Combinazioni lineari e base canonica

DEFINIZIONE 4.2 (Combinazione lineare)

Si dice che un vettore \mathbf{v} è *combinazione lineare* dei vettori $\mathbf{v}_1, \dots, \mathbf{v}_d$ se esistono degli scalari $t_1, \dots, t_d \in \mathbb{K}$ tali che

$$\mathbf{v} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \dots + t_d\mathbf{v}_d$$

Gli scalari t_1, \dots, t_d si dicono *coefficienti* della combinazione lineare.



Il vettore $\mathbf{v} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$ è combinazione lineare dei vettori $\mathbf{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ e $\mathbf{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ con coefficienti 3 e 5:

$$\begin{bmatrix} 3 \\ 5 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 5 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Evidentemente, come abbiamo visto nel capitolo 1, un vettore di \mathbb{R}^3 si scrive in modo unico come combinazione lineare dei tre versori fondamentali:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

dove

$$\mathbf{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

In \mathbb{K}^n un ruolo analogo ai versori fondamentali è svolto dai vettori che hanno una componente uguale a uno e tutte le altre uguali a zero. Per ogni $i = 1, 2, \dots, n$ denotiamo col simbolo \mathbf{e}_i il vettore che ha l' i -esima componente uguale a 1 e tutte le altre nulle:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

DEFINIZIONE 4.3' (Base canonica di \mathbb{K}^n)

L'insieme ordinato $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ si dice base canonica di \mathbb{K}^n .

Daremo più avanti un significato preciso al termine *base*: è l'analogo algebrico della nozione di sistema di riferimento. Dire che la base è un insieme ordinato significa che i vettori \mathbf{e}_k vanno presi nell'ordine specificato; esattamente come nel piano cartesiano l'ascissa x viene prima dell'ordinata y . Si osservi che

$$x_1\mathbf{e}_1 + \dots + x_n\mathbf{e}_n = \begin{bmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ x_2 \\ \vdots \\ 0 \end{bmatrix} + \dots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Quindi un arbitrario vettore $\mathbf{v} = [x_1, \dots, x_n]^T$ di \mathbb{K}^n si può scrivere in uno e un solo modo come *combinazione lineare* dei vettori della base canonica:

Ogni vettore è combinazione lineare dei vettori della base canonica:

$$(4.1) \quad [x_1, \dots, x_n]^T = x_1\mathbf{e}_1 + \dots + x_n\mathbf{e}_n.$$

I coefficienti della combinazione lineare al membro di destra della (4.1) sono esattamente le componenti del vettore \mathbf{v} : per questo la base si dice canonica.

5 SISTEMI LINEARI E MATRICI

Un sistema lineare di m equazioni in n incognite x_1, \dots, x_n è un insieme di m equazioni lineari in x_1, \dots, x_n

$$(5.1) \quad \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \qquad \vdots \qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

Supponiamo che i coefficienti a_{ij} e i termini noti b_h appartengano al campo \mathbb{K} . Una soluzione del sistema è un vettore $\mathbf{v} = [x_1, \dots, \bar{x}_n]^T$ di \mathbb{K}^n le cui componenti $\bar{x}_1, \dots, \bar{x}_n$ soddisfano tutte le equazioni del sistema.

Per semplificare le notazioni introduciamo la nozione di matrice.

DEFINIZIONE 5.1 Una *matrice* \mathbf{A} di tipo (m, n) (equivelentemente, una matrice $m \times n$) a elementi in \mathbb{K} è una tabella rettangolare di numeri

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad (a_{ij} \in \mathbb{K})$$

disposti in m righe e n colonne: a_{ij} è il numero che si trova sulla i -esima riga e sulla j -esima colonna di \mathbf{A} . Si dice che a_{ij} è l'elemento di posto (i, j) della matrice \mathbf{A} .

Si scrive anche

$$A = \left[a_{ij} \right]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$$

o semplicemente $\mathbf{A} = [a_{ij}]$. A volte si scrive $a_{i,j}$ al posto di a_{ij} ; per esempio, $a_{m-1,n}$ indica l'ultimo elemento della penultima riga di \mathbf{A} . Il simbolo $(\mathbf{A})_{ij}$ denota l'elemento di posto (i, j) della matrice \mathbf{A} .

Una matrice di tipo $(1, 1)$ è semplicemente uno scalare. Una matrice di tipo $(1, n)$ è un vettore riga con n componenti. Una matrice di tipo $(m, 1)$ è un vettore colonna con m componenti. La nostra convenzione di rappresentare gli elementi di \mathbb{K}^m come vettori colonna identifica \mathbb{K}^m con l'insieme delle matrici $m \times 1$.

Una matrice di tipo (m, n) ha m righe, ciascuna delle quali è un vettore con n componenti; e n colonne, ciascuna delle quali è un vettore con m componenti. Spesso scriveremo *riga i* invece dell'orrendo i -esima riga.

è una matrice 2×3 a coefficienti reali. L'elemento di posto $(1, 2)$ è 2, l'elemento di posto $(2, 1)$ è $\sqrt{2}$. La matrice è formata da due vettori riga con 3 componenti:

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} \sqrt{2} & \pi & 4 \end{bmatrix}$$

e da tre vettori colonna con 2 componenti.

Prodotto di una matrice per un vettore colonna

Definiamo ora il prodotto di un vettore riga per un vettore colonna, o anche *prodotto riga per colonna*, che sarà fondamentale nel seguito:

DEFINIZIONE 5.2 (Prodotto riga per colonna)

Dati un vettore riga \mathbf{a} e un vettore colonna \mathbf{v} con lo stesso numero di componenti, il loro prodotto $\mathbf{a} \cdot \mathbf{v}$ è lo scalare definito dalla formula:

$$(5.2) \quad \mathbf{a} \cdot \mathbf{v} = [a_1 \ a_2 \ \cdots \ a_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

Il prodotto si effettua quindi moltiplicando in \mathbb{K} le componenti di ugual indice e poi sommando su tutti gli indici. Utilizzando il simbolo di sommatoria si scrive

$$\mathbf{a} \cdot \mathbf{v} = \sum_{j=1}^n a_j x_j$$

Il prodotto di $\mathbf{a} = [1, -2, 1, 3]$ e $\mathbf{v} = [4, 3, 6, -1]^T$ è $\mathbf{a} \cdot \mathbf{v} = 1 \cdot 4 + (-2) \cdot 3 + 1 \cdot 6 + 3 \cdot (-1) = 1$.

Nel caso $n = 3$, il prodotto di un vettore riga \mathbf{v}^T con un vettore colonna \mathbf{w} coincide col prodotto scalare $\mathbf{v} \cdot \mathbf{w}$:

$$[x_1, y_1, z_1] \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = x_1x_2 + y_1y_2 + z_1z_2.$$

Il prodotto riga per colonna generalizza il prodotto scalare di due vettori di \mathbb{R}^3 .

Definiamo ora il prodotto di una matrice per un vettore colonna: per questo è necessario che il numero di colonne della matrice sia uguale al numero di componenti (ossia di righe) del vettore colonna

DEFINIZIONE 5.3 (Prodotto matrice per vettore)

Il prodotto \mathbf{Av} di una matrice $\mathbf{A} = [a_{ij}]$ di tipo (m, n) per un vettore colonna $\mathbf{v} = [x_1, \dots, x_n]^T$ con n componenti è il vettore colonna con m componenti così definito:

$$(5.3) \quad \mathbf{Av} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{bmatrix} \stackrel{\text{def}}{=} \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}$$

In altri termini, l' i -esima componente di \mathbf{Av} è il prodotto dell' i -esima riga di \mathbf{A} per \mathbf{v} . Il numero di componenti del vettore prodotto \mathbf{Av} è pari al numero di righe di \mathbf{A} .

Perc esempio

$$\begin{bmatrix} 1 & 2 & 3 \\ \sqrt{2} & \pi & 4 \end{bmatrix} \begin{bmatrix} 5 \\ -1 \\ 7 \end{bmatrix} = \begin{bmatrix} 24 \\ 5\sqrt{2} - \pi + 28 \end{bmatrix}$$

Il prodotto \mathbf{Av} si può esprimere anche in termini delle colonne di \mathbf{A} : infatti

$$\begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

Quindi:

Il prodotto \mathbf{Av} come combinazione lineare delle colonne di \mathbf{A}

Il prodotto \mathbf{Av} è la combinazione lineare delle colonne di \mathbf{A} che ha per coefficienti le componenti di \mathbf{v} : se $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n$ sono le colonne di \mathbf{A} e x_1, x_2, \dots, x_n sono le componenti di \mathbf{v} , allora

$$(5.4) \quad \mathbf{Av} = x_1\mathbf{w}_1 + x_2\mathbf{w}_2 + \cdots + x_n\mathbf{w}_n$$

Un altro modo di rappresentare la (5.4) è di pensare alla matrice \mathbf{A} come a un vettore riga le cui componenti sono i vettori colonna \mathbf{w}_i :

$$\mathbf{A} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_n]$$

La formula (5.4) indica, semplicemente, che il prodotto \mathbf{Av} si effettua con la regola del prodotto riga per colonna:

$$\mathbf{Av} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{w}_1 + x_2\mathbf{w}_2 + \cdots + x_n\mathbf{w}_n$$

Applicando la formula (5.4) a $\mathbf{v} = \mathbf{e}_k$, il k -esimo vettore della base canonica, otteniamo:

sia \mathbf{A} una matrice $m \times n$ e sia \mathbf{e}_k il k -esimo vettore della base canonica di \mathbb{K}^n . Allora $\mathbf{A}\mathbf{e}_k$ è la k -esima colonna di \mathbf{A} .

La proprietà più importante del prodotto matrice per vettore è di essere lineare in \mathbf{v} :

PROPOSIZIONE 5.4 Sia \mathbf{A} una matrice $m \times n$. Il prodotto \mathbf{Av} è *lineare* in \mathbf{v} . Questo significa che:

- a) è *additivo*: $\mathbf{A}(\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{Av}_1 + \mathbf{Av}_2$ per ogni \mathbf{v}_1 e \mathbf{v}_2 in \mathbb{K}^n ;
- b) è *omogeneo*: $\mathbf{A}(t\mathbf{v}) = t\mathbf{Av}$ per ogni scalare t in \mathbb{K} e ogni vettore \mathbf{v} in \mathbb{K}^n .

DIMOSTRAZIONE.

Sia $\mathbf{v}_1 = [x_1, x_2, \dots, x_n]^T$, $\mathbf{v}_2 = [y_1, y_2, \dots, y_n]^T$ e siano $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n$ le colonne di \mathbf{A} . Allora per la (5.4)

$$\begin{aligned}\mathbf{A}(\mathbf{v}_1 + \mathbf{v}_2) &= (x_1 + y_1)\mathbf{w}_1 + (x_2 + y_2)\mathbf{w}_2 + \cdots + (x_n + y_n)\mathbf{w}_n = \\ &= (x_1\mathbf{w}_1 + x_2\mathbf{w}_2 + \cdots + x_n\mathbf{w}_n) + (y_1\mathbf{w}_1 + y_2\mathbf{w}_2 + \cdots + y_n\mathbf{w}_n) = \\ &= \mathbf{Av}_1 + \mathbf{Av}_2\end{aligned}$$

La dimostrazione di b) è lasciata per esercizio.

Matrici associate a un sistema lineare

Dato un sistema lineare

$$(5.5) \quad \left\{ \begin{array}{lcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots & \vdots & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m \end{array} \right.$$

la matrice:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

si dice *matrice dei coefficienti del sistema lineare* (5.5). È una matrice con tante righe quante sono le equazioni del sistema e tante colonne quante sono le incognite. Raggruppiamo anche le incognite e i termini noti per formare il vettore \mathbf{x} delle incognite:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

e il vettore dei termini noti:

$$\mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

Utilizzando il prodotto matrice per vettore possiamo riscrivere il sistema (5.5) nella forma compatta

$$(5.6) \quad \mathbf{A}\mathbf{x} = \mathbf{b}$$

La matrice

$$(5.7) \quad [\mathbf{A}|\mathbf{b}]$$

ottenuta aggiungendo ad \mathbf{A} il vettore colonna \mathbf{b} dei termini noti si dice *matrice completa del sistema*. È una matrice con m righe, tante quante le equazioni del sistema, e $n+1$ colonne (numero delle incognite più uno). Si noti che il sistema (5.5) è determinato dalla sua matrice completa: diremo che il sistema è definito dalla matrice $[\mathbf{A}|\mathbf{b}]$.

La matrice completa del sistema

$$\left\{ \begin{array}{l} x_1 + 2x_2 + 3x_3 = 6 \\ x_2 + x_3 = 2 \\ 2x_2 + x_3 = 2 \end{array} \right. \quad \text{è}$$

$$[\mathbf{A}|\mathbf{b}] = \left[\begin{array}{ccc|c} 1 & 2 & 3 & 6 \\ 0 & 1 & 1 & 2 \\ 0 & 2 & 1 & 2 \end{array} \right]$$

Il sistema lineare definito dalla matrice completa $[\mathbf{A}|\mathbf{b}] = \left[\begin{array}{ccc|c} 1 & 2 & 3 & 6 \\ 2 & -1 & 1 & 2 \\ 3 & 8 & 10 & 20 \end{array} \right]$ è

$$\left\{ \begin{array}{l} x_1 + 2x_2 + 3x_3 = 6 \\ 2x_1 - x_2 + x_3 = 2 \\ 3x_1 + 8x_2 + 10x_3 = 20 \end{array} \right.$$

Insieme delle soluzioni e nucleo di una matrice

L'insieme delle soluzioni di un sistema lineare ha una semplice struttura algebrica e geometrica perché il membro di sinistra di $\mathbf{A}\mathbf{x} = \mathbf{b}$ è lineare in \mathbf{x} . Per descrivere tale struttura, diciamo che il sistema omogeneo

$$(5.8) \quad \mathbf{A}\mathbf{x} = \mathbf{0}$$

è il *sistema omogeneo associato* al sistema $\mathbf{A}\mathbf{x} = \mathbf{b}$; a secondo membro $\mathbf{0}$ denota il vettore nullo (con m componenti, tante quante le righe di \mathbf{A} , cioè le equazioni del sistema). Come negli esempi esaminati all'inizio del capitolo, le soluzioni di un

sistema lineare si ottengono sommando a una soluzione particolare le soluzioni del sistema omogeneo associato:

PROPOSIZIONE 5.5 Supponiamo che \mathbf{v}_0 sia una soluzione particolare del sistema lineare $\mathbf{Ax} = \mathbf{b}$. Allora le soluzioni di $\mathbf{Ax} = \mathbf{b}$ si ottengono sommando a \mathbf{v}_0 le soluzioni del sistema omogeneo $\mathbf{Ax} = \mathbf{0}$: sono i vettori della forma

$$\mathbf{v} = \mathbf{v}_0 + \mathbf{v}_h$$

dove \mathbf{v}_h denota una soluzione qualsiasi del sistema omogeneo associato.

DIMOSTRAZIONE. Supponiamo \mathbf{v} sia una soluzione di $\mathbf{Ax} = \mathbf{b}$ e consideriamo $\mathbf{v}_h = \mathbf{v} - \mathbf{v}_0$. Abbiamo

$$\mathbf{Av}_h = \mathbf{Av} - \mathbf{Av}_0 = \mathbf{b} - \mathbf{b} = \mathbf{0}$$

quindi \mathbf{v}_h è una soluzione del sistema omogeneo.

Viceversa, se \mathbf{v}_h è una soluzione del sistema omogeneo e poniamo $\mathbf{v} = \mathbf{v}_0 + \mathbf{v}_h$, allora

$$\mathbf{Av} = \mathbf{Av}_0 + \mathbf{Av}_h = \mathbf{b} + \mathbf{0} = \mathbf{b}$$

quindi \mathbf{v} è una soluzione del sistema completo.

L'insieme delle soluzioni \mathbf{v}_h del sistema omogeneo $\mathbf{Ax} = \mathbf{0}$ dipende solo dalla matrice \mathbf{A} :

DEFINIZIONE 5.6 (Nucleo di una matrice)

Data una matrice \mathbf{A} di tipo $m \times n$, l'insieme delle soluzioni del sistema omogeneo $\mathbf{Ax} = \mathbf{0}$ si dice *nucleo* della matrice. Lo denoteremo col simbolo $\text{Ker}(\mathbf{A})$ (dall'inglese *kernel*, che significa nucleo):

$$\text{Ker}(\mathbf{A}) = \{\mathbf{v} \in \mathbb{K}^n : \mathbf{Av} = \mathbf{0}\}.$$

Sinonimo di nucleo è *spazio nullo* (in inglese *null space*) e un altro simbolo comunemente usato per denotare il nucleo di una matrice è $\mathbf{N}(\mathbf{A})$.

Poiché $\mathbf{A}\mathbf{0} = \mathbf{0}$, il vettore nullo appartiene sempre al nucleo di una matrice. Attenzione il vettore nullo a primo membro ha n componenti, quello a destra m componenti; il vettore del nucleo di \mathbf{A} è quello a primo membro. Per essere precisi, dovremmo scrivere $\mathbf{A}\mathbf{0}_n = \mathbf{0}_m$.

Il nucleo delle matrici

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

contiene solamente il vettore nullo, nel caso di \mathbf{A}_1 è il vettore nullo con 3 componenti, nel caso di \mathbf{A}_2 le componenti sono 4. Il nucleo della matrice

$$\mathbf{A}_3 = \begin{bmatrix} 1 & 5 & 2 \\ 5 & 1 & -2 \\ 3 & 1 & -1 \end{bmatrix}$$

consiste dei vettori della forma $[t, -t, 2t]^T$.

Dati un vettore \mathbf{v}_0 in \mathbb{K}^n e un sottoinsieme \mathbf{W} di \mathbb{K}^n , definiamo

$$\mathbf{v}_0 + \mathbf{W} = \{\mathbf{v} \in \mathbb{K}^n : \mathbf{v} = \mathbf{v}_0 + \mathbf{w}, \mathbf{w} \in \mathbf{W}\}$$

Si dice che $\mathbf{v}_0 + \mathbf{W}$ è l'insieme ottenuto traslando \mathbf{W} del vettore \mathbf{v}_0 . Possiamo riscrivere l'enunciato della proposizione precedente nella forma:

se \mathbf{v}_0 è una soluzione particolare del sistema lineare $\mathbf{A}\mathbf{x} = \mathbf{b}$, l'insieme di tutte le soluzioni del sistema è $\mathbf{v}_0 + \text{Ker}(\mathbf{A})$.

In particolare, se il sistema ammette soluzioni, la soluzione è unica se e solo se il nucleo della matrice contiene unicamente il vettore nullo.



Consideriamo per esempio il sistema con l'unica equazione $x + y = 1$ nelle due variabili x e y . La matrice del sistema è $\mathbf{A} = [1 \ 1]$. Il nucleo di \mathbf{A} è la retta $y = -x$. Le soluzioni di $x + y = 1$ sono i punti della retta $y = -x + 1$, che è parallela alla retta $y = -x$ e si ottiene traslando $y = -x$ del vettore $\mathbf{v}_0 = [1, 0]^T$.

Vedremo nel prossimo paragrafo che, come negli esempi introduttivi, esistono un numero intero $s \geq 0$ e s vettori $\mathbf{v}_1, \dots, \mathbf{v}_s$ nel nucleo $\text{Ker}(\mathbf{A})$ tali che le soluzioni del sistema sono i vettori

$$\mathbf{v} = \mathbf{v}_0 + t_1 \mathbf{v}_1 + \cdots + t_s \mathbf{v}_s$$

al variare dei parametri t_1, \dots, t_s in \mathbb{K} .

Identificando $\mathbf{v}_1, \dots, \mathbf{v}_s$ con i vettori della base canonica di \mathbb{K}^s , si ottiene una corrispondenza biunivoca tra gli elementi del nucleo e \mathbb{K}^s . Quindi il nucleo è una copia di \mathbb{K}^s all'interno di \mathbb{K}^n e l'insieme delle soluzioni si ottiene traslando di \mathbf{v}_0 questa copia di \mathbb{K}^s .

Il fatto fondamentale da osservare è che, anche quando il sistema ammette infinite soluzioni, abbiamo ridotto il problema di determinare tutte le soluzioni al problema di determinare i vettori $\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_s$, che sono in numero finito e possono essere calcolati da un computer. Per esempio in MatLab il comando $\mathbf{A}\backslash\mathbf{b}$ fornisce la soluzione particolare \mathbf{v}_0 e il comando `null(A)` fornisce una base (ortonormale) del nucleo, cioè i vettori $\mathbf{v}_1, \dots, \mathbf{v}_s$. Nel prossimo paragrafo descriviamo l'algoritmo di eliminazione di Gauss per determinare i vettori $\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_s$.

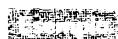
6 METODO DI ELIMINAZIONE DI GAUSS

In questo paragrafo illustriamo il metodo di eliminazione di Gauss, MEG per brevità, per la soluzione di un sistema lineare. L'idea alla base del metodo è la seguente: supponiamo che nella prima equazione del sistema la variabile x_1 abbia coefficiente non nullo allora possiamo *eliminare* da tutte le successive equazioni la variabile x_1 sottraendo un opportuno multiplo scalare della prima equazione (questo equivale a ricavare x_1 dalla prima equazione in funzione di x_2, \dots, x_n e sostituire x_1 con l'espressione ottenuta nelle altre equazioni). Se x_2 compare con coefficiente non nullo,

possiamo iterare il procedimento ed eliminare x_2 dalle equazioni successive alla seconda, fino a ridurci (sempre che i coefficienti delle incognite da eliminare siano sempre non nulli e che il numero di equazioni sia uguale al numero n delle incognite) a un sistema della forma

$$(6.1) \quad \begin{cases} x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = b_1 \\ x_2 + a_{23}x_3 + \cdots + a_{2n}x_n = b_2 \\ \vdots \quad \vdots \\ x_{n-1} + a_{n-1,n}x_n = b_{n-1} \\ x_n = b_n \end{cases}$$

Possiamo ricavare x_n dall'ultima equazione; sostituendo nella penultima equazione troviamo il valore di x_{n-1} e così via procedendo *per sostituzione all'indietro* determiniamo l'unica soluzione del sistema. Consideriamo ora alcuni esempi.



Abbiamo già utilizzato nell'esempio a pagina 74 il metodo di eliminazione per risolvere il sistema

$$(6.2) \quad \begin{cases} x_1 + 2x_2 + 3x_3 = 6 \\ 2x_1 - x_2 + x_3 = 2 \\ 3x_1 + 8x_2 + 10x_3 = 20 \end{cases}$$

Vogliamo ora ripetere il procedimento di eliminazione operando direttamente sulla matrice completa del sistema:

$$[\mathbf{A}|\mathbf{b}] = \left[\begin{array}{ccc|c} 1 & 2 & 3 & 6 \\ 2 & -1 & 1 & 2 \\ 3 & 8 & 10 & 20 \end{array} \right]$$

Per eliminare x_1 dalla seconda equazione abbiamo sottratto alla seconda equazione la prima equazione moltiplicata per due. Questo equivale a modificare la matrice completa del sistema sottraendo dalla seconda *riga* la prima riga moltiplicata per due. Analogamente, per eliminare x_1 dalla terza equazione occorre sottrarre dalla terza riga della matrice la prima riga moltiplicata per tre. Con queste due operazioni riduciamo il sistema originale a un sistema equivalente (stesse soluzioni) avente matrice completa:

$$[\mathbf{A}_1|\mathbf{b}_1] = \left[\begin{array}{ccc|c} 1 & 2 & 3 & 6 \\ 0 & -5 & -5 & -10 \\ 0 & 2 & 1 & 2 \end{array} \right]$$

Adesso sommiamo alla terza riga la seconda riga moltiplicata per $2/5$ in modo da annullare il coefficiente di x_2 nella terza equazione. Otteniamo la matrice completa:

$$[\mathbf{A}_2|\mathbf{b}_2] = \left[\begin{array}{ccc|c} 1 & 2 & 3 & 6 \\ 0 & -5 & -5 & -10 \\ 0 & 0 & -1 & -2 \end{array} \right]$$

Il sistema corrispondente

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 6 \\ -5x_2 - 5x_3 = 10 \\ -x_3 = -2 \end{cases}$$

si risolve immediatamente per *sostituzione all'indietro*: dall'ultima equazione ricaviamo $x_3 = 2$, sostituendo nella seconda troviamo $x_2 = 0$, infine dalla prima equazione otteniamo $x_1 = 0$. Il sistema ammette quindi l'unica soluzione $[0, 0, 2]^T$.

Prendiamo in considerazione un altro esempio, in cui l'algoritmo di soluzione che abbiamo utilizzato finora si interrompe.



Consideriamo il sistema

$$(6.3) \quad \begin{cases} x_1 + x_2 + 2x_3 + x_4 = 4 \\ x_1 + x_2 + 3x_3 + 2x_4 = 6 \\ x_1 + 2x_2 + 3x_3 - x_4 = -1 \\ 2x_1 + 3x_2 + x_3 - x_4 = 4 \end{cases}$$

La matrice completa del sistema è

$$[\mathbf{A}|\mathbf{b}] = \left[\begin{array}{cccc|c} 1 & 1 & 2 & 1 & 4 \\ 1 & 1 & 3 & 2 & 6 \\ 1 & 2 & 3 & -1 & -1 \\ 2 & 3 & 1 & -1 & 4 \end{array} \right]$$

Per eliminare x_1 dalle equazioni successive alla prima, occorre far comparire degli zeri nella prima colonna in tutte le righe sotto la prima. Per far questo basta sottrarre alle righe successive alla prima un opportuno multiplo della prima riga. Otteniamo così:

$$[\mathbf{A}_1|\mathbf{b}_1] = \left[\begin{array}{cccc|c} 1 & 1 & 2 & 1 & 4 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & -2 & -5 \\ 0 & 1 & -3 & -3 & -4 \end{array} \right]$$

Si è verificato un problema: nella seconda riga il coefficiente di x_2 è zero: non possiamo eliminare x_2 dalla terza e dalla quarta equazione sottraendo multipli della seconda equazione. Per ovviare a questo inconveniente basta scambiare tra loro la seconda e la terza riga, in quest'ultima il coefficiente di x_2 è non nullo. Scambiandole otteniamo

$$[\mathbf{A}_2|\mathbf{b}_2] = \left[\begin{array}{cccc|c} 1 & 1 & 2 & 1 & 4 \\ 0 & 1 & 1 & -2 & -5 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 1 & -3 & -3 & -4 \end{array} \right]$$

Ora possiamo riprendere l'algoritmo e ottenere coefficienti nulli sulla seconda colonna nella terza e quarta riga, è sufficiente sottrarre alla quarta riga la “nuova” seconda riga:

$$[\mathbf{A}_3|\mathbf{b}_3] = \left[\begin{array}{cccc|c} 1 & 1 & 2 & 1 & 4 \\ 0 & 1 & 1 & -2 & -5 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & -4 & -1 & 1 \end{array} \right]$$

Infine cancelliamo il coefficiente di x_3 dall'ultima riga sommandole la terza riga moltiplicata per 4:

$$[\mathbf{A}_4|\mathbf{b}_4] = \left[\begin{array}{cccc|c} 1 & 1 & 2 & 1 & 4 \\ 0 & 1 & 1 & -2 & -5 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 3 & 9 \end{array} \right]$$

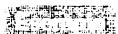
Adesso possiamo risolvere il sistema col metodo di sostituzione all'indietro. Dall'ultima equazione ricaviamo $x_4 = 3$, sostituendo nella terza equazione otteniamo $x_3 = -1$. Ricaviamo x_2 dalla seconda equazione:

$$x_2 = -5 - x_3 + 2x_4 = -5 + 1 + 6 = 2$$

e x_1 dalla prima

$$x_1 = 4 - x_2 - 2x_3 - x_4 = 4 - 2 + 2 - 3 = 1$$

In conclusione, il sistema ammette l'unica soluzione $[1, 2, -1, 3]^T$.



Consideriamo il sistema

$$(6.4) \quad \begin{cases} x_1 + 2x_2 + 3x_3 = 7 \\ x_1 + 2x_2 + 4x_3 = 6 \\ x_1 + 2x_2 + 5x_3 = k \end{cases}$$

in cui k è un parametro reale.

La matrice completa del sistema è

$$[\mathbf{A}|\mathbf{b}] = \left[\begin{array}{ccc|c} 1 & 2 & 3 & 7 \\ 1 & 2 & 4 & 6 \\ 1 & 2 & 5 & k \end{array} \right]$$

Il primo passo dell'algoritmo, con cui otteniamo degli zeri sotto il primo coefficiente della prima colonna, non presenta problemi e riduce la matrice a

$$[\mathbf{A}_2|\mathbf{b}_2] = \left[\begin{array}{ccc|c} 1 & 2 & 3 & 7 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 2 & k-7 \end{array} \right]$$

Ora però tutti i coefficienti sotto il primo nella seconda colonna sono nulli: la variabile x_2 non compare in nessuna equazione successiva alla prima. Questo significa che il sistema non impone alcuna condizione sulla variabile x_2 : se il sistema ammette soluzioni, ogni altro vettore ottenuto da una soluzione particolare cambiando la seconda coordinata è ancora una soluzione. Si dice che x_2 è una variabile libera del sistema. Assegnamole un valore arbitrario $x_2 = t$; infine, eliminiamo x_3 dall'ultima equazione sottraendo dall'ultima riga la seconda riga moltiplicata per due. Otteniamo:

$$[\mathbf{A}|\mathbf{b}] = \left[\begin{array}{ccc|c} 1 & 2 & 3 & 7 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & k-5 \end{array} \right]$$

Ora l'ultima equazione è $0 = k - 5$, che è verificata solo se $k = 5$. Quindi se $k \neq 5$ il sistema non ammette soluzioni; se invece $k = 5$, allora l'ultima riga è identicamente nulla e

corrisponde all'equazione $0 = 0$ che non impone alcuna condizione. Le uniche due equazioni rilevanti sono le prime due. Dalla seconda ricaviamo $x_3 = -1$, dalla prima

$$x_1 = 7 - 2x_2 - 3x_3 = 7 - 2t + 3 = 10 - 2t$$

ricordando che alla variabile x_2 possiamo assegnare un valore arbitrario t . In conclusione, se $k = 0$, il sistema ammette le infinite soluzioni

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 10 - 2t \\ t \\ -1 \end{bmatrix} = \begin{bmatrix} 10 \\ 0 \\ -1 \end{bmatrix} + t \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$$

Il metodo di soluzione di questi esempi si generalizza a un sistema arbitrario e si può schematizzare come segue:

1. le operazioni effettuate sulle righe delle matrici complete lasciano invariato l'insieme delle soluzioni del sistema;
2. tali operazioni vengono effettuate fino a ridurre la matrice completa del sistema a una “matrice scala”, ovvero a una matrice sulla quale non è più possibile compiere eliminazioni;
3. il sistema corrispondente a una matrice a scala si risolve col metodo di sostituzione all'indietro.

Analizziamo ora in dettaglio questi aspetti. Cominciamo con le operazioni che consentono la riduzione.

6.1 Operazioni elementari sulle righe

Innanzitutto per il processo di riduzione sono sufficienti due tipi di operazioni sulle righe della matrice completa del sistema:

- a) sostituire a una riga la sua somma con un multiplo di un'altra riga;
- b) scambiare due righe.

Queste operazioni si dicono *operazioni elementari* sulle righe di una matrice. Le operazioni elementari sono reversibili mediante operazioni elementari:

- i) Supponiamo che una matrice \mathbf{B} sia ottenuta da \mathbf{A} sommando all' i -esima riga la k -esima riga moltiplicata per t , con $i \neq k$. Allora \mathbf{A} si ottiene da \mathbf{B} sommando all' i -esima riga la k -esima riga moltiplicata per $-t$.
- ii) Se la matrice \mathbf{B} si ottiene scambiando l' i -esima e la k -esima riga di \mathbf{A} , scambiando l' i -esima e la k -esima riga di \mathbf{B} si ottiene nuovamente \mathbf{A} .

Possiamo ora mostrare che le operazioni elementari di riga sulla matrice completa di un sistema lasciano invariato l'insieme delle soluzioni. Due sistemi lineari si dicono *equivalenti* se hanno le stesse soluzioni.

PROPOSIZIONE 6.1 Sia $[\mathbf{A}|\mathbf{b}]$ la matrice completa del sistema lineare $\mathbf{Ax} = \mathbf{b}$. Supponiamo che $[\mathbf{A}'|\mathbf{b}']$ sia una matrice ottenuta da $[\mathbf{A}|\mathbf{b}]$ mediante operazioni elementari sulle righe. Allora il sistema lineare $\mathbf{Ax} = \mathbf{b}$ è equivalente al sistema $\mathbf{A}'\mathbf{x} = \mathbf{b}'$.

DIMOSTRAZIONE. Osserviamo che se \mathbf{v} soddisfa due equazioni lineari $\mathbf{a}_1\mathbf{v} = b_1$ e $\mathbf{a}_2\mathbf{v} = b_2$ (qui \mathbf{a}_1 e \mathbf{a}_2 denotano i vettori riga dei coefficienti delle due equazioni), allora

$$(\mathbf{a}_1 + t\mathbf{a}_2)\mathbf{v} = \mathbf{a}_1\mathbf{v} + t\mathbf{a}_2\mathbf{v} = b_1 + tb_2$$

Quindi, se \mathbf{v} soddisfa le due equazioni, allora soddisfa anche la somma della prima con un multiplo della seconda. Ne segue che se \mathbf{v} è una soluzione di $\mathbf{Ax} = \mathbf{b}$, allora è anche soluzione di $\mathbf{A}'\mathbf{x} = \mathbf{b}'$.

Siccome le operazioni elementari sono reversibili mediante operazioni elementari, anche $[\mathbf{A}'|\mathbf{b}']$ è ottenuta da $[\mathbf{A}|\mathbf{b}]$ mediante operazioni elementari. Possiamo quindi scambiare i ruoli delle matrici e concludere che, se \mathbf{v} è una soluzione di $\mathbf{A}'\mathbf{x} = \mathbf{b}'$, allora è anche soluzione di $\mathbf{Ax} = \mathbf{b}$. Quindi i due sistemi hanno le stesse soluzioni.

6.2 L'algoritmo di Gauss

Per risolvere un sistema lineare eliminiamo le incognite mediante le operazioni elementari sulle righe finché è possibile. Ci fermiamo quando non è possibile effettuare ulteriori eliminazioni. Questo succede quando la matrice del sistema assume la forma a scala che ora descriviamo.

DEFINIZIONE 6.2 (Matrice a scala)

Sia \mathbf{U} una matrice e siano $\mathbf{U}_1, \dots, \mathbf{U}_m$ le sue righe. Se una riga \mathbf{U}_i non è nulla, il *pivot* di \mathbf{U}_i è il primo elemento non nullo sulla riga. La matrice \mathbf{U} si dice *matrice a scala* se le seguenti condizioni sono soddisfatte:

- i) se due righe consecutive \mathbf{U}_i e \mathbf{U}_{i+1} sono non nulle, allora il pivot di \mathbf{U}_{i+1} si trova a destra del pivot di \mathbf{U}_i . Quindi, le righe \mathbf{U}_i e \mathbf{U}_{i+1} hanno l'aspetto

$$\begin{bmatrix} 0 & \cdots & 0 & p_i & * & * & * & * \\ 0 & \cdots & 0 & 0 & \cdots & 0 & p_{i+1} & * \end{bmatrix}$$

dove p_i e p_{i+1} sono i pivots di \mathbf{U}_i e \mathbf{U}_{i+1} , rispettivamente.

- ii) se $\mathbf{U}_i = \mathbf{0}$, anche $\mathbf{U}_{i+1} = \mathbf{0}$: le eventuali righe nulle della matrice si trovano tutte in fondo alla matrice.

Possiamo riformulare la definizione di matrice a scala in modo analitico, in termini degli indici di riga e colonna. Lo svantaggio è che la definizione diventa meno intuitiva, il vantaggio è di tipo operativo. In ogni caso, la condizione analitica è questa: una matrice $[u_{ij}]$ di tipo (m, n) è a scala se e solo se esistono un intero $r \leq m$ e una successione crescente $1 \leq j(1) < j(2) < \cdots < j(r) \leq n$ di indici colonna

tali che:

- a) $u_{i,j(i)} \neq 0$ per ogni $i = 1, \dots, r$;
- b) $u_{i,j} = 0$ se $j < j(i)$ per ogni $i = 1, \dots, r$;
- c) $u_{i,j} = 0$ se $i > r$.

La matrice ha allora r pivots e l'indice $j(i)$ è l'indice di colonna del pivot p_i dell' i -esima riga. In effetti, le condizioni a) e c) dicono che le righe non nulle della matrice sono le prime r e le condizioni a) e b) dicono che $p_i = u_{i,j(i)}$ è il pivot dell' i -esima riga. Il fatto che $j(i) < j(i+1)$ dice che l'indice di colonna dei pivots p_i cresce col crescere dell'indice di riga, quindi nello scendere lungo le righe i pivots si spostano verso destra e la matrice è a scala.

Dal fatto che $j(1) < j(2) < \dots < j(r)$ segue che $i \leq j(i)$ per ogni $i = 1, \dots, r$: *l'indice di riga di un pivot di una matrice a scala è minore o uguale del suo indice di colonna.* Inoltre

$$u_{k,j(i)} = 0 \quad \text{se } k > i.$$

Visualmente questo significa che gli elementi di una matrice a scala che si trovano sotto a un pivot sono nulli: questo perché con lo scendere delle righe i pivots scorrono verso destra. Diamo anche una dimostrazione analitica: $u_{k,j(i)} = 0$ se $k > r$ per la proprietà c); se $k \leq r$, allora:

$$i < k \quad \Rightarrow \quad j(i) < j(k) \quad \Rightarrow \quad u_{k,j(i)} = 0 \quad \text{per la proprietà b).}$$



Le seguenti matrici sono a scala:

$$\mathbf{U}_0 = \begin{bmatrix} 3 & 3 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{U}_1 = \begin{bmatrix} 1 & 2 & 5 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{U}_2 = \begin{bmatrix} 3 & 3 & 1 & 2 \\ 0 & 2 & 0 & 3 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

La matrice \mathbf{U}_0 ha $r = 3$ e gli indici di colonna dei pivots sono $j(1) = 1$, $j(2) = 2$, $j(3) = 3$. La matrice \mathbf{U}_1 ha $r = 2$ e gli indici di colonna dei pivots sono $j(1) = 1$, $j(2) = 3$. La matrice \mathbf{U}_2 ha $r = 3$ e gli indici di colonna dei pivots sono $j(1) = 1$, $j(2) = 2$, $j(3) = 3$. Le seguenti matrici non sono a scala:

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 2 & 5 & 3 \\ 0 & 0 & 1 & 3 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 3 & 3 & 1 & 2 \\ 0 & 2 & 0 & 3 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

perché in \mathbf{A}_1 abbiamo una riga nulla all'inizio della matrice e in \mathbf{A}_2 i pivots della seconda e terza riga si trovano sulla stessa colonna.

Il caso più importante è quello delle matrici quadrate, che corrispondono a sistemi in cui il numero di equazioni è uguale al numero di incognite, per i quali ci si aspetta un'unica soluzione. In questo caso una matrice a scala è *triangolare alta*:

DEFINIZIONE 6.3 (Matrice quadrata. Matrice triangolare)

Una matrice \mathbf{A} si dice *quadrata* se il numero di righe di \mathbf{A} è uguale al numero di colonne di \mathbf{A} . Una *matrice quadrata di ordine n* è una matrice $n \times n$.

La *diagonale principale* di una matrice quadrata $[a_{ij}]$ è l'insieme degli elementi a_{ii} i cui indici di riga e colonna sono uguali. Una matrice quadrata $[a_{ij}]$ si dice *triangolare alta* se tutti i suoi elementi sotto la diagonale principale sono nulli:

$$a_{ij} = 0 \quad \text{se } i > j$$

L'uso della lettera \mathbf{U} per le matrici triangolari alte è dovuto al fatto che in inglese triangolare alta si dice *upper triangular*. Nell'esempio precedente, la matrice \mathbf{U}_0 è triangolare alta, mentre \mathbf{U}_1 e \mathbf{U}_2 non sono triangolari alte perché non sono nemmeno quadrate. Le matrici a scala quadrate sono sempre triangolari alte:

PROPOSIZIONE 6.4 Sia \mathbf{U} una matrice quadrata. Se \mathbf{U} è a scala, allora \mathbf{U} è triangolare alta.

DIMOSTRAZIONE. Visualmente l'enunciato è evidente. Per dimostrarlo analiticamente, occorre mostrare $u_{ij} = 0$ se $i > j$. Se l' i -esima riga è nulla, questo è ovvio. Altrimenti si ricordi che in una matrice a scala il pivot $p_i = u_{i,j(i)}$ ha indice di riga minore o uguale dell'indice di colonna: $i \leq j(i)$. Per definizione di pivot $u_{ij} = 0$ se $j < j(i)$ e questo è certamente il caso se $j < i$ perché $i \leq j(i)$.

Vale un parziale viceversa:

Sia \mathbf{U} una matrice triangolare alta. Se tutti gli elementi $p_i = a_{ii}$ sulla diagonale principale di \mathbf{U} sono diversi da zero, allora \mathbf{U} è una matrice a scala:

$$\mathbf{U} = \begin{bmatrix} p_1 & * & * & * & * & * & * \\ 0 & p_2 & * & * & * & * & * \\ 0 & 0 & p_3 & * & * & * & * \\ \cdot & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \dots & p_n & & \end{bmatrix}$$

(* denota un valore qualsiasi).

Se però una matrice triangolare alta ha degli zeri sulla diagonale principale, allora non è detto che la matrice sia a scala. Per esempio:

$$\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

sono triangolari, ma non sono a scala, la prima perché ha due pivots sulla stessa colonna, la seconda perché una riga nulla precede una riga non nulla.

Come annunciato il metodo di eliminazione di Gauss (MEG) riduce una qualsiasi matrice alla forma a scala:

TEOREMA 6.5 (MEG: riduzione a scala di una matrice)

Esiste un algoritmo, detto algoritmo di eliminazione di Gauss, che riduce una matrice qualsiasi a una matrice a scala mediante un numero finito di operazioni elementari sulle righe.

DIMOSTRAZIONE. L'algoritmo procede per colonna, da sinistra verso destra, cancellando tutti gli elementi al di sotto del pivot. Precisamente: sia \mathbf{A} una matrice qualsiasi con m righe e n colonne. Se \mathbf{A} è a scala, stop. Altrimenti supponiamo che la prima colonna non nulla abbia indice $j(1)$ (di solito, la prima colonna è non nulla, quindi $j(1) = 1$); sia $p_1 = a_{l,j(1)}$ il primo coefficiente non nullo della colonna. L'elemento $a_{l,j(1)}$ è il primo pivot di \mathbf{A} , nel senso che tra i pivots che si trovano sulla prima colonna non nulla, è quello col minimo indice di riga.

Scambiamo la prima riga con l' l -esima riga, in modo da portare il primo pivot sulla prima riga: nella nuova matrice $p_1 = a_{1,j(1)}$ si trova sulla prima riga. Annulliamo ora tutti gli elementi $a_{i,j(1)}$ con $i > 1$, cioè tutti gli elementi sotto a p_1 . Per far questo a ogni riga \mathbf{A}_i con $i > 1$ sottraiamo la prima riga moltiplicata per

$$m_{i,1} = \frac{a_{i,j(1)}}{a_{1,j(1)}}$$

I numeri $m_{i,1}$ si dicono moltiplicatori. Effettuata questa operazione, a sinistra di p_1 tutte le colonne sono nulle e sotto p_1 tutti i coefficienti sono zero. Vediamo un esempio in cui $j(1) = 2$, $l = 2$ e $p_1 = 1$.

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 2 & 5 & 3 \\ 0 & 3 & 2 & 1 & 3 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 1 & 2 & 5 & 3 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 3 & 2 & 1 & 3 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 1 & 2 & 5 & 3 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & -4 & -14 & -6 \end{bmatrix}$$

Consideriamo la matrice \mathbf{A}' ottenuta dalla nuova matrice \mathbf{A} cancellando la prima riga. Si noti che \mathbf{A} è a scala se e solo se \mathbf{A}' è a scala.

Possiamo ora iterare l'algoritmo iniziando con la matrice \mathbf{A}' . Siccome \mathbf{A}' ha una riga in meno di \mathbf{A} , l'algoritmo termina in un numero finito di passi con una matrice a scala.

L'algoritmo di Gauss ammette variazioni, ma qui ci interessa, per le applicazioni teoriche, averlo definito in modo univoco. Si possono infatti fare scelte diverse quando si scambiano le righe, processo noto come *pivoting*; per esempio, per ridurre gli errori di arrotondamento, di solito si sceglie su una colonna l'elemento più grande in valore assoluto, anziché il primo elemento non nullo.

Nel caso di una generica matrice quadrata, gli elementi a_{ii} sulla diagonale principale rimangono non nulli a ogni passo, e l'eliminazione si svolge senza bisogno di scambi di righe. Val la pena esplicitare l'algoritmo in questo caso e contare il numero di operazioni che comporta. Per ipotesi, il primo pivot è a_{11} . L'algoritmo cancella tutti gli elementi sulla prima colonna al di sotto di a_{11} : questo avviene sottraendo all' i -esima riga, $i \geq 2$, la prima moltiplicata per il moltiplicatore

$$m_{i1} = \frac{a_{i1}}{a_{11}}.$$

Si compiono così $n-1$ operazioni elementari, tante quante sono le righe successive alla prima. A questo punto la prima colonna è sistemata e si ripete il procedimento con la seconda colonna: per cancellare gli elementi sotto il pivot a_{22} si sottrae all' i -esima riga la seconda moltiplicata per m_{i2} , per ogni $i \geq 3$. Si tratta di $n-2$ operazioni elementari. Così per sistemare la terza colonna occorrono $n-3$ operazioni elementari e così via fino alla penultima colonna, che viene sistemata con un'unica operazione elementare perché c'è solo l'elemento di posto $(n, n-1)$ da cancellare. L'algoritmo senza scambi di righe consiste quindi di

$$\sum_{k=1}^{n-1} k = \frac{1}{2}n(n-1)$$

operazioni elementari sulle righe, ogni operazione consiste nel sottrarre alla i -esima riga la j -esima riga moltiplicata per il moltiplicatore m_{ij} , con $1 \leq j < i \leq n$.

Se si vuole valutare il costo computazionale dell'eliminazione, per esempio nel caso quadrato $m = n$ e sotto l'ipotesi che non ci siano scambi di riga, occorre tener conto che il numero di operazioni numeriche diminuisce col procedere dell'algoritmo: mentre il primo passaggio, per sistemare la prima colonna, consiste di operazioni elementari su vettori riga con n componenti, nel secondo passaggio, quello per sistemare la seconda colonna, le operazioni si svolgono su righe che hanno in realtà $n-1$ componenti, perché l'elemento sulla prima colonna è nullo e così via. Alla fine si può mostrare che il numero di flops necessari per l'eliminazione è $2n^3/3$ a meno di termini di ordine inferiore. Questo numero andrebbe confrontato col numero, enormemente più grande perché dell'ordine di $(n+1)!$, di flops necessari per la risoluzione con la formula di Cramer (si veda il capitolo sul determinante). In pratica, anche un calcolatore molto potente impiegherebbe anni a risolvere un sistema 25×25 con la regola di Cramer, mentre, in meno di un secondo, risolverebbe il sistema con l'eliminazione di Gauss.

Negli esercizi fatti a mano è spesso utile introdurre un'ulteriore operazione di riga: moltiplicare una riga per uno scalare non nullo. Questo evidentemente non modifica le soluzioni del sistema associato. Per esempio nel ridurre la matrice

$$\begin{bmatrix} 3 & 1 & 2 \\ 4 & 2 & 3 \end{bmatrix}$$

il MEG richiede di sottrarre alla seconda riga la prima moltiplicata per $\frac{4}{3}$. Per evitare di introdurre frazioni, possiamo invece dapprima moltiplicare la seconda riga per 3 e poi sottrarre 4 volte la prima riga, ottenendo la matrice a scala

$$\begin{bmatrix} 3 & 1 & 2 \\ 0 & 2 & 1 \end{bmatrix}$$

Nelle applicazioni, il principale limite del MEG è che, se si parte da una matrice con molti zeri (come spesso succede nelle applicazioni che portano a considerare sistemi con un numero enorme di variabili), l'algoritmo di Gauss tende a ridurre il numero di zeri presenti, rendendo pesanti i conti. Esistono per questo altri algoritmi di soluzione.

6.3 Rango di una matrice. Teoremi di Cramer e Rouché-Capelli

Per la soluzione di un sistema lineare, rimane da descrivere il metodo di sostituzione all'indietro quando la matrice dei coefficienti sia a scala. Il numero r dei pivots della matrice ha un ruolo fondamentale e perciò merita un nome.

DEFINIZIONE 6.6 (Rango di una matrice)

Il *rango* $r(\mathbf{U})$ di una matrice a scala \mathbf{U} è il numero di righe non nulle di \mathbf{U} (quindi il numero dei pivots della matrice). Se \mathbf{A} è una matrice qualsiasi, il rango di \mathbf{A} è il rango della matrice a scala \mathbf{U} ottenuta a partire da \mathbf{A} con l'algoritmo di riduzione a scala di Gauss del Teorema 6.5.

Vedremo nel capitolo sugli spazi vettoriali che il rango di una matrice non dipende dall'algoritmo di Gauss. Questo significa che, se \mathbf{A}' e \mathbf{A}'' sono due matrici a scala ottenute da \mathbf{A} mediante operazioni elementari sulle righe, il numero di pivots di \mathbf{A}' coincide col numero di pivots di \mathbf{A}'' . Si pensi per esempio di ottenere \mathbf{A}' da \mathbf{A} con l'algoritmo di Gauss; invece per ottenere \mathbf{A}'' potremmo scambiare la prima e l'ultima riga di \mathbf{A} e poi applicare l'algoritmo di Gauss. In generale, \mathbf{A}' e \mathbf{A}'' saranno distinte: il lettore dovrebbe calcolarsi un esempio esplicito. Ma il numero di pivots di \mathbf{A}' e \mathbf{A}'' è lo stesso. Il motivo, come vedremo e preciseremo in seguito, è che il rango di una matrice è il massimo numero di righe *linearmente indipendenti* della matrice \mathbf{A} ; tale numero è lasciato invariato dalle operazioni elementari e coincide col numero di righe non nulle nel caso delle matrici a scala. Dal punto di vista dei sistemi lineari, l'algoritmo di climinazione seleziona le equazioni indipendenti del sistema ed elimina tutte quelle che non sono necessarie in quanto conseguenza delle altre.

OSSERVAZIONE Se \mathbf{A} è una matrice di tipo (m, n) ,

$$r(\mathbf{A}) \leq m \quad \text{e} \quad r(\mathbf{A}) \leq n :$$

il *rango di una matrice* è minore o uguale del numero delle righe e del numero di colonne. Per dimostrare queste diseguaglianze possiamo supporre che la matrice sia a scala, perché l'algoritmo di riduzione a scala lascia invariati i numeri r, m, n . Per una matrice a scala \mathbf{U} , r è il numero di righe non nulle, che è minore o uguale al numero m di tutte le righe di \mathbf{U} . Inoltre r è anche minore o uguale al numero n delle colonne di \mathbf{U} perché r è anche il numero dei pivots di \mathbf{U} , che si trovano su colonne distinte.

Calcoliamo il rango della matrice

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

Per questo riduciamo \mathbf{A} a scala:

$$\begin{bmatrix} 2 & 2 & 0 & 2 \\ 3 & -1 & 3 & 0 \\ -4 & 0 & -3 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 2 & 0 & 2 \\ 0 & -4 & 3 & -3 \\ 0 & 4 & -3 & 3 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 2 & 0 & 2 \\ 0 & -4 & 3 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Il rango della matrice è perciò due.

Teorema di Cramer

Cominciamo dal caso più semplice e più importante dei sistemi “quadrati”, quelli per cui il numero di equazioni è uguale al numero di incognite. In questo caso ci si aspetta un’unica soluzione, perché ogni equazione dovrebbe consentire di eliminare un’incognita. Questo succede a patto che al termine del processo di riduzione a scala tutte le equazioni “siano sopravvissute”, il che equivale a dire che la matrice ridotta abbia tutte le righe non nulle, o ancora che il rango della matrice sia uguale al numero delle righe (il massimo possibile).

TEOREMA 6.7 (Teorema di Cramer)

Sia $\mathbf{Ax} = \mathbf{b}$ un sistema lineare quadrato di n equazioni in n incognite. Se $r(\mathbf{A}) = n$, il sistema ammette una e una soluzione, qualunque sia $\mathbf{b} \in \mathbb{K}^n$.

DIMOSTRAZIONE. Sia \mathbf{U} la matrice ottenuta riducendo \mathbf{A} a scala. Il sistema $\mathbf{Ax} = \mathbf{b}$ è equivalente a un sistema della forma $\mathbf{Ux} = \mathbf{b}'$, perciò basta mostrare che quest’ultimo ha un’unica soluzione.

Siccome $r(\mathbf{U}) = r(\mathbf{A}) = n$, gli indici di colonna dei pivots devono essere $j(1) = 1$, $j(2) = 2, \dots, j(n) = n$: la matrice \mathbf{U} è triangolare alta con tutti i pivots $p_i = u_{ii}$ sulla diagonale principale.

$$\mathbf{U} = \begin{bmatrix} p_1 & * & * & \cdots & * \\ 0 & p_2 & * & \cdots & * \\ 0 & 0 & p_3 & \cdots & * \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & p_n \end{bmatrix}$$

Dall’ultima equazione del sistema $\mathbf{Ux} = \mathbf{b}'$ ricaviamo

$$x_n = \frac{b'_n}{p_n}$$

Sostituendo questa espressione di x_n nella penultima equazione

$$p_{n-1}x_{n-1} + u_{n-1,n}x_n = b'_{n-1}$$

possiamo ricavare x_{n-1} perché $p_{n-1} \neq 0$. Ora procedendo in maniera analoga all’indietro ricaviamo i valori di $x_{n-2}, x_{n-3}, \dots, x_1$. Quindi il sistema ammette un’unica soluzione. 



Risolviamo il sistema

$$(6.5) \quad \begin{cases} x_1 - 4x_2 + x_3 = 2 \\ 2x_1 - 6x_2 + 5x_3 = 8 \\ x_1 - x_2 + 5x_3 = 8 \end{cases}$$

Scriviamo la matrice completa del sistema:

$$[\mathbf{A}|\mathbf{b}] = \left[\begin{array}{ccc|c} 1 & -4 & 1 & 2 \\ 2 & -6 & 5 & 8 \\ 1 & -2 & 5 & 8 \end{array} \right]$$

Riduciamo ora la matrice a scala:

$$\left[\begin{array}{ccc|c} 1 & -4 & 1 & 2 \\ 2 & -6 & 5 & 8 \\ 1 & -2 & 5 & 8 \end{array} \right] \rightarrow \left[\begin{array}{ccc|c} 1 & -4 & 1 & 2 \\ 0 & 2 & 3 & 4 \\ 0 & 2 & 4 & 6 \end{array} \right] \rightarrow \left[\begin{array}{ccc|c} 1 & -4 & 1 & 2 \\ 0 & 2 & 3 & 4 \\ 0 & 0 & 1 & 2 \end{array} \right]$$

La matrice dei coefficienti del sistema ha perciò rango tre e il sistema ha un'unica soluzione. Dall'ultima riga deduciamo $x_3 = 2$. Sostituendo nell'equazione corrispondente alla seconda riga troviamo

$$2x_2 + 6 = 4$$

quindi, $x_2 = -1$. Sostituendo nell'equazione corrispondente alla prima riga troviamo infine

$$x_1 + 4 + 2 = 2$$

quindi, $x_1 = -4$. La soluzione del sistema è il vettore

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -4 \\ -1 \\ 2 \end{bmatrix}$$



In questo esempio risolviamo esplicitamente i sistemi 2×2 con matrice dei coefficienti di rango 2:

Formula di Cramer per i sistemi 2×2

Una matrice $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ ha rango due se e solo se $\det(\mathbf{A}) = ad - bc \neq 0$. In tal caso la soluzione del sistema $\begin{cases} ax + by = e \\ cx + dy = f \end{cases}$ è

$$(6.6) \quad \begin{cases} x = \frac{ed - bf}{ad - bc} = \frac{\det(\mathbf{A}_1)}{\det(\mathbf{A})} \\ y = \frac{af - ec}{ad - bc} = \frac{\det(\mathbf{A}_2)}{\det(\mathbf{A})} \end{cases}$$

Nella formula \mathbf{A}_1 (rispettivamente \mathbf{A}_2) denota la matrice ottenuta sostituendo la prima (rispettivamente la seconda) colonna di \mathbf{A} con il vettore dei termini noti.

Per provare l'enunciato riduciamo la matrice a scala. Supponiamo dapprima $a \neq 0$. La riduzione a scala si ottiene in un solo passo

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \rightarrow \begin{bmatrix} a & b \\ 0 & d - \frac{c}{a}b \end{bmatrix}$$

Il rango è due se solo se la seconda riga della matrice ridotta è diversa da zero, cioè $ad - bc \neq 0$. Se $a = 0$, ma $c \neq 0$, otteniamo lo stesso risultato scambiando la prima e la seconda riga.

Supponiamo infine $a = c = 0$. In questo caso $ad - bc = 0$. D'altra parte la prima colonna della matrice è nulla, quindi i pivots della matrice ridotta a scala si trovano sulle colonne

successive alla prima. Ma ci sono solo due colonne, quindi la matrice ridotta ha al massimo un pivot non nullo, cioè $\text{rango} \leq 1$. Abbiamo così mostrato che il rango è due se e solo se $ad - bc \neq 0$.

La soluzione del sistema si trova col metodo di eliminazione: se $a \neq 0$, l'algoritmo di Gauss riduce la matrice completa del sistema a

$$\left[\begin{array}{cc|c} a & b & e \\ 0 & d - \frac{c}{a}b & f - \frac{c}{a}e \end{array} \right]$$

L'equazione corrispondente all'ultima riga consente di ricavare

$$y = \frac{f - \frac{c}{a}e}{d - \frac{c}{a}b} = \frac{af - ec}{ad - bc}$$

Sostituendo nella prima equazione otteniamo

$$x = \frac{e - by}{a} = \frac{ed - bf}{ad - bc}$$

Nel capitolo sul determinante definiremo il determinante di una matrice quadrata di ordine n qualsiasi; esiste un'analogia formula di Cramer per la soluzione dei sistemi quadrati di ordine maggiore a 2, ma è di scarso interesse pratico a causa della complessità del calcolo dei determinanti.

Teorema di Rouché-Capelli. Struttura dell'insieme delle soluzioni di un sistema lineare

Trattiamo ora il caso di un sistema arbitrario: in generale, il numero di equazioni può essere diverso dal numero di incognite e anche nel caso dei sistemi quadrati è possibile che le equazioni non siano tutte indipendenti, cioè che la matrice dei coefficienti non abbia rango massimo. In ogni caso, i numeri significativi per determinare da quanti parametri dipendano le soluzioni del sistema sono il numero delle incognite e i ranghi della matrice dei coefficienti e della matrice completa del sistema:

TEOREMA 6.8 (Rouché-Capelli)

Sia $\mathbf{Ax} = \mathbf{b}$ un sistema lineare in n incognite a coefficienti in \mathbb{K} . Allora

- a) il sistema ammette soluzioni se e solo se il rango di \mathbf{A} è uguale al rango della matrice completa $[\mathbf{A}|\mathbf{b}]$;
- b) supponiamo che il sistema ammetta soluzioni. Allora l'insieme delle soluzioni dipende da $n-r$ parametri: esistono $\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_{n-r} \in \mathbb{K}^n$ tali che le soluzioni del sistema sono i vettori della forma

$$(6.7) \quad \mathbf{v} = \mathbf{v}_0 + t_1 \mathbf{v}_1 + \cdots + t_{n-r} \mathbf{v}_{n-r}$$

al variare di $t_1, \dots, t_{n-r} \in \mathbb{K}$. A valori distinti dei parametri t_1, \dots, t_{n-r} corrispondono soluzioni distinte;

Possiamo generalizzare questo procedimento al caso di una matrice a scala \mathbf{U} qualsiasi. Come nell'esempio, dividiamo le variabili in due categorie:

- chiamiamo *dipendenti* le r variabili corrispondenti alle r colonne contenenti i pivots;
- chiamiamo *libere* le rimanenti $n - r$ variabili.

Ora risolviamo il sistema per “sostituzione all’indietro”. Per iniziare, consideriamo l’ultima equazione significativa, quella che contiene il pivot $p_r = u_{rk}$. Essa ha la forma

$$p_r x_k + c_{k+1} x_{k+1} + \cdots + c_n x_n = b'_r$$

dove x_k è l’ultima variabile dipendente, così che x_{k+1}, \dots, x_n sono tutte libere.

Siccome $p_r \neq 0$, possiamo ricavare x_k in funzione di b'_r e delle variabili libere x_{k+1}, \dots, x_n .

$$x_k = \frac{b'_r}{p_r} - \frac{c_{k+1}}{p_r} x_{k+1} - \cdots - \frac{c_n}{p_r} x_n$$

Proseguiamo ora con la penultima equazione significativa, quella corrispondente al pivot p_{r-1} . Procedendo come sopra ricaviamo la penultima variabile dipendente, che è x_h , dove h è la colonna su cui giace p_{r-1} (quindi $h < k$):

$$x_h = \frac{b'_{r-1}}{p_{r-1}} - \frac{d_{h+1}}{p_{r-1}} x_{h+1} - \cdots - \frac{d_n}{p_{r-1}} x_n$$

Le variabili nel membro di destra dell’equazione sono tutte libere tranne x_k , che però abbiamo già ricavato in funzione di b'_r e delle variabili libere. Sostituendo x_k possiamo riscrivere x_h in funzione dei termini noti b'_{r-1}, b'_r e delle variabili libere.

Si continua in modo simile risolvendo le equazioni dal basso verso l’alto: dall’equazione corrispondente a un pivot p_j si ricava la corrispondente variabile dipendente in funzione del termine noto b'_j , delle variabili libere e delle variabili dipendenti che sono state già ricavate dalle equazioni sotto la j -esima.

Alla fine del procedimento ogni variabile dipendente viene determinata in funzione delle variabili libere e dei termini noti: assegnando valori arbitrari t_1, t_2, \dots, t_{n-r} alle variabili libere, per le variabili dipendenti x_i troviamo un’espressione della forma

$$x_i = c_i + a_{i,1}t_1 + a_{i,2}t_2 + \cdots + a_{i,n-r}t_{n-r}$$

dove c_i e $a_{i,j}$ sono costanti reali.

Per scrivere in maniera ordinata le soluzioni procediamo come nell’esempio. Denotiamo con \mathbf{v}_0 la soluzione ottenuta assegnando il valore zero a tutti i parametri t_1, t_2, \dots, t_{n-r} . Con \mathbf{v}_1 denotiamo il vettore dei coefficienti del parametro t_1 nell’espressione della soluzione generale del sistema. Equivalentemente

$$\mathbf{v}_1 = \mathbf{w}_1 - \mathbf{v}_0$$

dove \mathbf{w}_1 è la soluzione corrispondente a $t_1 = 1, t_2 = t_3 = \dots = t_{n-r} = 0$. Analogamente, definiamo \mathbf{v}_2 come il vettore dei coefficienti del parametro t_2 nell’espressione della soluzione generale del sistema e proseguendo così via definiamo $\mathbf{v}_3, \dots, \mathbf{v}_{n-r}$. L’espressione della soluzione generale del sistema diviene:

$$(6.12) \quad \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{v}_0 + t_1 \mathbf{v}_1 + t_2 \mathbf{v}_2 + \cdots + t_{n-r} \mathbf{v}_{n-r}$$

Tale espressione associa a ogni valore di $(t_1, t_2, \dots, t_{n-r})$ una soluzione del sistema. Inoltre, a valori distinti dei parametri corrispondono soluzioni distinte: supponiamo infatti che \mathbf{w} sia la soluzione corrispondente a $(t_1, t_2, \dots, t_{n-r})$ e \mathbf{w}' sia la soluzione corrispondente a $(t'_1, t'_2, \dots, t'_{n-r})$. Se per qualche indice j i parametri t_j e t'_j sono diversi, allora le corrispondenti variabili libere in \mathbf{w} e \mathbf{w}' sono diverse e, quindi, le due soluzioni sono distinte. Questo mostra che le soluzioni del sistema sono in corrispondenza biunivoca con l'insieme \mathbb{K}^{n-r} dei parametri.

Infine, siccome le soluzioni di $\mathbf{Ax} = \mathbf{b}$ si ottengono sommando a \mathbf{v}_0 le soluzioni del sistema omogeneo $\mathbf{Ax} = \mathbf{0}$ (per la proposizione 5.5), da (6.12) segue che le soluzioni del sistema omogeneo sono i vettori della forma

$$t_1 \mathbf{v}_1 + t_2 \mathbf{v}_2 + \cdots + t_{n-r} \mathbf{v}_{n-r}$$

(in alternativa, il lettore può osservare che quando il termine noto \mathbf{b} è nullo, col metodo di soluzione descritto si ottiene $\mathbf{v}_0 = \mathbf{0}$). 

Riassumendo:

sia $\mathbf{Ax} = \mathbf{b}$ un sistema in n incognite. Posto $r = r(\mathbf{A})$ e $r' = r([\mathbf{A}|\mathbf{b}])$, per il teorema abbiamo tre possibilità:

- a) se $n = r = r'$, il sistema ammette un'unica soluzione: si dice che il sistema è *determinato*;
- b) se $n > r = r'$, il sistema ammette infinite soluzioni, che dipendono da $n - r$ parametri: si scrive che il sistema ammette ∞^{n-r} soluzioni e si dice che il sistema è *sottodeterminato*;
- c) se $r \neq r'$, il sistema non ha soluzioni: si dice che il sistema è *impossibile* o *sovradeterminato*.

Vedremo che r' è il numero di equazioni indipendenti del sistema. Se il sistema ammette soluzioni ($r = r'$), il sistema è determinato quando il numero delle incognite n è uguale al numero di equazioni indipendenti; altrimenti il sistema è sottodeterminato, non ci sono abbastanza equazioni per determinare la soluzione. Infine si noti che:

un sistema quadrato di n equazioni in n incognite ha una e una sola soluzione se e solo se il rango della matrice dei coefficienti è n (il massimo possibile).

Infatti, se il rango r è uguale a n , la soluzione esiste ed è unica per il teorema di Cramer. Se $r < n$, per il teorema di Rouché-Capelli il sistema è impossibile oppure ammette infinite soluzioni.

OSSERVAZIONE I sistemi sottodeterminati (infinte soluzioni) e sovradeterminati (nessuna soluzione) occorrono naturalmente in molte applicazioni. Sistemi sottodeterminati, per esempio, sono quelli che intervengono nella ricerca degli *autovettori* di una matrice, un argomento che tratteremo successivamente nel testo e che è fondamentale in quasi tutte le applicazioni dell'algebra lineare. I sistemi sovradeterminati

hanno spesso origine da dati raccolti sperimentalmente, quando la teoria prevede l'esistenza di una soluzione, ma gli errori di misura producono un sistema impossibile. In questo caso si cerca di ottenere una soluzione il più possibile coerente con i dati sperimentali: si veda il paragrafo sulla soluzione ai minimi quadrati di un sistema sovradeterminato (equazioni normali) nel capitolo sugli spazi euclidei.



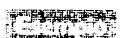
Consideriamo un sistema di un'unica equazione $ax = b$ nella variabile x . In questo caso $n = 1$, per cui r e r' possono essere solo 0 o 1. L'equazione ha un'unica soluzione se $1 = r = r'$, cioè se $a \neq 0$; l'equazione ha ∞^1 soluzioni se $r = r' = 0$, cioè se $a = b = 0$; infine, l'equazione non ha soluzioni se $r = 0$ e $r' = 1$, cioè se $a = 0$ e $b \neq 0$.



Supponiamo che il sistema consista di un'unica equazione:

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

L'equazione è non banale se almeno uno dei coefficienti a_i è diverso da zero; in tal caso il rango di $\mathbf{A} = [a_1 \ a_2 \ \cdots \ a_n]$ è 1, il rango della matrice completa è ancora 1. L'insieme delle soluzioni si dice *iperpiano*. I punti di un iperpiano dipendono da $n-1$ parametri: se l' i -esimo coefficiente a_i è diverso da zero, dall'equazione si ricava x_i in funzione delle altre $n-1$ variabili, a cui si possono attribuire valori arbitrari. Nel caso $\mathbb{K} = \mathbb{R}$ e $n = 2$ (rispettivamente, 3), un iperpiano è una retta (rispettivamente, un piano).



Questo è un esempio in cui è naturale cercare la soluzione di un sistema sottodeterminato. Il problema è di trovare l'equazione del piano che passa per tre punti non allineati dello spazio cartesiano. L'equazione del piano ha la forma

$$ax + by + cz = d$$

ma i coefficienti $[a, b, c, d]^T$ dell'equazione, che sono le incognite del problema, non sono univocamente determinati perché se moltiplichiamo l'equazione per uno scalare non nullo il piano non cambia. Quindi imponendo il passaggio per i tre punti è giusto ottenere un sistema indeterminato, di 3 equazioni nelle 4 incognite $[a, b, c, d]^T$: basterà poi scegliere una soluzione non nulla per scrivere l'equazione del piano.

Per un esempio specifico, cerchiamo un'equazione del piano che passa per i punti $A = (1, 0, -1)$, $B = (2, 1, 2)$, $C = (-1, 2, -1)$. Le coordinate dei tre punti devono soddisfare l'equazione $ax + by + cz = d$ del piano. Questo significa che i coefficienti a, b, c, d dell'equazione devono risolvere il sistema lineare omogeneo

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

Le soluzioni del sistema sono i vettori $t\mathbf{v}_0 = t[3, 3, -2, 5]^T$, per cui il piano ha un'equazione della forma

$$3x + 3y - 2z = 5$$

Questo problema era già stato risolto nel primo capitolo, ricavando l'equazione del piano dalla forma vettoriale

$$\overrightarrow{AX} \cdot (\overrightarrow{AB} \times \overrightarrow{AC}) = 0$$

La soluzione proposta qui è più elementare e si generalizza immediatamente a dimensione arbitraria.



Consideriamo il caso di un sistema di due equazioni in tre incognite a coefficienti reali:

$$(6.13) \quad \begin{cases} a_1x + b_1y + c_1z = d_1 \\ a_2x + b_2y + c_2z = d_2 \end{cases}$$

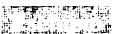
Supponiamo che le equazioni rappresentino due piani \mathbf{H}_1 e \mathbf{H}_2 in \mathbb{R}^3 , cioè che i vettori $\mathbf{n}_1 = [a_1, b_1, c_1]^T$ e $\mathbf{n}_2 = [a_2, b_2, c_2]^T$ siano non nulli. Denotiamo con r e r' i ranghi della matrice dei coefficienti e della matrice completa del sistema. Allora $1 \leq r \leq r' \leq 2$: la prima diseguaglianza perché la matrice dei coefficienti è non nulla, l'ultima perché la matrice completa ha due righe. I casi possibili sono quindi

- a) $r' = r = 2$
- b) $r' = 2, r = 1$
- c) $r' = r = 1$

Per il teorema di Rouché-Capelli, nel caso a) il sistema ha infinite soluzioni che dipendono da un parametro e quindi formano una retta di equazioni parametriche $\mathbf{v}_0 + t\mathbf{v}_1$ in \mathbb{R}^3 . Nel caso b) il sistema non ha soluzione e quindi i due piani sono paralleli. Nel caso c) il sistema ha ∞^2 soluzioni ed è equivalente al sistema formato dalla sola prima equazione: i due piani coincidono.

Vediamo così che i due piani sono paralleli se e solo se $r = 1$. In effetti, è semplice verificare direttamente che $r = 1$ se e solo se il vettore normale \mathbf{n}_2 del secondo piano è un multiplo del vettore normale \mathbf{n}_1 del primo piano, il che geometricamente significa che i due vettori normali e quindi i due piani, sono paralleli.

Infine, se i due piani non sono paralleli, vale a) e per Rouché-Capelli l'intersezione è una retta: ritroviamo così algebricamente il fatto che due piani non paralleli in \mathbb{R}^3 si intersecano in una retta.



Lo studio della posizione reciproca di due rette nello spazio, espresse mediante equazioni parametriche, porta a considerare un sistema di *tre equazioni lineari in due incognite*: questo fornisce un esempio naturale in cui il numero di equazioni supera il numero delle incognite e in cui non ci si aspetta di trovare delle soluzioni.

Infatti supponiamo che \mathbf{r}_1 sia la retta per il punto P con vettore direzione \mathbf{v} e che \mathbf{r}_2 sia la retta per il punto Q con vettore direzione \mathbf{w} . Le due rette hanno equazioni parametriche

$$\begin{cases} x(t_1) = x_P + t_1x(\mathbf{v}) \\ y(t_1) = y_P + t_1y(\mathbf{v}) \\ z(t_1) = z_P + t_1z(\mathbf{v}) \end{cases} \quad \text{e} \quad \begin{cases} x(t_2) = x_Q + t_2x(\mathbf{w}) \\ y(t_2) = y_Q + t_2y(\mathbf{w}) \\ z(t_2) = z_Q + t_2z(\mathbf{w}) \end{cases}$$

Sia $P(t_1)$ il punto corrispondente al valore t_1 del parametro sulla prima retta e sia $Q(t_2)$ il punto corrispondente al valore t_2 del parametro sulla seconda retta. Ora, $P(t_1) = Q(t_2)$ è un punto che appartiene a entrambe le rette se e solo se (t_1, t_2) è una soluzione del sistema lineare

$$\begin{cases} x_P + t_1 x(\mathbf{v}) = x_Q + t_2 x(\mathbf{w}) \\ y_P + t_1 y(\mathbf{v}) = y_Q + t_2 y(\mathbf{w}) \\ z_P + t_1 z(\mathbf{v}) = z_Q + t_2 z(\mathbf{w}) \end{cases}$$

che si può riscrivere nella forma equivalente

$$(6.14) \quad \mathbf{A} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = \mathbf{u}$$

dove \mathbf{A} è la matrice di tipo $(3, 2)$ che ha per colonne i vettori \mathbf{v} e $-\mathbf{w}$, mentre \mathbf{u} è il vettore \overrightarrow{QP} .

Il rango di \mathbf{A} è almeno 1 (perché la prima colonna \mathbf{v} non è il vettore nullo) e al massimo 2 (perché \mathbf{A} ha due colonne). Si verifica facilmente che il rango di \mathbf{A} è 1 se e solo se \mathbf{w} è un multiplo scalare di \mathbf{v} , cioè le due rette sono parallele.

D'altra parte, la matrice completa del sistema è una matrice di tipo $(3, 3)$ e quindi ha al massimo rango 3. Posto $r = r(\mathbf{A})$ e $r' = r([\mathbf{A}|\mathbf{u}])$, tenuto conto che $r' = r$ oppure $r' = r + 1$, i casi possibili sono:

- a) $r' = 3$;
- b) $r' = r = 2$;
- c) $r' = 2, r = 1$;
- d) $r' = r = 1$.

Nel caso a) il rango della matrice completa è $r' = 3$. Allora il rango di \mathbf{A} dev'essere due perché $r' - r \leq 1$ e $r \leq 2$. Per Rouché-Capelli il sistema (6.14) non ha soluzione e quindi le due rette non hanno punti in comune. Siccome $r = 2$, le due rette non sono parallele. Quindi nel caso a) le due rette sono sghembe.

Nel caso b), per Rouché -Capelli le due rette hanno un unico punto in comune e sono quindi incidenti.

Nel caso c), le due rette sono parallele ($r = 1$) e distinte (non hanno punti in comune perché $r \neq r'$).

Nel caso d) le due rette coincidono (per Rouché-Capelli hanno una retta in comune!)

Questi risultati sono in accordo con quelli ottenuti nel capitolo 1. Il caso a) si verifica quando la matrice completa del sistema ha rango massimo e vedremo che questo succede se e solo se il suo *determinante* è diverso da zero. Ora il determinante è, a meno del segno, il prodotto misto $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$. Ritroviamo così che le due rette sono sghembe se e solo se $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) \neq 0$.

Esercizi svolti

- 1 Eseguire - ove possibile - i seguenti prodotti matrice-vettore:

$$a) \begin{bmatrix} 1 & 2 & -1 \\ 2 & -3 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \quad b) \begin{bmatrix} 2 & 3 \\ 1 & -1 \\ 3 & -2 \end{bmatrix} \begin{bmatrix} \ln 2 \\ \ln 3 \end{bmatrix}, \quad c) \begin{bmatrix} 2 & 3 & 4 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} \pi \\ 0 \end{bmatrix}$$

Soluzione

a) Le colonne della matrice sono tre, tante quante le componenti del vettore, quindi è possibile eseguire il prodotto. Inoltre, la matrice ha due righe, pertanto il risultato del prodotto sarà un vettore a due componenti:

$$\begin{bmatrix} 1 \cdot 1 + 2 \cdot (-1) + -1 \cdot 2 \\ 2 \cdot 1 + -3 \cdot (-1) + 0 \cdot 2 \end{bmatrix} = \begin{bmatrix} -3 \\ 5 \end{bmatrix}$$

b) Le colonne della matrice sono due, tante quante le componenti del vettore, quindi è possibile eseguire il prodotto. Inoltre, la matrice ha tre righe, pertanto il risultato del prodotto sarà un vettore a tre componenti:

$$\begin{bmatrix} 2 \cdot \ln 2 + 3 \cdot \ln 3 \\ 1 \cdot \ln 2 + (-1) \cdot \ln 3 \\ 3 \cdot \ln 2 + (-2) \cdot \ln 3 \end{bmatrix} = \begin{bmatrix} \ln 108 \\ \ln \frac{2}{3} \\ \ln \frac{8}{9} \end{bmatrix}$$

c) Le colonne della matrice sono tre, mentre le componenti del vettore sono due, non è possibile eseguire il prodotto.

2 Ridurre a scala le seguenti matrici:

$$a) \begin{bmatrix} 1 & -1 & 0 & 3 \\ 3 & -1 & 1 & 10 \\ -1 & 5 & 2 & 1 \end{bmatrix}, \quad b) \begin{bmatrix} 1 & 2 & 1 & 0 & 3 \\ -3 & -4 & -4 & 3 & 0 \\ 2 & 2 & 3 & -3 & -6 \\ 3 & 10 & 3 & 3 & 1 \end{bmatrix}, \quad c) \begin{bmatrix} 1 & -1 & 3 & -2 & 0 \\ -2 & 2 & -4 & 0 & 2 \\ 3 & -4 & 8 & -3 & 1 \\ 2 & -1 & 8 & -4 & 1 \end{bmatrix}$$

Soluzione

a) Per prima cosa, sistemiamo la prima colonna: sottraiamo alla seconda riga il triplo della prima e sommiamo alla terza riga la prima:

$$\begin{aligned} II_R &\rightarrow II_R - 3I_R & \begin{bmatrix} 1 & -1 & 0 & 3 \\ 0 & 2 & 1 & 1 \\ 0 & 4 & 2 & 4 \end{bmatrix} \\ III_R &\rightarrow III_R + I_R \end{aligned}$$

usando il 2 nella seconda posizione della seconda riga, eliminiamo il 4 nella seconda posizione della terza riga, alla terza riga sottraiamo il doppio della seconda:

$$III_R \rightarrow III_R - 2II_R \begin{bmatrix} 1 & -1 & 0 & 3 \\ 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

In questo caso, la riduzione a scala ha potuto procedere senza intoppi, nei prossimi due casi esamineremo situazioni in cui è necessario ricorrere a qualche rimedio.

b) Sistemiamo la prima colonna: sommiamo alla seconda riga il triplo della prima, sottraiamo alla terza il doppio della prima e sottraiamo alla quarta riga il triplo della prima:

$$\begin{aligned} II_R &\rightarrow II_R + 3I_R & \begin{bmatrix} 1 & 2 & 1 & 0 & 3 \\ 0 & 2 & -1 & 3 & 9 \end{bmatrix} \\ III_R &\rightarrow III_R - 2I_R & \begin{bmatrix} 0 & -2 & 1 & -3 & -9 \\ 0 & 4 & 0 & 3 & 1 \end{bmatrix} \\ IV_R &\rightarrow IV_R - 3I_R \end{aligned}$$

procediamo con la seconda colonna, sommiamo alla terza riga la seconda e sottraiamo alla quarta il doppio della seconda:

$$\begin{array}{l} III_R \rightarrow III_R + II_R \\ IV_R \rightarrow IV_R - 2II_R \end{array} \left[\begin{array}{ccccc} 1 & 2 & 1 & 0 & 3 \\ 0 & 2 & -1 & 3 & 9 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -3 & -17 \end{array} \right]$$

poiché le righe di zeri devono trovarsi in fondo alla matrice, scambiamo terza e quarta riga:

$$\begin{array}{l} IV_R \rightarrow III_R \\ III_R \rightarrow IV_R \end{array} \left[\begin{array}{ccccc} 1 & 2 & 1 & 0 & 3 \\ 0 & 2 & -1 & 3 & 9 \\ 0 & 0 & 2 & -3 & -17 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

c) Sistemiamo la prima colonna: sommiamo alla seconda riga il doppio della prima, sottraiamo alla terza il triplo della prima e sottraiamo alla quarta il doppio della prima:

$$\begin{array}{l} II_R \rightarrow II_R + I_R \\ III_R \rightarrow III_R - 3I_R \\ IV_R \rightarrow IV_R - 2I_R \end{array} \left[\begin{array}{ccccc} 1 & -1 & 3 & -2 & 0 \\ 0 & 0 & 2 & -4 & 2 \\ 0 & -1 & -1 & 3 & 1 \\ 0 & 1 & 2 & 0 & 1 \end{array} \right]$$

la seconda riga non può essere usata per procedere nella riduzione, scambiamola con la terza:

$$\begin{array}{l} III_R \rightarrow II_R \\ II_R \rightarrow III_R \end{array} \left[\begin{array}{ccccc} 1 & -1 & 2 & -3 & 0 \\ 0 & -1 & -1 & 3 & 1 \\ 0 & 0 & 2 & -4 & 2 \\ 0 & 1 & 2 & 0 & 1 \end{array} \right]$$

adesso è possibile continuare, la terza riga è già a posto, sommiamo alla quarta la seconda:

$$\begin{array}{l} IV_R \rightarrow IV_R + II_R \end{array} \left[\begin{array}{ccccc} 1 & -1 & 2 & -3 & 0 \\ 0 & -1 & -1 & 3 & 1 \\ 0 & 0 & 2 & -4 & 2 \\ 0 & 0 & 1 & 3 & 2 \end{array} \right]$$

Concludiamo, sottraendo alla quarta riga la metà della terza:

$$\begin{array}{l} IV_R \rightarrow IV_R - \frac{1}{2}III_R \end{array} \left[\begin{array}{ccccc} 1 & -1 & 2 & -3 & 0 \\ 0 & -1 & -1 & 3 & 1 \\ 0 & 0 & 2 & -4 & 2 \\ 0 & 0 & 0 & 5 & 1 \end{array} \right]$$

 Stabilire il rango delle tre matrici dell'esercizio precedente.

Soluzione

Il rango di una matrice è il numero di pivot, una volta che è stata ridotta a scala; affianchiamo ciascuna matrice alla sua riduzione a scala, in cui i pivot sono segnati in grassetto.

a) Il rango della prima matrice è tre:

$$\left[\begin{array}{cccc} 1 & -1 & 0 & 3 \\ 3 & -1 & 1 & 10 \\ -1 & 5 & 2 & 1 \end{array} \right] \longrightarrow \left[\begin{array}{cccc} \mathbf{1} & -1 & 0 & 3 \\ 0 & \mathbf{2} & 1 & 1 \\ 0 & 0 & 0 & \mathbf{2} \end{array} \right]$$

b) Il rango della seconda matrice è tre:

$$\left[\begin{array}{ccccc} 1 & 2 & 1 & 0 & 3 \\ -3 & -4 & -4 & 3 & 0 \\ 2 & 2 & 3 & -3 & -6 \\ 3 & 10 & 3 & 3 & 1 \end{array} \right] \longrightarrow \left[\begin{array}{ccccc} \mathbf{1} & 2 & 1 & 0 & 3 \\ 0 & \mathbf{2} & -1 & 3 & 9 \\ 0 & 0 & \mathbf{2} & -3 & -17 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

b) Il rango della terza matrice è quattro:

$$\left[\begin{array}{ccccc} 1 & -1 & 3 & -2 & 0 \\ -2 & 2 & -4 & 0 & 2 \\ 3 & -4 & 8 & -3 & 1 \\ 2 & -1 & 8 & -4 & 1 \end{array} \right] \longrightarrow \left[\begin{array}{ccccc} \mathbf{1} & -1 & 2 & -3 & 0 \\ 0 & -\mathbf{1} & -1 & 3 & 1 \\ 0 & 0 & \mathbf{2} & -4 & 2 \\ 0 & 0 & 0 & \mathbf{5} & 1 \end{array} \right]$$

 Determinare al variare di $\alpha \in \mathbb{R}$ il rango della matrice

$$\left[\begin{array}{ccccc} 1 & 2 & -1 & -1 & 0 \\ 2 & 5 & 2 & -2 & 0 \\ -1 & -1 & 5 & 3 & 1 \\ 3 & 8 & 5 & -1 & 1 \\ 3 & 9 & 9 & 1 & \alpha \end{array} \right]$$

Soluzione

Riduciamo a scala la matrice, innanzitutto la prima colonna:

$$\left[\begin{array}{ccccc} 1 & 2 & -1 & -1 & 0 \\ 0 & 1 & 4 & 0 & 0 \\ 0 & 1 & 4 & 2 & 1 \\ 0 & 2 & 8 & 2 & 1 \\ 0 & 3 & 12 & 4 & \alpha \end{array} \right]$$

poi la seconda:

$$\left[\begin{array}{ccccc} 1 & 2 & -1 & -1 & 0 \\ 0 & 1 & 4 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 4 & \alpha \end{array} \right]$$

nella terza colonna non vi è alcun pivot, passando alla quarta colonna, il pivot si trova sulla terza riga, se continuiamo con la riduzione a scala troviamo

$$\left[\begin{array}{ccccc} 1 & 2 & -1 & -1 & 0 \\ 0 & 1 & 4 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha - 2 \end{array} \right]$$

la quarta riga dev'essere spostata sul fondo, la matrice diventa allora

$$\left[\begin{array}{ccccc} 1 & 2 & -1 & -1 & 0 \\ 0 & 1 & 4 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & \alpha - 2 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

pertanto: se $\alpha \neq 2$ nella quinta posizione della quarta riga troviamo il quarto pivot e $r(\mathbf{A}) = 4$, se $\alpha = 2$ la quarta e la quinta riga sono entrambe nulle, vi sono solo tre pivot e quindi $r(\mathbf{A}) = 3$.

5 Utilizzando il MEG, risolvere – ove possibile – i seguenti sistemi lineari:

$$a) \begin{cases} x_1 + 3x_2 + 2x_3 = 5 \\ 2x_1 + 5x_2 + 4x_3 + x_4 = 5 \\ x_1 + 5x_2 + 4x_3 - x_4 = 10 \\ 3x_1 + 10x_2 + 8x_3 + x_4 = 12 \end{cases}, \quad b) \begin{cases} x_1 + 2x_3 + x_4 = 3 \\ 2x_1 + 2x_2 + 5x_3 + x_4 = 7 \\ 3x_1 - 2x_2 + 5x_3 + 6x_4 = 10 \\ -x_1 + 6x_2 + x_3 = 3 \end{cases}$$

commentando il risultato ottenuto alla luce del teorema di Rouché-Capelli.

Soluzione

a) Riduciamo a scala la matrice completa, la prima colonna:

$$\left[\begin{array}{cccc|c} 1 & 3 & 2 & 0 & 5 \\ 0 & -1 & 0 & 1 & -5 \\ 0 & 2 & 2 & -1 & 5 \\ 0 & 1 & 2 & 1 & -3 \end{array} \right]$$

la seconda:

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

e la terza:

$$\left[\begin{array}{cccc|c} 1 & 3 & 2 & 0 & 5 \\ 0 & -1 & 0 & 1 & -5 \\ 0 & 0 & 2 & 1 & -5 \\ 0 & 0 & 0 & 1 & -3 \end{array} \right]$$

la riduzione è terminata, la matrice dei coefficienti ha rango 4, così come la matrice completa; anche le incognite sono 4, vi è pertanto una sola soluzione. Per sostituzione all'indietro, ricaviamo

$$x_4 = -3, \quad x_3 = -1, \quad x_2 = 2, \quad x_1 = 1$$

b) Anche in questo caso, riduciamo a scala la matrice completa; la prima colonna:

$$\left[\begin{array}{cccc|c} 1 & 0 & 2 & 1 & 3 \\ 0 & 2 & 1 & -1 & 1 \\ 0 & -2 & -1 & 3 & 1 \\ 0 & 6 & 3 & 1 & 6 \end{array} \right]$$

la seconda:

$$\left[\begin{array}{cccc|c} 1 & 0 & 2 & 1 & 3 \\ 0 & 2 & 1 & -1 & 1 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 4 & 3 \end{array} \right]$$

la terza è già sistemata, procediamo con la quarta:

$$\left[\begin{array}{cccc|c} 1 & 0 & 2 & 1 & 3 \\ 0 & 2 & 1 & -1 & 1 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 0 & -1 \end{array} \right]$$

il rango della matrice completa è 4, il rango della matrice dei coefficienti è 3, il sistema non ammette soluzione.

Utilizzando il MEG, risolvere il sistema lineare:

$$\left\{ \begin{array}{l} x_1 + 2x_2 - x_3 + 3x_4 + 2x_5 + x_6 = 0 \\ x_1 + 2x_2 + x_3 + 4x_4 - 2x_5 + 5x_6 = 1 \\ 2x_1 + 4x_2 + 4x_3 + 12x_4 + 3x_5 + 16x_6 = 2 \\ -x_1 - 2x_2 + 3x_3 - 8x_4 + 2x_6 = 5 \end{array} \right.$$

commentando il risultato ottenuto alla luce del teorema di Rouché-Capelli.

Soluzione

Riduciamo a scala la matrice completa, prima colonna:

$$\left[\begin{array}{cccccc|c} 1 & 2 & -1 & 3 & 2 & 1 & 0 \\ 0 & 0 & 2 & 1 & 0 & 4 & 1 \\ 0 & 0 & 6 & 6 & -1 & 14 & 2 \\ 0 & 0 & 2 & -5 & 2 & 3 & 5 \end{array} \right]$$

nella seconda non compaiono pivot, passiamo alla terza:

$$\left[\begin{array}{cccccc|c} 1 & 2 & -1 & 3 & 2 & 1 & 0 \\ 0 & 0 & 2 & 1 & 0 & 4 & 1 \\ 0 & 0 & 0 & 3 & -1 & 2 & -1 \\ 0 & 0 & 0 & -6 & 2 & -1 & 4 \end{array} \right]$$

e alla quarta:

$$\left[\begin{array}{cccccc|c} 1 & 2 & -1 & 3 & 2 & 1 & 0 \\ 0 & 0 & 2 & 1 & 0 & 4 & 1 \\ 0 & 0 & 0 & 3 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 3 & 2 \end{array} \right]$$

osserviamo che non vi sono pivot sulla seconda e sulla quarta colonna, quindi x_2 e x_5 sono variabili libere, mentre le altre sono variabili dipendenti; per sostituzione all'indietro, otteniamo:

$$\begin{aligned} x_6 &= \frac{2}{3}, & x_4 &= \frac{1}{3}(x_5 - 2x_6 - 1) = \frac{1}{3}x_5 - \frac{7}{9}, \\ x_3 &= \frac{1}{2}(-x_4 - 4x_6 + 1) = -\frac{1}{6}x_5 - \frac{4}{9}, \\ x_1 &= -2x_2 + x_3 - 3x_4 - 2x_5 - x_6 = -2x_2 - \frac{19}{6}x_5 + \frac{11}{9}; \end{aligned}$$

assegnando a x_2 e a x_5 i valori arbitrari t e s rispettivamente, possiamo scrivere la soluzione del sistema in forma vettoriale:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} \frac{11}{9} - 2t - \frac{19}{6}s \\ t \\ -\frac{4}{9} - \frac{1}{6}s \\ -\frac{7}{9} + \frac{1}{3}s \\ s \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} \frac{11}{9} \\ 0 \\ -\frac{4}{9} \\ -\frac{7}{9} \\ 0 \\ \frac{2}{3} \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} \frac{19}{6} \\ 0 \\ -\frac{1}{6} \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

o meglio ancora, per l'arbitrarietà di s

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} \frac{11}{9} \\ 0 \\ -\frac{4}{9} \\ -\frac{7}{9} \\ 0 \\ \frac{2}{3} \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 19 \\ 0 \\ -1 \\ 2 \\ 6 \\ 0 \end{bmatrix}$$

osserviamo che il sistema aveva 6 incognite, mentre il rango di \mathbf{A} e di $\mathbf{A}|\mathbf{b}$ era uguale a 4, il sistema doveva avere soluzioni, dipendenti da 2 parametri liberi.

 Utilizzando il MEG, risolvere il sistema lineare:

$$\begin{cases} x_1 + 2x_2 - x_3 + x_4 + 2x_6 + 2x_7 = 0 \\ -2x_1 - 3x_2 + 4x_3 - 3x_4 - 2x_5 - 3x_6 - 3x_7 = 0 \\ -2x_1 - 3x_2 + 4x_3 - x_4 - x_5 + x_6 + 3x_7 = 0 \\ 3x_1 + 4x_2 - 7x_3 + x_4 - 2x_5 - 4x_6 - 8x_7 = 0 \\ -2x_1 - 2x_2 + 6x_3 - 6x_4 - 5x_5 - 8x_6 = 0 \\ x_1 + x_2 - 3x_3 + 6x_4 + 4x_5 + 10x_6 + 9x_7 = 0 \end{cases}$$

commentando il risultato ottenuto alla luce del teorema di Rouché-Capelli.

Soluzione

Il sistema è omogeneo e pertanto ammette senz'altro soluzione. Riduciamo a scala la matrice dei coefficienti, partendo dalla prima colonna:

$$\left[\begin{array}{ccccccc} 1 & 2 & -1 & 1 & 0 & 2 & 2 \\ -2 & -3 & 4 & -3 & -2 & -3 & -3 \\ -2 & -3 & 4 & -1 & -1 & 1 & 3 \\ 3 & 4 & -7 & 1 & 2 & -4 & -8 \\ -2 & -2 & 6 & -6 & -5 & -8 & 0 \\ 1 & 1 & -3 & 6 & 4 & 10 & 9 \end{array} \right]$$

poi la seconda:

$$\left[\begin{array}{ccccccc} 1 & 2 & -1 & 1 & 0 & 2 & 2 \\ 0 & 1 & 2 & -1 & -2 & 1 & 1 \\ 0 & 1 & 2 & 1 & -1 & 5 & 7 \\ 0 & -2 & -4 & -2 & 2 & -10 & -14 \\ 0 & 2 & 4 & -4 & -5 & -4 & 4 \\ 0 & -1 & -2 & 5 & 4 & 8 & 7 \end{array} \right]$$

la terza:

$$\left[\begin{array}{ccccccc} 1 & 2 & -1 & 1 & 0 & 2 & 2 \\ 0 & 1 & 2 & -1 & -2 & 1 & 1 \\ 0 & 0 & 0 & 2 & 1 & 4 & 6 \\ 0 & 0 & 0 & -4 & -2 & -8 & -12 \\ 0 & 0 & 0 & -2 & -1 & -6 & 2 \\ 0 & 0 & 0 & 4 & 2 & 9 & 8 \end{array} \right]$$

la quarta:

$$\left[\begin{array}{ccccccc} 1 & 2 & -1 & 1 & 0 & 2 & 2 \\ 0 & 1 & 2 & -1 & -2 & 1 & 1 \\ 0 & 0 & 0 & 2 & 1 & 4 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & 8 \\ 0 & 0 & 0 & 0 & 0 & 1 & -4 \end{array} \right]$$

la quarta riga viene spostata in fondo:

$$\left[\begin{array}{ccccccc} 1 & 2 & -1 & 1 & 0 & 2 & 2 \\ 0 & 1 & 2 & -1 & -2 & 1 & 1 \\ 0 & 0 & 0 & 2 & 1 & 4 & 6 \\ 0 & 0 & 0 & 0 & 0 & -2 & 8 \\ 0 & 0 & 0 & 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

passiamo alla quinta colonna:

$$\left[\begin{array}{ccccccc} 1 & 2 & -1 & 1 & 0 & 2 & 2 \\ 0 & 1 & 2 & -1 & -2 & 1 & 1 \\ 0 & 0 & 0 & 2 & 1 & 4 & 6 \\ 0 & 0 & 0 & 0 & 0 & -2 & 8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

trascurando le ultime due righe della matrice, le equazioni si riducono a 4 (è il rango di \mathbf{A}), vi sono pertanto quattro variabili dipendenti (in corrispondenza dei pivot) e tre variabili libere: x_3, x_5 e x_7 ; per sostituzione all'indietro, si ricava

$$\begin{aligned} x_6 &= 4x_7, & x_4 &= -\frac{1}{2}x_5 - 11x_7 \\ x_2 &= -2x_3 + x_4 + 2x_5 - x_6 - x_7 = -2x_3 + \frac{3}{2}x_5 - 16x_7 \\ x_1 &= -2x_2 + x_3 - x_4 - 2x_6 - 2x_7 = 5x_3 - \frac{5}{2}x_5 + 17x_7 \end{aligned}$$

assegnando a x_3, x_5 e x_7 i valori arbitrari $t_1, 2t_2$ e t_3 rispettivamente, possiamo scrivere la soluzione del sistema in forma vettoriale (sfruttiamo l'arbitrarietà di t_2 per eliminare i denominatori):

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} 5t_1 - 5t_2 + 17t_3 \\ -2t_1 + 3t_2 - 16t_3 \\ t_1 \\ -t_2 + 11t_3 \\ 2t_2 \\ 4t_3 \\ t_3 \end{bmatrix} = t_1 \begin{bmatrix} 5 \\ -2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} -5 \\ 3 \\ 0 \\ -1 \\ 2 \\ 0 \\ 0 \end{bmatrix} + t_3 \begin{bmatrix} 17 \\ -16 \\ 0 \\ 11 \\ 0 \\ 4 \\ 1 \end{bmatrix}$$

Determinare al variare di $\alpha \in \mathbb{R}$ le soluzioni del sistema lineare

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

Interpretare geometricamente i risultati ottenuti.

Soluzione

La matrice associata al sistema è

$$[\mathbf{A}|\mathbf{b}] = \left[\begin{array}{ccc|c} 1 & -1 & \alpha & -1 \\ \alpha & 0 & 2 & -2 \\ 0 & 1 & 1 & -1 \end{array} \right]$$

prima di effettuare la riduzione a scala, osserviamo che la prima operazione consisterebbe nel sottrarre alla seconda riga la prima, moltiplicata per α , tale operazione è significativa se $\alpha \neq 0$, per $\alpha = 0$ il sistema si risolve immediatamente:

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

possiamo allora supporre $\alpha \neq 0$ e procedere nella riduzione a scala; incominciamo sottraendo alla seconda riga la prima, moltiplicata per α :

$$\left[\begin{array}{ccc|c} 1 & -1 & \alpha & -1 \\ 0 & \alpha & 2 - \alpha^2 & \alpha - 2 \\ 0 & 1 & 1 & -1 \end{array} \right]$$

è ora più prudente scambiare tra loro seconda e terza riga, poiché è più agevole usare un pivot pari a 1 anziché ad α :

$$\left[\begin{array}{ccc|c} 1 & -1 & \alpha & -1 \\ 0 & 1 & 1 & -1 \\ 0 & \alpha & 2 - \alpha^2 & \alpha - 2 \end{array} \right]$$

adesso sottraiamo alla terza riga la seconda moltiplicata per α :

$$\left[\begin{array}{ccc|c} 1 & -1 & \alpha & -1 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 2 - \alpha^2 - \alpha & 2\alpha - 2 \end{array} \right]$$

Il coefficiente di z nella terza equazione è $2 - \alpha^2 - \alpha$, che si annulla per $\alpha = 1$ o $\alpha = -2$, deduciamo immediatamente che, se $\alpha \neq 1$ e $\alpha \neq -2$, il sistema ha un'unica soluzione, ottenuta per sostituzione all'indietro:

$$z = \frac{2\alpha - 2}{2 - \alpha^2 - \alpha} = -\frac{2}{2 + \alpha}, \quad y = -1 - z = -\frac{\alpha}{2 + \alpha}, \quad x = y - \alpha z - 1 = -\frac{2}{2 + \alpha}$$

Esaminiamo separatamente i casi $\alpha = 1$ e $\alpha = -2$; se $\alpha = 1$ la matrice diventa

$$\left[\begin{array}{ccc|c} 1 & -1 & 1 & -1 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

x e y sono variabili dipendenti, z è una variabile libera, ponendo $z = t$ ricaviamo $y = -1 - z = -1 - t$ e $x = y - 1 - z = -2 - 2t$, le soluzioni sono pertanto

$$(6.15) \quad \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -2 \\ -1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix}$$

invece, se $\alpha = -2$ la matrice diventa

$$\left[\begin{array}{ccc|c} 1 & -1 & -2 & -1 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & -6 \end{array} \right]$$

e il sistema non ha soluzione.

Il caso $\alpha \neq 1, -2$ corrisponde a tre piani che si incontrano in un punto, di coordinate

$$(x, y, z) = \left(-\frac{2}{2+\alpha}, -\frac{\alpha}{2+\alpha}, -\frac{2}{2+\alpha} \right)$$

il caso $\alpha = 1$ corrisponde a tre piani che si intersecano lungo la stessa retta, la cui equazione è (6.15); il caso $\alpha = -2$ corrisponde a tre piani che si intersecano a due a due lungo tre rette parallele.

 Per quali valori di $\alpha \in \mathbb{R}$ i piani di \mathbb{R}^3 di equazioni

$$x + 3y - \alpha z = 4, \quad x + 8y + (\alpha - 1)z = 2 + 2\alpha, \quad 2x + \alpha y - 3z = 8$$

si intersecano in una retta?

Soluzione

La richiesta dell'esercizio corrisponde a chiedere per quali valori di α il sistema

$$\begin{cases} x + 3y - \alpha z = 4 \\ x + 8y + (\alpha - 1)z = 2 + 2\alpha \\ 2x + \alpha y - 3z = 8 \end{cases}$$

ammette infinite soluzioni, dipendenti da un parametro.

Riduciamo a scala la corrispondente matrice completa,

$$\begin{aligned} [\mathbf{A}|\mathbf{b}] &= \left[\begin{array}{ccc|c} 1 & 3 & -\alpha & 4 \\ 1 & 8 & \alpha - 1 & 2 + 2\alpha \\ 2 & \alpha & -3 & 8 \end{array} \right] \quad \longrightarrow \quad \left[\begin{array}{ccc|c} 1 & 3 & -\alpha & 4 \\ 0 & 5 & 2\alpha - 1 & 2\alpha - 2 \\ 0 & \alpha - 6 & 2\alpha - 3 & 0 \end{array} \right] \\ &\longrightarrow \quad \left[\begin{array}{ccc|c} 1 & 3 & -\alpha & 4 \\ 0 & 5 & 2\alpha - 1 & 2\alpha - 2 \\ 0 & 0 & -\frac{1}{5}(2\alpha - 21)(\alpha - 1) & -\frac{2}{5}(\alpha - 6)(\alpha - 1) \end{array} \right] \end{aligned}$$

Per $\alpha = 1$ $r(\mathbf{A}) = r(\mathbf{A}|\mathbf{b}) = 2$, pertanto il sistema ammette infinite soluzioni, dipendenti da un parametro.

Osserviamo che per $\alpha \neq 1, \frac{21}{2}$ il sistema ammette una sola soluzione (i tre piani si intersecano in un punto), mentre per $\alpha = \frac{21}{2}$ i tre piani non hanno alcun punto in comune.



 Sia \mathbf{A} una matrice di tipo (m, n) . Mostrare che il sistema omogeneo $\mathbf{Ax} = \mathbf{0}$ ha sempre almeno una soluzione: il vettore nullo $\mathbf{0}$. Mostrare anche che $r([\mathbf{A}|\mathbf{0}]) = r(\mathbf{A})$, coerentemente col teorema di Rouché-Capelli.

La linearità in \mathbf{v} del prodotto $\mathbf{A}\mathbf{v}$ è una versione matematica del *principio di sovrapposizione*: mostrare che se \mathbf{v}_1 e \mathbf{v}_2 sono soluzioni, rispettivamente, dei sistemi $\mathbf{Ax} = \mathbf{b}_1$ e $\mathbf{Ax} = \mathbf{b}_2$, allora $\mathbf{v}_1 + \mathbf{v}_2$ è soluzione di $\mathbf{Ax} = \mathbf{b}_1 + \mathbf{b}_2$.

Sia \mathbf{A} una matrice con due sole righe. Mostrare che, se $r(\mathbf{A}) < 2$, allora o la prima riga è nulla oppure la seconda è un multiplo scalare della prima.

Sia \mathbf{B} una matrice con due sole colonne. Mostrare che, se $r(\mathbf{B}) < 2$, allora o la prima colonna è nulla oppure la seconda è un multiplo scalare della prima.

Sia \mathbf{A} una matrice di tipo $(3, 3)$. Mostrare che $r(\mathbf{A}) < 3$ se e solo se le tre colonne di \mathbf{A} , come vettori di \mathbb{R}^3 , sono linearmente dipendenti. Concludere che la matrice ha rango tre se e solo se il suo determinante è diverso da zero (il determinante della matrice è il prodotto misto delle sue tre colonne).

Suggerimento: usando (5.4) mostrare che le tre colonne sono linearmente dipendenti se e solo se il sistema omogeneo $\mathbf{Ax} = \mathbf{0}$ ammette una soluzione non nulla; per Rouché-Capelli questo succede se e solo se $r < n = 3$.

Sia $ax + by = c$ la generica equazione di una retta nel piano cartesiano. Si scriva il sistema lineare nelle incognite a, b, c ottenuto richiedendo che la retta passi per due punti A e B . Sia \mathbf{A} la matrice 2×3 dei coefficienti di tale sistema. Si mostri che $r(\mathbf{A}) = 2$ se e solo se i due punti sono distinti. Dedurre usando Rouché-Capelli che esiste un'unica retta in \mathbb{R}^2 che passa per due punti distinti.

Sia $ax + by + cz = d$ la generica equazione di un piano dello spazio cartesiano. Si scriva il sistema lineare nelle incognite a, b, c, d ottenuto richiedendo che il piano passi per tre punti A, B e C . Sia \mathbf{A} la matrice 3×4 dei coefficienti di tale sistema. Si mostri che $r(\mathbf{A}) = 3$ se e solo se i tre punti non sono allineati. Dedurre usando Rouché-Capelli che esiste un unico piano in \mathbb{R}^3 che passa per tre punti non allineati.

Si determini un'equazione $ax_1 + bx_2 + cx_3 + dx_4 = e$ per l'iperpiano di \mathbb{R}^4 passante per i punti $(1, -1, 3, -2), (-2, 2, -4, 0), (3, -4, 8, -3), (2, -1, 8, -4)$.

Fissato $a \in \mathbb{R}$ si considerino le rette \mathbf{r}_1 e \mathbf{r}_2 di equazioni parametriche rispettivamente $(x, y, z) = (at, a^2t, 1-t)$ e $(x, y, z) = (2+u, 2-11u, -1+3u)$. Determinare i valori del parametro a per il quale le due rette sono incidenti, per tali valori trovare il punto P di incidenza. (*Risposta:* $a = 1, P = (2, 2, -1)$ e $a = -3, P = (3, -9, 2)$.)

Sia \mathbf{A} una matrice di tipo (m, n) di rango r . Si dica se le seguenti affermazioni sono vere o false, giustificando le risposte.

- Se il sistema lineare $\mathbf{Ax} = \mathbf{b}$ ammette soluzione per ogni $\mathbf{b} \in \mathbb{R}^m$, allora $r = m$. (*Risposta:* vero.)
- Se $n = m$ ed esiste $\mathbf{b}_1 \in \mathbb{R}^n$ per cui il sistema $\mathbf{Ax} = \mathbf{b}_1$ ammette più di una soluzione, allora esiste $\mathbf{b}_2 \in \mathbb{R}^n$ per cui il sistema $\mathbf{Ax} = \mathbf{b}_2$ non ammette alcuna soluzione. (*Risposta:* vero.)
- Se esiste $\mathbf{b}_1 \in \mathbb{R}^m$ per cui il sistema $\mathbf{Ax} = \mathbf{b}_1$ ammette infinite soluzioni, allora il sistema $\mathbf{Ax} = \mathbf{b}$ ammette infinite soluzioni per ogni $\mathbf{b} \in \mathbb{R}^m$. (*Risposta:* falso.)
- Se il sistema lineare $\mathbf{Ax} = \mathbf{b}$ ammette al più una soluzione per ogni $\mathbf{b} \in \mathbb{R}^m$, allora $r = n$. (*Risposta:* vero.)

- e) Se $n = m$ ed esiste $\mathbf{b}_1 \in \mathbb{R}^n$ per cui il sistema $\mathbf{Ax} = \mathbf{b}_1$ non ammette alcuna soluzione, allora esiste $\mathbf{b}_2 \in \mathbb{R}^n$ per cui il sistema $\mathbf{Ax} = \mathbf{b}_2$ ammette infinite soluzioni. (*Risposta: vero.*)
- f) Se esiste $\mathbf{b}_1 \in \mathbb{R}^m$ per cui il sistema $\mathbf{Ax} = \mathbf{b}_1$ ammette esattamente una soluzione, allora il sistema $\mathbf{Ax} = \mathbf{b}$ ammette esattamente una soluzione per ogni $\mathbf{b} \in \mathbb{R}^m$. (*Risposta: falso.*)

Verificare il teorema di Rouché-Capelli negli esempi del paragrafo 3.

Ridurre a scala le seguenti matrici.

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ -1 & -1 & -1 & -1 & -1 \\ -2 & 0 & 3 & 6 & 9 \\ 5 & 14 & 22 & 31 & 40 \\ 3 & 8 & 15 & 21 & 28 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 2 & 1 & 1 & 2 & -1 \\ 3 & 2 & 6 & 4 & 4 & 7 & -2 \\ 1 & -6 & 2 & -2 & -3 & 1 & -3 \\ -3 & 8 & -6 & 1 & 2 & -4 & 10 \end{bmatrix}$$

Risposta:

$$\mathbf{A} \rightarrow \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{B} \rightarrow \begin{bmatrix} 1 & 0 & 2 & 1 & 1 & 2 & -1 \\ 0 & 2 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & -1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

Determinare il rango delle seguenti matrici.

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 & 3 \\ 1 & 3 & 3 & 4 \\ 4 & 7 & 3 & 10 \\ -3 & -5 & -1 & -8 \\ 1 & 7 & 10 & 9 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 2 & -1 & 3 & 1 & 2 \\ 4 & 10 & -3 & 9 & 5 & 10 \\ -2 & 4 & 6 & -18 & 2 & 4 \\ -2 & -2 & 3 & -7 & 0 & 1 \\ 1 & 2 & -1 & 5 & 2 & 8 \end{bmatrix}$$

Risposta: $r(\mathbf{A}) = 3$ e $r(\mathbf{B}) = 4$.

Determinare le soluzioni dei seguenti sistemi lineari.

$$a) \begin{cases} x + 2y = 4 \\ 2x + 5y - 3z = 0 \\ x + 4y - 8z = -18 \end{cases}, \quad b) \begin{cases} x + 2y - z = 2 \\ 3x + 8y - 2z = 0 \\ 2x - 2z = 8 \\ 3x + 4y - 8z = 18 \end{cases}, \quad c) \begin{cases} 2x + y + 3z = -1 \\ 4x + 4y + 6z = 1 \\ 6x + 5y + 6z = -1 \\ 6x + y + 3z = -11 \end{cases}$$

Risposta: a) : $(x, y, z) = (2, 1, 3)$; b) : $(x, y, z) = (2, -1, -2)$; c) non ammette soluzione.

 Risolvere il seguente sistema lineare:

$$a) \begin{cases} x_1 + x_2 + x_3 - 3x_4 + 2x_5 = 5 \\ 3x_1 + 4x_2 + 3x_3 - 7x_4 + 7x_5 = 17 \\ x_1 - 4x_4 - x_5 = 2 \\ x_2 + 3x_3 + x_4 + 7x_5 = 5 \end{cases} \quad b) \begin{cases} x_1 + 2x_2 + x_3 - 3x_4 - 4x_5 + x_6 = 0 \\ 2x_1 + 5x_2 - 7x_4 - 7x_5 - x_6 = 0 \\ 3x_1 + 7x_2 + 2x_3 - 12x_4 - 12x_5 + 4x_6 = 0 \end{cases}$$

Risposta: Ponendo $x_4 = t$,

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ -3 \\ 0 \\ 2 \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}.$$

Risolvere il seguente sistema lineare:

$$\begin{cases} x_1 + 2x_2 + x_3 - 3x_4 - 4x_5 + x_6 = 0 \\ 2x_1 + 5x_2 - 7x_4 - 7x_5 - x_6 = 0 \\ 3x_1 + 7x_2 + 2x_3 - 12x_4 - 12x_5 + 4x_6 = 0 \end{cases}$$

Risposta: Ponendo $x_4 = t_1$, $x_5 = t_2$, $x_6 = t_3$,

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = t_1 \begin{bmatrix} -9 \\ 5 \\ 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + t_3 \begin{bmatrix} -19 \\ 11 \\ -4 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Tra le parabole di equazione

$$y = ax^2 + bx + c$$

determinare quella che passa per i punti $(-1, 2)$, $(0, 3)$ e $(1, 6)$.

Algebra delle matrici

1 INTRODUZIONE

Il protagonista indiscusso di questo capitolo è il *prodotto* di matrici. È proprio la possibilità di definire un prodotto che distingue le matrici $m \times n$ dai vettori con $m \times n$ componenti e che le rende utili. Nel capitolo precedente abbiamo definito il prodotto di una matrice \mathbf{A} di tipo (m, n) con un vettore \mathbf{w} di \mathbb{K}^n ; il risultato è un vettore $\mathbf{z} = \mathbf{Aw}$ di \mathbb{K}^m . Spesso nelle applicazioni \mathbf{w} è a sua volta il prodotto di una matrice \mathbf{B} di tipo (n, p) per un vettore \mathbf{v} di \mathbb{K}^p :

$$\mathbf{z} = \mathbf{Aw} = \mathbf{A}(\mathbf{Bv})$$

In queste circostanze vedremo che è possibile definire una matrice \mathbf{AB} di tipo (m, p) , che diremo prodotto di \mathbf{A} e \mathbf{B} , con la proprietà che

$$(\mathbf{AB})\mathbf{v} = \mathbf{A}(\mathbf{Bv}) \quad \text{per ogni } \mathbf{v} \in \mathbb{K}^p$$

Vogliamo reinterpretare questa uguaglianza in termini di funzioni. Una matrice \mathbf{A} di tipo (m, n) associa a un vettore $\mathbf{w} \in \mathbb{K}^n$ il vettore $\mathbf{Aw} \in \mathbb{K}^m$; una matrice \mathbf{B} di tipo (n, p) a sua volta associa a un vettore $\mathbf{v} \in \mathbb{K}^p$ il vettore $\mathbf{Bv} \in \mathbb{K}^n$; il prodotto \mathbf{AB} corrisponde alla funzione composta delle due precedenti, perché a un vettore $\mathbf{v} \in \mathbb{K}^p$ associa il vettore $(\mathbf{AB})\mathbf{v} = \mathbf{A}(\mathbf{Bv})$. Interpretare le matrici come funzioni offre un punto di vista per così dire dinamico sulle matrici che è ricco di conseguenze e che svilupperemo sistematicamente a partire dal capitolo sulle applicazioni lineari. Per il momento ci basta motivare la definizione del prodotto di matrici con la corrispondente operazione di composizione di funzioni.

Per vedere perché il prodotto possa essere utile, facciamo un paio di esempi. Un'equazione $ax = b$ si risolve dividendo per a (ammesso che sia diverso da zero) e la soluzione è $x = a^{-1}b$. Anche nel caso di un sistema quadrato $\mathbf{Ax} = \mathbf{b}$ è possibile trovare la soluzione *dividendo* per \mathbf{A} , ammesso questa volta che il rango di \mathbf{A} sia massimo. In questo caso dividere significa moltiplicare a sinistra (il prodotto di matrici non è commutativo) per la *matrice inversa* \mathbf{A}^{-1} , che è la matrice il cui prodotto per \mathbf{A} dà come risultato la matrice identità, e che esiste se e solo se il rango di \mathbf{A} è massimo. Moltiplicando entrambi i membri di $\mathbf{Ax} = \mathbf{b}$ a sinistra per la matrice inversa

\mathbf{A}^{-1} si ottiene la soluzione $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ del sistema. Purtroppo il calcolo della matrice inversa \mathbf{A}^{-1} è computazionalmente molto oneroso; per risolvere il sistema conviene per esempio utilizzare il MEG. Nel paragrafo sulla fattorizzazione **LU** vedremo che anche il MEG si può descrivere in termini di prodotto di matrici, col vantaggio che, se si deve risolvere $\mathbf{Ax} = \mathbf{b}$ per diversi termini noti \mathbf{b} , non occorre ripetere l'eliminazione per ogni \mathbf{b} .

Un altro esempio di applicazioni del prodotto di matrici è allo studio dei sistemi dinamici discreti. Nel caso più semplice, si tratta di un sistema dinamico il cui stato è descritto da n grandezze, quindi da un vettore $\mathbf{x} \in \mathbb{R}^n$ ed evolve a intervalli di tempo discreti secondo una legge lineare indipendente dal tempo. Supponiamo che $\mathbf{x}_0 \in \mathbb{R}^n$ descriva lo stato iniziale del sistema al tempo $t = 0$, e che nell'intervallo dal tempo $t = k-1$ al tempo $t = k$ il sistema passi dallo stato \mathbf{x}_{k-1} allo stato \mathbf{x}_k obbedendo alla legge di evoluzione

$$\mathbf{x}_k = \mathbf{B}\mathbf{x}_{k-1}$$

dove \mathbf{B} è una matrice $n \times n$ indipendente da k . In un'unità di tempo il sistema passa dallo stato \mathbf{x}_0 allo stato $\mathbf{x}_1 = \mathbf{B}\mathbf{x}_0$, in un'ulteriore unità di tempo dallo stato \mathbf{x}_1 allo stato $\mathbf{x}_2 = \mathbf{B}\mathbf{x}_1 = (\mathbf{B}\mathbf{B})\mathbf{x}_0$. Dopo k unità di tempo il sistema si troverà nello stato $\mathbf{x}_k = \mathbf{B}^k\mathbf{x}_0$, dove la potenza \mathbf{B}^k denota il prodotto di k fattori uguali a \mathbf{B} . Quindi le proprietà di \mathbf{B}^k riflettono il modo in cui il sistema evolve nel tempo. Per un esempio concreto, si immagini che la popolazione di una città si divida in due quartieri C e M (oppure gli studenti di una scuola si dividano in due classi coi diabolici insegnanti Caronte e Mefistofele). Lo stato iniziale è $\mathbf{x}_0 = [c_0, m_0]^T$ dove c_0 è la percentuale di abitanti nel primo quartiere, m_0 la percentuale di abitanti del secondo quartiere. Dopo un intervallo di tempo gli abitanti possono cambiare quartiere (o gli studenti docente). Supponiamo che lo facciano secondo una legge fissa nel tempo (che dipende dall'attrattiva dei quartieri o dei diabolici docenti): per esempio che il 40% degli abitanti del primo quartiere decida di rimanere, il resto si trasferisca nel secondo; mentre il 50% degli abitanti del secondo quartiere rimane e il 50% si trasferisce. La legge di evoluzione del sistema è allora $c_k = 2/5c_{k-1} + 1/2m_{k-1}$ (nel quartiere rimangono i 2/5 degli abitanti e in più arriva la metà degli abitanti dell'altro quartiere) e $m_k = 3/5c_{k-1} + 1/2m_{k-1}$. In termini di matrici:

$$\mathbf{x}_k = \mathbf{B}\mathbf{x}_{k-1} \quad \mathbf{B} = \begin{bmatrix} 2/5 & 1/2 \\ 3/5 & 1/2 \end{bmatrix}$$

Calcolando il limite per k che tende a più infinito di $\mathbf{x}_k = \mathbf{B}^k\mathbf{x}_0$ si ottiene (un conto che faremo nel capitolo sugli autovettori e autovalori), non sorprendentemente, lo stato di equilibrio $\mathbf{x}_\infty = [5/11, 6/11]^T$ del sistema:

$$\mathbf{x}_\infty = \mathbf{B}\mathbf{x}_\infty$$

indipendentemente dallo stato iniziale. Questo significa che, se lo stato iniziale è \mathbf{x}_∞ , il numero di abitanti nei due quartieri rimane costante nel tempo (anche se gli abitanti continuano a trasferirsi); ogni altro stato iniziale tende a trasformarsi col tempo nello stato di equilibrio. L'algoritmo con cui Google esegue le proprie ricerche si fonda sullo stesso principio: gli abitanti sono in questo caso navigatori sul web che, invece di cambiare quartiere, passano da una pagina web all'altra.

In analisi numerica, una situazione analoga si presenta per i *metodi iterativi* di soluzione di un sistema lineare. L'idea è quella di costruire una successione di vettori $\mathbf{x}_k \in \mathbb{R}^n$ che al crescere di k approssimi sempre meglio la soluzione esatta di $\mathbf{Ax} = \mathbf{b}$. Tipicamente i vettori \mathbf{x}_k sono definiti ricorsivamente da un vettore iniziale \mathbf{x}_0 per mezzo dell'equazione $\mathbf{x}_k = \mathbf{B}\mathbf{x}_{k-1}$ dove \mathbf{B} è una matrice quadrata di ordine n che si dice matrice di approssimazione. Col crescere di k , il vettore $\mathbf{B}^k\mathbf{x}_0$ approssima sempre meglio la soluzione.

2 SOMMA E PRODOTTO PER UNO SCALARE

Una matrice $m \times n$ può essere pensata come un vettore con $m \times n$ componenti, quindi come un elemento di $\mathbb{K}^{m \times n}$. In particolare, possiamo definire, per le matrici di tipo fissato, le operazioni di somma e prodotto per uno scalare.

Esplicitamente, se \mathbf{A} e \mathbf{B} sono due matrici $m \times n$, la loro somma $\mathbf{A} + \mathbf{B}$ è la matrice $m \times n$ che ha come elemento di posto (i, j) la somma $a_{ij} + b_{ij}$ degli elementi di posto (i, j) delle matrici \mathbf{A} e \mathbf{B} .

Il prodotto di una matrice per uno scalare si effettua moltiplicando gli elementi della matrice per lo scalare: se $\mathbf{A} = [a_{ij}]$, $t\mathbf{A} = [ta_{ij}]$.

Per esempio

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 3 & 5 \\ 7 & 9 \end{bmatrix} \quad \text{e} \quad 7 \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 14 & 21 \\ 28 & 35 \end{bmatrix}$$

Denotiamo col simbolo $\mathbb{M}_{\mathbb{K}}(m, n)$ l'insieme delle matrici di tipo (m, n) a coefficienti nel campo \mathbb{K} . Se consideriamo solo le operazioni di somma e prodotto per uno scalare, $\mathbb{M}_{\mathbb{K}}(m, n)$ può essere identificato con $\mathbb{K}^{m \times n}$. Il ruolo della base canonica è svolto dalle matrici $\{\mathbf{E}_{ij}\}$ $1 \leq i \leq m$, $1 \leq j \leq n$ così definite: \mathbf{E}_{ij} è la matrice di tipo (m, n) che ha 1 come elemento di posto (i, j) e tutti gli altri elementi nulli. Per esempio, la base canonica per lo spazio delle matrici di tipo $(2, 2)$ consiste delle quattro matrici:

$$\mathbf{E}_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{E}_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \mathbf{E}_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \mathbf{E}_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Una matrice generica 2×2 si scrive in modo unico come combinazione lineare delle matrici della base canonica:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + a_{12} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + a_{21} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + a_{22} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Analogamente, nel caso generale $m \times n$, la matrice $\mathbf{A} = [a_{ij}]$ si scrive in modo unico come combinazione lineare delle matrici della base canonica:

$$\mathbf{A} = \sum_{i=1}^m \sum_{j=1}^n a_{ij} \mathbf{E}_{ij}$$

Il coefficiente con cui \mathbf{E}_{ij} compare nella combinazione lineare è l'elemento di posto (i, j) in \mathbf{A} .

3 IL PRODOTTO RIGHE PER COLONNE

Una matrice \mathbf{A} di tipo (m, n) definisce una funzione $\mathcal{L}_{\mathbf{A}} : \mathbb{K}^n \rightarrow \mathbb{K}^m$: a un vettore $\mathbf{v} \in \mathbb{K}^n$ la funzione $\mathcal{L}_{\mathbf{A}}$ associa il vettore $\mathbf{Av} \in \mathbb{K}^m$. La funzione $\mathcal{L}_{\mathbf{A}}$ determina la matrice: infatti, se \mathbf{e}_k è il k -esimo vettore della base canonica di \mathbb{K}^n , il vettore $\mathcal{L}_{\mathbf{A}}(\mathbf{e}_k) = \mathbf{Ae}_k$ è la colonna k di \mathbf{A} . Vorremmo ora definire il prodotto \mathbf{AB} in modo che

$$\mathcal{L}_{\mathbf{AB}} = \mathcal{L}_{\mathbf{A}} \circ \mathcal{L}_{\mathbf{B}}$$

Questo significa

$$(3.1) \quad (\mathbf{AB})\mathbf{v} = \mathbf{A}(\mathbf{B}\mathbf{v}) \quad \text{per ogni } \mathbf{v} \in \mathbb{K}^p$$

Perché questo sia possibile, \mathbf{AB} dev'essere una matrice di tipo (m, p) , e la sua colonna k dev'essere

$$(\mathbf{AB})\mathbf{e}_k = \mathbf{A}(\mathbf{Be}_k)$$

A secondo membro abbiamo il prodotto della matrice \mathbf{A} per la k -esima colonna \mathbf{Be}_k di \mathbf{B} . Quindi il prodotto \mathbf{AB} si definisce come la matrice la cui colonna k è il prodotto della matrice \mathbf{A} per la colonna k di \mathbf{B} . Vedremo più avanti che con questa definizione la (3.1) è verificata non solo per i vettori della base canonica, ma per ogni vettore \mathbf{v} e quindi il prodotto di matrici corrisponde davvero alla composizione delle funzioni associate, come desiderato. Ricordando che \mathbf{Av} è il vettore ottenuto moltiplicando le righe di \mathbf{A} per la colonna \mathbf{v} , possiamo rendere più esplicita la definizione di \mathbf{AB} .

DEFINIZIONE 3.1 (Prodotto di matrici)

Siano \mathbf{A} e \mathbf{B} matrici di tipo (m, n) e (n, p) rispettivamente. Il prodotto *righe per colonne* di \mathbf{A} e \mathbf{B} è la matrice \mathbf{AB} di tipo (m, p) il cui elemento di posto (i, k) è il prodotto della riga i di \mathbf{A} per la colonna k di \mathbf{B} . Se $\mathbf{A} = [a_{ij}]$ e $\mathbf{B} = [b_{jk}]$, l'elemento di posto (i, k) della matrice prodotto è quindi

$$(3.2) \quad (\mathbf{AB})_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk} = \sum_{j=1}^n a_{ij}b_{jk}$$

Si osservi che il prodotto \mathbf{AB} è definito *solo* se *il numero di colonne di \mathbf{A} coincide col numero di righe di \mathbf{B}* : solo in questo caso infatti le righe di \mathbf{A} e le colonne di \mathbf{B} hanno lo stesso numero di componenti e possono perciò essere moltiplicate tra loro. Dalla definizione segue subito che:

Righe e colonne della matrice prodotto

La riga i di \mathbf{AB} è il prodotto della riga i di \mathbf{A} per \mathbf{B}

La colonna k di \mathbf{AB} è il prodotto di \mathbf{A} per la colonna k di \mathbf{B}

Il numero di righe di \mathbf{AB} è uguale al numero di righe di \mathbf{A} e il numero di colonne di \mathbf{AB} è uguale al numero di colonne di \mathbf{B} :

Il prodotto di due matrici di tipo (m, n) e (n, p) è di tipo (m, p) .



Calcoliamo il prodotto delle matrici $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ e $\mathbf{B} = \begin{bmatrix} 5 & 3 \\ 2 & 1 \end{bmatrix}$. Innanzitutto, il prodotto \mathbf{AB} è ben definito perché \mathbf{A} ha 2 colonne e \mathbf{B} ha 2 righe, ed è una matrice di tipo $(2, 2)$ perché \mathbf{A} ha due righe e \mathbf{B} ha due colonne. L'elemento di posto $(1, 1)$ del prodotto si trova moltiplicando la prima riga di \mathbf{A} per la prima colonna di \mathbf{B} :

$$(\mathbf{AB})_{11} = [1 \ 2] \begin{bmatrix} 5 \\ 2 \end{bmatrix} = 5 + 4 = 9$$

L'elemento di posto $(1, 2)$ del prodotto si trova moltiplicando la prima riga di \mathbf{A} per la seconda colonna di \mathbf{B} :

$$(\mathbf{AB})_{12} = [1 \ 2] \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 3 + 2 = 5$$

L'elemento di posto $(2, 1)$ del prodotto si trova moltiplicando la seconda riga di \mathbf{A} per la prima colonna di \mathbf{B} :

$$(\mathbf{AB})_{21} = [3 \ 4] \begin{bmatrix} 5 \\ 2 \end{bmatrix} = 15 + 8 = 23$$

L'elemento di posto $(2, 2)$ del prodotto si trova moltiplicando la seconda riga di \mathbf{A} per la prima colonna di \mathbf{B} :

$$(\mathbf{AB})_{22} = [3 \ 4] \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 9 + 4 = 13$$

La matrice prodotto è perciò:

$$\mathbf{AB} = \begin{bmatrix} 9 & 5 \\ 23 & 13 \end{bmatrix}$$

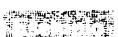
In questo esempio possiamo calcolare anche il prodotto \mathbf{BA} perché il numero di colonne di \mathbf{B} è uguale al numero di righe di \mathbf{A} . Svolgendo i conti otteniamo

$$\mathbf{BA} = \begin{bmatrix} 5 & 3 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 14 & 22 \\ 5 & 8 \end{bmatrix}$$

Quindi \mathbf{AB} è diversa da \mathbf{BA} .

Quando \mathbf{A} e \mathbf{B} sono matrici quadrate dello stesso ordine n , entrambi i prodotti \mathbf{AB} e \mathbf{BA} sono definiti e sono matrici quadrate di ordine n . Come mostra l'esempio precedente, le matrici prodotto \mathbf{AB} e \mathbf{BA} possono essere diverse. Questo non deve sorprendere perché la definizione di prodotto di matrici è asimmetrica rispetto ai fattori. Quindi il prodotto di matrici *dipende dall'ordine dei fattori*: si dice che

Il prodotto di matrici non è commutativo.



Per due matrici \mathbf{A} e \mathbf{B} di tipo (m, n) ed (n, m) , entrambi i prodotti \mathbf{AB} e \mathbf{BA} sono definiti, ma \mathbf{AB} è una matrice $m \times m$, mentre \mathbf{BA} è una matrice $n \times n$. Per un esempio concreto, consideriamo il vettore riga $\mathbf{A} = [2 \ 3]$ e il vettore colonna $\mathbf{B} = [7 \ 5]$. Si tratta di matrici 1×2 e 2×1 rispettivamente. Quindi i prodotti \mathbf{AB} e \mathbf{BA} sono entrambi ben definiti, ma \mathbf{AB} è una matrice di tipo $(1, 1)$, cioè un numero, mentre \mathbf{BA} è una matrice di tipo $(2, 2)$:

$$\mathbf{AB} = [2 \ 3] \begin{bmatrix} 7 \\ 5 \end{bmatrix} = 14 + 15 = 29, \quad \mathbf{BA} = \begin{bmatrix} 7 \\ 5 \end{bmatrix} [2 \ 3] = \begin{bmatrix} 7 \times 2 & 7 \times 3 \\ 5 \times 2 & 5 \times 3 \end{bmatrix} = \begin{bmatrix} 14 & 21 \\ 10 & 15 \end{bmatrix}$$

Elenchiamo ora le principali proprietà del prodotto di matrici: il prodotto di matrici è associativo e bilineare. Questo significa:

- **Proprietà associativa:**

date tre matrici \mathbf{A} , \mathbf{B} , \mathbf{C} di tipo (m, n) , (n, p) e (p, q) rispettivamente, vale l'uguaglianza:

$$(3.3) \quad (\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$$

- **Proprietà distributiva a sinistra:**

date due matrici \mathbf{A}_1 e \mathbf{A}_2 di tipo (m, n) e una matrice \mathbf{B} di tipo (n, p) , vale l'uguaglianza:

$$(3.4) \quad (\mathbf{A}_1 + \mathbf{A}_2)\mathbf{B} = \mathbf{A}_1\mathbf{B} + \mathbf{A}_2\mathbf{B}$$

- **Proprietà distributiva a destra:**

date una matrice \mathbf{A} di tipo (m, n) e due matrici \mathbf{B}_1 e \mathbf{B}_2 di tipo (n, p) , vale l'uguaglianza:

$$(3.5) \quad \mathbf{A}(\mathbf{B}_1 + \mathbf{B}_2) = \mathbf{AB}_1 + \mathbf{AB}_2$$

- **Omogeneità**

Per ogni scalare $t \in \mathbb{K}$ e ogni coppia di matrici \mathbf{A} e \mathbf{B} di tipo (m, n) e (n, p) , valgono le uguaglianze:

$$(3.6) \quad t(\mathbf{AB}) = (t\mathbf{A})\mathbf{B} = \mathbf{A}(t\mathbf{B})$$

Si noti che le proprietà distributive sono due e una non segue dall'altra, perché il prodotto di matrici non è commutativo.

DIMOSTRAZIONE. Se $\mathbf{A} = [a_{ij}]$, $\mathbf{B} = [b_{jk}]$, $\mathbf{C} = [c_{kl}]$, allora

$$\begin{aligned} ((\mathbf{AB})\mathbf{C})_{il} &= \sum_k (\mathbf{AB})_{ik} c_{kl} = \sum_k (\sum_j a_{ij} b_{jk}) c_{kl} = \\ &= \sum_k \sum_j a_{ij} b_{jk} c_{kl} = \sum_j a_{ij} (\sum_k b_{jk} c_{kl}) = (\mathbf{A}(\mathbf{BC}))_{il} \end{aligned}$$

Questo dimostra la proprietà associativa $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$. Come nel caso dei numeri, tale proprietà ci consente di scrivere \mathbf{ABC} al posto di $\mathbf{A}(\mathbf{BC})$ e di $(\mathbf{AB})\mathbf{C}$: il risultato non dipende da quale prodotto eseguiamo prima. Si noti che, nel caso particolare in cui $\mathbf{C} = \mathbf{v}$ sia un vettore colonna, otteniamo una dimostrazione della formula (3.1): il prodotto \mathbf{AB} corrisponde davvero alla funzione composta $\mathcal{L}_{\mathbf{A}} \circ \mathcal{L}_{\mathbf{B}}$ come desiderato.

La dimostrazione delle altre proprietà è più semplice e viene lasciata al lettore.

Matrice nulla e matrice identità

La *matrice nulla* di tipo (m, n) è la matrice $m \times n$ che ha tutti gli elementi uguali a zero. La denoteremo col simbolo \mathbf{O} . Ha proprietà analoghe a quelle di 0 tra i numeri: è elemento neutro per la somma e annulla tutti i prodotti. Questo significa: sommando \mathbf{O} a un'altra matrice \mathbf{A} di tipo (m, n) si ottiene la matrice \mathbf{A} ; moltiplicando \mathbf{O} a destra per una matrice $n \times p$ (risp. a sinistra per una matrice $l \times m$) si ottiene la matrice nulla $m \times p$ (risp. $l \times n$).

Il prodotto tra numeri ha una unità, il numero uno, che moltiplicato per ogni numero x dà come risultato il numero x . Nel caso delle matrici un ruolo analogo è giocato dalla matrice identità. La *matrice identità* \mathbf{I} (o \mathbf{I}_n) di ordine n è la matrice $n \times n$ che ha gli elementi sulla diagonale principale tutti uguali a 1 e gli altri elementi uguali a zero: l'elemento di posto (i, j) nella matrice identità è 1 se $i = j$ ed è 0 se $i \neq j$. Quindi $\mathbf{I} = [\delta_{ij}]$, dove δ_{ij} è il simbolo di Kronecker. Per esempio:

$$\mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{I}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Le colonne della matrice identità \mathbf{I}_n di ordine n sono i vettori della base canonica di \mathbb{K}^n :

$$\mathbf{I}_n = [\mathbf{e}_1 \ \mathbf{e}_2 \ \cdots \ \mathbf{e}_n]$$

Analogamente, le righe di \mathbf{I}_n sono $\mathbf{e}_1^T, \dots, \mathbf{e}_n^T$.

La matrice identità è l'elemento neutro rispetto al prodotto di matrici:

Se $\mathbf{I} = \mathbf{I}_n$ è la matrice identità di ordine n , allora

$$\mathbf{AI} = \mathbf{A} \quad \text{per ogni matrice } \mathbf{A} \text{ con } n \text{ colonne;}$$

$$\mathbf{IB} = \mathbf{B} \quad \text{per ogni matrice } \mathbf{B} \text{ con } n \text{ righe.}$$

DIMOSTRAZIONE. Verifichiamo solo la prima formula. Scriviamo $\mathbf{A} = [a_{ij}]$ e verifichiamo che l'elemento di posto (i, k) nella matrice prodotto \mathbf{AI} coincida con a_{ik} :

$$(\mathbf{AI})_{ik} = \sum_{j=1}^n a_{ij} \delta_{jk} = a_{ik}$$

125

Matrici diagonali

Le matrici diagonali sono le più semplici matrici quadrate:

DEFINIZIONE 3.2 Una matrice *diagonale* è una matrice quadrata $\mathbf{D} = [d_{ij}]$ che ha tutti gli elementi al di fuori dalla diagonale principale nulli:

$$d_{ij} = 0 \quad \text{se } i \neq j$$

Denotiamo col simbolo $\text{diag}(\lambda_1, \dots, \lambda_n)$ la matrice diagonale in cui $d_{ii} = \lambda_i$:

$$\text{diag}(\lambda_1, \dots, \lambda_n) = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

Per esempio

$$\text{diag}(7, 5, -3) = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -3 \end{bmatrix}$$

La matrice nulla e la matrice identità sono diagonali; più in generale, tutti i multipli scalari della matrice identità sono matrici diagonali:

$$\lambda \mathbf{I} = \text{diag}(\lambda, \dots, \lambda)$$

(per $\lambda = 0$ si tratta della matrice nulla, per $\lambda = 1$ della matrice identità).

Prodotto di matrici diagonali

È immediato verificare che il prodotto di due matrici diagonali è ancora diagonale, e più precisamente:

$$\text{diag}(\lambda_1, \dots, \lambda_n) \text{diag}(\mu_1, \dots, \mu_n) = \text{diag}(\lambda_1\mu_1, \dots, \lambda_n\mu_n)$$

Potenze di una matrice quadrata

Due matrici quadrate \mathbf{A} e \mathbf{B} di ordine n hanno lo stesso numero di righe e di colonne e quindi sono ben definiti i prodotti \mathbf{AB} e \mathbf{BA} , che sono ancora matrici quadrate di ordine n . In particolare, possiamo moltiplicare una matrice quadrata con se stessa e definire

$$\mathbf{A}^2 = \mathbf{AA}$$

Siccome \mathbf{A}^2 è ancora una matrice quadrata di ordine n , possiamo moltiplicarla per \mathbf{A} e definire $\mathbf{A}^3 = \mathbf{A}^2\mathbf{A}$; per la proprietà associativa $\mathbf{A}^3 = \mathbf{AA}^2$. Quindi \mathbf{A}^3 è il prodotto di tre fattori tutti uguali ad \mathbf{A} , non importa in che ordine grazie alla proprietà associativa.

Più in generale, per ogni intero $k \geq 2$ possiamo definire la *potenza k -esima* \mathbf{A}^k di una matrice quadrata \mathbf{A} come il prodotto di \mathbf{A}^{k-1} per \mathbf{A} o, equivalentemente, come il prodotto di k fattori tutti uguali ad \mathbf{A} ; di nuovo è irrilevante l'ordine in cui svolgiamo il prodotto grazie alla proprietà associativa:

$$\mathbf{A}^k = \mathbf{A}^{k-1}\mathbf{A} = \mathbf{AA} \cdots \mathbf{A} \quad (k \text{ fattori tutti uguali ad } \mathbf{A})$$

Come per i numeri, si pone $\mathbf{A}^0 = \mathbf{I}$ e $\mathbf{A}^1 = \mathbf{A}$.



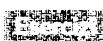
- Per la matrice identità $\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ si ha $\mathbf{I}^k = \mathbf{I}$ per ogni $k \geq 1$.
- Anche per la matrice nulla $\mathbf{O} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ si ha $\mathbf{O}^k = \mathbf{O}$ per ogni $k \geq 1$.
- Più in generale, per una matrice diagonale si ha:

$$(\text{diag}(\lambda_1, \dots, \lambda_n))^k = \text{diag}(\lambda_1^k, \dots, \lambda_n^k)$$

- Per la matrice $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ si ha $\mathbf{A}^k = \mathbf{A}$ per ogni $k \geq 1$: basta verificare che $\mathbf{A}^2 = \mathbf{A}$, e questo è immediato.
- La matrice $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ non è nulla, ma $\mathbf{A}^2 = \mathbf{O}$ è la matrice nulla. Quindi $\mathbf{A}^k = \mathbf{O}$ per ogni $k \geq 2$.
- La matrice

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

soddisfa $\mathbf{A}^2 \neq \mathbf{O}$ e $\mathbf{A}^k = \mathbf{O}$ per ogni $k \geq 3$.



Esercizio 1 Moltiplicare la matrice $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix}$ a destra per i vettori \mathbf{e}_1 , \mathbf{e}_2 e \mathbf{e}_3 e verificare che così si ottengono le colonne di \mathbf{A} . Moltiplicare poi \mathbf{A} a sinistra per \mathbf{e}_1^T , \mathbf{e}_2^T e \mathbf{e}_3^T e verificare che così si ottengono le righe di \mathbf{A} .

Esercizio 2 Calcolare i prodotti indicati, dopo aver controllato che si possano effettuare e aver scritto il numero di righe e di colonne che la matrice prodotto deve avere in ciascun caso:

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

Esercizio 3 Calcolare le potenze \mathbf{A}^k della matrice $\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$
(Risposta: $\mathbf{A}^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ e $\mathbf{A}^k = \mathbf{0}$ per $k \geq 3$).

Esercizio 4 Sia $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ Determinare tutte le matrici $\mathbf{B} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ tali che $\mathbf{AB} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Suggerimento: le colonne di \mathbf{B} appartengono al nucleo di \mathbf{A} . Risposta: $\mathbf{B} = \begin{bmatrix} 2x & 2y \\ -x & -y \end{bmatrix}$ con $x, y \in \mathbb{K}$.

Esercizio 5 Per il prodotto di matrici *non vale la legge di annullamento*: il prodotto di due matrici può essere la matrice nulla senza che alcuno dei fattori sia nullo. Fornire dei controesempi: a) trovare una matrice quadrata \mathbf{A} di ordine 2 tale che $\mathbf{A}^2 = \mathbf{O}$, b) trovare due distinte matrici quadrate \mathbf{A} e \mathbf{B} di ordine due tali che $\mathbf{AB} = \mathbf{O}$.

Esercizio 6 Per il prodotto di matrici *non vale la legge di cancellazione*: da $\mathbf{AB} = \mathbf{AC}$ non segue $\mathbf{B} = \mathbf{C}$. Trovare tre matrici quadrate \mathbf{A} , \mathbf{B} e \mathbf{C} di ordine due tali che $\mathbf{B} \neq \mathbf{C}$, ma $\mathbf{AB} = \mathbf{AC}$. Mostrare però che la legge di cancellazione vale se $\text{Ker}(\mathbf{A})$ contiene solo il vettore nullo.

Sia \mathbf{A} una matrice quadrata di ordine n . Spiegare perché $\mathbf{A}^4 = \mathbf{A}\mathbf{A}^2\mathbf{A}$.

Sia \mathbf{A} una matrice quadrata di ordine n . Mostrare che $\mathbf{A}^k\mathbf{A}^l = \mathbf{A}^{k+l}$ per ogni $k, l \geq 0$.

Dimostrare che per il prodotto di due matrici diagonali vale la formula

$$\text{diag}(\lambda_1, \dots, \lambda_n) \text{ diag}(\mu_1, \dots, \mu_n) = \text{diag}(\lambda_1\mu_1, \dots, \lambda_n\mu_n)$$

Concludere che il prodotto tra matrici diagonali è commutativo.

Sia $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$. Determinare tutte le matrici \mathbf{A} di tipo $(2, 2)$ che commutano con \mathbf{B} : $\mathbf{AB} = \mathbf{BA}$.

Suggerimento: occorre risolvere un sistema lineare che ha come incognite i 4 elementi della matrice \mathbf{A} . Risposta: $\mathbf{A} = x\mathbf{I} + y\mathbf{B}$ con $x, y \in \mathbb{K}$.

Trovare due matrici \mathbf{A} e \mathbf{D} di tipo $(2, 2)$, con \mathbf{D} diagonale, tali che $\mathbf{DA} \neq \mathbf{AD}$.

Sia \mathbf{I} la matrice identità di ordine n . Mostrare che

$$(\lambda\mathbf{I})\mathbf{B} = \mathbf{B}(\lambda\mathbf{I})$$

per ogni matrice quadrata \mathbf{B} di ordine n .

Sia \mathbf{A} una matrice quadrata di ordine 2 che commuta con tutte le matrici quadrate di ordine 2:

$$\mathbf{AB} = \mathbf{BA} \quad \text{per ogni matrice } \mathbf{B} \text{ quadrata di ordine 2}$$

Dimostrare che \mathbf{A} è un multiplo scalare dell'identità: esiste uno scalare $\lambda \in \mathbb{K}$ tale che $\mathbf{A} = \lambda\mathbf{I}$. Dimostrare l'enunciato analogo per le matrici quadrate di ordine n .

Mostrare che il prodotto di due matrici triangolari alte è una matrice triangolare alta.

(Radici quadrate di $-\mathbf{I}$)

- Mostrare che $\begin{bmatrix} -2 & 1 \\ -5 & 2 \end{bmatrix}^2 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} = -\mathbf{I}$. Più in generale $\begin{bmatrix} a & 1 \\ -1-a^2 & -a \end{bmatrix}^2 = -\mathbf{I}$.
- Mostrare che ci sono infinite matrici quadrate \mathbf{A} di ordine 2 a coefficienti reali (perfino interi) tali che $\mathbf{A}^2 = -\mathbf{I}$, e che tutte soddisfano $a_{22} = -a_{11}$.
- Mostrare che esistono quattro matrici diagonali $\text{diag}(\lambda, \mu)$ a coefficienti complessi (e nessuna a coefficienti reali) il cui quadrato è $-\mathbf{I}$.

4 MATRICI INVERTIBILI

In questo paragrafo assumeremo che tutte le matrici siano quadrate dello stesso ordine n . Ci chiediamo se è possibile definire l'inversa di una matrice \mathbf{A} . L'inverso a^{-1} di un numero a è l'unico numero b tale che

$$ab = 1$$

Si noti che l'inverso di un numero a esiste se e solo se $a \neq 0$.

Vogliamo definire l'inversa di una matrice sostituendo a 1 la matrice identità. Dobbiamo però tenere conto del fatto che il prodotto di matrici non è commutativo. Quindi poniamo:

DEFINIZIONE 4.1 Siano \mathbf{A} e \mathbf{B} due matrici quadrate di ordine n . Si dice che \mathbf{B} è una *matrice inversa* di \mathbf{A} se

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}$$

Si dice che \mathbf{A} è *invertibile* se esiste una matrice inversa di \mathbf{A} .

Si osservi che la definizione è simmetrica in \mathbf{A} e \mathbf{B} : se \mathbf{B} è una *matrice inversa* di \mathbf{A} , allora \mathbf{A} è una *matrice inversa* di \mathbf{B} .

Più in generale, una *inversa sinistra* di una matrice \mathbf{A} di tipo (m, n) è una matrice \mathbf{B} di tipo (n, m) tale che $\mathbf{BA} = \mathbf{I}_n$; un'*inversa destra* di \mathbf{A} è una matrice \mathbf{B} di tipo (n, m) tale che $\mathbf{AB} = \mathbf{I}_m$. Se $m \neq n$, le nozioni di inversa destra e sinistra non coincidono: si vedano gli esercizi alla fine del paragrafo. Ma nel caso di matrici quadrate vedremo che la situazione è semplice: se \mathbf{B} è un'inversa destra (o sinistra) di una matrice quadrata \mathbf{A} , allora automaticamente \mathbf{B} è un'inversa di \mathbf{A} e, inoltre, l'inversa, se esiste, è unica. Cominciamo a mostrare l'unicità.

PROPOSIZIONE 4.2 Sia \mathbf{A} una matrice quadrata.

- a) Se \mathbf{B} e \mathbf{C} sono, rispettivamente, un'inversa destra e un'inversa sinistra di \mathbf{A} , allora $\mathbf{B} = \mathbf{C}$ e, quindi, \mathbf{B} è un'inversa di \mathbf{A} .
- b) Se \mathbf{A} è invertibile, \mathbf{A} ha un'unica matrice inversa che si denota col simbolo \mathbf{A}^{-1} .

DIMOSTRAZIONE. L'ipotesi in a) significa che $\mathbf{AB} = \mathbf{I} = \mathbf{CA}$. La tesi $\mathbf{B} = \mathbf{C}$ segue dalla proprietà associativa del prodotto e dal fatto che la matrice identità è l'elemento neutro per il prodotto di matrici:

$$\mathbf{B} = \mathbf{IB} = (\mathbf{CA})\mathbf{B} = \mathbf{C}(\mathbf{AB}) = \mathbf{CI} = \mathbf{C}$$

Se \mathbf{B} e \mathbf{C} sono due matrici inverse di \mathbf{A} , allora \mathbf{B} è un'inversa destra, \mathbf{C} un'inversa sinistra, e quindi per quanto abbiamo appena mostrato $\mathbf{B} = \mathbf{C}$.

Avendo stabilito l'unicità dell'inversa, ci chiediamo ora quando esista. Non ci possiamo aspettare che esista sempre: già nel caso scalare (matrici 1×1), un numero a è invertibile in \mathbb{K} se e solo se è *diverso da zero*. L'analogo enunciato per le matrici quadrate è che una matrice quadrata di ordine n è invertibile se e solo se *ha rango n* (il massimo possibile).

TEOREMA 4.3 (Condizioni di invertibilità)

Per una matrice \mathbf{A} quadrata di ordine n le seguenti condizioni sono equivalenti:

- a) \mathbf{A} ha rango massimo: $r(\mathbf{A}) = n$;
- b) l'unica soluzione di $\mathbf{A}\mathbf{x} = \mathbf{0}$ è $\mathbf{x} = \mathbf{0}$: $\text{Ker}(\mathbf{A}) = \{\mathbf{0}\}$;
- c) \mathbf{A} è invertibile;
- d) \mathbf{A} ha un'inversa sinistra;
- e) \mathbf{A} ha un'inversa destra.

DIMOSTRAZIONE. Mostriamo l'equivalenza delle prime quattro condizioni mediante la catena di implicazioni

$$d) \Rightarrow b) \Rightarrow a) \Rightarrow c) \Rightarrow d)$$

Proviamo che d) implica b). La condizione d) significa che esiste una matrice \mathbf{C} tale che $\mathbf{CA} = \mathbf{I}$. Supponiamo che $\mathbf{Ax} = \mathbf{0}$. Moltiplichiamo entrambi i membri di questa uguaglianza a sinistra per \mathbf{C} : otteniamo

$$\mathbf{C}(\mathbf{Ax}) = \mathbf{C}\mathbf{0} = \mathbf{0}$$

D'altra parte

$$\mathbf{C}(\mathbf{Ax}) = (\mathbf{CA})\mathbf{x} = \mathbf{Ix} = \mathbf{x}$$

Confrontando le due uguaglianze ottenute troviamo $\mathbf{x} = \mathbf{0}$. Questo mostra che l'unico vettore del nucleo di \mathbf{A} è $\mathbf{x} = \mathbf{0}$, quindi d) implica b).

Proviamo che b) implica a). Se $\mathbf{Ax} = \mathbf{0}$ ha un'unica soluzione $\mathbf{x} = \mathbf{0}$, per il teorema di Rouché-Capelli $r(\mathbf{A}) = n$ (non ci possono essere variabili libere). Quindi b) implica a).

Mostriamo ora che a) implica c): supponiamo $r(\mathbf{A}) = n$ e proviamo che esiste una matrice inversa di \mathbf{A} : questo è il punto centrale della dimostrazione. Per il teorema di Cramer, dato che \mathbf{A} è quadrata di ordine n e $r(\mathbf{A}) = n$, per ogni $\mathbf{b} \in \mathbb{K}^n$ esiste un'unica soluzione del sistema $\mathbf{Ax} = \mathbf{b}$. Risolvendo i sistemi $\mathbf{Ax} = \mathbf{e}_k$ che hanno come termine noto le colonne $\mathbf{e}_1, \dots, \mathbf{e}_n$ della matrice identità \mathbf{I} troviamo vettori $\mathbf{v}_1, \dots, \mathbf{v}_n$ tali che

$$\mathbf{Av}_1 = \mathbf{e}_1, \quad \mathbf{Av}_2 = \mathbf{e}_2, \dots, \quad \mathbf{Av}_n = \mathbf{e}_n$$

Sia \mathbf{B} la matrice che ha per colonne i vettori $\mathbf{v}_1, \dots, \mathbf{v}_n$. Allora

$$\mathbf{AB} = \mathbf{A}[\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n] = [\mathbf{e}_1 \quad \mathbf{e}_2 \quad \dots \quad \mathbf{e}_n] = \mathbf{I}$$

Quindi \mathbf{A} è un'inversa sinistra di \mathbf{B} . Abbiamo già mostrato d) implica b), e che b) è equivalente ad a). Quindi \mathbf{B} , che ha un'inversa sinistra, ha rango massimo: $r(\mathbf{B}) = n$. Ma allora possiamo ripetere il ragionamento con \mathbf{B} al posto di \mathbf{A} e concludere che esiste una matrice \mathbf{C} tale che $\mathbf{BC} = \mathbf{I}$. Quindi \mathbf{B} ha un'inversa sinistra \mathbf{A} e un'inversa destra \mathbf{C} . Per la proposizione 4.2 \mathbf{A} è un'inversa di \mathbf{B} , e questo è equivalente a dire che \mathbf{B} è un'inversa di \mathbf{A} . Questo conclude la dimostrazione che a) implica c). Ovviamente c) implica d), e quindi le prime quattro condizioni sono equivalenti.

Mostriamo infine l'equivalenza di c) ed e). Ovviamente c) implica e). Viceversa, supponiamo che \mathbf{A} abbia un'inversa destra \mathbf{B} . Allora \mathbf{B} ha un'inversa sinistra, e quindi per quanto abbiamo appena visto \mathbf{B} è invertibile con inversa \mathbf{A} , ovverosia \mathbf{A} è invertibile con inversa \mathbf{B} .

DEFINIZIONE 4.4 (Matrici nonsingolari)

Una matrice che soddisfi le condizioni equivalenti del teorema 4.3 si dice *nonsingolare*. Quindi nonsingolare è sinonimo di invertibile, e *singolare* significa non invertibile.

Dal teorema segue immediatamente:

COROLLARIO 4.5 Se \mathbf{A} e \mathbf{B} sono due matrici quadrate dello stesso ordine e $\mathbf{AB} = \mathbf{I}$, allora \mathbf{A} e \mathbf{B} sono invertibili, $\mathbf{B} = \mathbf{A}^{-1}$ e $\mathbf{A} = \mathbf{B}^{-1}$.

Il corollario afferma che da $\mathbf{AB} = \mathbf{I}$ segue $\mathbf{BA} = \mathbf{I}$: per concludere che \mathbf{B} è un'inversa basta verificare che sia un'inversa destra o un'inversa sinistra.

Nel caso di matrici di ordine 2 il calcolo dell'inversa è semplice e vale la pena conoscere la formula esplicita per l'inversa.

Inversa di una matrice 2×2

Una matrice $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ è invertibile se e solo se $ad - bc \neq 0$, e in tal caso

$$(4.1) \quad \mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

DIMOSTRAZIONE. La matrice \mathbf{A} è invertibile se e solo se ha rango 2. Abbiamo già visto nel capitolo sui sistemi lineari che questo succede se e solo se il determinante $ad - bc$ è diverso da zero. Per trovare una formula esplicita per l'inversa, procediamo come nella dimostrazione del teorema 4.3: le colonne di \mathbf{A}^{-1} soddisfano le due uguaglianze

$$\mathbf{Av}_1 = \mathbf{e}_1 \quad \mathbf{Av}_2 = \mathbf{e}_2$$

Dalla formula di Cramer segue

$$\mathbf{v}_1 = \begin{bmatrix} \frac{d}{ad - bc} \\ \frac{-c}{ad - bc} \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} \frac{-b}{ad - bc} \\ \frac{-a}{ad - bc} \end{bmatrix}$$

e quindi

$$\mathbf{A}^{-1} = \begin{bmatrix} \frac{d}{ad - bc} & \frac{-b}{ad - bc} \\ \frac{-c}{ad - bc} & \frac{-a}{ad - bc} \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Diamo ora una formula per l'inversa del prodotto di due matrici invertibili:

PROPOSIZIONE 4.6 (Inversa di un prodotto)

Siano \mathbf{A} e \mathbf{B} due matrici quadrate di ordine n . La matrice prodotto \mathbf{AB} è invertibile se e solo se \mathbf{A} e \mathbf{B} sono invertibili. In tal caso

$$(4.2) \quad (\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

DIMOSTRAZIONE. Se \mathbf{A} e \mathbf{B} sono invertibili, allora

$$(\mathbf{B}^{-1}\mathbf{A}^{-1})(\mathbf{AB}) = \mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\mathbf{B} = \mathbf{B}^{-1}\mathbf{I}\mathbf{B} = \mathbf{B}^{-1}\mathbf{B} = \mathbf{I}$$

Quindi \mathbf{AB} è invertibile con matrice inversa $\mathbf{B}^{-1}\mathbf{A}^{-1}$.

Viceversa, supponiamo che il prodotto \mathbf{AB} sia invertibile. Allora

$$((\mathbf{AB})^{-1}\mathbf{A})\mathbf{B} = (\mathbf{AB})^{-1}(\mathbf{AB}) = \mathbf{I}$$

Questo mostra che \mathbf{B} ha un'inversa sinistra e, quindi, è invertibile per il teorema 4.3; inoltre

$$\mathbf{A} = \mathbf{A}(\mathbf{BB}^{-1}) = (\mathbf{AB})\mathbf{B}^{-1}$$

è invertibile in quanto prodotto delle due matrici invertibili \mathbf{AB} e \mathbf{B}^{-1} .

La prima applicazione della matrice inversa è alla soluzione di sistemi lineari quadrati. Come l'esistenza dell'inverso a^{-1} consente di risolvere l'equazione $ax = b$ rispetto all'incognita x , così l'esistenza della matrice inversa \mathbf{A}^{-1} consente in linea di principio di risolvere il sistema $\mathbf{Ax} = \mathbf{b}$. In linea di principio soltanto, perché dal punto di vista computazionale il calcolo dell'inversa risulta più oneroso della soluzione del sistema. In ogni caso otteniamo una versione più esplicita del teorema di Cramer:

PROPOSIZIONE 4.7 Sia \mathbf{A} una matrice quadrata di ordine n di rango massimo. Allora l'unica soluzione del sistema lineare $\mathbf{Ax} = \mathbf{b}$ è $\mathbf{v} = \mathbf{A}^{-1}\mathbf{b}$.

DIMOSTRAZIONE. La verifica è immediata: infatti se \mathbf{v} è soluzione di $\mathbf{Ax} = \mathbf{b}$, cioè $\mathbf{Av} = \mathbf{b}$, allora

$$\mathbf{v} = \mathbf{I}\mathbf{v} = (\mathbf{A}^{-1}\mathbf{A})\mathbf{v} = \mathbf{A}^{-1}(\mathbf{Av}) = \mathbf{A}^{-1}\mathbf{b}$$

Viceversa, se $\mathbf{v} = \mathbf{A}^{-1}\mathbf{b}$,

$$\mathbf{Av} = \mathbf{A}(\mathbf{A}^{-1}\mathbf{b}) = (\mathbf{AA}^{-1})\mathbf{b} = \mathbf{I}\mathbf{b} = \mathbf{b}$$

Quindi $\mathbf{A}^{-1}\mathbf{b}$ è l'unica soluzione di $\mathbf{Ax} = \mathbf{b}$. Si noti che abbiamo usato entrambe le uguaglianze $\mathbf{AA}^{-1} = \mathbf{I}$ e $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$.



Se \mathbf{A} è invertibile, si mostri che valgono le leggi di cancellazione:

$$\mathbf{AB} = \mathbf{AC} \Rightarrow \mathbf{B} = \mathbf{C} \quad \text{e} \quad \mathbf{BA} = \mathbf{CA} \Rightarrow \mathbf{B} = \mathbf{C}$$

Per una matrice invertibile \mathbf{A} si definiscono le potenze con esponente negativo nel modo usuale:

$$\mathbf{A}^{-k} = (\mathbf{A}^{-1})^k \quad \text{per ogni intero } k \geq 1$$

Mostrare che $\mathbf{A}^h \mathbf{A}^k = \mathbf{A}^{h+k}$ per ogni $h, k \in \mathbb{Z}$. In particolare, l'inversa di \mathbf{A}^k è \mathbf{A}^{-k} .

Mostrare che una matrice diagonale $\text{diag}(\lambda_1, \dots, \lambda_n)$ è invertibile se e solo se gli elementi λ_i sono tutti diversi da zero. In tal caso l'inversa è la matrice diagonale $\text{diag}(\lambda_1^{-1}, \dots, \lambda_n^{-1})$.

Una matrice triangolare alta è invertibile se e solo se tutti i suoi elementi sulla diagonale principale sono diversi da zero. Se questo è il caso, anche l'inversa è una matrice triangolare alta.

Se \mathbf{A} è una matrice di tipo (m, n) , non necessariamente quadrata, si mostri imitando la dimostrazione del teorema 4.3 che \mathbf{A} ammette un'inversa destra se e solo se $r(\mathbf{A}) = m$ (è vero anche che \mathbf{A} ammette un'inversa sinistra se e solo se $r(\mathbf{A}) = n$, ma per dimostrare questo occorre un po' più di teoria).

Per ogni $k \in \mathbb{K}$ risolvere il sistema

$$\begin{cases} 2x + 3y = 12 \\ 3x + 2y = k \end{cases}$$

usando la formula esplicita per l'inversa di una matrice 2×2 .

5 MATRICE TRASPOSTA. MATRICI SIMMETRICHE

Un'ulteriore operazione sulle matrici è quella di *trasposizione* che scambia le righe con le colonne:

DEFINIZIONE 5.1 La *trasposta* di una matrice \mathbf{A} è la matrice \mathbf{A}^T che ha per righe le colonne di \mathbf{A} : l'elemento di posto (i, j) di \mathbf{A} è l'elemento di posto (j, i) di \mathbf{A}^T .

Altri simboli di uso comune per denotare la matrice trasposta sono ${}^t\mathbf{A}$, \mathbf{A}^t e \mathbf{A}_T .



Se \mathbf{v} è un vettore colonna, \mathbf{v}^T è il vettore riga che ha le stesse componenti di \mathbf{v} . Per esempio:

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}^T = [1 \ 2 \ 3]$$

La trasposta della matrice

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

è la matrice

$$\mathbf{A}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

Le principali proprietà della matrice trasposta sono:

PROPOSIZIONE 5.2 (Proprietà della matrice trasposta)

a) Se A e B sono matrici $m \times n$ e $t \in \mathbb{K}$, allora $(A + B)^T = A^T + B^T$ e $(tA)^T = tA^T$

b) Per ogni matrice \mathbf{A} :

$$(\mathbf{A}^T)^T = \mathbf{A}$$

c) Supponiamo che \mathbf{AB} sia definito. Allora anche $\mathbf{B}^T \mathbf{A}^T$ è definito e

$$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$$

d) Sia \mathbf{A} una matrice quadrata invertibile. Allora anche \mathbf{A}^T è invertibile, e

$$(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$$

DIMOSTRAZIONE. La verifica di a) e b) è lasciata al lettore. Il prodotto \mathbf{AB} è definito se il numero di colonne di \mathbf{A} è uguale al numero di righe di \mathbf{B} : questo equivale a dire che il numero di righe di \mathbf{A}^T è uguale al numero di colonne di \mathbf{B}^T , cioè che il prodotto $\mathbf{B}^T \mathbf{A}^T$ è definito. Inoltre $(\mathbf{AB})^T$ ha lo stesso numero di righe e di colonne $\mathbf{B}^T \mathbf{A}^T$. Verifichiamo che le due matrici coincidono: se $\mathbf{A} = [a_{ij}]$ e $\mathbf{B} = [b_{jk}]$, allora

$$\begin{aligned} (\mathbf{AB})_{ki}^T &= (\mathbf{AB})_{ik} = \sum_j a_{ij} b_{jk} \\ (\mathbf{B}^T \mathbf{A}^T)_{ki} &= \sum_j \mathbf{B}_{kj}^T \mathbf{A}_{ji}^T = \sum_j b_{jk} a_{ij} \end{aligned}$$

e le due espressioni sono uguali perché il prodotto di due scalari è commutativo.

Infine mostriamo che $(\mathbf{A}^{-1})^T$ è l'inversa di \mathbf{A}^T . Basta verificare che sia un'inversa sinistra. Questo segue dal punto c):

$$(\mathbf{A}^{-1})^T \mathbf{A}^T = (\mathbf{A} \mathbf{A}^{-1})^T = \mathbf{I}^T = \mathbf{I}.$$

DEFINIZIONE 5.3 Una matrice \mathbf{A} si dice *simmetrica* se $\mathbf{A}^T = \mathbf{A}$. Una matrice \mathbf{A} si dice *antisimmetrica* o *emisimmetrica* se $\mathbf{A}^T = -\mathbf{A}$.

OSSERVAZIONE Possiamo riformulare la definizione in questo modo: una matrice $\mathbf{A} = [a_{ij}]$ è simmetrica se e solo se è quadrata, diciamo di ordine n , e $a_{ij} = a_{ji}$ per

ogni $i, j = 1, 2, \dots, n$. La simmetria è rispetto alla diagonale principale: scambiando gli indici di riga e di colonna gli elementi non cambiano. Analogamente, una matrice $\mathbf{A} = [a_{ij}]$ è antisimmetrica se e solo se è quadrata, diciamo di ordine n , e $a_{ij} = -a_{ji}$ per ogni $i, j = 1, 2, \dots, n$.



- Ogni matrice diagonale è simmetrica.
- Tutti gli elementi sulla diagonale principale di una matrice antisimmetrica sono nulli perché $a_{ii} = -a_{ii}$ implica $a_{ii} = 0$.
- Una matrice quadrata di ordine 2 è simmetrica se e solo se ha la forma:

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

Quindi le matrici simmetriche di ordine 2 dipendono da 3 parametri.

- Una matrice quadrata di ordine 2 è antisimmetrica se e solo se ha la forma:

$$\begin{bmatrix} 0 & a \\ -a & 0 \end{bmatrix}$$

Le matrici antisimmetriche di ordine 2 dipendono da un unico parametro.

- Una matrice quadrata di ordine 3 è simmetrica se e solo se ha la forma:

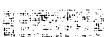
$$\begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}$$

Le matrici simmetriche di ordine 3 dipendono da 6 parametri.

- Una matrice quadrata di ordine 3 è antisimmetrica se e solo se ha la forma:

$$\begin{bmatrix} 0 & b & c \\ -b & 0 & e \\ -c & -e & 0 \end{bmatrix}$$

Le matrici antisimmetriche di ordine 3 dipendono da 2 parametri.



Scrivere le matrici trasposte delle matrici

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 3 \\ 0 & 1 & 2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 2 & 0 \\ -1 & 2 \end{bmatrix}, \quad \mathbf{AB}$$



Per quale valore di a la matrice $\begin{bmatrix} 0 & 2 \\ a & 0 \end{bmatrix}$ è simmetrica (risp. antisimmetrica)?



Mostrare che, data una qualunque matrice quadrata \mathbf{A} , la matrice $\mathbf{A} + \mathbf{A}^T$ è simmetrica e la matrice $\mathbf{A} - \mathbf{A}^T$ è antisimmetrica.

Mostrare che ogni matrice quadrata si scrive in uno e un solo modo come somma di una matrice simmetrica e di una matrice antisimmetrica.

Suggerimento: $\mathbf{A} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^T) + \frac{1}{2}(\mathbf{A} - \mathbf{A}^T)$.

Mostrare che il prodotto di due matrici simmetriche \mathbf{A} e \mathbf{B} è una matrice simmetrica se e solo se $\mathbf{AB} = \mathbf{BA}$. Trovare un esempio in cui il prodotto \mathbf{AB} non è una matrice simmetrica.

Mostrare che la matrice inversa di una matrice simmetrica è ancora una matrice simmetrica.

Mostrare che, se \mathbf{A} è una matrice $m \times n$, la matrice $\mathbf{A}^T \mathbf{A}$ è una matrice simmetrica $n \times n$, e \mathbf{AA}^T è una matrice simmetrica $m \times m$.

Sia \mathbf{A} la matrice simmetrica $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$, e sia $\mathbf{v} = [x, y]^T$ un generico vettore di \mathbb{K}^2 . Mostrare che

$$\mathbf{v}^T \mathbf{A} \mathbf{v} = ax^2 + 2bxy + cy^2$$

Sia $f(x_1, \dots, x_n)$ una funzione con derivate seconde continue su \mathbb{R}^n . Mostrare che la matrice Hessiana:

$$\mathbf{H}_f(\mathbf{x}) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) \right]$$

è una matrice simmetrica.

Al generico vettore $\mathbf{v} = [x, y, z]^T$ di \mathbb{R}^3 si associa la matrice antisimmetrica

$$\mathbf{E}(\mathbf{v}) = \begin{bmatrix} 0 & x & -y \\ -x & 0 & z \\ y & -z & 0 \end{bmatrix}$$

Si mostri che la matrice associata al prodotto vettoriale di due vettori è

$$\mathbf{E}(\mathbf{v} \times \mathbf{w}) = \mathbf{E}(\mathbf{v})\mathbf{E}(\mathbf{w}) - \mathbf{E}(\mathbf{w})\mathbf{E}(\mathbf{v})$$

Scrivere la forma generale di una matrice simmetrica di ordine n e concludere che le matrici simmetriche di ordine n dipendono da $\frac{n(n+1)}{2}$ parametri.

Scrivere la forma generale di una matrice antisimmetrica di ordine n e concludere che le matrici antisimmetriche di ordine n dipendono da $\frac{n(n-1)}{2}$ parametri.

Se \mathbf{A} è una matrice di tipo (m, n) , non necessariamente quadrata, si mostri che \mathbf{A} ammette un'inversa sinistra se e solo se $r(\mathbf{A}^T) = n$.

Suggerimento: per un esercizio precedente \mathbf{A}^T ha un'inversa destra se e solo se il suo rango è n . Commento: nel capitoli sugli spazi vettoriali vedremo che $r(\mathbf{A}^T) = r(\mathbf{A})$. Quindi l'esistenza dell'inversa sinistra equivale a $r(\mathbf{A}) = n$.

6 L'ALGORITMO DI GAUSS-JORDAN E IL CALCOLO DELL'INVERSA

In questo paragrafo descriviamo un algoritmo che consente di calcolare l'inversa di una matrice quadrata \mathbf{A} di rango massimo. Si tratta di evidenziare come la dimostrazione del teorema 4.3 consenta effettivamente di costruire l'inversa di \mathbf{A} . Per ottenere un algoritmo dobbiamo inoltre ricondurre il metodo di sostituzione all'indietro a operazioni sulle matrici del sistema.

Il problema è: data una matrice quadrata \mathbf{A} di rango massimo, scrivere un algoritmo che costruisca una matrice quadrata $\mathbf{B} = \mathbf{A}^{-1}$ tale che

$$\mathbf{AB} = \mathbf{I}$$

Come osservato nella dimostrazione del teorema 4.3, questo equivale a dire che le colonne $\mathbf{v}_1, \dots, \mathbf{v}_n$ di \mathbf{B} risolvono i sistemi lineari

$$(6.1) \quad \mathbf{Ax} = \mathbf{e}_1, \dots, \mathbf{Ax} = \mathbf{e}_n$$

dove $\mathbf{e}_1, \dots, \mathbf{e}_n$ sono le colonne della matrice identità \mathbf{I} . Calcolare la matrice inversa è quindi equivalente a risolvere simultaneamente gli n sistemi lineari (6.1). Descriviamo ora un algoritmo che risolve simultaneamente questi sistemi.

Consideriamo un sistema quadrato con matrice completa

$$[\mathbf{A}|\mathbf{b}]$$

e supponiamo che \mathbf{A} abbia rango n . Il sistema ammette allora un'unica soluzione. Per determinarla, procediamo con l'algoritmo di Gauss, che produce un sistema equivalente con matrice completa

$$[\mathbf{U}|\mathbf{b}_1]$$

dove \mathbf{U} è triangolare alta. Siccome $r(\mathbf{U}) = r(\mathbf{A}) = n$, i pivots di \mathbf{U} sono gli elementi $p_i = (\mathbf{U})_{ii}$ della diagonale principale di \mathbf{U} e sono tutti non nulli:

$$\mathbf{U} = \begin{bmatrix} p_1 & * & * & * & * & * & * \\ 0 & p_2 & * & * & * & * & * \\ 0 & 0 & p_3 & * & * & * & * \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & . & p_n \end{bmatrix}$$

Per avere un esempio esplicito, come nel capitolo 2 consideriamo il sistema lineare associato alla matrice completa

$$[\mathbf{A}|\mathbf{b}] = \left[\begin{array}{ccc|c} 1 & -4 & 1 & 2 \\ 2 & -6 & 5 & 8 \\ 1 & -2 & 5 & 8 \end{array} \right]$$

L'algoritmo di Gauss riduce la matrice alla forma

$$[\mathbf{U}|\mathbf{b}'] = \left[\begin{array}{ccc|c} 1 & -4 & 1 & 2 \\ 0 & 2 & 3 & 4 \\ 0 & 0 & 1 & 2 \end{array} \right]$$

La soluzione del sistema si trova velocemente per sostituzione all'indietro. Le operazioni di sostituzione all'indietro si possono svolgere sulla matrice dei coefficienti con il procedimento seguente, che riduce \mathbf{U} a una matrice diagonale. Sottraendo un opportuno multiplo dell'ultima riga alle altre righe, si cancellano tutti i termini sopra p_n nell' n -esima colonna. Queste operazioni non modificano le altre colonne della matrice dei coefficienti. Poi sottraendo un opportuno multiplo della penultima riga alle righe sovrastanti, si cancellano tutti i termini sopra p_{n-1} nella penultima colonna; queste operazioni modificano solo la penultima colonna della matrice dei coefficienti, che ora sopra p_{n-1} ha tutti i coefficienti nulli. Proseguendo così si cancellano tutti i coefficienti sopra i pivots c , quindi, alla fine della procedura la matrice \mathbf{U} è ridotta alla matrice diagonale $\mathbf{D} = \text{diag}(p_1, \dots, p_n)$.

Nell'esempio la riduzione si effettua coi passaggi

$$\left[\begin{array}{ccc|c} 1 & -4 & 1 & 2 \\ 0 & 2 & 3 & 4 \\ 0 & 0 & 1 & 2 \end{array} \right] \rightarrow \left[\begin{array}{ccc|c} 1 & -4 & 0 & 0 \\ 0 & 2 & 0 & -2 \\ 0 & 0 & 1 & 2 \end{array} \right] \rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & -4 \\ 0 & 2 & 0 & -2 \\ 0 & 0 & 1 & 2 \end{array} \right]$$

e $\mathbf{D} = \text{diag}(1, 2, 1)$.

In generale si ottiene una matrice della forma

$$[\mathbf{D} | \mathbf{b}_2]$$

con \mathbf{D} diagonale di rango massimo. Ora si divide ogni riga per il corrispondente pivot ottenendo la matrice completa

$$[\mathbf{I} | \mathbf{b}_3]$$

dove \mathbf{I} denota come sempre la matrice identità. Nell'esempio:

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & -4 \\ 0 & 2 & 0 & -2 \\ 0 & 0 & 1 & 2 \end{array} \right] \rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & -4 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{array} \right]$$

In questo passaggio abbiamo utilizzato un nuovo tipo di operazione sulle righe: *moltiplicare una riga per uno scalare non nullo*. Questa operazione evidentemente non cambia l'insieme delle soluzioni del sistema. Quindi il sistema associato alla matrice $[\mathbf{I} | \mathbf{b}_3]$ è equivalente al sistema dato. Ma il sistema associato alla matrice $[\mathbf{I} | \mathbf{b}_3]$ è

$$\mathbf{x} = \mathbf{b}_3$$

e quindi \mathbf{b}_3 è la soluzione del sistema. Nell'esempio la soluzione è $\mathbf{b}_3 = [-4, -1, 2]^T$. Il procedimento che abbiamo utilizzato per ridurre la matrice $[\mathbf{A} | \mathbf{b}]$ alla forma $[\mathbf{I} | \mathbf{b}_3]$ è noto come *algoritmo di Gauss-Jordan*. Riassumendo:

Sia $\mathbf{Ax} = \mathbf{b}$ un sistema lineare quadrato. Supponiamo che \mathbf{A} abbia rango massimo. L'algoritmo di Gauss-Jordan riduce la matrice completa $[\mathbf{A} | \mathbf{b}]$ alla matrice $[\mathbf{I} | \mathbf{b}_3]$. Il vettore \mathbf{b}_3 è la soluzione del sistema $\mathbf{Ax} = \mathbf{b}$.

Torniamo al problema di determinare l'inversa di una matrice A . Dobbiamo risolvere simultaneamente i sistemi lineari (6.1):

$$\mathbf{Ax} = \mathbf{e}_1, \dots, \mathbf{Ax} = \mathbf{e}_n$$

Possiamo riportare i dati degli n sistemi (le loro matrici complete) in un'unica matrice aggiungendo alla matrice dei coefficienti n colonne coi termini noti di tutti gli n sistemi:

$$[\mathbf{A}|\mathbf{e}_1 \quad \mathbf{e}_n] = [\mathbf{A}|\mathbf{I}]$$

Applicando l'algoritmo di Gauss-Jordan riduciamo questa matrice alla forma

$$[\mathbf{I}|\mathbf{v}_1 \quad \mathbf{v}_n] = [\mathbf{I}|\mathbf{B}]$$

Per quanto visto sopra, il vettore \mathbf{v}_i risolve $\mathbf{Ax} = \mathbf{e}_i$ per ogni $i = 1, 2, \dots, n$. Quindi la matrice

$$\mathbf{B} = [\mathbf{v}_1 \quad \mathbf{v}_n]$$

è l'inversa di \mathbf{A} .

Sia \mathbf{A} una matrice quadrata di rango massimo. Per trovare l'inversa di \mathbf{A} si accosta ad \mathbf{A} la matrice identità \mathbf{I} , quindi con l'algoritmo di Gauss-Jordan si riduce la matrice $[\mathbf{A}|\mathbf{I}]$ alla matrice $[\mathbf{I}|\mathbf{B}]$. La matrice \mathbf{B} è l'inversa di \mathbf{A} .



Calcoliamo l'inversa della matrice

$$\mathbf{A} = \begin{bmatrix} 1 & -4 & 1 \\ 2 & -6 & 5 \\ 1 & -2 & 5 \end{bmatrix}$$

Accostiamo ad \mathbf{A} la matrice identità:

$$[\mathbf{A}|\mathbf{I}] = \left[\begin{array}{ccc|ccc} 1 & -4 & 1 & 1 & 0 & 0 \\ 2 & -6 & 5 & 0 & 1 & 0 \\ 1 & -2 & 5 & 0 & 0 & 1 \end{array} \right]$$

Riduciamo \mathbf{A} a forma triangolare:

$$\left[\begin{array}{ccc|ccc} 1 & -4 & 1 & 1 & 0 & 0 \\ 2 & -6 & 5 & 0 & 1 & 0 \\ 1 & -2 & 5 & 0 & 0 & 1 \end{array} \right] \rightarrow \left[\begin{array}{ccc|ccc} 1 & -4 & 1 & 1 & 0 & 0 \\ 0 & 2 & 3 & -2 & 1 & 0 \\ 0 & 2 & 4 & -1 & 0 & 1 \end{array} \right] \rightarrow \left[\begin{array}{ccc|ccc} 1 & -4 & 1 & 1 & 0 & 0 \\ 0 & 2 & 3 & -2 & 1 & 0 \\ 0 & 0 & 1 & 1 & -1 & 1 \end{array} \right]$$

e poi a forma diagonale:

$$\left[\begin{array}{ccc|ccc} 1 & -4 & 1 & 1 & 0 & 0 \\ 0 & 2 & 3 & -2 & 1 & 0 \\ 0 & 0 & 1 & 1 & -1 & 1 \end{array} \right] \rightarrow \left[\begin{array}{ccc|ccc} 1 & -4 & 0 & 0 & 1 & -1 \\ 0 & 2 & 0 & -5 & 4 & -3 \\ 0 & 0 & 1 & 1 & -1 & 1 \end{array} \right] \rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -10 & 9 & -7 \\ 0 & 2 & 0 & -5 & 4 & -3 \\ 0 & 0 & 1 & 1 & -1 & 1 \end{array} \right]$$

Infine dividiamo per il pivot 2 la seconda riga

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -10 & 9 & -7 \\ 0 & 1 & 0 & -5/2 & 2 & -3/2 \\ 0 & 0 & 1 & 1 & -1 & 1 \end{array} \right]$$

per cui

$$\mathbf{A}^{-1} = \begin{bmatrix} 10 & 9 & -7 \\ -5/2 & 2 & -3/2 \\ 1 & -1 & 1 \end{bmatrix}$$

Controlliamo di non aver sbagliato i conti in due modi diversi. La prima verifica consiste nel calcolo del prodotto $\mathbf{A}^{-1}\mathbf{A}$:

$$\begin{aligned} & \begin{bmatrix} -10 & 9 & -7 \\ -5/2 & 2 & -3/2 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -4 & 1 \\ 2 & -6 & 5 \\ 1 & -2 & 5 \end{bmatrix} = \\ & = \begin{bmatrix} -10 + 18 - 7 & 40 - 54 + 14 & -10 + 45 - 35 \\ -5/2 + 4 - 3/2 & 10 - 12 + 3 & -5/2 + 10 - 15/2 \\ 1 - 2 + 1 & -4 + 6 - 2 & 1 - 5 + 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

Possiamo, alternativamente, calcolare $\mathbf{A}\mathbf{A}^{-1}$:

$$\begin{aligned} & \begin{bmatrix} 1 & -4 & 1 \\ 2 & -6 & 5 \\ 1 & -2 & 5 \end{bmatrix} \begin{bmatrix} -10 & 9 & -7 \\ -5/2 & 2 & -3/2 \\ 1 & -1 & 1 \end{bmatrix} = \\ & = \begin{bmatrix} -10 + 10 + 1 & 9 - 8 - 1 & -7 + 6 + 1 \\ -20 + 15 + 5 & 18 - 12 - 5 & -14 + 9 + 5 \\ -10 + 5 + 5 & 9 - 4 - 5 & -7 + 3 + 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

Nel capitolo sui sistemi abbiamo ricavato la soluzione $\mathbf{v} = \begin{bmatrix} -4 \\ -1 \\ 2 \end{bmatrix}$ del sistema lineare $\mathbf{Ax} = \begin{bmatrix} 2 \\ 8 \\ 8 \end{bmatrix}$. Tale soluzione deve coincidere con $\mathbf{A}^{-1} \begin{bmatrix} 2 \\ 8 \\ 8 \end{bmatrix}$. Verifichiamo:

$$\begin{bmatrix} -10 & 9 & -7 \\ -5/2 & 2 & -3/2 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 8 \\ 8 \end{bmatrix} = \begin{bmatrix} -20 + 72 - 56 \\ -5 + 16 - 12 \\ 2 - 8 + 8 \end{bmatrix} = \begin{bmatrix} -4 \\ -1 \\ 2 \end{bmatrix}$$



Calcoliamo nuovamente l'inversa di una matrice $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ soddisfacente $ad - bc \neq 0$, questa volta col metodo di Gauss-Jordan. Assumiamo per semplicità $a \neq 0$ e applichiamo l'algoritmo:

$$\left[\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array} \right] \rightarrow \left[\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & d - \frac{c}{a}b & -\frac{c}{a} & 1 \end{array} \right]$$

Poniamo $e = ad - bc$ e proseguiamo l'algoritmo sottraendo dalla prima riga la seconda riga moltiplicata per $\frac{ab}{e}$

$$\left[\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & \frac{e}{a} & -\frac{c}{a} & 1 \end{array} \right] \rightarrow \left[\begin{array}{cc|cc} a & 0 & \frac{ad}{e} & -\frac{ab}{e} \\ 0 & \frac{e}{a} & -\frac{c}{a} & 1 \end{array} \right]$$

Dividiamo infine la prima riga per il suo pivot a e la seconda riga per e/a : otteniamo

$$\left[\begin{array}{cc|cc} 1 & 0 & \frac{d}{e} & -\frac{b}{e} \\ 0 & 1 & -\frac{c}{e} & \frac{a}{e} \end{array} \right]$$

quindi

$$\mathbf{A}^{-1} = \begin{bmatrix} \frac{d}{e} & -\frac{b}{e} \\ -\frac{c}{e} & \frac{a}{e} \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Nel capitolo sui sistemi lineari trovare tre sistemi quadrati con matrice dei coefficienti \mathbf{A} di rango massimo. Risolvere nuovamente tali sistemi calcolando prima \mathbf{A}^{-1} e poi la soluzione $\mathbf{A}^{-1}\mathbf{v}$.

7 FATTORIZZAZIONE LU

L'algoritmo di Gauss consente di risolvere un sistema lineare $\mathbf{Ax} = \mathbf{b}$ riducendo a scala la matrice $[\mathbf{A}|\mathbf{b}]$. Supponiamo di dover risolvere il sistema al variare del termine noto \mathbf{b} , mentre la matrice dei coefficienti \mathbf{A} rimane fissata. Ci chiediamo se sia possibile mantenere memoria delle operazioni effettuate in modo da non doverle ripetere per ogni nuova scelta del termine noto. La chiave per trovare la risposta a questo problema è l'osservazione che ciascuna operazione elementare sulle righe di una matrice si può ottenere moltiplicando la matrice a sinistra per una opportuna *matrice elementare*. Questo conduce alla fattorizzazione **LU** della matrice \mathbf{A} che ora descriviamo.

7.1 La fattorizzazione nel caso non siano necessari scambi di righe

Una matrice quadrata $\mathbf{A} = [a_{ij}]$ si dice *triangolare bassa* se i suoi elementi sopra la diagonale principale sono nulli: $a_{ij} = 0$ se $i < j$; si dice *triangolare alta* se i suoi elementi sotto la diagonale principale sono nulli: $a_{ij} = 0$ se $i > j$. L'uso delle lettere **L** (rispettivamente **U**) per denotare matrici triangolari basse (rispettivamente alte) viene dall'inglese *lower triangular* (rispettivamente *upper triangular*). Si ricordi che una matrice **U** che sia al tempo stesso a scala e quadrata è triangolare alta.

Vogliamo ora mostrare che l'algoritmo di Gauss, quando non comporta scambi di righe, produce una fattorizzazione $\mathbf{A} = \mathbf{LU}$ con **L** triangolare bassa e **U** a scala (quindi triangolare alta nel caso quadrato). Inoltre **L** ha tutti gli elementi sulla diagonale principale uguali a 1 e, al di sotto della diagonale principale gli elementi di **L** sono i moltiplicatori dell'algoritmo di eliminazione. Per esempio, la riduzione a scala

$$\mathbf{A} = \begin{bmatrix} 1 & -4 & 1 \\ 2 & -6 & 5 \\ 1 & -2 & 5 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -4 & 1 \\ 0 & 2 & 3 \\ 0 & 2 & 4 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -4 & 1 \\ 0 & 2 & 3 \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{U}$$

consiste di tre operazioni elementari. Il primo passaggio consiste di due operazioni elementari: sottrarre alla seconda riga la prima moltiplicata per $m_{21} = 2$ e quindi sottrarre alla terza riga la prima moltiplicata per $m_{31} = 1$. L'ultimo passaggio consiste nel sottrarre alla terza riga la seconda moltiplicata per $m_{32} = 1$. I coefficienti m_{ik} sono i moltiplicatori dell'algoritmo.

La matrice \mathbf{A} si fattorizza come prodotto

$$(7.1) \quad \mathbf{A} = \mathbf{L}\mathbf{U} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -4 & 1 \\ 0 & 2 & 3 \\ 0 & 0 & 1 \end{bmatrix}$$

Come annunciato \mathbf{L} è la matrice triangolare bassa che ha gli elementi al di sotto della diagonale principale uguali ai moltiplicatori m_{ik} e gli elementi sulla diagonale principale uguali a uno.

Per trattare il caso generale, riesaminiamo il processo di eliminazione sotto quest'ottica. L'ipotesi che facciamo è che l'algoritmo di Gauss si possa portare a termine senza scambi di riga. Quindi le operazioni elementari che riducono la matrice sono tutte della forma: sottrarre alla riga i un certo multiplo m_{ik} della riga k , con $i > k$ (la riga k precede la riga i). Più precisamente, supponiamo che $j(1) < j(2) < \dots < j(r)$ siano gli indici di colonna dei pivots della matrice ridotta. L'eliminazione si svolge sulle colonne $j(k)$ e consiste nella cancellazione degli elementi sotto il pivot di posto $(k, j(k))$ su queste colonne: l'operazione che permette la cancellazione è sottrarre alla riga i un certo multiplo m_{ik} della riga k , con $i > k$. Il moltiplicatore m_{ik} è il quoziente

$$m_{ik} = \frac{a_{i,j(k)}}{a_{k,j(k)}}$$

tra l'elemento di posto $(i, j(k))$ e il pivot di posto $(k, j(k))$.

Per completezza, quando $m > r + 1$, definiamo m_{ik} anche per $k > r$ ponendolo uguale a zero:

$$m_{ik} = 0 \quad \text{se } i > k > r$$

Come nell'esempio precedente, l'eliminazione senza scambi di righe fornisce una fattorizzazione $\mathbf{A} = \mathbf{L}\mathbf{U}$.

TEOREMA 7.1 (Fattorizzazione LU)

Sia \mathbf{A} una matrice di tipo (m, n) , e supponiamo che l'algoritmo di Gauss riduca \mathbf{A} a scala senza scambi di righe. Allora è possibile fattorizzare \mathbf{A} nella forma

$$(7.2) \quad \mathbf{A} = \mathbf{L}\mathbf{U}$$

dove \mathbf{L} è una matrice di tipo (m, m) triangolare bassa, con tutte le entrate sulla diagonale principale uguali a 1 e \mathbf{U} è una matrice a scala di tipo (m, n) . Nel caso $m = n$, la matrice \mathbf{U} è triangolare alta. Gli elementi di \mathbf{L} al di sotto della diagonale principale sono i moltiplicatori m_{ik} del processo di eliminazione ($i > k$).

DIMOSTRAZIONE. L'osservazione fondamentale è che l'operazione *sommare a una riga un multiplo di una riga precedente* si può realizzare moltiplicando a sinistra per un'opportuna matrice triangolare bassa:

Dati due indici distinti $i > k$, compresi tra 1 ed m , e uno scalare b , sia $\mathbf{F} = \mathbf{F}_{ik}(b)$ la matrice quadrata di ordine m che ha tutti gli elementi sulla diagonale principale uguali a 1, l'elemento di posto (i, k) uguale a b e tutti gli altri elementi uguali a zero. Per ogni matrice \mathbf{A} con m righe, il prodotto $\mathbf{F}\mathbf{A}$ è la matrice ottenuta da \mathbf{A} sommando alla riga i la riga k moltiplicata per lo scalare b .

La verifica di questo fatto non presenta difficoltà ed è lasciata al lettore. Quando $m = 3$ le matrici \mathbf{F} sono

$$\mathbf{F}_{21}(b) = \begin{bmatrix} 1 & 0 & 0 \\ b & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{F}_{31}(b) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ b & 0 & 1 \end{bmatrix}, \quad \mathbf{F}_{32}(b) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & b & 1 \end{bmatrix}$$

Siccome sommare alla riga i la riga k moltiplicata per $(-b)$ è l'operazione inversa di sommare alla riga i la riga k moltiplicata per b , la matrice $\mathbf{F}_{ik}(b)$ è invertibile e la sua inversa è $\mathbf{F}_{ik}(-b)$:

$$(\mathbf{F}_{ik}(b))^{-1} = \mathbf{F}_{ik}(-b)$$

Abbiamo ora tutti gli elementi per ottenere la fattorizzazione LU di una matrice. Per ipotesi, è possibile ridurre a scala \mathbf{A} con operazioni elementari del tipo: sottrarre a una riga un multiplo scalare di una riga precedente; tale operazione consiste nel moltiplicare a sinistra per una matrice $\mathbf{F}_{ik}(-m_{ik})$ con $i > k$. Per ridurre \mathbf{A} a scala, dobbiamo effettuare un certo numero di operazioni di questo tipo, corrispondenti a matrici elementari $\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_p$: il processo completo di riduzione è quindi

$$(7.3) \quad \mathbf{F}_p \mathbf{F}_{p-1} \cdots \mathbf{F}_1 \mathbf{A} = \mathbf{U}$$

dove \mathbf{U} è a scala. Da questo uguaglianza segue $\mathbf{A} = \mathbf{LU}$ dove

$$\mathbf{L} = \mathbf{F}_1^{-1} \mathbf{F}_2^{-1} \cdots \mathbf{F}_p^{-1}$$

Per finire occorre ancora mostrare che \mathbf{L} è triangolare bassa, con tutte le entrate sulla diagonale principale uguali a 1 e con i moltiplicatori m_{ik} al di sotto della diagonale principale. Questo segue dal fatto che \mathbf{L} si ottiene come prodotto delle matrici $\mathbf{F}_{ik}(m_{ik})$, con $i > k$, e che nel prodotto compaiono prima i fattori F_{i1} con $i = 2, 3, \dots, m$, poi i fattori F_{i2} e così via fino ad arrivare a $k = r$. Dato che l'effetto della moltiplicazione a sinistra per $\mathbf{F}_{ik}(m_{ik})$ è sommare alla riga i la riga k moltiplicata per m_{ik} , il prodotto \mathbf{L} è la matrice triangolare bassa descritta nell'enunciato. Per esempio, nel caso $m = 3$

$$\begin{aligned} \mathbf{L} &= \begin{bmatrix} 1 & 0 & 0 \\ m_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ m_{31} & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & m_{32} & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 1 & 0 & 0 \\ m_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ m_{31} & m_{32} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ m_{21} & 1 & 0 \\ m_{31} & m_{32} & 1 \end{bmatrix} \end{aligned}$$

Si faccia attenzione: l'inversa $\mathbf{F}_p \mathbf{F}_{p-1} \cdots \mathbf{F}_1$ di \mathbf{L} è triangolare bassa con tutti gli elementi sulla diagonale principale uguale a 1, ma *non* ha come elementi, sotto la diagonale principale gli opposti $-m_{ik}$ dei moltiplicatori: è importante l'ordine con cui si moltiplicano i fattori \mathbf{F}_{ik} .



Una volta calcolata la fattorizzazione $\mathbf{A} = \mathbf{LU}$, il sistema lineare $\mathbf{Ax} = \mathbf{b}$ si risolve velocemente per *ogni* vettore \mathbf{b} : possiamo infatti risolvere il sistema nei due passi

- trovare \mathbf{c} tale che $\mathbf{Lc} = \mathbf{b}$;
- trovare \mathbf{x} tale che $\mathbf{Ux} = \mathbf{c}$.

Il primo passo consiste nel risolvere un sistema lineare con matrice \mathbf{L} triangolare bassa; il secondo nel risolvere un sistema a scala. Entrambi i sistemi si risolvono velocemente: nel caso \mathbf{A} sia una matrice quadrata, ciascuno dei passi richiede $n^2/2$ flops, mentre il processo di eliminazione, che produce la fattorizzazione, ne richiede $2n^3/3$ (a meno di termini di ordine inferiore). Quindi il costo computazionale della soluzione risiede nel calcolo della fattorizzazione.



Supponiamo di voler risolvere il sistema lineare

$$\begin{bmatrix} 1 & -4 & 1 \\ 2 & -6 & 5 \\ 1 & -2 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Come abbiamo visto all'inizio del paragrafo, la fattorizzazione \mathbf{LU} della matrice dei coefficienti del sistema è

$$(7.4) \quad \mathbf{A} = \mathbf{LU} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -4 & 1 \\ 0 & 2 & 3 \\ 0 & 0 & 1 \end{bmatrix}$$

Cominciamo col risolvere $\mathbf{Lc} = \mathbf{b}$:

$$\begin{cases} c_1 = 1 \\ 2c_1 + c_2 = 2 \\ c_1 + c_2 + c_3 = 3 \end{cases}$$

Il sistema si risolve per sostituzione *in avanti*, ricavando c_1 dalla prima equazione, quindi c_2 dalla seconda equazione, e infine c_3 dalla terza equazione:

$$\begin{cases} c_1 = 1 \\ c_2 = 2 - 2c_1 = 0 \\ c_3 = 3 - c_1 - c_2 = 2 \end{cases}$$

Ora risolviamo $\mathbf{Ux} = \mathbf{c}$:

$$\begin{cases} x_1 - 4x_2 + x_3 = 1 \\ 2x_2 + 3x_3 = 0 \\ x_3 = 2 \end{cases}$$

per sostituzione all'indietro:

$$\begin{cases} x_3 = 2 \\ x_2 = \frac{-3x_3}{2} = -3 \\ x_1 = 1 + 4x_2 - x_3 = 1 - 12 - 2 = -13 \end{cases}$$

La soluzione del sistema $\mathbf{Ax} = \mathbf{b}$ è quindi $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -13 \\ -3 \\ 2 \end{bmatrix}$.

Complementi

La fattorizzazione LU, così come descritta nel teorema 7.1, è unica.

DIMOSTRAZIONE. In dettaglio trattiamo solo il caso di una matrice $n \times n$ di rango massimo. In questo caso possiamo mostrare facilmente che, se

$$\mathbf{L}_1 \mathbf{U}_1 = \mathbf{L}_2 \mathbf{U}_2,$$

allora $\mathbf{L}_1 = \mathbf{L}_2$ e $\mathbf{U}_1 = \mathbf{U}_2$. L'ipotesi è che le matrici \mathbf{L}_i siano triangolari basse con gli elementi sulla diagonale principale uguali a 1 e che le matrici \mathbf{U}_i siano triangolari alte di rango massimo. Per mostrare la tesi, moltiplichiamo l'uguaglianza precedente a destra per \mathbf{U}_1^{-1} e a sinistra per \mathbf{L}_2^{-1} , ottenendo

$$\mathbf{L}_2^{-1} \mathbf{L}_1 = \mathbf{U}_2 \mathbf{U}_1^{-1}$$

La matrice a primo membro è triangolare bassa con gli elementi sulla diagonale principale uguali a uno (si vedano gli esercizi alla fine del paragrafo), la matrice a secondo membro è triangolare alta. Quindi $\mathbf{L}_2^{-1} \mathbf{L}_1$ è una matrice diagonale con gli elementi sulla diagonale uguali a uno, cioè la matrice identità. Questo significa $\mathbf{L}_1 = \mathbf{L}_2$; moltiplicando l'uguaglianza di partenza per \mathbf{L}_1^{-1} otteniamo anche $\mathbf{U}_1 = \mathbf{U}_2$.



Nel caso $\mathbf{A} = \mathbf{LU}$ sia quadrata di rango massimo, la fattorizzazione si può rendere più simmetrica:

$$\mathbf{A} = \mathbf{LDU}_1$$

dove \mathbf{D} è diagonale e \mathbf{U}_1 è triangolare alta con tutti gli elementi sulla diagonale principale uguali a 1.

DIMOSTRAZIONE. L'ipotesi di rango massimo significa che gli elementi u_{ii} sulla diagonale principale di \mathbf{U} sono diversi da zero, per cui $\mathbf{U} = \text{diag}(u_{11}, \dots, u_{nn}) \mathbf{U}_1$:

$$\begin{bmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ 0 & u_{22} & \dots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & u_{nn} \end{bmatrix} = \begin{bmatrix} u_{11} & 0 & \dots & 0 \\ 0 & u_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & u_{nn} \end{bmatrix} \begin{bmatrix} 1 & u_{12}/u_{11} & \dots & u_{1n}/u_{11} \\ 0 & 1 & \dots & u_{2n}/u_{22} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$



Consideriamo ora il caso di matrici simmetriche a coefficienti reali.

Sia \mathbf{A} una matrice simmetrica di rango massimo che ammette la fattorizzazione $\mathbf{A} = \mathbf{LU}$: allora $\mathbf{A} = \mathbf{LDL}^T$ con \mathbf{D} diagonale.

DIMOSTRAZIONE. Occorre combinare le due osservazioni precedenti: da $\mathbf{A} = \mathbf{LU}$ ricaviamo $\mathbf{A} = \mathbf{LDU}_1$. Trasponendo questa uguaglianza si ottiene

$$\mathbf{A} = \mathbf{A}^T = \mathbf{U}_1^T \mathbf{D}^T \mathbf{L}^T = \mathbf{U}_1^T \mathbf{DL}^T$$

Si osservi ora che \mathbf{U}_1^T è triangolare bassa con tutti gli elementi sulla diagonale principale uguali a uno, mentre \mathbf{DL}^T è triangolare alta. Dall'unicità della fattorizzazione \mathbf{LU} segue che $\mathbf{L} = \mathbf{U}_1^T$ e $\mathbf{DL}^T = \mathbf{U}$. Quindi

$$\mathbf{A} = \mathbf{LU} = \mathbf{LDL}^T$$

Fattorizzazione di Cholesky

Sia \mathbf{A} una matrice simmetrica di rango massimo che ammette la fattorizzazione $\mathbf{A} = \mathbf{LDL}^T$ con *i pivots tutti positivi*:

$$\mathbf{D} = \text{diag}(p_1, \dots, p_n), \quad p_i > 0 \quad \text{per ogni } i = 1, 2, \dots, n$$

Allora esiste una matrice triangolare bassa \mathbf{C} , con tutti i pivots positivi, tale che

$$\mathbf{A} = \mathbf{CC}^T$$

DIMOSTRAZIONE. Siccome

$$\mathbf{D} = \text{diag}(\sqrt{p_1}, \dots, \sqrt{p_n})^2 = \text{diag}(\sqrt{p_1}, \dots, \sqrt{p_n})(\text{diag}(\sqrt{p_1}, \dots, \sqrt{p_n}))^T,$$

basta (e occorre) prendere

$$\mathbf{C} = \mathbf{L} \text{diag}(\sqrt{p_1}, \dots, \sqrt{p_n})$$

La decomposizione $\mathbf{A} = \mathbf{CC}^T$ è nota come *fattorizzazione di Cholesky*. Le matrici simmetriche che ammettono tale decomposizione (eliminazione senza scambi e con pivots positivi) si dicono *definite positive*.

Possiamo infine esplicitare, nel caso di una matrice invertibile, la condizione che l'algoritmo di Gauss termini senza bisogno di scambi di righe. Per questo, data una matrice \mathbf{A} quadrata di ordine n , per ogni $k = 1, 2, \dots, n$ definiamo $\mathbf{A}_{(k)}$ come la matrice $k \times k$ ottenuta cancellando le ultime $n-k$ righe e le ultime $n-k$ colonne di \mathbf{A} .

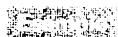
Supponiamo che \mathbf{A} sia una matrice $n \times n$ invertibile. Allora \mathbf{A} ammette la fattorizzazione \mathbf{LU} se e solo se $\mathbf{A}_{(k)}$ ha rango k per ogni $k = 1, 2, \dots, n$.

DIMOSTRAZIONE. Se \mathbf{A} ammette la fattorizzazione \mathbf{LU} , allora il rango di \mathbf{U} , che è uguale a quello di \mathbf{A} , è n . Quindi tutti gli elementi sulla diagonale principale di \mathbf{U} sono non nulli, e perciò $\mathbf{U}_{(k)}$ ha rango k . Siccome

$$\mathbf{A}_{(k)} = \mathbf{L}_{(k)} \mathbf{U}_{(k)}$$

la matrice $\mathbf{U}_{(k)}$ è la riduzione a scala di $\mathbf{A}_{(k)}$ che ha perciò rango k .

Viceversa, supponiamo che $\mathbf{A}_{(k)}$ abbia rango k per ogni k . Supponiamo per assurdo che l'algoritmo di Gauss non si possa portare a termine senza scambi di righe: questo significa che esiste un k tale che, dopo aver sistemato senza scambi le prime $k-1$ colonne, l'algoritmo produca una matrice con l'elemento di posto (k, k) nullo. Ma allora la matrice $\mathbf{A}_{(k)}$ si riduce a una matrice a scala con la riga k (l'ultima) uguale a zero. Quindi il rango di $\mathbf{A}_{(k)}$ è al massimo $k-1$, contraddicendo l'ipotesi.



La matrice $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ non ammette la fattorizzazione **LU** perché $\mathbf{A}_{(1)} = [0]$ ha rango 0.

Risolvere il sistema lineare $\begin{bmatrix} 1 & -4 & 1 \\ 2 & -6 & 5 \\ 1 & -2 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$ per ogni valore di b_1, b_2, b_3 .

Scrivere la fattorizzazione **LDL**^T della matrice simmetrica $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$.

Scrivere la fattorizzazione di Cholesky **CC**^T della matrice simmetrica definita positiva $\begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$.

Scrivere la fattorizzazione **LU** per le matrici che nel capitolo sui sistemi lineari sono state ridotte a scala senza scambi di righe. Verificare il risultato controllando che il prodotto **LU** coincida davvero con **A**.

Mostrare che il prodotto di due matrici triangolari alte (rispettivamente basse) è ancora una matrice triangolare alta (rispettivamente bassa). Mostrare che il prodotto di due matrici triangolari alte con gli elementi sulla diagonale principale uguali a uno è ancora una matrice triangolare alta con gli elementi sulla diagonale uguali a uno.

Suggerimento: trattare prima i casi $n = 2$ e $n = 3$ per capire cosa succede.

Mostrare che l'inversa di una matrice triangolare alta (rispettivamente bassa) di rango massimo è ancora una matrice triangolare alta (rispettivamente bassa).

7.2 La fattorizzazione nel caso generale

In generale per poter portare a termine il MEG può essere necessario scambiare delle righe. Anche l'operazione elementare di scambiare due righe può essere implementata mediante la moltiplicazione a sinistra per una matrice, ma tale matrice non è triangolare bassa e questo comporta che non sia più possibile arrivare a una fattorizzazione **LU** come nel paragrafo precedente. Possiamo però pensare di effettuare tutti gli scambi necessari per evitare i pivots nulli *prima* di procedere alle cancellazioni: possiamo cioè permutare le righe della matrice **A** in modo da ottenere una matrice **A'** la cui riduzione non richieda scambi di righe. Per quanto visto nel paragrafo precedente, la matrice **A'** ammette una fattorizzazione **LU**, mentre vediamo ora come si possa esprimere il legame tra **A** e **A'**.

La matrice **A'** si ottiene da **A** con una permutazione delle righe. Per essere precisi, supponiamo che la riga i di **A'** sia la riga $\sigma(i)$ di **A**: otteniamo così una *permutazione* $(\sigma(1), \sigma(2), \dots, \sigma(m))$ di $\{1, 2, \dots, m\}$. Alla permutazione σ di $\{1, 2, \dots, m\}$ associamo la matrice **P** $_{\sigma}$ di tipo (m, m) , ottenuta dalla matrice identità di ordine m come **A'** è ottenuta da: la riga i di **P** $_{\sigma}$ è la riga $\sigma(i)$ della matrice identità. Quindi ogni riga di **P** $_{\sigma}$ ha un elemento uguale a 1, quello con indice di colonna $\sigma(i)$ e tutti gli altri elementi nulli. Per esempio:

$$\mathbf{P}_{(2,3,1)} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Le matrici \mathbf{P}_σ si dicono *matrici di permutazione*. Una matrice di permutazione \mathbf{P}_σ ha esattamente un 1 su ogni riga e su ogni colonna, tutti gli altri elementi della matrice sono nulli. Viceversa, ogni matrice che abbia esattamente un 1 su ogni riga e su ogni colonna e tutti gli altri elementi nulli, è una matrice di permutazione \mathbf{P}_σ : la permutazione σ si trova ponendo $\sigma(i)$ uguale all'indice di colonna dell'unico 1 dell' i -esima riga. Per esempio la matrice:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

definisce $\sigma = (1, 3, 2)$.

Mostriamo ora che $\mathbf{A}' = \mathbf{P}_\sigma \mathbf{A}$: moltiplicare \mathbf{A} a sinistra per una matrice di permutazione ha l'effetto di permutare le righe di \mathbf{A} .

Permutazione delle righe di una matrice

Sia \mathbf{A} una matrice di tipo (m, n) , sia σ una permutazione di $\{1, 2, \dots, m\}$ e sia \mathbf{P}_σ la matrice di permutazione associata a σ . Allora $\mathbf{P}_\sigma \mathbf{A}$ è la matrice ottenuta permutando le righe di \mathbf{A} come prescritto da σ :

$$(7.5) \quad \mathbf{P}_\sigma \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{\sigma(1)} \\ \mathbf{a}_{\sigma(2)} \\ \vdots \\ \mathbf{a}_{\sigma(m)} \end{bmatrix}$$

DIMOSTRAZIONE. L' i -esima riga della matrice prodotto \mathbf{PA} è il prodotto dell' i -esima riga di \mathbf{P} per la matrice \mathbf{A} . L' i -esima riga di \mathbf{P}_σ ha tutti gli elementi nulli, tranne quello sulla colonna σ_i , che è uguale a 1; moltiplicando l' i -esima riga di \mathbf{P}_σ per la matrice \mathbf{A} si ottiene perciò la riga σ_i di \mathbf{A} .

Ne deduciamo:

TEOREMA 7.2 (Fattorizzazione $\mathbf{PA} = \mathbf{LU}$)

Sia \mathbf{A} una matrice di tipo (m, n) . Allora esistono una matrice di permutazione \mathbf{P} di ordine m , una matrice triangolare bassa \mathbf{L} di ordine m con tutte le entrate sulla diagonale principale uguali a 1 e una matrice a scala \mathbf{U} di tipo (m, n) , tali che

$$(7.6) \quad \mathbf{PA} = \mathbf{LU}$$

DIMOSTRAZIONE. Possiamo riordinare le righe di \mathbf{A} in modo da ottenere una matrice \mathbf{A}' su cui l'algoritmo di Gauss termina senza bisogno di scambi di riga. Per quanto visto in precedenza, possiamo fattorizzare \mathbf{A}' nella forma \mathbf{LU} . D'altra parte esiste una matrice di permutazione \mathbf{P} tale che $\mathbf{A}' = \mathbf{PA}$ perché le righe di \mathbf{A}' sono una permutazione delle righe di \mathbf{A} .

Consideriamo la riduzione a scala

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 3 \\ 0 & 3 & 2 & 1 \end{bmatrix} \xrightarrow{a} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 2 & 2 & 2 & 3 \\ 0 & 3 & 2 & 1 \end{bmatrix} \xrightarrow{b} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 3 & 2 & 1 \end{bmatrix} \xrightarrow{c} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 3 & 2 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \xrightarrow{d} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 3 & 2 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Nel passaggio a , si scambiano la prima e la seconda riga, in c si scambiano la seconda e la quarta riga, in d la terza e la quarta. Una volta compiuti tutti gli scambi la prima riga corrisponde alla seconda riga della matrice iniziale, la seconda alla quarta, la terza alla prima e l'ultima alla terza. La permutazione da effettuare è quindi $\sigma = (2, 4, 1, 3)$, per cui

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad \text{e} \quad \mathbf{A}' = \mathbf{PA} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 3 & 2 & 1 \\ 0 & 0 & 1 & 0 \\ 2 & 2 & 2 & 3 \end{bmatrix}$$

Il MEG su \mathbf{A}' consiste di un'unica operazione (il passaggio b di sopra), per cui la fattorizzazione \mathbf{LU} di \mathbf{A}' è

$$\mathbf{A}' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 3 & 2 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Osserviamo che:

- l'inversa di una matrice di permutazione \mathbf{P} è la sua trasposta \mathbf{P}^T ed è ancora una matrice di permutazione;
- nelle notazioni del teorema, la matrice \mathbf{A} si fattorizza nella forma

$$\mathbf{A} = \mathbf{P}^T \mathbf{LU}$$

DIMOSTRAZIONE. Il fatto che $\mathbf{P}^T = \mathbf{P}^{-1}$ si può verificare direttamente, ma si può spiegare concettualmente come segue. Per ogni permutazione σ esiste una permutazione *inversa* σ^{-1} che rimette a posto quello che σ ha messo in disordine. Per essere precisi, $\sigma^{-1} : \{1, 2, \dots, m\} \rightarrow \{1, 2, \dots, m\}$ è la funzione inversa di σ ed è, quindi, definita dalla proprietà:

$$\sigma^{-1}(j) = i \iff \sigma(i) = j$$

Per esempio, l'inversa della permutazione $\sigma = (2, 3, 1)$ è $\sigma^{-1} = (3, 1, 2)$.

Dal fatto che $\mathbf{P}_\sigma \mathbf{A}$ è la matrice \mathbf{A} con le righe permutate come prescritto da σ , segue immediatamente che

$$\mathbf{P}_\sigma \mathbf{P}_{\sigma^{-1}} = \mathbf{I}$$

e quindi $(\mathbf{P}_\sigma)^{-1} = \mathbf{P}_{\sigma^{-1}}$. Infine l'elemento di posto (j, i) di $\mathbf{P}_{\sigma^{-1}}$ è $\delta_{\sigma^{-1}(j)i} = \delta_{\sigma(i)j}$, quindi coincide con l'elemento di posto (i, j) in \mathbf{P}_σ . Questo significa $\mathbf{P}_{\sigma^{-1}} = \mathbf{P}_\sigma^T$.

La fattorizzazione $\mathbf{A} = \mathbf{P}^T \mathbf{LU}$ si ottiene moltiplicando l'uguaglianza $\mathbf{PA} = \mathbf{LU}$ a sinistra per $\mathbf{P}^T = \mathbf{P}^{-1}$.

Siano $\mathbf{A} = \begin{bmatrix} 1 & 2 & 4 \\ 3 & 1 & 5 \\ 2 & 0 & 7 \end{bmatrix}$ e $\mathbf{B} = \begin{bmatrix} 2 & 0 & 7 \\ 3 & 1 & 5 \\ 1 & 2 & 4 \end{bmatrix}$. Scrivere la matrice di permutazione \mathbf{P} tale che $\mathbf{PA} = \mathbf{B}$. Verificare il risultato eseguendo la moltiplicazione \mathbf{PA} .

 Trovare tra gli esempi del capitolo sui sistemi lineari tre sistemi $\mathbf{Ax} = \mathbf{b}$ per cui il MEG richieda scambi di righe. Scrivere la fattorizzazione $\mathbf{PA} = \mathbf{LU}$ per le matrici \mathbf{A} di tali sistemi. Riscrivere poi i sistemi nella forma $\mathbf{LUx} = \mathbf{Pb}$, spezzarli nella coppia di sistemi $\mathbf{Lc} = \mathbf{Pb}$ e $\mathbf{Ux} = \mathbf{c}$ e risolverli.

8 PRODOTTO DI MATRICI A BLOCCHI

Spesso è utile suddividere una matrice in blocchi più piccoli, tracciando o immaginando di tracciare delle linee orizzontali e/o verticali, come abbiamo fatto per la matrice completa $[\mathbf{A}|\mathbf{b}]$ di un sistema. Se \mathbf{A} e \mathbf{B} sono suddivisi in blocchi che si possono moltiplicare tra loro, il prodotto \mathbf{AB} si può effettuare come se i blocchi fossero dei numeri.

Supponiamo, per esempio, che la matrice $\mathbf{A} = [\mathbf{A}_1 \quad \mathbf{A}_2]$ sia costituita da due blocchi \mathbf{A}_1 e \mathbf{A}_2 di tipo (m, n_1) e (m, n_2) e che la matrice $\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix}$ sia costituita da due blocchi di tipo \mathbf{B}_1 e \mathbf{B}_2 di tipo (n_1, p) e (n_2, p) . Allora il prodotto \mathbf{AB} , che è di tipo (m, p) , è la matrice

$$[\mathbf{A}_1 \quad \mathbf{A}_2] \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix} = \mathbf{A}_1 \mathbf{B}_1 + \mathbf{A}_2 \mathbf{B}_2$$

Un esempio più generale, che è quello di uso più frequente, è il seguente:

Moltiplicazione a blocchi

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix}$$

La formula è valida quando ha senso, cioè quando $m = m_1 + m_2$, $n = n_1 + n_2$, $p = p_1 + p_2$ e

- i blocchi \mathbf{A}_{ij} sono matrici di tipo (m_i, n_j) ;
- i blocchi \mathbf{B}_{jk} sono matrici di tipo (n_j, p_k) .



Supponiamo che \mathbf{C} e \mathbf{D} siano due matrici di tipo $(3, 3)$, \mathbf{v} e \mathbf{w} siano due matrici di tipo $(3, 1)$, $\mathbf{0}$ sia la matrice nulla di tipo $(1, 3)$ e c e d siano due numeri. Allora il prodotto delle due matrici di tipo $(4, 4)$

$$\mathbf{A} = \begin{bmatrix} \mathbf{C} & \mathbf{v} \\ \mathbf{0} & c \end{bmatrix} \quad \text{e} \quad \mathbf{B} = \begin{bmatrix} \mathbf{D} & \mathbf{w} \\ \mathbf{0} & d \end{bmatrix}$$

è la matrice di tipo (4, 4)

$$\begin{bmatrix} \mathbf{C} & \mathbf{v} \\ \mathbf{0} & c \end{bmatrix} \begin{bmatrix} \mathbf{D} & \mathbf{w} \\ \mathbf{0} & d \end{bmatrix} = \begin{bmatrix} \mathbf{CD} & \mathbf{Cw} + dv \\ \mathbf{0} & cd \end{bmatrix}$$

Per passare a un esempio numerico, prendiamo $\mathbf{C} = \mathbf{I}$, la matrice identità di ordine 3, $\mathbf{v} = [1, 2, 3]^T$, $c = 2$, in modo da ottenere:

$$\mathbf{A} = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \\ \hline 0 & 0 & 0 & 1 \end{array} \right] = \left[\begin{array}{c|c} \mathbf{I} & \mathbf{v} \\ \hline \mathbf{0} & 1 \end{array} \right]$$

Usando la formula per la moltiplicazione a blocchi calcoliamo le potenze della matrice \mathbf{A} :

$$\mathbf{A}^2 = \begin{bmatrix} \mathbf{I} & \mathbf{v} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{v} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{I}^2 & \mathbf{I}\mathbf{v} + \mathbf{v} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & 2\mathbf{v} \\ \mathbf{0} & 1 \end{bmatrix}$$

$$\mathbf{A}^3 = \begin{bmatrix} \mathbf{I} & 2\mathbf{v} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{v} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{I}^2 & \mathbf{I}\mathbf{v} + 2\mathbf{v} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & 3\mathbf{v} \\ \mathbf{0} & 1 \end{bmatrix}$$

$$\mathbf{A}^k = \begin{bmatrix} \mathbf{I} & k\mathbf{v} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & k \\ 0 & 1 & 0 & 2k \\ 0 & 0 & 1 & 3k \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (\forall k \geq 1)$$



Siano $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{K}^n$. Sia A la matrice $n \times n$ le cui colonne sono i vettori $\mathbf{v}_1, \dots, \mathbf{v}_n$:

$$A = [\mathbf{v}_1 \dots \mathbf{v}_n]$$

Allora

$$AA^T = [\mathbf{v}_1 \dots \mathbf{v}_n] \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} = \mathbf{v}_1\mathbf{v}_1^T + \dots + \mathbf{v}_n\mathbf{v}_n^T$$

Si osservi che $\mathbf{v}_k\mathbf{v}_k^T$ è una matrice $n \times n$. Per esempio, se $\mathbf{v}_1 = [1, 2]^T$ e $\mathbf{v}_2 = [3, 4]^T$,

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} [1 \ 2] + \begin{bmatrix} 3 \\ 4 \end{bmatrix} [3 \ 4] = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} + \begin{bmatrix} 9 & 12 \\ 12 & 4 \end{bmatrix} = \begin{bmatrix} 10 & 14 \\ 14 & 20 \end{bmatrix}$$



Spazi vettoriali

1 INTRODUZIONE

In questo capitolo introduciamo la nozione astratta di *spazio vettoriale*. Si tratta semplicemente di un insieme su cui sono definite le operazioni di somma e di prodotto per uno scalare, con la richiesta che tali operazioni godano delle proprietà delle analoghe operazioni sui vettori liberi dello spazio euclideo. Tra gli spazi vettoriali studieremo più in dettaglio quelli che hanno *dimensione finita*. Questi spazi hanno una *base* finita, il che intuitivamente significa che ammettono un sistema di riferimento: fissata una base, un elemento dello spazio vettoriale è individuato dalla n -upla delle sue coordinate rispetto alla base, in modo del tutto analogo a come un vettore libero è individuato, una volta fissato un sistema di riferimento dello spazio euclideo, dalla terna delle sue coordinate. Il lettore deve quindi pensare a uno spazio vettoriale come a una generalizzazione dell'insieme dei vettori liberi e a una base come all'analogo di un sistema di riferimento; se esiste una base, è possibile identificare lo spazio astratto con \mathbb{K}^n . L'introduzione di questa struttura astratta ha due grossi vantaggi. Il primo vantaggio di introdurre uno spazio \mathbf{V} che stia a \mathbb{K}^n come l'insieme dei vettori liberi sta a \mathbb{R}^3 è quello di svincolare la teoria da una particolare scelta di coordinate. Per esempio è così possibile dare una definizione di rango di una matrice più concettuale, dalla quale sarà immediato dedurre che il rango *non dipende* dall'ordine in cui si considerano le righe della matrice. Inoltre, se non è imposta a priori una scelta di coordinate, questa può essere compiuta a posteriori, a seconda del problema che occorre risolvere, il che di regola consente drastiche semplificazioni dei conti da effettuare. La ricerca del miglior sistema di coordinate è il nocciolo della teoria degli autovettori e autovalori che illustreremo in un capitolo successivo, dopo che avremo introdotto le applicazioni lineari e avremo così svincolato anche le matrici dal sistema di coordinate.

Il secondo vantaggio, nello sviluppare una teoria astratta degli spazi vettoriali, è di ottenere risultati che sono validi indipendentemente dal particolare contesto (che sia quello dei sistemi lineari o quello dei vettori geometrici) in cui sono stati inizialmente scoperti e che quindi siano applicabili, più in generale, a ogni altra situazione in cui le operazioni rilevanti sono la somma e il prodotto per scalare. L'esempio più importante in questo caso è quello delle funzioni e delle equazioni differenziali lineari. In effetti, anche le funzioni possono essere sommate e moltiplicate per uno scalare,

e quindi formano uno spazio vettoriale. Inoltre la derivata è un operatore lineare, o come si dice in gergo, rispetta le operazioni di spazio vettoriale: la derivata di una somma di funzioni è la somma delle derivate e la derivata di una funzione moltiplicata per uno scalare è la funzione derivata moltiplicata per quello stesso scalare. Queste semplici osservazioni permettono l'uso di metodi di algebra lineare per la soluzione delle *equazioni differenziali lineari*, equazioni che sono alla base di moltissime applicazioni della matematica alle altre scienze. Per fare un esempio concreto, consideriamo l'equazione differenziale dell'oscillatore armonico:

$$(1.1) \quad y''(x) + y(x) = 0$$

Una soluzione della (1.1) è una funzione $y(x) : \mathbb{R} \rightarrow \mathbb{R}$ che sia derivabile due volte e per cui l'equazione (1.1) è soddisfatta per ogni x in \mathbb{R} . Siccome la derivata di una somma è la somma delle derivate, la somma di due soluzioni della (1.1) è ancora una soluzione; e moltiplicando una soluzione per una costante si ottiene ancora una soluzione. Quindi l'insieme \mathbf{V} delle soluzioni della (1.1) è uno spazio vettoriale. Le funzioni $\cos(x)$ e $\sin(x)$ sono evidentemente due soluzioni di (1.1) e in analisi si mostra che ogni soluzione si scrive in uno e un solo modo come combinazione lineare

$$y(x) = a \cos(x) + b \sin(x)$$

delle funzioni $\cos(x)$ e $\sin(x)$. Diremo che $\{\cos(x), \sin(x)\}$ è una base dell'insieme delle soluzioni e che la soluzione $a \cos(x) + b \sin(x)$ ha coordinate $[a, b]^T$ rispetto a tale base. L'insieme delle soluzioni si identifica, quindi, con \mathbb{R}^2 : è un piano all'interno dell'insieme di tutte le funzioni.

2 ASSIOMI ED ESEMPI

Come nei capitoli precedenti, il simbolo \mathbb{K} denota il campo degli scalari, che a seconda delle applicazioni potrà essere \mathbb{R} oppure \mathbb{C} , cioè gli scalari saranno numeri reali oppure complessi. I risultati di questo capitolo sono in realtà validi per un campo qualsiasi. Nel capitolo 1 abbiamo introdotto le operazioni di somma e prodotto per uno scalare sull'insieme dei vettori liberi dello spazio euclideo e mostrato che godono di certe proprietà **V1 – V8**. Nel capitolo 2 abbiamo introdotto analoghe operazioni sull'insieme dei vettori colonna \mathbb{K}^n . In questo capitolo trattiamo in generale la teoria degli *spazi vettoriali*, che sono insiemi dotati di due operazioni, la somma e il prodotto per uno scalare, che godono delle proprietà **V1 – V8**. Ne diamo ora la definizione formale.



DEFINIZIONE 2.1 (Spazio vettoriale sul campo \mathbb{K})

Uno *spazio vettoriale* sul campo \mathbb{K} è un insieme \mathbf{V} , i cui elementi si dicono vettori, dotato di:

- un'operazione *interna* che a una coppia di vettori \mathbf{v} e \mathbf{w} di \mathbf{V} associa un vettore di \mathbf{V} che si dice vettore *somma* di \mathbf{v} e \mathbf{w} e si denota col simbolo $\mathbf{v} + \mathbf{w}$;

- un'operazione *esterna* che a uno scalare $t \in \mathbb{K}$ e a un vettore $\mathbf{v} \in \mathbf{V}$ associa un vettore di \mathbf{V} che si dice *prodotto di \mathbf{v} per lo scalare t* e si denota col simbolo $t\mathbf{v}$;

per le quali sono soddisfatti i seguenti assiomi **V1 – V8**:

V1 Proprietà commutativa della somma:

$$\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v} \quad \text{per ogni } \mathbf{v}, \mathbf{w} \text{ in } \mathbf{V}$$

V2 Proprietà associativa della somma:

$$(\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 = \mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3) \quad \text{per ogni } \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \text{ in } \mathbf{V}$$

V3 Esistenza elemento neutro della somma: esiste un vettore $\mathbf{0}$ in \mathbf{V} , detto vettore nullo, tale che

$$\mathbf{v} + \mathbf{0} = \mathbf{v} \quad \text{per ogni } \mathbf{v} \in \mathbf{V}$$

V4 Esistenza dell'opposto di un dato vettore: per ogni vettore \mathbf{v} in \mathbf{V} , esiste un vettore $-\mathbf{v}$, detto opposto di \mathbf{v} , tale che

$$\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$$

V5 Proprietà distributiva del prodotto per uno scalare rispetto alla somma di vettori:

$$t(\mathbf{v} + \mathbf{w}) = t\mathbf{v} + t\mathbf{w} \quad \text{per ogni } t \in \mathbb{K} \text{ e ogni } \mathbf{v}, \mathbf{w} \in \mathbf{V}$$

V6 Proprietà distributiva del prodotto per uno scalare rispetto alla somma di scalari:

$$(t + u)\mathbf{v} = t\mathbf{v} + u\mathbf{v} \quad \text{per ogni } t, u \in \mathbb{K} \text{ e ogni } \mathbf{v} \in \mathbf{V}$$

V7 Proprietà associativa del prodotto per uno scalare:

$$t(u\mathbf{v}) = (tu)\mathbf{v} \quad \text{per ogni } t, u \in \mathbb{K} \text{ e ogni } \mathbf{v} \in \mathbf{V}$$

V8 Normalizzazione del prodotto per uno scalare:

$$1\mathbf{v} = \mathbf{v} \quad \text{per ogni } \mathbf{v} \in \mathbf{V}$$

Esempio

L'insieme \mathbb{K}^n è uno spazio vettoriale su \mathbb{K} rispetto alle operazioni di somma e prodotto per uno scalare definite (componente per componente) nel capitolo 2.

Esempio

L'insieme $M_{\mathbb{K}}(m, n)$ delle matrici di tipo (m, n) a coefficienti in \mathbb{K} è uno spazio vettoriale: possiamo infatti sommare due matrici dello stesso tipo e moltiplicare una matrice per uno scalare. Il vettore nullo di $M_{\mathbb{K}}(m, n)$ è la matrice che ha tutte le entrate uguali allo scalare zero. Si noti che possiamo identificare \mathbb{K}^n con $M_{\mathbb{K}}(n, 1)$: i vettori colonna con n componenti sono la stessa cosa delle matrici con n righe e una colonna. Infine, si osservi che i vettori dello spazio vettoriale $M_{\mathbb{K}}(m, n)$ sono matrici! D'ora in poi il termine vettore sarà usato in astratto per denotare un elemento di uno spazio vettoriale.

Esempio

Esempi sostanzialmente diversi dai precedenti si ottengono rimuovendo la condizione che i vettori abbiano un numero finito di componenti. In prima battuta possiamo considerare vettori con un'infinità numerabile di componenti: le componenti sono indicizzate dall'insieme dei numeri naturali $N = \{0, 1, 2, \dots\}$. Si ottiene così l'insieme \mathbb{K}^N delle successioni di numeri reali (o complessi):

$$\mathbb{K}^N = \{\mathbf{a} = \{a_k\} : k \in N, a_k \in \mathbb{K}\}$$

Analogamente al caso di \mathbb{K}^n , in \mathbb{K}^N si definiscono componente per componente le operazioni di somma e prodotto per uno scalare:

$$\begin{aligned} \{a_k\} + \{b_k\} &= \{a_k + b_k\} && \text{per ogni } k \in N \\ t\{a_k\} &= \{ta_k\} && \text{per ogni } k \in N \end{aligned}$$

Le operazioni in \mathbb{K}^N compaiono a primo membro di queste uguaglianze e sono definite per mezzo delle operazioni di somma e prodotto del campo \mathbb{K} che compaiono a secondo membro. Si verifica senza difficoltà che rispetto a tali operazioni \mathbb{K}^N è uno spazio vettoriale.

Esempio

Passiamo ora al caso in cui le componenti siano indicizzate da un insieme qualsiasi S . Un elemento di \mathbb{K}^S consiste perciò di una funzione che a ogni elemento $s \in S$ associa uno scalare $x_s \in \mathbb{K}$. Invece di scrivere così adottiamo la notazione comune per le funzioni: scriviamo f per la funzione e $f(s)$ per lo scalare che corrisponde all'elemento s di S :

$$\mathbb{K}^S = \{f = f(s) : s \in S, f(s) \in \mathbb{K}\}$$

L'esempio da tener presente, certamente familiare al lettore, è quello in cui $S = \mathbb{R}$ e $\mathbb{K} = \mathbb{R}$: si tratta delle funzioni reali di una variabile reale, quali per esempio $f(x) = \cos(x)$ o $f(x) = e^x$.

Definiamo, generalizzando quanto fatto negli esempi precedenti, la somma di due funzioni f e g e il prodotto di una funzione f per uno scalare t componente per componente:

$$\begin{aligned} (f + g)(s) &= f(s) + g(s) && \text{per ogni } s \in S \\ (tf)(s) &= tf(s) && \text{per ogni } s \in S \end{aligned}$$

Anche qui le operazioni a primo membro sono quelle nello spazio delle funzioni, e vengono definite per mezzo delle operazioni di somma e prodotto del campo \mathbb{K} che compaiono a secondo membro.

Il lettore è invitato a verificare che rispetto a tali operazioni \mathbb{K}^S è uno spazio vettoriale su \mathbb{K} . In questo caso i vettori sono funzioni e lo spazio si dice anche *spazio funzionale*. Il vettore nullo di \mathbb{K}^S è la funzione nulla, cioè la funzione che a ogni elemento $s \in S$ associa lo scalare 0.

Vediamo adesso alcune proprietà algebriche che seguono formalmente dagli assiomi.

Legge di cancellazione della somma

Dall'uguaglianza $\mathbf{v} + \mathbf{w}_1 = \mathbf{v} + \mathbf{w}_2$ segue $\mathbf{w}_1 = \mathbf{w}_2$.

DIMOSTRAZIONE. Per dimostrare la legge di cancellazione, sommiamo $-\mathbf{v}$ a entrambi i membri di $\mathbf{v} + \mathbf{w}_1 = \mathbf{v} + \mathbf{w}_2$ ottenendo:

$$(2.1) \quad -\mathbf{v} + (\mathbf{v} + \mathbf{w}_1) = -\mathbf{v} + (\mathbf{v} + \mathbf{w}_2)$$

Per la proprietà associativa il primo membro di (2.1) è uguale a

$$(-\mathbf{v} + \mathbf{v}) + \mathbf{w}_1 = \mathbf{0} + \mathbf{w}_1 = \mathbf{w}_1$$

Per lo stesso motivo il membro di destra di (2.1) è uguale a \mathbf{w}_2 , per cui $\mathbf{w}_1 = \mathbf{w}_2$ e la legge di cancellazione è dimostrata.

Come conseguenza, osserviamo che il vettore opposto di \mathbf{v} è univocamente determinato da \mathbf{v} . Infatti, se \mathbf{w}_1 e \mathbf{w}_2 sono due vettori opposti di \mathbf{v} , allora

$$\mathbf{v} + \mathbf{w}_1 = \mathbf{0} = \mathbf{v} + \mathbf{w}_2$$

e dalla legge di cancellazione segue $\mathbf{w}_1 = \mathbf{w}_2$.

La *differenza* $\mathbf{v} - \mathbf{w}$ di due vettori \mathbf{v} e \mathbf{w} si definisce ponendo

$$\mathbf{v} - \mathbf{w} = \mathbf{v} + (-\mathbf{w})$$

Naturalmente $\mathbf{w} + (\mathbf{v} - \mathbf{w}) = \mathbf{v}$: analogamente al caso dell'aritmetica, la differenza di \mathbf{v} e \mathbf{w} è quel vettore che sommato a \mathbf{w} dà come risultato \mathbf{v} .

Vale anche la

Legge di annullamento del prodotto per uno scalare

Il vettore $t\mathbf{v}$ è nullo se e solo se $t = 0$ oppure $\mathbf{v} = \mathbf{0}$.

DIMOSTRAZIONE. Mostriamo che $0\mathbf{v} = \mathbf{0}$ per ogni vettore \mathbf{v} : infatti

$$\mathbf{v} + 0\mathbf{v} = 1\mathbf{v} + 0\mathbf{v} = (1 + 0)\mathbf{v} = 1\mathbf{v} = \mathbf{v} = \mathbf{v} + \mathbf{0}$$

e quindi $0\mathbf{v} = \mathbf{0}$ per la legge di cancellazione della somma.

Mostriamo che $t\mathbf{0} = \mathbf{0}$ per ogni scalare t : infatti per ogni vettore \mathbf{v}

$$t\mathbf{0} = t(\mathbf{0} + \mathbf{0}) = t\mathbf{0} + t\mathbf{0}$$

e sommando $-t\mathbf{0}$ otteniamo $\mathbf{0} = t\mathbf{0}$.

Mostriamo che, viceversa, da $t\mathbf{v} = \mathbf{0}$ segue $t = 0$ oppure $\mathbf{v} = \mathbf{0}$. Infatti, se $t\mathbf{v} = \mathbf{0}$, ma $t \neq 0$, allora

$$\mathbf{0} = t^{-1}\mathbf{0} = t^{-1}(t\mathbf{v}) = (t^{-1}t)\mathbf{v} = 1\mathbf{v} = \mathbf{v}$$

■■■

Infine osserviamo che il vettore opposto $-\mathbf{v}$ coincide col prodotto $(-1)\mathbf{v}$ del vettore \mathbf{v} per lo scalare -1 . Infatti

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



Esempio Sia \mathbf{V} uno spazio vettoriale, siano $\mathbf{v}_1, \dots, \mathbf{v}_d$ vettori di \mathbf{V} e siano t_1, \dots, t_d degli scalari. Supponiamo

$$t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_d\mathbf{v}_d = \mathbf{0}$$

Mostrare che, se $t_d \neq 0$, allora

$$\mathbf{v}_d = -\frac{t_1}{t_d}\mathbf{v}_1 - \frac{t_2}{t_d}\mathbf{v}_2 - \cdots - \frac{t_{d-1}}{t_d}\mathbf{v}_{d-1}$$

Esempio Sia \mathbf{V} uno spazio vettoriale su \mathbb{K} , sia S un insieme qualsiasi e sia

$$\mathbf{V}^S = \{\mathbf{F} : S \rightarrow \mathbf{V}\}$$

l'insieme delle funzioni con dominio S e codominio \mathbf{V} . Definire su \mathbf{V}^S una struttura di spazio vettoriale (cioè somma e prodotto per uno scalare) come nell'esempio di \mathbb{K}^S , quindi componente per componente, utilizzando le operazioni definite in \mathbf{V} al posto di quelle di \mathbb{K} . Per esempio, l'insieme delle funzioni $\mathbf{F}(t) = (x(t), y(t), z(t))$ di una variabile reale t a valori in \mathbb{R}^3 è uno spazio vettoriale.

3 SOTTOSPAZI

Gli esempi del paragrafo precedente ne generano moltissimi altri con un semplice procedimento: dato uno spazio vettoriale \mathbf{V} , si cercano quei sottoinsiemi \mathbf{H} di \mathbf{V} che rispetto alle operazioni di somma e prodotto per uno scalare definite in \mathbf{V} sono essi stessi degli spazi vettoriali. Si pensi, per esempio, al caso in cui \mathbf{H} è un piano per l'origine nello spazio $\mathbf{V} = \mathbb{R}^3$. Uno degli obiettivi di questo capitolo è di descrivere i sottospazi di \mathbb{K}^n mediante equazioni cartesiane e mediante equazioni parametriche. Le equazioni parametriche definiranno delle coordinate nei sottospazi, per cui per esempio potremo descrivere i punti di un piano obliqui in \mathbb{R}^3 mediante due coordinate esattamente come per il piano cartesiano. Nel resto del capitolo \mathbf{V} denoterà uno spazio vettoriale sul campo \mathbb{K} .

DEFINIZIONE 3.1 (Sottospazio vettoriale)

Un sottoinsieme \mathbf{H} di \mathbf{V} si dice *sottospazio vettoriale* se \mathbf{H} è uno spazio vettoriale rispetto alle operazioni di somma e prodotto per uno scalare di \mathbf{V} . Questo è il caso se e solo se \mathbf{H} soddisfa le seguenti proprietà:

1) il vettore nullo $\mathbf{0}$ appartiene ad \mathbf{H} ;

2) \mathbf{H} è chiuso rispetto alla somma:

se \mathbf{v} e \mathbf{w} appartengono ad \mathbf{H} , anche $\mathbf{v} + \mathbf{w}$ appartiene ad \mathbf{H} ;

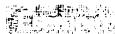
3) \mathbf{H} è chiuso rispetto al prodotto per uno scalare:

se t è uno scalare qualsiasi e \mathbf{v} appartiene ad \mathbf{H} , anche $t\mathbf{v}$ appartiene ad \mathbf{H} .

OSSERVAZIONE Se un sottoinsieme \mathbf{H} di \mathbf{V} è chiuso rispetto al prodotto scalare e contiene un vettore \mathbf{v} , allora contiene anche il vettore nullo perché $0\mathbf{v} = \mathbf{0}$. Nelle proprietà che caratterizzano un sottospazio si può quindi sostituire 1) con la richiesta che \mathbf{H} sia *non vuoto*.



Dato un qualunque spazio vettoriale \mathbf{V} , il sottoinsieme $\{\mathbf{0}\}$ di \mathbf{V} costituito dal solo vettore nullo è un sottospazio, detto sottospazio nullo. L'intero spazio \mathbf{V} è anch'esso un sottospazio. Questi due sottospazi si dicono sottospazi *banali* di \mathbf{V} . Gli altri sottospazi, se esistono, sono quelli interessanti.



In questi esempi, e nel resto del capitolo, identifieremo \mathbb{R}^2 e \mathbb{R}^3 con il piano e lo spazio cartesiano, rispettivamente.

- La retta di equazione $y = x + 1$ nel piano cartesiano non è un sottospazio perché non contiene l'origine $[0, 0]^T$.
- Il primo quadrante del piano cartesiano \mathbb{R}^2 :

$$\{[x, y]^T \in \mathbb{R}^2 : x \geq 0, y \geq 0\}$$

è chiuso rispetto alla somma, ma non rispetto al prodotto per uno scalare (si consideri il prodotto per lo scalare -1). Perciò non è un sottospazio di \mathbb{R}^2 .

- L'unione dei due assi cartesiani

$$\{[x, y]^T \in \mathbb{R}^2 : xy = 0\}$$

è chiuso rispetto al prodotto per uno scalare, ma non rispetto alla somma. Perciò non è un sottospazio di \mathbb{R}^2 .

- I seguenti insiemi sono sottospazi di \mathbb{R}^2 : l'origine, le rette per l'origine, tutto \mathbb{R}^2 . Vedremo in seguito che non ce ne sono altri.

- I seguenti insiemi sono sottospazi di \mathbb{R}^3 : l'origine, le rette per l'origine, i piani per l'origine, tutto \mathbb{R}^3 . Vedremo in seguito che non ce ne sono altri.

Nel capitolo sui sistemi lineari abbiamo introdotto il *nucleo* $\text{Ker}(\mathbf{A})$ di una matrice: si tratta dell'insieme delle soluzioni del sistema lineare omogeneo $\mathbf{Ax} = \mathbf{0}$. Se \mathbf{A} ha n colonne, il nucleo è un sottoinsieme di \mathbb{K}^n :

$$\text{Ker}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{K}^n : \mathbf{Ax} = \mathbf{0}\}$$

PROPOSIZIONE 3.2 (Il nucleo di una matrice è un sottospazio)

Sia \mathbf{A} una matrice con n colonne. Il nucleo $\text{Ker}(\mathbf{A})$ è un sottospazio vettoriale di \mathbb{K}^n .

DIMOSTRAZIONE. Dobbiamo dimostrare che $\mathbf{0}$, il vettore nullo di \mathbb{K}^n , appartiene al nucleo e che il nucleo è chiuso rispetto alla somma e al prodotto per uno scalare.

Siccome

$$\mathbf{A}\mathbf{0} = \mathbf{0}$$

il vettore nullo appartiene al nucleo.

Supponiamo che \mathbf{v} e \mathbf{w} siano due vettori in $\text{Ker}(\mathbf{A})$. Allora

$$\mathbf{A}(\mathbf{v} + \mathbf{w}) = \mathbf{Av} + \mathbf{Aw} = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

quindi anche $\mathbf{v} + \mathbf{w} \in \text{Ker}(\mathbf{A})$.

Supponiamo che $t \in \mathbb{K}$ e $\mathbf{v} \in \text{Ker}(\mathbf{A})$: allora

$$\mathbf{A}(t\mathbf{v}) = t\mathbf{Av} = t\mathbf{0} = \mathbf{0}$$

quindi anche $t\mathbf{v} \in \text{Ker}(\mathbf{A})$.

Esplicitamente, il sottospazio $\mathbf{H} = \text{Ker}(\mathbf{A})$ consiste dei vettori $[x_1, \dots, x_n]^T$ che soddisfano il sistema

$$(3.1) \quad \left\{ \begin{array}{lcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & 0 \\ \vdots & \vdots & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & 0 \end{array} \right.$$

Le equazioni del sistema si dicono *equazioni cartesiane* del sottospazio \mathbf{H} . Dimostremo più avanti (proposizione 9.1) che tutti i sottospazi di \mathbb{K}^n ammettono equazioni cartesiane, cioè sono il nucleo di una matrice con n colonne.



Possiamo dedurre che rette e piani per l'origine sono sottospazi di \mathbb{R}^3 dalla proposizione precedente. Infatti un piano \mathbf{H} per l'origine ha un'equazione omogenea del tipo

$$ax + by + cz = 0$$

Quindi, posto $\mathbf{A} = [a \ b \ c]$, abbiamo $\mathbf{H} = \text{Ker}(\mathbf{A})$, per cui \mathbf{H} è un sottospazio di \mathbb{R}^3 .

Analogamente, le equazioni cartesiane di una retta \mathbf{r} per l'origine sono due equazioni omogenee e, quindi, se \mathbf{A} è la matrice che ha per righe i coefficienti delle due equazioni di \mathbf{r} , $\mathbf{r} = \text{Ker}(\mathbf{A})$.

Prima di proseguire con la teoria, vediamo importanti esempi in cui \mathbf{V} è uno spazio di matrici o di funzioni.



Sia $\mathbb{M} = \mathbb{M}_{\mathbb{K}}(n, n)$ lo spazio vettoriale delle matrici *quadrate* di ordine n a coefficienti in \mathbb{K} . Una matrice $\mathbf{A} = [a_{ij}] \in \mathbb{M}$ è *simmetrica* se $a_{ij} = a_{ji}$ per ogni $1 \leq i < j \leq n$. L'insieme delle matrici simmetriche \mathbb{S} è un sottospazio vettoriale di \mathbb{M} : il lettore lo verifichi direttamente. Per vedere che \mathbb{S} è un sottospazio si può anche ragionare come segue: identifichiamo \mathbb{M} con $\mathbb{K}^{n \times n}$, cioè pensiamo una matrice quadrata come un vettore con $n \times n$ coordinate a_{ij} . L'insieme delle matrici simmetriche è allora l'insieme delle soluzioni del sistema lineare omogeneo

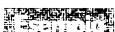
$$a_{ij} - a_{ji} = 0 \quad \text{per ogni } 1 \leq i < j \leq n$$

Abbiamo appena mostrato che l'insieme delle soluzioni di un sistema omogeneo è un sottospazio, per cui l'insieme delle matrici simmetriche è un sottospazio.

Una matrice $\mathbf{A} \in \mathbb{M}$ si dice invece *antisimmetrica* o *emisimmetrica* se $a_{ij} = -a_{ji}$ per ogni $1 \leq i \leq j \leq n$. Anche le matrici antisimmetriche formano un sottospazio vettoriale.



Sempre in $\mathbb{M}_{\mathbb{K}}(n, n)$ consideriamo l'insieme delle matrici invertibili. Tale insieme non è un sottospazio vettoriale. In effetti nessuna delle tre proprietà richieste dalla definizione di sottospazio è verificata: la matrice nulla non è invertibile; la somma di due matrici invertibili può non essere invertibile (per esempio, \mathbf{I} e $-\mathbf{I}$ sono invertibili, ma la loro somma è la matrice nulla che non è invertibile); infine, il prodotto di uno scalare t per una matrice invertibile è una matrice invertibile se e solo se $t \neq 0$.



Un polinomio a coefficienti in \mathbb{K} è una funzione $P(x) : \mathbb{K} \rightarrow \mathbb{K}$ della forma:

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

I coefficienti del polinomio sono gli scalari $a_0, a_1, \dots, a_n \in \mathbb{K}$. Il grado di $P(x)$ è n se $a_n \neq 0$; in generale, è la più alta potenza di x che compare nell'espressione di $P(x)$ con coefficiente non nullo. Il polinomio nullo è quello con tutti i coefficienti nulli: la funzione nulla $P(x) = 0$ per ogni $x \in \mathbb{K}$; il grado del polinomio nullo è per convenzione $-\infty$. Si osservi che il grado del prodotto di due polinomi è sempre la somma dei gradi dei due polinomi.

L'insieme $\mathbb{K}[x]$ dei polinomi è un sottoinsieme dello spazio funzionale $\mathbb{K}^{\mathbb{K}} = \{f : \mathbb{K} \rightarrow \mathbb{K}\}$. Non è solo un sottoinsieme, ma è anche un sottospazio perché la funzione nulla è un polinomio, la somma di due polinomi è un polinomio, così come il prodotto di uno scalare t per un polinomio è un polinomio.

Fissiamo un intero $n \geq 0$. L'insieme $\mathbb{K}[x]_{\leq n}$ dei polinomi $P(x)$ di grado minore o uguale a n :

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

è un sottospazio vettoriale di $\mathbb{K}[x]$. La somma di due polinomi di grado minore o uguale a n si effettua sommando i corrispondenti coefficienti:

$$(a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0) + (b_n x^n + b_{n-1} x^{n-1} + \cdots + b_0) = \\ (a_n + b_n) x^n + (a_{n-1} + b_{n-1}) x^{n-1} + \cdots + (a_0 + b_0)$$

e il prodotto di $P(x)$ per uno scalare t si effettua moltiplicando i coefficienti di $P(x)$ per t . Possiamo allora associare al polinomio $P(x)$ il vettore $[a_n, a_{n-1}, \dots, a_0]^T$ dei suoi $n+1$ coefficienti e ottenere così una corrispondenza biunivoca tra $\mathbb{K}[x]_{\leq n}$ e \mathbb{K}^{n+1} . Tale corrispondenza rispetta le operazioni di somma e di prodotto per uno scalare. Lo spazio $\mathbb{K}[x]_{\leq n}$ è uno spazio di funzioni che, in quanto spazio vettoriale, è identico a \mathbb{K}^{n+1} .



4 Quali dei seguenti sottoinsiemi di \mathbb{R}^3 sono dei sottospazi vettoriali?

- $\mathbf{H}_1 = \{\mathbf{v} = [x, y, z]^T \in \mathbb{R}^3 : z = 0\};$
- $\mathbf{H}_2 = \{\mathbf{v} = [x, y, z]^T \in \mathbb{R}^3 : z \geq 0\};$
- $\mathbf{H}_3 = \{\mathbf{v} = [x, y, z]^T \in \mathbb{R}^3 : xy = 0\};$
- $\mathbf{H}_4 = \{\mathbf{v} = [x, y, z]^T \in \mathbb{R}^3 : x + y + z = y = 0\};$
- $\mathbf{H}_5 = \{\mathbf{v} = [x, y, z]^T \in \mathbb{R}^3 : x, y, z \in \mathbb{Q}\};$
- $\mathbf{H}_6 = \{\mathbf{v} = [x, y, z]^T \in \mathbb{R}^3 : y - z^2 = 0\};$
- $\mathbf{H}_7 = \{\mathbf{v} = [x, y, z]^T \in \mathbb{R}^3 : x + y + z = z - 1 = 0\}$

5 Sia \mathcal{F} lo spazio delle funzioni reali di variabile reale. Si considerino i sottoinsiemi: $\mathbf{H}_1 = \{f \in \mathcal{F} : f(3) = 0\}$, $\mathbf{H}_2 = \{g \in \mathcal{F} : g(2) = g(4)\}$ e $\mathbf{H}_3 = \{h \in \mathcal{F} : h(7) = 1 + h(0)\}$. Quali di questi sottoinsiemi sono sottospazi vettoriali di \mathcal{F} ?

6 Sia $\mathcal{F} = \mathbb{R}^{\mathbb{R}}$ lo spazio vettoriale delle funzioni reali a valori reali. Mostrare che i seguenti sottoinsiemi di \mathcal{F} sono dei sottospazi:

1. l'insieme delle funzioni continue in un punto x_0 ;
2. l'insieme delle funzioni derivabili in un punto x_0 ;
3. l'insieme $\mathcal{C}(\mathbb{R})$ delle funzioni continue in ogni punto $x \in \mathbb{R}$;
4. l'insieme $\mathcal{F}^1(\mathbb{R})$ delle funzioni derivabili in ogni punto $x \in \mathbb{R}$;
5. l'insieme $\mathcal{C}^1(\mathbb{R})$ delle funzioni derivabili con derivata continua;
6. l'insieme delle funzioni integrabili in $[0, 1]$.

7 Sia \mathbb{M} lo spazio vettoriale delle matrici quadrate di ordine n a coefficienti in \mathbb{K} . Mostrare che l'insieme delle matrici triangolari alte e l'insieme delle matrici diagonali sono sottospazi di \mathbb{M} e determinarne delle equazioni come si è fatto nel testo per le matrici simmetriche e antisimmetriche.

8 Una matrice quadrata \mathbf{N} si dice *nilpotente* se una potenza di \mathbf{N} è uguale alla matrice nulla. Sia \mathfrak{N} l'insieme delle matrici nilpotenti di ordine n . Mostrare che \mathfrak{N} è chiuso rispetto al prodotto per uno scalare, ma che \mathfrak{N} non è chiuso rispetto alla somma e quindi non è un sottospazio.

Suggerimento: considerare prima il caso $n = 2$ e cercare gli esempi tra le matrici che hanno tutti gli elementi tranne uno uguale a zero.

 Sia \mathbf{V} lo spazio vettoriale delle successioni a_k di numeri reali. Sia ℓ_1 il sottoinsieme delle successioni a_k tali che

$$\sum_{k=0}^{+\infty} |a_k| < +\infty$$

Mostrare che ℓ_1 è un sottospazio vettoriale di \mathbf{V} .

 Sia $\mathcal{F} = \mathbb{R}^I$ l'insieme delle funzioni reali definite su un intervallo $I \subseteq \mathbb{R}$ e sia $L^1(I)$ il sottoinsieme delle funzioni $f(x)$ il cui modulo $|f(x)|$ è una funzione integrabile in I . Mostrare che $L^1(I)$ è un sottospazio vettoriale di \mathcal{F} .

4 COMBINAZIONI LINEARI

Nello spazio cartesiano la retta

$$(4.1) \quad \begin{cases} x + 2y = 0 \\ y - z = 0 \end{cases}$$

ha equazioni parametriche $[x, y, z]^T = t[-2, 1, 1]^T$ che si ottengono risolvendo il sistema (4.1) rispetto alla variabile libera z . Più in generale, supponiamo di conoscere le equazioni cartesiane di un sottospazio \mathbf{H} di \mathbb{K}^n : questo significa che $\mathbf{H} = \text{Ker}(\mathbf{A})$ è l'insieme delle soluzioni del sistema omogeneo $\mathbf{Ax} = \mathbf{0}$. Abbiamo visto nel capitolo sui sistemi lineari come, risolvendo il sistema lineare $\mathbf{Ax} = \mathbf{0}$, possiamo scrivere ogni soluzione $\mathbf{v} \in \mathbf{H}$ nella forma

$$(4.2) \quad t_1 \mathbf{v}_1 + \cdots + t_{n-r} \mathbf{v}_{n-r}$$

dove $\mathbf{v}_1, \dots, \mathbf{v}_{n-r}$ sono vettori di $\text{Ker}(\mathbf{A})$ e t_1, \dots, t_{n-r} sono degli scalari. Le equazioni (4.2) sono delle equazioni parametriche del sottospazio. Vogliamo ora analizzarne l'espressione. Per questo diamo la definizione di combinazione lineare.

DEFINIZIONE 4.1 (Combinazione lineare)

Si dice che un vettore $\mathbf{v} \in \mathbf{V}$ è *combinazione lineare* dei vettori $\mathbf{v}_1, \dots, \mathbf{v}_d \in \mathbf{V}$ se esistono degli scalari $t_1, \dots, t_d \in \mathbb{K}$ tali che

$$\mathbf{v} = t_1 \mathbf{v}_1 + t_2 \mathbf{v}_2 + \cdots + t_d \mathbf{v}_d$$

Gli scalari t_1, \dots, t_d si dicono *coefficienti* della combinazione lineare.

Si osservi che una combinazione lineare si ottiene combinando le due operazioni fondamentali dello spazio vettoriale: la somma e il prodotto per uno scalare.



In \mathbb{R}^3 il vettore $[-5, 2, 3]^T$ è combinazione lineare di $\mathbf{v}_1 = [-1, 1, 0]^T$ e $\mathbf{v}_2 = [-1, 0, 1]^T$:

$$\begin{bmatrix} -5 \\ 2 \\ 3 \end{bmatrix} = 2 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

Esempio

Una matrice simmetrica di ordine 2 ha la forma

$$\mathbf{A} = \begin{bmatrix} a & b \\ b & c \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Quindi ogni matrice simmetrica di ordine 2 è combinazione lineare delle tre matrici simmetriche

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Esempio

Il polinomio $(1+x)^2$ è combinazione lineare dei polinomi 1, x e x^2 :

$$(1+x)^2 = 1 + 2x + x^2$$

Esempio

La funzione seno iperbolico

$$\text{Sh}(x) = \frac{1}{2}e^x - \frac{1}{2}e^{-x}$$

è una combinazione lineare delle funzioni e^x e e^{-x} .

Esempio

Per il teorema di Rouché-Capelli, dato un sistema lineare omogeneo $\mathbf{Ax} = \mathbf{0}$, esistono vettori $\mathbf{v}_1, \dots, \mathbf{v}_{n-r}$ tali che ogni soluzione del sistema è una combinazione lineare

$$t_1 \mathbf{v}_1 + \cdots + t_{n-r} \mathbf{v}_{n-r}$$

di $\mathbf{v}_1, \dots, \mathbf{v}_{n-r}$.

Esempio

Si può mostrare che le soluzioni dell'equazione differenziale

$$y''(t) + y(t) = 0$$

sono tutte e solo le funzioni della forma

$$y(t) = a \cos(t) + b \sin(t)$$

dove a e b sono costanti reali. Si tratta delle combinazioni lineari delle funzioni $\cos(t)$ e $\sin(t)$ nello spazio vettoriale delle funzioni reali di variabile reale.

PROPOSIZIONE 4.2 (Sottospazio generato da un insieme di vettori)

Siano $\mathbf{v}_1, \dots, \mathbf{v}_d$ vettori di \mathbf{V} .

a) L'insieme

$$\mathcal{L}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d)$$

di tutte le combinazioni lineari di $\mathbf{v}_1, \dots, \mathbf{v}_d$ è un sottospazio vettoriale di \mathbf{V} , che si dice *sottospazio generato* da $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d$.

- b) Se $\mathbf{v}_1, \dots, \mathbf{v}_d$ appartengono ad \mathbf{H} e \mathbf{H} è un sottospazio di \mathbf{V} , allora ogni combinazione lineare di $\mathbf{v}_1, \dots, \mathbf{v}_d$ appartiene ad \mathbf{H} :

$$\mathcal{L}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d) \subseteq \mathbf{H}$$

DIMOSTRAZIONE. Poniamo $\mathbf{W} = \mathcal{L}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d)$. Dobbiamo dimostrare che $\mathbf{0} \in \mathbf{W}$, e che \mathbf{W} è chiuso rispetto alla somma e al prodotto per uno scalare. Per definizione, un vettore appartiene a \mathbf{W} se e solo se è combinazione lineare di $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d$.

Il vettore nullo appartiene a \mathbf{W} poiché

$$\mathbf{0} = 0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_d$$

Supponiamo che \mathbf{v} e \mathbf{w} siano due vettori in \mathbf{W} :

$$\mathbf{v} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_d\mathbf{v}_d, \quad \mathbf{w} = u_1\mathbf{v}_1 + u_2\mathbf{v}_2 + \cdots + u_d\mathbf{v}_d$$

Allora

$$\mathbf{v} + \mathbf{w} = (t_1 + u_1)\mathbf{v}_1 + (t_2 + u_2)\mathbf{v}_2 + \cdots + (t_d + u_d)\mathbf{v}_d$$

quindi anche $\mathbf{v} + \mathbf{w} \in \mathbf{W}$.

Infine supponiamo che $t \in \mathbb{K}$ e $\mathbf{v} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_d\mathbf{v}_d \in \mathbf{W}$. Allora

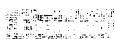
$$t\mathbf{v} = (tt_1)\mathbf{v}_1 + (tt_2)\mathbf{v}_2 + \cdots + (tt_d)\mathbf{v}_d$$

quindi anche $t\mathbf{v} \in \mathbf{W}$.

Mostriamo b): supponiamo che \mathbf{H} sia un sottospazio e che $\mathbf{v}_1, \dots, \mathbf{v}_d \in \mathbf{H}$. La tesi è che ogni combinazione lineare

$$t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_d\mathbf{v}_d$$

appartenga ad \mathbf{H} . Per questo osserviamo che ogni addendo $t_k\mathbf{v}_k$ appartiene ad \mathbf{H} perché $\mathbf{v}_k \in \mathbf{H}$ e \mathbf{H} è chiuso rispetto al prodotto per uno scalare. Ma \mathbf{H} è chiuso anche rispetto alla somma per cui contiene la somma \mathbf{v} di tutti gli addendi $t_k\mathbf{v}_k$.



Se $\mathbf{v} \neq 0$, il sottospazio $\mathcal{L}(\mathbf{v})$ si dice *retta* generata da \mathbf{v} : consiste dei vettori della forma $t\mathbf{v}$ al variare di $t \in \mathbb{K}$. Per questo si scrive anche $\mathbb{K}\mathbf{v}$ al posto di $\mathcal{L}(\mathbf{v})$. Intuitivamente si tratta della retta per l'origine parallela a \mathbf{v} . Più in generale, un altro simbolo per denotare lo spazio generato da $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d$ è

$$\mathbb{K}\mathbf{v}_1 + \mathbb{K}\mathbf{v}_2 + \cdots + \mathbb{K}\mathbf{v}_d$$

Questo simbolo ha il vantaggio di ricordare che i vettori del sottospazio generato da $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d$ si ottengono come combinazioni lineari di $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d$. La retta $\mathbb{R}\mathbf{i}$ generata da $\mathbf{i} = [1, 0, 0]^T$ in \mathbb{R}^3 consiste dei vettori della forma

$$\mathbf{v} = t\mathbf{i} = [t, 0, 0]^T \quad (t \in \mathbb{R})$$

ed è perciò l'asse delle ascisse. Il sottospazio $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j}$ di \mathbb{R}^3 generato da $\mathbf{i} = [1, 0, 0]^T$ e $\mathbf{j} = [0, 1, 0]^T$ consiste dei vettori della forma

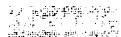
$$\mathbf{v} = t_1\mathbf{i} + t_2\mathbf{j} = [t_1, t_2, 0]^T$$

ed è perciò il piano xy (il piano di equazione $z = 0$).

DEFINIZIONE 4.3 (Insieme di generatori)

Si dice che $\mathbf{v}_1, \dots, \mathbf{v}_d$ generano \mathbf{H} , o che sono un *insieme di generatori* di \mathbf{H} , se \mathbf{H} è il sottospazio di \mathbf{V} generato da $\mathbf{v}_1, \dots, \mathbf{v}_d$: $\mathbf{H} = \mathcal{L}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d)$.

In altre parole, $\mathbf{v}_1, \dots, \mathbf{v}_d$ generano \mathbf{H} se appartengono a \mathbf{H} e ogni vettore di \mathbf{H} è combinazione lineare di $\mathbf{v}_1, \dots, \mathbf{v}_d$.



I vettori $\mathbf{v}_1 = [-1, 1, 0]^T$ e $\mathbf{v}_2 = [-1, 0, 1]^T$ generano il piano \mathbf{H} di equazione $x + y + z = 0$. Infatti $x = -y - z$ se e solo se, posto $t_1 = y$ e $t_2 = z$, si ha

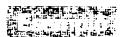
$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = t_1 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2$$



Le tre matrici simmetriche

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

generano lo spazio delle matrici simmetriche di ordine 2.



Consideriamo gli $n + 1$ monomi $x^n, x^{n-1}, \dots, x, 1 = x^0$ di grado $\leq n$ nello spazio vettoriale delle funzioni $f : \mathbb{K} \rightarrow \mathbb{K}$. Il sottospazio $\mathcal{L}(x^n, x^{n-1}, \dots, x, 1)$ generato da questi monomi consiste delle combinazioni lineari

$$t_1x^n + t_2x^{n-1} + \cdots + t_nx + t_{n+1}$$

ed è, quindi, lo spazio $\mathbb{K}[x]_{\leq n}$ dei polinomi di grado $\leq n$ a coefficienti in \mathbb{K} .

Il prossimo corollario contiene un utile espeditivo tecnico: per mostrare che un sottospazio \mathbf{H} contenga un altro sottospazio \mathbf{W} , è sufficiente verificare che \mathbf{H} contenga un insieme di generatori di \mathbf{W} .

COROLLARIO 4.4 Siano \mathbf{W} e \mathbf{H} sottospazi di \mathbf{V} . Se \mathbf{H} contiene un insieme di generatori di \mathbf{W} , allora \mathbf{H} contiene \mathbf{W} .

DIMOSTRAZIONE. Supponiamo che $\mathbf{v}_1, \dots, \mathbf{v}_d$ sia un insieme di generatori di \mathbf{W} contenuto in \mathbf{H} . Allora $\mathbf{W} = \mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_d)$ è contenuto in \mathbf{H} per la proposizione 4.2. ■

COROLLARIO 4.5 Un vettore \mathbf{w} è combinazione lineare di $\mathbf{v}_1, \dots, \mathbf{v}_d$ se e solo se

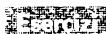
$$\mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_d) = \mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_d, \mathbf{w})$$

DIMOSTRAZIONE. Se \mathbf{w} è combinazione lineare di $\mathbf{v}_1, \dots, \mathbf{v}_d$, allora $\mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_d)$ contiene \mathbf{w} e, quindi, contiene i generatori $\mathbf{v}_1, \dots, \mathbf{v}_d, \mathbf{w}$ di $\mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_d, \mathbf{w})$. Perciò

$$\mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_d, \mathbf{w}) \subseteq \mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_d)$$

Evidentemente vale anche l'inclusione opposta, per cui i due sottospazi sono uguali.

Viceversa, se $\mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_d) = \mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_d, \mathbf{w})$, allora $\mathbf{w} \in \mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_d)$ e questo significa proprio che \mathbf{w} è combinazione lineare di $\mathbf{v}_1, \dots, \mathbf{v}_d$. ■



(1) Scrivere il vettore $\mathbf{v} = [1, -2, 5]^T$ come combinazione lineare dei vettori $\mathbf{v}_1 = [1, 1, 1]^T$, $\mathbf{v}_2 = [1, 2, 3]^T$, $\mathbf{v}_3 = [2, -1, 1]^T$.

(2) Determinare per quali valori del parametro $k \in \mathbb{R}$, il vettore $\mathbf{v} = [-2, 1, k]^T$ si scrive come combinazione lineare dei vettori $\mathbf{v}_1 = [0, 3, -2]^T$ e $\mathbf{v}_2 = [-1, 2, 5]^T$.

(3) Scrivere il polinomio $(1+x)^3$ e la sua derivata $3(1+x)^2$ come combinazione lineare dei polinomi 1 , x , x^2 e x^3 .

(4) Scrivere il polinomio $\mathbf{p}(x) = -x^2 + 7$ come combinazione lineare dei vettori $\mathbf{p}_1(x) = x^2 - 2x + 5$, $\mathbf{p}_2(x) = 2x^2$, $\mathbf{p}_3(x) = x + 1$.

(5) Mostrare che i vettori $\mathbf{v}_1 = [1, 2, 3]^T$, $\mathbf{v}_2 = [0, 1, 2]^T$, $\mathbf{v}_3 = [0, 0, 1]^T$ sono un insieme di generatori di \mathbb{R}^3 .

(6) Spiegare perché $\mathcal{L}(\mathbf{v}_1, \mathbf{v}_2) = \mathcal{L}(\mathbf{v}_2, \mathbf{v}_1)$. Generalizzare a un numero arbitrario di vettori: il sottospazio $\mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_d)$ non dipende dall'ordine in cui si prendono i vettori $\mathbf{v}_1, \dots, \mathbf{v}_d$. *Suggerimento:* Una possibile soluzione è l'osservazione che $\mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_d)$ è il più piccolo sottospazio che contenga $\mathbf{v}_1, \dots, \mathbf{v}_d$; questa proprietà è indipendente dall'ordine dei vettori.

(7) Spiegare perché $\mathcal{L}(\mathbf{v}, \mathbf{v}, \mathbf{v}) = \mathcal{L}(\mathbf{v})$ e $\mathcal{L}(\mathbf{v}, \mathbf{0}) = \mathcal{L}(\mathbf{v})$.

(8) Mostrare che per ogni \mathbf{v} e \mathbf{w}

$$\mathcal{L}(\mathbf{v}, \mathbf{w}) = \mathcal{L}(\mathbf{v}, \mathbf{v} + \mathbf{w})$$

(9) Scrivere il vettore $\begin{bmatrix} i \\ 1+i \end{bmatrix} \in \mathbb{C}^2$ come combinazione lineare (a coefficienti complessi!) dei vettori della base canonica di \mathbb{C}^2 .

(10) Scrivere la funzione costante 1 come combinazione lineare delle funzioni $\cos^2(x)$ e $\sin^2(x)$.

Scrivere la funzione coseno iperbolico come combinazione lineare di funzioni esponenziali.

Scrivere la funzione $\cos(x)$ come combinazione lineare a coefficienti complessi di funzioni esponenziali complesse

Suggerimento: $\exp(ix) = \cos(x) + i\sin(x)$ per ogni $x \in \mathbb{R}$.

Mostrare che l'insieme delle soluzioni dell'equazione differenziale $y'(x) = y(x)$ è $\mathbb{R}e^x$, la retta generata dalla funzione esponenziale nello spazio vettoriale delle funzioni $f(x) : \mathbb{R} \rightarrow \mathbb{R}$.

Mostrare che il sottospazio di $\mathbb{M}(2, 2)$ formato dalle matrici triangolari alte è generato dalle matrici $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ e $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$.

Sia \mathbf{V} uno spazio vettoriale e sia \mathbf{S} un sottoinsieme (non necessariamente finito) di \mathbf{V} . Mostrare che l'insieme di tutte le combinazioni lineari di vettori di \mathbf{S} è un sottospazio di \mathbf{V} , ed è il più piccolo sottospazio di \mathbf{V} che contenga \mathbf{S} . Sia $\mathbb{R}^{\mathbb{N}}$ lo spazio vettoriale delle successioni di numeri reali. Per ogni $n \in \mathbb{N}$ sia \mathbf{e}_n la successione che è identicamente nulla tranne per l' n -esima componente, che è uguale a 1. Si mostri che il più piccolo sottospazio di $\mathbb{R}^{\mathbb{N}}$ che contenga tutte le successioni \mathbf{e}_n è il sottospazio delle successioni finite (cioè delle successioni $\mathbf{a} = \{a_k\}$ tali che $a_k = 0$ per k sufficientemente grande).

5 INDIPENDENZA LINEARE

Se conosciamo un insieme di generatori $\mathbf{v}_1, \dots, \mathbf{v}_d$ di uno spazio vettoriale \mathbf{V} , possiamo esprimere i vettori di \mathbf{V} in funzione di d parametri t_1, \dots, t_d . Può succedere però che uno *stesso* vettore corrisponda a valori *distinti* dei parametri. Per esempio, consideriamo in \mathbb{R}^2 i vettori $\mathbf{v}_1 = [1, 2]^T$ e $\mathbf{v}_2 = [2, 4]^T$. Il vettore $[7, 14]^T$ si può scrivere in modi diversi come combinazione lineare di \mathbf{v}_1 e \mathbf{v}_2 , per esempio

$$\begin{bmatrix} 7 \\ 14 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 2 \begin{bmatrix} 2 \\ 4 \end{bmatrix} = 7 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 0 \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

Questo problema sorge perché $\mathbf{v}_2 = [2, 4]^T$ è uguale a 2 volte $\mathbf{v}_1 = [1, 2]^T$, ovvero tra i due vettori intercorre la relazione lineare

$$2\mathbf{v}_1 + (-1)\mathbf{v}_2 = \mathbf{0}$$

Siamo così indotti a formulare la seguente definizione:

DEFINIZIONE 5.1 (Insiemi di vettori linearmente dipendenti [indipendenti])

Si dice che un insieme $\{\mathbf{v}_1, \dots, \mathbf{v}_d\}$ di vettori di \mathbf{V} è *linearmente dipendente* se esistono degli scalari t_1, \dots, t_d non tutti nulli tali che

$$(5.1) \quad t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_d\mathbf{v}_d = \mathbf{0}$$

Si dice che l'insieme $\{\mathbf{v}_1, \dots, \mathbf{v}_d\}$ è *linearmente indipendente* se non è linearmente dipendente.

Si dice che (5.1) è una *relazione lineare* tra i vettori $\mathbf{v}_1, \dots, \mathbf{v}_d$. Si osservi che l'insieme $\{\mathbf{v}_1, \dots, \mathbf{v}_d\}$ è linearmente indipendente se l'uguaglianza

$$t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_d\mathbf{v}_d = \mathbf{0}$$

implica $t_1 = t_2 = \cdots = t_d = 0$. Questo significa che tra i vettori $\mathbf{v}_1, \dots, \mathbf{v}_d$ non c'è alcuna relazione lineare, tranne quella banale per la quale $t_1 = t_2 = \cdots = t_d = 0$.

OSSERVAZIONE Per abuso di linguaggio, si dice che i vettori $\mathbf{v}_1, \dots, \mathbf{v}_d$ sono linearmente (in)dipendenti se l'insieme $\{\mathbf{v}_1, \dots, \mathbf{v}_d\}$ è linearmente (in)dipendente. Si tratta di un abuso di linguaggio perché la dipendenza lineare non è una proprietà dei singoli vettori (come, per esempio, sarebbero la proprietà di avere n componenti oppure componenti complesse oppure l'ultima componente nulla), ma dell'insieme che essi formano. Dire che i vettori sono linearmente indipendenti non significa che ciascuno di essi ha la proprietà di essere linearmente indipendente, significa invece che tra di essi non c'è nessuna relazione di dipendenza lineare.

Per esempio, l'insieme $\{\mathbf{v}_1, \mathbf{v}_2\}$ è linearmente dipendente perché esiste la relazione lineare non banale:

$$2\mathbf{v}_1 + (-1)\mathbf{v}_2 = \mathbf{0}$$

Per esempio, l'insieme $\{\mathbf{v}_1, \mathbf{v}_2\}$ è linearmente indipendente perché tra i due vettori intercorre la relazione lineare non banale:

$$t_1\mathbf{v}_1 + t_2\mathbf{v}_2 = \mathbf{0}$$

allora $t_1 + 2t_2 = 2t_1 + 3t_2 = 0$. Questo sistema ha solo la soluzione banale $t_1 = t_2 = 0$, quindi i due vettori $\mathbf{v}_1, \mathbf{v}_2$ sono indipendenti.

Per esempio, l'insieme $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ è linearmente dipendente perché esiste la relazione lineare non banale:

Ci chiediamo se i vettori $\mathbf{v}_1 = [1, 2, 3]^T$, $\mathbf{v}_2 = [3, 2, 1]^T$, $\mathbf{v}_3 = [1, -2, -5]^T$ siano linearmente dipendenti. Questo significa chiedersi se esiste tra di essi una relazione lineare

$$t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + t_3\mathbf{v}_3 = \mathbf{0}$$

con $[t_1, t_2, t_3]^T \neq [0, 0, 0]^T$. Se introduciamo la matrice $\mathbf{A} = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3]$ che ha per colonne i tre vettori, la relazione cercata si riscrive

$$\mathbf{A} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Quindi i tre vettori sono linearmente dipendenti se e solo se il nucleo di \mathbf{A} contiene un vettore non nullo $[t_1, t_2, t_3]^T$. Ora, risolvendo il sistema $\mathbf{Ax} = \mathbf{0}$ si trova che il nucleo di \mathbf{A} contiene vettori non nulli, per esempio $[2, -1, 1]^T$, che corrisponde alla relazione lineare:

$$2\mathbf{v}_1 - \mathbf{v}_2 + \mathbf{v}_3 = \mathbf{0}$$

Concludiamo che i tre vettori sono linearmente dipendenti.

Esempio

Nel capitolo 1, proposizione 7.5, abbiamo mostrato che tre vettori dello spazio cartesiano \mathbb{R}^3 sono linearmente indipendenti se e solo se il loro prodotto misto è diverso da zero.

Esempio

Nello spazio vettoriale dei polinomi di grado ≤ 2 , i polinomi $(1+x)^2$, 1 , x e x^2 sono linearmente dipendenti perché:

$$(1+x)^2 + (-1)1 + (-2)x + (-1)x^2 = 0 \quad \text{per ogni } x \in \mathbb{K}$$

Esempio

Nello spazio vettoriale delle funzioni reali di una variabile reale, siano: \mathbf{v}_1 la funzione costante 1 , \mathbf{v}_2 la funzione $\cos(x)$ e \mathbf{v}_3 la funzione $\sin(x)$. Vogliamo mostrare che l'insieme $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ è linearmente indipendente. Supponiamo infatti che

$$t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + t_3\mathbf{v}_3 = \mathbf{0}$$

Questo significa che

$$t_1 + t_2 \cos(x) + t_3 \sin(x) = 0 \quad \text{per ogni } x \in \mathbb{R}$$

Sostituendo a x i tre valori $x = 0$, $x = \pi/2$ e $x = \pi$, si vede che t_1, t_2, t_3 devono soddisfare il sistema lineare

$$\begin{cases} t_1 + t_2 = 0 \\ t_1 + t_3 = 0 \\ t_1 - t_2 = 0 \end{cases}$$

Questo sistema ha solo la soluzione nulla, quindi $t_1 = t_2 = t_3 = 0$. Concludiamo che le tre funzioni 1 , $\cos(x)$ e $\sin(x)$ sono linearmente indipendenti.

Esempio

Se \mathbf{w} è combinazione lineare di $\mathbf{v}_1, \dots, \mathbf{v}_n$, allora l'insieme $\{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{w}\}$ è linearmente dipendente. Infatti dall'uguaglianza

$$\mathbf{w} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_n\mathbf{v}_n$$

segue

$$t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_n\mathbf{v}_n + (-1)\mathbf{w} = \mathbf{0}$$

che è una relazione di dipendenza lineare, certamente non banale perché l'ultimo coefficiente (-1) non è nullo.

È vero anche il viceversa dell'esempio precedente: un insieme di vettori è linearmente dipendente se e solo se uno di essi è combinazione lineare degli altri. Precisamente:

PROPOSIZIONE 5.2

- a) L'insieme costituito dal solo vettore \mathbf{v} è linearmente indipendente se e solo se $\mathbf{v} \neq 0$.

- b) Se $d \geq 2$, l'insieme $\{\mathbf{v}_1, \dots, \mathbf{v}_d\}$ è linearmente indipendente se e solo se $\{\mathbf{v}_1, \dots, \mathbf{v}_{d-1}\}$ è linearmente indipendente e \mathbf{v}_d non è combinazione lineare di $\mathbf{v}_1, \dots, \mathbf{v}_{d-1}$.

DIMOSTRAZIONE. Per mostrare a) basta mostrare che l'insieme $\{\mathbf{v}\}$ è linearmente dipendente se e solo se $\mathbf{v} = \mathbf{0}$.

Supponiamo che $\{\mathbf{v}\}$ sia linearmente dipendente: esiste uno scalare non nullo t tale che $t\mathbf{v} = \mathbf{0}$. Allora per la legge di annullamento del prodotto per uno scalare \mathbf{v} è il vettore nullo. Viceversa, il vettore nullo costituisce un insieme linearmente dipendente perché $1\mathbf{0} = \mathbf{0}$ e lo scalare 1 è non nullo.

Mostriamo ora che un insieme $\{\mathbf{v}_1, \dots, \mathbf{v}_d\}$ con $d \geq 2$ è linearmente indipendente se e solo se $\mathbf{v}_1, \dots, \mathbf{v}_{d-1}$ sono linearmente indipendenti e \mathbf{v}_d non è combinazione lineare di $\mathbf{v}_1, \dots, \mathbf{v}_{d-1}$.

Supponiamo dapprima che $\mathbf{v}_1, \dots, \mathbf{v}_d$ siano linearmente indipendenti. Siccome ogni relazione di dipendenza lineare tra $\mathbf{v}_1, \dots, \mathbf{v}_{d-1}$ è anche una relazione di dipendenza tra $\mathbf{v}_1, \dots, \mathbf{v}_d$, i vettori $\mathbf{v}_1, \dots, \mathbf{v}_{d-1}$ sono linearmente indipendenti: formalmente, se

$$t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_{d-1}\mathbf{v}_{d-1} = \mathbf{0}$$

allora anche

$$t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_{d-1}\mathbf{v}_{d-1} + t_d\mathbf{v}_d = \mathbf{0}$$

Siccome $\mathbf{v}_1, \dots, \mathbf{v}_d$ sono linearmente indipendenti, concludiamo che $t_1 = t_2 = \dots = t_{d-1} = 0$. Quindi $\mathbf{v}_1, \dots, \mathbf{v}_{d-1}$ sono linearmente indipendenti.

Mostriamo ora che \mathbf{v}_d non è combinazione lineare di $\mathbf{v}_1, \dots, \mathbf{v}_{d-1}$: altrimenti

$$\mathbf{v}_d = t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_{d-1}\mathbf{v}_{d-1}$$

e quindi

$$t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_{d-1}\mathbf{v}_{d-1} + (-1)\mathbf{v}_d = \mathbf{0}$$

Quest'ultima è una relazione di dipendenza lineare non banale tra $\mathbf{v}_1, \dots, \mathbf{v}_d$, il che è assurdo perché $\mathbf{v}_1, \dots, \mathbf{v}_d$ sono linearmente indipendenti.

Abbiamo mostrato che, se $\mathbf{v}_1, \dots, \mathbf{v}_d$ sono linearmente indipendenti, allora $\mathbf{v}_1, \dots, \mathbf{v}_{d-1}$ sono linearmente indipendenti e \mathbf{v}_d non è combinazione lineare di $\mathbf{v}_1, \dots, \mathbf{v}_{d-1}$.

Viceversa, supponiamo che $\mathbf{v}_1, \dots, \mathbf{v}_{d-1}$ siano linearmente indipendenti e che \mathbf{v}_d non sia combinazione lineare di $\mathbf{v}_1, \dots, \mathbf{v}_{d-1}$. Dobbiamo mostrare che $\mathbf{v}_1, \dots, \mathbf{v}_d$ sono indipendenti. Quindi supponiamo

$$(5.2) \quad t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_{d-1}\mathbf{v}_{d-1} + t_d\mathbf{v}_d = \mathbf{0}$$

Se $t_d \neq 0$, dividendo l'equazione precedente per t_d otteniamo

$$\mathbf{v}_d = -\frac{t_1}{t_d}\mathbf{v}_1 - \frac{t_2}{t_d}\mathbf{v}_2 - \cdots - \frac{t_{d-1}}{t_d}\mathbf{v}_{d-1}$$

il che mostra che \mathbf{v}_d è combinazione lineare di $\mathbf{v}_1, \dots, \mathbf{v}_{d-1}$, in contraddizione con le ipotesi. Quindi $t_d = 0$ e allora l'equazione (5.2) si riduce a

$$t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_{d-1}\mathbf{v}_{d-1} = \mathbf{0}$$

Ora dall'ipotesi che $\mathbf{v}_1, \dots, \mathbf{v}_{d-1}$ siano linearmente indipendenti segue che $t_1 = t_2 = \dots = t_{d-1} = 0$. Siccome abbiamo già mostrato che $t_d = 0$, possiamo concludere che $\mathbf{v}_1, \dots, \mathbf{v}_d$ sono linearmente indipendenti.



Prendendo $d = 2$ nella proposizione precedente, otteniamo che due vettori \mathbf{v}_1 e \mathbf{v}_2 sono linearmente dipendenti se e solo se $\mathbf{v}_1 = \mathbf{0}$ oppure $\mathbf{v}_2 = t\mathbf{v}_1$ è un multiplo scalare di \mathbf{v}_1 . In particolare, due vettori nello spazio cartesiano sono linearmente dipendenti se e solo se sono paralleli.



Mostrare che i vettori $[a, b]^T$ e $[c, d]^T$ sono linearmente indipendenti se e solo se $ad - bc \neq 0$.

Per quali valori dei parametri a, b, c l'insieme

$$\{[1, 0, 0]^T, [1, 1, 0]^T, [a, b, c]^T\}$$

è linearmente dipendente?

Stabilire se i seguenti vettori di \mathbb{R}^4 sono linearmente indipendenti

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ -2 \\ -1 \\ 2 \end{bmatrix}.$$

Determinare se i seguenti vettori di $\mathbb{M}_{\mathbb{K}}(3, 2)$ sono linearmente indipendenti

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 3 \\ 2 & 1 \\ 1 & 4 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} -1 & 0 \\ 3 & 5 \\ -1 & 6 \end{bmatrix}, \quad \mathbf{A}_3 = \begin{bmatrix} 1 & 1 \\ -1 & -1 \\ 1 & -1 \end{bmatrix}.$$

Siano $\mathbf{u}, \mathbf{v}, \mathbf{w}$ tre vettori linearmente indipendenti di uno spazio vettoriale \mathbf{V} . Si mostri che i vettori $\mathbf{p} = \mathbf{u} + \mathbf{v}$, $\mathbf{r} = \mathbf{u} - \mathbf{v}$ e $\mathbf{s} = \mathbf{u} - 2\mathbf{v} + \mathbf{w}$ sono linearmente indipendenti.

Supponiamo che $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ sia un insieme linearmente dipendente. Allora qualunque insieme di vettori \mathcal{S} che contenga $\mathbf{v}_1, \dots, \mathbf{v}_r$ è un insieme di vettori linearmente dipendenti.

Sia $\mathcal{S} = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ un insieme di vettori linearmente indipendenti. Allora ogni sottoinsieme di \mathcal{S} è un insieme di vettori linearmente indipendenti.

Spiegare perché il fatto che dei vettori siano linearmente indipendenti o meno non dipende dall'ordine in cui sono presi.



Mostrare che $\mathbf{v}_1, \dots, \mathbf{v}_d$ sono linearmente indipendenti se e solo se

$$\{0\} \subsetneq \mathcal{L}(\mathbf{v}_1) \subsetneq \mathcal{L}(\mathbf{v}_1, \mathbf{v}_2) \subsetneq \dots \subsetneq \mathcal{L}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d).$$

Il simbolo di inclusione stretta \subsetneq significa *contenuto in e non uguale a*.

Suggerimento: combinare la proposizione 5.2 con il corollario 4.5.

Mostrare che i vettori $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ e $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ sono linearmente dipendenti nello spazio vettoriale \mathbb{C}^2 .

Nello spazio vettoriale delle funzioni reali di una variabile reale, siano: \mathbf{v}_1 la funzione costante 1, \mathbf{v}_2 la funzione la funzione $\cos^2(x)$, \mathbf{v}_3 la funzione $\sin^2(x)$. Mostrare che \mathbf{v}_1 , \mathbf{v}_2 e \mathbf{v}_3 sono linearmente dipendenti.

Sia $f(x) : \mathbb{R} \rightarrow \mathbb{R}$ una funzione derivabile in ogni $x \in \mathbb{R}$. Si dimostri che la funzione f e la sua derivata f' sono linearmente dipendenti se e solo se esistono due costanti reali c e λ tali che $f(x) = ce^{\lambda x}$.

6 BASI E DIMENSIONE

In questo paragrafo introduciamo il concetto fondamentale di *base* di uno spazio vettoriale: una base sta a uno spazio vettoriale come un sistema di riferimento sta allo spazio euclideo, perché fissata una base è possibile identificare gli elementi dello spazio vettoriale astratto mediante delle coordinate.

DEFINIZIONE 6.1 (Base di uno spazio vettoriale)

Sia \mathbf{V} uno spazio vettoriale sul campo \mathbb{K} . Un insieme ordinato $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ di vettori di \mathbf{V} si dice *base* di \mathbf{V} se ogni vettore \mathbf{v} di \mathbf{V} si scrive in uno e un solo modo come combinazione lineare

$$\mathbf{v} = t_1 \mathbf{v}_1 + t_2 \mathbf{v}_2 + \cdots + t_n \mathbf{v}_n$$

di $\mathbf{v}_1, \dots, \mathbf{v}_n$.

Più esplicitamente: $\mathbf{v}_1, \dots, \mathbf{v}_n$ è una base di \mathbf{V} se e solo se

a) per ogni \mathbf{v} in \mathbf{V} esistono degli scalari $t_1, \dots, t_n \in \mathbb{K}$ tali che

$$\mathbf{v} = t_1 \mathbf{v}_1 + t_2 \mathbf{v}_2 + \cdots + t_n \mathbf{v}_n$$

b) dall'uguaglianza

$$t_1 \mathbf{v}_1 + \cdots + t_n \mathbf{v}_n = u_1 \mathbf{v}_1 + \cdots + u_n \mathbf{v}_n$$

segue $t_1 = u_1, \dots, t_n = u_n$.

Gli scalari t_1, \dots, t_n sono quindi univocamente determinati da \mathbf{v} e dalla base $\mathbf{v}_1, \dots, \mathbf{v}_n$ e si chiamano *componenti* o *coordinate* del vettore \mathbf{v} rispetto alla base $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$.

Più formalmente, dato un insieme ordinato di vettori $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$, possiamo considerare la funzione $\mathcal{P} : \mathbb{K}^n \rightarrow \mathbf{V}$ definita dalla formula

$$\mathcal{P}([t_1, t_2, \dots, t_n]^T) = t_1 \mathbf{v}_1 + t_2 \mathbf{v}_2 + \cdots + t_n \mathbf{v}_n$$

Quindi \mathcal{P} associa alla n -upla dei parametri t_i la corrispondente combinazione lineare di $\mathbf{v}_1, \dots, \mathbf{v}_n$. La lettera \mathcal{P} dovrebbe ricordare il senso di quello che stiamo facendo: parametrizzare con le coordinate t_i i vettori di \mathbf{V} . Per definizione, l'insieme \mathcal{B} è una base di \mathbf{V} se \mathcal{P} è una corrispondenza biunivoca, cioè iniettiva (a valori distinti dei parametri corrispondono vettori distinti) e suriettiva (ogni vettore corrisponde a una

n -upla di parametri). La funzione inversa $\mathcal{P}^{-1} : \mathbf{V} \rightarrow \mathbb{K}^n$ associa a un vettore \mathbf{v} la n -upla delle sue componenti rispetto alla base \mathcal{B} .

OSSERVAZIONE È necessario richiedere che i vettori della base siano presi con un ordine specifico per poter determinare il vettore delle coordinate, distinguendo la prima coordinata dalla seconda e così via. Esattamente come quando si fissano le coordinate cartesiane nel piano il ruolo dell'asse x e dell'asse y non può essere scambiato.

PROPOSIZIONE 6.2 Sia \mathbf{V} uno spazio vettoriale. I vettori $\mathbf{v}_1, \dots, \mathbf{v}_n$ formano una base di \mathbf{V} se e solo se

- i) $\mathbf{v}_1, \dots, \mathbf{v}_n$ sono un insieme di generatori di \mathbf{V} ;
- ii) $\mathbf{v}_1, \dots, \mathbf{v}_n$ sono linearmente indipendenti.

DIMOSTRAZIONE. Supponiamo che $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ sia una base di \mathbf{V} . Siccome per ogni \mathbf{v} in \mathbf{V} esistono degli scalari t_1, \dots, t_n tali che

$$\mathbf{v} = t_1 \mathbf{v}_1 + t_2 \mathbf{v}_2 + \cdots + t_n \mathbf{v}_n$$

i vettori $\mathbf{v}_1, \dots, \mathbf{v}_n$ sono un insieme di generatori di \mathbf{V} .

Se

$$t_1 \mathbf{v}_1 + \cdots + t_n \mathbf{v}_n = \mathbf{0}$$

confrontando questa espressione del vettore nullo con

$$0 \mathbf{v}_1 + \cdots + 0 \mathbf{v}_n = \mathbf{0}$$

concludiamo che $t_1 = t_2 = \cdots = t_n = 0$ per l'unicità delle coordinate del vettore nullo. Quindi $\mathbf{v}_1, \dots, \mathbf{v}_n$ sono linearmente indipendenti.

Viceversa, supponiamo che $\mathbf{v}_1, \dots, \mathbf{v}_n$ siano un insieme di generatori di \mathbf{V} e siano linearmente indipendenti. Allora ogni vettore di \mathbf{V} è combinazione lineare di $\mathbf{v}_1, \dots, \mathbf{v}_n$. Se

$$t_1 \mathbf{v}_1 + \cdots + t_n \mathbf{v}_n = u_1 \mathbf{v}_1 + \cdots + u_n \mathbf{v}_n$$

allora sottraendo il secondo membro dell'equazione al primo troviamo

$$(t_1 - u_1) \mathbf{v}_1 + \cdots + (t_n - u_n) \mathbf{v}_n = \mathbf{0}$$

Siccome i vettori $\mathbf{v}_1, \dots, \mathbf{v}_n$ sono linearmente indipendenti, l'ultima equazione implica $t_1 - u_1 = t_2 - u_2 = \cdots = t_n - u_n = 0$, ovvero i coefficienti della combinazione lineare sono univocamente determinati e i vettori $\mathbf{v}_1, \dots, \mathbf{v}_n$ formano una base. ■

Il concetto di base è fondamentale perché consente di parametrizzare i vettori di uno spazio vettoriale. Data l'importanza della nozione, diamo numerosi esempi.

Esempio**Base canonica di \mathbb{R}^3**

In \mathbb{R}^3 definiamo come di consuetudine

$$\mathbf{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Allora $[x, y, z]^T = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ per ogni vettore $[x, y, z]^T$ di \mathbb{R}^3 . Siccome due vettori di \mathbb{R}^3 sono uguali se e solo se le loro tre componenti sono uguali, ne deduciamo che ogni vettore di \mathbb{R}^3 si scrive in uno e un sol modo come combinazione lineare dei tre vettori $\mathbf{i}, \mathbf{j}, \mathbf{k}$. Quindi $\mathbf{i}, \mathbf{j}, \mathbf{k}$ formano una base di \mathbb{R}^3 che si chiama *base canonica*.

Esempio**Base canonica di \mathbb{K}^n**

In \mathbb{K}^n gli n vettori:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

formano la *base canonica* di \mathbb{K}^n : il vettore \mathbf{e}_i ha tutte le componenti nulle tranne la i -esima che è 1. Un arbitrario vettore $\mathbf{v} = [x_1, \dots, x_n]^T$ di \mathbb{K}^n si può scrivere in uno e un sol modo come *combinazione lineare* dei vettori della base canonica perché

$$[x_1, \dots, x_n]^T = x_1\mathbf{e}_1 + \dots + x_n\mathbf{e}_n$$

Quindi le coordinate di $\mathbf{v} = [x_1, \dots, x_n]^T$ rispetto alla base canonica sono proprio x_1, \dots, x_n : per questo la base $\mathbf{e}_1, \dots, \mathbf{e}_n$ si dice *canonica*.

Esempio**Base canonica di $\mathbb{M}_{\mathbb{K}}(m, n)$**

Analogamente al caso di \mathbb{K}^n , anche lo spazio delle matrici $\mathbb{M}_{\mathbb{K}}(m, n)$ ha una base canonica, rispetto alla quale le coordinate della matrice $[a_{ij}]$ sono gli scalari a_{ij} . La base canonica consiste delle matrici \mathbf{E}_{ij} che hanno l'elemento di posto (i, j) uguale a uno e tutti gli altri elementi uguali a zero: l'indice i varia tra 1 e m , l'indice j tra 1 e n , quindi la base consiste di $m \times n$ matrici. Per esempio, la base canonica di $\mathbb{M}_{\mathbb{K}}(2, 2)$ consiste delle quattro matrici:

$$\mathbf{E}_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{E}_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \mathbf{E}_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \mathbf{E}_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Esempio**Base canonica dello spazio dei polinomi $\mathbb{K}[x]_{\leq n}$**

Per definizione l'insieme

$$\{x^n, x^{n-1}, x^{n-2}, \dots, x, x^0 = 1\}$$

dei monomi di grado $\leq n$ genera l'insieme $\mathbb{K}[x]_{\leq n}$ dei polinomi di grado $\leq n$ a coefficienti in \mathbb{K} . Le funzioni x^k sono anche linearmente indipendenti: se

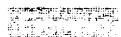
$$t_1 x^n + t_2 x^{n-1} + \cdots + t_n x + t_{n+1} = 0 \quad \text{per ogni } x \in \mathbb{K}$$

allora $t_1 = t_2 = \cdots = t_{n+1} = 0$ perché un polinomio non nullo ha un numero finito di radici (per il teorema di Ruffini). Quindi

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

è una base di $\mathbb{K}[x]_{\leq n}$.

Si noti che l'argomento che abbiamo usato non sarebbe più valido se invece di considerare come in questo libro $\mathbb{K} = \mathbb{R}$ o \mathbb{C} stessimo lavorando con un campo con un numero finito di elementi. In tal caso, occorrerebbe modificare la definizione di polinomio: un polinomio non è più pensato come una funzione di x , ma come una combinazione lineare dei simboli formali x^k , che sarebbero per definizione linearmente indipendenti.



Le tre matrici

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

formano una base dello spazio delle matrici simmetriche di ordine 2 dato che ogni matrice simmetrica si scrive in uno e un sol modo come combinazione lineare di queste tre matrici:

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Il vettore delle coordinate di $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$ rispetto alla base considerata è $[a, b, c]^T$. La mappa di parametrizzazione \mathcal{P} in questo caso è

$$\mathcal{P}([a, b, c]^T) = a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}.$$



Mostriamo che due vettori $\mathbf{v}_1 = \begin{bmatrix} a \\ b \end{bmatrix}$ e $\mathbf{v}_2 = \begin{bmatrix} c \\ d \end{bmatrix}$ formano una base di \mathbb{K}^2 se e solo se $ad - bc \neq 0$.

Per definizione, i due vettori formano una base se e solo se ogni vettore $\begin{bmatrix} x \\ y \end{bmatrix}$ si scrive in uno e un solo modo nella forma

$$(6.1) \quad \begin{bmatrix} x \\ y \end{bmatrix} = t_1 \begin{bmatrix} a \\ b \end{bmatrix} + t_2 \begin{bmatrix} c \\ d \end{bmatrix}$$

L'equazione (6.1) è equivalente al sistema lineare nelle incognite t_1 e t_2

$$(6.2) \quad \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

Per il teorema di Cramer questo sistema ha una e una sola soluzione, qualunque sia il vettore $\begin{bmatrix} x \\ y \end{bmatrix}^T$, se e solo se $ad - bc \neq 0$. Quindi $\begin{bmatrix} x \\ y \end{bmatrix}^T$ si scrive in uno e un solo modo nella forma (6.1) se e solo se $ad - bc \neq 0$.

Base di $\text{Ker}(\mathbf{A})$

Sia \mathbf{A} una matrice con n colonne di rango r . Sappiamo che l'insieme $\text{Ker}(\mathbf{A})$ delle soluzioni del sistema omogeneo $\mathbf{Ax} = \mathbf{0}$ è un sottospazio vettoriale di \mathbb{K}^n e quindi è uno spazio vettoriale, per cui ha senso chiedersi se ne possiamo trovare una base. In effetti risolvere il sistema significa proprio trovare una base di $\text{Ker}(\mathbf{A})$: per il teorema di Rouché-Capelli, detto r il rango di \mathbf{A} , esistono dei vettori $\mathbf{v}_1, \dots, \mathbf{v}_{n-r}$ in \mathbb{K}^n tali che un vettore $\mathbf{v} \in \text{Ker}(\mathbf{A})$ si scrive in uno e un sol modo come combinazione lineare

$$\mathbf{v} = t_1 \mathbf{v}_1 + \cdots + t_{n-r} \mathbf{v}_{n-r}$$

di $\mathbf{v}_1, \dots, \mathbf{v}_{n-r}$. Quindi $\mathbf{v}_1, \dots, \mathbf{v}_{n-r}$ formano una base di $\text{Ker}(\mathbf{A})$. Il MEG fornisce così un metodo per trovare una base di $\text{Ker}(\mathbf{A})$.

Per un esempio esplicito, sia \mathbf{H} il piano di \mathbb{R}^3 di equazione $x + y + z = 0$, di modo che $\mathbf{H} = \text{Ker}([1 \ 1 \ 1])$. Se risolviamo l'equazione usando y e z come variabili libere, troviamo la base di \mathbf{H} formata dai vettori $\mathbf{v}_1 = [-1, 1, 0]^T$ e $\mathbf{v}_2 = [-1, 0, 1]^T$. Il vettore delle coordinate del vettore $[-5, 2, 3]^T$ di \mathbf{H} rispetto a tale base è $[2, 3]^T$: fissare la base $\{\mathbf{v}_1, \mathbf{v}_2\}$ significa usare y e z come coordinate per \mathbf{H} .

Un'altra base di \mathbf{H} si ottiene usando x e y come variabili libere: $\mathbf{w}_1 = [1, 0, -1]^T$ e $\mathbf{w}_2 = [0, 1, -1]^T$. Il vettore delle coordinate di $[-5, 2, 3]^T$ rispetto a questa seconda base è $[-5, 2]^T$.

Il nostro prossimo obiettivo è mostrare che *ogni sottospazio vettoriale di \mathbb{K}^n ammette una base*. Lo faremo in due passi: prima osserviamo che un insieme *massimale*, nel senso specificato qui sotto di vettori linearmente indipendenti è una base, poi (lemma fondamentale) mostriamo che un insieme di vettori linearmente indipendenti di \mathbb{K}^n ha al massimo n elementi, per cui esiste un insieme massimale di vettori linearmente indipendenti.

DEFINIZIONE 6.3 Sia \mathbf{V} uno spazio vettoriale. Si dice che $\mathbf{v}_1, \dots, \mathbf{v}_d$ sono un insieme *massimale* di vettori linearmente indipendenti di \mathbf{V} se $\mathbf{v}_1, \dots, \mathbf{v}_d$ sono linearmente indipendenti e per ogni \mathbf{w} in \mathbf{V} , l'insieme

$$\mathbf{v}_1, \dots, \mathbf{v}_d, \mathbf{w}$$

è linearmente *dipendente*.

Mostriamo ora che un insieme massimale di vettori linearmente indipendenti è una base.

PROPOSIZIONE 6.4 Sia \mathbf{V} uno spazio vettoriale. Sia $\mathbf{v}_1, \dots, \mathbf{v}_d$ un insieme massimale di vettori linearmente indipendenti di \mathbf{V} . Allora $\mathbf{v}_1, \dots, \mathbf{v}_d$ è una base di \mathbf{V} .

DIMOSTRAZIONE. Per ipotesi i vettori $\mathbf{v}_1, \dots, \mathbf{v}_d$ sono linearmente indipendenti, quindi basta mostrare che sono un insieme di generatori di \mathbf{V} , cioè che ogni vettore di \mathbf{V} è combinazione lineare di $\mathbf{v}_1, \dots, \mathbf{v}_d$.

\mathbb{K}^n ha dimensione finita perché tutti i suoi elementi sono combinazione lineare dei vettori della base canonica. Più in generale, uno spazio vettoriale che ha una base è di dimensione finita. Daremo più avanti esempi di spazi che non hanno dimensione finita.

Vogliamo ora mostrare che ogni spazio vettoriale di dimensione finita ha una base. Se \mathbf{V} è ridotto al solo vettore nullo, per cui non contiene alcun insieme linearmente indipendente, si conviene che l'insieme vuoto sia una base, anzi l'unica base, di \mathbf{V} .

TEOREMA 6.7 (Esistenza di una base)

Sia \mathbf{V} uno spazio vettoriale di dimensione finita e sia $\mathbf{H} \subseteq \mathbf{V}$ un suo sottospazio.

Allora esiste una base di \mathbf{H} .

DIMOSTRAZIONE. Se $\mathbf{H} = \{\mathbf{0}\}$, per definizione l'insieme vuoto è una base di \mathbf{H} . Altrimenti \mathbf{H} contiene un vettore \mathbf{w}_1 non nullo. Se \mathbf{w}_1 è un insieme massimale di vettori linearmente indipendenti di \mathbf{H} , allora \mathbf{w}_1 è una base di \mathbf{H} e abbiamo finito. Altrimenti, esiste in \mathbf{H} un vettore \mathbf{w}_2 tale che \mathbf{w}_1 e \mathbf{w}_2 siano linearmente indipendenti.

Se $\mathbf{w}_1, \mathbf{w}_2$ è un insieme massimale di vettori linearmente indipendenti di \mathbf{H} , allora è una base di \mathbf{H} e abbiamo finito. Altrimenti possiamo ripetere il ragionamento: dopo d iterazioni, troviamo una base $\mathbf{w}_1, \dots, \mathbf{w}_d$ di \mathbf{H} oppure $d + 1$ vettori linearmente indipendenti. Supponiamo che \mathbf{V} sia generato da m vettori. Allora il procedimento descritto produce una base di \mathbf{H} dopo al più m passi, perché per il lemma fondamentale ogni insieme di $m + 1$ vettori in \mathbf{V} è linearmente dipendente.

COROLLARIO 6.8 Sia \mathbf{V} uno spazio vettoriale di dimensione finita non ridotto al solo vettore nullo. Allora:

- a) esiste una base di \mathbf{V} ;
- b) le basi di \mathbf{V} hanno tutte lo stesso numero di elementi.

DIMOSTRAZIONE. Per la proposizione precedente \mathbf{V} , che è un sottospazio di se stesso, ammette una base. Siano \mathcal{B}_1 e \mathcal{B}_2 due basi di \mathbf{V} e siano n_1 il numero di elementi di \mathcal{B}_1 e n_2 il numero di elementi di \mathcal{B}_2 .

Per le proprietà di una base, gli elementi di \mathcal{B}_1 sono linearmente indipendenti e gli elementi di \mathcal{B}_2 sono dei generatori di \mathbf{V} . Dal lemma fondamentale segue che $n_1 \leq n_2$.

Ma è anche vero che gli elementi di \mathcal{B}_2 sono linearmente indipendenti e gli elementi di \mathcal{B}_1 sono dei generatori di \mathbf{V} , per cui $n_2 \leq n_1$. Dalle due diseguaglianze segue $n_1 = n_2$ come volevasi dimostrare.

DEFINIZIONE 6.9 (Dimensione di uno spazio vettoriale)

Sia \mathbf{V} uno spazio vettoriale di dimensione finita. Il numero di elementi di una (e quindi di ogni) base di \mathbf{V} si dice *dimensione* di \mathbf{V} , e si denota col simbolo $\dim \mathbf{V}$.

La dimensione di uno spazio vettoriale è pertanto il minimo numero di variabili che sono necessarie per poter parametrizzare tutti i vettori dello spazio: un fisico direbbe che la dimensione è il numero dei gradi di libertà dei vettori di \mathbf{V} .

Possiamo ricavare la dimensione di tutti gli spazi di cui abbiamo scritto una base. Lo spazio \mathbb{K}^n ha dimensione n perché la base canonica consiste di n elementi. Analogamente $M_{\mathbb{K}}(m, n)$ ha dimensione $m \times n$. Lo spazio $\mathbb{K}[x]_{\leq n}$ dei polinomi di grado minore o uguale a n ha dimensione $n + 1$. Lo spazio delle matrici simmetriche di ordine 2 ha dimensione 3. Uno spazio vettoriale ha dimensione 0 se e solo se consiste del solo vettore nullo.

Non tutti gli spazi vettoriali hanno dimensione finita. Di norma, gli spazi funzionali non hanno dimensione finita. Per esempio lo spazio vettoriale $\mathbb{K}[x]$ dei polinomi a coefficienti in \mathbb{K} non ha dimensione finita. In effetti, se avesse dimensione finita d , per il lemma fondamentale ogni insieme costituito da $d + 1$ vettori di $\mathbb{K}[x]$ sarebbe linearmente dipendente. Ma questo è assurdo perché l'insieme $\{x^d, x^{d-1}, \dots, x, x^0 = 1\}$ dei monomi di grado $\leq d$ è linearmente indipendente e ha cardinalità $d + 1$. A fortiori, lo spazio vettoriale delle funzioni $f : \mathbb{K} \rightarrow \mathbb{K}$ non ha dimensione finita perché contiene lo spazio dei polinomi.

Mostriamo ora che, in uno spazio di dimensione finita, da ogni insieme di generatori si può estrarre una base, mentre ogni insieme di vettori linearmente indipendenti si può estendere a una base.

PROPOSIZIONE 6.10 Sia \mathbf{V} uno spazio vettoriale di dimensione finita. Allora:

- a) ogni insieme finito di generatori di \mathbf{V} contiene una base;
- b) ogni insieme di vettori linearmente indipendenti di \mathbf{V} è contenuto in una base.

DIMOSTRAZIONE. Supponiamo che $\mathbf{w}_1, \dots, \mathbf{w}_m$ generino \mathbf{V} . Se sono linearmente indipendenti, essi formano una base di \mathbf{V} . Altrimenti non sono linearmente indipendenti e allora per la proposizione 5.2 uno di essi è combinazione lineare degli altri. Riordinando eventualmente i vettori, possiamo supporre che \mathbf{w}_m sia combinazione lineare di $\mathbf{w}_1, \dots, \mathbf{w}_{m-1}$. Allora per il corollario 4.5

$$\mathbf{V} = \mathcal{L}(\mathbf{w}_1, \dots, \mathbf{w}_{m-1}, \mathbf{w}_m) = \mathcal{L}(\mathbf{w}_1, \dots, \mathbf{w}_{m-1})$$

cioè $\mathbf{w}_1, \dots, \mathbf{w}_{m-1}$ sono ancora un insieme di generatori di \mathbf{V} . Possiamo quindi iterare il procedimento: se $\mathbf{w}_1, \dots, \mathbf{w}_{m-1}$ sono linearmente indipendenti, essi formano una base; altrimenti possiamo eliminarne uno e ottenerne un insieme di generatori con $m - 2$ elementi. Proseguendo si arriva a costruire una base di \mathbf{V} contenuta in $\mathbf{w}_1, \dots, \mathbf{w}_m$ (nel caso limite in cui si possano eliminare tutti i generatori, si ottiene l'insieme vuoto ma anche $\mathbf{V} = \{\mathbf{0}\}$, e l'insieme vuoto è una base del sottospazio nullo).

La dimostrazione del secondo enunciato è del tutto analoga a quella della proposizione 6.7: se l'insieme di vettori linearmente indipendenti è massimale, allora è una base; se non è massimale, è contenuto in un insieme di vettori linearmente indipendenti con un vettore in

più. Si itera il procedimento finché non si raggiunge una base; il fatto che l'algoritmo termini è garantito dal fatto che ogni insieme di vettori linearmente indipendenti consiste al massimo di $\dim \mathbf{V}$ vettori.

OSSERVAZIONE Il punto b) della proposizione 6.9 è noto come teorema del completamento della base.

Il prossimo corollario mostra che, in uno spazio di dimensione n , un insieme di n vettori linearmente indipendenti (rispettivamente un insieme di generatori costituito da n vettori) è automaticamente una base. Per esempio, un insieme di 3 vettori linearmente indipendenti di \mathbb{R}^3 (rispettivamente, un insieme di generatori di \mathbb{R}^3 costituito da 3 vettori) è automaticamente una base di \mathbb{R}^3 .

COROLLARIO 6.11 Supponiamo $\dim \mathbf{V} = n$. Allora

- a) ogni insieme di generatori di \mathbf{V} costituito da n elementi è una base di \mathbf{V} ;
- b) ogni insieme di n vettori linearmente indipendenti è una base di \mathbf{V} .

DIMOSTRAZIONE. Tenuto conto che ogni base di \mathbf{V} è formata da n vettori, il primo enunciato segue dal fatto che ogni insieme finito di generatori contiene una base e il secondo enunciato dal fatto che ogni insieme di vettori linearmente indipendenti è contenuto in una base. \blacksquare

Esempio

Per il corollario, tre vettori di \mathbb{R}^3 formano una base di \mathbb{R}^3 se e solo se sono linearmente indipendenti. Sappiamo che tre vettori sono linearmente indipendenti se e solo se il loro prodotto misto è diverso da zero. Concludiamo che tre vettori di \mathbb{R}^3 formano una base di \mathbb{R}^3 se e solo se il loro prodotto misto è diverso da zero. Nel capitolo sui determinanti generalizzeremo questa affermazione al caso di \mathbb{K}^n : un insieme di n vettori di \mathbb{K}^n è una base di \mathbb{K}^n se e solo se è diverso da zero il determinante della matrice che ha per colonne gli n vettori.

COROLLARIO 6.12 Siano $\mathbf{H} \subseteq \mathbf{V}$ spazi vettoriali di dimensione finita. Allora $\dim \mathbf{H} \leq \dim \mathbf{V}$. Inoltre, se $\dim \mathbf{H} = \dim \mathbf{V}$, allora $\mathbf{H} = \mathbf{V}$.

DIMOSTRAZIONE. Se $\mathcal{C} = \{\mathbf{w}_1, \dots, \mathbf{w}_d\}$ è una base di \mathbf{H} , allora $\mathbf{w}_1, \dots, \mathbf{w}_d$ sono vettori linearmente indipendenti di \mathbf{V} e per il lemma fondamentale

$$\dim \mathbf{H} = d \leq n = \dim \mathbf{V}$$

Se $d = n$, dal corollario 6.11 segue che \mathcal{C} è una base anche di \mathbf{V} , per cui $\mathbf{w}_1, \dots, \mathbf{w}_d$ sono un insieme di generatori di \mathbf{V} . Ma allora \mathbf{V} è il sottospazio generato da $\mathbf{w}_1, \dots, \mathbf{w}_d$ e quindi coincide con \mathbf{H} . \blacksquare

Per mostrare che due spazi vettoriali \mathbf{H} e \mathbf{V} sono uguali, è sufficiente far vedere $\mathbf{H} \subseteq \mathbf{V}$ e $\dim \mathbf{H} = \dim \mathbf{V}$.

Possiamo ora mostrare che gli unici sottospazi vettoriali di \mathbb{R}^3 sono l'origine, le rette per l'origine, i piani per l'origine e tutto \mathbb{R}^3 .

Sia \mathbf{H} un sottospazio di \mathbb{R}^3 . Allora $0 \leq \dim \mathbf{H} \leq 3$ perché in \mathbb{R}^3 ogni insieme di più di tre vettori è linearmente dipendente. Se $\dim \mathbf{H} = 0$, per definizione $\mathbf{H} = \{0\}$ consiste della sola origine. Se $\dim \mathbf{H} = 1$ e \mathbf{v} è una base di \mathbf{H} , allora \mathbf{H} è l'insieme dei multipli scalari di \mathbf{v} , cioè la retta per l'origine diretta come \mathbf{v} . Se $\dim \mathbf{H} = 2$ e \mathbf{v}, \mathbf{w} formano una base di \mathbf{H} , allora \mathbf{H} è il piano per l'origine contenente \mathbf{v} e \mathbf{w} . Infine, se $\dim \mathbf{H} = 3$, per il corollario precedente $\mathbf{H} = \mathbb{R}^3$.

Sia \mathbf{H} il piano di \mathbb{R}^3 di equazione $2x - y + z = 0$. Trovare quattro basi distinte di \mathbf{H} . Scrivere le componenti del vettore $[0, 1, 1]^T$ rispetto a ciascuna di tali basi (si osservi che le componenti sono due perché \mathbf{H} ha dimensione 2, e che, se $\{\mathbf{v}_1, \mathbf{v}_2\}$ è una base di \mathbf{H} , allora $\{\mathbf{v}_2, \mathbf{v}_1\}$ è un'altra base di \mathbf{H}).

Sia \mathbf{V} lo spazio vettoriale dei polinomi $P(x)$ di grado minore o uguale a 2 e sia \mathbf{H} il sottospazio dei polinomi tali che $P(1) = 0$. Trovare una base di \mathbf{H} e concludere che \mathbf{H} ha dimensione 2. Scrivere le componenti di $P(x) = x^2 - 1$ rispetto alla base trovata.

Suggerimento: un polinomio $P(x) = ax^2 + bx + c$ è determinato dalla sua terna di coordinate $[a, b, c]^T$ rispetto alla base canonica $\{x^2, x, 1\}$; è quindi possibile tradurre l'esercizio in un esercizio sui vettori $[a, b, c]^T$.

Mostrare che l'insieme \mathbb{C} dei numeri complessi è uno spazio vettoriale su \mathbb{R} di dimensione 2, e uno spazio vettoriale su \mathbb{C} di dimensione 1.

Mostrare che $\mathbf{v}_1, \dots, \mathbf{v}_d$ sono linearmente indipendenti se e solo se $\dim \mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_d) = d$.

Suggerimento: una direzione è semplice perché se i vettori sono linearmente indipendenti, allora formano una base dello spazio da essi generato; per l'altra direzione usare il corollario 6.11.

Sia $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ una base dello spazio vettoriale \mathbf{V} e siano $\mathbf{v}_1, \dots, \mathbf{v}_d$ vettori linearmente indipendenti di \mathbf{V} . Mostrare che esiste una base di \mathbf{V} formata da $\mathbf{v}_1, \dots, \mathbf{v}_d$ e da $n-d$ vettori della base \mathcal{B} .

Trovare una base di \mathbb{R}^3 che contenga i vettori $[1, 1, 1]^T$ e $[1, 2, 5]^T$. Scrivere le componenti di $[1, 0, 0]^T$ rispetto a tale base.

Mostrare che le quattro matrici

$$\mathbf{B}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{B}_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \mathbf{B}_3 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{B}_4 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

formano una base dello spazio vettoriale delle matrici quadrate di ordine 2. Scrivere le coordinate della matrice identità rispetto a tale base.

Suggerimento: basta verificare che le matrici sono linearmente indipendenti.

Sia $P(x) = x^2$. Mostrare che $P(x)$, $P'(x)$ e $P''(x)$ formano una base dello spazio dei polinomi di grado ≤ 2 a coefficienti reali.

Sia \mathbb{M} lo spazio vettoriale delle matrici quadrate di ordine n e sia $\mathbb{D} \subset \mathbb{M}$ il sottospazio delle matrici diagonali. Trovare una base di \mathbb{D} e dedurre che $\dim(\mathbb{D}) = n$. Scrivere le coordinate della matrice identità rispetto alla base trovata.

Sia \mathbb{S} lo spazio vettoriale delle matrici quadrate *simmetriche* di ordine n . Trovare una base di \mathbb{S} e mostrare che $\dim(\mathbb{S}) = \frac{1}{2}n(n+1)$

Suggerimento: trattare prima i casi $n = 2$ e $n = 3$.

Sia \mathbb{E} lo spazio vettoriale delle matrici quadrate *antisimmetriche* di ordine n . Trovare una base di \mathbb{E} e mostrare che $\dim(\mathbb{E}) = \frac{1}{2}n(n-1)$.

Sia \mathbb{T} lo spazio vettoriale delle matrici triangolari alte di ordine n . Trovare una base di \mathbb{T} e mostrare che $\dim(\mathbb{T}) = \frac{1}{2}n(n+1)$.

Sia ℓ_1 lo spazio vettoriale delle successioni a_k tali che

$$\sum_{k=0}^{+\infty} |a_k| < +\infty$$

Mostrare che ℓ_1 non ha dimensione finita.

■ 7 COORDINATE

Fissata una base di uno spazio vettoriale \mathbf{V} , possiamo identificare un vettore di \mathbf{V} con la n -upla delle sue *coordinate*, analogamente a come l'introduzione di un sistema di riferimento consente di identificare un punto dello spazio cartesiano con la terna delle sue coordinate. Questo permette di fare i conti esplicitamente. In questo paragrafo spieghiamo come ricondurre problemi astratti a problemi sui vettori delle coordinate. Nel prossimo paragrafo ci occuperemo invece di come fare i conti in coordinate. Chiudiamo il paragrafo mostrando come cambiano le coordinate al variare della base. Questo è estremamente utile perché una scelta oculata di una base (come per il sistema di riferimento) consente spesso una drastica semplificazione dei conti; il problema di come scegliere una base che semplifichi i conti verrà affrontato nel capitolo su autovettori e autovalori.

Fissiamo quindi uno spazio vettoriale \mathbf{V} e una sua base $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$. Dato $\mathbf{v} \in \mathbf{V}$, denotiamo il vettore delle coordinate di \mathbf{v} rispetto alla base \mathcal{B} col simbolo $\mathbf{x}(\mathbf{v})$: quindi $\mathbf{x}(\mathbf{v}) = [x_1, \dots, x_n]^T$ significa

$$\mathbf{v} = x_1 \mathbf{b}_1 + \cdots + x_n \mathbf{b}_n$$

Una notazione più precisa, ma più pesante, è $\mathbf{x}_{\mathcal{B}}(\mathbf{v})$: ha l'unico vantaggio di ricordare che si tratta delle coordinate rispetto alla base \mathcal{B} . Altri autori scrivono $[\mathbf{v}]_{\mathcal{B}}$.

In termini più formali, la funzione $\mathbf{x}(\mathbf{v}) : \mathbf{V} \rightarrow \mathbb{K}^n$ è la funzione inversa della mappa di parametrizzazione $\mathcal{P} : \mathbb{K}^n \rightarrow \mathbf{V}$:

$$\mathcal{P}([x_1, \dots, x_n]^T) = x_1 \mathbf{b}_1 + \dots + x_n \mathbf{b}_n$$

Esempio

Il vettore $\mathbf{x}(\mathbf{b}_k)$ delle coordinate del k -esimo vettore della base \mathcal{B} rispetto alla base \mathcal{B} è il k -esimo vettore \mathbf{e}_k della base canonica di \mathbb{K}^n poiché

$$\mathbf{b}_k = 0 \mathbf{b}_1 + \dots + 1 \mathbf{b}_k + \dots + 0 \mathbf{b}_n$$

Esempio

Consideriamo $\mathbf{V} = \mathbb{M}_{\mathbb{K}}(2, 2)$ con la base canonica:

$$\mathbf{E}_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{E}_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \mathbf{E}_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \mathbf{E}_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Allora

$$\mathbf{x}\left(\begin{bmatrix} a & b \\ c & c \end{bmatrix}\right) = [a, b, c, d]^T$$

perché

$$\begin{bmatrix} a & b \\ c & c \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Esempio

Sia \mathbf{V} lo spazio vettoriale delle matrici simmetriche di ordine 2, con la base

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

Allora

$$\mathbf{x}\left(\begin{bmatrix} a & b \\ b & c \end{bmatrix}\right) = [a, b, c]^T$$

Esempio

Sia \mathbf{V} lo spazio vettoriale $\mathbb{K}[x]_{\leq 2}$ dei polinomi di grado ≤ 3 con la base $\{x^3, x^2, x, 1\}$. Allora

$$\mathbf{x}(ax^3 + bx^2 + cx + d) = [a, b, c, d]^T.$$

Per esempio le coordinate di $P(x) = (1+x)^2$ sono

$$\mathbf{x}((1+x)^2) = [0, 1, 2, 1]^T$$

Esempio

Sia \mathbf{V} il piano di \mathbb{R}^3 di equazione $x + y + z = 0$ con la base $\{[-1, 1, 0]^T, [-1, 0, 1]^T\}$. Allora

$$\mathbf{x}((-y - z, y, z)) = [y, z]^T$$

La funzione che associa a un vettore le sue coordinate è lineare:

PROPOSIZIONE 7.1 La funzione $\mathbf{x}(\mathbf{v})$ è lineare:

a) rispetta la somma:

$$\mathbf{x}(\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{x}(\mathbf{v}_1) + \mathbf{x}(\mathbf{v}_2) \quad \text{per ogni } \mathbf{v}_1, \mathbf{v}_2 \in \mathbf{V}$$

b) rispetta il prodotto scalare:

$$\mathbf{x}(t\mathbf{v}) = t\mathbf{x}(\mathbf{v}) \quad \text{per ogni } t \in \mathbb{K} \text{ e ogni } \mathbf{v} \in \mathbf{V}$$

c) e quindi rispetta le combinazioni lineari:

$$\mathbf{x}(t_1\mathbf{v}_1 + \cdots + t_d\mathbf{v}_d) = t_1\mathbf{x}(\mathbf{v}_1) + \cdots + t_d\mathbf{x}(\mathbf{v}_d)$$

per ogni $t_1, \dots, t_d \in \mathbb{K}$ e ogni $\mathbf{v}_1, \dots, \mathbf{v}_d \in \mathbf{V}$.

DIMOSTRAZIONE. Lasciamo la semplice dimostrazione come esercizio al lettore (che la può sostanzialmente trovare all'interno della dimostrazione del lemma fondamentale e nel primo capitolo nel caso \mathbf{V} sia lo spazio dei vettori liberi).

La proposizione seguente riconduce il problema di determinare se un insieme di vettori di \mathbf{V} è linearmente indipendente, o una base di un sottospazio, a un problema in \mathbb{K}^n . Nel prossimo paragrafo vedremo come risolvere questi problemi in \mathbb{K}^n .

PROPOSIZIONE 7.2 Sia \mathbf{V} uno spazio vettoriale di dimensione n con base \mathcal{B} e sia $\mathbf{x} : \mathbf{V} \rightarrow \mathbb{K}^n$ la funzione che a un vettore \mathbf{v} di \mathbf{V} associa la n -pla delle sue coordinate rispetto alla base \mathcal{B} . Allora:

a) se \mathbf{H} è un sottospazio di \mathbf{V} , la sua immagine in \mathbb{K}^n

$$\mathbf{x}(\mathbf{H}) = \{\mathbf{y} \in \mathbb{K}^n : \text{esiste } \mathbf{v} \in \mathbf{H} \text{ tale che } \mathbf{y} = \mathbf{x}(\mathbf{v})\}$$

è un sottospazio di \mathbb{K}^n ;

b) i vettori $\mathbf{v}_1, \dots, \mathbf{v}_d$ di \mathbf{V} sono linearmente indipendenti se e solo se i vettori $\mathbf{x}(\mathbf{v}_1), \dots, \mathbf{x}(\mathbf{v}_d)$ di \mathbb{K}^n sono linearmente indipendenti;

c) i vettori $\mathbf{v}_1, \dots, \mathbf{v}_d$ generano il sottospazio \mathbf{H} di \mathbf{V} se e solo se $\mathbf{x}(\mathbf{v}_1), \dots, \mathbf{x}(\mathbf{v}_d)$ generano $\mathbf{x}(\mathbf{H})$;

d) un insieme $\{\mathbf{v}_1, \dots, \mathbf{v}_d\}$ è una base di un sottospazio \mathbf{H} di \mathbf{V} se e solo se $\{\mathbf{x}(\mathbf{v}_1), \dots, \mathbf{x}(\mathbf{v}_d)\}$ è una base di $\mathbf{x}(\mathbf{H})$. In particolare, \mathbf{H} e $\mathbf{x}(\mathbf{H})$ hanno la stessa dimensione.

DIMOSTRAZIONE. L'enunciato è vero per ogni funzione che, come \mathbf{x} , sia lineare e iniettiva. La dimostrazione verrà data nel capitolo sulle applicazioni lineari (il lettore dovrebbe farla in questo caso come esercizio). ■

ESERCIZIO 1

Siano $\mathbf{u}, \mathbf{v}, \mathbf{w}$ tre vettori linearmente indipendenti di uno spazio vettoriale \mathbf{V} . Mostriamo che i vettori $\mathbf{p} = \mathbf{u} + \mathbf{v}$, $\mathbf{r} = \mathbf{u} - \mathbf{v}$ e $\mathbf{s} = \mathbf{u} - 2\mathbf{v} + \mathbf{w}$ sono linearmente indipendenti. Per questo consideriamo il sottospazio \mathbf{H} generato da $\mathbf{u}, \mathbf{v}, \mathbf{w}$ con la base formata da questi tre vettori. Le coordinate dei vettori in questione sono

$$\mathbf{x}(\mathbf{p}) = [1, 1, 0]^T, \quad \mathbf{x}(\mathbf{r}) = [1, -1, 0]^T, \quad \mathbf{x}(\mathbf{s}) = [1, -2, 1]^T$$

Per la proposizione precedente basta mostrare che i vettori delle coordinate sono indipendenti. Per questo osserviamo che

$$t_1 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + t_3 \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}$$

è il vettore nullo solo se $t_1 = t_2 = t_3 = 0$. Quindi i tre vettori $\mathbf{p}, \mathbf{r}, \mathbf{s}$ sono linearmente indipendenti.

Cambiamento di base

Sia \mathbf{V} uno spazio vettoriale di dimensione n e siano

$$\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\} \quad \text{e} \quad \mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_n\}$$

due basi di \mathbf{V} . Dato un vettore $\mathbf{v} \in \mathbf{V}$, siano

$$\mathbf{x}(\mathbf{v}) = [x_1, \dots, x_n]^T \quad \text{e} \quad \mathbf{X}(\mathbf{v}) = [X_1, \dots, X_n]^T$$

i vettori delle coordinate di \mathbf{v} rispetto alle basi \mathcal{B} e \mathcal{C} , rispettivamente. Questo significa che:

$$\mathbf{v} = x_1 \mathbf{b}_1 + \cdots + x_n \mathbf{b}_n = X_1 \mathbf{c}_1 + \cdots + X_n \mathbf{c}_n$$

È possibile e comodo estendere il prodotto righe per colonne al caso in cui il vettore riga non sia un vettore vero e proprio, ma un vettore le cui componenti siano a loro volta dei vettori; in questo modo possiamo riscrivere l'uguaglianza $\mathbf{v} = x_1 \mathbf{b}_1 + \cdots + x_n \mathbf{b}_n$ nella forma compatta:

$$\mathbf{v} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_n] \mathbf{x}(\mathbf{v})$$

Per i vettori \mathbf{c}_k della base \mathcal{C} otteniamo

$$\mathbf{c}_k = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_n] \mathbf{x}(\mathbf{c}_k) \quad \text{per ogni } k = 1, 2, \dots, n$$

La relazione che lega le due basi si può quindi scrivere nella forma matriciale

$$(7.1) \quad [\mathbf{c}_1 \ \mathbf{c}_2 \ \cdots \ \mathbf{c}_n] = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_n] [\mathbf{x}(\mathbf{c}_1) \ \mathbf{x}(\mathbf{c}_2) \ \cdots \ \mathbf{x}(\mathbf{c}_n)]$$

Introduciamo, quindi, la matrice

$$\mathbf{S} = [\mathbf{x}(\mathbf{c}_1) \ \mathbf{x}(\mathbf{c}_2) \ \cdots \ \mathbf{x}(\mathbf{c}_n)]$$

la cui k -esima colonna è il vettore $\mathbf{x}(\mathbf{c}_k)$ delle coordinate del k -esimo vettore della base \mathcal{C} rispetto alla \mathcal{B} . Si dice che \mathbf{S} è la *matrice di passaggio* dalla base \mathcal{B} alla base \mathcal{C} . L'equazione (7.1) diventa così maneggevole:

$$(7.2) \quad [\mathbf{c}_1 \ \mathbf{c}_2 \ \cdots \ \mathbf{c}_n] = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_n] \mathbf{S}$$



Sia \mathbf{V} il sottospazio dello spazio vettoriale delle funzioni $f(x) : \mathbb{R} \rightarrow \mathbb{R}$ generato dalle funzioni e^x e e^{-x} . Queste due funzioni sono linearmente indipendenti perché e^{-x} non è un multiplo scalare di e^x . Quindi $\mathcal{B} = \{e^x, e^{-x}\}$ è una base di \mathbf{V} . Le funzioni iperboliche

$$\text{Ch}(x) = \frac{e^x + e^{-x}}{2} \quad \text{e} \quad \text{Sh}(x) = \frac{e^x - e^{-x}}{2}$$

appartengono a \mathbf{V} perché sono combinazioni lineari di e^x ed e^{-x} e sono linearmente indipendenti. Quindi formano una seconda base $\mathcal{C} = \{\text{Ch}(x), \text{Sh}(x)\}$ di \mathbf{V} . Rispetto alla base $\{e^x, e^{-x}\}$, il vettore delle coordinate di $\text{Ch}(x)$ è $[1/2, 1/2]^T$ e quello di $\text{Sh}(x)$ è $[1/2, -1/2]^T$, per cui la matrice di passaggio dalla base \mathcal{B} alla base \mathcal{C} è

$$\mathbf{S} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}$$

e l'uguaglianza (7.1) è

$$[\text{Ch}(x) \ \text{Sh}(x)] = [e^x \ e^{-x}] \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}$$



Supponiamo $\mathbf{V} = \mathbb{K}^n$. Allora i vettori \mathbf{b}_i sono vettori colonna e $\mathbf{B} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_n]$ è la matrice che ha per colonne i vettori \mathbf{b}_i . Analogamente $\mathbf{C} = [\mathbf{c}_1 \ \mathbf{c}_2 \ \cdots \ \mathbf{c}_n]$ è la matrice che per colonne i vettori \mathbf{c}_k . L'equazione (7.2) diviene

$$\mathbf{C} = \mathbf{BS}$$

Il caso di uso più frequente è quello in cui la base \mathcal{B} sia la base canonica; allora la matrice $\mathbf{B} = \mathbf{I}$ è la matrice identità e $\mathbf{C} = \mathbf{S}$. Vediamo così che

Quando $\mathbf{V} = \mathbb{K}^n$ e \mathcal{B} è la base canonica, la matrice di passaggio \mathbf{S} ha per colonne i vettori della base \mathcal{C} :

$$\mathbf{S} = [\mathbf{c}_1 \ \mathbf{c}_2 \ \cdots \ \mathbf{c}_n]$$

Per un esempio esplicito, si consideri $\mathbf{V} = \mathbb{R}^2$, $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2\}$ la base canonica e

$$\mathcal{C} = \left\{ \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix}, \begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \right\}$$

la base \mathcal{C} si ottiene ruotando la base canonica di 45° in senso antiorario. La matrice di passaggio ha come colonne i vettori della base \mathcal{C} :

$$\mathbf{S} = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}$$

Diamo ora la formula per il cambiamento di coordinate in termini della matrice di passaggio \mathbf{S} :

PROPOSIZIONE 7.3 Sia \mathbf{V} uno spazio vettoriale e siano

$$\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\} \quad \text{e} \quad \mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_n\}$$

basi di \mathbf{V} . Sia \mathbf{S} la matrice di passaggio la cui k -esima colonna è il vettore delle coordinate di \mathbf{c}_k rispetto alla base \mathcal{B} . Allora \mathbf{S} è invertibile e le coordinate \mathbf{x} e \mathbf{X} di un vettore \mathbf{v} rispetto alle due basi sono legate dalle formule:

$$\mathbf{x} = \mathbf{S}\mathbf{X} \quad \text{e} \quad \mathbf{X} = \mathbf{S}^{-1}\mathbf{x}$$

DIMOSTRAZIONE. Il legame tra le due basi è fornito dall'equazione

$$[\mathbf{c}_1 \ \mathbf{c}_2 \ \cdots \ \mathbf{c}_n] = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_n] \mathbf{S}$$

Da quest'equazione ricaviamo il legame tra le coordinate \mathbf{x} e \mathbf{X} di un vettore \mathbf{v} . In termini della base \mathcal{C} il vettore si scrive:

$$\mathbf{v} = X_1\mathbf{c}_1 + \cdots + X_n\mathbf{c}_n = [\mathbf{c}_1 \ \mathbf{c}_2 \ \cdots \ \mathbf{c}_n] \mathbf{X}$$

Sostituendo l'espressione della base \mathcal{C} rispetto alla base \mathcal{B} troviamo

$$\mathbf{v} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_n] \mathbf{S}\mathbf{X}$$

D'altra parte, per definizione di coordinate

$$\mathbf{v} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_n] \mathbf{x}$$

e dall'unicità delle coordinate rispetto alla base \mathcal{B} segue

$$\mathbf{x} = \mathbf{S}\mathbf{X}$$

Ora osserviamo che la matrice \mathbf{S} è invertibile: questo segue dal teorema 4.3 del capitolo sull'algebra delle matrici, se facciamo vedere che $\text{Ker}(\mathbf{S}) = \{\mathbf{0}\}$. Supponiamo che $\mathbf{X} \in \text{Ker}(\mathbf{S})$: allora $\mathbf{x} = \mathbf{S}\mathbf{X} = \mathbf{0}$. Quindi il vettore \mathbf{v} , che ha coordinate $\mathbf{x} = \mathbf{0}$ e \mathbf{X} rispetto alle basi \mathcal{B} e \mathcal{C} , è il vettore nullo di \mathbf{V} ; ma allora \mathbf{X} dev'essere il vettore nullo di \mathbb{K}^n per l'unicità delle coordinate rispetto alla base \mathcal{C} . Perciò $\text{Ker}(\mathbf{S}) = \{\mathbf{0}\}$ e \mathbf{S} è invertibile. Moltiplicando l'uguaglianza $\mathbf{x} = \mathbf{S}\mathbf{X}$ a sinistra per \mathbf{S}^{-1} otteniamo:

$$\mathbf{X} = \mathbf{S}^{-1}\mathbf{x}$$



OSSERVAZIONE Vi è un'importante difficoltà nell'utilizzo della formula della proposizione precedente: la matrice di passaggio \mathbf{S} esprime la base \mathcal{C} in termini della base \mathcal{B} :

$$[\mathbf{c}_1 \ \mathbf{c}_2 \ \cdots \ \mathbf{c}_n] = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_n] \mathbf{S}$$

mentre è la matrice \mathbf{S}^{-1} che trasforma le coordinate \mathbf{x} rispetto a \mathcal{B} nelle coordinate \mathbf{X} rispetto a \mathcal{C} :

$$\mathbf{X} = \mathbf{S}^{-1} \mathbf{x}$$

La legge di trasformazione delle coordinate è l'inversa della legge di trasformazione delle basi. Purtroppo il calcolo dell'inversa è computazionalmente molto costoso.

L'esempio più comune è quello in cui $\mathbf{V} = \mathbb{K}^n$, la base \mathcal{B} è la base canonica e il problema che vogliamo risolvere suggerisce di considerare una base \mathcal{C} diversa da quella canonica. Possiamo immediatamente scrivere la matrice \mathbf{S} , che ha per colonne i vettori della base \mathcal{C} , ma per trovare le nuove coordinate in funzione di quelle canoniche dobbiamo invertire \mathbf{S} .

Esempio

Torniamo all'esempio dello spazio vettoriale generato da e^x ed e^{-x} . Abbiamo visto che la matrice di passaggio dalla base $\mathcal{B} = \{e^x, e^{-x}\}$ alla base $\mathcal{C} = \{\text{Ch}(x), \text{Sh}(x)\}$ è

$$\mathbf{S} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}$$

Usando la formula per l'inversa di una matrice 2×2 troviamo

$$\mathbf{S}^{-1} = -2 \begin{bmatrix} -1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

La funzione $f(x) = ae^x + be^{-x}$ ha coordinare $\mathbf{x}(f) = [a, b]^T$ rispetto alla base \mathcal{B} . Quindi le sue coordinate $\mathbf{X} = [A, B]^T$ rispetto alla base \mathcal{C} sono

$$\begin{bmatrix} A \\ B \end{bmatrix} = \mathbf{S}^{-1} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a+b \\ a-b \end{bmatrix}$$

Per esempio, la funzione $\text{Ch}(x)$ ha coordinare $a = b = \frac{1}{2}$ rispetto a \mathcal{B} e coordinate $[A, B]^T = [1, 0]$ rispetto a \mathcal{C} ; la funzione e^x ha coordinare $[a, b]^T = [1, 0]^T$ rispetto a \mathcal{B} e coordinate $[A, B]^T = [1, 1]$ rispetto alla base \mathcal{C} . Effettivamente

$$e^x = \text{Ch}(x) + \text{Sh}(x)$$

Esempio

Un'importante classe di matrici quadrate a coefficienti reali è quella delle matrici ortogonali: una matrice quadrata reale \mathbf{Q} si dice ortogonale se

$$\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$$

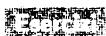
(vedremo che questa condizione significa che le colonne di \mathbf{Q} formano una base ortonormale di \mathbb{R}^n : sono a due a due perpendicolari e hanno tutte lunghezza 1). La matrice trasposta

\mathbf{Q}^T è allora l'inversa di \mathbf{Q} : in questo caso il calcolo dell'inversa è immediato! Per esempio, il lettore può verificare che

$$\mathbf{Q} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

è ortogonale perché la sua inversa è

$$\mathbf{Q}^T = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}.$$



Per quali valori del parametro reale k il polinomio $3x^2 + kx + 2$ è combinazione lineare dei polinomi $x^2 + 2x + 1$ e $x - 1$? Per tali valori trovare i coefficienti della combinazione lineare.

Suggerimento: risolvere il problema sui vettori delle coordinate rispetto alla base $\{x^2, x, 1\}$.

Trovare le coordinate di $[x, y]^T$ rispetto alla base $\{[1, 2]^T, [2, 1]^T\}$ di \mathbb{R}^2 .

Supponiamo di effettuare nel piano \mathbb{R}^2 il cambiamento di coordinate lineare

$$\begin{cases} X = 3x - y \\ Y = x + 2y \end{cases}$$

Trovare la base \mathcal{C} di \mathbb{R}^2 tale che $[X, Y]^T$ sia il vettore delle coordinate di $[x, y]^T$ rispetto a \mathcal{C} .
Suggerimento: riscrivere il cambiamento di coordinate nella forma matriciale $\mathbf{X} = \mathbf{S}^{-1}\mathbf{x}$; la matrice \mathbf{S} ha per colonne i vettori della base cercata.

Si verifichi che i vettori $\{[1, 0, 1]^T, [2, 0, 0]^T, [3, 1, 1]^T\}$ sono linearmente indipendenti e quindi formano una base di \mathbb{R}^3 . Determinare le componenti del vettore $[0, 2, 1]^T$ rispetto a tale base.

Suggerimento: risolvere il sistema lineare $[0, 2, 1]^T = X_1\mathbf{c}_1 + X_2\mathbf{c}_2 + X_3\mathbf{c}_3$.

Verificare che $\mathcal{B} = \{[1, 1, -2]^T, [2, -1, -1]^T\}$ è una base del piano di \mathbb{R}^3 di equazione $x + y + z = 0$. Trovare le coordinate di $[-y - z, y, z]^T$ rispetto alla base \mathcal{B} .

Verificare che $\mathcal{B} = \{(x+1)^2, x+1, 1\}$ è una base dello spazio dei polinomi di grado ≤ 2 . Trovare le coordinate di

$$P(x) = ax^2 + bx + c$$

rispetto alla base \mathcal{B} .

8 RANGO DI UNA MATRICE

In questo paragrafo mostriamo il primo risultato importante dell'algebra lineare: il rango di una matrice è la dimensione dello spazio generato dalle righe della matrice e anche la dimensione dello spazio generato dalle colonne. Quindi il rango è il massimo numero di righe linearmente indipendenti e anche il massimo numero di colonne linearmente indipendenti. Inoltre reinterpretiamo il teorema di Rouché-Capelli come

teorema di nullità più rango: la somma della dimensione del nucleo e del rango è il numero di colonne della matrice. In termini di sistemi lineari, questo significa che le soluzioni di un sistema lineare si esprimono in funzione di $n-r$ parametri, dove n è il numero di incognite e r è il numero di equazioni linearmente *indipendenti* del sistema.

Sia \mathbf{A} una matrice di tipo (m, n) a coefficienti in \mathbb{K} . Abbiamo già associato ad \mathbf{A} un sottospazio di \mathbb{K}^n : il nucleo di \mathbf{A} . In questo paragrafo consideriamo altri due spazi naturalmente associati ad \mathbf{A} :

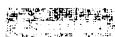
DEFINIZIONE 8.1 (Spazio riga e spazio colonna)

Sia \mathbf{A} una matrice di tipo (m, n) .

- a) Lo spazio riga $\text{Row}(\mathbf{A})$ di \mathbf{A} è il sottospazio di \mathbb{K}^n generato dalle righe di \mathbf{A} ;
- b) lo spazio colonna $\text{Col}(\mathbf{A})$ di \mathbf{A} è il sottospazio di \mathbb{K}^m generato dalle colonne di \mathbf{A} .

Nota tecnica: siccome in questo libro \mathbb{K}^n denota l'insieme dei vettori *colonna* con n componenti, per pensare le righe della matrice come elementi di \mathbb{K}^n dobbiamo trasporle in modo da ottenere dei vettori colonna.

Lo spazio riga consiste per definizione dei vettori che sono combinazione lineare delle righe della matrice; non contiene solo le righe di A , ma anche tutte le loro infinite combinazioni lineari. Siccome le righe di \mathbf{A} sono vettori con n componenti, lo spazio riga è un sottospazio di \mathbb{K}^n . Per capire l'importanza del concetto, è utile pensare a una riga della matrice come a un'equazione del sistema $\mathbf{Ax} = \mathbf{0}$. Una combinazione lineare delle righe corrisponde così a una combinazione lineare delle equazioni del sistema. Lo spazio riga consente quindi di considerare insieme alle equazioni originali del sistema, anche tutte quelle che da queste si ottengono come combinazioni lineari. Si noti che se \mathbf{v} è una soluzione di $\mathbf{Ax} = \mathbf{0}$, cioè un vettore del nucleo di \mathbf{A} , allora \mathbf{v} soddisfa ogni equazione del sistema $\mathbf{Ax} = \mathbf{0}$ e quindi ogni equazione che sia combinazione lineare delle equazioni del sistema, cioè ogni equazione il cui vettore dei coefficienti appartenga allo spazio riga. Si può mostrare che qualcosa di più è vero: un'equazione è soddisfatta da tutti i vettori del nucleo se e solo se il vettore dei suoi coefficienti appartiene allo spazio riga; lo spazio riga di \mathbf{A} consiste quindi (dei vettori dei coefficienti) di tutte le equazioni che sono soddisfatte dalle soluzioni di $\mathbf{Ax} = \mathbf{0}$.



Consideriamo la matrice

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

Lo spazio riga di \mathbf{A} consiste dei vettori che sono combinazione lineare delle due righe di \mathbf{A} , quindi dei vettori della forma

$$t_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + t_2 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} t_1 + t_2 \\ t_1 + 2t_2 \\ t_1 + 3t_2 \end{bmatrix}$$

Il nucleo di \mathbf{A} è la retta generata da $[1, -2, 1]^T$. Si può mostrare (esercizio!) che

$$a_1x_1 + a_2x_2 + a_3x_3 = 0$$

è soddisfatta da $[1, -2, 1]^T$ se e solo se $[a_1, a_2, a_3]^T$ appartiene allo spazio riga di \mathbf{A} .

Si noti che il sistema $\mathbf{Ax} = \mathbf{0}$ si risolve col MEG sottraendo alla seconda equazione la prima; il risultato è ancora nello spazio riga:

$$-\text{Riga}_1 + \text{Riga}_2 = [0, 1, 2]^T \in \text{Row}(\mathbf{A})$$

Geometricamente, il piano di equazione $a_1x_1 + a_2x_2 + a_3x_3 = 0$ appartiene al fascio che ha per sostegno la retta generata da $[1, -2, 1]^T$ se e solo se la sua equazione è combinazione lineare delle equazioni cartesiane della retta.

Lo spazio colonna consiste dei vettori di \mathbb{K}^m che sono combinazione lineare delle colonne di \mathbf{A} . Il suo significato in termini di sistemi lineari è:

PROPOSIZIONE 8.2 Sia \mathbf{A} una matrice di tipo (m, n) . Lo spazio colonna di \mathbf{A} coincide con l'insieme dei vettori \mathbf{b} di \mathbb{K}^m per i quali il sistema lineare $\mathbf{Ax} = \mathbf{b}$ ammette soluzione.

DIMOSTRAZIONE. Se $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ sono le colonne di \mathbf{A} , il prodotto

$$\mathbf{Ax} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$$

è la combinazione lineare delle colonne di \mathbf{A} che ha per coefficienti le componenti di \mathbf{x} . Quindi un vettore \mathbf{b} di \mathbb{K}^m è della forma \mathbf{Ax} se e solo se è combinazione lineare delle colonne di \mathbf{A} , cioè appartiene allo spazio colonna di \mathbf{A} . ■

Lo spazio riga e lo spazio colonna di una matrice, in quanto sottospazi di \mathbb{K}^n e \mathbb{K}^m rispettivamente, hanno dimensione finita. Il nostro obiettivo è ora mostrare che la dimensione di entrambi coincide col rango della matrice: questo è il risultato su cui si fonda l'algebra lineare. Ricordiamo che abbiamo definito il rango $r(\mathbf{A})$ di \mathbf{A} come il numero di righe non nulle della matrice a scala \mathbf{U} ottenuta da \mathbf{A} mediante l'algoritmo di eliminazione di Gauss.

DEFINIZIONE 8.3 (Rango per righe e rango per colonne)

Il *rango per righe* di una matrice \mathbf{A} è la *dimensione* di $\text{Row}(\mathbf{A})$, quindi il massimo numero di righe linearmente indipendenti di \mathbf{A} .

Il *rango per colonne* di una matrice \mathbf{A} è la *dimensione* di $\text{Col}(\mathbf{A})$, quindi il massimo numero di colonne linearmente indipendenti di \mathbf{A} .

Cominciamo col mostrare che il rango di \mathbf{A} coincide con il rango per righe. Sostanzialmente, questo significa che il MEG seleziona le righe indipendenti della matrice, eliminando quelle che sono combinazioni lineari delle precedenti.

PROPOSIZIONE 8.4 Sia \mathbf{A} una matrice qualsiasi e sia \mathbf{U} una matrice a scala ottenuta da \mathbf{A} con operazioni elementari sulle righe. Allora gli spazi riga di \mathbf{A} e \mathbf{U} coincidono:

$$\text{Row}(\mathbf{A}) = \text{Row}(\mathbf{U})$$

Inoltre, le righe non nulle della matrice a scala \mathbf{U} formano una base di $\text{Row}(\mathbf{A})$.

In particolare, il rango per righe di \mathbf{A} è uguale al rango di \mathbf{A} .

DIMOSTRAZIONE. Denotiamo con $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ le righe di \mathbf{A} . Per definizione

$$\text{Row}(\mathbf{A}) = \mathcal{L}(\mathbf{a}_1, \dots, \mathbf{a}_m)$$

è l'insieme di tutte le combinazioni lineari di $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$.

Cominciamo a mostrare che $\text{Row}(\mathbf{A}) = \text{Row}(\mathbf{U})$. Per questo occorre mostrare che, se \mathbf{B} è una matrice ottenuta da \mathbf{A} mediante un'operazione elementare sulle righe, allora $\text{Row}(\mathbf{A}) = \text{Row}(\mathbf{B})$. Abbiamo due casi da considerare:

Caso 1 La matrice \mathbf{B} è ottenuta scambiando due righe di \mathbf{A} .

In questo caso l'insieme delle righe di \mathbf{A} coincide con l'insieme delle righe di \mathbf{B} per cui $\text{Row}(\mathbf{A}) = \text{Row}(\mathbf{B})$.

Caso 2 La matrice \mathbf{B} è ottenuta sostituendo la riga \mathbf{a}_i con $\mathbf{a}_i + t\mathbf{a}_j$, dove $t \in \mathbb{K}$ e $i \neq j$.

In questo caso le righe di \mathbf{B} sono combinazione lineare delle righe di \mathbf{A} , e quindi appartengono allo spazio riga di \mathbf{A} . Ma allora lo spazio riga di \mathbf{A} contiene dei generatori dello spazio riga di \mathbf{B} e quindi contiene lo spazio riga di \mathbf{B} per il corollario 4.4.

D'altra parte, \mathbf{A} si ottiene da \mathbf{B} sostituendo la riga \mathbf{b}_i con $\mathbf{b}_i - t\mathbf{b}_j$. Il ragionamento precedente mostra $\text{Row}(\mathbf{A}) \subseteq \text{Row}(\mathbf{B})$. Confrontando le due inclusioni concludiamo $\text{Row}(\mathbf{B}) = \text{Row}(\mathbf{A})$.

Abbiamo mostrato che le operazioni elementari sulle righe lasciano lo spazio riga invariato. In particolare, se \mathbf{U} è una matrice a scala ottenuta da \mathbf{A} con operazioni elementari delle righe, $\text{Row}(\mathbf{A}) = \text{Row}(\mathbf{U})$, e perciò le righe non nulle di \mathbf{U} formano un insieme di generatori di $\text{Row}(\mathbf{A})$.

Rimane da dimostrare che le righe non nulle di \mathbf{U} sono linearmente indipendenti. Facciamo la dimostrazione nel caso in cui i pivots p_1, p_2, \dots, p_r di \mathbf{U} si trovino sulle prime r colonne:

$$\mathbf{U} = \begin{bmatrix} p_1 & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & p_2 & a_{23} & \dots & a_{2n} \\ 0 & 0 & p_3 & \dots & a_{3n} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & & \ddots \end{bmatrix}$$

La dimostrazione del caso generale (i pivots possono essere spostati più a destra), che è del tutto analoga, è lasciata al lettore.

Supponiamo che

$$t_1 \mathbf{u}_1 + t_2 \mathbf{u}_2 + t_3 \mathbf{u}_3 + \dots + t_r \mathbf{u}_r = \mathbf{0}$$

sia una relazione di dipendenza lineare tra le prime r righe di \mathbf{U} . Scrivendo la relazione componenti per componente si ottiene:

$$\left\{ \begin{array}{l} t_1 p_1 = 0 \\ t_1 a_{12} + t_2 p_2 = 0 \\ t_1 a_{13} + t_2 a_{23} + t_3 p_3 = 0 \\ \vdots \\ t_1 a_{1r} + t_2 a_{2r} + \cdots + t_r p_r = 0 \end{array} \right.$$

Dalla prima equazione, siccome il pivot p_1 è diverso da zero, otteniamo $t_1 = 0$. Sostituendo nella seconda equazione troviamo $t_2 p_2 = 0$. Ma $p_2 \neq 0$, quindi $t_2 = 0$. Sostituendo $t_1 = t_2 = 0$ nella terza equazione troviamo $t_3 p_3 = 0$ e quindi $t_3 = 0$. Continuando così si mostra che

$$t_1 = t_2 = \dots = t_r = 0$$

Le righe non nulle di U sono quindi linearmente indipendenti e questo conclude la dimostrazione. \blacksquare

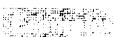
OSSERVAZIONE È ora evidente che il rango di \mathbf{A} non cambia se permutiamo le righe di \mathbf{A} , cosa non affatto ovvia invece dalla definizione di rango via MEG. Questo è uno dei vantaggi di pensare al rango come dimensione dello spazio riga, in modo quindi più concettuale, anche se meno operativo.

OSSERVAZIONE La proposizione 8.4 fornisce un algoritmo per trovare una base di un sottospazio di \mathbb{K}^n di cui siano noti dei generatori; in particolare consente di stabilire se i generatori sono linearmente indipendenti.

Supponiamo infatti che $\mathbf{v}_1, \dots, \mathbf{v}_m$ siano vettori di \mathbb{K}^n . Il sottospazio $\mathbf{H} = \mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_m)$ generato da $\mathbf{v}_1, \dots, \mathbf{v}_m$ è lo spazio riga della matrice \mathbf{A} che ha per righe i vettori $\mathbf{v}_1^T, \dots, \mathbf{v}_m^T$. Per la proposizione precedente, se \mathbf{U} è la matrice a scala ottenuta da \mathbf{A} col MEG, le righe non nulle di \mathbf{U} formano una base di $\mathbf{H} = \text{Row}(\mathbf{A})$. Sempre per la proposizione precedente la dimensione di \mathbf{H} è uguale al rango della matrice \mathbf{A} :

$$(8.1) \quad \dim \mathcal{L}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m) = r \left(\begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_m^T \end{bmatrix} \right)$$

In particolare, i vettori $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ sono linearmente indipendenti se e solo se $r(\mathbf{A}) = m$.



Sia \mathbf{H} il sottospazio di \mathbb{R}^4 generato dai vettori

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 3 \\ 1 \\ -1 \\ 4 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 4 \\ 1 \\ -2 \\ 3 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} 4 \\ 2 \\ 0 \\ 5 \end{bmatrix}$$

Allora \mathbf{H} è lo spazio riga della matrice

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 3 \\ 3 & 1 & -1 & 4 \\ 4 & 1 & -2 & 3 \\ 4 & 2 & 0 & 5 \end{bmatrix}$$

Riduciamo la matrice \mathbf{A} a scala:

$$\begin{bmatrix} 1 & 2 & 3 & 3 \\ 3 & 1 & -1 & 4 \\ 4 & 1 & -2 & 3 \\ 4 & 2 & 0 & 5 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 3 & 3 \\ 0 & -5 & -10 & -5 \\ 0 & -7 & -14 & -9 \\ 0 & -6 & -12 & -7 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 3 & 3 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 3 & 3 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Si noti che nel passare dalla seconda alla terza matrice abbiamo diviso la seconda riga per -5 ; questa non è un'operazione elementare del metodo di Gauss, ma la si può compiere perché non modifica lo spazio riga. Analogamente, nel passare dalla terza alla quarta matrice abbiamo diviso la terza riga per -2 .

Concludiamo che

$$\dim(\mathbf{H}) = r(\mathbf{A}) = 3$$

e una base di \mathbf{H} è formata dai vettori

$$\mathbf{u}_1 = \mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

In particolare, i quattro vettori dati sono linearmente dipendenti. La riduzione a scala di \mathbf{A} mostra anche che i primi tre sono indipendenti (se avessimo omesso l'ultima riga, avremmo ugualmente ottenuto una matrice a scala con tre righe non nulle). Quindi anche $\mathbf{v}_1, \mathbf{v}_2$ e \mathbf{v}_3 sono linearmente indipendenti e formano una base di \mathbf{H} (i vettori \mathbf{u}_i sono però più semplici da usare per le molte componenti nulle).

Abbiamo così stabilito che il rango di una matrice è il massimo numero di righe linearmente indipendenti. Nel capitolo sui sistemi lineari abbiamo visto che l'insieme delle soluzioni di un sistema omogeneo dipende da $n-r$ parametri, dove n è il numero delle incognite e r è il rango della matrice del sistema: ogni equazione indipendente dalle precedenti permette di determinare una delle incognite in funzione delle altre. Nel linguaggio di questo capitolo otteniamo una relazione tra la dimensione dello spazio riga (il rango) e la dimensione del suo nucleo, che per definizione è la *nullità* della matrice: la somma di rango e nullità è uguale al numero di colonne della matrice.

TEOREMA 8.5 (Teorema di nullità più rango)

Se \mathbf{A} ha n colonne, allora

$$\dim \text{Ker}(\mathbf{A}) + r(\mathbf{A}) = n$$

DIMOSTRAZIONE. Per la proposizione 8.4 il rango per righe di \mathbf{A} coincide col numero r delle righe non nulle della sua riduzione a scala \mathbf{U} . Per il teorema 6.8 di Rouché-Capelli esistono $n-r$ vettori $\mathbf{v}_1, \dots, \mathbf{v}_{n-r}$ di \mathbb{K}^n tali che ogni vettore $\mathbf{v} \in \text{Ker}(\mathbf{A})$ si scrive in uno e un solo modo nella forma

$$\mathbf{v} = t_1 \mathbf{v}_1 + \cdots + t_{n-r} \mathbf{v}_{n-r}.$$

Questo significa che $\mathbf{v}_1, \dots, \mathbf{v}_{n-r}$ formano una base di $\text{Ker}(\mathbf{A})$ e perciò $\dim \text{Ker}(\mathbf{A}) = n - r$.

OSSERVAZIONE Si osservi il fatto che il rango di \mathbf{A} dipenda solo da $\text{Ker}(\mathbf{A})$, cioè dall'insieme delle soluzioni del sistema lineare $\mathbf{Ax} = \mathbf{0}$, è conseguenza dei risultati di questo capitolo: infatti solo in questo capitolo abbiamo mostrato che le basi di $\text{Ker}(\mathbf{A})$ hanno tutte lo stesso numero di elementi.

Il nostro prossimo obiettivo è di mostrare che il rango è anche *il massimo numero di colonne linearmente indipendenti della matrice*. In altre parole, lo spazio colonna e lo spazio riga di una matrice hanno la stessa dimensione. Questo fatto non è per nulla banale: lo spazio riga e lo spazio colonna possono essere molto diversi. Se la matrice non è quadrata, le righe hanno addirittura un numero di componenti diverso da quello delle colonne. Ma anche nel caso di una matrice quadrata i due spazi in generale non coincidono: quello che hanno in comune è la dimensione. Per esempio, se $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$, lo spazio riga consiste dei vettori di \mathbb{K}^2 che hanno le coordinate uguali, mentre lo spazio colonna consiste dei vettori che hanno la seconda coordinata uguale all'opposto della prima: si tratta delle due rette $x = y$ e $x = -y$.

Per procedere dobbiamo mettere in evidenza che intercorre tra nucleo e spazio colonna di una matrice: l'osservazione fondamentale è che *i vettori del nucleo sono precisamente i vettori dei coefficienti delle relazioni lineari tra le colonne*. Per essere precisi, sia \mathbf{A} una matrice di tipo (m, n) e siano

$$\mathbf{v}_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}, \quad \dots, \quad \mathbf{v}_n = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

le sue colonne. Dato un vettore $\mathbf{x} \in \mathbb{K}^n$ di componenti x_1, \dots, x_n , il prodotto \mathbf{Ax} è la combinazione lineare delle colonne di \mathbf{A} che ha come coefficienti le componenti di \mathbf{x} :

$$\mathbf{Ax} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2 + \cdots + x_n \mathbf{v}_n$$

La condizione che \mathbf{x} appartenga al nucleo è perciò:

$$x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2 + \cdots + x_n \mathbf{v}_n = \mathbf{0}$$

Quindi $\mathbf{x} \in \text{Ker}(\mathbf{A})$ se e solo se le componenti di \mathbf{x} sono i coefficienti di una relazione lineare tra le colonne di \mathbf{A} .

Esempio

Consideriamo la matrice

Il vettore $\mathbf{v} = [2, -1]^T$ appartiene al nucleo della matrice e corrisponde alla relazione tra le colonne:

$$2\mathbf{v}_1 + (-1)\mathbf{v}_2 = [0, 0]^T$$

ovvero al fatto che la seconda colonna è uguale alla prima colonna moltiplicata per due.

Possiamo finalmente dimostrare:

TEOREMA 8.6 Il rango per righe e il rango per colonne di una matrice \mathbf{A} coincidono:

$$\dim \text{Row}(\mathbf{A}) = \dim \text{Col}(\mathbf{A})$$

DIMOSTRAZIONE. Sia \mathbf{U} la matrice ottenuta riducendo a scala la matrice \mathbf{A} col metodo di eliminazione di Gauss. Grazie alla proposizione 8.4 basta verificare che $\dim \text{Col}(\mathbf{A}) = r$, dove r è il numero di pivots di \mathbf{U} . Dividiamo la dimostrazione in due passi. Nel primo passo mostriamo che $\dim \text{Col}(\mathbf{A}) = \dim \text{Col}(\mathbf{U})$. Nel secondo passo mostriamo che una base di $\text{Col}(\mathbf{U})$ è formata dalle r colonne a cui appartengono gli r pivots di \mathbf{U} . Quindi

$$\dim \text{Col}(\mathbf{A}) = \dim \text{Col}(\mathbf{U}) = r$$

come volevasi dimostrare.

Primo passo Facciamo vedere che $\dim \text{Col}(\mathbf{A}) = \dim \text{Col}(\mathbf{U})$. La difficoltà è che in generale gli spazi colonna di \mathbf{A} e \mathbf{U} sono diversi: per esempio se

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

allora

$$\mathbf{U} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

e lo spazio colonna di \mathbf{A} è la retta di equazione $y = x$ in \mathbb{K}^2 , mentre lo spazio colonna di \mathbf{U} è l'asse delle ascisse, di equazione $y = 0$.

Quello che \mathbf{A} e \mathbf{U} hanno in comune è il nucleo:

$$\text{Ker}(\mathbf{A}) = \text{Ker}(\mathbf{U})$$

perché il nucleo $\text{Ker}(\mathbf{A})$ è l'insieme delle soluzioni del sistema omogeneo $\mathbf{Ax} = \mathbf{0}$ e tale insieme è lasciato immutato dalle operazioni elementari sulle righe.

Sappiamo che i vettori del nucleo corrispondono alle relazioni di dipendenza lineare delle colonne. Nell'esempio il vettore

$$\mathbf{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

appartiene a $\text{Ker}(\mathbf{A}) = \text{Ker}(\mathbf{U})$ e corrisponde alla relazione che la prima colonna è uguale alla seconda, che vale tanto per \mathbf{A} quanto per \mathbf{U} . Come conseguenza, un insieme di colonne di \mathbf{A} è linearmente indipendente se e solo se il corrispondente insieme di colonne di \mathbf{U} è linearmente indipendente. Per esempio le prime tre colonne di \mathbf{A} sono indipendenti se e solo se le prime tre colonne di \mathbf{U} sono indipendenti, perché questo significa che nel nucleo non si trova nessun vettore con le prime tre componenti non tutte nulle e le altre uguali a zero.

Siccome la dimensione dello spazio colonna è il massimo numero di colonne linearmente indipendenti, possiamo concludere che $\dim \text{Col}(\mathbf{A}) = \dim \text{Col}(\mathbf{U})$.

Secondo passo Mostriamo che una base di $\text{Col}(\mathbf{U})$ è formata dalle r colonne a cui appartengono gli r pivots di \mathbf{U} . Osserviamo che \mathbf{U} ha r righe non nulle. Cancellando eventualmente le righe nulle, cosa che non influisce sull'enunciato da dimostrare, possiamo supporre che \mathbf{U} sia una matrice con r righe. Quindi le colonne di \mathbf{U} sono vettori di \mathbb{K}^r e al massimo r di esse possono essere indipendenti: basta allora dimostrare che le r colonne coi pivots sono linearmente indipendenti per concludere che la dimensione dello spazio colonna è r .

Per dimostrare che le colonne coi pivots di \mathbf{U} sono linearmente indipendenti, consideriamo la matrice \mathbf{T} ottenuta da \mathbf{U} cancellando le colonne senza pivots. La matrice \mathbf{T} è quadrata di ordine r , è triangolare alta e sulla diagonale principale ha gli r pivots di \mathbf{U} . Quindi \mathbf{T} ha rango massimo e il suo nucleo è ridotto al vettore nullo (teorema di Cramer, non ci sono variabili libere!). Siccome il nucleo di \mathbf{T} è ridotto al vettore nullo, le r colonne di \mathbf{T} sono linearmente indipendenti. Ma le colonne di \mathbf{T} sono le colonne coi pivots di \mathbf{U} . Quindi le colonne coi pivots di \mathbf{U} sono linearmente indipendenti e questo conclude la dimostrazione. 

COROLLARIO 8.7 Il rango di una matrice \mathbf{A} coincide col rango della matrice trasposta \mathbf{A}^T .

DIMOSTRAZIONE. Questo corollario è un altro modo di enunciare il teorema 8.6. La matrice trasposta ha per righe le colonne di \mathbf{A} , per cui lo spazio riga di \mathbf{A}^T coincide con lo spazio colonna di \mathbf{A} . Il rango di \mathbf{A}^T è la dimensione di $\text{Row}(\mathbf{A}^T) = \text{Col}(\mathbf{A})$, che coincide col rango di \mathbf{A} per il teorema 8.6. 

OSSERVAZIONE In conclusione, il rango di una matrice coincide sia con la dimensione dello spazio riga sia con la dimensione dello spazio colonna. In particolare, il rango di una matrice non cambia se si scambiano delle colonne della matrice, perché lo spazio colonna di una matrice non dipende dall'ordine in cui si prendono le colonne. 

OSSERVAZIONE Supponiamo che \mathbf{A} sia una matrice e che \mathbf{U} sia una matrice a scala ottenuta da \mathbf{A} con operazioni elementari sulle righe. La dimostrazione del teorema 8.6 mostra che le colonne di \mathbf{U} su cui giacciono i pivots sono linearmente indipendenti, e che questo è vero anche delle corrispondenti colonne della matrice \mathbf{A} , perché le relazioni lineari tra le colonne delle due matrici sono le stesse. Quindi:

Siano $j(1), \dots, j(r)$ gli indici di colonna dei pivots di \mathbf{U} . Le colonne della matrice \mathbf{A} di indice $j(1), \dots, j(r)$ sono linearmente indipendenti e formano una base dello spazio colonna $\text{Col}(\mathbf{A})$.

Si ottiene così un algoritmo per estrarre una base da un insieme di generatori $\mathbf{v}_1, \dots, \mathbf{v}_m$ di un sottospazio $\mathbf{H} \subseteq \mathbb{K}^n$ (cf. il corollario 6.10.a). Si considera la matrice \mathbf{A} che ha per colonne i vettori $\mathbf{v}_1, \dots, \mathbf{v}_m$, in modo che $\mathbf{H} = \text{Col}(\mathbf{A})$. Si riduce quindi \mathbf{A} a una matrice a scala \mathbf{U} . Se $j(1), \dots, j(r)$ sono gli indici di colonna dei pivots di \mathbf{U} , i vettori $\mathbf{v}_{j(1)}, \dots, \mathbf{v}_{j(r)}$ formano una base di $\text{Col}(\mathbf{A}) = \mathbf{H}$.



Sia \mathbf{H} il sottospazio di \mathbb{R}^4 generato dalle colonne della matrice

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 3 \\ 3 & 1 & -1 & 4 \\ 4 & 1 & -2 & 3 \\ 4 & 2 & 0 & 5 \end{bmatrix}$$

Come abbiamo visto in precedenza, la matrice \mathbf{A} si riduce con operazioni elementari di riga alla matrice a scala

$$\mathbf{U} = \begin{bmatrix} 1 & 2 & 3 & 3 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Questo calcolo permette di concludere che

- a) $\dim(\mathbf{H}) = r(\mathbf{A}) = 3$;
- b) una base di \mathbf{H} è formata dalla prima, seconda e quarta colonna della matrice \mathbf{A} , perché i pivots di \mathbf{U} si trovano sulla prima, seconda e quarta colonna;
- c) la terza colonna di \mathbf{A} è combinazione lineare delle prime due: infatti, se si elimina la quarta colonna, si ottiene una matrice di rango 2, per cui le prime tre colonne sono linearmente dipendenti. Le prime due sono indipendenti, per cui la terza dev'essere una combinazione lineare delle prime due.

Come corollario otteniamo un criterio per stabilire se le colonne di una matrice sono linearmente indipendenti e se un vettore $\mathbf{b} \in \mathbb{K}^m$ è combinazione lineare delle colonne:

COROLLARIO 8.8 Siano $\mathbf{v}_1, \dots, \mathbf{v}_n$ vettori di \mathbb{K}^m e sia \mathbf{A} la matrice $m \times n$ che ha questi vettori come colonne. Allora

- a) i vettori $\mathbf{v}_1, \dots, \mathbf{v}_n$ sono linearmente indipendenti se e solo se $r(\mathbf{A}) = n$;
- b) un vettore $\mathbf{b} \in \mathbb{K}^m$ è combinazione lineare di $\mathbf{v}_1, \dots, \mathbf{v}_n$ se e solo se $r([\mathbf{A}|\mathbf{b}]) = r([\mathbf{A}])$.

DIMOSTRAZIONE. I vettori $\mathbf{v}_1, \dots, \mathbf{v}_n$ sono linearmente indipendenti se e solo se lo spazio da essi generato ha dimensione n ; lo spazio generato dai vettori è lo spazio colonna di \mathbf{A} che ha dimensione $r(\mathbf{A}) = n$. Questo mostra il primo punto.

Il secondo punto naturalmente segue dal teorema di Rouché-Capelli, ma ne vogliamo dare una dimostrazione più diretta (che usa il fatto che il rango è la dimensione dello spazio colonna, mentre nella dimostrazione di Rouché-Capelli il rango è sostanzialmente la dimensione dello spazio riga). Per il corollario 4.5, il vettore \mathbf{b} è combinazione lineare di $\mathbf{v}_1, \dots, \mathbf{v}_n$ se e solo se il sottospazio generato da $\mathbf{v}_1, \dots, \mathbf{v}_n$ e da \mathbf{b} coincide col sottospazio (in generale più piccolo) generato da $\mathbf{v}_1, \dots, \mathbf{v}_n$. Questi sottospazi coincidono (per il corollario 6.12) se hanno la stessa dimensione e le loro dimensioni sono, rispettivamente, $r(\mathbf{A})$ e $r([\mathbf{A}|\mathbf{b}])$. ■



Sia \mathbf{A} una matrice quadrata di ordine n e rango n . Mostrare che l'insieme delle righe (rispettivamente l'insieme delle colonne) di \mathbf{A} è una base di \mathbb{K}^n .

 Sia \mathbf{A} una matrice 2×3 di rango 2. Mostrare che le due righe di \mathbf{A} sono linearmente indipendenti e che le tre colonne di \mathbf{A} sono linearmente dipendenti. Mostrare con un esempio che non è detto né che le prime due colonne di \mathbf{A} siano linearmente indipendenti né che la terza colonna sia combinazione lineare delle prime due.

 Sia \mathbf{A} la matrice:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 3 & 4 \\ -3 & 4 & 3 & 4 \\ 1 & 3 & 6 & 9 \end{bmatrix}$$

Trovare una base dello spazio riga e una base dello spazio colonna di \mathbf{A} . Ci sono delle righe che sono combinazioni lineari delle righe precedenti? E colonne?

 Trovare una base per lo spazio riga e una base per lo spazio colonna della matrice

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

 Nella matrice

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 2 \\ 1 & 0 & 1 \end{bmatrix}$$

la terza colonna è la somma delle prime due. Possiamo dedurre che la terza riga è combinazione lineare delle prime due? In caso affermativo trovare i coefficienti di tale combinazione lineare.

Suggerimento: la matrice ha rango 2 e le prime due righe sono linearmente indipendenti.

 Siano \mathbf{A} e \mathbf{B} due matrici per cui il prodotto \mathbf{AB} è definito. Mostrare che $r(\mathbf{AB}) \leq r(\mathbf{A})$ e $r(\mathbf{AB}) \leq r(\mathbf{B})$. Mostrare anche che, se una delle due matrici è invertibile, allora vale l'uguaglianza in entrambe le diseguaglianze.

Suggerimento: mostrare che $\text{Ker}(\mathbf{B}) \subseteq \text{Ker}(\mathbf{AB})$ e poi usare il teorema di nullità più rango per concludere $r(\mathbf{AB}) \leq r(\mathbf{B})$; la prima diseguaglianza segue da questa passando alle matrici trasposte.

 Stabilire se i vettori che compongono i seguenti insiemi sono linearmente indipendenti o meno.

$$a) \left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} -2 \\ -4 \\ 2 \end{bmatrix} \right\}, \quad b) \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} \right\}.$$

 Determinare una base per il sottospazio di \mathbb{R}^4 generato dai vettori

$$\left\{ \begin{bmatrix} 1 \\ -3 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ -4 \\ -2 \\ -4 \end{bmatrix}, \begin{bmatrix} -1 \\ 4 \\ -3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -7 \\ 7 \\ 2 \end{bmatrix} \right\}.$$

 Sia \mathbf{V} lo spazio vettoriale dei polinomi a coefficienti reali. Dati $g(x) = 1+x$, $h(x) = x+x^2$, $k(x) = 3+2x-x^2$, $l(x) = 2$, verificare che il sottospazio di \mathbf{V} generato da g, h, k, l ha

dimensione 3 ed è quindi l'insieme dei polinomi di grado ≤ 2 . Estrarre da g, h, k, l una base di tale sottospazio

Suggerimento: fissare una base di \mathbf{V} e identificare il sottospazio con lo spazio colonna della matrice che ha per colonne i vettori delle coordinate dei quattro polinomi.

Determinare i valori da attribuire a α, β e γ affinché ciascuno degli insiemi sottoelencati sia formato da vettori linearmente dipendenti.

$$\left\{ \begin{bmatrix} \alpha^2 + \alpha \\ \alpha^2 - 1 \end{bmatrix} \right\}, \quad \left\{ \begin{bmatrix} 2 \\ -3 \\ 6 \end{bmatrix}, \begin{bmatrix} -4 \\ 6 \\ \beta \end{bmatrix} \right\}, \quad \left\{ \begin{bmatrix} 3 \\ 8 \\ -11 \end{bmatrix}, \begin{bmatrix} 1 \\ \gamma \\ -3 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -9 \end{bmatrix} \right\}$$

(Risposta: $\alpha = -1, \beta = -12, \gamma = 3$).

Determinare il massimo numero di vettori linearmente indipendenti nei due insiemi:

$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \\ 5 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 3 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 0 \\ -4 \\ -3 \end{bmatrix}, \begin{bmatrix} 1 \\ 5 \\ -4 \\ 18 \\ -3 \end{bmatrix} \right\}, \quad \mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 8 \\ 4 \\ 2 \\ -1 \\ 6 \end{bmatrix}, \begin{bmatrix} -3 \\ -2 \\ -1 \\ 6 \\ -2 \\ -5 \end{bmatrix}, \begin{bmatrix} -1 \\ 6 \\ 3 \\ 4 \\ -4 \\ -4 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ -1 \\ 3 \\ 6 \\ 1 \end{bmatrix} \right\}$$

9 EQUAZIONI CARTESIANE DI UN SOTTOSPAZIO

In questo paragrafo spieghiamo come ottenere equazioni cartesiane di un sottospazio di \mathbb{K}^n quando se ne conosca una base. Come i punti di una retta in \mathbb{R}^3 possono essere descritti da un parametro oppure da due equazioni cartesiane, così un sottospazio di dimensione d in \mathbb{K}^n può essere descritto da d parametri oppure da $n - d$ equazioni cartesiane. Si tratta di descrizioni diverse: le equazioni parametriche esprimono le coordinate dei vettori del sottospazio in funzione dei parametri e sono perciò utili per produrre elementi del sottospazio; le equazioni cartesiane esprimono invece delle condizioni che devono essere soddisfatte dai vettori del sottospazio e sono utili quando si deve verificare se un dato vettore appartiene al sottospazio. Tuttavia è chiaro che quando d è piccolo rispetto a n , il numero dei parametri è piccolo rispetto a quello delle equazioni cartesiane e, quindi, le equazioni parametriche sono di regola più semplici da usare nei conti; viceversa quando d è grande rispetto a n , le equazioni cartesiane sono molte meno dei parametri e vanno di regola preferite: già nel caso di \mathbb{R}^3 il più delle volte si usano le equazioni parametriche per le rette ($d = 1$) e l'equazione cartesiana ($n - d = 1$) per i piani.

Vogliamo mostrare che un sottospazio \mathbf{H} di dimensione d di \mathbb{K}^n ammette delle equazioni cartesiane e, più precisamente, che è l'insieme delle soluzioni di un sistema di $n - d$ equazioni lineari omogenee e quindi il nucleo di una matrice di rango $n - d$. Si pensi all'esempio di una retta per l'origine in \mathbb{R}^3 : per individuarla sono necessarie 2 = 3 - 1 equazioni cartesiane. In generale, le $r = n - d$ equazioni di un sottospazio di

\mathbb{K}^n avranno la forma:

$$(9.1) \quad \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0 \\ \vdots \quad \vdots \quad \vdots \quad \vdots \\ a_{r,1}x_1 + a_{r,2}x_2 + \cdots + a_{r,n}x_n = 0 \end{cases}$$

Dire che (9.1) sono le equazioni cartesiane di \mathbf{H} è equivalente a dire che è il nucleo della matrice $\mathbf{A} = [a_{ij}]$.

Nel cercare le equazioni cartesiane di \mathbf{H} , il punto fondamentale è che un'equazione lineare

$$(9.2) \quad a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0$$

può essere vista come un'equazione nelle incognite x_1, \dots, x_n con coefficienti a_1, \dots, a_n , ma anche come un'equazione nelle incognite a_1, \dots, a_n con coefficienti x_1, \dots, x_n . Per trovare un'equazione cartesiana di \mathbf{H} occorre considerare $\mathbf{a} = [a_1, a_2, \dots, a_n]^T$ come vettore delle incognite e imporre che \mathbf{a} soddisfi la (9.2) per *ogni* vettore $\mathbf{v} = [x_1, x_2, \dots, x_n]^T$ di \mathbf{H} .

PROPOSIZIONE 9.1 (Equazioni cartesiane di un sottospazio)

Ogni sottospazio \mathbf{H} di dimensione d di \mathbb{K}^n è definito da $n-d$ equazioni lineari: esiste una matrice \mathbf{A} a coefficienti in \mathbb{K} di tipo $(n-d, n)$ e rango $n-d$ tale che

$$\mathbf{H} = \text{Ker}(\mathbf{A}).$$

DIMOSTRAZIONE. Diamo un algoritmo per la costruzione della matrice \mathbf{A} a partire da una base $\mathbf{v}_1, \dots, \mathbf{v}_d$ di \mathbf{H} :

- a) Si considera la matrice $\mathbf{B} = \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_d^T \end{bmatrix}$ che ha come righe i vettori della base. È una matrice di tipo (d, n) ; siccome le righe sono linearmente indipendenti, $r(\mathbf{B}) = d$. Il nucleo $\text{Ker}(\mathbf{B})$ è perciò un sottospazio di \mathbb{K}^n di dimensione $n - d$.
- b) Risolvendo il sistema lineare $\mathbf{B}\mathbf{x} = \mathbf{0}$ si determina una base $\mathbf{a}_1, \dots, \mathbf{a}_{n-d}$ di $\text{Ker}(\mathbf{B})$.
- c) I vettori $\mathbf{a}_1, \dots, \mathbf{a}_{n-d}$ sono i vettori dei coefficienti delle equazioni cartesiane di \mathbf{H} : in altri termini, posto $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_{n-d}^T \end{bmatrix}$, si ha $\mathbf{H} = \text{Ker}(\mathbf{A})$.

Dobbiamo spiegare perché \mathbf{H} coincide con l'insieme delle soluzioni $\text{Ker}(\mathbf{A})$ del sistema lineare $\mathbf{Ax} = \mathbf{0}$. Osserviamo che \mathbf{A} ha rango $n-d$ perché per costruzione le sue righe sono linearmente indipendenti. Quindi $\text{Ker}(\mathbf{A})$ è un sottospazio di \mathbb{K}^n di dimensione $n - (n - d) = d$, la stessa di \mathbf{H} .

Per costruzione, i vettori $\mathbf{a}_1, \dots, \mathbf{a}_{n-d}$ appartengono al nucleo di \mathbf{B} :

$$\mathbf{B}\mathbf{a}_k = \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_d^T \end{bmatrix} \mathbf{a}_k = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{per ogni } k = 1, 2, \dots, n-d$$

cioè

$$(9.3) \quad \mathbf{v}_1^T \mathbf{a}_k = \mathbf{v}_2^T \mathbf{a}_k = \cdots = \mathbf{v}_d^T \mathbf{a}_k = 0 \quad \text{per ogni } k = 1, 2, \dots, n-d$$

Ora osserviamo che, posto

$$\mathbf{a} = [a_1, \dots, a_n]^T \quad \text{e} \quad \mathbf{v} = [x_1, \dots, x_n]^T$$

possiamo scrivere l'equazione $a_1x_1 + a_2x_2 + \dots + a_nx_n = 0$ sia nella forma $\mathbf{a}^T \mathbf{v} = 0$ sia nella forma $\mathbf{v}^T \mathbf{a} = 0$ perché

$$(9.4) \quad \mathbf{a}^T \mathbf{v} = a_1x_1 + a_2x_2 + \dots + a_nx_n = \mathbf{v}^T \mathbf{a}$$

Possiamo perciò riscrivere le equazioni (9.3) nella forma equivalente

$$\mathbf{a}_k^T \mathbf{v}_1 = \mathbf{a}_k^T \mathbf{v}_2 = \cdots = \mathbf{a}_k^T \mathbf{v}_d = 0 \quad \text{per ogni } k = 1, 2, \dots, n-d$$

Questo mostra che il prodotto di una riga di \mathbf{A} con ciascuno dei vettori colonna $\mathbf{v}_1, \dots, \mathbf{v}_d$ è nullo e quindi che il nucleo di \mathbf{A} contiene i vettori $\mathbf{v}_1, \dots, \mathbf{v}_d$. Siccome i vettori $\mathbf{v}_1, \dots, \mathbf{v}_d$ generano \mathbf{H} , il nucleo $\text{Ker}(\mathbf{A})$ contiene \mathbf{H} (corollario 4.4). Infine \mathbf{H} e $\text{Ker}(\mathbf{A})$ hanno la stessa dimensione d , perciò coincidono per il corollario 6.12.

Esempio

In \mathbb{R}^4 cerchiamo le equazioni cartesiane del sottospazio \mathbf{H} generato dai due vettori $\mathbf{v}_1 = [1, 2, 3, 4]^T$ e $\mathbf{v}_2 = [4, 3, 2, 1]^T$. Come spiegato nella dimostrazione occorre risolvere il sistema lineare

$$\begin{cases} a_1 + 2a_2 + 3a_3 + 4a_4 &= 0 \\ 4a_1 + 3a_2 + 2a_3 + a_4 &= 0 \end{cases}$$

la cui matrice dei coefficienti ha per righe i vettori \mathbf{v}_1 e \mathbf{v}_2 :

$$\mathbf{B} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{bmatrix}$$

Risolviamo il sistema riducendo \mathbf{B} a scala

$$\mathbf{U} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -5 & -10 & -15 \end{bmatrix}$$

Scegliendo a_3 e a_4 come variabili libere si ottiene la base di $\text{Ker}(\mathbf{B})$ formata dai vettori $\mathbf{a}_1 = [1, -2, 1, 0]^T$ e $\mathbf{a}_2 = [2, -3, 0, 1]^T$. Delle equazioni cartesiane di \mathbf{H} sono perciò

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



Trovare delle equazioni cartesiane per il sottospazio \mathbf{H} di \mathbb{R}^4 generato dai vettori $\mathbf{v}_1 = [1, 1, 0, 0]^T$, $\mathbf{v}_2 = [1, 0, 0, 1]^T$, $\mathbf{v}_3 = [0, 1, 1, 0]^T$ e stabilire per quale valore di k il vettore $[3, 12, -9, k]^T$ appartiene ad \mathbf{H} . Trovare delle equazioni cartesiane per il sottospazio generato da \mathbf{v}_1 e \mathbf{v}_2 (rispettivamente, da \mathbf{v}_1 e \mathbf{v}_3 , rispettivamente, da \mathbf{v}_2 e \mathbf{v}_3).

10 OPERAZIONI SUI SOTTOSPAZI

In questo paragrafo \mathbf{H} e \mathbf{K} denotano due sottospazi di uno spazio vettoriale \mathbf{V} sul campo \mathbb{K} . Possiamo formare l'intersezione e l'unione di \mathbf{H} e \mathbf{K} . L'intersezione $\mathbf{H} \cap \mathbf{K}$ consiste di quei vettori di \mathbf{V} che appartengono sia ad \mathbf{H} sia a \mathbf{K} . Per esempio, in \mathbb{R}^3 l'intersezione del piano xy col piano xz è l'asse delle x . L'unione $\mathbf{H} \cup \mathbf{K}$ consiste dei vettori di \mathbf{V} che appartengono ad almeno uno dei due sottospazi. Per esempio, in \mathbb{R}^2 l'unione dell'asse x e dell'asse y è l'insieme dei punti che appartengono a uno dei due assi cartesiani: evidentemente non si tratta di un sottospazio di \mathbb{R}^2 . L'intersezione di due sottospazi è invece sempre un sottospazio.

PROPOSIZIONE 10.1 (Intersezione di due sottospazi)

Siano \mathbf{H} e \mathbf{K} due sottospazi di uno spazio vettoriale \mathbf{V} .

L'intersezione

$$\mathbf{H} \cap \mathbf{K} = \{\mathbf{v} \in \mathbf{V} : \mathbf{v} \in \mathbf{H} \text{ e } \mathbf{v} \in \mathbf{K}\}$$

è un sottospazio di \mathbf{V} .

DIMOSTRAZIONE. Occorre verificare che $\mathbf{0} \in \mathbf{H} \cap \mathbf{K}$ e che $\mathbf{H} \cap \mathbf{K}$ è chiuso rispetto a somma e prodotto per uno scalare. Verifichiamo solo che $\mathbf{H} \cap \mathbf{K}$ è chiuso rispetto alla somma, visto che il resto della dimostrazione è simile. Supponiamo quindi che \mathbf{v} e \mathbf{w} siano due vettori di $\mathbf{H} \cap \mathbf{K}$. Allora \mathbf{v} e \mathbf{w} appartengono a \mathbf{H} e, quindi, $\mathbf{v} + \mathbf{w}$ appartiene a \mathbf{H} perché \mathbf{H} è un sottospazio. Per lo stesso motivo la somma $\mathbf{v} + \mathbf{w}$ appartiene anche a \mathbf{K} e, quindi, $\mathbf{v} + \mathbf{w} \in \mathbf{H} \cap \mathbf{K}$ come volevasi dimostrare. 

Autore: Giacomo Belotti
Titolo: Matematica per le scienze applicate
Capitolo: 10. Operazioni sui sottospazi

In \mathbb{R}^3 l'intersezione del piano xy e del piano yz è l'asse y . L'equazione del piano xy è $z = 0$, l'equazione del piano yz è $x = 0$, le equazioni cartesiane dell'asse y sono $x = z = 0$. L'intersezione del piano xy con l'asse z è l'origine.

Autore: Giacomo Belotti
Titolo: Matematica per le scienze applicate
Capitolo: 10. Operazioni sui sottospazi

Sia \mathbb{M} lo spazio vettoriale delle matrici quadrate di ordine n a coefficienti in \mathbb{K} . Siano: \mathbb{S} il sottospazio delle matrici simmetriche e \mathbb{T} il sottospazio delle matrici triangolari alte. Allora $\mathbb{S} \cap \mathbb{T}$ è il sottospazio delle matrici diagonali. Per convincersene, si consideri dapprima il caso $n = 2$: le matrici simmetriche hanno la forma $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$, quelle triangolari alte la forma $\begin{bmatrix} a & b \\ 0 & c \end{bmatrix}$ e quindi una matrice è sia simmetrica sia triangolare alta se e solo se ha la forma $\begin{bmatrix} a & 0 \\ 0 & c \end{bmatrix}$, cioè è diagonale. Il caso $n > 2$ è simile e lasciato al lettore.

Consideriamo ora il sottospazio \mathbb{E} formato dalle matrici antisimmetriche: si ricordi che si tratta delle matrici \mathbf{A} tali che $\mathbf{A}^T = -\mathbf{A}$. Per esempio, nel caso $n = 2$, le matrici antisimmetriche sono quelle della forma $\begin{bmatrix} 0 & b \\ -b & 0 \end{bmatrix}$. Una matrice \mathbf{A} che sia simmetrica e antisimmetrica al tempo stesso soddisfa le due uguaglianze $\mathbf{A}^T = \mathbf{A}$ e $\mathbf{A}^T = -\mathbf{A}$. Quindi $\mathbf{A} = -\mathbf{A}$ ed è perciò la matrice nulla: concludiamo che $\mathbb{S} \cap \mathbb{E}$ contiene solo la matrice nulla. Il lettore è invitato a verificare che anche $\mathbb{T} \cap \mathbb{E}$ contiene solo la matrice nulla: una matrice che sia triangolare alta e antisimmetrica è nulla.

Supponiamo che $\mathbf{H} = \text{Ker}(\mathbf{A})$ sia l'insieme delle soluzioni del sistema lineare omogeneo

$$(10.1) \quad \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0 \\ \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n = 0 \end{cases}$$

Denotiamo coi simboli $\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_m$ gli insiemi delle soluzioni delle singole equazioni del sistema. Allora

$$\mathbf{H} = \mathbf{H}_1 \cap \mathbf{H}_2 \cap \cdots \cap \mathbf{H}_m$$

Se le equazioni non sono banali, ciascun \mathbf{H}_i è un iperpiano per l'origine, cioè un sottospazio vettoriale di dimensione $n - 1$. Il fatto che ogni sottospazio di \mathbb{K}^n di dimensione d sia definito da $n - d$ equazioni lineari ha quindi la seguente interpretazione geometrica: *ogni sottospazio di \mathbb{K}^n di dimensione d è intersezione di $n - d$ iperpiani passanti per l'origine.*



Se \mathbf{H} e \mathbf{K} sono sottospazi di \mathbb{K}^n e conosciamo delle equazioni cartesiane di entrambi, è immediato scrivere delle equazioni per l'intersezione $\mathbf{H} \cap \mathbf{K}$: basta accostare le equazioni di \mathbf{H} e \mathbf{K} .

In termini di matrici, se $\mathbf{H} = \text{Ker}(\mathbf{A})$ e $\mathbf{K} = \text{Ker}(\mathbf{B})$, allora \mathbf{A} e \mathbf{B} hanno entrambe n colonne, e possiamo formare la matrice $\begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix}$ scrivendo le righe di \mathbf{B} sotto le righe di \mathbf{A} . Allora

$$\mathbf{H} \cap \mathbf{K} = \text{Ker} \left(\begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \right)$$

Si noti però che le equazioni di $\mathbf{H} \cap \mathbf{K}$ ottenute in questo modo possono essere troppe, cioè linearmente dipendenti. Per un esempio banale, possiamo prendere per \mathbf{H} il sottospazio di equazioni $x_1 = 0$ e $x_2 = 0$ e per \mathbf{K} il sottospazio di equazioni $x_2 = 0$ e $x_3 = 0$; accostando le equazioni otteniamo l'equazione $x_2 = 0$ due volte; per individuare il sottospazio $\mathbf{H} \cap \mathbf{K}$ bastano le tre equazioni $x_1 = 0$, $x_2 = 0$ e $x_3 = 0$. Se invece \mathbf{K}' è il sottospazio di equazioni $x_3 = 0$ e $x_4 = 0$, l'intersezione $\mathbf{H} \cap \mathbf{K}'$ non può essere definita da meno di 4 equazioni.

Questi esempi mostrano che la dimensione di $\mathbf{H} \cap \mathbf{K}$ non dipende soltanto dalla dimensione di \mathbf{H} e \mathbf{K} ; approfondiremo questo punto più avanti.

L'unione di due sottospazi non è in generale un sottospazio perché può non essere chiusa rispetto alla somma, come mostra l'esempio dell'unione dei due assi cartesiani di \mathbb{R}^2 mostra: i vettori $[1, 0]^T$ e $[0, 1]^T$ appartengono all'unione dei due assi cartesiani, ma la loro somma $[1, 1]^T$ non vi appartiene. Per rimediare a questo fatto si introduce il *sottospazio somma $\mathbf{H} + \mathbf{K}$* che è l'insieme costituito dai vettori che si possono scrivere come somma di un vettore di \mathbf{H} e di un vettore di \mathbf{K} :

$$\mathbf{H} + \mathbf{K} = \{\mathbf{v} \in \mathbf{V} : \text{esistono } \mathbf{h} \in \mathbf{H} \text{ e } \mathbf{k} \in \mathbf{K} \text{ tali che } \mathbf{v} = \mathbf{h} + \mathbf{k}\}$$

PROPOSIZIONE 10.2 (Somma di due sottospazi)

Siano \mathbf{H} e \mathbf{K} due sottospazi di uno spazio vettoriale \mathbf{V} .

- La somma $\mathbf{H} + \mathbf{K}$ è il più piccolo sottospazio di \mathbf{V} che contenga sia \mathbf{H} sia \mathbf{K} . Questo significa: $\mathbf{H} + \mathbf{K}$ è un sottospazio di \mathbf{V} , e, se \mathbf{W} è un sottospazio di \mathbf{V} che contiene \mathbf{H} e \mathbf{K} , allora \mathbf{W} contiene $\mathbf{H} + \mathbf{K}$.
- Se $\mathbf{h}_1, \dots, \mathbf{h}_d$ sono dei generatori di \mathbf{H} e $\mathbf{k}_1, \dots, \mathbf{k}_e$ sono dei generatori di \mathbf{K} , allora $\mathbf{h}_1, \dots, \mathbf{h}_d, \mathbf{k}_1, \dots, \mathbf{k}_e$ sono dei generatori di $\mathbf{H} + \mathbf{K}$.

DIMOSTRAZIONE. Mostriamo innanzitutto che $\mathbf{H} + \mathbf{K}$ è un sottospazio: dobbiamo verificare che $\mathbf{0} \in \mathbf{H} + \mathbf{K}$ e che $\mathbf{H} + \mathbf{K}$ è chiuso rispetto alla somma e al prodotto per uno scalare. Il vettore nullo appartiene sia a \mathbf{H} sia a \mathbf{K} e, siccome $\mathbf{0} = \mathbf{0} + \mathbf{0}$, appartiene anche a $\mathbf{H} + \mathbf{K}$. Verifichiamo che $\mathbf{H} + \mathbf{K}$ è chiuso rispetto alla somma. Consideriamo due vettori \mathbf{v}_1 e \mathbf{v}_2 di $\mathbf{H} + \mathbf{K}$. Possiamo scrivere $\mathbf{v}_1 = \mathbf{h}_1 + \mathbf{k}_1$ come somma di $\mathbf{h}_1 \in \mathbf{H}$ e $\mathbf{k}_1 \in \mathbf{K}$, e $\mathbf{v}_2 = \mathbf{h}_2 + \mathbf{k}_2$ con $\mathbf{h}_2 \in \mathbf{H}$ e $\mathbf{k}_2 \in \mathbf{K}$. Allora

$$\mathbf{v}_1 + \mathbf{v}_2 = (\mathbf{h}_1 + \mathbf{h}_2) + (\mathbf{k}_1 + \mathbf{k}_2)$$

Siccome \mathbf{h}_1 e \mathbf{h}_2 appartengono ad \mathbf{H} , anche la loro somma $\mathbf{h}_1 + \mathbf{h}_2$ appartiene ad \mathbf{H} perché \mathbf{H} è un sottospazio. Analogamente $\mathbf{k}_1 + \mathbf{k}_2 \in \mathbf{K}$ e quindi $\mathbf{v}_1 + \mathbf{v}_2$ appartiene ad $\mathbf{H} + \mathbf{K}$. La dimostrazione che $\mathbf{H} + \mathbf{K}$ è chiuso rispetto al prodotto per uno scalare è simile e viene lasciata al lettore.

Supponiamo ora che \mathbf{W} sia un sottospazio di \mathbf{V} che contenga sia \mathbf{H} sia \mathbf{K} e mostriamo che \mathbf{W} contiene $\mathbf{H} + \mathbf{K}$. Questo è immediato: se \mathbf{W} è un sottospazio, allora è chiuso rispetto alla somma e quindi contiene tutte le somme $\mathbf{w}_1 + \mathbf{w}_2$ con $\mathbf{w}_1, \mathbf{w}_2 \in \mathbf{W}$; se $\mathbf{H} \subseteq \mathbf{W}$ e $\mathbf{K} \subseteq \mathbf{W}$, allora \mathbf{W} contiene in particolare tutte le somme di un vettore di \mathbf{H} con un vettore di \mathbf{K} , cioè contiene $\mathbf{H} + \mathbf{K}$.

Infine, se ogni vettore di \mathbf{H} è combinazione lineare di $\mathbf{h}_1, \dots, \mathbf{h}_d$ e ogni vettore di \mathbf{K} è combinazione lineare di $\mathbf{k}_1, \dots, \mathbf{k}_e$, allora ogni vettore $\mathbf{h} + \mathbf{k}$ di $\mathbf{H} + \mathbf{K}$ è combinazione lineare di $\mathbf{h}_1, \dots, \mathbf{h}_d, \mathbf{k}_1, \dots, \mathbf{k}_e$: questi vettori sono perciò dei generatori del sottospazio somma.

In \mathbb{R}^3 la somma del piano xy e del piano yz è tutto \mathbb{R}^3 perché ogni vettore di \mathbb{R}^3 si può scrivere come somma

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ z \end{bmatrix}$$

di un vettore del piano xy e di uno del piano yz ; in realtà in più modi perché per esempio

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ y \\ z \end{bmatrix}$$

è un altro modo di scrivere $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ come somma di un vettore del piano xy e di uno del piano yz .

La somma del piano xy con l'asse z è di nuovo \mathbb{R}^3 , ma questa volta vi è un unico modo di decomporre un vettore di \mathbb{R}^3 come somma di un vettore del piano xy e di uno dell'asse z :

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ z \end{bmatrix}$$



Sia M lo spazio vettoriale delle matrici quadrate di ordine 2 a coefficienti in \mathbb{K} . Per ogni matrice $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M$ vale l'uguaglianza

$$(10.2) \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & \frac{b+c}{2} \\ \frac{b+c}{2} & d \end{bmatrix} + \begin{bmatrix} 0 & \frac{b-c}{2} \\ \frac{c-b}{2} & 0 \end{bmatrix}$$

Il primo addendo a secondo membro è una matrice simmetrica, il secondo addendo è una matrice antisimmetrica. Concludiamo che ogni matrice quadrata di ordine 2 è somma di una matrice simmetrica e di una matrice antisimmetrica. Quindi, se S (rispettivamente E) denota il sottospazio delle matrici simmetriche (rispettivamente antisimmetriche), il sottospazio somma $S + E$ è l'intero spazio M .

Questo esempio si generalizza al caso di matrici quadrate di un arbitrario ordine n . Infatti per ogni matrice quadrata \mathbf{A} vale l'uguaglianza

$$(10.3) \quad \mathbf{A} = \frac{\mathbf{A} + \mathbf{A}^T}{2} + \frac{\mathbf{A} - \mathbf{A}^T}{2}$$

A secondo membro, il primo addendo $S = \frac{\mathbf{A} + \mathbf{A}^T}{2}$ è una matrice simmetrica; il secondo addendo $E = \frac{\mathbf{A} - \mathbf{A}^T}{2}$ è una matrice antisimmetrica. Concludiamo che ogni matrice quadrata è somma di una matrice simmetrica e di una matrice antisimmetrica.

Anche quando $\mathbf{h}_1, \dots, \mathbf{h}_d$ formano una base di \mathbf{H} e $\mathbf{k}_1, \dots, \mathbf{k}_e$ una base di \mathbf{K} non è detto che $\mathbf{h}_1, \dots, \mathbf{h}_d, \mathbf{k}_1, \dots, \mathbf{k}_e$ siano ancora linearmente indipendenti e quindi formino una base di $\mathbf{H} + \mathbf{K}$. Per esempio in \mathbb{R}^3 possiamo prendere per \mathbf{H} il piano xy con la base $\{\mathbf{h}_1, \mathbf{h}_2\} = \{\mathbf{i}, \mathbf{j}\}$ e per \mathbf{K} il piano yz con la base $\{\mathbf{k}_1, \mathbf{k}_2\} = \{\mathbf{j}, \mathbf{k}\}$. I quattro vettori $\{\mathbf{h}_1, \mathbf{h}_2, \mathbf{k}_1, \mathbf{k}_2\} = \{\mathbf{i}, \mathbf{j}, \mathbf{j}, \mathbf{k}\}$ generano $\mathbf{H} + \mathbf{K}$, che è l'intero spazio vettoriale \mathbb{R}^3 , ma non ne formano una base.

Il problema è che la dimensione del sottospazio somma può essere minore della somma delle dimensioni di \mathbf{H} e di \mathbf{K} . Dimostriamo ora la *formula di Grassmann* che lega tra loro le dimensioni di $\mathbf{H} \cap \mathbf{K}$ e $\mathbf{H} + \mathbf{K}$.

TEOREMA 10.3 (Formula di Grassmann)

Siano \mathbf{H} e \mathbf{K} due sottospazi di uno spazio vettoriale \mathbf{V} . Supponiamo che \mathbf{H} e \mathbf{K} abbiano dimensione finita. Allora

$$\dim(\mathbf{H} \cap \mathbf{K}) + \dim(\mathbf{H} + \mathbf{K}) = \dim \mathbf{H} + \dim \mathbf{K}$$

DIMOSTRAZIONE. Per ipotesi \mathbf{H} e \mathbf{K} ammettono entrambi un numero finito di generatori. Per la proposizione precedente anche il sottospazio somma $\mathbf{H} + \mathbf{K}$ è generato da un numero finito di vettori. Quindi $\mathbf{H} + \mathbf{K}$ ha dimensione finita. Sia $m = \dim(\mathbf{H} + \mathbf{K})$. Fissata una base

di $\mathbf{H} + \mathbf{K}$, possiamo identificare un vettore di $\mathbf{H} + \mathbf{K}$ con la m -upla delle sue coordinate e quindi supporre $\mathbf{H} + \mathbf{K} = \mathbb{K}^m$ (si veda la proposizione 7.2). In particolare, possiamo scrivere gli elementi di \mathbf{H} e \mathbf{K} come vettori colonna con m componenti. Fissiamo una base $\mathbf{v}_1, \dots, \mathbf{v}_d$ di \mathbf{H} e consideriamo la matrice

$$\mathbf{P} = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_d]$$

Si tratta di una matrice di tipo (m, d) dove $d = \dim(\mathbf{H})$. Siccome le colonne di \mathbf{P} sono i vettori di una base di \mathbf{H} , lo spazio colonna $\text{Col}(\mathbf{P})$ coincide con \mathbf{H} . Analogamente, utilizzando un base di \mathbf{K} , costruiamo una matrice \mathbf{Q} di tipo (m, e) tale che $e = \dim(\mathbf{K})$ e $\mathbf{K} = \text{Col}(\mathbf{Q})$.

Per la proposizione precedente, il sottospazio somma $\mathbf{H} + \mathbf{K}$ è generato dalle colonne di \mathbf{P} e dalle colonne di \mathbf{Q} : in altre parole il sottospazio somma è la spazio colonna della matrice $[\mathbf{P} \ \mathbf{Q}]$ ottenuta aggiungendo le colonne di \mathbf{Q} a quelle di \mathbf{P} :

$$(10.4) \quad \mathbf{H} + \mathbf{K} = \text{Col}([\mathbf{P} \ \mathbf{Q}])$$

Siccome il rango di una matrice è la dimensione dello spazio colonna, otteniamo una formula per la dimensione di $\mathbf{H} + \mathbf{K}$:

$$(10.5) \quad \dim(\mathbf{H} + \mathbf{K}) = r([\mathbf{P} \ \mathbf{Q}])$$

Applicando il teorema di nullità più rango alla matrice $[\mathbf{P} \ \mathbf{Q}]$, che ha $d+e$ colonne otteniamo

$$(10.6) \quad \dim(\mathbf{H} + \mathbf{K}) = d + e - \dim \text{Ker}([\mathbf{P} \ \mathbf{Q}])$$

Poiché $\dim \mathbf{H} = d$ e $\dim \mathbf{K} = e$, la tesi è dimostrata se verifichiamo

$$(10.7) \quad \dim(\mathbf{H} \cap \mathbf{K}) = \dim \text{Ker}([\mathbf{P} \ \mathbf{Q}])$$

L'ultima equazione indica che la dimensione di $\mathbf{H} \cap \mathbf{K}$ è uguale al numero di relazioni di dipendenza lineare tra i vettori che formano le basi di \mathbf{H} e di \mathbf{K} , ed è il punto essenziale della dimostrazione. Per provarla, consideriamo un vettore \mathbf{v} dell'intersezione $\mathbf{H} \cap \mathbf{K}$. Poiché $\mathbf{v} \in \mathbf{H} = \text{Col}(\mathbf{P})$, esiste $\mathbf{t} \in \mathbb{K}^d$ tale che $\mathbf{v} = \mathbf{Pt}$; tale \mathbf{t} inoltre è univocamente determinato da \mathbf{v} perché le colonne di \mathbf{P} sono linearmente indipendenti (formano una base di \mathbf{H}). Analogamente, esiste uno e un solo $\mathbf{u} \in \mathbb{K}^e$ tale che $\mathbf{v} = \mathbf{Qu}$. Il vettore $[\begin{smallmatrix} \mathbf{t} \\ -\mathbf{u} \end{smallmatrix}] \in \mathbb{K}^{d+e}$ appartiene al nucleo della matrice $[\mathbf{P} \ \mathbf{Q}]$:

$$[\mathbf{P} \ \mathbf{Q}] \begin{bmatrix} \mathbf{t} \\ -\mathbf{u} \end{bmatrix} = \mathbf{Pt} - \mathbf{Qu} = \mathbf{v} - \mathbf{v} = \mathbf{0}$$

Viceversa, se il vettore $[\begin{smallmatrix} \mathbf{t} \\ -\mathbf{u} \end{smallmatrix}]$ appartiene al nucleo di $[\mathbf{P} \ \mathbf{Q}]$, il vettore $\mathbf{v} = \mathbf{Pt}$ appartiene ad $\mathbf{H} \cap \mathbf{K}$ perché:

$$\mathbf{0} = [\mathbf{P} \ \mathbf{Q}] \begin{bmatrix} \mathbf{t} \\ -\mathbf{u} \end{bmatrix} = \mathbf{Pt} - \mathbf{Qu} \quad \Rightarrow \quad \mathbf{Pt} = \mathbf{Qu} \in \text{Col}(\mathbf{P}) \cap \text{Col}(\mathbf{Q}) = \mathbf{H} \cap \mathbf{K}$$

In conclusione, la funzione

$$\text{Ker}([\mathbf{P} \ \mathbf{Q}]) \rightarrow \mathbf{H} \cap \mathbf{K}$$

che a un vettore $[\begin{smallmatrix} \mathbf{t} \\ -\mathbf{u} \end{smallmatrix}]$ del nucleo di $[\mathbf{P} \ \mathbf{Q}]$ associa il vettore \mathbf{Pt} è invertibile (iniettiva e suriettiva).

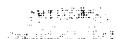
In particolare, se $[\begin{smallmatrix} \mathbf{t}_1 \\ -\mathbf{u}_1 \end{smallmatrix}], \dots, [\begin{smallmatrix} \mathbf{t}_s \\ -\mathbf{u}_s \end{smallmatrix}]$ è una base di $\text{Ker}([\mathbf{P} \ \mathbf{Q}])$, allora $\mathbf{Pt}_1, \dots, \mathbf{Pt}_s$ è una base di $\mathbf{H} \cap \mathbf{K}$. Quindi $\text{Ker}([\mathbf{P} \ \mathbf{Q}])$ e $\mathbf{H} \cap \mathbf{K}$ hanno la stessa dimensione e la dimostrazione è completa. ■



In \mathbb{R}^3 prendiamo per \mathbf{H} il piano xy e per \mathbf{K} il piano yz . L'intersezione $\mathbf{H} \cap \mathbf{K}$ è l'asse y che ha dimensione 1, quindi

$$\dim(\mathbf{H} + \mathbf{K}) = \dim(\mathbf{H}) + \dim(\mathbf{K}) - \dim(\mathbf{H} \cap \mathbf{K}) = 3$$

Quindi $\mathbf{H} + \mathbf{K}$ ha dimensione 3 e deve coincidere con \mathbb{R}^3 .



Nello spazio vettoriale \mathbb{M} delle matrici quadrate di ordine 2 consideriamo il sottospazio \mathbb{S} delle matrici simmetriche, il sottospazio \mathbb{T} delle matrici triangolari, il sottospazio \mathbb{E} delle matrici antisimmetriche e il sottospazio \mathbb{D} delle matrici diagonali.

a) Abbiamo visto che $\mathbb{S} \cap \mathbb{T} = \mathbb{D}$. Per la formula di Grassmann

$$\dim(\mathbb{S} + \mathbb{T}) = \dim(\mathbb{S}) + \dim(\mathbb{T}) - \dim(\mathbb{D}) = 3 + 3 - 2 = 4$$

Quindi $\mathbb{S} + \mathbb{T}$ ha dimensione 4 e deve perciò coincidere con l'intero spazio \mathbb{M} . Questo significa che ogni matrice quadrata di ordine 2 si può scrivere come somma di una matrice simmetrica e di una matrice triangolare. Il lettore è invitato a trovare esplicitamente una tale decomposizione.

b) Abbiamo visto che $\mathbb{S} \cap \mathbb{E} = \{\mathbf{0}\}$. Per la formula di Grassmann

$$\dim(\mathbb{S} + \mathbb{E}) = \dim(\mathbb{S}) + \dim(\mathbb{E}) - \dim(\mathbb{S} \cap \mathbb{E}) = 3 + 1 - 0 = 4$$

Quindi $\mathbb{S} + \mathbb{E}$ ha dimensione 4 e deve perciò coincidere con l'intero spazio \mathbb{M} . Ritroviamo così il fatto che ogni matrice quadrata di ordine 2 si può scrivere come somma di una matrice simmetrica e di una matrice antisimmetrica, senza bisogno di ricorrere alla decomposizione esplicita (10.2).

c) Anche $\mathbb{E} \cap \mathbb{T} = \{\mathbf{0}\}$: una matrice che sia antisimmetrica e triangolare alta dev'essere nulla. Per la formula di Grassmann

$$\dim(\mathbb{E} + \mathbb{T}) = \dim(\mathbb{E}) + \dim(\mathbb{T}) - \dim(\mathbb{E} \cap \mathbb{T}) = 1 + 3 - 0 = 4$$

Quindi $\mathbb{E} + \mathbb{T}$ ha dimensione 4 e deve perciò coincidere con l'intero spazio \mathbb{M} . Questo significa che ogni matrice quadrata di ordine 2 si può scrivere come somma di una matrice antisimmetrica e di una matrice triangolare. Il lettore è invitato a trovare esplicitamente una tale decomposizione.

Questi esempi si generalizzano al caso di matrici quadrate di ordine n (si tenga presente che $\dim(\mathbb{M}) = n^2$, $\dim(\mathbb{S}) = \dim(\mathbb{T}) = \frac{1}{2}n(n+1)$, $\dim(\mathbb{E}) = \frac{1}{2}n(n-1)$, $\dim(\mathbb{D}) = n$).

OSSERVAZIONE Come nella dimostrazione della formula di Grassmann, siano \mathbf{H} e \mathbf{K} sottospazi di \mathbb{K}^n , sia \mathbf{P} (rispettivamente \mathbf{Q}) una matrice che ha per colonne i vettori una base di \mathbf{H} (rispettivamente \mathbf{K}). Allora $\mathbf{H} + \mathbf{K}$ è lo spazio colonna della matrice $[\mathbf{P} \ \mathbf{Q}]$. Si osservi:

- a) la formula (10.5) $\dim(\mathbf{H} + \mathbf{K}) = r([\mathbf{P} \ \mathbf{Q}])$ consente di calcolare la dimensione di $\mathbf{H} + \mathbf{K}$ e la dimensione dell'intersezione si ricava, quindi, mediante la formula di Grassmann;

- b) la dimostrazione del teorema 10.3 fornisce un metodo per trovare una base dell'intersezione $\mathbf{H} \cap \mathbf{K}$: con l'algoritmo di Gauss si ricava una base $\left[\begin{smallmatrix} \mathbf{t}_1 \\ -\mathbf{u}_1 \end{smallmatrix} \right], \dots, \left[\begin{smallmatrix} \mathbf{t}_s \\ -\mathbf{u}_s \end{smallmatrix} \right]$ di $\text{Ker}([\mathbf{P} \ \mathbf{Q}])$; una base di $\mathbf{H} \cap \mathbf{K}$ è allora formata da $\mathbf{Pt}_1, \dots, \mathbf{Pt}_s$ (o anche $\mathbf{Qu}_1, \dots, \mathbf{Qu}_s$ visto che $\mathbf{Pt}_i = \mathbf{Qu}_i$).

In \mathbb{R}^4 consideriamo i vettori

$$\mathbf{h}_1 = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 4 \end{bmatrix}, \quad \mathbf{h}_2 = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 2 \end{bmatrix}, \quad \mathbf{k}_1 = \begin{bmatrix} 3 \\ 4 \\ 3 \\ 5 \end{bmatrix}, \quad \mathbf{k}_2 = \begin{bmatrix} 0 \\ 5 \\ 5 \\ 5 \end{bmatrix}$$

e i sottospazi $\mathbf{H} = \mathcal{L}(\mathbf{h}_1, \mathbf{h}_2)$ e $\mathbf{K} = \mathcal{L}(\mathbf{k}_1, \mathbf{k}_2)$. Come prima poniamo

$$\mathbf{P} = [\mathbf{h}_1 \ \mathbf{h}_2] = \begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 4 & 1 \\ 4 & 2 \end{bmatrix} \quad \text{e} \quad \mathbf{Q} = [\mathbf{k}_1 \ \mathbf{k}_2] = \begin{bmatrix} 3 & 0 \\ 4 & 5 \\ 3 & 5 \\ 5 & 5 \end{bmatrix}$$

in modo che $\mathbf{H} = \text{Col}(\mathbf{P})$ e $\mathbf{K} = \text{Col}(\mathbf{Q})$. Il sottospazio somma $\mathbf{H} + \mathbf{K}$ è allora lo spazio colonna della matrice

$$[\mathbf{P} \ \mathbf{Q}] = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 3 & 1 & 4 & 5 \\ 4 & 1 & 3 & 5 \\ 4 & 2 & 5 & 5 \end{bmatrix}$$

Riduciamo la matrice $[\mathbf{P} \ \mathbf{Q}]$ a scala:

$$\begin{bmatrix} 1 & 2 & 3 & 0 \\ 3 & 1 & 4 & 5 \\ 4 & 1 & 3 & 5 \\ 4 & 2 & 5 & 5 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 3 & 3 \\ 0 & -5 & -5 & 5 \\ 0 & -7 & -9 & 5 \\ 0 & -6 & -7 & 5 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 3 & 3 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & -2 & -2 \\ 0 & 0 & -1 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 3 & 3 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \mathbf{U}$$

Quindi $[\mathbf{P} \ \mathbf{Q}]$ ha rango 3 e questa è la dimensione del sottospazio somma. Per la formula di Grassmann, l'intersezione $\mathbf{H} \cap \mathbf{K}$ ha dimensione $4 - 3 = 1$.

Per trovare una base dell'intersezione, usiamo il procedimento descritto nell'osservazione precedente. Determiniamo dapprima una base di $\text{Ker}([\mathbf{P} \ \mathbf{Q}]) = \text{Ker}(\mathbf{U})$, che è costituita da un solo vettore \mathbf{w} , risolvendo il sistema $\mathbf{Ux} = \mathbf{0}$. La variabile libera è x_4 e ponendo $x_4 = 1$ troviamo $\mathbf{w} = [1, -2, 1, -1]^T$. Nelle notazioni dell'osservazione precedente, abbiamo $s = 1$ e $\mathbf{t}_1 = [-1, -2]^T$, $\mathbf{u}_1 = [-1, 1]^T$. Una base per $\mathbf{H} \cap \mathbf{K}$ è allora fornita dal vettore

$$\mathbf{Pt}_1 = 1\mathbf{h}_1 + (-2)\mathbf{h}_2 = [-3, 1, 2, 0]^T$$

o equivalentemente

$$\mathbf{Qu}_1 = -\mathbf{k}_1 + 1\mathbf{k}_2 = [-3, 1, 2, 0]^T$$

DEFINIZIONE 10.4 (Somma diretta)

Siano \mathbf{H} e \mathbf{K} sottospazi di uno spazio vettoriale \mathbf{V} . Si dice che \mathbf{V} è *somma diretta* di \mathbf{H} e \mathbf{K} e si scrive

$$\mathbf{V} = \mathbf{H} \oplus \mathbf{K}$$

se ogni vettore di $\mathbf{v} \in \mathbf{V}$ si scrive in uno e un solo modo come somma $\mathbf{v} = \mathbf{h} + \mathbf{k}$ di un vettore di $\mathbf{h} \in \mathbf{H}$ e di un vettore di $\mathbf{k} \in \mathbf{K}$.

... nella proiezione su \mathbf{K}

Il piano cartesiano \mathbb{R}^2 è somma diretta dell'asse x e dell'asse y . Lo spazio cartesiano \mathbb{R}^3 è somma diretta del piano xy e dell'asse z ; non è invece somma diretta del piano xy e del piano yz .

OSSERVAZIONE Se $\mathbf{V} = \mathbf{H} \oplus \mathbf{K}$, dal punto di vista insiemistico \mathbf{V} si identifica con il *prodotto cartesiano* $\mathbf{H} \times \mathbf{K}$, che è l'insieme delle coppie ordinate (\mathbf{h}, \mathbf{k}) di un vettore $\mathbf{h} \in \mathbf{H}$ e di un vettore $\mathbf{k} \in \mathbf{K}$. Infatti ogni vettore $\mathbf{v} \in \mathbf{V}$ si scrive in uno e un sol modo nella forma $\mathbf{v} = \mathbf{h} + \mathbf{k}$ con $\mathbf{h} \in \mathbf{H}$ e $\mathbf{k} \in \mathbf{K}$, si può così identificare con la coppia ordinata (\mathbf{h}, \mathbf{k}) . Si noti che un vettore $\mathbf{h} \in \mathbf{H} \subseteq \mathbf{V}$ corrisponde alla coppia $(\mathbf{h}, \mathbf{0})$, mentre un vettore di $\mathbf{k} \in \mathbf{K}$ corrisponde alla coppia $(\mathbf{0}, \mathbf{k})$. In particolare il vettore nullo corrisponde alla coppia $(\mathbf{0}, \mathbf{0})$.

Diamo ora due criteri per verificare se \mathbf{V} è somma diretta di due suoi sottospazi:

PROPOSIZIONE 10.5 Uno spazio vettoriale \mathbf{V} è somma diretta dei suoi sottospazi \mathbf{H} e \mathbf{K} se e solo se:

- a) $\mathbf{H} + \mathbf{K} = \mathbf{V}$;
- b) $\mathbf{H} \cap \mathbf{K} = \{\mathbf{0}\}$.

In particolare, se \mathbf{V} ha dimensione finita e $\mathbf{V} = \mathbf{H} \oplus \mathbf{K}$, allora

$$\dim(\mathbf{V}) = \dim(\mathbf{H}) + \dim(\mathbf{K})$$

DIMOSTRAZIONE. Supponiamo dapprima $\mathbf{V} = \mathbf{H} \oplus \mathbf{K}$ e mostriamo che valgono a) e b). L'ipotesi è che ogni vettore di \mathbf{V} si scriva in uno e un solo modo come somma di un vettore di \mathbf{H} e di un vettore di \mathbf{K} . In particolare, $\mathbf{V} = \mathbf{H} + \mathbf{K}$. Se $\mathbf{v} \in \mathbf{H} \cap \mathbf{K}$, poniamo $\mathbf{h}_1 = \mathbf{v}$, $\mathbf{h}_2 = \mathbf{0}$, $\mathbf{k}_1 = \mathbf{0}$, $\mathbf{k}_2 = \mathbf{v}$. Allora

$$\mathbf{v} = \mathbf{h}_1 + \mathbf{k}_1 = \mathbf{h}_2 + \mathbf{k}_2$$

sono due modi di scrivere \mathbf{v} come somma di un vettore di \mathbf{H} e di un vettore di \mathbf{K} . Siccome tale scrittura per ipotesi è unica, dobbiamo avere $\mathbf{h}_1 = \mathbf{h}_2$ e $\mathbf{k}_1 = \mathbf{k}_2$, cioè $\mathbf{v} = \mathbf{0}$. Questo mostra $\mathbf{H} \cap \mathbf{K} = \{\mathbf{0}\}$.

Viceversa, supponiamo che valgano a) e b). Siccome $\mathbf{V} = \mathbf{H} + \mathbf{K}$, possiamo scrivere ogni vettore $\mathbf{v} \in \mathbf{V}$ come somma di un vettore di \mathbf{H} e di un vettore di \mathbf{K} . Se

$$\mathbf{v} = \mathbf{h}_1 + \mathbf{k}_1 = \mathbf{h}_2 + \mathbf{k}_2$$

con $\mathbf{h}_1, \mathbf{h}_2 \in \mathbf{H}$ e $\mathbf{k}_1, \mathbf{k}_2 \in \mathbf{K}$, consideriamo il vettore

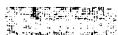
$$\mathbf{w} = \mathbf{h}_1 - \mathbf{h}_2 = \mathbf{k}_2 - \mathbf{k}_1$$

La prima uguaglianza mostra $\mathbf{w} \in \mathbf{H}$, la seconda mostra $\mathbf{w} \in \mathbf{K}$. Quindi $\mathbf{w} \in \mathbf{H} \cap \mathbf{K}$, e da b) segue $\mathbf{w} = \mathbf{0}$. Ma $\mathbf{w} = \mathbf{0}$ significa $\mathbf{h}_1 = \mathbf{h}_2$ e $\mathbf{k}_1 = \mathbf{k}_2$ e abbiamo così mostrato che \mathbf{v} si scrive in un unico modo come somma di un vettore di \mathbf{H} e di un vettore di \mathbf{K} .

Infine, la formula per la dimensione segue dalla formula di Grassmann, tenuto conto di a) e di b).



Per ogni coefficiente angolare $m \in \mathbb{R}$, il piano cartesiano \mathbb{R}^2 è somma diretta della retta $y = mx$ e dell'asse y . Infatti l'intersezione consiste della sola origine e per la formula di Grassmann il sottospazio somma dev'essere tutto \mathbb{R}^2 .



Consideriamo nuovamente lo spazio vettoriale \mathbb{M} delle matrici quadrate di ordine n e i suoi sottospazi \mathbb{S} (matrici simmetriche), \mathbb{E} (matrici antisimmetriche) e \mathbb{T} (matrici triangolari alte). Abbiamo già mostrato che $\mathbb{S} + \mathbb{E} = \mathbb{M}$ e $\mathbb{S} \cap \mathbb{E} = \{\mathbf{0}\}$. Quindi $\mathbb{M} = \mathbb{S} \oplus \mathbb{E}$: ogni matrice quadrata si scrive in uno e un sol modo come somma di una matrice simmetrica e di una antisimmetrica. Analogamente, $\mathbb{M} = \mathbb{T} \oplus \mathbb{E}$. Invece la somma $\mathbb{M} = \mathbb{S} + \mathbb{T}$ non è diretta perché $\mathbb{S} \cap \mathbb{T}$ consiste delle matrici diagonali.

PROPOSIZIONE 10.6 Sia \mathbf{V} uno spazio vettoriale di dimensione finita, siano \mathbf{H} e \mathbf{K} sottospazi di \mathbf{V} , e siano $\{\mathbf{h}_1, \dots, \mathbf{h}_d\}$ e $\{\mathbf{k}_1, \dots, \mathbf{k}_e\}$ basi di \mathbf{H} e \mathbf{K} , rispettivamente. Allora $\mathbf{V} = \mathbf{H} \oplus \mathbf{K}$ se e solo se $\{\mathbf{h}_1, \dots, \mathbf{h}_d, \mathbf{k}_1, \dots, \mathbf{k}_e\}$ è una base di \mathbf{V} .

DIMOSTRAZIONE. Supponiamo che $\mathbf{V} = \mathbf{H} \oplus \mathbf{K}$ e mostriamo che $\{\mathbf{h}_1, \dots, \mathbf{h}_d, \mathbf{k}_1, \dots, \mathbf{k}_e\}$ è una base di \mathbf{V} . Per la proposizione 10.2

$$\{\mathbf{h}_1, \dots, \mathbf{h}_d, \mathbf{k}_1, \dots, \mathbf{k}_e\}$$

è un insieme di generatori di $\mathbf{H} + \mathbf{K} = \mathbf{V}$. Inoltre

$$\dim(\mathbf{V}) = \dim(\mathbf{H}) + \dim(\mathbf{K}) = d + e$$

e quindi $\{\mathbf{h}_1, \dots, \mathbf{h}_d, \mathbf{k}_1, \dots, \mathbf{k}_e\}$ è una base di \mathbf{V} per il corollario 6.11.

Viceversa, supponiamo che $\{\mathbf{h}_1, \dots, \mathbf{h}_d, \mathbf{k}_1, \dots, \mathbf{k}_e\}$ sia una base di \mathbf{V} . Facciamo vedere che $\mathbf{V} = \mathbf{H} \oplus \mathbf{K}$ mostrando che $\mathbf{V} = \mathbf{H} + \mathbf{K}$ e che $\mathbf{H} \cap \mathbf{K} = \{\mathbf{0}\}$. Ogni vettore di $\mathbf{v} \in \mathbf{V}$ si scrive come combinazione lineare dei vettori della base:

$$\mathbf{v} = (t_1 \mathbf{h}_1 + \cdots + t_d \mathbf{h}_d) + (u_1 \mathbf{k}_1 + \cdots + u_e \mathbf{k}_e)$$

Questo mostra che \mathbf{v} è somma di un vettore di \mathbf{H} e di un vettore di \mathbf{K} , quindi $\mathbf{V} = \mathbf{H} + \mathbf{K}$. Inoltre,

$$\dim(\mathbf{H} + \mathbf{K}) = \dim(\mathbf{H}) + \dim(\mathbf{K})$$

perché $\mathbf{H} + \mathbf{K}$ ha una base formata da $d + e$ vettori. Per la formula di Grassmann $\dim(\mathbf{H} \cap \mathbf{K}) = 0$, quindi $\mathbf{H} \cap \mathbf{K} = \{\mathbf{0}\}$ e questo conclude la dimostrazione.



Siano \mathbf{H} il sottospazio di \mathbb{K}^n generato da d vettori della base canonica e sia \mathbf{K} il sottospazio generato dai rimanenti $n - d$ vettori della base canonica. Per la proposizione precedente $\mathbb{K}^n = \mathbf{H} \oplus \mathbf{K}$.



Sia \mathbf{V} uno spazio vettoriale di dimensione finita e sia \mathbf{H} un suo sottospazio. Allora esiste un sottospazio \mathbf{K} di \mathbf{V} tale che $\mathbf{V} = \mathbf{H} \oplus \mathbf{K}$: si dice che \mathbf{K} è un *sottospazio complementare* di \mathbf{H} in \mathbf{V} .

Per dimostrare questa affermazione, fissiamo una base \mathcal{B} di \mathbf{H} . I vettori di \mathcal{B} sono linearmente indipendenti e sono quindi contenuti in una base dell'intero spazio \mathbf{V} . Sia \mathcal{C} l'insieme dei vettori di tale base che non appartengono a \mathcal{B} e sia \mathbf{K} il sottospazio generato dai vettori in \mathcal{C} . Per la proposizione 6 $\mathbf{V} = \mathbf{H} \oplus \mathbf{K}$.

Per esempio, se $\mathbf{V} = \mathbb{K}^2$ e \mathbf{H} è l'asse y , la retta \mathbf{K} di equazione $y = mx$ è un complementare di \mathbf{H} . In questo caso $\mathcal{B} = \{\begin{pmatrix} 0 \\ 1 \end{pmatrix}\}$ e $\mathcal{C} = \{\begin{pmatrix} 1 \\ m \end{pmatrix}\}$.



Siano \mathbf{U} e \mathbf{V} i piani di \mathbb{R}^3 di equazione $x + 2y + 3z = 0$ e $3x + 2y + z = 0$ rispettivamente. Determinare una base di $\mathbf{U} + \mathbf{V}$ e $\mathbf{U} \cap \mathbf{V}$.

In \mathbb{R}^4 si considerino il sottospazio \mathbf{V} generato da $\mathbf{v}_1 = [1, 2, 3, 4]^T$ e $\mathbf{v}_2 = [3, 1, 3, 4]^T$ e il sottospazio \mathbf{W} generato da $\mathbf{w}_1 = [3, 4, 3, 4]^T$ e $\mathbf{w}_2 = [1, 3, 6, 9]^T$.

- Trovare una base di $\mathbf{V} + \mathbf{W}$.
- Trovare la dimensione di $\mathbf{V} \cap \mathbf{W}$.
- Trovare una base di $\mathbf{V} \cap \mathbf{W}$.

Sia \mathcal{F} lo spazio delle funzioni reali di variabile reale. Si considerino i sottospazi delle funzioni pari $\mathbf{U} = \{f \in \mathcal{F} : f(x) = f(-x) \forall x \in \mathbb{R}\}$ e delle funzioni dispari $\mathbf{V} = \{f \in \mathcal{F} : f(-x) = -f(x) \forall x \in \mathbb{R}\}$. Si mostri che $\mathcal{F} = \mathbf{U} \oplus \mathbf{V}$.

Mostrare che l'unione $\mathbf{H} \cup \mathbf{K}$ di due sottospazi di \mathbf{V} è un sottospazio se e solo se $\mathbf{H} \subseteq \mathbf{K}$ oppure $\mathbf{K} \subseteq \mathbf{H}$.

Suggerimento: se \mathbf{h} appartiene a \mathbf{H} ma non a \mathbf{K} e \mathbf{k} appartiene a \mathbf{K} ma non a \mathbf{H} , allora $\mathbf{h} + \mathbf{k}$ non appartiene né a \mathbf{H} né a \mathbf{K} .

Dimostrare la formula di Grassmann nel modo seguente: fissare una base $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ di $\mathbf{H} \cap \mathbf{K}$, estenderla a una base $\{\mathbf{u}_1, \dots, \mathbf{u}_r, \mathbf{v}_1, \dots, \mathbf{v}_s\}$ di \mathbf{H} e a una base $\{\mathbf{u}_1, \dots, \mathbf{u}_r, \mathbf{w}_1, \dots, \mathbf{w}_t\}$ di \mathbf{K} . Mostrare che l'insieme

$$\{\mathbf{u}_1, \dots, \mathbf{u}_r, \mathbf{v}_1, \dots, \mathbf{v}_s, \mathbf{w}_1, \dots, \mathbf{w}_t\}$$

è una base di $\mathbf{H} + \mathbf{K}$. Quindi

$$\dim(\mathbf{H} + \mathbf{K}) = r + s + t = (r + s) + (r + t) - r = \dim \mathbf{H} + \dim \mathbf{K} - \dim(\mathbf{H} \cap \mathbf{K})$$

Il difetto di questa dimostrazione è che, a differenza di quella presentata nel testo, non è costruttiva (non spiega come trovare una base di $\mathbf{H} \cap \mathbf{K}$).

Siano A e B due sottoinsiemi di un insieme finito C . Mostrare che

$$\#(A \cup B) + \#(A \cap B) = \#(A) + \#(B)$$

dove il simbolo $\#(S)$ denota il numero di elementi di S . Confrontare questa formula con la formula di Grassmann (si veda anche l'esercizio precedente).

Estendere il concetto di somma diretta al caso di più addendi: si dice che \mathbf{V} è somma diretta dei suoi sottospazi $\mathbf{H}_1, \dots, \mathbf{H}_r$ se ogni vettore \mathbf{v} di \mathbf{V} si scrive in uno e un solo modo nella forma

$$\mathbf{v} = \mathbf{h}_1 + \cdots + \mathbf{h}_r$$

con $\mathbf{h}_i \in \mathbf{H}_i$ per ogni $i = 1, 2, \dots, r$. In tal caso si scrive $\mathbf{V} = \mathbf{H}_1 \oplus \mathbf{H}_2 \oplus \cdots \oplus \mathbf{H}_r$. Mostrare che

- a) se \mathcal{B}_i è una base di \mathbf{H}_i per ogni i e

$$\mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_r$$

è una base di \mathbf{V} , allora $\mathbf{V} = \mathbf{H}_1 \oplus \mathbf{H}_2 \oplus \cdots \oplus \mathbf{H}_r$.

- b) Se $\mathbf{V} = \mathbf{H}_1 + \mathbf{H}_2 + \cdots + \mathbf{H}_r$ e, per ogni i , l'intersezione di \mathbf{H}_i con la somma dei rimanenti \mathbf{H}_j è ridotta al solo vettore nullo, allora $\mathbf{V} = \mathbf{H}_1 \oplus \mathbf{H}_2 \oplus \cdots \oplus \mathbf{H}_r$.
- c) L'insieme $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ è una base di \mathbf{V} se e solo se

$$\mathbf{V} = \mathcal{L}(\mathbf{v}_1) \oplus \mathcal{L}(\mathbf{v}_2) \oplus \cdots \oplus \mathcal{L}(\mathbf{v}_n)$$



Applicazioni lineari

1 INTRODUZIONE

L'algebra lineare ha come oggetto di studio gli *spazi vettoriali* e le *funzioni lineari* tra di essi: una funzione $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{W}$ tra spazi vettoriali si dice *lineare* se rispetta le operazioni di somma e prodotto per scalare; questo significa che $\mathfrak{L}(\mathbf{v}_1 + \mathbf{v}_2) = \mathfrak{L}(\mathbf{v}_1) + \mathfrak{L}(\mathbf{v}_2)$ per ogni coppia di vettori del dominio \mathbf{V} e $\mathfrak{L}(t\mathbf{v}) = t\mathfrak{L}(\mathbf{v})$ per ogni scalare t e per ogni vettore $\mathbf{v} \in \mathbf{V}$. Sinonimi di funzione lineare sono applicazione, trasformazione e operatore lineare. Dato il ruolo centrale di questa nozione, non è sorprendente che si sia già presentata nei capitoli precedenti: nel primo capitolo, per esempio, abbiamo sfruttato la linearità del prodotto scalare e del prodotto vettoriale per ricavarne l'espressione in coordinate; la dimostrazione del lemma fondamentale del capitolo sugli spazi vettoriali si basa sulla linearità della mappa che, attraverso le coordinate, identifica uno spazio vettoriale con lo spazio \mathbb{K}^n dei vettori colonna.

Le funzioni lineari sono le funzioni più semplici dopo quelle costanti (il sogno della matricola, si dice, è che tutte le funzioni siano lineari, per esempio che il logaritmo di una somma sia la somma dei logaritmi). Questo traspare dall'espressione in coordinate delle funzioni lineari: nel caso monodimensionale una funzione $y = f(x)$ è lineare se è della forma $y = ax$ con a costante, cioè se y è direttamente proporzionale a x ; nel caso multidimensionale, una funzione $\mathbb{K}^n \rightarrow \mathbb{K}^m$ è lineare se è della forma $\mathbf{y} = \mathbf{Ax}$ dove \mathbf{A} è una matrice $m \times n$. Ma in realtà la nozione di linearità *dipende soltanto* dalla struttura di spazio vettoriale e non dalla particolare scelta di coordinate, perché una funzione è lineare se e solo se rispetta la somma e il prodotto per uno scalare. Il fatto che la definizione non dipenda dalle coordinate comporta gli usuali vantaggi: (1) poter scegliere le coordinate a posteriori, in modo da semplificare i conti e (2) poter applicare i risultati a contesti più generali. In questo modo, molto di quanto abbiamo visto per un sistema lineare $\mathbf{Ax} = \mathbf{b}$ si generalizza a qualsiasi equazione della forma $\mathfrak{L}(\mathbf{v}) = \mathbf{w}$ a patto che la funzione \mathfrak{L} a primo membro sia una funzione lineare dell'incognita \mathbf{v} ; per esempio, il fatto che le soluzioni si ottengano come somma di una soluzione particolare con le soluzioni dell'equazione omogenea $\mathfrak{L}(\mathbf{v}) = \mathbf{0}$ rimane vero per ogni equazione lineare. L'esempio più importante di equazione lineare è quello delle

equazioni differenziali lineari, in cui l'incognita \mathbf{v} è una funzione $y(t)$ e la funzione \mathfrak{L} è

$$\mathfrak{L}(y(t)) = a_n(t)y^{(n)}(t) + a_{n-1}(t)y^{(n-1)}(t) + \cdots + a_1(t)y'(t) + a_0(t)y(t)$$

Un esempio esplicito è l'equazione delle vibrazioni meccaniche:

$$my''(t) + hy'(t) + ky(t) = f(t)$$

che governa i fenomeni oscillatori:

L'importanza della nozione di applicazione lineare va al di là del fatto che fenomeni fondamentali siano governati da leggi lineari. L'analisi matematica consente infatti di approssimare equazioni non lineari con equazioni lineari: l'idea di base, efficace, come lo studente avrà già avuto modo di sperimentare per le funzioni di una variabile, è di ricondurre in prima approssimazione lo studio di una funzione qualsiasi allo studio del suo differenziale, che in ciascun punto è una funzione lineare e, quindi, semplice, dell'incremento. Questo metodo va sotto il nome di *linearizzazione*. Un ottimo esempio è quello delle piccole oscillazioni del pendolo: le oscillazioni di un pendolo ideale obbediscono a un'equazione del tipo $y'' + \omega^2 \sin(y) = 0$, dove y denota l'angolo di oscillazione del pendolo rispetto alla posizione di equilibrio. Si tratta di un'equazione non lineare in y , ma per piccoli valori di y si può approssimare $\sin(y)$ con y : la funzione non lineare $\Delta f(y) = \sin(y) - \sin(0) = \sin(y)$ viene sostituita dal suo differenziale nell'origine $f'(0)\Delta y = y$. Più in generale, il differenziale $d\mathbf{F}(\mathbf{x}_0)$ in \mathbf{x}_0 di una funzione $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ è (quando esiste) l'*approssimazione lineare* di $\Delta\mathbf{F}$ vicino a \mathbf{x}_0 . Questo significa che

$$\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{x}_0) = d\mathbf{F}(\mathbf{x}_0)(\mathbf{h}) + o(\mathbf{h}) \quad \text{per } \mathbf{h} = \mathbf{x} - \mathbf{x}_0 \rightarrow \mathbf{0}$$

dove l'errore di approssimazione $o(\mathbf{h})$ è un infinitesimo di ordine superiore ad \mathbf{h} per $\mathbf{h} \rightarrow \mathbf{0}$ e il differenziale $d\mathbf{F}(\mathbf{x}_0)$ è una funzione lineare dell'incremento \mathbf{h} . In coordinate il differenziale è la funzione lineare che corrisponde alla moltiplicazione per la matrice jacobiana di \mathbf{F} :

$$d\mathbf{F}(\mathbf{x}_0)(\mathbf{h}) = \left[\frac{\partial \mathbf{F}_i}{\partial x_j}(\mathbf{x}_0) \right] \mathbf{h}$$

Dal punto di vista geometrico, si sta approssimando il grafico $\mathbf{y} = \mathbf{F}(\mathbf{x})$ di \mathbf{F} vicino a \mathbf{x}_0 con il sottospazio tangente $\mathbf{y} = \mathbf{y}_0 + d\mathbf{F}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$. Questo metodo consente di approssimare un problema non lineare con uno lineare ed è di fondamentale importanza nelle applicazioni perché di norma il problema non lineare è difficile da risolvere, mentre quello lineare è spesso trattabile.

Un teorema dimostrato per una funzione lineare ha invariabilmente applicazioni allo studio di funzioni differenziabili qualsiasi. Facciamo due esempi specifici. In questo capitolo mostriamo che, se un'applicazione lineare $\mathfrak{L} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ha rango massimo n , allora è invertibile; l'analogo per una funzione differenziabile $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ è il *teorema della funzione inversa*: se il differenziale $d\mathbf{F}(\mathbf{x}_0)$ ha rango n , allora \mathbf{F} è localmente invertibile vicino a \mathbf{x}_0 e l'inversa è differenziabile. Un altro importante teorema di questo capitolo, il teorema di nullità più rango, afferma che se $\mathfrak{L} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ha rango r , allora l'insieme degli \mathbf{x} tali che $\mathfrak{L}(\mathbf{x}) = \mathbf{0}$ è un sottospazio di \mathbb{R}^n di dimensione $n - r$. L'analogo risultato per una funzione $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ con derivate parziali continue è il

teorema della funzione implicita: se il differenziale $d\mathbf{F}(\mathbf{x}_0)$ ha rango r , allora l'insieme degli \mathbf{x} tali che $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ è, vicino a \mathbf{x}_0 , una sottovarietà di \mathbb{R}^n di dimensione $n - r$.

Riassumendo, lo studio delle applicazioni e delle equazioni lineari è importante di per sé, ma è anche prerequisito indispensabile allo studio di fenomeni e di equazioni più complessi.

2 APPLICAZIONI LINEARI

Il modo più semplice in cui una variabile y può dipendere da un'altra variabile x è di essere *direttamente proporzionale* a x . Questo significa che

$$y = ax$$

dove a è una costante fissata. Si dice anche che y dipende *linearmente* da x perché la curva di equazione $y = ax$ è una retta. Generalizzando al caso di più variabili indipendenti, si dice che y dipende linearmente da x_1, \dots, x_n se esistono degli scalari a_1, a_2, \dots, a_n tali che

$$y = a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

Dal punto di vista algebrico, y è un polinomio *omogeneo di primo grado* in x_1, x_2, \dots, x_n (un polinomio è omogeneo se è somma di monomi che hanno tutti lo stesso grado; omogeneo di primo grado vuol dire somma di monomi di primo grado *senza termine costante*). Nel caso ci siano più variabili dipendenti, diremo che il vettore $\mathbf{y} = [y_1, \dots, y_m]^T$ dipende linearmente da $\mathbf{x} = [x_1, \dots, x_n]^T$ se ciascuna componente y_i di \mathbf{y} è una funzione lineare di \mathbf{x} : questo significa che esistono degli scalari a_{ij} tali che

$$(2.1) \quad \begin{cases} y_1 = a_{11}x_1 + \cdots + a_{1n}x_n \\ y_2 = a_{21}x_1 + \cdots + a_{2n}x_n \\ \vdots \quad \vdots \quad \vdots \\ y_m = a_{m1}x_1 + \cdots + a_{mn}x_n \end{cases}$$

Sia \mathbf{A} la matrice che ha come elemento di posto (i, j) lo scalare a_{ij} . Possiamo riscrivere la (2.1) nella forma compatta

$$(2.2) \quad \mathbf{y} = \mathbf{Ax}$$

Questa scrittura ricorda, suggestivamente, il caso monodimensionale $y = ax$: il vettore \mathbf{y} si ottiene da \mathbf{x} mediante moltiplicazione a sinistra per la matrice \mathbf{A} . Si osservi che $\mathbf{y} \in \mathbb{K}^m$ perché il prodotto di una matrice di tipo (m, n) per un vettore colonna con n componenti è un vettore colonna con m componenti. Denotiamo col simbolo $\mathcal{L}_{\mathbf{A}}$ la funzione definita da (2.2):

DEFINIZIONE 2.1 (Applicazione rappresentata da una matrice)

Data una matrice \mathbf{A} di tipo (m, n) , la funzione $\mathcal{L}_{\mathbf{A}} : \mathbb{K}^n \rightarrow \mathbb{K}^m$ definita dalla formula:

$$\mathcal{L}_{\mathbf{A}}(\mathbf{x}) = \mathbf{Ax} \quad \text{per ogni } \mathbf{x} \in \mathbb{K}^n$$

si dice applicazione rappresentata dalla matrice \mathbf{A}

Abbiamo ora a disposizione un nuovo punto di vista sulle matrici: finora abbiamo pensato a una matrice come a un oggetto per così dire *statico*: una tabella di numeri. D'ora innanzi potremo pensare a una matrice anche come a un oggetto *dinamico*: una funzione che prende come input un vettore e produce come output un altro vettore. Questo nuovo punto di vista rende lo studio delle matrici molto più interessante e ne moltiplica le applicazioni.



Consideriamo la funzione $\mathcal{L} : \mathbb{K}^2 \rightarrow \mathbb{K}^3$ che associa al vettore $[x_1, x_2]^T$ di \mathbb{K}^2 il vettore $[y_1, y_2, y_3]^T$ di \mathbb{K}^3 definito da

$$\begin{cases} y_1 = 2x_1 + 3x_2 \\ y_2 = x_1 + 4x_2 \\ y_3 = -x_1 + 5x_2 \end{cases}$$

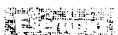
Procediamo come sopra: la matrice \mathbf{A} di tipo $(3, 2)$ ha come elementi i coefficienti con cui compaiono a secondo membro le variabili indipendenti:

$$\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 1 & 4 \\ -1 & 5 \end{bmatrix}$$

Se moltiplichiamo \mathbf{A} per il vettore $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ otteniamo

$$\begin{bmatrix} 2 & 3 \\ 1 & 4 \\ -1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 + 3x_2 \\ x_1 + 4x_2 \\ -x_1 + 5x_2 \end{bmatrix}$$

per cui $\mathcal{L} = \mathcal{L}_{\mathbf{A}}$.



Sia $\mathcal{S} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ la riflessione ortogonale rispetto alla bisettrice del primo e terzo quadrante. In coordinate

$$\mathcal{S}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}$$

Si tratta di una funzione lineare perché le sue due componenti $y_1 = 0x_1 + 1x_2$, e $y_2 = 1x_1 + 0x_2$ sono polinomi omogenei di primo grado in x_1 e x_2 . Se \mathbf{A} è la matrice che rappresenta \mathcal{S} , la prima riga di \mathbf{A} è $[0 \ 1]$ perché 0 e 1 sono i coefficienti delle variabili x_1 e x_2 nell'espressione di y_1 . Analogamente, la seconda riga di \mathbf{A} è $[1 \ 0]$. Ovvero:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Abbiamo dato una definizione di linearità in coordinate: note le coordinate di \mathbf{x} , una funzione è del tipo $\mathcal{L}_{\mathbf{A}}$ se le coordinate del vettore trasformato sono polinomi omogenei di primo grado nelle coordinate di \mathbf{x} . È invece fondamentale ottenere una definizione di

linearità che *non dipenda* dalle coordinate. Questo per lo stesso motivo per cui è utile avere una nozione astratta di spazio vettoriale oltre alla versione concreta \mathbb{K}^n : solo una volta che ci si è svincolati dalle coordinate è possibile chiedersi se le problematiche che si pongono sono effettivamente indipendenti dalla scelta delle coordinate e in caso di risposta positiva, quale sia il miglior sistema di coordinate in cui fare i conti. Occorre quindi trovare una caratterizzazione delle funzioni \mathcal{L}_A che non dipenda dalle coordinate, ma soltanto dalle operazioni di somma di vettori e di prodotto per uno scalare. Il primo semplice passo in questa direzione è la seguente proposizione (il passo successivo è il teorema di rappresentazione 5.2).

PROPOSIZIONE 2.2 Sia A una matrice di tipo (m, n) . La funzione $\mathcal{L}_A : \mathbb{K}^n \rightarrow \mathbb{K}^m$ gode delle seguenti proprietà:

a) \mathcal{L}_A è *additiva*:

$$\mathcal{L}_A(\mathbf{x}_1 + \mathbf{x}_2) = \mathcal{L}_A(\mathbf{x}_1) + \mathcal{L}_A(\mathbf{x}_2) \quad \text{per ogni } \mathbf{x}_1 \text{ e } \mathbf{x}_2 \text{ in } \mathbb{K}^n$$

b) \mathcal{L}_A è *omogenea*:

$$\mathcal{L}_A(t\mathbf{x}) = t\mathcal{L}_A(\mathbf{x}) \quad \text{per ogni } t \in \mathbb{K} \text{ e ogni } \mathbf{x} \in \mathbb{K}^n$$

DIMOSTRAZIONE. Il punto a) è equivalente a

$$A(\mathbf{x}_1 + \mathbf{x}_2) = A\mathbf{x}_1 + A\mathbf{x}_2$$

e questa identità non è altro che la proprietà distributiva del prodotto di matrici. Analogamente, l'omogeneità è equivalente all'identità

$$A(t\mathbf{x}) = tA(\mathbf{x})$$

che segue anch'essa dalle proprietà del prodotto di matrici.

Motivati dalla proposizione precedente, possiamo dare una definizione di *funzione lineare* indipendente dalla scelta di una base e delle relative coordinate, nel contesto astratto degli spazi vettoriali:

DEFINIZIONE 2.3 (Applicazione o funzione lineare)

Siano V e W due spazi vettoriali sullo stesso campo \mathbb{K} . Una funzione $\mathcal{L} : V \rightarrow W$ si dice *lineare* se

a) \mathcal{L} è *additiva*:

$$\mathcal{L}(\mathbf{v}_1 + \mathbf{v}_2) = \mathcal{L}(\mathbf{v}_1) + \mathcal{L}(\mathbf{v}_2) \quad \text{per ogni } \mathbf{v}_1 \text{ e } \mathbf{v}_2 \text{ in } V$$

b) \mathcal{L} è *omogenea*:

$$\mathcal{L}(t\mathbf{v}) = t\mathcal{L}(\mathbf{v}) \quad \text{per ogni } t \in \mathbb{K} \text{ e ogni } \mathbf{v} \in V$$

Una funzione lineare si dice anche *applicazione lineare* od *operatore lineare* o *trasformazione lineare*.



Se \mathbf{A} è una matrice $m \times n$, l'applicazione $\mathcal{L}_{\mathbf{A}} : \mathbb{K}^n \rightarrow \mathbb{K}^m$ è lineare per la proposizione 2.2. Vedremo più avanti (teorema di rappresentazione 5.2) che *ogni* funzione lineare da $\mathbb{K}^n \rightarrow \mathbb{K}^m$ è della forma $\mathcal{L}_{\mathbf{A}}$, non ce ne sono altre.



Per ogni spazio vettoriale \mathbf{V} possiamo considerare la funzione identità: $\mathcal{I} : \mathbf{V} \rightarrow \mathbf{V}$ definita da

$$\mathcal{I}(\mathbf{v}) = \mathbf{v}$$

Evidentemente è una funzione lineare. Se $\mathbf{V} = \mathbb{K}^n$, si tratta dell'applicazione rappresentata dalla matrice identità.



Pensiamo \mathbb{R}^2 come l'insieme dei vettori del piano cartesiano applicati nell'origine. Sia $\mathcal{R} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ la rotazione di un angolo θ in senso antiorario. Siccome una rotazione manda parallelogrammi in parallelogrammi, \mathcal{R} è additiva; inoltre, poiché \mathcal{R} lascia invariati il verso e la lunghezza di un vettore, \mathcal{R} è anche omogenea. Quindi \mathcal{R} è lineare.

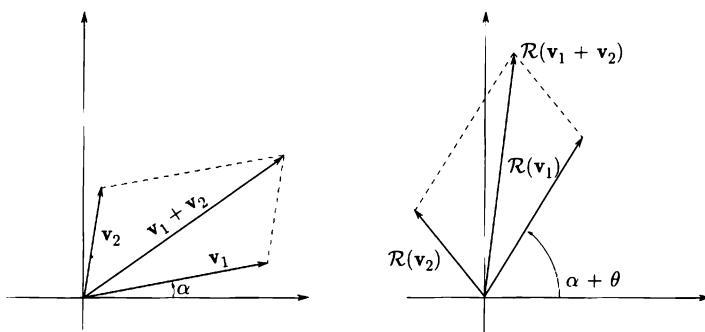


Figura 2.1. Rotazione del piano cartesiano.

Possiamo dimostrare anche algebricamente che \mathcal{R} è lineare, facendo vedere che è l'applicazione rappresentata da una matrice. Per questo usiamo le coordinate polari che sono più adatte a descrivere la rotazione \mathcal{R} : se $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} r \cos \alpha \\ r \sin \alpha \end{bmatrix}$, allora

$$\mathcal{R}(\mathbf{x}) = \begin{bmatrix} r \cos(\alpha + \theta) \\ r \sin(\alpha + \theta) \end{bmatrix}$$

Usando le formule di addizione delle funzioni trigonometriche troviamo

$$\begin{aligned} \mathcal{R}(\mathbf{x}) &= \begin{bmatrix} r(\cos \alpha \cos \theta - \sin \alpha \sin \theta) \\ r(\cos \alpha \sin \theta + \sin \alpha \cos \theta) \end{bmatrix} = \\ &= \begin{bmatrix} x \cos \theta - y \sin \theta \\ x \sin \theta + y \cos \theta \end{bmatrix} \end{aligned}$$

ovvero

$$\mathcal{R} \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Questo mostra che \mathcal{R} è l'applicazione rappresentata dalla matrice

$$\mathbf{Q} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

Vedremo più avanti che è possibile rovesciare il ragionamento: il teorema di rappresentazione consente di determinare in altro modo la matrice \mathbf{Q} e, quindi, di dedurre le formule trigonometriche dalla linearità di \mathcal{R} .



Linearità delle proiezioni su un asse coordinato

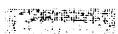
Sia $\epsilon_1 : \mathbb{K}^n \rightarrow \mathbb{K}$ la proiezione che a un vettore di \mathbb{K}^n associa la sua prima componente:

$$\epsilon_1([x_1, x_2, \dots, x_n]^T) = x_1$$

Si tratta dell'applicazione rappresentata dalla matrice $1 \times n$

$$\mathbf{A} = [1 \quad 0 \quad \dots \quad 0]$$

ed è perciò lineare. Analogamente sono lineari le proiezioni sugli altri assi coordinati.



Linearità delle coordinate

Questo esempio generalizza il precedente al caso astratto. Sia \mathbf{V} uno spazio vettoriale di dimensione n , e sia $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ una base di \mathbf{V} . Fissato un indice $k = 1, 2, \dots, n$, sia $\beta_k : \mathbf{V} \rightarrow \mathbb{K}$ la funzione che a un vettore \mathbf{v} di \mathbf{V} associa la sua k -esima coordinata:

$$\beta_k(x_1 \mathbf{b}_1 + \dots + x_k \mathbf{b}_k + \dots + x_n \mathbf{b}_n) = x_k$$

È immediato verificare che β_k è una funzione lineare: abbiamo visto la dimostrazione nel caso dei vettori liberi nel capitolo 1 (proposizione 4.4); tale dimostrazione rimane valida nel caso generale.

Da questo segue che è lineare la funzione $\mathbf{x} : \mathbf{V} \rightarrow \mathbb{K}^n$ che a un vettore di \mathbf{V} associa la n -upla delle sue coordinate: questo perché le componenti di \mathbf{x} sono le funzioni β_k (si veda anche la proposizione 4.4).



L'operatore di derivazione è un'applicazione lineare

Sia \mathcal{F} lo spazio vettoriale delle funzioni $f : \mathbb{R} \rightarrow \mathbb{R}$ e sia \mathcal{F}^1 il sottospazio delle funzioni che sono derivabili in ogni punto x di \mathbb{R} . Sia $\mathfrak{D} : \mathcal{F}^1 \rightarrow \mathcal{F}$ la trasformazione che associa a ogni funzione derivabile f la sua derivata f' :

$$\mathfrak{D}(f) = f'$$

La trasformazione \mathfrak{D} è lineare. Infatti

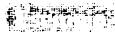
a) \mathfrak{D} è additiva:

$$\mathfrak{D}(f + g) = (f + g)' = f' + g' = \mathfrak{D}(f) + \mathfrak{D}(g)$$

b) \mathfrak{D} è omogenea:

$$\mathfrak{D}(tf) = (tf)' = tf' = t\mathfrak{D}(f)$$

Non si può sottolineare abbastanza l'importanza di questo esempio: da solo giustificherebbe l'introduzione degli spazi vettoriali astratti e del concetto di applicazione lineare. L'osservazione che la derivata è lineare apre la porta all'uso dei metodi dell'algebra lineare nello studio delle equazioni differenziali lineari.



L'integrale è un'applicazione lineare

Fissiamo un intervallo chiuso e limitato $[a, b]$ sulla retta reale. Sia $\mathcal{R}([a, b])$ l'insieme delle funzioni che sono Riemann integrabili su $[a, b]$. La somma di due funzioni integrabili è ancora integrabile, così come è integrabile il prodotto di una costante per una funzione integrabile. Quindi $\mathcal{R}([a, b])$ è un sottospazio vettoriale dello spazio delle funzioni $f : [a, b] \rightarrow \mathbb{R}$. L'integrale

$$\mathcal{I}(f) = \int_a^b f(x)dx$$

è una funzione $\mathcal{I} : \mathcal{R}([a, b]) \rightarrow \mathbb{R}$ ed è lineare perché

$$\mathcal{I}(f_1 + f_2) = \mathcal{I}(f_1) + \mathcal{I}(f_2) \quad \text{e} \quad \mathcal{I}(tf) = t\mathcal{I}(f)$$

È comodo, per esempio quando si deve mostrare che un'applicazione è lineare, riscrivere le due condizioni di additività e omogeneità come un'unica condizione:

Una funzione $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{W}$ è lineare se e solo se

$$(2.3) \quad \mathfrak{L}(t_1\mathbf{v}_1 + t_2\mathbf{v}_2) = t_1\mathfrak{L}(\mathbf{v}_1) + t_2\mathfrak{L}(\mathbf{v}_2) \quad \text{per ogni } \mathbf{v}_1, \mathbf{v}_2 \in \mathbf{V} \text{ e } t_1, t_2 \in \mathbb{K}.$$

In parole la (2.3) dice che \mathfrak{L} mappa una combinazione lineare di \mathbf{v}_1 e \mathbf{v}_2 nella corrispondente combinazione lineare dei vettori trasformati $\mathfrak{L}(\mathbf{v}_1)$ e $\mathfrak{L}(\mathbf{v}_2)$.

DIMOSTRAZIONE. Per verificare che la (2.3) implica la linearità di \mathfrak{L} , si osservi che l'additività si ritrova ponendo $t_1 = t_2 = 1$, l'omogeneità ponendo $t_2 = 0$. Viceversa, se \mathfrak{L} è lineare, allora usando prima l'additività e poi l'omogeneità di \mathfrak{L} ritroviamo (2.3):

$$\mathfrak{L}(t_1\mathbf{v}_1 + t_2\mathbf{v}_2) = \mathfrak{L}(t_1\mathbf{v}_1) + \mathfrak{L}(t_2\mathbf{v}_2) = t_1\mathfrak{L}(\mathbf{v}_1) + t_2\mathfrak{L}(\mathbf{v}_2)$$

OSSERVAZIONE Lo stesso ragionamento mostra che \mathbf{H} è un sottospazio di \mathbf{V} se e solo se, oltre a contenere il vettore nullo, è chiuso rispetto alle combinazioni lineari, cioè se $t_1\mathbf{v}_1 + t_2\mathbf{v}_2 \in \mathbf{H}$ per ogni $\mathbf{v}_1, \mathbf{v}_2 \in \mathbf{H}$ e $t_1, t_2 \in \mathbb{K}$.

Prendendo $t_1 = 1$ e $t_2 = -1$ nella (2.3), troviamo

$$(2.4) \quad \mathfrak{L}(\mathbf{v}_2 - \mathbf{v}_1) = \mathfrak{L}(\mathbf{v}_2) - \mathfrak{L}(\mathbf{v}_1)$$

Quindi un'applicazione lineare mappa la differenza di due vettori nella differenza dei vettori trasformati. Prendendo $\mathbf{v}_1 = \mathbf{v}_2$ nella (2.4) otteniamo

Se $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ è lineare, allora

$$(2.5) \quad \mathcal{L}(\mathbf{0}) = \mathbf{0}$$

Si osservi che $\mathbf{0}$ a primo membro della (2.5) denota il vettore nullo di \mathbf{V} , mentre $\mathbf{0}$ a secondo membro denota il vettore nullo di \mathbf{W} . Si tratta di una condizione *necessaria* per la linearità: se una funzione non mappa il vettore nullo di \mathbf{V} nel vettore nullo di \mathbf{W} , certamente non è lineare.

Un polinomio di primo grado *non omogeneo*, cioè con termine costante non nullo, non è lineare perché la (2.5) non è soddisfatta: come esempio esplicito, $y = x + 1$ non è una funzione lineare $\mathbb{R} \rightarrow \mathbb{R}$ perché $y(0) = 1 \neq 0$.

Infine, la (2.3) si generalizza a una combinazione lineare qualsiasi:

Se $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ è lineare, per ogni $\mathbf{v}_1, \dots, \mathbf{v}_d \in \mathbf{V}$ e ogni $t_1, \dots, t_d \in \mathbb{K}$

$$(2.6) \quad \mathcal{L}(t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \dots + t_d\mathbf{v}_d) = t_1\mathcal{L}(\mathbf{v}_1) + t_2\mathcal{L}(\mathbf{v}_2) + \dots + t_d\mathcal{L}(\mathbf{v}_d)$$

Mostrare che una matrice è determinata dall'applicazione che rappresenta: se \mathbf{A} e \mathbf{B} sono matrici $m \times n$ e $\mathcal{L}_{\mathbf{A}} = \mathcal{L}_{\mathbf{B}}$, allora $\mathbf{A} = \mathbf{B}$.

Dati uno spazio vettoriale \mathbf{V} e un vettore $\mathbf{v}_0 \in \mathbf{V}$, mostrare che la traslazione $\mathbf{T} : \mathbf{V} \rightarrow \mathbf{V}$ definita da $\mathbf{T}(\mathbf{v}) = \mathbf{v} + \mathbf{v}_0$ è lineare se e solo se $\mathbf{v}_0 = \mathbf{0}$.

Traccia di una matrice La *traccia* di una matrice quadrata $\mathbf{A} = [a_{ij}]$ è la somma degli elementi sulla diagonale principale:

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}$$

Mostrare che $\text{tr} : \mathbb{M}_{\mathbb{K}}(n, n) \rightarrow \mathbb{K}$ è lineare.

Dati due spazi vettoriali \mathbf{V} e \mathbf{W} sul campo \mathbb{K} , per ogni vettore $\mathbf{w}_0 \in \mathbf{W}$ possiamo considerare la funzione costante

$$\mathbf{T}(\mathbf{v}) = \mathbf{w}_0 \quad \text{per ogni } \mathbf{v} \in \mathbf{V}$$

Mostrare che \mathbf{T} è lineare se e solo se $\mathbf{w}_0 = \mathbf{0}$ (nel qual caso \mathbf{T} si dice applicazione lineare nulla).

 Sia \mathcal{F} lo spazio vettoriale delle funzioni $y(t) : \mathbb{R} \rightarrow \mathbb{R}$ e sia $\mathcal{F}^{(2)}$ il sottospazio delle funzioni che sono derivabili due volte in ogni punto x di \mathbb{R} . Sia $\mathcal{L} : \mathcal{F}^{(2)} \rightarrow \mathcal{F}$ la funzione:

$$\mathcal{L}(y(t)) = 3y''(t) + e^t y(t)$$

Mostrare che \mathcal{L} è lineare.

 L'insieme dei numeri complessi \mathbb{C} è uno spazio vettoriale su \mathbb{R} ma anche uno spazio vettoriale su \mathbb{C} . Mostrare che la funzione $\mathbb{C} \rightarrow \mathbb{C}$ che al numero complesso $z = x + iy$ associa il suo coniugato

$$\overline{x + iy} = x - iy \quad (x \in \mathbb{R}, y \in \mathbb{R})$$

è lineare su \mathbb{R} , ma non è lineare su \mathbb{C} .

 Una fonte importante di esempi di funzioni lineari è costituita dai *prodotti*: si dice che un prodotto, i cui fattori siano elementi di uno spazio vettoriale, è bilineare se, fissato arbitrariamente il primo dei fattori, il prodotto è una funzione lineare del secondo, e, fissato arbitrariamente il secondo dei fattori, il prodotto è una funzione lineare del primo.

Mostrare che i seguenti prodotti sono bilineari: il prodotto di numeri reali, il prodotto di numeri complessi, il prodotto di matrici, il prodotto scalare di due vettori di \mathbb{R}^3 e il prodotto vettoriale di due vettori di \mathbb{R}^3 .

 Siano \mathbf{A} e \mathbf{B} due matrici quadrate di ordine n . Si consideri il commutatore o parentesi di Lie di \mathbf{A} e \mathbf{B} :

$$[\mathbf{A}, \mathbf{B}] = \mathbf{AB} - \mathbf{BA}$$

Si dimostri che $[\mathbf{A}, \mathbf{B}]$ è un prodotto bilineare, ma non è associativo ($[\mathbf{A}, [\mathbf{B}, \mathbf{C}]] \neq [[\mathbf{A}, \mathbf{B}], \mathbf{C}]$ in genere).

3 NUCLEO, FIBRE E IMMAGINE

Un'equazione lineare è, in astratto, un'equazione del tipo

$$(3.1) \quad \mathcal{L}(\mathbf{v}) = \mathbf{w}$$

in cui $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ è un'applicazione lineare, \mathbf{w} è il termine noto e \mathbf{v} è l'incognita. I sistemi lineari sono il caso particolare in cui $\mathcal{L} = \mathcal{L}_{\mathbf{A}} : \mathbb{K}^n \rightarrow \mathbb{K}^m$ è l'applicazione rappresentata dalla matrice \mathbf{A} , $\mathbf{v} = \mathbf{x} \in \mathbb{K}^n$ è il vettore delle incognite, mentre $\mathbf{w} = \mathbf{b} \in \mathbb{K}^m$ è il vettore dei termini noti: l'equazione (3.1) è in questo caso $\mathbf{Ax} = \mathbf{b}$. Un altro esempio fondamentale è costituito dalle *equazioni differenziali lineari* (si veda il volume *Analisi matematica 2* di M. Bramanti, C. D. Pagani e S. Salsa, Zanichelli 2009). In questo caso l'incognita è una funzione $y(t)$, il termine noto è una funzione $g(t)$ e la funzione \mathcal{L} è un operatore differenziale lineare, tipicamente una combinazione lineare di derivate del primo e secondo ordine. Un esempio è fornito dall'equazione dell'*oscillatore armonico*

$$y''(t) + y(t) = 0$$

in cui $\mathcal{L}(y(t)) = y''(t) + y(t)$.

In questo paragrafo vogliamo generalizzare al contesto astratto quanto già visto sulla struttura dell'insieme delle soluzioni di un sistema lineare: le soluzioni di

un'equazione lineare si ottengono sommando una soluzione particolare alle soluzioni dell'equazione omogenea associata.

Fissiamo la terminologia della teoria delle funzioni che useremo nel seguito. Data un'applicazione $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$, l'insieme \mathbf{V} si dice *dominio* dell'applicazione e l'insieme \mathbf{W} *codominio*. Se vale la relazione $\mathcal{L}(\mathbf{v}) = \mathbf{w}$, si dice che \mathcal{L} mappa \mathbf{v} in \mathbf{w} , che \mathbf{w} è l'immagine di \mathbf{v} mediante \mathcal{L} e che \mathbf{v} è una controimmagine di \mathbf{w} . Un vettore \mathbf{w} di \mathbf{W} può avere più controimmagini o non averne neanche una. In ogni caso è interessante considerare l'insieme delle controimmagini di \mathbf{w} dato che è l'insieme delle soluzioni dell'equazione $\mathcal{L}(\mathbf{v}) = \mathbf{w}$.

DEFINIZIONE 3.1 (Fibre di un'applicazione)

Dati $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ e $\mathbf{w} \in \mathbf{W}$, l'insieme delle controimmagini di \mathbf{w} è un sottoinsieme di \mathbf{V} che si denota col simbolo $\mathcal{L}^{-1}(\mathbf{w})$ e si dice *fiba* di \mathcal{L} sopra \mathbf{w} :

$$\mathcal{L}^{-1}(\mathbf{w}) = \{\mathbf{v} \in \mathbf{V} : \mathcal{L}(\mathbf{v}) = \mathbf{w}\}$$

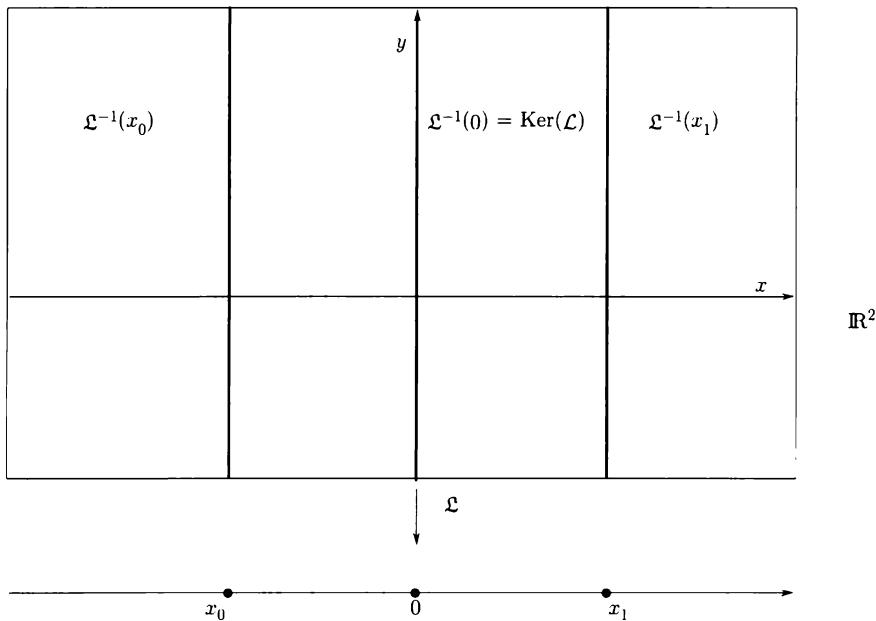


Figura 3.1. Nucleo e fibre di $\mathcal{L}([x, y]^T) = x$.

Una funzione \mathcal{L} si dice *iniettiva* se $\mathbf{v}_1 \neq \mathbf{v}_2$ in \mathbf{V} implica $\mathcal{L}(\mathbf{v}_1) \neq \mathcal{L}(\mathbf{v}_2)$ in \mathbf{W} ; questo equivale a dire che per ogni $\mathbf{w} \in \mathbf{W}$ la fibra $\mathcal{L}^{-1}(\mathbf{w})$ o è vuota o consiste di un unico elemento. Quindi \mathcal{L} è iniettiva se e solo se, per ogni $\mathbf{w} \in \mathbf{W}$, l'equazione $\mathcal{L}(\mathbf{v}) = \mathbf{w}$ ammette al più una soluzione.

La funzione \mathcal{L} si dice *suriettiva* se ogni $\mathbf{w} \in \mathbf{W}$ è immagine mediante \mathcal{L} di almeno un vettore \mathbf{v} di \mathbf{V} ; questo equivale a dire che per ogni $\mathbf{w} \in \mathbf{W}$ la fibra $\mathcal{L}^{-1}(\mathbf{w})$ non

è l'insieme vuoto. Quindi \mathfrak{L} è suriettiva se e solo se, per ogni $\mathbf{w} \in \mathbf{W}$, l'equazione $\mathfrak{L}(\mathbf{v}) = \mathbf{w}$ ammette almeno una soluzione.

La funzione \mathfrak{L} si dice *bijettiva biunivoca* se è iniettiva e suriettiva; questo equivale a dire che per ogni $\mathbf{w} \in \mathbf{W}$ la fibra $\mathfrak{L}^{-1}(\mathbf{w})$ consiste di uno e un solo vettore $\mathbf{v} \in \mathbf{V}$, ovvero che per ogni $\mathbf{w} \in \mathbf{W}$ l'equazione $\mathfrak{L}(\mathbf{v}) = \mathbf{w}$ ammette una e una sola soluzione \mathbf{v} .

Queste definizioni sono valide per una funzione qualsiasi. Nel caso particolare delle funzioni lineari, vogliamo ora mostrare che le diverse fibre di \mathfrak{L} sono strettamente legate tra loro; per capire in che modo, introduciamo il *nucleo* di \mathfrak{L} , che è la fibra sopra il vettore nullo $\mathbf{0}_\mathbf{W}$ di \mathbf{W} , ovvero l'insieme delle soluzioni dell'equazione omogenea $\mathfrak{L}(\mathbf{v}) = \mathbf{0}_\mathbf{W}$ (usiamo il simbolo $\mathbf{0}_\mathbf{W}$ per il vettore nullo di \mathbf{W} quando sia necessario distinguerlo dal vettore nullo $\mathbf{0}_\mathbf{V}$ di \mathbf{V}).

DEFINIZIONE 3.2 (Nucleo di un'applicazione lineare)

Sia $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare. Il nucleo $\text{Ker}(\mathfrak{L})$ è la fibra di \mathfrak{L} sopra $\mathbf{0}_\mathbf{W}$:

$$\text{Ker}(\mathfrak{L}) = \{\mathbf{v} \in \mathbf{V} : \mathfrak{L}(\mathbf{v}) = \mathbf{0}_\mathbf{W}\}$$

Esempio

Il nucleo dell'applicazione lineare $\mathfrak{L}_\mathbf{A}$ rappresentata dalla matrice \mathbf{A} coincide con il nucleo della matrice \mathbf{A} .

Esempio

Consideriamo la proiezione ortogonale $\mathfrak{P} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ del piano cartesiano sull'asse x :

$$\mathfrak{P} \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} x \\ 0 \end{bmatrix}$$

Si tratta di un'applicazione lineare perché le sue componenti sono polinomi omogenei di primo grado in x e y . Il nucleo di \mathfrak{P} consiste dei vettori con $x = 0$, ed è quindi l'asse y . Un vettore $\mathbf{w} := [x_0, y_0]^T$ è immagine di un vettore di \mathbb{R}^2 se e solo se $y_0 = 0$, cioè se e solo se appartiene all'asse x . La fibra $\mathfrak{P}^{-1}(\mathbf{w})$ è quindi l'insieme vuoto se \mathbf{w} non appartiene all'asse x ; se invece $\mathbf{w} = [x_0, 0]^T$ appartiene all'asse x , la fibra su \mathbf{w} è la retta di equazione $x = x_0$ perché:

$$\mathfrak{P}^{-1} \left(\begin{bmatrix} x_0 \\ 0 \end{bmatrix} \right) = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 : x = x_0 \right\}$$

Osserviamo che le fibre sono rette parallele al nucleo, che è la retta di equazione $x = 0$.

Il nucleo di una matrice $m \times n$ è un sottospazio vettoriale di \mathbb{K}^n ; più in generale:

PROPOSIZIONE 3.3 Il nucleo di un'applicazione lineare $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{W}$ è un sottospazio vettoriale di \mathbf{V} .

DIMOSTRAZIONE. Dobbiamo dimostrare che $\text{Ker}(\mathfrak{L})$ contiene il vettore nullo di \mathbf{V} ed è chiuso rispetto alla somma e al prodotto per scalare. Il vettore nullo di \mathbf{V} appartiene al

nucleo per (2.5). Se \mathbf{v}_1 e \mathbf{v}_2 sono due vettori del nucleo, anche la loro somma appartiene al nucleo:

$$\mathcal{L}(\mathbf{v}_1 + \mathbf{v}_2) = \mathcal{L}(\mathbf{v}_1) + \mathcal{L}(\mathbf{v}_2) = \mathbf{0}_W + \mathbf{0}_W = \mathbf{0}_W$$

Infine, se $t \in \mathbb{K}$ e $\mathbf{v} \in \text{Ker}(\mathcal{L})$: allora

$$\mathcal{L}(t\mathbf{v}) = t\mathcal{L}(\mathbf{v}) = t\mathbf{0}_W = \mathbf{0}_W$$

quindi anche $t\mathbf{v} \in \text{Ker}(\mathcal{L})$: il nucleo è chiuso rispetto al prodotto scalare. 

Analogamente al caso dei sistemi lineari, le soluzioni di un'equazione lineare $\mathcal{L}(\mathbf{v}) = \mathbf{w}$ si ottengono sommando a una soluzione particolare le soluzioni dell'equazione omogenea associata:

PROPOSIZIONE 3.4 Sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare e sia $\mathbf{w} \in \mathbf{W}$.

Supponiamo che $\mathbf{v}_0 \in \mathbf{V}$ sia una soluzione particolare dell'equazione $\mathcal{L}(\mathbf{v}) = \mathbf{w}$.

Allora le soluzioni di $\mathcal{L}(\mathbf{v}) = \mathbf{w}$ si ottengono sommando a \mathbf{v}_0 i vettori del nucleo di \mathcal{L} : sono i vettori della forma

$$\mathbf{v} = \mathbf{v}_0 + \mathbf{v}_h$$

con $\mathbf{v}_h \in \text{Ker}(\mathcal{L})$.

DIMOSTRAZIONE. Supponiamo che $\mathbf{v} \in \mathbf{V}$ soddisfi $\mathcal{L}(\mathbf{v}) = \mathbf{w}$ e consideriamo $\mathbf{v}_h = \mathbf{v} - \mathbf{v}_0$. Allora usando prima l'additività e poi l'omogeneità di \mathcal{L} possiamo scrivere le uguaglianze:

$$\mathcal{L}(\mathbf{v}_h) = \mathcal{L}(\mathbf{v}) + \mathcal{L}(-\mathbf{v}_0) = \mathcal{L}(\mathbf{v}) - \mathcal{L}(\mathbf{v}_0) = \mathbf{w} - \mathbf{w} = \mathbf{0}$$

quindi \mathbf{v}_h è un vettore del nucleo.

Viceversa, se \mathbf{v}_h è un vettore del nucleo, poniamo $\mathbf{v} = \mathbf{v}_0 + \mathbf{v}_h$. Allora

$$\mathcal{L}(\mathbf{v}) = \mathcal{L}(\mathbf{v}_0) + \mathcal{L}(\mathbf{v}_h) = \mathbf{w} + \mathbf{0} = \mathbf{w}$$

quindi \mathbf{v} è una soluzione di $\mathcal{L}(\mathbf{v}) = \mathbf{w}$. 

Traduciamo in termini più geometrici l'enunciato della proposizione 3.4. Se $\mathbf{w} \in \mathbf{W}$, la fibra $\mathcal{L}^{-1}(\mathbf{w})$ può essere l'insieme vuoto. Se non è l'insieme vuoto, allora contiene un vettore \mathbf{v}_0 e

$$\mathcal{L}^{-1}(\mathbf{w}) = \mathbf{v}_0 + \text{Ker}(\mathcal{L})$$

Le fibre non vuote di \mathcal{L} si ottengono, quindi, per traslazione dal nucleo, che è la fibra sopra il vettore nullo: si dice che sono *parallele* al nucleo e che il nucleo è il loro *sottospazio direttore* per analogia con il caso dello spazio cartesiano. Il dominio \mathbf{V} di \mathcal{L} è l'unione di queste fibre parallele: è come affettato, le fette sono le fibre non vuote di \mathcal{L} e sono sostanzialmente identiche tra loro perché coincidono a meno di una traslazione.



Riprendiamo l'esempio dell'operatore $\mathcal{D} : \mathcal{F}^1 \rightarrow \mathcal{F}$ che a una funzione derivabile $y(x)$ associa la sua derivata $y'(x)$. L'equazione lineare $\mathcal{D}(y) = g$ ha come soluzioni le funzioni y che sono primitive di g . Siccome il vettore nullo di \mathcal{F} è la funzione nulla, il nucleo di \mathcal{D} , cioè l'insieme

delle soluzioni dell'equazione omogenea $\mathfrak{D}(y) = \mathbf{0}$, è costituito dalle funzioni costanti. Si tratta di uno spazio vettoriale di dimensione 1, e una sua base è costituita dalla funzione costante uguale a 1.

Le soluzioni di $\mathfrak{D}(y) = g$ sono le funzioni del tipo

$$y(x) = z(x) + c$$

dove z è una particolare primitiva di g e c è una costante, cioè un elemento di $\text{Ker}(\mathfrak{D})$. Per esempio, se $g(x) = \cos(x)$, una soluzione particolare di $\mathfrak{D}(y) = \cos(x)$ è $z(x) = \sin(x)$ e una soluzione arbitraria ha la forma $\sin(x) + c$.

COROLARIO 3.5 (Criterio di iniettività)

Un'applicazione lineare $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{W}$ è iniettiva se e solo se

$$\text{Ker}(\mathfrak{L}) = \{\mathbf{0}_\mathbf{V}\}. \quad \begin{array}{l} \text{Ker}(\mathfrak{L}) = \text{Nucleo di } \mathfrak{L} \\ \text{non contiene vettori} \end{array}$$

DIMOSTRAZIONE. La funzione $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{W}$ è *iniettiva*, se per ogni $\mathbf{w} \in \mathbf{W}$ la fibra $\mathfrak{L}^{-1}(\mathbf{w})$ contiene al più un vettore \mathbf{v}_0 . Se \mathfrak{L} è lineare e $\mathfrak{L}^{-1}(\mathbf{w})$ è non vuoto e contiene il vettore \mathbf{v}_0 , allora

$$\mathfrak{L}^{-1}(\mathbf{w}) = \mathbf{v}_0 + \text{Ker}(\mathfrak{L})$$

Quindi $\mathfrak{L}^{-1}(\mathbf{w})$ consiste del solo vettore \mathbf{v}_0 se e solo se il nucleo consiste del solo vettore nullo ed \mathfrak{L} è iniettiva se e solo se $\text{Ker}(\mathfrak{L}) = \{\mathbf{0}_\mathbf{V}\}$.

L'utilità della proposizione è che, per stabilire che \mathfrak{L} è iniettiva, *a priori* dovremmo controllare che *ogni* fibra $\mathfrak{L}^{-1}(\mathbf{w})$ non contiene più di un vettore; se \mathfrak{L} è lineare, basta verificare questo solo per la fibra $\mathfrak{L}^{-1}(\mathbf{0}_\mathbf{W}) = \text{Ker}(\mathbf{0})$. Il motivo è che le altre fibre non vuote si ottengono dal nucleo per traslazione.

Spostiamo ora la nostra attenzione sul codominio \mathbf{W} .

DEFINIZIONE 3.6 Sia $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{W}$ una funzione. Per ogni sottoinsieme \mathbf{U} di \mathbf{V} , l'insieme delle immagini di elementi di \mathbf{U} mediante \mathfrak{L} si dice *immagine di \mathbf{U}* e si denota con il simbolo $\mathfrak{L}(\mathbf{U})$:

$$\mathfrak{L}(\mathbf{U}) = \{\mathbf{w} \in \mathbf{W} : \text{esiste } \mathbf{v} \in \mathbf{U} \text{ tale che } \mathbf{w} = \mathfrak{L}(\mathbf{v})\}$$

L'immagine di \mathbf{V} si denota anche con il simbolo $\text{Im}(\mathfrak{L})$ e si dice *immagine di \mathfrak{L}* :

$$\text{Im}(\mathfrak{L}) = \mathfrak{L}(\mathbf{V})$$

Nel caso di un'applicazione lineare, l'immagine di un sottospazio vettoriale di \mathbf{V} è un sottospazio vettoriale di \mathbf{W} :

PROPOSIZIONE 3.7 Sia $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare. Se \mathbf{H} è un sottospazio vettoriale di \mathbf{V} , la sua immagine $\mathfrak{L}(\mathbf{H})$ è un sottospazio vettoriale di \mathbf{W} . In particolare, l'immagine $\text{Im}(\mathfrak{L})$ è un sottospazio vettoriale di \mathbf{W} .

DIMOSTRAZIONE. Dobbiamo dimostrare che il vettore nullo $\mathbf{0}_W$ appartiene a $\mathcal{L}(\mathbf{H})$, e che $\mathcal{L}(\mathbf{H})$ è chiuso rispetto alle combinazioni lineari.

Il vettore nullo $\mathbf{0}_V$ appartiene a \mathbf{H} perché \mathbf{H} è un sottospazio, e $\mathcal{L}(\mathbf{0}_V) = \mathbf{0}_W$, quindi $\mathbf{0}_W$ appartiene a $\mathcal{L}(\mathbf{H})$. Mostriamo ora che $\mathcal{L}(\mathbf{H})$ è chiuso rispetto alle combinazioni lineari. Siamo \mathbf{w}_1 e \mathbf{w}_2 due vettori di $\mathcal{L}(\mathbf{H})$ e siano t_1 e t_2 due scalari. Dobbiamo dimostrare che $t_1\mathbf{w}_1 + t_2\mathbf{w}_2 \in \mathcal{L}(\mathbf{H})$. Siccome \mathbf{w}_1 e \mathbf{w}_2 appartengono a $\mathcal{L}(\mathbf{H})$, esistono \mathbf{v}_1 e \mathbf{v}_2 in \mathbf{H} tali che $\mathcal{L}(\mathbf{v}_1) = \mathbf{w}_1$ e $\mathcal{L}(\mathbf{v}_2) = \mathbf{w}_2$. Allora

$$\mathcal{L}(t_1\mathbf{v}_1 + t_2\mathbf{v}_2) = t_1\mathcal{L}(\mathbf{v}_1) + t_2\mathcal{L}(\mathbf{v}_2) = t_1\mathbf{w}_1 + t_2\mathbf{w}_2$$

Poiché \mathbf{H} è un sottospazio, la combinazione lineare $t_1\mathbf{v}_1 + t_2\mathbf{v}_2$ è un vettore di \mathbf{H} , e quindi l'uguaglianza precedente mostra che

$$t_1\mathbf{w}_1 + t_2\mathbf{w}_2 \in \mathcal{L}(\mathbf{H})$$

come dovevansi dimostrare.

Concludiamo questo paragrafo mostrando che, quando \mathcal{L} è l'applicazione rappresentata dalla matrice \mathbf{A} , il nucleo di \mathcal{L} coincide con quello di \mathbf{A} , mentre l'immagine di \mathcal{L} è lo spazio colonna della matrice.

PROPOSIZIONE 3.8 Sia \mathbf{A} una matrice di tipo (m, n) e sia $\mathcal{L}_{\mathbf{A}} : \mathbb{K}^n \rightarrow \mathbb{K}^m$ l'applicazione lineare rappresentata da \mathbf{A} . Allora:

- a) $\text{Ker}(\mathcal{L}_{\mathbf{A}}) = \text{Ker}(\mathbf{A})$;
- b) $\text{Im}(\mathcal{L}_{\mathbf{A}}) = \text{Col}(\mathbf{A})$.

DIMOSTRAZIONE. Siccome $\mathcal{L}_{\mathbf{A}}(\mathbf{v}) = \mathbf{Av}$, il nucleo di \mathcal{L} coincide con quello di \mathbf{A} . Un vettore $\mathbf{w} \in \mathbb{K}^m$ appartiene a $\text{Im}(\mathcal{L}_{\mathbf{A}})$ se e solo se esiste $\mathbf{v} \in \mathbb{K}^n$ tale che $\mathbf{w} = \mathbf{Av}$. D'altra parte, il prodotto \mathbf{Av} è la combinazione lineare delle colonne di \mathbf{A} che ha per coefficienti le componenti del vettore \mathbf{v} . Ne segue che l'immagine di $\mathcal{L}_{\mathbf{A}}$ consiste di quei vettori che sono combinazione lineare delle colonne di \mathbf{A} e, quindi, coincide con lo spazio colonna della matrice. 

 Sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare. Mostrare che, se i vettori $\mathbf{v}_1, \dots, \mathbf{v}_n$ sono linearmente dipendenti in \mathbf{V} , allora le loro immagini $\mathcal{L}(\mathbf{v}_1), \dots, \mathcal{L}(\mathbf{v}_n)$ sono linearmente dipendenti in \mathbf{W} .

Sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare. Mostrare che se i vettori $\mathbf{v}_1, \dots, \mathbf{v}_n$ sono linearmente indipendenti in \mathbf{V} e il nucleo di \mathcal{L} consiste del solo vettore nullo, allora $\mathcal{L}(\mathbf{v}_1), \dots, \mathcal{L}(\mathbf{v}_n)$ sono linearmente indipendenti in \mathbf{W} . Mostrare con un esempio che l'ipotesi $\text{Ker}(\mathcal{L}) = \{\mathbf{0}\}$ è necessaria: se i vettori $\mathbf{v}_1, \dots, \mathbf{v}_n$ sono linearmente indipendenti in \mathbf{V} , non è detto che i vettori $\mathcal{L}(\mathbf{v}_1), \dots, \mathcal{L}(\mathbf{v}_n)$ siano linearmente indipendenti in \mathbf{W} (per esempio, considerare l'immagine della base canonica mediante l'applicazione nulla o mediante la proiezione del piano sull'asse x).

Sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare. Sia \mathbf{Z} un sottospazio di \mathbf{W} . Mostrare che la controimmagine di \mathbf{Z} :

$$\mathcal{L}^{-1}(\mathbf{Z}) = \{\mathbf{v} \in \mathbf{V} : \mathcal{L}(\mathbf{v}) \in \mathbf{Z}\}$$

è un sottospazio di \mathbf{V} .

4 ALGEBRA DELLE APPLICAZIONI LINEARI

È utile introdurre sulle applicazioni lineari le operazioni analoghe a quelle che abbiamo definito per le matrici. Come nel caso delle matrici l'operazione più ricca di applicazioni è il prodotto, così nel caso delle applicazioni l'operazione corrispondente, *il prodotto di composizione*, ha un ruolo fondamentale. Ma procediamo con ordine. Come l'insieme delle matrici $m \times n$, con *un numero di righe e colonne fissati*, è uno spazio vettoriale, così l'insieme delle applicazioni lineari con *dominio e codominio fissati* è uno spazio vettoriale. Per spiegare questo, fissiamo due spazi vettoriali \mathbf{V} e \mathbf{W} sul campo \mathbb{K} , e denotiamo col simbolo $\text{Hom}_{\mathbb{K}}(\mathbf{V}, \mathbf{W})$ l'insieme delle applicazioni lineari da \mathbf{V} in \mathbf{W} :

$$\text{Hom}_{\mathbb{K}}(\mathbf{V}, \mathbf{W}) = \{\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W} : \mathcal{L} \text{ è lineare}\}$$

Date due applicazioni lineari \mathcal{L}_1 e \mathcal{L}_2 da \mathbf{V} in \mathbf{W} , la somma $\mathcal{L}_1 + \mathcal{L}_2 : \mathbf{V} \rightarrow \mathbf{W}$ è la funzione che a un vettore $\mathbf{v} \in \mathbf{V}$ associa il vettore

$$(\mathcal{L}_1 + \mathcal{L}_2)(\mathbf{v}) = \mathcal{L}_1(\mathbf{v}) + \mathcal{L}_2(\mathbf{v}).$$

Dati uno scalare $t \in \mathbb{K}$ e un'applicazione lineare $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$, il prodotto di \mathcal{L} per lo scalare t è la funzione $t\mathcal{L}$ che a un vettore $\mathbf{v} \in \mathbf{V}$ associa il vettore

$$(t\mathcal{L})(\mathbf{v}) = t(\mathcal{L}(\mathbf{v}))$$

È immediato verificare che $\mathcal{L}_1 + \mathcal{L}_2$ e $t\mathcal{L}$ sono lineari, per cui sull'insieme $\text{Hom}_{\mathbb{K}}(\mathbf{V}, \mathbf{W})$ abbiamo definito le operazioni di somma e prodotto per uno scalare ed è facile verificare che rispetto a tali operazioni $\text{Hom}_{\mathbb{K}}(\mathbf{V}, \mathbf{W})$ è uno spazio vettoriale; si osservi che il vettore nullo di $\text{Hom}_{\mathbb{K}}(\mathbf{V}, \mathbf{W})$ è la funzione che associa a ogni vettore $\mathbf{v} \in \mathbf{V}$ il vettore nullo di \mathbf{W} .

Il simbolo Hom viene da *omomorfismo*: in algebra un omomorfismo è una funzione che preserva le operazioni della struttura algebrica che si sta trattando; in questo caso la struttura algebrica è quella di spazio vettoriale e un omomorfismo tra spazi vettoriali è un'operazione che preserva somma e prodotto per scalare, cioè un'applicazione lineare (cosa sarà un omomorfismo di campi?).

Passiamo ora all'operazione che corrisponde al prodotto di matrici. Siano \mathbf{V} , \mathbf{W} e \mathbf{Z} spazi vettoriali sul campo \mathbb{K} . Date due funzioni $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ e $\mathfrak{M} : \mathbf{W} \rightarrow \mathbf{Z}$, il loro *prodotto di composizione* $\mathfrak{M} \circ \mathcal{L} : \mathbf{V} \rightarrow \mathbf{Z}$ (o funzione composta) è la funzione che a $\mathbf{v} \in \mathbf{V}$ associa $\mathfrak{M}(\mathcal{L}(\mathbf{v})) \in \mathbf{Z}$:

$$\mathfrak{M} \circ \mathcal{L}(\mathbf{v}) = \mathfrak{M}(\mathcal{L}(\mathbf{v}))$$

Questa definizione naturalmente ha senso per funzioni qualsiasi, non necessariamente lineari. Il punto è:

PROPOSIZIONE 4.1 Se $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{W}$ e $\mathfrak{M} : \mathbf{W} \rightarrow \mathbf{Z}$ sono lineari, allora anche $\mathfrak{M} \circ \mathfrak{L} : \mathbf{V} \rightarrow \mathbf{Z}$ è lineare.

DIMOSTRAZIONE. La dimostrazione è semplice, si sfrutta prima la linearità di \mathfrak{L} e poi quella di \mathfrak{M} :

$$\begin{aligned}\mathfrak{M} \circ \mathfrak{L}(t_1 \mathbf{v}_1 + t_2 \mathbf{v}_2) &= \mathfrak{M}(\mathfrak{L}(t_1 \mathbf{v}_1 + t_2 \mathbf{v}_2)) = \mathfrak{M}(t_1 \mathfrak{L}(\mathbf{v}_1) + t_2 \mathfrak{L}(\mathbf{v}_2)) = \\ &= t_1 \mathfrak{M}(\mathfrak{L}(\mathbf{v}_1)) + t_2 \mathfrak{M}(\mathfrak{L}(\mathbf{v}_2)) = t_1 \mathfrak{M} \circ \mathfrak{L}(\mathbf{v}_1) + t_2 \mathfrak{M} \circ \mathfrak{L}(\mathbf{v}_2)\end{aligned}$$

Nel caso che dominio e codominio coincidano, che è il caso più frequente e importante, possiamo introdurre le potenze di un'applicazione lineare: data $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{V}$, definiamo \mathfrak{L}^n per $n \geq 1$ ponendo

$$\mathfrak{L}^n = \mathfrak{L} \circ \mathfrak{L} \circ \cdots \circ \mathfrak{L} \quad n \text{ fattori}$$

e

$$\mathfrak{L}^0 = \mathcal{I}_{\mathbf{V}} \quad (\text{la funzione identità})$$

Per la proposizione precedente, \mathfrak{L}^n è ancora lineare. Un'applicazione lineare da \mathbf{V} in se stesso si dice *endomorfismo* e lo spazio vettoriale $\text{Hom}_{\mathbb{K}}(\mathbf{V}, \mathbf{V})$ si denota anche col simbolo $\text{End}_{\mathbb{K}}(\mathbf{V})$. Siccome $\text{End}_{\mathbb{K}}(\mathbf{V})$ è uno spazio vettoriale, possiamo considerare combinazioni lineari di potenze di \mathfrak{L} e ottenere così *polinomi* in \mathfrak{L} : dato un polinomio

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

nella variabile x i cui coefficienti a_i appartengano al campo \mathbb{K} , possiamo definire

$$P(\mathfrak{L}) = a_n \mathfrak{L}^n + a_{n-1} \mathfrak{L}^{n-1} + \cdots + a_1 \mathfrak{L} + a_0 \mathcal{I}_{\mathbf{V}}$$

sostituendo alla variabile x l'applicazione \mathfrak{L} . Si osservi che il termine noto a_0 va moltiplicato per l'applicazione identità $\mathcal{I}_{\mathbf{V}} = \mathfrak{L}^0$. Per qualunque polinomio $P(x)$, la funzione $P(\mathfrak{L})$ è un'applicazione lineare.

L'esempio più importante è quello dei *polinomi differenziali*, quando \mathbf{V} è lo spazio vettoriale $C^\infty(\mathbb{R})$ delle funzioni $y(t) : \mathbb{R} \rightarrow \mathbb{R}$ che sono derivabili infinite volte, e l'endomorfismo è l'operatore derivata $\mathfrak{D}(y) = y'$. Allora $\mathfrak{D}^n(y)$ è la derivata n -esima $y^{(n)}$ di y , un polinomio differenziale $P(\mathfrak{D})$ ha quindi la forma

$$P(\mathfrak{D})(y) = a_n y^{(n)} + a_{n-1} y^{(n-1)} + \cdots + a_1 y' + a_0 y$$

Per quanto osservato sopra, $P(\mathfrak{D})$ è una funzione lineare di y . In particolare, all'equazione differenziale

$$P(\mathfrak{D})(y(t)) = g(t)$$

si applicano tutte le considerazioni fatte per l'equazione lineare astratta $\mathfrak{L}(\mathbf{v}) = \mathbf{w}$. Per esempio, le soluzioni dell'equazione si trovano sommando a una soluzione particolare le soluzioni dell'equazione omogenea associata $P(\mathfrak{D})(y(t)) = \mathbf{0}$.



Sia $\mathcal{R} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ la rotazione del piano cartesiano \mathbb{R}^2 di un angolo retto in senso antiorario. Allora \mathcal{R}^n è la rotazione dell'angolo $n\pi/2$. In particolare, \mathcal{R}^4 è l'applicazione identità.

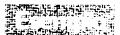
Il prodotto di composizione di applicazioni corrisponde al prodotto di matrici in un senso preciso:

PROPOSIZIONE 4.2 Siano \mathbf{A} e \mathbf{B} due matrici di tipo (m, n) e (n, p) , rispettivamente. Allora

$$\mathcal{L}_{\mathbf{AB}} = \mathcal{L}_{\mathbf{A}} \circ \mathcal{L}_{\mathbf{B}}$$

DIMOSTRAZIONE. Si noti che $\mathcal{L}_{\mathbf{A}}$ mappa \mathbb{K}^n in \mathbb{K}^m e $\mathcal{L}_{\mathbf{B}}$ mappa \mathbb{K}^p in \mathbb{K}^n , quindi $\mathcal{L}_{\mathbf{A}} \circ \mathcal{L}_{\mathbf{B}}$ mappa \mathbb{K}^p in \mathbb{K}^m . Anche $\mathcal{L}_{\mathbf{AB}}$ mappa \mathbb{K}^p in \mathbb{K}^m perché \mathbf{AB} è di tipo (m, p) . Dobbiamo verificare che queste due funzioni hanno lo stessa immagine per ciascun $\mathbf{v} \in \mathbb{K}^p$: in effetti

$$\mathcal{L}_{\mathbf{A}} \circ \mathcal{L}_{\mathbf{B}}(\mathbf{v}) = \mathcal{L}_{\mathbf{A}}(\mathcal{L}_{\mathbf{B}}(\mathbf{v})) = \mathbf{A}(\mathbf{B}\mathbf{v}) = (\mathbf{AB})(\mathbf{v}) = \mathcal{L}_{\mathbf{AB}}(\mathbf{v})$$



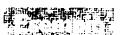
La rotazione \mathcal{R} del piano cartesiano \mathbb{R}^2 di un angolo retto in senso antiorario è rappresentata dalla matrice

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

Le matrici

$$\mathbf{A}^2 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \quad \mathbf{A}^4 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

rappresentano, rispettivamente, la rotazione \mathcal{R}^2 di una angolo π e la rotazione \mathcal{R}^4 di una angolo 2π , cioè l'applicazione identità. In questo caso, potendo visualizzare l'effetto della trasformazione lineare, possiamo più facilmente comprendere le potenze dell'applicazione che non le corrispondenti potenze di matrici.



Sia \mathcal{R} la rotazione del piano cartesiano \mathbb{R}^2 di un angolo retto in senso antiorario ed \mathcal{S} la riflessione ortogonale che ha per asse la bisettrice del primo e terzo quadrante. Le matrici che rappresentano \mathcal{R} ed \mathcal{S} sono, rispettivamente

$$\mathbf{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad \text{e} \quad \mathbf{B} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

il che equivale a dire che

$$\mathcal{R} \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} -y \\ x \end{bmatrix} \quad \text{ed} \quad \mathcal{S} \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} y \\ x \end{bmatrix}$$

L'applicazione composta $\mathcal{R} \circ \mathcal{S}$ si ottiene riflettendo prima rispetto alla bisettrice e ruotando poi di un angolo retto; secondo la proposizione precedente si tratta dell'applicazione rappresentata dalla matrice prodotto

$$\mathbf{AB} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

quindi

$$\mathcal{R} \circ \mathcal{S} \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -x \\ y \end{bmatrix}$$

Questo mostra che $\mathcal{R} \circ \mathcal{S}$ è la riflessione ortogonale che ha per asse l'asse delle ordinate. Possiamo comporre le due funzioni in ordine opposto ottenendo $\mathcal{S} \circ \mathcal{R}$, che è l'applicazione rappresentata dalla matrice $\mathbf{BA} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Quindi:

$$\mathcal{S} \circ \mathcal{R} \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ -y \end{bmatrix}$$

Concludiamo che $\mathcal{S} \circ \mathcal{R}$ è la riflessione ortogonale che ha per asse l'asse delle ascisse. Vediamo così che, esattamente come il prodotto di matrici, il prodotto di composizione di applicazioni non è commutativo. Con un vantaggio: abbiamo ora un'intuizione geometrica del perché l'operazione non sia commutativa.

Isomorfismi

Avendo ora a disposizione un prodotto di applicazioni, possiamo chiederci se ci sia un'identità rispetto a tale prodotto. La risposta è affermativa: per ogni spazio vettoriale \mathbf{V} , consideriamo la funzione $\mathcal{I}_{\mathbf{V}} : \mathbf{V} \rightarrow \mathbf{V}$ che mappa ogni vettore di \mathbf{V} in se stesso:

$$\mathcal{I}_{\mathbf{V}}(\mathbf{v}) = \mathbf{v} \quad \text{per ogni } \mathbf{v} \in \mathbf{V}$$

La funzione $\mathcal{I}_{\mathbf{V}}$ è lineare e si dice *applicazione identità* dello spazio \mathbf{V} . Se $\mathbf{V} = \mathbb{K}^n$, $\mathcal{I}_{\mathbf{V}}$ è l'applicazione rappresentata dalla matrice identità di ordine n . L'applicazione identità è un'identità per il prodotto di composizione nel senso che per ogni $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ valgono le uguaglianze:

$$\mathcal{L} \circ \mathcal{I}_{\mathbf{V}} = \mathcal{L} \quad \text{e} \quad \mathcal{I}_{\mathbf{W}} \circ \mathcal{L} = \mathcal{L}$$

In teoria degli insiemi, una funzione $F : A \rightarrow B$ si dice *invertibile* se ha un'inversa rispetto al prodotto di composizione. Questo significa che esiste una funzione $F^{-1} : B \rightarrow A$, detta *funzione inversa* di F , tale che

$$F^{-1} \circ F = \mathcal{I}_A \quad \text{e} \quad F \circ F^{-1} = \mathcal{I}_B$$

Meno formalmente, la condizione $F^{-1} \circ F = \mathcal{I}_A$ significa che

$$F^{-1}(F(a)) = a \quad \text{per ogni } a \in A$$

mentre $F \circ F^{-1} = \mathcal{I}_B$ significa che

$$F(F^{-1}(b)) = b \quad \text{per ogni } b \in B$$

Insieme queste condizioni significano che $b = F(a)$ se e solo se $a = F^{-1}(b)$.

Una funzione è invertibile se e solo se è iniettiva e suriettiva. Supponiamo infatti che F sia iniettiva e suriettiva. Allora per ogni elemento b di B , esiste un unico elemento $a \in A$ tale che $b = F(a)$, e possiamo definire $F^{-1}(b) = a$. Otteniamo così una funzione $F^{-1} : B \rightarrow A$ che soddisfa $F(F^{-1}(b)) = b$ per ogni $b \in B$, quindi $F \circ F^{-1} = \mathcal{I}_B$. Dato un elemento $a \in A$, poniamo $b = F(a)$. Allora $a = F^{-1}(b)$ e quindi $a = F^{-1}(F(a))$. Questo mostra che $F^{-1} \circ F = \mathcal{I}_A$, per cui F è invertibile. Viceversa, supponiamo che F sia invertibile, e mostriamo che F è iniettiva e suriettiva. Se $F(a_1) = F(a_2)$,

applicando F^{-1} a entrambi i membri troviamo $a_1 = a_2$, quindi F è iniettiva. Se $b \in B$, allora, posto $a = F^{-1}(b)$, abbiamo $F(a) = F(F^{-1}(b)) = b$ quindi F è suriettiva. Mostriamo ora che l'inversa di un'applicazione lineare invertibile è lineare:

PROPOSIZIONE 4.3 Sia $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare iniettiva e suriettiva. Allora la funzione inversa $\mathfrak{L}^{-1} : \mathbf{W} \rightarrow \mathbf{V}$ è lineare.

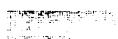
DIMOSTRAZIONE. Dati \mathbf{w}_1 e \mathbf{w}_2 in \mathbf{W} e t_1 e t_2 in \mathbb{K} , occorre verificare che

$$(4.1) \quad \mathfrak{L}^{-1}(t_1\mathbf{w}_1 + t_2\mathbf{w}_2) = t_1\mathfrak{L}^{-1}(\mathbf{w}_1) + t_2\mathfrak{L}^{-1}(\mathbf{w}_2)$$

Poniamo $\mathbf{v}_1 = \mathfrak{L}^{-1}(\mathbf{w}_1)$ e $\mathbf{v}_2 = \mathfrak{L}^{-1}(\mathbf{w}_2)$, in modo che $\mathbf{w}_1 = \mathfrak{L}(\mathbf{v}_1)$ e $\mathbf{w}_2 = \mathfrak{L}(\mathbf{v}_2)$. Il secondo membro della (4.1) diviene $t_1\mathbf{v}_1 + t_2\mathbf{v}_2$, mentre il primo membro è

$$\mathfrak{L}^{-1}(t_1\mathbf{w}_1 + t_2\mathbf{w}_2) = \mathfrak{L}^{-1}(t_1\mathfrak{L}(\mathbf{v}_1) + t_2\mathfrak{L}(\mathbf{v}_2)) = \mathfrak{L}^{-1}(\mathfrak{L}(t_1\mathbf{v}_1 + t_2\mathbf{v}_2)) = t_1\mathbf{v}_1 + t_2\mathbf{v}_2$$

(abbiamo usato la linearità di \mathfrak{L} nel secondo passaggio). Quindi la (4.1) è verificata e \mathfrak{L}^{-1} è lineare.



Supponiamo che \mathbf{A} sia una matrice quadrata di ordine n invertibile. Allora l'applicazione $\mathfrak{L}_{\mathbf{A}} : \mathbb{K}^n \rightarrow \mathbb{K}^n$ rappresentata da \mathbf{A} è invertibile e la sua inversa è l'applicazione $\mathfrak{L}_{\mathbf{A}^{-1}}$ rappresentata da \mathbf{A}^{-1} . Infatti

$$\mathfrak{L}_{\mathbf{A}^{-1}} \circ \mathfrak{L}_{\mathbf{A}}(\mathbf{x}) = \mathbf{A}^{-1}\mathbf{Ax} = \mathbf{x} \quad \text{per ogni } \mathbf{x} \in \mathbb{K}^n$$

$$\mathfrak{L}_{\mathbf{A}} \circ \mathfrak{L}_{\mathbf{A}^{-1}}(\mathbf{y}) = \mathbf{AA}^{-1}\mathbf{y} = \mathbf{y} \quad \text{per ogni } \mathbf{y} \in \mathbb{K}^n$$

Quando una funzione $F : A \rightarrow B$ tra due insiemi è invertibile, si dice che i due insiemi sono in corrispondenza biunivoca. I due insiemi si possono identificare: un elemento a di A si identifica con l'elemento $f(a)$ di B , e viceversa un elemento b di B si identifica con $f^{-1}(b)$. Per esempio, tutti gli insiemi B costituiti da due elementi sono tra loro in corrispondenza biunivoca: basta scegliere qual è il primo dei due elementi di B e chiamarlo $f(1)$, e qual è il secondo e chiamarlo $f(2)$; si definisce così una funzione invertibile $f : \{1, 2\} \rightarrow B$. La differenza tra i due insiemi è che $\{1, 2\}$ è un insieme di numeri, mentre non abbiamo fatto ipotesi sulla natura degli elementi di B , che potrebbero essere vettori, persone o qualunque altra cosa. Dal punto di vista, però, delle proprietà degli insiemi, $\{1, 2\}$ e B sono indistinguibili e possono essere identificati.

Analogamente, se \mathbf{V} e \mathbf{W} sono spazi vettoriali ed $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{W}$ è lineare e invertibile, \mathfrak{L} mette in corrispondenza biunivoca i vettori di \mathbf{V} con quelli di \mathbf{W} . Come insiemi possiamo quindi identificare \mathbf{V} e \mathbf{W} non solo come insiemi, ma anche come spazi vettoriali: infatti \mathfrak{L} è lineare, il che significa che la somma di due vettori in \mathbf{V} corrisponde alla somma dei vettori corrispondenti in \mathbf{W} e il prodotto di uno scalare t per un vettore $\mathbf{v} \in \mathbf{V}$ corrisponde al prodotto di t per il vettore $\mathbf{w} = \mathfrak{L}(\mathbf{v})$. Si tratta quindi di una corrispondenza biunivoca che in più rispetta le due operazioni di somma

e prodotto per uno scalare che caratterizzano gli spazi vettoriali: in quanto spazi vettoriali \mathbf{V} e \mathbf{W} sono identificati da \mathfrak{L} . Un'applicazione lineare invertibile $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{W}$ è detta anche *isomorfismo*, perché in algebra un isomorfismo è una funzione invertibile che rispetta le operazioni di una particolare struttura algebrica. Per la proposizione 4.3, l'inversa di un isomorfismo è un isomorfismo. Diremo che due spazi vettoriali \mathbf{V} e \mathbf{W} su \mathbb{K} sono *isomorfi* se esiste un isomorfismo $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{W}$. Come spazi vettoriali, due spazi isomorfi sono identici.

Come ogni insieme finito formato da n elementi può essere messo in corrispondenza biunivoca con l'insieme dei primi n numeri interi, così tutti gli spazi vettoriali su \mathbb{K} di dimensione n sono isomorfi a \mathbb{K}^n : l'identificazione con \mathbb{K}^n si ottiene scegliendo una base. Quindi, a meno di isomorfismi, \mathbb{K}^n è l'unico spazio vettoriale su \mathbb{K} di dimensione n .

PROPOSIZIONE 4.4 Sia \mathbf{V} uno spazio vettoriale su \mathbb{K} di dimensione n e sia $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ una base di \mathbf{V} . La funzione $\mathcal{P} : \mathbb{K}^n \rightarrow \mathbf{V}$ definita da

$$\mathcal{P}([x_1, x_2, \dots, x_n]^T) = x_1 \mathbf{b}_1 + x_2 \mathbf{b}_2 + \cdots + x_n \mathbf{b}_n$$

è un isomorfismo. La funzione inversa \mathcal{P}^{-1} , che a un vettore \mathbf{v} di \mathbf{V} associa la n -upla $\mathbf{x}(\mathbf{v})$ delle sue coordinate rispetto alla base fissata, è perciò anch'essa lineare.

DIMOSTRAZIONE. Per definizione $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ è una base se e solo se \mathcal{P} è iniettiva e suriettiva. La verifica che \mathcal{P} sia lineare è immediata e il lettore la può ritrovare nel capitolo sugli spazi vettoriali nella dimostrazione del lemma fondamentale. Quindi \mathcal{P} è un isomorfismo. □

COROLLARIO 4.5 (Criterio di isomorfismo)

Ogni spazio vettoriale su \mathbb{K} di dimensione n è isomorfo a \mathbb{K}^n . Due spazi vettoriali di dimensione finita su \mathbb{K} sono isomorfi se e solo se hanno la stessa dimensione.

DIMOSTRAZIONE. Sia \mathbf{V} uno spazio vettoriale di dimensione n . Per la proposizione precedente è isomorfo a \mathbb{K}^n . Dati due spazi vettoriali \mathbf{V}_1 e \mathbf{V}_2 di dimensione n , esistono isomorfismi $\mathcal{P}_1 : \mathbb{K}^n \rightarrow \mathbf{V}_1$ e $\mathcal{P}_2 : \mathbb{K}^n \rightarrow \mathbf{V}_2$. Allora $\mathcal{P}_2 \circ \mathcal{P}_1^{-1} : \mathbf{V}_1 \rightarrow \mathbf{V}_2$ è un isomorfismo, per cui \mathbf{V}_1 e \mathbf{V}_2 sono isomorfi (stiamo sostanzialmente dimostrando che la relazione di isomorfismo è una relazione di equivalenza). Viceversa, supponiamo che $\mathfrak{L} : \mathbf{V}_1 \rightarrow \mathbf{V}_2$ sia un isomorfismo di spazi vettoriali e che \mathbf{V}_1 abbia dimensione finita. Nel capitolo sugli spazi vettoriali abbiamo dimostrato che esiste una base di \mathbf{V}_1 e che ogni base di \mathbf{V}_1 ha $n = \dim \mathbf{V}_1$ elementi. Siccome \mathfrak{L} è un isomorfismo, è immediato mostrare che \mathfrak{L} mappa una base di \mathbf{V}_1 in una base di \mathbf{V}_2 (perché nelle notazioni di prima $\mathcal{P}_2 = \mathfrak{L} \circ \mathcal{P}_1$ è iniettiva e suriettiva). Quindi anche \mathbf{V}_2 ha dimensione n . □

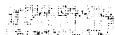
L'insieme $M_{\mathbb{K}}(2, 2)$ delle matrici di tipo $(2, 2)$ a coefficienti in \mathbb{K} è uno spazio vettoriale di dimensione 4. Fissata la base canonica formata dalle quattro matrici

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

l'applicazione \mathcal{P}^{-1} che identifica $M_{\mathbb{K}}(2, 2)$ e \mathbb{K}^4 è

$$\mathcal{P}^{-1} \left(\begin{bmatrix} t_1 & t_2 \\ t_3 & t_4 \end{bmatrix} \right) = [t_1, t_2, t_3, t_4]^T$$

Gli elementi di questi due insiemi sono diversi tra loro, matrici in un caso e vettori nell'altro, ma come spazi vettoriali i due insiemi sono identici.



Se $m \neq n$, lo spazio vettoriale \mathbb{K}^m non è isomorfo a \mathbb{K}^n : infatti

$$\dim(\mathbb{K}^m) = m \neq n = \dim(\mathbb{K}^n)$$

mentre due spazi isomorfi hanno la stessa dimensione.



Descrivere geometricamente le applicazioni lineari $\mathbb{R}^2 \rightarrow \mathbb{R}^2$ rappresentate dalle matrici

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Calcolare il quadrato delle tre matrici. Come si spiega geometricamente il risultato?

In questo esercizio si completa la dimostrazione che $\text{Hom}_{\mathbb{K}}(\mathbf{V}, \mathbf{W})$ è uno spazio vettoriale.

- a) Verificare che la somma di due applicazioni lineari è un'applicazione lineare.
- b) Verificare che il prodotto di un'applicazione lineare per uno scalare è un'applicazione lineare.
- c) Mostrare che le operazioni di somma e prodotto per scalare definite in $\text{Hom}_{\mathbb{K}}(\mathbf{V}, \mathbf{W})$ soddisfano gli assiomi di spazio vettoriale.
- d) Sia \mathcal{F} la funzione che a una matrice \mathbf{A} di tipo (m, n) a coefficienti in \mathbb{K} associa la funzione $\mathcal{L}_{\mathbf{A}}$. Mostrare che $\mathcal{F} : M_{\mathbb{K}}(m, n) \rightarrow \text{Hom}_{\mathbb{K}}(\mathbb{K}^n, \mathbb{K}^m)$ è lincare. Meno formalmente, questo significa che la somma delle applicazioni rappresentate da due matrici è l'applicazione rappresentata dalla somma delle due matrici:

$$\mathcal{L}_{\mathbf{A}_1} + \mathcal{L}_{\mathbf{A}_2} = \mathcal{L}_{\mathbf{A}_1 + \mathbf{A}_2}$$

e analogamente

$$\mathcal{L}_{t\mathbf{A}} = t\mathcal{L}_{\mathbf{A}}$$

Forme lineari e spazio duale Una *forma lineare* su uno spazio vettoriale \mathbf{V} è una funzione lineare $\alpha : \mathbf{V} \rightarrow \mathbb{K}$. Lo *spazio duale* di \mathbf{V} è lo spazio vettoriale $\mathbf{V}^* = \text{Hom}_{\mathbb{K}}(\mathbf{V}, \mathbb{K})$ delle forme lineari su \mathbf{V} . Sia $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ una base di \mathbf{V} . Fissato un indice $k = 1, 2, \dots, n$, sia $\beta_k : \mathbf{V} \rightarrow \mathbb{K}$ la funzione che a un vettore \mathbf{v} di \mathbf{V} associa la sua k -esima coordinata:

$$\beta_k(x_1\mathbf{b}_1 + \dots + x_n\mathbf{b}_n) = x_k$$

- a) Mostrare che β_k è una forma lineare.

b) Se $\mathbf{V} = \mathbb{K}^n$ e $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ è la base canonica, mostrare che β_k è l'applicazione lineare rappresentata dal vettore riga che ha la k -esima componente uguale a 1 e le altre componenti nulle.

c) Definire un isomorfismo tra lo spazio dei vettori riga $M_{\mathbb{K}}(1, n)$ e lo spazio duale di \mathbb{K}^n .

Suggerimento: un vettore riga è una matrice e quindi rappresenta, per moltiplicazione, un'applicazione lineare.

Mostrare che:

a) se \mathbf{A} è una matrice quadrata e $\mathcal{L} = \mathcal{L}_{\mathbf{A}}$, allora $\mathcal{L}^n = \mathcal{L}_{\mathbf{A}^n}$;

b) più in generale, dati un polinomio $P(x) = P(x) = a_n x^n + \dots + a_1 x + a_0$ e una matrice quadrata \mathbf{A} , si definisca

$$P(\mathbf{A}) = a_n \mathbf{A}^n + \dots + a_1 \mathbf{A} + a_0 \mathbf{I}$$

e si mostri che $P(\mathcal{L}_{\mathbf{A}}) = \mathcal{L}_{P(\mathbf{A})}$;

c) se $\mathcal{R} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ è la rotazione di un angolo θ in senso antiorario, $\mathcal{R}^n : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ è la rotazione di un angolo $n\theta$ in senso antiorario, e \mathcal{R}^{-1} è la rotazione di un angolo θ in senso orario.

Sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare, e sia $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ una base di \mathbf{V} . Mostrare che \mathcal{L} è un isomorfismo se e solo se $\{\mathcal{L}(\mathbf{b}_1), \dots, \mathcal{L}(\mathbf{b}_n)\}$ è una base di \mathbf{W} .

Verificare che:

a) se $\mathcal{L}_1 : \mathbf{V} \rightarrow \mathbf{W}$ e $\mathcal{L}_2 : \mathbf{W} \rightarrow \mathbf{Z}$ sono isomorfismi, anche $\mathcal{L}_2 \circ \mathcal{L}_1 : \mathbf{V} \rightarrow \mathbf{Z}$ è un isomorfismo, e

$$(\mathcal{L}_2 \circ \mathcal{L}_1)^{-1} = \mathcal{L}_1^{-1} \circ \mathcal{L}_2^{-1}$$

b) la relazione di isomorfismo è una relazione di equivalenza;

c) se \mathbf{A} è una matrice quadrata, l'applicazione $\mathcal{L}_{\mathbf{A}}$ è un isomorfismo se e solo se \mathbf{A} è invertibile;

d) se $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ è un isomorfismo, un insieme di vettori $\mathbf{v}_1, \dots, \mathbf{v}_d$ di \mathbf{V} è una base di \mathbf{V} (rispettivamente, un insieme di generatori o un insieme linearmente indipendente) se e solo se $\{\mathcal{L}(\mathbf{v}_1), \dots, \mathcal{L}(\mathbf{v}_d)\}$ formano una base di \mathbf{W} (rispettivamente, un insieme di generatori o un insieme linearmente indipendente). In particolare, \mathbf{V} e \mathbf{W} hanno la stessa dimensione, e più in generale, ogni sottospazio \mathbf{H} di \mathbf{V} ha la stessa dimensione di $\mathcal{L}(\mathbf{H})$.

Mostrare che lo spazio vettoriale dei polinomi $\mathbb{K}[x]$ è isomorfo allo spazio vettoriale delle successioni *finite* di elementi di \mathbb{K} (una successione $\{a_k\}_{k \geq 0}$ è finita se esiste $N \geq 0$ tale che $a_k = 0$ per ogni $k > N$).

In algebra astratta, una \mathbb{K} -algebra è uno spazio vettoriale su \mathbb{K} in cui è definita una operazione *prodotto* che gode delle proprietà associative, distributiva (a destra e sinistra) e di omogeneità che abbiamo definito nel paragrafo sul prodotto di matrici. Mostrare che $\text{End}_{\mathbb{K}}(\mathbf{V})$, col prodotto definito dal prodotto di composizione, è una \mathbb{K} -algebra.

5 IL TEOREMA DI RAPPRESENTAZIONE

In questo paragrafo illustriamo e dimostriamo il teorema di rappresentazione, che afferma che una funzione $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ tra spazi vettoriali di dimensione finita è lineare se e solo se le coordinate $\mathbf{y} = [y_1, \dots, y_m]^T$ di $\mathcal{L}(\mathbf{v})$ (rispetto a una qualunque base di \mathbf{W}) sono polinomi omogenei di primo grado nelle coordinate $\mathbf{x} = [x_1, \dots, x_n]^T$ di \mathbf{v} (rispetto a una qualunque base di \mathbf{V}): questo equivale a dire, come abbiamo visto nel paragrafo 2, che esiste una matrice \mathbf{A} tale che

$$(5.1) \quad \mathbf{y} = \mathbf{Ax}$$

Quindi $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ è lineare se e solo se la sua espressione in coordinate è una funzione della forma $\mathcal{L}_{\mathbf{A}}$. Questo è fondamentale perché se da un lato la rappresentazione in coordinate è necessaria per fare i conti, dall'altro la forma astratta \mathcal{L} consente di poter scegliere le coordinate e quindi la matrice \mathbf{A} a posteriori: più semplice sarà \mathbf{A} , più semplice sarà eseguire i conti.

Il teorema di rappresentazione si fonda sull'osservazione che un'applicazione lineare $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ è determinata dalle immagini dei vettori di una base di \mathbf{V} ; inoltre tali immagini possono essere scelte arbitrariamente in \mathbf{W} :

PROPOSIZIONE 5.1 Sia \mathbf{V} uno spazio vettoriale di dimensione n , sia $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ una base di \mathbf{V} , e siano $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n$ vettori arbitrari di uno spazio vettoriale \mathbf{W} . Allora esiste una e una sola applicazione lineare $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ tale che

$$\mathcal{L}(\mathbf{b}_1) = \mathbf{w}_1, \mathcal{L}(\mathbf{b}_2) = \mathbf{w}_2, \dots, \mathcal{L}(\mathbf{b}_n) = \mathbf{w}_n$$

L'applicazione \mathcal{L} è definita dalla formula:

$$(5.2) \quad \mathcal{L}(x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \dots + x_n\mathbf{b}_n) = x_1\mathbf{w}_1 + x_2\mathbf{w}_2 + \dots + x_n\mathbf{w}_n$$

DIMOSTRAZIONE. Supponiamo che esista un'applicazione \mathcal{L} con la proprietà desiderata. Dato un vettore $\mathbf{v} \in \mathbf{V}$, siano x_1, x_2, \dots, x_n le sue coordinate rispetto alla base $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$, in modo che:

$$\mathbf{v} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \dots + x_n\mathbf{b}_n$$

Allora per la linearità di \mathcal{L} si ha

$$(5.3) \quad \begin{aligned} \mathcal{L}(\mathbf{v}) &= \mathcal{L}(x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \dots + x_n\mathbf{b}_n) = x_1\mathcal{L}(\mathbf{b}_1) + x_2\mathcal{L}(\mathbf{b}_2) + \dots + x_n\mathcal{L}(\mathbf{b}_n) = \\ &= x_1\mathbf{w}_1 + x_2\mathbf{w}_2 + \dots + x_n\mathbf{w}_n \end{aligned}$$

Quindi, se \mathcal{L} esiste, è determinata dalla formula (5.2)

Questo mostra l'unicità di \mathcal{L} . Per verificare l'esistenza, definiamo la funzione $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ mediante la formula (5.2): si osservi che \mathcal{L} è ben definita perché le coordinate x_1, \dots, x_n di un vettore sono univocamente determinate. Allora evidentemente $\mathcal{L}(\mathbf{b}_k) = \mathbf{w}_k$ per ogni $k = 1, 2, \dots, n$. Occorre ancora mostrare che \mathcal{L} è lineare: questo è un esercizio che abbiamo già svolto nella dimostrazione del lemma fondamentale 6.5 nel capitolo sugli spazi vettoriali.

Sia $\{\mathbf{e}_1, \mathbf{e}_2\}$ la base canonica di \mathbb{R}^2 , e sia $\mathcal{L} : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ l'applicazione lineare tale che $\mathcal{L}(\mathbf{e}_1) = [1, 2, 3]^T$ e $\mathcal{L}(\mathbf{e}_2) = [3, 2, 1]^T$. Allora l'immagine di un arbitrario vettore $[x, y]^T$ è

$$\mathcal{L}([x, y]^T) = \mathcal{L}(x\mathbf{e}_1 + y\mathbf{e}_2) = x[1, 2, 3]^T + y[3, 2, 1]^T = [x + 3y, 2x + 2y, 3x + y]^T$$

Nelle notazioni della proposizione 5.1, se $\mathbf{w}_1 = \mathbf{w}_2 = \dots = \mathbf{w}_n = \mathbf{0}_{\mathbf{W}}$, allora \mathcal{L} è l'applicazione nulla. All'estremo opposto, se $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n$ è una base di \mathbf{W} , allora \mathcal{L} mappa una base di \mathbf{V} in una base di \mathbf{W} ed è perciò un isomorfismo.

Data una matrice \mathbf{A} di tipo (m, n) , siano $\mathbf{w}_1, \dots, \mathbf{w}_n$ le colonne di \mathbf{A} e sia $\mathcal{L}_{\mathbf{A}} : \mathbb{K}^n \rightarrow \mathbb{K}^m$ la funzione definita $\mathcal{L}_{\mathbf{A}}(\mathbf{x}) = \mathbf{Ax}$. Tale funzione è lineare e mappa il k -esimo vettore \mathbf{e}_k della base canonica di \mathbb{K}^n nella k -esima colonna \mathbf{w}_k di \mathbf{A} . Per la proposizione 5.1, la funzione $\mathcal{L}_{\mathbf{A}}$ è quindi l'unica applicazione lineare da \mathbb{K}^n a \mathbb{K}^m che mappi \mathbf{e}_k in \mathbf{w}_k per ogni k .

Sia \mathbf{V} uno spazio vettoriale su \mathbb{K} , e siano $\mathbf{v}_1, \dots, \mathbf{v}_d$ vettori di \mathbf{V} . Abbiamo già considerato più volte la funzione $\mathcal{P} : \mathbb{K}^d \rightarrow \mathbf{V}$ definita da

$$\mathcal{P}([t_1, t_2, \dots, t_d]^T) = t_1 \mathbf{v}_1 + t_2 \mathbf{v}_2 + \dots + t_d \mathbf{v}_d$$

Si tratta dell'applicazione lineare che al k -esimo vettore \mathbf{e}_k della base canonica di \mathbb{K}^d associa \mathbf{v}_k poiché

$$[t_1, t_2, \dots, t_d]^T = t_1 \mathbf{e}_1 + t_2 \mathbf{e}_2 + \dots + t_d \mathbf{e}_d$$

Sia $\mathcal{R} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ la rotazione di un angolo θ in senso antiorario. Ricaviamo nuovamente l'espressione in coordinate di \mathcal{R} utilizzando questa volta (5.2). Per questo occorre semplicemente conoscere poiché l'effetto della rotazione sui due vettori $\mathbf{e}_1 = [1, 0]^T$ e $\mathbf{e}_2 = [0, 1]^T$ della base canonica, \mathcal{R} ruota i vettori del piano di un angolo θ in senso antiorario,

$$\mathcal{R}(\mathbf{e}_1) = \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix} \quad \text{e} \quad \mathcal{R}(\mathbf{e}_2) = \begin{bmatrix} -\sin(\theta) \\ \cos(\theta) \end{bmatrix}$$

Quindi da (5.2) ricaviamo

$$\mathcal{R}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = x\mathcal{R}(\mathbf{e}_1) + y\mathcal{R}(\mathbf{e}_2) = \begin{bmatrix} \cos(\theta)x - \sin(\theta)y \\ \sin(\theta)x + \cos(\theta)y \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Come applicazione, da questa espressione è possibile ricavare le formule di addizione per il seno e coseno invertendo il ragionamento fatto a pagina 222.

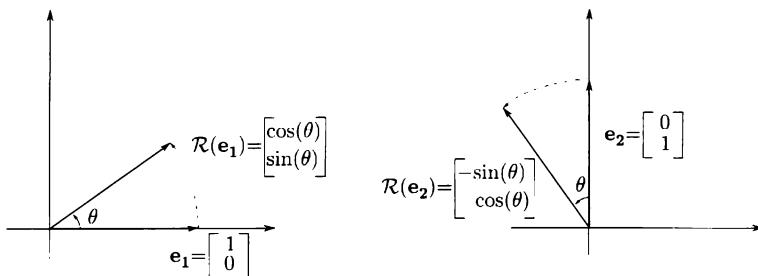


Figura 5.1. Rotazione dei vettori della base canonica.

Sia \mathcal{R} la rotazione del piano cartesiano \mathbb{R}^2 di un angolo retto in senso antiorario, e sia \mathcal{S} la riflessione ortogonale che ha per asse la bisettrice del primo e terzo quadrante: si veda l'esempio a pagina 234. L'applicazione lineare composta $\mathcal{L} = \mathcal{S} \circ \mathcal{R}$ è determinata una volta che si conoscano le immagini $\mathcal{L}(e_1)$ e $\mathcal{L}(e_2)$ dei due vettori della base canonica. Osserviamo che \mathcal{R} mappa e_1 in e_2 e che \mathcal{S} mappa e_2 in e_1 . Quindi \mathcal{L} mappa e_1 in se stesso. D'altra parte, \mathcal{R} mappa e_2 in $-e_1$ ed \mathcal{S} mappa $-e_1$ in $-e_2$. Quindi \mathcal{L} mappa e_2 in $-e_2$. Concludiamo che \mathcal{L} è l'applicazione lineare che mappa e_1 in se stesso e e_2 nel suo opposto $-e_2$. Siccome anche la riflessione ortogonale che ha per asse l'asse delle ascisse mappa e_1 in se stesso ed e_2 nel suo opposto, ritroviamo il fatto che \mathcal{L} coincide con tale riflessione. Un analogo ragionamento mostra che $\mathcal{R} \circ \mathcal{S}$ è la riflessione ortogonale che ha per asse l'asse delle ordinate.

Diamo ora l'enunciato formale e una dimostrazione completa del teorema di rappresentazione:

TEOREMA 5.2 (Teorema di rappresentazione)

Sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare tra spazi vettoriali di dimensione finita.

Supponiamo che \mathcal{B} sia una base di \mathbf{V} e \mathcal{C} sia una base di \mathbf{W} . Esiste un'unica matrice \mathbf{A} a coefficienti in \mathbb{K} con $m = \dim(\mathbf{W})$ righe ed $n = \dim(\mathbf{V})$ colonne con la seguente proprietà: per ogni $\mathbf{v} \in \mathbf{V}$, se \mathbf{x} è il vettore delle coordinate di \mathbf{v} rispetto alla base \mathcal{B} e \mathbf{y} è il vettore delle coordinate di $\mathcal{L}(\mathbf{v})$ rispetto alla base \mathcal{C} , allora

$$(5.4) \quad \mathbf{y} = \mathbf{Ax}$$

Si dice che la matrice \mathbf{A} rappresenta l'applicazione lineare \mathcal{L} rispetto alle basi \mathcal{B} di \mathbf{V} e \mathcal{C} di \mathbf{W} .

DIMOSTRAZIONE. Sia $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ e sia $\mathcal{P} : \mathbb{K}^n \rightarrow \mathbf{V}$ l'isomorfismo che manda il k -esimo vettore \mathbf{e}_k della base canonica in \mathbf{b}_k :

$$\mathcal{P} \left([x_1, \dots, x_n]^T \right) = x_1 \mathbf{b}_1 + \cdots + x_n \mathbf{b}_n$$

L'applicazione inversa $\mathbf{x} = \mathcal{P}^{-1} : \mathbf{V} \rightarrow \mathbb{K}^n$ associa a un vettore \mathbf{v} di \mathbf{V} la n -upla $\mathbf{x}(\mathbf{v})$ delle sue coordinate rispetto alla base \mathcal{B} . Analogamente la base \mathcal{C} definisce un isomorfismo $\mathcal{Q} : \mathbb{K}^m \rightarrow \mathbf{W}$, e la funzione inversa $\mathbf{y} : \mathbf{W} \rightarrow \mathbb{K}^m$ associa a un vettore \mathbf{w} di \mathbf{W} la m -upla $\mathbf{y}(\mathbf{w})$ delle sue coordinate rispetto alla base \mathcal{C} . La tesi da dimostrare è che esiste una e una sola matrice \mathbf{A} tale che

$$\mathbf{y}(\mathcal{L}(\mathbf{v})) = \mathbf{Ax}(\mathbf{v}) \quad \text{per ogni } \mathbf{v} \in \mathbf{V}$$

Sostituendo $\mathbf{v} = \mathcal{P}(\mathbf{x})$ otteniamo l'espressione equivalente

$$\mathbf{y}(\mathcal{L}(\mathcal{P}(\mathbf{x}))) = \mathbf{Ax} \quad \text{per ogni } \mathbf{x} \in \mathbb{K}^n$$

L'applicazione $\mathcal{L}_c = \mathbf{y} \circ \mathcal{L} \circ \mathcal{P} : \mathbb{K}^n \rightarrow \mathbb{K}^m$ che compare a primo membro è l'*espressione in coordinate* di \mathcal{L} ; è lineare perché composta di applicazioni lineari. In sostanza: se usando le coordinate si identificano \mathbf{V} con \mathbb{K}^n e \mathbf{W} con \mathbb{K}^m , l'applicazione \mathcal{L}_c si identifica allora con \mathcal{L} , come illustrato dal diagramma della figura 5.2, in cui le frecce verticali sono isomorfismi.

$$\begin{array}{ccc} \mathbf{V} & \xrightarrow{\mathcal{L}} & \mathbf{W} \\ \mathcal{P} \uparrow & & \downarrow \mathbf{y} \\ \mathbb{K}^n & \xrightarrow{\mathcal{L}_c = \mathbf{y} \circ \mathcal{L} \circ \mathcal{P}} & \mathbb{K}^m \end{array}$$

Figura 5.2. L'espressione in coordinate di un'applicazione lineare.

Per finire, occorre mostrare che, se $\mathcal{L}_c : \mathbb{K}^n \rightarrow \mathbb{K}^m$ è lineare, esiste un'unica matrice \mathbf{A} di tipo (m, n) tale che

$$(5.5) \quad \mathcal{L}_c(\mathbf{x}) = \mathbf{Ax} \quad \text{per ogni } \mathbf{x} \in \mathbb{K}^n$$

Sia $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ la base canonica di \mathbb{K}^n . Per ogni $k = 1, 2, \dots, n$ sia $\mathbf{y}_k = \mathcal{L}_c(\mathbf{e}_k)$ l'immagine mediante \mathcal{L}_c di \mathbf{e}_k . Per la proposizione 5.1, \mathcal{L}_c è l'unica applicazione lineare da \mathbb{K}^n a \mathbb{K}^m che mandi \mathbf{e}_k in \mathbf{y}_k per ogni k . D'altro canto, data una matrice \mathbf{A} , la funzione $\mathcal{L}_A(\mathbf{x}) = \mathbf{Ax}$ è l'unica applicazione lineare da \mathbb{K}^n a \mathbb{K}^m che mandi \mathbf{e}_k nella k -esima colonna di \mathbf{A} per ogni k . Quindi (5.5) è soddisfatta se e solo se

$$\mathbf{A} = [\mathbf{y}_1 \ \mathbf{y}_2 \ \cdots \ \mathbf{y}_n]$$

è la matrice che ha per colonne i vettori \mathbf{y}_k . Questo conclude la dimostrazione. Vale la pena di ripetere l'argomento della dimostrazione della proposizione 5.1 e mostrare direttamente che \mathbf{A} verifica la (5.5):

$$\begin{aligned} \mathcal{L}_c(\mathbf{x}) &= \mathcal{L}_c(x_1 \mathbf{e}_1 + \cdots + x_n \mathbf{e}_n) = x_1 \mathcal{L}_c(\mathbf{e}_1) + \cdots + x_n \mathcal{L}_c(\mathbf{e}_n) = \\ &= x_1 \mathbf{y}_1 + \cdots + x_n \mathbf{y}_n = \mathbf{Ax} \end{aligned}$$



OSSERVAZIONE Dal teorema di rappresentazione segue che *ogni* funzione lineare $\mathcal{L} : \mathbb{K}^n \rightarrow \mathbb{K}^m$ è della forma \mathcal{L}_A , per un'unica matrice \mathbf{A} di tipo (m, n) : basta porre nell'enunciato $\mathbf{V} = \mathbb{K}^n$, $\mathbf{W} = \mathbb{K}^m$ e prendere le coordinate rispetto alle basi canoniche. In particolare, \mathbf{A} è la matrice che rappresenta \mathcal{L}_A rispetto alle basi canoniche e la terminologia è coerente con quella introdotta in precedenza: quando si dice che \mathbf{A} rappresenta \mathcal{L} senza specificare rispetto a quali basi, si intende rispetto alle basi canoniche.

OSSERVAZIONE Il caso più importante è quello in cui il dominio e il codominio coincidono: $\mathbf{V} = \mathbf{W}$. In questo caso di norma si sceglie $\mathcal{B} = \mathcal{C}$, per avere un unico sistema di coordinate su \mathbf{V} e si dice che \mathbf{A} rappresenta \mathfrak{L} rispetto alla base \mathcal{B} .

OSSERVAZIONE Nelle applicazioni è importante saper costruire la matrice rappresentativa \mathbf{A} . Per questo osserviamo che la k -esima colonna di \mathbf{A} è, nelle notazioni della dimostrazione del teorema,

$$\mathbf{y}_k = \mathfrak{L}_c(\mathbf{e}_k) = \mathbf{y}(\mathfrak{L}(\mathcal{P}(\mathbf{e}_k))) = \mathbf{y}(\mathfrak{L}(\mathbf{b}_k))$$

Quindi la k -esima colonna di \mathbf{A} è il vettore delle coordinate di $\mathfrak{L}(\mathbf{b}_k)$ rispetto alla base \mathcal{C} . La regola per scrivere \mathbf{A} è pertanto:

La k -esima colonna della matrice \mathbf{A} è il vettore \mathbf{y}_k delle coordinate di $\mathfrak{L}(\mathbf{b}_k)$ rispetto alla base \mathcal{C} . Questo significa: se

$$\mathfrak{L}(\mathbf{b}_k) = t_1 \mathbf{c}_1 + t_2 \mathbf{c}_2 + \cdots + t_m \mathbf{c}_m$$

allora la k -esima colonna della matrice \mathbf{A} è il vettore $[t_1, \dots, t_m]^T$.

OSSERVAZIONE

Un altro modo di scrivere (5.4) è

$$(5.6) \quad \mathfrak{L}(x_1 \mathbf{b}_1 + x_2 \mathbf{b}_2 + \cdots + x_n \mathbf{b}_n) = y_1 \mathbf{c}_1 + y_2 \mathbf{c}_2 + \cdots + y_m \mathbf{c}_m$$

$$\text{con } y_i = \sum_{j=1}^n a_{ij} x_j.$$

Scrivendo

Sia $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ la base canonica di \mathbb{R}^3 e sia \mathfrak{L} la funzione lineare $\mathbb{R}^3 \rightarrow \mathbb{R}^2$ tale che $\mathfrak{L}(\mathbf{e}_1) = \mathbf{0}$, $\mathfrak{L}(\mathbf{e}_2) = [2, 3]^T$ e $\mathfrak{L}(\mathbf{e}_3) = [7, 5]^T$. La matrice che rappresenta \mathfrak{L} rispetto alle basi canoniche di \mathbb{R}^3 e \mathbb{R}^2 è

$$\mathbf{A} = \begin{bmatrix} 0 & 2 & 7 \\ 0 & 3 & 5 \end{bmatrix}$$

Scrivendo

Supponiamo che \mathbf{V} sia uno spazio vettoriale di dimensione 3 con base $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$. Sia $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{V}$ l'applicazione lineare tale che

$$\begin{cases} \mathfrak{L}(\mathbf{b}_1) = 3\mathbf{b}_1 \\ \mathfrak{L}(\mathbf{b}_2) = \mathbf{b}_1 + 3\mathbf{b}_2 \\ \mathfrak{L}(\mathbf{b}_3) = \mathbf{b}_2 + 3\mathbf{b}_3 \end{cases}$$

Sia \mathbf{A} la matrice che rappresenta \mathfrak{L} rispetto alla base \mathcal{B} . La prima colonna di \mathbf{A} è il vettore delle coordinate di

$$\mathfrak{L}(\mathbf{b}_1) = 3\mathbf{b}_1 = 3\mathbf{b}_1 + 0\mathbf{b}_2 + 0\mathbf{b}_3$$

rispetto alla base \mathcal{B} , ed è perciò $[3, 0, 0]^T$. Analogamente si determinano la seconda e la terza colonna di \mathbf{A} :

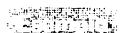
$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix}$$



La matrice identità \mathbf{I} rappresenta l'applicazione identità $\mathcal{I} : \mathbf{V} \rightarrow \mathbf{V}$ rispetto a qualunque base $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ di \mathbf{V} . Questo perché $\mathcal{I}(\mathbf{v}) = \mathbf{v}$ per ogni $\mathbf{v} \in \mathbf{V}$ significa che la matrice rappresentativa \mathbf{A} deve soddisfare

$$\mathbf{Ax} = \mathbf{x} \quad \text{per ogni } \mathbf{x} \in \mathbb{K}^n$$

quindi dev'essere la matrice identità \mathbf{I} .



Consideriamo il sottospazio \mathbf{V} di $C^\infty(\mathbb{R})$ generato dalle funzioni $\cos(t)$ e $\sin(t)$, cioè l'insieme delle funzioni della forma:

$$y(t) = a \cos(t) + b \sin(t)$$

Siccome $\cos(t)$ e $\sin(t)$ sono linearmente indipendenti, $\mathcal{B} = \{\cos(t), \sin(t)\}$ è una base di \mathbf{V} . Il vettore delle coordinate di $y(t) = a \cos(t) + b \sin(t)$ rispetto alla base \mathcal{B} è

$$\mathbf{x} = \begin{bmatrix} a \\ b \end{bmatrix}$$

L'operatore derivata \mathfrak{D} mappa \mathbf{V} in \mathbf{V} perché:

$$\mathfrak{D}(a \cos(t) + b \sin(t)) = b \cos(t) - a \sin(t)$$

La matrice \mathbf{A} che rappresenta $\mathfrak{D} : \mathbf{V} \rightarrow \mathbf{V}$ rispetto alla base \mathcal{B} ha come prima colonna il vettore \mathbf{y}_1 delle coordinate di $\mathfrak{D}(\cos(t)) = -\sin(t)$ e come seconda colonna il vettore \mathbf{y}_2 delle coordinate di $\mathfrak{D}(\sin(t)) = \cos(t)$. Quindi:

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

Questo significa che il vettore delle coordinate della derivata $\mathfrak{D}(y(t))$ si ottiene moltiplicando \mathbf{A} per il vettore delle coordinate di $y(t)$. Verifichiamo:

$$\mathbf{A} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} b \\ -a \end{bmatrix}$$

in accordo col fatto che

$$\mathfrak{D}(a \cos(t) + b \sin(t)) = b \cos(t) - a \sin(t)$$

ha coordinate $\begin{bmatrix} b \\ -a \end{bmatrix}$.



In questo esempio determiniamo la matrice di una riflessione ortogonale \mathcal{S} del piano. Sia $\mathcal{S} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ la riflessione ortogonale che ha per asse una retta \mathbf{r} passante per l'origine. Si

tratta di un'applicazione lineare. Sia α l'angolo acuto che l'asse di riflessione r forma col primo vettore \mathbf{e}_1 della base canonica. La riflessione S manda \mathbf{e}_1 nel suo simmetrico $S(\mathbf{e}_1)$ rispetto all'asse. Il vettore $S(\mathbf{e}_1)$ è quindi un vettore di lunghezza uno che forma con r un angolo $-\alpha$, per cui:

$$S(\mathbf{e}_1) = \begin{bmatrix} \cos(2\alpha) \\ \sin(2\alpha) \end{bmatrix}$$

D'altra parte, l'asse di riflessione forma un angolo $\pi/2 - \alpha$ col secondo vettore \mathbf{e}_2 della base canonica, per cui $S(\mathbf{e}_2)$ forma un angolo pari a $\pi - 2\alpha$ con \mathbf{e}_2 e quindi un angolo pari a $2\alpha - \pi/2$ con \mathbf{e}_1 . Perciò

$$S(\mathbf{e}_2) = \begin{bmatrix} \cos(2\alpha - \pi/2) \\ \sin(2\alpha - \pi/2) \end{bmatrix} = \begin{bmatrix} \sin(2\alpha) \\ -\cos(2\alpha) \end{bmatrix}$$

Concludendo, la matrice che rappresenta S rispetto alla base canonica è

$$\mathbf{A} = \begin{bmatrix} \cos(2\alpha) & \sin(2\alpha) \\ \sin(2\alpha) & -\cos(2\alpha) \end{bmatrix}$$

C'è però una base molto più naturale per rappresentare S : è la base \mathcal{B} costituita dal versore $\mathbf{b}_1 = \begin{bmatrix} \cos(\alpha) \\ \sin(\alpha) \end{bmatrix}$ dell'asse di riflessione, e dal versore $\mathbf{b}_2 = \begin{bmatrix} -\sin(\alpha) \\ \cos(\alpha) \end{bmatrix}$ perpendicolare all'asse. Evidentemente, $S(\mathbf{b}_1) = \mathbf{b}_1$ e $S(\mathbf{b}_2) = -\mathbf{b}_2$. Siccome $\mathbf{b}_1 = 1\mathbf{b}_1 + 0\mathbf{b}_2$, il vettore \mathbf{y}_1 delle coordinate di $S(\mathbf{b}_1)$ rispetto alla base \mathcal{B} è $\mathbf{y}_1 = [1, 0]^T$. Analogamente, il vettore \mathbf{y}_2 delle coordinate di $S(\mathbf{b}_2)$ rispetto alla base \mathcal{B} è $\mathbf{y}_2 = [0, -1]^T$. Quindi la matrice che rappresenta S rispetto alla base \mathcal{B} è:

$$\mathbf{B} = [\mathbf{y}_1 \quad \mathbf{y}_2] = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

La matrice *diagonale* \mathbf{B} è molto più semplice della matrice \mathbf{A} che rappresenta la stessa applicazione rispetto alla base canonica: il motivo è che la base \mathcal{B} è quella naturale dal punto di vista della riflessione S . Nel capitolo sugli autovettori e autovalori tratteremo il problema di come determinare una base altrettanto naturale per una trasformazione di cui si conosca soltanto la matrice rappresentativa \mathbf{A} rispetto alla base canonica.



Sia \mathcal{R} la rotazione dello spazio cartesiano di un angolo θ intorno all'asse r .

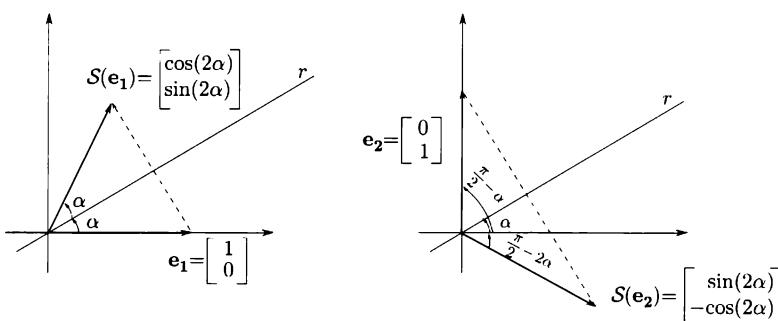


Figura 5.3. Riflessione ortogonale con asse $y = \operatorname{tg}(\alpha)$.

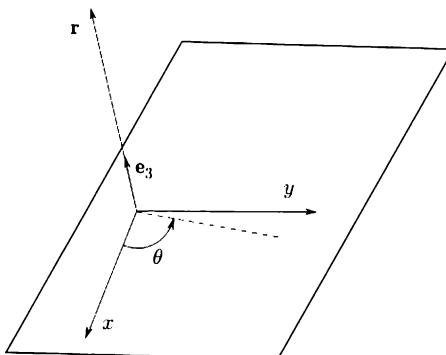


Figura 5.4. Rotazione di un angolo θ intorno all'asse r nello spazio.

Si tratta di un'applicazione lineare, ma la sua espressione in coordinate (la sua matrice rappresentativa) è in generale complicata, a meno che non si scelgano le coordinate in modo opportuno. Una scelta intelligente di coordinate in questo caso consiste nel prendere il versore e_3 lungo l'asse di rotazione, in modo che l'asse z coincida con l'asse di rotazione. Scegliamo poi nel piano perpendicolare a e_3 gli assi x e y (equivalentemente i versori e_1 ed e_2) in modo che la rotazione ruoti il semiasse positivo x verso il semiasse positivo y di un angolo θ . L'effetto della rotazione sui vettori della base è allora:

$$\mathcal{R}(e_1) = \cos(\theta)e_1 + \sin(\theta)e_2, \quad \mathcal{R}(e_2) = -\sin(\theta)e_1 + \cos(\theta)e_2, \quad \mathcal{R}(e_3) = e_3.$$

La matrice A che rappresenta \mathcal{R} rispetto alla base $B = \{e_1, e_2, e_3\}$ è perciò

$$A = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

e l'espressione di \mathcal{R} rispetto alle coordinate scelte è:

$$\mathcal{R}_c \left([x, y, z]^T \right) = A \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \cos(\theta)x - \sin(\theta)y \\ \sin(\theta)x + \cos(\theta)y \\ z \end{bmatrix}$$

Sia V lo spazio vettoriale delle matrici 2×2 a coefficienti reali, con la base canonica B formata dalle quattro matrici:

$$E_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad E_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Consideriamo l'applicazione $\mathfrak{L} : V \rightarrow V$ definita da

$$\mathfrak{L}(A) = AE_1 - E_1A$$

L'applicazione \mathfrak{L} è lineare perché il prodotto di matrici è bilineare. La matrice A che rappresenta \mathfrak{L} rispetto alla base canonica B è una matrice 4×4 perché V ha dimensione 4. Per determinare A scriviamo l'espressione di \mathfrak{L} in coordinate: siccome

$$\mathfrak{L} \left(\begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \right) = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} = \begin{bmatrix} 0 & -x_2 \\ x_3 & 0 \end{bmatrix}$$

l'espressione in coordinate di \mathcal{L} è

$$\mathcal{L}_c \left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \right) = \begin{bmatrix} 0 \\ -x_2 \\ x_3 \\ 0 \end{bmatrix}$$

La matrice \mathbf{A} è quindi

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

Il teorema di rappresentazione permette di identificare lo spazio vettoriale delle applicazioni lineari da \mathbf{V} in \mathbf{W} con lo spazio vettoriale delle matrici $\dim \mathbf{W} \times \dim \mathbf{V}$. Questa identificazione rispetta le operazioni di somma e prodotto per scalare ed è quindi un isomorfismo:

PROPOSIZIONE 5.3 Siano \mathbf{V} e \mathbf{W} spazi vettoriale di dimensione n ed m rispettivamente, \mathcal{B} una base di \mathbf{V} e \mathcal{C} una base di \mathbf{W} . Sia

$$\mathcal{G} : \text{Hom}_{\mathbb{K}}(\mathbf{V}, \mathbf{W}) \rightarrow \mathbb{M}_{\mathbb{K}}(m, n)$$

la funzione che a un'applicazione \mathcal{L} associa la matrice \mathbf{A} che rappresenta \mathcal{L} rispetto alle basi \mathcal{B} e \mathcal{C} . Allora \mathcal{G} è un isomorfismo.

DIMOSTRAZIONE. Dobbiamo dimostrare che \mathcal{G} è invertibile e lineare. Sia \mathcal{F} la funzione che a una matrice $\mathbf{A} \in \mathbb{M}_{\mathbb{K}}(m, n)$ associa l'applicazione lineare la cui espressione in coordinate è $\mathcal{L}_{\mathbf{A}}$. Per il teorema di rappresentazione \mathcal{G} è l'inversa di \mathcal{F} . Basta quindi dimostrare che \mathcal{G} è lineare. Questo significa che, se \mathbf{A} e \mathbf{B} sono le matrici che rappresentano le applicazioni lineari $\mathcal{L}, \mathcal{M} : \mathbf{V} \rightarrow \mathbf{W}$, allora:

- a) la matrice $\mathbf{A} + \mathbf{B}$ rappresenta $\mathcal{L} + \mathcal{M}$;
- b) per ogni scalare $t \in \mathbb{K}$, la matrice $t\mathbf{A}$ rappresenta l'applicazione $t\mathcal{L}$.

La semplice verifica di a) e b) è lasciata al lettore.

La matrice che rappresenta il prodotto di composizione $\mathcal{M} \circ \mathcal{L}$ è il prodotto delle matrici che rappresentano \mathcal{M} e \mathcal{L} :

PROPOSIZIONE 5.4 Siano $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ ed $\mathcal{M} : \mathbf{W} \rightarrow \mathbf{Z}$ applicazioni lineari.

Siano $\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3$ basi di \mathbf{V}, \mathbf{W} e \mathbf{Z} , rispettivamente. Supponiamo che la matrice \mathbf{A} rappresenti \mathcal{L} rispetto alle basi \mathcal{B}_1 e \mathcal{B}_2 e che la matrice \mathbf{B} rappresenti \mathcal{M} rispetto alle basi \mathcal{B}_2 e \mathcal{B}_3 . Allora la matrice prodotto \mathbf{BA} rappresenta $\mathcal{M} \circ \mathcal{L}$ rispetto alle basi \mathcal{B}_1 e \mathcal{B}_3 .

DIMOSTRAZIONE. Fissato $\mathbf{v} \in \mathbf{V}$, siano \mathbf{x}, \mathbf{y} e \mathbf{z} i vettori delle coordinate di \mathbf{v} , $\mathcal{L}(\mathbf{v})$ e $\mathcal{M}(\mathcal{L}(\mathbf{v}))$ rispetto alle basi $\mathcal{B}_1, \mathcal{B}_2$ e \mathcal{B}_3 . Allora

$$\mathbf{z} = \mathbf{By} = \mathbf{B}(\mathbf{Ax}) = \mathbf{BAx}$$

Siccome $\mathfrak{M} \circ \mathcal{L} = \mathfrak{M}(\mathcal{L}(\mathbf{v}))$, questo mostra che $\mathbf{B}\mathbf{A}$ rappresenta $\mathfrak{M} \circ \mathcal{L}$ rispetto alle basi \mathcal{B}_1 e \mathcal{B}_3 . ■

Esempio

Dati un endomorfismo \mathcal{L} di uno spazio vettoriale \mathbf{V} , cioè un'applicazione lineare $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{V}$, e un polinomio $P(x)$, abbiamo definito nel paragrafo precedente l'applicazione lineare $P(\mathcal{L})$: se $P(x) = a_m x^m + \dots + a_1 x + a_0$, allora

$$P(\mathcal{L}) = a_m \mathcal{L}^m + \dots + a_1 \mathcal{L} + a_0 \mathcal{I}_{\mathbf{V}}$$

Se \mathbf{A} è una matrice quadrata, possiamo anche definire

$$P(\mathbf{A}) = a_m \mathbf{A}^m + \dots + a_1 \mathbf{A} + a_0 \mathbf{I}$$

Combinando i risultati di questo paragrafo, si vede che, se \mathbf{A} rappresenta \mathcal{L} rispetto a una base \mathcal{B} di \mathbf{V} , allora $P(\mathbf{A})$ rappresenta $P(\mathcal{L})$ rispetto alla stessa base \mathcal{B} .

Esempio

Cerchiamo una soluzione particolare dell'equazione differenziale

$$(5.7) \quad y''(t) - 5y'(t) + 6y(t) = \cos(t)$$

della forma $y(t) = a \cos(t) + b \sin(t)$, cioè appartenente allo spazio vettoriale \mathbf{V} dell'esercizio precedente. A primo membro di (5.7) compare il polinomio differenziale $P(\mathcal{D})$, dove $P(x) = x^2 - 5x + 6$ e $\mathcal{D}(y) = y'$ è l'operatore derivata. Siccome \mathcal{D} è lineare in y , anche $P(\mathcal{D})$ è lineare in y . La matrice che rappresenta \mathcal{D} rispetto alla base $\{\cos(t), \sin(t)\}$ è $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, per cui la matrice che rappresenta $P(\mathcal{D})$ è

$$\mathbf{B} = P(\mathbf{A}) = \mathbf{A}^2 - 5\mathbf{A} + 6\mathbf{I} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} - 5 \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} + 6 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & -5 \\ 5 & 5 \end{bmatrix}$$

L'equazione differenziale (5.7) nelle coordinate $\mathbf{x} = \begin{bmatrix} a \\ b \end{bmatrix}$ diviene quindi il sistema lineare

$$(5.8) \quad \mathbf{B}\mathbf{x} = \begin{bmatrix} 5 & -5 \\ 5 & 5 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

che ha per soluzione

$$\begin{bmatrix} a \\ b \end{bmatrix} = \mathbf{B}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{50} \begin{bmatrix} 5 & 5 \\ -5 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/10 \\ -1/10 \end{bmatrix}$$

Quindi una soluzione particolare della (5.7), l'unica soluzione ad appartenere a \mathbf{V} , è

$$y_0(t) = \frac{1}{10} \cos(t) - \frac{1}{10} \sin(t)$$

Il lettore dovrebbe verificare che $y_0(t)$ soddisfa l'equazione (5.7).

Cambiamo il termine noto dell'equazione mantenendolo però in \mathbf{V} . Otteniamo un'equazione del tipo:

$$(5.9) \quad y''(t) - 5y'(t) + 6y(t) = c \cos(t) + d \sin(t)$$

Come prima, l'equazione ha un'unica soluzione $a_0 \cos(t) + b_0 \sin(t)$ in \mathbf{V} , con a_0 e b_0 assegnati dalla formula

$$\begin{bmatrix} a_0 \\ b_0 \end{bmatrix} = \mathbf{B}^{-1} \begin{bmatrix} c \\ d \end{bmatrix} = \frac{1}{50} \begin{bmatrix} 5 & 5 \\ -5 & 5 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \frac{1}{10} \begin{bmatrix} c+d \\ d-c \end{bmatrix}$$

Cambiamento della matrice rappresentativa al variare delle basi

Vediamo ora come cambia la matrice rappresentativa di un'applicazione lineare al variare delle basi. Si tratta di una formula importante: tipicamente l'applicazione lineare è assegnata rispetto a una base canonica e si vuole invece rappresentarla rispetto a una base che meglio le si adatta. Questo è il principio su cui si basa il metodo degli autovettori e autovalori di cui ci occuperemo più avanti.

Si ricordi che nel paragrafo 7 del capitolo 4 abbiamo introdotto la matrice \mathbf{S} di passaggio da una base \mathcal{B}_1 a una base \mathcal{B}_2 di uno spazio vettoriale \mathbf{V} . La matrice di passaggio esprime il legame tra le coordinate di un vettore rispetto alle due basi: se \mathbf{x} e \mathbf{X} sono i vettori delle coordinate di $\mathbf{v} \in \mathbf{V}$ rispetto a \mathcal{B}_1 e \mathcal{B}_2 rispettivamente, allora

$$(5.10) \quad \mathbf{x} = \mathbf{S}\mathbf{X}$$

PROPOSIZIONE 5.5 (Formula cambiamento matrice rappresentativa)

Sia $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare. Siano: \mathcal{B}_1 e \mathcal{B}_2 basi di \mathbf{V} , \mathcal{C}_1 e \mathcal{C}_2 basi di \mathbf{W} , \mathbf{S} la matrice di passaggio da \mathcal{B}_1 a \mathcal{B}_2 e \mathbf{T} la matrice di passaggio da \mathcal{C}_1 a \mathcal{C}_2 .

Se \mathbf{A} è la matrice che rappresenta \mathfrak{L} rispetto alle basi \mathcal{B}_1 e \mathcal{C}_2 , la matrice

$$\mathbf{B} = \mathbf{T}^{-1}\mathbf{AS}$$

rappresenta \mathfrak{L} rispetto alle basi \mathcal{B}_2 e \mathcal{C}_2 .

DIMOSTRAZIONE. Se \mathbf{x} e \mathbf{X} sono i vettori delle coordinate di $\mathbf{v} \in \mathbf{V}$ rispetto alle basi \mathcal{B}_1 e \mathcal{B}_2 , allora $\mathbf{x} = \mathbf{S}\mathbf{X}$. Se \mathbf{y} e \mathbf{Y} sono i vettori delle coordinate di $\mathbf{w} \in \mathbf{W}$ rispetto alle basi \mathcal{B}_2 e \mathcal{C}_2 , allora $\mathbf{y} = \mathbf{T}\mathbf{Y}$. La matrice \mathbf{A} è l'unica matrice tale che per ogni $\mathbf{v} \in \mathbf{V}$ si abbia

$$(5.11) \quad \mathbf{y} = \mathbf{Ax}$$

dove \mathbf{y} denota il vettore delle coordinate di $\mathbf{w} = \mathfrak{L}(\mathbf{v})$ rispetto alla base \mathcal{C}_1 . Passando alle coordinate rispetto alle basi \mathcal{B}_2 e \mathcal{C}_2 avremo allora:

$$(5.12) \quad \mathbf{Y} = \mathbf{T}^{-1}\mathbf{y} = \mathbf{T}^{-1}\mathbf{Ax} = \mathbf{T}^{-1}\mathbf{AS}\mathbf{X}$$

Questo mostra che la matrice $\mathbf{T}^{-1}\mathbf{AS}$ rappresenta \mathfrak{L} rispetto alle basi \mathcal{B}_2 e \mathcal{C}_2 .

Nella maggior parte delle applicazioni, la funzione $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{V}$ è un endomorfismo di \mathbf{V} . Se \mathbf{S} è la matrice di passaggio da una base \mathcal{B}_1 a una base \mathcal{B}_2 di \mathbf{V} e \mathbf{A} rappresenta \mathfrak{L} rispetto alla base \mathcal{B}_1 , allora

$$(5.13) \quad \mathbf{S}^{-1}\mathbf{AS}$$

rappresenta \mathfrak{L} rispetto alla base \mathcal{B}_2 .

Consideriamo la riflessione ortogonale \mathcal{S} del piano \mathbb{R}^2 che ha come asse la retta per l'origine diretta come il versore $[\cos(\alpha), \sin(\alpha)]^T$. In questo caso $\mathbf{V} = \mathbf{W} = \mathbb{R}^2$, e prendiamo per \mathcal{B}_1 la base canonica, mentre poniamo

$$\mathcal{B}_2 = \left\{ \begin{bmatrix} \cos(\alpha) \\ \sin(\alpha) \end{bmatrix}, \begin{bmatrix} -\sin(\alpha) \\ \cos(\alpha) \end{bmatrix} \right\}$$

La matrice \mathbf{S} di passaggio dalla base canonica alla base \mathcal{B}_2 ha per colonne i vettori della nuova base:

$$\mathbf{S} = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix}$$

Abbiamo determinato precedentemente le matrici \mathbf{A} e \mathbf{B} che rappresentano \mathcal{S} rispetto alla base canonica e alla base \mathcal{B}_2 :

$$\mathbf{A} = \begin{bmatrix} \cos(2\alpha) & \sin(2\alpha) \\ \sin(2\alpha) & -\cos(2\alpha) \end{bmatrix} \quad \text{e} \quad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Per la proposizione precedente

$$\mathbf{B} = \mathbf{S}^{-1} \mathbf{A} \mathbf{S}$$

Verifichiamo il caso in cui $\alpha = \pi/4$:

$$\begin{aligned} \mathbf{S}^{-1} \mathbf{A} \mathbf{S} &= \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} = \\ &= \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ \sqrt{2}/2 & -\sqrt{2}/2 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \end{aligned}$$

Sia $\{\mathbf{e}_1, \mathbf{e}_2\}$ la base canonica di \mathbb{R}^2 e sia $\mathfrak{L} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ l'applicazione lineare tale che $\mathfrak{L}(\mathbf{e}_1) = [1, 2]^T$ e $\mathfrak{L}(\mathbf{e}_2) = [4, 3]^T$. Mostrare che

$$\mathfrak{L} \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} x + 4y \\ 2x + 3y \end{bmatrix}$$

Scrivere la matrice che rappresenta \mathfrak{L} rispetto alla base canonica.

Sia $\mathfrak{L} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ l'applicazione lineare

$$\mathfrak{L} \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} 3x + y \\ 3x + y \end{bmatrix}$$

Scrivere la matrice che rappresenta \mathfrak{L} rispetto alla base canonica e la matrice che rappresenta \mathfrak{L} rispetto alla base $\{[1, 1]^T, [1, -3]^T\}$. Trovare una base del nucleo e dell'immagine di \mathfrak{L} . Trovare equazioni cartesiane per il nucleo e per l'immagine di \mathfrak{L} .

Sia $\mathfrak{L} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ l'applicazione lineare tale che:

$$\mathfrak{L} \left(\begin{bmatrix} 2 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \text{e} \quad \mathfrak{L} \left(\begin{bmatrix} 2 \\ 3 \end{bmatrix} \right) = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

Mostrare che:

$$\mathfrak{L} \left(\begin{bmatrix} 2 \\ -1 \end{bmatrix} \right) = \begin{bmatrix} -1 \\ 4 \end{bmatrix}$$

 Supponiamo che \mathbf{V} sia uno spazio vettoriale con base $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ e sia a uno scalare. Sia $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{V}$ l'applicazione lineare tale che $\mathfrak{L}(\mathbf{b}_1) = a\mathbf{b}_1$ e $\mathfrak{L}(\mathbf{b}_k) = \mathbf{b}_{k-1} + a\mathbf{b}_k$ per $k = 2, \dots, n$. Mostrare che la matrice che rappresenta \mathfrak{L} rispetto alla base \mathcal{B} è il *blocco di Jordan* $\mathbf{J}_n(a)$, cioè la matrice quadrata $n \times n$ che ha gli elementi sulla diagonale principale uguali ad a , gli elementi sulla diagonale immediatamente sopra alla diagonale principale uguali a 1 e tutti gli altri elementi nulli.

 Sia \mathbf{V} lo spazio vettoriale delle funzioni $y(x)$ che sono combinazioni lineari di $\cos(x)$ e $\sin(x)$. Si consideri l'operatore differenziale:

$$\mathfrak{L}(y) = y'' + ay' + by$$

dove a e b sono numeri reali fissati. Mostrare che \mathfrak{L} è una funzione lineare da \mathbf{V} in \mathbf{V} . Trovare la matrice che rappresenta \mathfrak{L} rispetto alla base $\{\cos(x), \sin(x)\}$ di \mathbf{V} . Mostrare che l'equazione differenziale

$$\mathfrak{L}(y) = \cos(x)$$

ha una e una sola soluzione in \mathbf{V} tranne nel caso $a = 0, b = 1$. Per trattare questo caso, sia \mathbf{Z} lo spazio vettoriale delle funzioni $y(x)$ che sono combinazioni lineari di $x\cos(x)$ e $x\sin(x)$. Mostrare che se $y(x) \in \mathbf{Z}$, allora $\mathfrak{L}(y) = y'' + y$ appartiene a \mathbf{V} e che la funzione $\mathfrak{L} : \mathbf{Z} \rightarrow \mathbf{V}$ è invertibile. Trovare una soluzione $y(x)$ di $L(y) = \cos(x)$.

Suggerimento: La matrice che rappresenta l'operatore derivata $\mathfrak{D}(y) = y'$ è $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. La matrice di \mathfrak{L} è quindi:

$$\mathbf{M} = \mathbf{A}^2 + a\mathbf{A} + b\mathbf{I} = \begin{bmatrix} b-1 & a \\ -a & b-1 \end{bmatrix}$$

che ha determinante $(b-1)^2 + a^2$. L'applicazione \mathfrak{L} è invertibile (e quindi esiste una e una sola soluzione di $\mathfrak{L}(y) = \cos(x)$ in \mathbf{V}) se e solo se tale determinante è diverso da zero, quindi in tutti i casi tranne quello in cui $a = 0$ e $b = 1$.

 Sia \mathbf{V} lo spazio vettoriale delle matrici di tipo $(2, 2)$ a coefficienti reali. Mostrare che, se una matrice \mathbf{B} non è un multiplo della matrice identità, allora le uniche matrici che commutano con \mathbf{B} sono le combinazioni lineari

$$t_1\mathbf{I} + t_2\mathbf{B}$$

utilizzando la seguente traccia.

1. Mostrare che le quattro matrici

$$\mathbf{E}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{E}_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{E}_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{E}_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

formano una base \mathcal{B} di \mathbf{V} .

2. Si fissi una matrice $\mathbf{B} \in \mathbf{V}$. Sia $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{V}$ la funzione definita da

$$\mathfrak{L}(\mathbf{A}) = \mathbf{AB} - \mathbf{BA}$$

Si mostri che \mathfrak{L} è lineare.

3. Fissati gli elementi di \mathbf{A} e \mathbf{B} (cioè le coordinate rispetto alla base \mathcal{B})

$$\mathbf{A} = \begin{bmatrix} x & y \\ z & w \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

si calcoli $\mathfrak{L}(\mathbf{A})$.

4. Si scriva la matrice \mathbf{M} di tipo $(4,4)$ che rappresenta \mathcal{L} rispetto alla base \mathcal{B} di \mathbf{V} : occorre calcolare le coordinate di $\mathcal{L}(\mathbf{E}_1), \dots, \mathcal{L}(\mathbf{E}_4)$.
5. Si mostri che il rango di \mathbf{M} è 2, a meno che \mathbf{B} non sia un multiplo della matrice identità, nel qual caso \mathbf{M} è la matrice nulla.
6. Il nucleo di \mathbf{M} consiste delle coordinate delle matrici che commutano con \mathbf{B} . Per il punto precedente ha dimensione 2 se \mathbf{B} non è un multiplo della matrice identità. Concludere che le uniche matrici che commutano con \mathbf{B} sono le combinazioni lineari

$$t_1\mathbf{I} + t_2\mathbf{B}$$

Suggerimento: il nucleo ha dimensione 2, e contiene le due matrici \mathbf{I} e \mathbf{B} , perché \mathbf{I} e \mathbf{B} commutano con \mathbf{B} . L'ipotesi che \mathbf{B} non sia un multiplo della matrice identità significa che \mathbf{I} e \mathbf{B} sono linearmente indipendenti, quindi generano un sottospazio di dimensione 2, che deve coincidere col nucleo.

Si consideri la funzione $\mathcal{T} : \mathbb{M}_{\mathbb{K}}(m,n) \rightarrow \mathbb{M}_{\mathbb{K}}(n,m)$ che a una matrice \mathbf{A} associa la matrice trasposta \mathbf{A}^T . Si mostri che \mathcal{T} è un isomorfismo.

Suggerimento: $(\mathbf{A}^T)^T = \mathbf{A}$). Si trovi la matrice che rappresenta \mathcal{T} rispetto alle basi canoniche nel caso $m = n = 2$ (*Risposta:* si tratta della matrice di permutazione di ordine 4, ottenuta scambiando la seconda riga della matrice identità con la terza).

 Mostrare che la matrice \mathbf{S} di passaggio dalla base \mathcal{B} alla base \mathcal{C} di \mathbf{V} rappresenta l'applicazione identità $\mathcal{I}_{\mathbf{V}} : \mathbf{V} \rightarrow \mathbf{V}$ rispetto alle basi \mathcal{C} per il dominio e \mathcal{B} per il codominio, in ordine inverso quindi rispetto a come le abbiamo introdotte.

Suggerimento: la matrice che rappresenta $\mathcal{I}_{\mathbf{V}}$ ha come k -esima colonna il vettore delle coordinate di $\mathcal{I}_{\mathbf{V}}(\mathbf{c}_k) = \mathbf{c}_k$ rispetto alla base \mathcal{B} .

 Se $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ è un'applicazione invertibile e \mathbf{A} è la matrice che rappresenta \mathcal{L} rispetto alle basi \mathcal{B} e \mathcal{C} , allora \mathbf{A} è invertibile e l'applicazione $\mathcal{L}^{-1} : \mathbf{W} \rightarrow \mathbf{V}$ è rappresentata da \mathbf{A}^{-1} rispetto alle basi \mathcal{C} e \mathcal{B} .

 **Base duale** Sia \mathbf{V} uno spazio vettoriale, e sia $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ una base di \mathbf{V} . Lo *spazio duale* $\mathbf{V}^* = \text{Hom}_{\mathbb{K}}(\mathbf{V}, \mathbb{K})$ contiene le forme lineari $\beta_k : \mathbf{V} \rightarrow \mathbb{K}$

$$\beta_k(x_1\mathbf{b}_1 + \dots + x_n\mathbf{b}_n) = x_k \quad (k = 1, 2, \dots, n)$$

Mostrare che

- a) la funzione β_k è l'unica applicazione lineare $\mathbf{V} \rightarrow \mathbb{K}$ tale che

$$\beta_k(\mathbf{b}_h) = \delta_{hk}$$

(qui δ_{hk} è il simbolo di Kronecker che vale 1 se $h = k$ e 0 altrimenti).

- b) l'insieme $\{\beta_1, \dots, \beta_n\}$ è una base di \mathbf{V}^* : si dice che è la base duale della base $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$. In particolare, $\dim \mathbf{V}^* = \dim \mathbf{V}$. Suggerimento: se una forma lineare β è rappresentata dal vettore riga $[a_1, a_2, \dots, a_n]$, allora $\beta = \sum_{i=1}^n a_i \beta_i$.
- c) se \mathbf{v} ha coordinate \mathbf{x} rispetto alla base \mathcal{B} e β ha coordinate \mathbf{a} rispetto alla base duale, allora

$$\beta(\mathbf{v}) = \mathbf{a}^T \mathbf{x} = a_1x_1 + \dots + a_nx_n$$

N.B. Questo esercizio mostra che una forma lineare in coordinate è un polinomio omogeneo di primo grado; questo è il significato originale di forma (= polinomio omogeneo) lineare (= di primo grado).

 Sia \mathbf{H} un sottospazio di uno spazio vettoriale \mathbf{V} di dimensione finita e sia $\mathfrak{M} : \mathbf{H} \rightarrow \mathbf{W}$ un'applicazione lineare. Mostrare che esiste un'estensione lineare di \mathfrak{M} a tutto \mathbf{V} : esiste cioè un'applicazione lineare $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{W}$ tale che $\mathfrak{L}(\mathbf{v}) = \mathfrak{M}(\mathbf{v})$ per ogni $\mathbf{v} \in \mathbf{H}$.

Suggerimento: fissata una base $\{\mathbf{b}_1, \dots, \mathbf{b}_d\}$ di \mathbf{H} , la si estenda a una base $\{\mathbf{b}_1, \dots, \mathbf{b}_d, \mathbf{b}_{d+1}, \dots, \mathbf{b}_n\}$ di \mathbf{V} ; Si definisca $\mathfrak{L}(\mathbf{b}_k) = \mathfrak{M}(\mathbf{b}_k)$ per $k \leq d$ ed $\mathfrak{L}(\mathbf{b}_k)$ in modo qualsiasi per $k > d$.

 Sia \mathbf{V} uno spazio vettoriale di dimensione finita e sia \mathbf{v} un vettore di \mathbf{V} . Se $\beta(\mathbf{v}) = 0$ per ogni $\beta \in \mathbf{V}^*$, allora $\mathbf{v} = 0$.

Suggerimento: se $\mathbf{v} \neq 0$, estendere a \mathbf{V} la forma lineare $\mathfrak{M} : \mathbb{K}\mathbf{v} \rightarrow \mathbb{K}$ che manda la base \mathbf{v} di $\mathbb{K}\mathbf{v}$ in 1 (cioè $\mathfrak{M}(t\mathbf{v}) = t$).

Dato un vettore \mathbf{v} in uno spazio vettoriale \mathbf{V} , mostrare che la funzione $\mathbf{v}^{**} : \mathbf{V}^* \rightarrow \mathbb{K}$ definita da

$$\mathbf{v}^{**}(\beta) = \beta(\mathbf{v})$$

è una forma lineare su \mathbf{V}^* , quindi un elemento di \mathbf{V}^{**} . Mostrare che la funzione $\mathbf{V} \rightarrow \mathbf{V}^{**}$ che a \mathbf{v} associa \mathbf{v}^{**} è lineare e che è un isomorfismo se \mathbf{V} ha dimensione finita (questa ipotesi è necessaria).

 Se \mathbf{H} è un sottospazio di uno spazio vettoriale \mathbf{V} , mostrare che

$$\mathbf{H}^\perp = \{\beta \in \mathbf{V}^* : \beta(\mathbf{v}) = 0 \text{ per ogni } \mathbf{v} \in \mathbf{V}\}$$

è un sottospazio di \mathbf{V}^* ; se \mathbf{V} ha dimensione finita, mostrare che $\dim \mathbf{H}^\perp = \dim \mathbf{V} - \dim \mathbf{H}$ e dedurre che $(\mathbf{H}^\perp)^\perp = \mathbf{H}$ (modulo l'identificazione $\mathbf{V}^{**} = \mathbf{V}$ dell'esercizio precedente).

Suggerimento: la versione in coordinate di questo esercizio è la proposizione 9.1 del capitolo sugli spazi vettoriali.

 Sia $\mathfrak{L} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare tra spazi vettoriali di dimensione finita. Si dice che una funzione lineare $\mathfrak{M} : \mathbf{W} \rightarrow \mathbf{V}$ è un'inversa sinistra (rispettivamente destra) di \mathfrak{L} se $\mathfrak{M} \circ \mathfrak{L} = \mathcal{I}_{\mathbf{V}}$ (rispettivamente $\mathfrak{L} \circ \mathfrak{M} = \mathcal{I}_{\mathbf{W}}$). Mostrare che \mathfrak{L} ha un'inversa sinistra se e solo se è iniettiva, e ha un'inversa destra se e solo se è suriettiva.

Suggerimento: se \mathfrak{L} è iniettiva, $\mathfrak{L} : \mathbf{V} \rightarrow \text{Im}(\mathfrak{L})$ è invertibile, per cui esiste $\mathfrak{L}^{-1} : \text{Im}(\mathfrak{L}) \rightarrow \mathbf{V}$; per ottenere un'inversa sinistra estendere, usando un esercizio precedente, \mathfrak{L}^{-1} a una funzione lineare $\mathbf{W} \rightarrow \mathbf{V}$; si noti che l'inversa sinistra non è unica se $\dim(\mathbf{W}) > \dim(\mathbf{V})$; se \mathfrak{L} è suriettiva, si fissi una base $\{\mathbf{b}_1, \dots, \mathbf{b}_m\}$ di \mathbf{W} e per ogni $k = 1, \dots, m$ si scelga \mathbf{v}_k tale che $\mathfrak{L}(\mathbf{v}_k) = \mathbf{b}_k$; si definisca $\mathfrak{M} : \mathbf{W} \rightarrow \mathbf{V}$ come la funzione lineare tale che $\mathfrak{M}(\mathbf{b}_k) = \mathbf{v}_k$ per ogni k .

Se \mathbf{V} ha dimensione n , la funzione

$$\mathcal{G} : \text{End}_{\mathbb{K}}(\mathbf{V}) \rightarrow \mathbb{M}_{\mathbb{K}}(n, n)$$

che a un endomorfismo associa la matrice che lo rappresenta (rispetto a una base fissata) è un isomorfismo di \mathbb{K} -algebre (questo significa che \mathcal{G} rispetta somma, prodotto per scalare e prodotto ed è iniettiva e suriettiva).

6 TEOREMA DI NULLITÀ PIÙ RANGO E APPLICAZIONI

In questo paragrafo traduciamo il teorema di nullità più rango nel contesto delle applicazioni lineari. Il rango di una matrice \mathbf{A} è la dimensione dello spazio colonna; lo spazio colonna a sua volta coincide con l'immagine dell'applicazione lineare $\mathcal{L}_{\mathbf{A}}$ rappresentata dalla matrice. È quindi naturale definire il rango di un'applicazione lineare come la dimensione dell'immagine.

DEFINIZIONE 6.1 (Rango di un'applicazione lineare)

Sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare. Il rango $r(\mathcal{L})$ di \mathcal{L} è la dimensione dell'immagine $\text{Im}(\mathcal{L})$ di \mathcal{L} .

Se $\mathcal{L} = \mathcal{L}_{\mathbf{A}}$, allora $r(\mathcal{L}) = r(\mathbf{A})$ perché $\text{Im}(\mathcal{L}_{\mathbf{A}}) = \text{Col}(\mathbf{A})$.

Come il rango di una matrice è minore o uguale del minimo tra il numero delle righe e delle colonne, così :

PROPOSIZIONE 6.2 Sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare. Allora

$$r(\mathcal{L}) \leq \dim(\mathbf{V}) \quad \text{e} \quad r(\mathcal{L}) \leq \dim(\mathbf{W})$$

In particolare, $r(\mathcal{L}) < +\infty$ se almeno uno dei due spazi \mathbf{V} e \mathbf{W} ha dimensione finita.

DIMOSTRAZIONE. Supponiamo che \mathbf{V} abbia dimensione finita e che $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ sia una base di \mathbf{V} . Per mostrare che $r(\mathcal{L}) \leq n$ basta mostrare che i vettori $\mathcal{L}(\mathbf{b}_1), \dots, \mathcal{L}(\mathbf{b}_n)$ sono un insieme di generatori dell'immagine; questo significa che ogni vettore $\mathcal{L}(\mathbf{v})$ dell'immagine è combinazione lineare di $\mathcal{L}(\mathbf{b}_1), \dots, \mathcal{L}(\mathbf{b}_n)$. Certamente possiamo scrivere \mathbf{v} come combinazione lineare

$$\mathbf{v} = t_1 \mathbf{b}_1 + \cdots + t_n \mathbf{b}_n$$

dei vettori della base di \mathbf{V} . Applicando \mathcal{L} ai due membri di questa uguaglianza, tenuto conto della linearità di \mathcal{L} , otteniamo

$$\mathcal{L}(\mathbf{v}) = t_1 \mathcal{L}(\mathbf{b}_1) + \cdots + t_n \mathcal{L}(\mathbf{b}_n)$$

Quindi $\mathcal{L}(\mathbf{b}_1), \dots, \mathcal{L}(\mathbf{b}_n)$ generano l'immagine, per cui

$$r(\mathcal{L}) = \dim(\text{Im}(\mathcal{L})) \leq n = \dim(\mathbf{V})$$

La seconda diseguaglianza $r(\mathcal{L}) \leq \dim(\mathbf{W})$ segue dal fatto che l'immagine $\text{Im}(\mathcal{L})$ è un sottospazio del codominio \mathbf{W} .

Il teorema di nullità più rango per un'applicazione lineare è:

TEOREMA 6.3 (di nullità più rango)

Sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare. Se \mathbf{V} ha dimensione finita, allora

$$\dim(\mathbf{V}) = \dim \text{Ker}(\mathcal{L}) + \dim \text{Im}(\mathcal{L})$$

DIMOSTRAZIONE. Il sottospazio immagine $\text{Im}(\mathcal{L})$ ha dimensione finita per la proposizione precedente. Possiamo quindi sostituire \mathbf{W} con il suo sottospazio $\text{Im}(\mathcal{L})$ e assumere che anche \mathbf{W} abbia dimensione finita.

Fissate una base di \mathbf{V} e una di \mathbf{W} , sia $\mathbf{x} : \mathbf{V} \rightarrow \mathbb{K}^n$ (rispettivamente, $\mathbf{y} : \mathbf{W} \rightarrow \mathbb{K}^m$) l'applicazione lineare che a un vettore associa il vettore delle sue coordinate e sia \mathbf{A} la matrice che rappresenta \mathcal{L} rispetto alle basi fissate. Siccome $\mathbf{x}(\text{Ker}(\mathcal{L})) = \text{Ker}(\mathbf{A})$ e \mathbf{x} è un isomorfismo, la dimensione del nucleo di \mathcal{L} è uguale alla dimensione del nucleo di \mathbf{A} . Analogamente, poiché $\mathbf{y}(\text{Im}(\mathcal{L})) = \text{Im}(\mathbf{A}) = \text{Col}(\mathbf{A})$, la dimensione dell'immagine di \mathcal{L} è uguale al rango di \mathbf{A} . La tesi segue a questo punto dal teorema di nullità più rango per le matrici.

Il teorema di nullità più rango ha la seguente interpretazione geometrica. Il dominio \mathbf{V} è l'unione disgiunta delle fibre $\mathcal{L}^{-1}(\mathbf{w})$ al variare di \mathbf{w} nell'immagine di \mathcal{L} . Ciascuna di tali fibre è ottenuta per traslazione dalla fibra $\mathcal{L}^{-1}(\mathbf{0}_\mathbf{W}) = \text{Ker}(\mathcal{L})$ e ha perciò dimensione uguale alla dimensione del nucleo (la nullità). Per rendere l'idea, è come se \mathbf{V} si ottenessesse mettendo insieme tanti fogli paralleli (le fibre) e ciascun foglio corrispondesse esattamente a un vettore \mathbf{w} dell'immagine. Il teorema di nullità più rango afferma che la dimensione di \mathbf{V} è uguale alla dimensione delle fibre più la dimensione dello spazio che parametrizza le fibre. Per intuire perché questo dovrebbe essere vero si pensi al caso particolare dello spazio cartesiano, che ha dimensione 3, e che si può pensare come l'unione di piani paralleli parametrizzati da una retta. Per essere specifici, si consideri la proiezione dello spazio cartesiano \mathbb{R}^3 sull'asse x : l'immagine della proiezione è l'asse delle ascisse; il nucleo è il piano yz ; la fibra su un punto \mathbf{w} dell'asse delle ascisse è il piano parallelo al piano yz passante per w ; il dominio \mathbb{R}^3 è l'unione disgiunta dei piani paralleli al piano xy .

Come primo corollario del teorema di nullità più rango ricaviamo un criterio numerico per l'iniettività e la suriettività di un'applicazione lineare:

COROLLARIO 6.4 Sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare tra spazi di dimensione finita. Siano n la dimensione di \mathbf{V} , m la dimensione di \mathbf{W} ed r il rango di \mathcal{L} . Allora

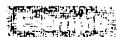
- a) \mathcal{L} è iniettiva se e solo se $r = n$;
- b) \mathcal{L} è suriettiva se e solo se $r = m$;
- c) \mathcal{L} è un isomorfismo se e solo se $r = n = m$.

DIMOSTRAZIONE. Mostriamo a): per il criterio di iniettività, \mathcal{L} è iniettiva se e solo se $\text{Ker}(\mathcal{L}) = \{\mathbf{0}_\mathbf{V}\}$, cioè se la nullità di \mathcal{L} è zero. Per il teorema di nullità più rango, la nullità è zero se e solo se il rango è uguale alla dimensione del dominio \mathbf{V} .

L'applicazione è suriettiva se e solo se l'immagine $\text{Im}(\mathcal{L})$ coincide con il codominio \mathbf{W} e, siccome $\text{Im}(\mathcal{L}) \subseteq \mathbf{W}$, questo equivale a dire che

$$\dim(\text{Im}(\mathcal{L})) = \dim(\mathbf{W})$$

Questo mostra b), e il punto c) segue da a) e b) perché un isomorfismo è un'applicazione lineare iniettiva e suriettiva.



Applichiamo il corollario per studiare grossolanamente la risolubilità di un sistema lineare di m equazioni in n incognite $\mathbf{Ax} = \mathbf{b}$. L'insieme delle soluzioni del sistema è la fibra di \mathcal{L}_A su \mathbf{b} . Quindi \mathcal{L}_A è iniettiva se e solo se per ogni $\mathbf{b} \in \mathbb{K}^m$ il sistema ammette al più una soluzione; \mathcal{L}_A è suriettiva se e solo se per ogni $\mathbf{b} \in \mathbb{K}^m$ il sistema ammette almeno una soluzione; \mathcal{L}_A è biunivoca se e solo se per ogni $\mathbf{b} \in \mathbb{K}^m$ il sistema ammette una e una sola soluzione. Dal corollario precedente segue:

- per ogni $\mathbf{b} \in \mathbb{K}^m$ il sistema ammette al più una soluzione se e solo se $r(\mathbf{A}) = n$;
- per ogni $\mathbf{b} \in \mathbb{K}^m$ il sistema ammette almeno una soluzione se e solo se $r(\mathbf{A}) = m$;
- per ogni $\mathbf{b} \in \mathbb{K}^m$ il sistema ammette una e una sola soluzione se e solo se $r(\mathbf{A}) = n = m$.

Possiamo dare una nuova dimostrazione del fatto fondamentale che una matrice quadrata è invertibile se e solo se ha rango massimo:

COROLLARIO 6.5 Una matrice quadrata \mathbf{A} di ordine n è invertibile se e solo se \mathbf{A} ha rango n .

DIMOSTRAZIONE. La matrice \mathbf{A} è invertibile se e solo se l'applicazione lineare $\mathcal{L}_A : \mathbb{K}^n \rightarrow \mathbb{K}^n$ è invertibile (se \mathbf{A} è invertibile, la funzione $\mathcal{L}_{\mathbf{A}^{-1}}$ è l'inversa di \mathcal{L}_A ; viceversa, se \mathcal{L}_A è invertibile, la matrice che rappresenta \mathcal{L}_A^{-1} rispetto alla base canonica è la matrice inversa di \mathbf{A}). Per il corollario 6.4 \mathcal{L}_A è invertibile se e solo se il suo rango è uguale a n . Infine il rango di \mathcal{L}_A è uguale a quello di \mathbf{A} . ■



La traccia $\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}$ di una matrice quadrata $\mathbf{A} = [a_{ij}]$ è un'applicazione lineare: $\mathbb{M}_{\mathbb{K}}(n, n) \rightarrow \mathbb{K}$. Determinarne il rango e la nullità.

Suggerimento: il rango è 1). Determinarne anche una base del nucleo (che è l'insieme delle matrici a traccia nulla).

Sia $\mathcal{L} : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ (rispettivamente, $\mathbb{R} \rightarrow \mathbb{R}^3$) un'applicazione iniettiva. Mostrare che $\text{Im}(\mathcal{L})$ è un piano (rispettivamente, una retta) per l'origine in \mathbb{R}^3 .

Sia $\mathcal{L} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ l'applicazione lineare definita da

$$\mathcal{L}([x, y, z]^T) = [x + 3y + 4z, 2x + y + 3z, -x + 2y + z]^T$$

- Trovare la matrice \mathbf{A} che rappresenta \mathcal{L} rispetto alla base canonica di \mathbb{R}^3 .
- Trovare la dimensione dello spazio immagine $\text{Im } \mathcal{L}$.
- Trovare una base di $\text{Ker } \mathcal{L}$.
- Per quali valori di h il vettore $[2, 3, h]^T$ appartiene all'immagine di \mathcal{L} ?
- L'applicazione lineare \mathcal{L} è iniettiva o suriettiva?

Sia $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$ la base canonica di \mathbb{R}^4 e sia $\mathcal{L} : \mathbb{R}^4 \rightarrow \mathbb{R}^4$ l'applicazione lineare tale che

$$\mathcal{L}(\mathbf{e}_1) = \mathbf{e}_1, \quad \mathcal{L}(\mathbf{e}_2) = \mathbf{0}, \quad \mathcal{L}(\mathbf{e}_3) = \mathbf{e}_2, \quad \mathcal{L}(\mathbf{e}_4) = \mathbf{e}_3$$

- Scrivere la matrice \mathbf{A} che rappresenta \mathcal{L} rispetto alla base canonica.
- Determinare la dimensione dell'immagine di \mathcal{L} e una base del nucleo di \mathcal{L} .
- Scrivere la matrice che rappresenta $\mathcal{L}^2 = \mathcal{L} \circ \mathcal{L}$ rispetto alla base canonica.
- Determinare la dimensione dell'immagine di \mathcal{L}^2 e una base del nucleo di \mathcal{L}^2 .
- Determinare la dimensione dell'immagine di \mathcal{L}^n e una base del nucleo di \mathcal{L}^n , per ogni $n \geq 3$.

Sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare suriettiva. Mostrare che per ogni sottospazio \mathbf{Z} di \mathbf{W} la controimmagine:

$$\mathcal{L}^{-1}(\mathbf{Z}) = \{\mathbf{v} \in \mathbf{V} : \mathcal{L}(\mathbf{v}) \in \mathbf{Z}\}$$

è un sottospazio di \mathbf{V} di dimensione $\dim \mathbf{Z} + \dim \text{Ker}(\mathcal{L})$. Come va modificata questa formula se \mathcal{L} non è suriettiva e \mathbf{Z} non è contenuto in $\text{Im}(\mathcal{L})$?

Siano $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ e $\mathfrak{M} : \mathbf{W} \rightarrow \mathbf{Z}$ applicazioni lineari tra spazi vettoriali di dimensione finita. Mostrare che

$$r(\mathfrak{M} \circ \mathcal{L}) \leq r(\mathcal{L}) \quad \text{e} \quad r(\mathfrak{M} \circ \mathcal{L}) \leq r(\mathfrak{M})$$

In particolare, $r(\mathfrak{M} \circ \mathcal{L}) \leq \min\{\dim \mathbf{V}, \dim \mathbf{W}, \dim \mathbf{Z}\}$. Inoltre, se \mathfrak{M} è un isomorfismo, allora $r(\mathfrak{M} \circ \mathcal{L}) = r(\mathcal{L})$ e se \mathcal{L} è un isomorfismo, allora $r(\mathfrak{M} \circ \mathcal{L}) = r(\mathfrak{M})$. Dedurne che, se \mathbf{A} e \mathbf{S} sono due matrici quadrate di ordine n e \mathbf{S} è invertibile, allora il rango di \mathbf{A} è uguale al rango di $\mathbf{S}^{-1} \mathbf{A} \mathbf{S}$.

Suggerimento: $\text{Im}(\mathfrak{M} \circ \mathcal{L}) = \mathfrak{M}(\text{Im}(\mathcal{L}))$.

Sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{V}$ un endomorfismo di uno spazio vettoriale \mathbf{V} di dimensione finita. Mostrare che

- per ogni $k \geq 1$ vale l'inclusione $\text{Ker}(\mathcal{L}^k) \subseteq \text{Ker}(\mathcal{L}^{k+1})$ ed esiste $N \geq 1$ tale che $\text{Ker}(\mathcal{L}^k) = \text{Ker}(\mathcal{L}^{k+1})$ per ogni $k \geq N$;
- per ogni $k \geq 1$ vale l'inclusione $\text{Im}(\mathcal{L}^{k+1}) \subseteq \text{Im}(\mathcal{L}^k)$ ed esiste $N \geq 1$ tale che $\text{Im}(\mathcal{L}^k) = \text{Im}(\mathcal{L}^{k+1})$ per ogni $k \geq N$.

Suggerimento: vi è un numero finito di possibilità per la dimensione di un sottospazio di \mathbf{V} .

Sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare tra spazi vettoriali di dimensione finita. Si dimostrò il teorema di nullità più rango seguendo la seguente traccia: si fissi una base $\{\mathbf{b}_1, \dots, \mathbf{b}_s\}$ del nucleo di \mathcal{L} e la si estenda a una base

$$\{\mathbf{b}_1, \dots, \mathbf{b}_s, \mathbf{c}_1, \dots, \mathbf{c}_t\}$$

di \mathbf{V} . Si mostri che $\{\mathcal{L}(\mathbf{c}_1), \dots, \mathcal{L}(\mathbf{c}_t)\}$ è una base di $\text{Im} \mathcal{L}$, quindi $r(\mathcal{L}) = t = \dim \mathbf{V} - \dim \text{Ker}(\mathcal{L})$. Questa dimostrazione è più semplice di quella presentata nel testo, ma non è costruttiva: non spiega come ricavare basi per nucleo e immagine.

Sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare tra spazi vettoriali di dimensione finita. Sia $\mathcal{B} = \{\mathbf{c}_1, \dots, \mathbf{c}_r, \mathbf{b}_1, \dots, \mathbf{b}_s\}$ una base di \mathbf{V} tale che $\{\mathbf{b}_1, \dots, \mathbf{b}_s\}$ sia una base del nucleo come nell'esercizio precedente. Sia \mathcal{C} una base di \mathbf{W} i cui primi r vettori siano la base $\{\mathcal{L}(\mathbf{c}_1), \dots, \mathcal{L}(\mathbf{c}_r)\}$ di $\text{Im} \mathcal{L}$. Si mostri che la matrice che rappresenta \mathcal{L} rispetto a tali basi è la matrice a blocchi

$$\mathbf{A} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

dove \mathbf{I}_r denota la matrice identità di ordine r ; cioè $a_{ii} = 1$ per $i = 1, \dots, r$ e gli altri coefficienti di \mathbf{A} sono tutti nulli. Concludere che, a meno di un cambiamento di basi, due applicazioni lineari coincidono se e solo se hanno lo stesso rango: in gergo matematico si dice che *il rango è l'unico invariante di un'applicazione lineare*.

Matrice e rango dell'applicazione duale Sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare. L'applicazione duale o trasposta di \mathcal{L} è la funzione $\mathcal{L}^* : \mathbf{W}^* \rightarrow \mathbf{V}^*$ così definita: un elemento di $\mathbf{W}^* = \text{Hom}_{\mathbb{K}}(\mathbf{W}, \mathbb{K})$ è una forma lineare $\gamma : \mathbf{W} \rightarrow \mathbb{K}$; a γ l'applicazione \mathcal{L}^* associa la forma lineare $\mathcal{L}^*(\gamma) : \mathbf{V} \rightarrow \mathbb{K}$ ottenuta componendo γ con \mathcal{L} :

$$\mathcal{L}^*(\gamma)(\mathbf{v}) = \gamma(\mathcal{L}(\mathbf{v}))$$

Si supponga che \mathbf{V} e \mathbf{W} abbiano dimensione finita e, fissate basi \mathcal{B} e \mathcal{C} di \mathbf{V} e \mathbf{W} rispettivamente, siano \mathcal{B}^* e \mathcal{C}^* le corrispondenti basi duali. Mostrare che

- a) se \mathbf{A} rappresenta \mathcal{L} rispetto a \mathcal{B} e \mathcal{C} , allora la matrice trasposta \mathbf{A}^T rappresenta \mathcal{L}^* rispetto a \mathcal{C}^* e \mathcal{B}^* ; in particolare, \mathcal{L} e \mathcal{L}^* hanno lo stesso rango;
- b) vale l'uguaglianza

$$\text{Ker}(\mathcal{L}^*) = \text{Im}(\mathcal{L})^\perp$$

Suggerimento: se γ ha coordinate \mathbf{b} rispetto a \mathcal{C}^* e \mathbf{v} ha coordinate \mathbf{x} rispetto a \mathcal{B} , allora per un esercizio precedente:

$$\mathcal{L}^*(\gamma)(\mathbf{v}) = \gamma(\mathcal{L}(\mathbf{v})) = \mathbf{b}^T(\mathbf{A}\mathbf{x}) = (\mathbf{A}^T\mathbf{b})^T\mathbf{x}$$

Questo mostra che il vettore delle coordinate di $\mathcal{L}^*(\gamma)$ rispetto alla base \mathcal{B}^* è $\mathbf{A}^T\mathbf{b}$, cioè che \mathbf{A}^T rappresenta \mathcal{L}^* rispetto a \mathcal{C}^* e \mathcal{B}^* .

Sia \mathbf{A} una matrice di tipo (m, n) di rango r . Si dica se le seguenti affermazioni sono vere o false, giustificando le risposte.

- a) Se $n = m$ ed esiste $\mathbf{b}_1 \in \mathbb{R}^n$ per cui il sistema $\mathbf{Ax} = \mathbf{b}_1$ ammette più di una soluzione, allora esiste $\mathbf{b}_2 \in \mathbb{R}^n$ per cui il sistema $\mathbf{Ax} = \mathbf{b}_2$ non ammette alcuna soluzione (vero).
- b) Se esiste $\mathbf{b}_1 \in \mathbb{R}^m$ per cui il sistema $\mathbf{Ax} = \mathbf{b}_1$ ammette infinite soluzioni, allora il sistema $\mathbf{Ax} = \mathbf{b}$ ammette infinite soluzioni per ogni $\mathbf{b} \in \mathbb{R}^m$ (falso).
- c) Se $n = m$ ed esiste $\mathbf{b}_1 \in \mathbb{R}^n$ per cui il sistema $\mathbf{Ax} = \mathbf{b}_1$ non ammette alcuna soluzione, allora esiste $\mathbf{b}_2 \in \mathbb{R}^n$ per cui il sistema $\mathbf{Ax} = \mathbf{b}_2$ ammette infinite soluzioni (vero).
- d) Se esiste $\mathbf{b}_1 \in \mathbb{R}^m$ per cui il sistema $\mathbf{Ax} = \mathbf{b}_1$ ammette esattamente una soluzione, allora il sistema $\mathbf{Ax} = \mathbf{b}$ ammette esattamente una soluzione per ogni $\mathbf{b} \in \mathbb{R}^m$ (falso).

6

Determinante

1 INTRODUZIONE

Nel primo capitolo abbiamo già definito il determinante di matrici 2×2 e 3×3 a coefficienti reali. Nel caso 2×2 il valore assoluto del determinante è l'area del parallelogramma generato dai vettori riga della matrice (il parallelogramma che ha un vertice nell'origine e i cui lati uscenti dall'origine sono i vettori riga della matrice). Nel caso 3×3 il determinante è il prodotto misto dei vettori riga della matrice e, quindi, il suo valore assoluto è il volume del parallelepipedo generato dai vettori riga della matrice (il parallelepipedo che ha un vertice nell'origine e i cui spigoli uscenti dall'origine sono i vettori riga della matrice). In entrambi i casi il determinante è diverso da zero se e solo se le righe della matrice sono linearmente indipendenti. È naturale chiedersi se si possa trovare un numero con proprietà analoghe per matrici quadrate di ordine arbitrario: intuitivamente ci aspettiamo che n vettori in \mathbb{R}^n siano linearmente indipendenti se e solo se il parallelepipedo da essi generato ha volume diverso da zero (cioè non è schiacciato in un sottospazio proprio di \mathbb{R}^n). La risposta è positiva: a ogni matrice quadrata \mathbf{A} si può associare un numero, $\det(\mathbf{A})$, che dipende dalle righe di \mathbf{A} come dovrebbe dipendere il *volume con segno del parallelepipedo generato dalle righe di una matrice*, con la proprietà fondamentale che $\det(\mathbf{A}) \neq 0$ se e solo se le righe di \mathbf{A} sono linearmente indipendenti. Per arrivare a trovare una formula esplicita per il determinante, occorre precisare le proprietà che il *volume con segno del parallelepipedo generato dalle righe di una matrice* dovrebbe avere, e si mostra che esiste un'unica funzione con tali proprietà: faremo vedere che il determinante è l'unica funzione delle righe della matrice che sia (a) *multilineare*, cioè lineare in ciascuna riga fissate le altre, (b) *alternante* (il determinante cambia segno se si scambiano due righe), (c) *normalizzata* in modo da valere 1 per la matrice identità (geometricamente, 1 è il volume del cubo di lato unitario, che è il parallelepipedo generato dai vettori della base canonica).

In questo modo arriveremo a scrivere una formula esplicita per il determinante in termini degli elementi della matrice e anche una formula di calcolo del determinante per induzione (sviluppo di Laplace), che riconduce il calcolo del determinante di una matrice $n \times n$ a una combinazione lineare di n determinanti di matrici $(n-1) \times (n-1)$. Tuttavia queste formule hanno un costo computazionale altissimo anche per valori di

n relativamente piccoli. Pertanto, l'interesse del determinante risiede prevalentemente nelle sue applicazioni teoriche e l'attenzione del lettore dovrebbe rivolgersi più alle proprietà che non ai metodi di calcolo. Un esempio da tener presente è l'elegante formula di Cramer, secondo la quale le componenti della soluzione $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ di un sistema lineare $\mathbf{Ax} = \mathbf{b}$, con \mathbf{A} quadrata non singolare, sono:

$$x_i = \frac{\det(\mathbf{A}_i)}{\det(\mathbf{A})}$$

dove \mathbf{A}_i è ottenuta sostituendo la colonna i di \mathbf{A} col termine noto \mathbf{b} . Questa formula è tanto elegante quanto computazionalmente inutile per calcolare la soluzione: utilizzando la formula di Cramer e gli sviluppi di Laplace per il calcolo del determinante anche il computer più veloce impiegherebbe decine di anni per risolvere un sistema con un numero relativamente piccolo di equazioni come 25, mentre col Meg (e altri algoritmi) la soluzione si troverebbe in un batter d'occhio. Se oggi è possibile risolvere sistemi lineari con centinaia di migliaia di incognite non è certo per merito della formula di Cramer.

Per quel che riguarda questo libro, useremo il determinante principalmente per ottenere informazioni sugli autovalori di una matrice. Da questo punto di vista, le proprietà fondamentali che ci interessano sono che (a) il determinante è diverso da zero se e solo se la matrice è non singolare, (b) il determinante è un polinomio omogeneo di grado n (l'ordine della matrice) negli elementi della matrice, (c) il determinante di un prodotto di matrici è il prodotto dei determinanti: $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$.

Questo capitolo contiene molti dettagli tecnici su cui il lettore non dovrebbe perdere. Consiglio, pertanto, di saltare, almeno in prima lettura, tutte le dimostrazioni e di concentrarsi sull'apprendimento delle proprietà del determinante, con l'eccezione della prima metà della dimostrazione del teorema 4.5 sull'esistenza del determinante: è interessante vedere come dalle proprietà astratte di multilinearità e alternanza si possa ricavare la formula analitica del determinante.

2 DETERMINANTE E MOSSE DI GAUSS

Come abbiamo fatto per il rango, introduciamo il determinante facendo riferimento al metodo di climinazione di Gauss: a parte il segno, il determinante è il prodotto dei pivots della matrice; mentre il segno dipende dal numero di scambi di righe che il MEG richiede. Il risultato preciso, che dimostreremo nei paragrafi successivi, è:

TEOREMA 2.1 (Determinante e MEG)

Esiste un'unica funzione

$$\det : \mathbb{M}_{\mathbb{K}}(n, n) \rightarrow \mathbb{K}$$

con le seguenti proprietà:

- a) Invarianza per scorrimento: se \mathbf{B} è ottenuta da \mathbf{A} sommando a una riga un multiplo di un'altra riga, $\det(\mathbf{B}) = \det(\mathbf{A})$.
- b) Alternanza: se \mathbf{B} è ottenuta da \mathbf{A} scambiando due righe, allora $\det(\mathbf{B}) = -\det(\mathbf{A})$.

- c) Normalizzazione: se \mathbf{U} è una matrice triangolare alta, $\det(\mathbf{U})$ è il prodotto degli elementi sulla diagonale principale di \mathbf{U} :

$$(2.1) \quad \det \begin{pmatrix} p_1 & * & * & * & * \\ 0 & p_2 & * & * & * \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & 0 & . & p_n \end{pmatrix} = p_1 p_2 \cdots p_n$$

L'unicità della funzione determinante segue dal fatto che le proprietà a), b) e c) consentono il calcolo del determinante di una matrice \mathbf{A} : col metodo di Gauss si riduce la matrice a una matrice triangolare \mathbf{U} ; per le proprietà a) e b) il determinante di \mathbf{A} è $\pm \det(\mathbf{U})$; per la proprietà c) il determinante di \mathbf{U} è il prodotto dei pivots. Il punto del teorema è che non è affatto evidente che una funzione con le proprietà richieste esista: scambiando le righe di \mathbf{A} in modi diversi si ottengono matrici triangolari diverse, e non è evidente che non cambi anche il prodotto dei pivots di tali matrici. Rimandiamo la dimostrazione dell'esistenza della funzione determinante ai paragrafi successivi.

Il determinante di una matrice si indica anche sostituendo le parentesi della matrice con delle barre verticali: per esempio

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

OSSERVAZIONE Il determinante, quando $\mathbb{K} = \mathbb{R}$, va pensato come il *volume con segno del parallelepipedo che ha un vertice nell'origine e, come spigoli uscenti dall'origine, i vettori riga della matrice*. Spieghiamo questa affermazione nel caso di \mathbb{R}^3 . In questo caso, come già visto nel primo capitolo, il segno del determinante è positivo se e solo se i vettori riga della matrice formano una base destrorsa. Esaminiamo le tre proprietà dell'enunciato del teorema 2.1. La prima proprietà esprime l'invarianza del volume di un parallelepipedo quando se ne lasciano fissi la base e l'altezza; per vederlo, prendiamo come base del parallelepipedo il parallelogramma generato dalla seconda e dalla terza riga della matrice. La prima riga è quindi lo spigolo del parallelepipedo che esce dal piano della base; ora un multiplo della seconda o della terza riga è un vettore parallelo alla base e aggiungendo un vettore parallelo alla base alla prima riga non si modifica l'altezza del parallelogramma rispetto alla base; il volume rimane perciò invariato. Per quanto riguarda la proprietà b), scambiare due spigoli lascia invariato il parallelepipedo e, in particolare, non varia il volume ma trasforma una terna destrorsa in sinistrorsa e viceversa, per cui il segno del determinante deve cambiare. Consideriamo, infine, la proprietà c): supponiamo quindi che la matrice sia triangolare alta:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}$$

La terza riga è un vettore dell'asse z , la seconda un vettore del piano yz , la cui proiezione sull'asse y è a_{22} . La base del parallelepipedo, che è il parallelogramma generato dalla seconda e terza riga, è contenuta nel piano yz e ha la base lunga $|a_{33}|$ e altezza $|a_{22}|$. L'area di base è quindi $|a_{22}a_{33}|$. L'altezza del parallelepipedo è il valore assoluto della proiezione a_{11} della prima riga sull'asse x . Perciò il volume del parallelepipedo è il valore assoluto del prodotto degli elementi sulla diagonale principale. Dal punto di vista geometrico, le operazioni di riga trasformano il parallelepipedo mantenendone invariato il volume, in modo tale che alla fine del processo il calcolo del volume sia immediato.



Calcoliamo il determinante della matrice

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

Il determinante non cambia sommando a una riga un multiplo di un'altra riga:

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

Scambiando la seconda e la terza riga il determinante cambia di segno:

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

Completando il processo di eliminazione otteniamo infine

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

Possiamo procedere in ordine diverso. Per esempio, possiamo prima scambiare la prima e la terza riga:

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

e poi proseguire con l'eliminazione:

$$- \begin{vmatrix} 1 & 2 & 3 & -1 \\ 1 & 1 & 3 & 2 \\ 1 & 1 & 2 & 1 \\ 2 & 3 & 1 & -1 \end{vmatrix} = - \begin{vmatrix} 1 & 2 & 3 & -1 \\ 0 & -1 & 0 & 3 \\ 0 & -1 & -1 & 2 \\ 0 & -1 & -5 & 1 \end{vmatrix} = - \begin{vmatrix} 1 & 2 & 3 & -1 \\ 0 & -1 & 0 & 3 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & -5 & -2 \end{vmatrix} = - \begin{vmatrix} 1 & 2 & 3 & -1 \\ 0 & -1 & 0 & 3 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 3 \end{vmatrix} = -3$$

Il teorema 2.1 garantisce che in qualunque modo si proceda nella riduzione a matrice triangolare si ottiene sempre lo stesso risultato -3 per il determinante di \mathbf{A} .

Dal teorema 2.1 si deduce immediatamente la proprietà fondamentale del determinante: il determinante di una matrice quadrata è diverso da zero se e solo se le righe della matrice sono linearmente indipendenti. Intuitivamente questo è chiaro: le n righe sono indipendenti se e solo se lo spazio che generano ha dimensione n e questo equivale a dire che il volume del parallelepipedo generato dalle righe è diverso da zero: si pensi al caso di \mathbb{R}^3 , in cui un parallelepipedo è schiacciato in un piano se e solo se ha volume nullo. Tenuto conto che le righe di una matrice $n \times n$ sono linearmente indipendenti se e solo se la matrice ha rango n , la proprietà cruciale del determinante si esprime come segue:

TEOREMA 2.2 (Matrici non singolari)

Per una matrice quadrata \mathbf{A} di ordine n le seguenti affermazioni sono equivalenti:

- 1) $r(\mathbf{A}) = n$;
- 2) $\det(\mathbf{A}) \neq 0$;
- 3) \mathbf{A} è invertibile.

DIMOSTRAZIONE. Abbiamo già dimostrato che \mathbf{A} è invertibile se e solo se ha rango massimo, quindi basta mostrare che $r(\mathbf{A}) = n$ se e solo se $\det(\mathbf{A}) \neq 0$. Col MEG riduciamo \mathbf{A} alla matrice triangolare \mathbf{U} . La matrice \mathbf{U} ha lo stesso rango di \mathbf{A} e, a meno del segno, lo stesso determinante. Quindi basta dimostrare l'enunciato per \mathbf{U} . Per finire osserviamo che il rango di \mathbf{U} è n se e solo se tutti gli elementi sulla diagonale principale di \mathbf{U} sono non nulli; e questo succede se e solo se il prodotto di tali elementi, che è il determinante di \mathbf{U} , è diverso da zero. □

COROLLARIO 2.3 (Annullamento del determinante)

Il determinante di \mathbf{A} è nullo in ciascuno dei seguenti casi:

- 1) una riga o una colonna è nulla;
- 2) due righe o due colonne sono uguali;
- 3) una riga è combinazione lineare delle altre righe (o una colonna è combinazione lineare delle altre colonne).

DIMOSTRAZIONE. In ciascuno dei casi le righe (o le colonne) di \mathbf{A} sono linearmente dipendenti e quindi \mathbf{A} non ha rango massimo. □

Il determinante della matrice

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 1 \\ 3 & 2 & 4 \end{bmatrix}$$

è nullo perché la terza riga è la somma delle prime due.

Calcolare il determinante delle matrici quadrate che nei capitoli precedenti sono state ridotte a forma triangolare.

Suggerimento: il determinante è \pm il prodotto dei pivots.

Mostrare che, se \mathbf{A} è una matrice quadrata di ordine n e t è uno scalare, allora $\det(t\mathbf{A}) = t^n \det(\mathbf{A})$.

Suggerimento: moltiplicando \mathbf{A} per t si moltiplica ogni pivot di \mathbf{A} per t). Se \mathbf{A} ha ordine 3 e $\det(\mathbf{A}) = -2$, quanto vale $\det(3\mathbf{A})$?

Se \mathbf{v} è un vettore di \mathbb{K}^n , il prodotto $\mathbf{v}\mathbf{v}^T$ è una matrice $n \times n$ di rango ≤ 1 (esattamente 1 se $\mathbf{v} \neq 0$). Concludere che $\det(\mathbf{v}\mathbf{v}^T) = 0$. Più in generale, se \mathbf{B} è una matrice di tipo (n, p) con $p < n$, mostrare che $\mathbf{B}\mathbf{B}^T$ è una matrice $n \times n$ di rango $\leq p$. Concludere che $\det(\mathbf{B}\mathbf{B}^T) = 0$. Verificare questa formula nel caso $\mathbf{B} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$.

3 DETERMINANTE DI MATRICI DI PERMUTAZIONE

Questo paragrafo è dedicato a mostrare che la richiesta che *scambiando due righe il determinante cambi di segno* non conduce a una contraddizione. Si tratta di un fatto tecnico che il lettore dovrebbe dare per buono, almeno in prima battuta. Facciamo un esempio. Sia \mathbf{P} la matrice ottenuta dalla matrice identità 3×3 scambiando la prima e la terza riga:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \mathbf{P}$$

La matrice \mathbf{P} si può anche ottenere da \mathbf{I} scambiando inizialmente le prime due righe, poi la prima e la terza riga e, infine, la seconda e la terza riga

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \mathbf{P}$$

Questo mostra che \mathbf{P} si può ottenere da \mathbf{I} con un solo scambio di righe, e anche con tre scambi di righe. Siccome 1 e 3 sono entrambi dispari, in entrambi i casi la nostra regola impone $\det(\mathbf{P}) = -1$ e non c'è problema. Però se \mathbf{P} si potesse ottenere da \mathbf{I} con un numero *pari* di scambi di righe, la regola imporrebbe $\det(\mathbf{P}) = 1$, una contraddizione. Quindi occorre escludere che una stessa matrice si possa ottenere da \mathbf{I} sia con un numero *pari* sia con un numero *dispari* di scambi di righe. Se il lettore è disposto ad accettare questo fatto, può saltare il resto del paragrafo.

La matrice \mathbf{P} è un esempio di *matrice di permutazione*. Una *permutazione* dei numeri da 1 a n è un ordinamento dell'insieme $\{1, \dots, n\}$. Per esempio, ci sono sei permutazioni dei numeri da 1 a 3:

$$(1, 2, 3), \quad (1, 3, 2), \quad (2, 1, 3), \quad (2, 3, 1), \quad (3, 1, 2), \quad (3, 2, 1)$$

Possiamo pensare a una permutazione $\sigma = (\sigma(1), \sigma(2), \dots, \sigma(n))$ anche come a una funzione biiettiva $\sigma : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$. Per la permutazione $\sigma = (2, 3, 1)$ abbiamo $\sigma(1) = 2$, $\sigma(2) = 3$ e $\sigma(3) = 1$. Il fatto che la funzione sia biiettiva significa che ogni intero tra 1 e n compare una e una sola volta nella lista $\sigma(1), \sigma(2), \dots, \sigma(n)$. In generale, le permutazioni dei numeri da 1 a n sono $n!$ (un numero enorme) perché possiamo scegliere il primo intero in n modi diversi, poi il secondo in $n - 1$ modi perché il primo numero è stato fissato, il terzo in $n - 2$ modi perché i primi due numeri sono stati fissati, eccetera; quindi il numero delle permutazioni, che corrisponde a tutte le scelte possibili, è

$$n \cdot (n - 1) \cdot (n - 2) \cdots 2 \cdot 1 = n!$$

L'insieme di tutte le permutazioni di $\{1, \dots, n\}$ si denota col simbolo S_n . Uno *scambio* è una permutazione che scambia tra loro due numeri e lascia gli altri invariati. Fissati $1 \leq i < j \leq n$, lo scambio $\epsilon = [i, j]$ è la permutazione che scambia i e j . Quindi $\epsilon(i) = j$, $\epsilon(j) = i$ e $\epsilon(k) = k$ per $k \neq i, j$, ovvero

$$\epsilon = (1, \dots, i - 1, j, i + 1, \dots, j - 1, i, j + 1, \dots, n)$$

Per esempio, fissato $n = 5$, lo scambio $[2, 3]$ è la permutazione $(1, 3, 2, 4, 5)$ che scambia tra loro 2 e 3, e lascia fissi 1, 4 e 5.

Data una coppia di permutazioni σ e τ in S_n , la permutazione prodotto $\sigma\tau$ si ottiene applicando prima τ e poi σ (attenzione all'ordine dei fattori; altri autori usano la convenzione opposta):

$$\sigma\tau(k) = \sigma(\tau(k)) \quad \text{per ogni } k = 1, \dots, n$$

Se pensiamo a σ e τ come a funzioni $\{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$, il prodotto $\sigma\tau$ non è altro che la funzione composta. Per esempio, fissiamo $n = 3$ e calcoliamo il prodotto dei due scambi $\sigma = [1, 3] = (3, 2, 1)$ e $\tau = [2, 3] = (1, 3, 2)$: siccome $\sigma(\tau(1)) = \sigma(1) = 3$, $\sigma(\tau(2)) = \sigma(3) = 1$, $\sigma(\tau(3)) = \sigma(2) = 2$, la permutazione prodotto è

$$\sigma\tau = (3, 1, 2)$$

PROPOSIZIONE 3.1 (Parità di una permutazione)

- a) Ogni permutazione σ è il prodotto di un certo numero di scambi.
- b) Se una permutazione σ è il prodotto di un numero pari (rispettivamente dispari di scambi), allora ogni altra decomposizione di σ come prodotto di scambi ha un numero pari (rispettivamente dispari) di fattori.

DIMOSTRAZIONE. Il fatto che σ si possa ottenere come prodotto di scambi è evidente: se $\sigma(1) \neq 1$, con lo scambio $[1, \sigma(1)]$ si mette $\sigma(1)$ come primo elemento, poi con un altro scambio (se necessario) si mette $\sigma(2)$ al secondo posto e procedendo così con al massimo $n-1$ scambi si trasforma l'ordinamento naturale nella permutazione $\sigma = (\sigma(1), \sigma(2), \dots, \sigma(n))$. Per esempio, per ottenere $(2, 3, 1)$ dall'ordinamento naturale $(1, 2, 3)$, possiamo scambiare prima 1 e 2 ottenendo $(2, 1, 3)$, e poi 1 e 3.

Per mostrare il punto b), introduciamo la nozione di *inversione* di una permutazione: data una permutazione $\sigma = (\sigma(1), \sigma(2), \dots, \sigma(n))$, una coppia di indici (i, j) compresi tra 1 e n è un'inversione di σ se

- $i < j$
- $\sigma(i) > \sigma(j)$

Computazionalmente è lungo (oltre che per fortuna inutile) determinare quali o quante siano le inversioni di una permutazione: occorre infatti controllare tutte le coppie (i, j) con $1 \leq i < j \leq n$ e tali coppie sono $\binom{n}{2} = \frac{1}{2}n(n - 1)$. Per esempio, per trovare le inversioni della permutazione $(\sigma(1), \sigma(2), \sigma(3)) = (2, 3, 1)$ occorre controllare le 3 coppie $(1, 2)$, $(1, 3)$ e $(2, 3)$. La permutazione σ manda queste tre coppie rispettivamente in $(2, 3)$, $(2, 1)$ e $(3, 1)$. Quindi le inversioni di σ sono le due coppie $(1, 3)$ e $(2, 3)$.

Il punto cruciale è l'unica cosa che ci interessa, è che il prodotto di uno scambio ϵ con una permutazione σ cambia la parità del numero di inversioni di σ ; questo significa:

se ϵ è uno scambio e σ ha un numero pari (rispettivamente dispari) di inversioni, allora $\epsilon\sigma$ ha un numero dispari (rispettivamente pari) di inversioni.

Da questa affermazione segue subito il punto b). Supponiamo infatti che una permutazione σ si scriva come un prodotto di d scambi. La permutazione identità $(1, 2, \dots, n)$ ha zero inversioni, quindi un numero pari di inversioni. Per ipotesi σ si ottiene dalla permutazione identità con d scambi; la parità del numero di inversioni viene cambiata d volte, e quindi σ ha un numero pari (rispettivamente dispari) di inversioni se e solo se d è pari (rispettivamente dispari). Siccome il numero di inversioni dipende solo da σ e non dal modo in cui decomponiamo σ in un prodotto di scambi, il punto b) è dimostrato.

Rimane da mostrare che il prodotto a sinistra per uno scambio ϵ cambia la parità del numero di inversioni di σ . Poniamo $\tau = \epsilon\sigma$. Sia D_1 l'insieme delle coppie (i, j) che sono inversioni di σ ma non di τ e sia D_2 l'insieme delle coppie (i, j) che sono inversioni di τ ma non di σ . Se d_1 (rispettivamente d_2) è il numero di elementi di D_1 (rispettivamente D_2), allora la differenza tra il numero di inversioni di σ e il numero di inversioni di τ è $d_1 - d_2$. Occorre quindi dimostrare che $d_1 - d_2$ è dispari. Siccome $d_1 - d_2 = (d_1 + d_2) - 2d_2$, basta mostrare che $d_1 + d_2$ è dispari. Ora $d_1 + d_2$ è il numero di elementi di $D = D_1 \cup D_2$, cioè delle coppie che sono inversioni per σ o τ , ma non per entrambe.

A questo punto è semplice mostrare che D contiene un numero dispari di coppie. Lo scambio ϵ scambia due numeri $\sigma(a)$ e $\sigma(b)$, con $1 \leq a < b \leq n$. Per una coppia (i, j) con i e j entrambi distinti da a e da b si ha $(\sigma(i), \sigma(j)) = (\tau(i), \tau(j))$, quindi la coppia è un'inversione per σ se e solo se lo è per τ e non appartiene D . Quindi le coppie in D hanno almeno un elemento uguale ad a e b . Consideriamo tutte le possibilità.

1. Coppia del tipo (i, a) con $i < a$. Viene mandata da σ in $(\sigma(i), \sigma(a))$ e da τ in $(\sigma(i), \sigma(b))$. Appartiene a D se e solo se $\sigma(i)$ è compreso tra $\sigma(a)$ e $\sigma(b)$.
2. Coppia del tipo (i, b) con $i < a$. Viene mandata da σ in $(\sigma(i), \sigma(b))$ e da τ in $(\sigma(i), \sigma(a))$. Appartiene a D se e solo se $\sigma(i)$ è compreso tra $\sigma(a)$ e $\sigma(b)$.
3. Coppia del tipo (a, i) con $a < i < b$. Viene mandata da σ in $(\sigma(a), \sigma(i))$ e da τ in $(\sigma(b), \sigma(i))$. Appartiene a D se e solo se $\sigma(i)$ è compreso tra $\sigma(a)$ e $\sigma(b)$.
4. Coppia del tipo (i, b) con $a < i < b$. Viene mandata da σ in $(\sigma(i), \sigma(b))$ e da τ in $(\sigma(i), \sigma(a))$. Appartiene a D se e solo se $\sigma(i)$ è compreso tra $\sigma(a)$ e $\sigma(b)$.

5. Coppia del tipo (a, i) con $i > b$. Viene mandata da σ in $(\sigma(a), \sigma(i))$ e da τ in $(\sigma(b), \sigma(i))$. Appartiene a D se e solo se $\sigma(i)$ è compreso tra $\sigma(a)$ e $\sigma(b)$.
6. Coppia del tipo (b, i) con $i > b$. Viene mandata da σ in $(\sigma(b), \sigma(i))$ e da τ in $(\sigma(a), \sigma(i))$. Appartiene a D se e solo se $\sigma(i)$ è compreso tra $\sigma(a)$ e $\sigma(b)$.
7. Coppia (a, b) . Viene mandata da σ in $(\sigma(a), \sigma(b))$ e da τ in $(\sigma(b), \sigma(a))$. Appartiene a D .

Osserviamo che una coppia (i, a) di tipo 1 appartiene a D se e solo se la coppia (i, b) di tipo 2 appartiene a D ; quindi D contiene un numero pari di coppie di tipo 1 o 2; analogamente D contiene un numero pari di coppie di tipo 3 o 4, e un numero pari di coppie di tipo 5 o 6. In più D contiene la singola coppia (a, b) . Quindi D contiene un numero dispari di coppie, e questo conclude la dimostrazione. 

DEFINIZIONE 3.2 (Permutazioni pari e dispari)

Una permutazione σ si dice pari (rispettivamente dispari) se è il prodotto di un numero pari (rispettivamente dispari) di scambi. Si pone $(-1)^\sigma = 1$ se σ è pari, $(-1)^\sigma = -1$ se σ è dispari.

La proposizione precedente mostra che una permutazione non può essere pari e dispari allo stesso tempo.



La permutazione identità è pari. Uno scambio è una permutazione dispari. La permutazione

$$\sigma = (2, 3, \dots, n, 1)$$

è pari se n è dispari e dispari se n è pari, perché

$$\sigma = [1, n-1] \cdots [1, 3][1, 2]$$

è il prodotto di $n-1$ scambi.

A una permutazione σ di $\{1, 2, \dots, n\}$ associamo, come nel capitolo sull'algebra delle matrici, la matrice di permutazione \mathbf{P}_σ : si tratta della matrice di tipo (n, n) , ottenuta dalla matrice identità permutando le righe come prescritto da σ ; questo significa che la riga i di \mathbf{P}_σ è la riga $\sigma(i)$ della matrice identità. Quindi, se $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ è la base canonica di \mathbb{K}^n , la riga i di \mathbf{P}_σ è $\mathbf{e}_{\sigma(i)}^T$.

Per esempio

$$\mathbf{P}_{(2,3,1)} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Le matrici \mathbf{P}_σ si dicono *matrici di permutazione*. Una matrice di permutazione \mathbf{P}_σ ha esattamente un 1 su ogni riga e su ogni colonna e tutti gli altri elementi della matrice sono nulli; l'elemento non nullo della riga i è quello con indice di colonna $\sigma(i)$. Viceversa, ogni matrice che abbia esattamente un 1 su ogni riga e su ogni colonna e tutti gli altri elementi nulli, è una matrice di permutazione \mathbf{P}_σ : la permutazione σ

si trova ponendo $\sigma(i)$ uguale all'indice di colonna dell'unico 1 dell' i -esima riga. Per esempio, la matrice

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

definisce $\sigma = (1, 3, 2)$.

Se la permutazione σ si decompone come prodotto di d scambi, la matrice \mathbf{P}_σ si ottiene dalla matrice identità con d scambi di righe e quindi per la proprietà di alternanza b) del teorema 2.1 il determinante di \mathbf{P}_σ dev'essere $(-1)^d$ (il determinante della matrice identità è 1 per la proprietà c) di normalizzazione). Per la proposizione 3.1 questo numero è ben definito e non dipende dalla decomposizione di σ come prodotto di scambi: è uguale a 1 se σ è pari, a -1 se σ è dispari. Abbiamo così calcolato il determinante per le matrici di permutazione:

PROPOSIZIONE 3.3 (Determinante di una matrice di permutazione)

Supponiamo che esista la funzione determinante con le proprietà richieste dal teorema 2.1. Allora il determinante della matrice di permutazione \mathbf{P}_σ è uguale a 1 se σ è pari e a -1 se σ è dispari:

$$(3.1) \quad \det(\mathbf{P}_\sigma) = (-1)^\sigma$$

Stabilire se le seguenti permutazioni sono pari o dispari:

$$\sigma_1 = (5, 4, 3, 2, 1), \sigma_2 = (2, 1, 4, 3, 5), \sigma_3 = (2, 3, 1, 5, 4), \sigma_4 = (1, 2, 3, 5, 4)$$

Scrivere le corrispondenti matrici di permutazione (sono matrici 5×5) e calcolarne il determinante.

Calcolare il determinante della matrice

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Scrivere la permutazione σ tale che $\mathbf{P} = \mathbf{P}_\sigma$ e dire se σ è pari o dispari.

Mostrare che, se $\mathbf{A} = \mathbf{LU}$ è una fattorizzazione LU (rispettivamente $\mathbf{A} = \mathbf{PLU}$ con \mathbf{P} matrice di permutazione), allora $\det(\mathbf{A}) = \det(\mathbf{L}) \det(\mathbf{U}) = \det(\mathbf{U})$ (rispettivamente $\det(\mathbf{A}) = \det(\mathbf{P}) \det(\mathbf{L}) \det(\mathbf{U})$).

Mostrare che ogni permutazione σ ha un'inversa σ^{-1} definita dalla proprietà

$$\sigma(i) = j \iff i = \sigma^{-1}(j)$$

Equivalentemente, l'inversa è l'unica permutazione σ^{-1} tale che il prodotto $\sigma^{-1}\sigma$ sia la permutazione identità. Mostrare che (a) se ϵ è uno scambio, $\epsilon^{-1} = \epsilon$; (b) σ^{-1} è pari se e solo se σ è pari (se σ è scritta come prodotto di scambi, σ^{-1} è il prodotto degli stessi scambi in ordine inverso); (c) se \mathbf{P} è la matrice di permutazione associata a σ , allora \mathbf{P}^T è la matrice di permutazione associata a σ^{-1} . Concludere che per ogni matrice di permutazione \mathbf{P} valgono le formule

$$\mathbf{P}^{-1} = \mathbf{P}^T, \quad \det(\mathbf{P}^T) = \det(\mathbf{P})$$

Data $\sigma = (5, 4, 3, 2, 1)$, scrivere σ^{-1} , la matrice \mathbf{P}_σ e la sua inversa.

4 FORMULA ESPPLICITA PER IL DETERMINANTE

In questo paragrafo dimostriamo la parte difficile del teorema 2.1, cioè l'esistenza della funzione determinante, ricavandone una formula esplicita, che ora anticipiamo. Il lettore è però avvertito che, tranne che per i casi $n = 2$ e $n = 3$, si tratta di una formula inutilizzabile per il calcolo effettivo del determinante; l'interesse della formula è per le applicazioni teoriche. La formula è

DEFINIZIONE 4.1 (Formula analitica del determinante)

Sia $\mathbf{A} = [a_{ij}]$ una matrice $n \times n$. Il determinante di \mathbf{A} è lo scalare

$$(4.1) \quad \det(\mathbf{A}) = \sum_{\sigma \in S_n} (-1)^\sigma a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}$$

Vedremo nella dimostrazione del teorema 4.5 quale sia l'origine di questa formula. Il simbolo $\sum_{\sigma \in S_n}$ significa che la sommatoria a secondo membro di (4.1) ha un addendo per ogni permutazione σ di $\{1, 2, \dots, n\}$ (il simbolo S_n denota l'insieme di tali permutazioni). Ci sono quindi $n!$ addendi in questa sommatoria, il che rende la formula inutile in pratica se n non è molto piccolo. Esplicitiamo la formula per $n = 2$ e $n = 3$, gli unici due casi in cui la si può utilizzare nei conti.

Quando $n = 2$, le permutazioni sono solo due, l'identità $(1, 2)$ che è pari, e lo scambio $(2, 1)$ che è dispari. Quindi la formula (4.1) diventa:

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

in accordo con la formula dei primi capitoli per il determinante delle matrici 2×2 .

Quando $n = 3$, ci sono 3 permutazioni pari $(1, 2, 3)$, $(2, 3, 1)$ e $(3, 1, 2)$, e 3 permutazioni dispari (scambi) $(3, 2, 1)$, $(1, 3, 2)$ e $(2, 1, 3)$. La formula (4.1) è in questo caso nota come *regola di Sarrus*:

$$(4.2) \quad \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} + \\ - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}$$

Anche questa formula coincide con quella introdotta nel primo capitolo per il determinante di una matrice 3×3 e, quindi, in questo caso il determinante è il prodotto misto dei vettori riga della matrice.

Per ricordare la regola di Sarrus si può procedere così: si aggiungono alla matrice una quarta e una quinta colonna uguali alla prima e alla seconda; nella formula del determinante i termini col segno $+$ si ottengono moltiplicando lungo le diagonali dall'alto verso il basso:

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

mentre i termini col segno $-$ si ottengono moltiplicando lungo le diagonali dal basso verso l'alto:

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



Calcoliamo con la regola di Sarrus il determinante della matrice $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 1 \\ 3 & 2 & 4 \end{bmatrix}$:

$$\begin{vmatrix} 1 & 2 & 3 \\ 2 & 0 & 1 \\ 3 & 2 & 4 \end{vmatrix} = 1 \times 0 \times 4 + 2 \times 1 \times 3 + 3 \times 2 \times 2 + \\ - 3 \times 0 \times 3 - 2 \times 1 \times 1 - 4 \times 2 \times 2 = \\ = 6 + 12 - 2 - 16 = 0$$

Già per $n = 4$ la formula (4.1) ha $4! = 24$ termini e, per una matrice generica, è inutilizzabile. Ciononostante la formula ha molte applicazioni teoriche. Cominciamo qui con l'osservare che il determinante è un *polinomio omogeneo di grado n* negli elementi della matrice. Più precisamente, a parte il segno $(-1)^\sigma = \pm 1$, ogni addendo di (4.1) è un monomio

$$a_{1\sigma(1)}a_{2\sigma(2)} \cdots a_{n\sigma(n)}$$

di grado n nelle variabili a_{ij} . Inoltre, in questo prodotto compare un unico fattore per ogni riga e per ogni colonna: il fattore sulla riga i è $a_{i\sigma_i}$, il fattore sulla colonna j è $a_{\sigma^{-1}(j), j}$. In altre parole, il prodotto $a_{1\sigma(1)}a_{2\sigma(2)} \cdots a_{n\sigma(n)}$ si ottiene scegliendo un elemento $a_{i\sigma_i}$ per ogni riga, avendo cura che gli elementi scelti stiano su colonne distinte, e facendo poi il prodotto di tali elementi. Il determinante di $[a_{ij}]$, in quanto somma di monomi di grado n , è un polinomio *omogeneo* di grado n in a_{ij} . In particolare:

$$(4.3) \quad \det(t\mathbf{A}) = t^n \det(\mathbf{A})$$

per ogni scalare $t \in \mathbb{K}$: se moltiplichiamo tutti gli elementi della matrice per t , il determinante viene moltiplicato per t^n .

Per dimostrare il teorema 2.1 è opportuno considerare il determinante come una funzione delle righe di una matrice: per questo scriviamo $D(\mathbf{v}_1, \dots, \mathbf{v}_n)$ per denotare il determinante della matrice che ha per righe i vettori $\mathbf{v}_1^T, \dots, \mathbf{v}_n^T$ di \mathbb{K}^n :

$$D(\mathbf{v}_1, \dots, \mathbf{v}_n) = \det \left(\begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} \right)$$

Le proprietà a) e b) del teorema 2.1 in queste nuove notazioni divengono:

- a) **Invarianza per scorrimento**: il determinante non cambia se a una riga si aggiunge un multiplo di un'altra riga:

$$(4.4) \quad D(\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_k + t\mathbf{v}_j, \dots, \mathbf{v}_n) = D(\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_k, \dots, \mathbf{v}_n)$$

per ogni $j, k \in \{1, \dots, n\}$ con $j \neq k$.

- b) **Alternanza**: il determinante cambia segno se si scambiano due righe:

$$(4.5) \quad D(\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n) = -D(\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_i, \dots, \mathbf{v}_n)$$

per ogni $1 \leq i < j \leq n$.

Per ricavare la formula (4.1), cominciamo col mostrare che la funzione D è lineare in ciascuna delle sue variabili \mathbf{v}_k ; si tratta di una delle proprietà fondamentali del determinante, e merita un'apposita definizione.

DEFINIZIONE 4.2 (Funzione multilinearare)

Sia \mathbf{V} uno spazio vettoriale sul campo \mathbb{K} . Una funzione

$$D = D(\mathbf{v}_1, \dots, \mathbf{v}_n) : \mathbf{V} \times \dots \times \mathbf{V} \rightarrow \mathbb{K}$$

(cioè una funzione che a una n -upla di vettori di \mathbf{V} associa uno scalare) si dice multilinear se, fissate arbitrariamente tutte le variabili tranne una, D è una funzione lineare della rimanente variabile:

$$\begin{aligned} D(\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, t\mathbf{v} + u\mathbf{w}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n) &= \\ &= tD(\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n) + uD(\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{w}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n) \end{aligned}$$

OSSERVAZIONE Più in generale si può definire la nozione di multilinearità per una funzione $\mathbf{F}(\mathbf{v}_1, \dots, \mathbf{v}_n)$ in cui le variabili \mathbf{v}_k appartengano a spazi vettoriali \mathbf{V}_k non necessariamente coincidenti, e a valori in uno spazio vettoriale \mathbf{W} , purché il campo degli scalari \mathbb{K} sia lo stesso per tutti gli spazi vettoriali coinvolti. Quando $n = 2$, si parla di funzione bilineare.

Di norma i prodotti sono esempi di funzioni bilineari. Il prodotto scalare nello spazio euclideo è un esempio di una funzione bilineare $\mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}$. Il prodotto vettoriale in \mathbb{R}^3 è una funzione bilineare $\mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^3$ a valori vettoriali; con il determinante ha in comune l'alternanza: scambiando i due fattori, il prodotto vettoriale cambia di segno. Il prodotto $\mathbf{A}_1 \mathbf{A}_2 \dots \mathbf{A}_n$ di n matrici $m \times m$ è una funzione multilineare.

PROPOSIZIONE 4.3 (Il determinante è una funzione multilineare)

Supponiamo che esista la funzione determinante $D(\mathbf{v}_1, \dots, \mathbf{v}_n)$ del teorema 2.1.

Allora $D(\mathbf{v}_1, \dots, \mathbf{v}_n)$ è multilineare.

DIMOSTRAZIONE. Basta dimostrare la linearità nella prima variabile; la linearità in una variabile successiva k segue dalla proprietà di alternanza b) scambiando la variabile k con la prima. Per dimostrare che una funzione è lineare, occorre mostrare che è additiva e omogenea. Cominciamo col mostrare che il determinante è una funzione omogenea della prima riga:

$$(4.6) \quad D(t\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = tD(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$$

Per questo osserviamo che, per il teorema 2.1, il determinante è il prodotto dei pivots della matrice, e che, quando si moltiplica la prima riga per t , il pivot della prima riga viene moltiplicato per t . Mostriamo ora che il determinante è una funzione additiva nella prima riga:

$$(4.7) \quad D(\mathbf{v} + \mathbf{w}, \mathbf{v}_2, \dots, \mathbf{v}_n) = D(\mathbf{v}, \mathbf{v}_2, \dots, \mathbf{v}_n) + D(\mathbf{w}, \mathbf{v}_2, \dots, \mathbf{v}_n)$$

Se $\mathbf{v}_2, \dots, \mathbf{v}_n$ sono linearmente dipendenti, allora entrambi i membri di (4.7) sono nulli, e quindi uguali, per la proprietà di annullamento del determinante. Possiamo quindi supporre che l'insieme $\{\mathbf{v}_2, \dots, \mathbf{v}_n\}$ sia linearmente indipendente. Possiamo allora aggiungere un vettore \mathbf{v}_1 a questo insieme e ottenere una base di \mathbb{K}^n e scrivere \mathbf{v} e \mathbf{w} come combinazioni lineari dei vettori della base:

$$\mathbf{v} = t\mathbf{v}_1 + \sum_{j=2}^n t_j \mathbf{v}_j, \quad \mathbf{w} = u\mathbf{v}_1 + \sum_{j=2}^n u_j \mathbf{v}_j$$

Allora

$$\mathbf{v} + \mathbf{w} = (t+u)\mathbf{v}_1 + \sum_{j=2}^n (t_j + u_j) \mathbf{v}_j.$$

Osserviamo che, usando la (4.4) e la (4.6), otteniamo l'uguaglianza

$$D(\mathbf{v}, \mathbf{v}_2, \dots, \mathbf{v}_n) = D\left(t\mathbf{v}_1 + \sum_{j=2}^n t_j \mathbf{v}_j, \mathbf{v}_2, \dots, \mathbf{v}_n\right) = D(t\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = tD(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$$

Analogamente

$$D(\mathbf{w}, \mathbf{v}_2, \dots, \mathbf{v}_n) = uD(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n), \quad D(\mathbf{v} + \mathbf{w}, \mathbf{v}_2, \dots, \mathbf{v}_n) = (t+u)D(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$$

Ora (4.7) segue dalla proprietà distributiva del campo \mathbb{K} :

$$(t_1 + u_1)D(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = t_1D(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) + u_1D(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$$

Per ricavare la formula per il determinante, serve ancora una proprietà del determinante: il determinante è nullo se due righe sono uguali. Abbiamo già mostrato che questa proprietà di annullamento è vera se il determinante soddisfa le richieste del teorema 2.1. Però per lo sviluppo logico della teoria occorre mostrare che questa proprietà di annullamento è conseguenza delle sole proprietà di (1) multilinearità e (2) alternanza:

LEMMA 4.4 (Annullamento)

Supponiamo che $D(\mathbf{v}_1, \dots, \mathbf{v}_n)$ sia una funzione multilineare e alternante. Se due dei vettori $\mathbf{v}_1, \dots, \mathbf{v}_n$ coincidono, cioè se esistono due indici $i < j$ con $\mathbf{v}_i = \mathbf{v}_j$, allora

$$D(\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_i, \dots, \mathbf{v}_n) = 0$$

DIMOSTRAZIONE. Supponiamo che $i < j$ e $\mathbf{v}_i = \mathbf{v}_j$. Per la proprietà di alternanza

$$D(\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n) = -D(\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_i, \dots, \mathbf{v}_n)$$

D'altra parte, siccome $\mathbf{v}_i = \mathbf{v}_j$, vale anche l'uguaglianza

$$D(\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n) = D(\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_i, \dots, \mathbf{v}_n)$$

Sommendo le due uguaglianze troviamo

$$2D(\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n) = 0$$

e quindi $D(\mathbf{v}_1, \dots, \mathbf{v}_n) = 0$.

Abbiamo ora a disposizione tutte le proprietà del determinante che servono per ricavarne l'espressione esplicita in funzione degli elementi della matrice. Dal punto di vista algebrico, la multilinearità è più semplice da utilizzare dell'invarianza per scorrimento. Dimostriamo, quindi, prima una versione più algebrica (e standard) del teorema 2.1: il determinante è l'unica funzione multilineare alternante che vale 1 per la matrice identità.

TEOREMA 4.5 (Caratterizzazione algebrica del determinante)

Il determinante

$$(4.8) \quad \det([a_{ij}]) = \sum_{\sigma \in S_n} (-1)^\sigma a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}$$

è l'unica funzione $M_{\mathbb{K}}(n, n) \rightarrow \mathbb{K}$ che

- i) sia multilineare e alternante come funzione delle righe della matrice;
- ii) valga 1 per la matrice identità (normalizzazione).

DIMOSTRAZIONE. Scriviamo come prima $\det(\mathbf{A}) = D(\mathbf{v}_1, \dots, \mathbf{v}_n)$ dove $\mathbf{v}_1, \dots, \mathbf{v}_n$ sono le righe della matrice \mathbf{A} . Nella prima parte della dimostrazione facciamo vedere che, se D

è una funzione multilineare alternante che vale 1 sulla matrice identità, allora D è data dalla formula (4.8). Nella seconda parte mostreremo che la funzione definita da (4.8) ha le proprietà richieste.

Si noti che la normalizzazione richiesta in questo teorema è molto più debole di quella richiesta dal teorema 2.1: basta che il determinante della matrice identità sia 1, in simboli

$$1 = \det(\mathbf{I}) = D(\mathbf{e}_1, \dots, \mathbf{e}_n)$$

dove $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ denota come al solito la base canonica di \mathbb{K}^n . Questa richiesta, insieme alla proprietà di alternanza, forza $\det \mathbf{P}_\sigma = (-1)^\sigma$ come spiegato nel paragrafo precedente, perché \mathbf{P}_σ si ottiene dalla matrice identità con un numero pari o dispari di scambi di righe a seconda che σ sia pari o dispari.

Per ricavare la formula, scriviamo i vettori riga della matrice come combinazioni lineari dei vettori della base canonica. Per la prima riga abbiamo

$$\mathbf{v}_1 = a_{11}\mathbf{e}_1 + a_{12}\mathbf{e}_2 + \cdots + a_{1n}\mathbf{e}_n = \sum_{j_1=1}^n a_{1j_1}\mathbf{e}_{j_1}$$

Usiamo la notazione j_1 al posto di j perché poi avremo bisogno di un indice j_2 per la seconda riga, di un indice j_3 per la terza fino ad arrivare a j_n per l'ultima riga. Dalla linearità del determinante nella prima riga segue:

$$(4.9) \quad D(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = D\left(\sum_{j_1=1}^n a_{1j_1}\mathbf{e}_{j_1}, \mathbf{v}_2, \dots, \mathbf{v}_n\right) = \sum_{j_1=1}^n a_{1j_1}D(\mathbf{e}_{j_1}, \mathbf{v}_2, \dots, \mathbf{v}_n)$$

Ripetendo questo argomento per le righe successive troviamo

$$(4.10) \quad D(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = \sum_{j_1=1}^n \sum_{j_2=1}^n \cdots \sum_{j_n=1}^n a_{1j_1}a_{2j_2} \cdots a_{nj_n} D(\mathbf{e}_{j_1}, \mathbf{e}_{j_2}, \dots, \mathbf{e}_{j_n})$$

Questa somma ha $n \times n \times \cdots \times n = n^n$ addendi. Per fortuna molti di questi addendi sono nulli: per il lemma di annullamento $D(\mathbf{e}_{j_1}, \mathbf{e}_{j_2}, \dots, \mathbf{e}_{j_n}) = 0$ se due indici sono uguali; quindi gli unici addendi non nulli sono quelli per cui gli indici j_k sono tutti distinti e costituiscono una permutazione σ :

$$(4.11) \quad (j_1, \dots, j_n) = (\sigma(1), \dots, \sigma(n))$$

Possiamo quindi limitare la somma agli addendi che corrispondono alle permutazioni σ :

$$(4.12) \quad D(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = \sum_{\sigma \in S_n} a_{1\sigma(1)}a_{2\sigma(2)} \cdots a_{n\sigma(n)} D(\mathbf{e}_{\sigma(1)}, \mathbf{e}_{\sigma(2)}, \dots, \mathbf{e}_{\sigma(n)})$$

La matrice di permutazione P_σ ha come k -esima riga il vettore $\mathbf{e}_{\sigma(k)}$, quindi

$$(4.13) \quad D(\mathbf{e}_{\sigma(1)}, \mathbf{e}_{\sigma(2)}, \dots, \mathbf{e}_{\sigma(n)}) = \det(\mathbf{P}_\sigma) = (-1)^\sigma$$

Sostituendo in (4.12) troviamo la formula dell'enunciato

$$(4.14) \quad \det(\mathbf{A}) = \sum_{\sigma \in S_n} (-1)^\sigma a_{1\sigma(1)}a_{2\sigma(2)} \cdots a_{n\sigma(n)}$$

Questo conclude la dimostrazione dell'unicità. Occorre dimostrare ancora che la funzione \det così definita è multilineare e alternante nelle righe della matrice, e vale 1 per la matrice identità.

Cominciamo a mostrare $\det(\mathbf{I}) = 1$. Per la matrice identità, l'unico elemento non nullo sulla riga i è quello con indice di colonna i , e $a_{ii} = 1$. Quindi, nella formula del determinante l'unico addendo non nullo è quello che corrisponde alla permutazione identità $(\sigma(1), \dots, \sigma(n)) = (1, \dots, n)$. Siccome la permutazione identità è pari e $a_{ii} = 1$ per ogni i , la formula del determinante assegna alla matrice identità il numero 1.

Supponiamo che la matrice \mathbf{B} sia ottenuta da \mathbf{A} scambiando la riga i con la riga j . Allora $b_{hk} = a_{hk}$ se $h \neq i, j$, mentre $b_{ik} = a_{jk}$ e $b_{jk} = a_{ik}$. Quindi fissata una permutazione σ

$$\begin{aligned} a_{1\sigma(1)} \cdots a_{i\sigma(i)} \cdots a_{j\sigma(j)} \cdots a_{n\sigma(n)} &= b_{1\sigma(1)} \cdots b_{j\sigma(i)} \cdots b_{i\sigma(j)} \cdots b_{n\sigma(n)} = \\ &= b_{1\tau(1)} \cdots b_{i\tau(i)} \cdots b_{j\tau(j)} \cdots b_{n\tau(n)} \end{aligned}$$

dove $\tau = \sigma\epsilon$ e ϵ è la permutazione che scambia i e j in modo che $\tau(i) = \sigma(j)$ e $\tau(j) = \sigma(i)$. Osserviamo che $(-1)^\tau = -(-1)^\sigma$ perché τ è il prodotto di σ con uno scambio. Inoltre, al variare di σ in S_n la permutazione $\tau = \sigma\epsilon$ assume una e una sola volta tutti i valori possibili: la funzione $S_n \rightarrow S_n$ che a σ associa $\tau = \sigma\epsilon$ è biiettiva (l'inversa associa a τ la permutazione $\tau\epsilon$). Quindi

$$\begin{aligned} \det(\mathbf{A}) &= \sum_{\sigma \in S_n} (-1)^\sigma a_{1\sigma(1)} \cdots a_{i\sigma(i)} \cdots a_{j\sigma(j)} \cdots a_{n\sigma(n)} = \\ &= - \sum_{\tau \in S_n} (-1)^\tau b_{1\tau(1)} \cdots b_{i\tau(i)} \cdots b_{j\tau(j)} \cdots b_{n\tau(n)} = -\det(\mathbf{B}) \end{aligned}$$

Questo mostra che la funzione det definita da (4.8) è alternante. Rimane da dimostrare la multilinearità. Siccome la funzione è alternante, basta mostrare che è lineare nella prima riga. Dalla formula è evidente che il determinante è un polinomio omogeneo di primo grado nelle componenti a_{1j} della prima riga. Sappiamo che un polinomio omogeneo di primo grado nelle componenti di un vettore di \mathbb{K}^n è lineare. Quindi il determinante è lineare nella prima riga, e questo conclude la dimostrazione. 

Possiamo finalmente dimostrare il teorema 2.1:

DIMOSTRAZIONE (del teorema 2.1). Dobbiamo dimostrare che la funzione det definita dalla (4.8) verifica le tre proprietà di invarianza per scorrimento, alternanza e normalizzazione (per le matrici triangolari alte) del teorema 2.1, e che è l'unica funzione a verificarle. L'unicità, come abbiamo già osservato, segue dal fatto che queste tre proprietà consentono di calcolare il determinante mediante riduzione a scala della matrice. Rimane da mostrare che la funzione det verifica le tre proprietà. Abbiamo già verificato l'alternanza nel teorema precedente. Per quanto riguarda l'invarianza per scorrimento, basta dimostrarla per la prima riga: l'invarianza per un'altra riga segue scambiando tale riga con la prima e usando l'alternanza. Ora, se aggiungiamo alla prima riga \mathbf{v}_1^T un multiplo $t\mathbf{v}_j$ della riga j (con $j > 1$), il determinante della matrice così ottenuta è:

$$\begin{aligned} D(\mathbf{v}_1 + t\mathbf{v}_j, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n) &= && \text{(linearità nella prima riga)} \\ &= D(\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n) + tD(\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n) = && \text{(annullamento)} \\ &= D(\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n) \end{aligned}$$

Questo mostra la proprietà di invarianza per scorrimento; si osservi che per la prima uguaglianza abbiamo usato la linearità del determinante nella prima riga, fatto che abbiamo dimostrato nel teorema precedente; la seconda uguaglianza segue dal lemma di annullamento 4.4, che possiamo usare perché sappiamo che la funzione det definita dalla (4.8) è multilineare e alternante.

Dobbiamo infine mostrare che il determinante di una matrice triangolare alta è il prodotto degli elementi sulla diagonale principale. Una matrice triangolare alta $[a_{ij}]$ ha tutti gli elementi sotto la diagonale principale nulli, cioè $a_{ij} = 0$ per $i > j$; nella somma (4.8) affinché un termine

$$a_{1\sigma(1)}a_{2\sigma(2)} \cdots a_{n\sigma(n)}$$

sia non nullo occorre che $i \leq \sigma(i)$ per ogni $i = 1, 2, \dots, n$. Ma allora $n \leq \sigma(n)$, quindi $\sigma(n) = n$. Poi $n-1 \leq \sigma(n-1)$, quindi $\sigma(n-1)$ può essere solo $n-1$ oppure n ; ma non può essere n perché $\sigma(n) = n$, quindi $\sigma(n-1) = n-1$. Proseguendo così si vede $\sigma(i) = i$ per ogni i : l'unica permutazione per cui il prodotto $a_{1\sigma(1)}a_{2\sigma(2)} \cdots a_{n\sigma(n)}$ può essere diverso da zero è l'identità, per cui dalla formula (4.8) segue

$$\det(\mathbf{A}) = a_{11}a_{22} \cdots a_{nn}$$

come volevasi dimostrare.

Come prima applicazione della formula per il determinante, mostriamo che il determinante non cambia se trasponiamo la matrice:

PROPOSIZIONE 4.6 (Determinante della matrice trasposta)

Per ogni matrice quadrata \mathbf{A}

$$\det(\mathbf{A}^T) = \det(\mathbf{A})$$

DIMOSTRAZIONE. Occorre osservare che la formula (4.8) è simmetrica rispetto alle righe e alle colonne di \mathbf{A} ; per questo ricordiamo che ogni permutazione σ ha un'inversa σ^{-1} , con la proprietà

$$\sigma(i) = j \Leftrightarrow i = \sigma^{-1}(j)$$

L'inversa è anche l'unica permutazione tale che il prodotto $\sigma^{-1}\sigma$ sia la permutazione identità. Se ϵ è uno scambio, allora $\epsilon^{-1} = \epsilon$. Se σ si decompone come prodotto di scambi

$$\sigma = \epsilon_1 \dots \epsilon_d$$

allora

$$\sigma^{-1} = \epsilon_d \dots \epsilon_1$$

In particolare σ e σ^{-1} hanno la stessa parità:

$$(-1)^\sigma = (-1)^d = (-1)^{\sigma^{-1}}$$

Il generico termine della formula (4.8) si può allora riscrivere nella forma:

$$(-1)^\sigma a_{1\sigma(1)}a_{2\sigma(2)} \cdots a_{n\sigma(n)} = (-1)^{\sigma^{-1}} a_{\sigma^{-1}(1)1}a_{\sigma^{-1}(2)2} \cdots a_{\sigma^{-1}(n)n}$$

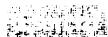
(si osservi che abbiamo permuto i fattori a secondo membro scrivendo per primo quello sulla prima colonna: $a_{\sigma^{-1}(1)1} = a_{\sigma^{-1}(1)\sigma(\sigma^{-1}(1))}$). La funzione che a una permutazione $\sigma \in S_n$ associa la sua inversa $\tau = \sigma^{-1}$ è biiettiva perché ogni permutazione $\tau \in S_n$ è l'inversa di una e una sola permutazione $\sigma = \tau^{-1}$: quindi sommare al variare di $\sigma \in S_n$ è la stessa cosa che

sommare al variare di $\sigma^{-1} \in S_n$, e

$$\begin{aligned}\det(\mathbf{A}) &= \sum_{\sigma \in S_n} (-1)^\sigma a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} = \\ &= \sum_{\sigma \in S_n} (-1)^{\sigma^{-1}} a_{\sigma^{-1}(1)1} a_{\sigma^{-1}(2)2} \cdots a_{\sigma^{-1}(n)n} = \\ &= \sum_{\tau \in S_n} (-1)^\tau a_{\tau(1)1} a_{\tau(2)2} \cdots a_{\tau(n)n} = \det(\mathbf{A}^T)\end{aligned}$$

dove l'ultimo passaggio è la formula del determinante applicata a \mathbf{A}^T , il cui elemento di posto (i, j) è a_{ji} .

Come conseguenza, tutto quanto visto per le righe vale anche per le colonne: in particolare il determinante di una matrice è una funzione multilineare alternante delle colonne; il determinante si annulla se e solo se le colonne della matrice sono linearmente dipendenti; il determinante di una matrice triangolare *bassa* è il prodotto degli elementi sulla diagonale principale.



Data la matrice $\mathbf{A} = \begin{bmatrix} 3 & 1 & -3 \\ 1 & 3 & 0 \\ 1 & 2 & -1 \end{bmatrix}$, usare la regola di Sarrus per calcolare $\det(\mathbf{A})$ e $\det(\mathbf{A}^T)$ e verificare che sono uguali. Calcolare anche i determinanti delle matrici $\mathbf{A}^T \mathbf{A}$, \mathbf{A}^2 e $\mathbf{A} \mathbf{A}^T$ e verificare che sono tutti uguali a $\det(\mathbf{A})^2$ (questo segue dalla formula di Binet $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$ che dimostreremo più avanti).

Verificare che $\det(\mathbf{P}_\sigma) = (-1)^\sigma$ usando la formula (4.1).

Suggerimento: nella sommatoria solo un termine è diverso da zero.

Spiegare perché

$$\begin{vmatrix} 3 & 27 & 4 \\ 1 & 0 & 2 \\ 5 & 36 & 6 \end{vmatrix} = 27 \begin{vmatrix} 3 & 1 & 4 \\ 1 & 0 & 2 \\ 5 & 0 & 6 \end{vmatrix} + 36 \begin{vmatrix} 3 & 0 & 4 \\ 1 & 0 & 2 \\ 5 & 1 & 6 \end{vmatrix}$$

e

$$\begin{vmatrix} 3 & 27 & 4 \\ 1 & 0 & 2 \\ 5 & 36 & 6 \end{vmatrix} = \begin{vmatrix} 3 & 27 & 4 \\ 1 & 0 & 0 \\ 5 & 36 & 6 \end{vmatrix} + 2 \begin{vmatrix} 3 & 27 & 4 \\ 0 & 0 & 1 \\ 5 & 26 & 6 \end{vmatrix}$$

Suggerimento: il determinante è lineare nelle colonne e nelle righe della matrice.

Calcolare il determinante della matrice $\begin{bmatrix} 3 & 0 & 0 & 0 \\ -4 & 1 & 0 & 0 \\ -5 & 26 & 2 & 0 \\ 1 & 3 & -7 & -2 \end{bmatrix}$

Mostrare che, per $n \geq 2$, il determinante $\det(\mathbf{A})$ non è una funzione lineare di \mathbf{A} : trovare due matrici tali che $\det(\mathbf{A} + \mathbf{B}) \neq \det(\mathbf{A}) + \det(\mathbf{B})$. Mostrare anche che $\det(t\mathbf{A}) = t^n \det(\mathbf{A})$ è in generale diverso da $t \det(\mathbf{A})$.

Sia \mathbf{V} uno spazio vettoriale su \mathbb{R} e sia $B = B(\mathbf{v}, \mathbf{w}) : \mathbf{V} \times \mathbf{V} \rightarrow \mathbb{R}$ una funzione bilineare. Mostrare che B è alternante, cioè $B(\mathbf{v}, \mathbf{w}) = -B(\mathbf{w}, \mathbf{v})$ per ogni $\mathbf{v}, \mathbf{w} \in \mathbf{V}$, se e solo se $B(\mathbf{u}, \mathbf{u}) = 0$ per ogni $\mathbf{u} \in \mathbf{V}$.

Suggerimento: siccome B è bilineare,

$$B(\mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w}) = B(\mathbf{v}, \mathbf{v}) + B(\mathbf{v}, \mathbf{w}) + B(\mathbf{w}, \mathbf{v}) + B(\mathbf{w}, \mathbf{w})$$

5 SVILUPPI DI LAPLACE

Nel primo capitolo abbiamo introdotto il determinante di una matrice 3×3 mediante la formula

$$(5.1) \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Dal punto di vista geometrico questo mostra che il determinante è il prodotto misto dei tre vettori riga della matrice. Dal punto di vista algebrico, che è quello che ci interessa ora, la formula consente di ridurre il calcolo del determinante di una matrice 3×3 al calcolo di una combinazione lineare di 3 determinanti di matrici 2×2 . Per una matrice $n \times n$ il determinante ammette ancora uno *sviluppo di Laplace*

$$\det(\mathbf{A}) = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}$$

perché il determinante è una funzione lineare degli elementi della prima riga; il coefficiente C_{11} si trova raccogliendo a_{11} nella formula (4.8) del determinante, C_{12} raccogliendo a_{12} e così via. A parte il segno, vedremo che C_{1j} è il determinante della sottomatrice di \mathbf{A} ottenuta cancellando la prima riga e la j -esima colonna. Per questo diamo la seguente definizione:

DEFINIZIONE 5.1 (Complemento algebrico)

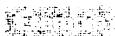
Sia $\mathbf{A} = [a_{ij}]$ una matrice quadrata di ordine n . Per ogni coppia (i, j) con $1 \leq i, j \leq n$, sia \mathbf{A}_{ij} la sottomatrice quadrata di \mathbf{A} ottenuta cancellando la riga i e la colonna j di \mathbf{A} ; il determinante

$$C_{ij} = (-1)^{i+j} \det(\mathbf{A}_{ij})$$

si dice *complemento algebrico o cofattore* dell'elemento a_{ij} .

Il segno $(-1)^{i+j}$ da anteporre a $\det(\mathbf{A}_{ij})$ è alternativamente 1 e -1 ; lo si visualizza con la matrice

$$\begin{bmatrix} + & - & + & \cdots \\ - & + & - & \cdots \\ + & - & + & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$



Se $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ è una matrice 2×2 , la sottomatrice ottenuta cancellando la prima riga e la prima colonna consiste dell'unico elemento a_{22} , per cui $C_{11} = a_{22}$; cancellando la prima riga e la seconda colonna si trova $\mathbf{A}_{12} = a_{21}$, per cui $C_{12} = -a_{21}$. Osserviamo che

$$\det(\mathbf{A}) = a_{11}a_{22} - a_{12}a_{21} = a_{11}C_{11} + a_{12}C_{12}$$

Per calcolare il complemento algebrico C_{11} di a_{11} in una matrice 3×3

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

occorre cancellare la prima riga e la prima colonna ottenendo la matrice

$$\mathbf{A}_{11} = \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}$$

e poi calcolare il determinante di \mathbf{A}_{11} :

$$C_{11} = \det(\mathbf{A}_{11}) = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = a_{22}a_{33} - a_{23}a_{32}$$

Analogamente

$$C_{12} = -\det(\mathbf{A}_{12}) = -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} = -(a_{21}a_{33} - a_{23}a_{31})$$

e

$$C_{13} = \det(\mathbf{A}_{13}) = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} = a_{21}a_{32} - a_{22}a_{31}$$

La formula (5.1) si può quindi riscrivere nella forma

$$\det(\mathbf{A}) = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13}$$

TEOREMA 5.2 (Sviluppi di Laplace)

Sia \mathbf{A} una matrice quadrata di ordine n e sia C_{ij} il complemento algebrico dell'elemento a_{ij} . Allora

a) fissato un indice di riga i , vale la formula

$$\det(\mathbf{A}) = \sum_{j=1}^n a_{ij}C_{ij} = \sum_{j=1}^n (-1)^{i+j}a_{ij}\det(\mathbf{A}_{ij})$$

che si dice *sviluppo di Laplace del determinante rispetto alla riga i*;

b) fissato un indice di colonna j , vale la formula

$$\det(\mathbf{A}) = \sum_{i=1}^n a_{ij} C_{ij} = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det(\mathbf{A}_{ij})$$

che si dice *sviluppo di Laplace del determinante rispetto alla colonna j*.

DIMOSTRAZIONE. La prima riga della matrice è

$$\mathbf{v}_1^T = [a_{11} \quad a_{12} \quad \dots \quad a_{1n}] = \sum_{j=1}^n a_{1j} \mathbf{e}_j^T$$

Per la linearità nella prima riga del determinante

$$(5.2) \quad \det(\mathbf{A}) = D(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = \sum_{j=1}^n a_{1j} D(\mathbf{e}_j, \mathbf{v}_2, \dots, \mathbf{v}_n)$$

Ora calcoliamo $D(\mathbf{e}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$: per la proprietà di invarianza per scorrimento il determinante non cambia se alla riga i sommiamo la prima riga moltiplicata per $-a_{i1}$, per cui

$$D(\mathbf{e}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = \begin{vmatrix} 1 & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = \begin{vmatrix} 1 & 0 & \dots & 0 \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & a_{n2} & \dots & a_{nn} \end{vmatrix}$$

Osserviamo ora che

$$\begin{vmatrix} 1 & 0 & \dots & 0 \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & a_{n2} & \dots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots \\ a_{n2} & \dots & a_{nn} \end{vmatrix} = \det(\mathbf{A}_{11})$$

l'uguaglianza a sinistra si ottiene calcolando il determinante delle due matrici con il metodo di eliminazione di Gauss come prodotto dei pivots.

Per calcolare il secondo addendo di (5.2) osserviamo che

$$D(\mathbf{e}_2, \mathbf{v}_2, \dots, \mathbf{v}_n) = \begin{vmatrix} 0 & 1 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = - \begin{vmatrix} 1 & 0 & 0 & \dots & 0 \\ a_{22} & a_{21} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n2} & a_{n1} & a_{n3} & \dots & a_{nn} \end{vmatrix}$$

dove nell'ultimo passaggio abbiamo scambiato la prima e la seconda colonna, il che produce un cambio di segno del determinante. Per quanto abbiamo appena dimostrato

$$\begin{vmatrix} 1 & 0 & \dots & 0 \\ a_{22} & a_{11} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n2} & a_{n1} & \dots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & \dots & a_{2n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} = \det(\mathbf{A}_{12})$$

Analogamente, per calcolare

$$D(\mathbf{e}_j, \mathbf{v}_2, \dots, \mathbf{v}_n) = \begin{vmatrix} 0 & 0 & 1 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nj} & \cdots & a_{nn} \end{vmatrix}$$

facciamo $j-1$ scambi di colonna per ottenere una matrice con le stesse colonne ma nell'ordine $(j, 1, 2, \dots, j-1, j+1, \dots, n)$: per questo scambiamo la colonna j con la colonna $j-1$, poi la nuova colonna $j-1$ (la colonna j originale) con la colonna $j-2$, e così via fino a scambiare la colonna 2 (la colonna j originale) con la prima colonna. Quindi

$$D(\mathbf{e}_j, \mathbf{v}_2, \dots, \mathbf{v}_n) = (-1)^{j-1} \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{2j} & a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{nj} & a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = (-1)^{1+j} \det(\mathbf{A}_{1j}) = C_{1j}$$

Sostituendo nella (5.2) troviamo lo sviluppo di Laplace rispetto alla prima riga

$$(5.3) \quad \det(\mathbf{A}) = \sum_{j=1}^n a_{1j} C_{1j}$$

Lo sviluppo di Laplace rispetto a un'altra riga i di \mathbf{A} si riconduce a quello della prima riga permutando opportunamente le righe di \mathbf{A} : sia \mathbf{B} la matrice che ha le stesse righe di \mathbf{A} , prese però nell'ordine

$$(i, 1, \dots, i-1, i+1, \dots, n)$$

La prima riga di \mathbf{B} è la riga i di \mathbf{A} , le righe successive di \mathbf{B} sono le righe di \mathbf{A} dalla prima all'ultima saltando la riga i . La matrice \mathbf{B} si ottiene da \mathbf{A} con $i-1$ scambi di righe: si porta la riga i nella prima riga scambiandola con la riga precedente e ripetendo il procedimento $i-1$ volte. La matrice \mathbf{B}_{1j} ottenuta da \mathbf{B} cancellando la prima riga e la j -esima colonna coincide con la matrice \mathbf{A}_{ij} . Possiamo ora dedurre lo sviluppo di Laplace del determinante di \mathbf{A} rispetto alla riga i :

$$\begin{aligned} \det(\mathbf{A}) &= (-1)^{i-1} \det(\mathbf{B}) = (-1)^{i-1} \sum_{j=1}^n (-1)^{1+j} b_{1j} \det(\mathbf{B}_{1j}) = \\ &= \sum_{j=1}^n (-1)^{i+j} a_{ij} \det(\mathbf{A}_{1j}) \end{aligned}$$

la prima uguaglianza vale perché \mathbf{A} e \mathbf{B} differiscono per $i-1$ scambi di righe: la seconda è lo sviluppo di Laplace di $\det(\mathbf{B})$ rispetto alla prima riga.

Infine lo sviluppo di Laplace di $\det(\mathbf{A})$ rispetto alla colonna j si ottiene sviluppando $\det(\mathbf{A}^T)$ rispetto alla riga j e ricordano l'uguaglianza $\det(\mathbf{A}) = \det(\mathbf{A}^T)$.

Lo sviluppo di Laplace è molto conveniente se fatto rispetto a una riga (o a una colonna) di \mathbf{A} che contenga molti zeri: Per esempio, sviluppando il seguente determinante rispetto alla prima colonna, troviamo

$$\begin{vmatrix} 5 & a_2 & a_3 \\ 0 & b_2 & b_3 \\ 0 & c_2 & c_3 \end{vmatrix} = 5 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} = 5(b_2c_3 - b_3c_2)$$

Analogamente

$$\begin{vmatrix} 0 & a_2 & a_3 \\ 5 & b_2 & b_3 \\ 0 & c_2 & c_3 \end{vmatrix} = -5 \begin{vmatrix} a_2 & a_3 \\ c_2 & c_3 \end{vmatrix} = -5(a_2c_3 - a_3c_2)$$

Per calcolare il seguente determinante conviene sviluppare rispetto alla seconda riga:

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ 0 & 27 & 0 \\ c_1 & c_2 & c_3 \end{vmatrix} = 27 \begin{vmatrix} a_1 & a_3 \\ c_1 & c_3 \end{vmatrix} = 27(a_1c_3 - a_3c_1)$$

Sviluppando rispetto alla prima colonna si trova

$$\begin{vmatrix} 0 & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & a_{n2} & \cdots & a_{nn} \end{vmatrix} = (-1)^{n+1} \begin{vmatrix} a_{12} & & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n-1,2} & \cdots & a_{n-1,n} \end{vmatrix}$$

Esempio

Gli sviluppi di Laplace sono particolarmente utili quando il determinante da calcolare dipende da un parametro, il che rende difficile l'uso del MEG. L'esempio più importante, che studieremo in dettaglio nel prossimo capitolo, è quello del *polinomio caratteristico* $P(\lambda)$ di una matrice quadrata \mathbf{A} : per definizione, $P(\lambda)$ è il determinante della matrice ottenuta sottraendo la variabile λ degli elementi sulla diagonale principale:

$$P(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I}) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix}$$

Per esempio il polinomio caratteristico della matrice

$$\mathbf{A} = \begin{bmatrix} 9 & 2 & -2 \\ 2 & 2 & 0 \\ -2 & 0 & 2 \end{bmatrix}$$

si calcola facilmente sviluppando rispetto alla terza colonna:

$$\begin{aligned} \det(\mathbf{A} - \lambda \mathbf{I}) &= \begin{vmatrix} 9 - \lambda & 2 & -2 \\ 2 & 2 - \lambda & 0 \\ -2 & 0 & 2 - \lambda \end{vmatrix} = -2 \begin{vmatrix} 2 & 2 - \lambda \\ -2 & 0 \end{vmatrix} + (2 - \lambda) \begin{vmatrix} 9 - \lambda & 2 \\ 2 & 2 - \lambda \end{vmatrix} = \\ &= -4(2 - \lambda) + (2 - \lambda)(\lambda^2 - 11\lambda + 18 - 4) = (2 - \lambda)(\lambda - 1)(\lambda - 10) \end{aligned}$$

Questo conto mostra che la matrice $\mathbf{A} - \lambda \mathbf{I}$ è singolare, precisamente, per i valori del parametro $\lambda = 2$, $\lambda = 1$ e $\lambda = 10$.

Gli sviluppi di Laplace forniscono una formula esplicita per la matrice inversa di una matrice non singolare; tale formula è così complessa dal punto di vista computazionale da essere inutile per il calcolo dell'inversa, ma è interessante dal punto di vista teorico.

TEOREMA 5.3 (Formula di Laplace per la matrice inversa)

Sia \mathbf{A} una matrice quadrata e sia \mathbf{A}^* la matrice che ha come elemento di posto (j, i) il complemento algebrico C_{ij} di \mathbf{A} . Allora

$$\mathbf{AA}^* = \mathbf{A}^*\mathbf{A} = \det(\mathbf{A})\mathbf{I}$$

In particolare, se $\det(\mathbf{A}) \neq 0$,

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \mathbf{A}^*$$

DIMOSTRAZIONE. Per definizione di prodotto di matrici

$$(5.4) \quad (\mathbf{AA}^*)_{ik} = \sum_{j=1}^n (\mathbf{A})_{ij} (\mathbf{A}^*)_{jk} = \sum_{j=1}^n a_{ij} C_{kj}$$

Se $i = k$, troviamo

$$(\mathbf{AA}^*)_{ii} = \sum_{j=1}^n a_{ij} C_{ij} = \det(\mathbf{A})$$

Quindi gli elementi sulla diagonale principale di \mathbf{AA}^* sono tutti uguali a $\det(\mathbf{A})$.

Se $i \neq k$, sia \mathbf{A}' la matrice ottenuta da \mathbf{A} sostituendo la riga k con la riga i : nella matrice \mathbf{A}' le righe i e k sono entrambe uguali alla riga i di \mathbf{A} , le altre righe di \mathbf{A}' sono uguali a quelle di \mathbf{A} ; in particolare le sottomatrici di \mathbf{A} e \mathbf{A}' ottenute rimuovendo la riga k e una colonna j sono uguali, per cui i cofattori C'_{kj} sono uguali ai cofattori C_{kj} di \mathbf{A} . Il determinante di \mathbf{A}' è zero perché la matrice ha due righe uguali. D'altra parte sviluppando il determinante rispetto alla colonna k troviamo:

$$0 = \det(\mathbf{A}') = \sum_{j=1}^n (\mathbf{A}')_{kj} C'_{kj} = \sum_{j=1}^n a_{ij} C_{kj}$$

Confrontando con l'equazione (5.4) troviamo $(\mathbf{AA}^*)_{ik} = 0$ per $i \neq k$: gli elementi al di fuori della diagonale principale in \mathbf{AA}^* sono nulli. La matrice \mathbf{AA}^* coincide quindi con la matrice identità moltiplicata per lo scalare $\det(\mathbf{A})$.

La dimostrazione dell'uguaglianza $\mathbf{A}^*\mathbf{A} = \det(\mathbf{A})\mathbf{I}$ è identica se si sostituiscono gli sviluppi rispetto alle righe con gli sviluppi rispetto alle colonne.

Se $\det(\mathbf{A}) \neq 0$, allora dividendo per $\det(\mathbf{A})$ si vede che $\frac{1}{\det(\mathbf{A})} \mathbf{A}^*$ è la matrice inversa di \mathbf{A} .

la matrice \mathbf{A}^* è

$$\mathbf{A}^* = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Se $ad - bc \neq 0$, ritroviamo la formula per l'inversa

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

COROLLARIO 5.4 (Formula di Cramer)

Un sistema quadrato $\mathbf{Ax} = \mathbf{b}$ con $\det(\mathbf{A}) \neq 0$ ammette l'unica soluzione $\mathbf{v} = [x_1, \dots, x_n]^T$ di componenti

$$x_i = \frac{\det(\mathbf{A}_i)}{\det(\mathbf{A})}$$

dove \mathbf{A}_i è ottenuta sostituendo la colonna i di \mathbf{A} col termine noto \mathbf{b} .

DIMOSTRAZIONE. Poiché \mathbf{A} ha determinante non nullo, il sistema ammette un'unica soluzione

$$\mathbf{v} = \mathbf{A}^{-1}\mathbf{b} = \frac{1}{\det(\mathbf{A})} \mathbf{A}^*\mathbf{b}$$

Sviluppando il determinante di \mathbf{A}_i rispetto alla colonna i si trova che $\det(\mathbf{A}_i)$ è uguale al prodotto della riga i di \mathbf{A}^* con \mathbf{b} , quindi

$$x_i(\mathbf{v}) = \frac{\det(\mathbf{A}_i)}{\det(\mathbf{A})}$$

Determinante di Vandermonde

Fissati d scalari x_1, \dots, x_d , consideriamo la matrice, detta di Vandermonde

$$(5.5) \quad \mathbf{V}(x_1, \dots, x_d) = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{d-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{d-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_d & x_d^2 & \dots & x_d^{d-1} \end{bmatrix}$$

Il determinante di questa matrice, detto *determinante di Vandermonde*, è il prodotto di tutte le differenze $x_j - x_i$ con $j > i$:

$$(5.6) \quad \det(\mathbf{V}(x_1, \dots, x_d)) = \begin{vmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{d-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{d-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_d & x_d^2 & \dots & x_d^{d-1} \end{vmatrix} = \prod_{1 < i < j \leq d} (x_j - x_i)$$

In particolare, le righe della matrice di Vandermonde sono linearmente indipendenti se e solo se gli scalari x_1, \dots, x_d sono distinti. La formula (5.6) si può dimostrare per induzione con un calcolo brutale. I primi casi sono:

Caso $d = 2$

$$\begin{vmatrix} 1 & x_1 \\ 1 & x_2 \end{vmatrix} = x_2 - x_1$$

Caso $d = 3$

$$\begin{aligned}
 & \left| \begin{array}{ccc} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{array} \right| = \quad (\text{sottraggo la prima riga alle altre}) \\
 & = \left| \begin{array}{ccc} 1 & x_1 & x_1^2 \\ 0 & x_2 - x_1 & x_2^2 - x_1^2 \\ 0 & x_3 - x_1 & x_3^2 - x_1^2 \end{array} \right| = \quad (\text{Laplace rispetto alla prima colonna}) \\
 & = \left| \begin{array}{ccc} x_2 - x_1 & x_2^2 - x_1^2 \\ x_3 - x_1 & x_3^2 - x_1^2 \end{array} \right| = \quad (\text{sottraggo alla terza colonna la seconda moltiplicata per } x_1) \\
 & = \left| \begin{array}{cc} x_2 - x_1 & x_2(x_2 - x_1) \\ x_3 - x_1 & x_3(x_3 - x_1) \end{array} \right| = \quad (\text{linearità nella seconda e terza riga}) \\
 & = (x_2 - x_1)(x_3 - x_1) \left| \begin{array}{c} 1 & x_2 \\ 1 & x_3 \end{array} \right| = (x_2 - x_1)(x_3 - x_1)(x_3 - x_2)
 \end{aligned}$$

Caso $d = 4$

$$\begin{aligned}
 & \left| \begin{array}{cccc} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ 1 & x_3 & x_3^2 & x_3^3 \\ 1 & x_4 & x_4^2 & x_4^3 \end{array} \right| = \quad (\text{sottraggo la prima riga alle altre}) \\
 & = \left| \begin{array}{cccc} 1 & x_1 & x_1^2 & x_1^3 \\ 0 & x_2 - x_1 & x_2^2 - x_1^2 & x_2^3 - x_1^3 \\ 0 & x_3 - x_1 & x_3^2 - x_1^2 & x_3^3 - x_1^3 \\ 0 & x_4 - x_1 & x_4^2 - x_1^2 & x_4^3 - x_1^3 \end{array} \right| = \quad (\text{Laplace rispetto alla prima colonna}) \\
 & = \left| \begin{array}{ccc} x_2 - x_1 & x_2^2 - x_1^2 & x_2^3 - x_1^3 \\ x_3 - x_1 & x_3^2 - x_1^2 & x_3^3 - x_1^3 \\ x_4 - x_1 & x_4^2 - x_1^2 & x_4^3 - x_1^3 \end{array} \right| = \quad (\text{sottraggo alla quarta colonna la terza moltiplicata per } x_1) \\
 & = \left| \begin{array}{ccc} x_2 - x_1 & x_2^2 - x_1^2 & x_2^2(x_2 - x_1) \\ x_3 - x_1 & x_3^2 - x_1^2 & x_3^2(x_3 - x_1) \\ x_4 - x_1 & x_4^2 - x_1^2 & x_4^2(x_4 - x_1) \end{array} \right| = \quad (\text{sottraggo alla terza colonna la seconda moltiplicata per } x_1) \\
 & = \left| \begin{array}{ccc} x_2 - x_1 & x_2(x_2 - x_1) & x_2^2(x_2 - x_1) \\ x_3 - x_1 & x_3(x_3 - x_1) & x_3^2(x_3 - x_1) \\ x_4 - x_1 & x_4(x_4 - x_1) & x_4^2(x_4 - x_1) \end{array} \right| = \quad (\text{multilinearità nelle righe}) \\
 & = (x_2 - x_1)(x_3 - x_1)(x_4 - x_1) \det(\mathbf{V}(x_2, x_3, x_4)) = \\
 & = (x_2 - x_1)(x_3 - x_1)(x_4 - x_2)(x_3 - x_2)(x_4 - x_2)(x_4 - x_3) = \prod_{1 \leq i < j \leq 4} (x_j - x_i).
 \end{aligned}$$

Caso generale: con passaggi analoghi si mostra

$$\begin{aligned}
 \det(\mathbf{V}(x_1, \dots, x_d)) &= \prod_{j=2}^d (x_j - x_1) \det(\mathbf{V}(x_2, \dots, x_d)) = \prod_{j=2}^d (x_j - x_1) \prod_{2 \leq i < j \leq d} (x_j - x_i) = \\
 &= \prod_{1 \leq i < j \leq d} (x_j - x_i)
 \end{aligned}$$

Si possono dare dimostrazioni più intelligenti. Per esempio, la formula (4.8) mostra che il determinante di Vandermonde è un polinomio di grado al massimo

$$\sum_{k=0}^{d-1} k = \frac{1}{2}d(d-1)$$

nella variabili x_i . D'altra parte il determinante è 0 se $x_i = x_j$ con $i < j$ perché in tal caso la matrice ha due righe uguali. Per Ruffini il determinante è divisibile per $x_j - x_i$ per ogni coppia di indici con $i < j$, e quindi $\prod_{1 \leq i < j \leq d} (x_j - x_i)$ divide $\det(\mathbf{V}(x_1, \dots, x_d))$. Siccome il grado del prodotto è proprio $\frac{1}{2}d(d-1)$, ne segue che esiste uno scalare c_d tale che $\det(\mathbf{V}(x_1, \dots, x_d)) = c_d \prod_{1 \leq i < j \leq d} (x_j - x_i)$. A questo punto occorre mostrare che $c_d = 1$: per questo si può sviluppare secondo Laplace rispetto all'ultima riga e controllare che il coefficiente di x_d^{d-1} in $\det(\mathbf{V}(x_1, \dots, x_d))$ è $\det(\mathbf{V}(x_1, \dots, x_{d-1}))$. Confrontando col coefficiente di x_d^{d-1} in $c_d \prod_{1 \leq i < j \leq d} (x_j - x_i)$ si scopre $c_d = c_{d-1}$. Siccome $c_2 = 1$, per induzione $c_d = c_2$ e la formula (5.6) è completamente dimostrata.

Esercizi

- 15** Scrivere gli sviluppi di Laplace del determinante della matrice $\mathbf{A} = \begin{bmatrix} 3 & 1 & -3 \\ 1 & 3 & 0 \\ 1 & 2 & -1 \end{bmatrix}$ rispetto a ciascuna delle righe e delle colonne (6 sviluppi in tutto, devono dare lo stesso risultato).

- 16** Calcolare il determinante della matrice

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 2 & 0 & 1 \\ 3 & 0 & 1 & 2 \\ 1 & 0 & 0 & 5 \end{bmatrix}$$

- 17** Calcolare il polinomio caratteristico $\det(\mathbf{A} - \lambda \mathbf{I})$ della matrice

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & 0 \\ 3 & 0 & 1 \end{bmatrix}$$

- 18** Calcolare l'inversa della matrice $\begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & 0 \\ 4 & 0 & 1 \end{bmatrix}$ dopo aver verificato che il determinante è non nullo. Usare la formula del teorema 5.3 e poi ripetere l'esercizio usando il metodo di Gauss-Jordan.

- 19** Risolvere con la formula di Cramer il sistema lineare

$$\begin{cases} 3x + y - 3z = 6 \\ x + 3y = 8 \\ x + 2y - z = 7 \end{cases}$$

- 20** Mostrare che esiste un'unica parabola di equazione $y = Ax^2 + Bx + C$ passante per i punti (x_1, y_1) , (x_2, y_2) e (x_3, y_3) a condizione che x_1, x_2, x_3 siano distinti.

6 IL TEOREMA DI BINET E IL DETERMINANTE DI UN'APPLICAZIONE LINEARE

Il determinante di un prodotto è il prodotto dei determinanti:

TEOREMA 6.1 (Teorema di Binet)

Se \mathbf{A} e \mathbf{B} sono matrici quadrate di ordine n , allora

$$\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$$

DIMOSTRAZIONE. La dimostrazione è una semplice applicazione del teorema 4.5, ed è un buon esempio di come un teorema astratto possa essere utile: è una pessima idea cercare di dimostrare la formula di Binet direttamente dalla formula del determinante. Supponiamo dapprima $\det(\mathbf{B}) \neq 0$, e consideriamo la funzione $D : \mathbb{M}_{\mathbb{K}}(n, n) \rightarrow \mathbb{K}$ definita da

$$(6.1) \quad D(\mathbf{A}) = \frac{\det(\mathbf{AB})}{\det(\mathbf{B})}$$

Basta mostrare che $D(\mathbf{A}) = \det(\mathbf{A})$, perché, data questa uguaglianza, moltiplicando la (6.1) per $\det(\mathbf{B})$ si ottiene la tesi. Per il teorema 4.5, per mostrare $D(\mathbf{A}) = \det(\mathbf{A})$ è sufficiente far vedere che D è una funzione multilinearare e alternante delle righe di \mathbf{A} che soddisfa $D(\mathbf{I}) = 1$. Quest'ultima uguaglianza è evidente:

$$D(\mathbf{I}) = \frac{\det(\mathbf{IB})}{\det(\mathbf{B})} = \frac{\det(\mathbf{B})}{\det(\mathbf{B})} = 1$$

Per quanto riguarda l'alternanza, ricordiamo che la riga i della matrice prodotto \mathbf{AB} è il prodotto della riga i di \mathbf{A} per la matrice \mathbf{B} . Quindi, se si scambiano due righe di \mathbf{A} , si scambiano anche le medesime righe in \mathbf{AB} e il determinante di \mathbf{AB} cambia segno e perciò cambia segno anche $D(\mathbf{A})$. Questo mostra l'alternanza. Visto che la funzione è alternante, per mostrare la multilinearità è sufficiente mostrare la linearità nella prima riga. Per questo scriviamo una matrice $n \times n$ nella forma $\mathbf{A} = [\mathbf{a}]$ dove \mathbf{a} è la prima riga e \mathbf{C} è la matrice formata dalle altre righe. Allora

$$\begin{aligned} \det \left(\begin{bmatrix} t_1 \mathbf{a}_1 + t_2 \mathbf{a}_2 \\ \mathbf{C} \end{bmatrix} \mathbf{B} \right) &= \det \left(\begin{bmatrix} t_1 \mathbf{a}_1 \mathbf{B} + t_2 \mathbf{a}_2 \mathbf{B} \\ \mathbf{C} \mathbf{B} \end{bmatrix} \right) = \\ &= t_1 \det \left(\begin{bmatrix} \mathbf{a}_1 \mathbf{B} \\ \mathbf{C} \mathbf{B} \end{bmatrix} \right) + t_2 \det \left(\begin{bmatrix} \mathbf{a}_2 \mathbf{B} \\ \mathbf{C} \mathbf{B} \end{bmatrix} \right) = t_1 \det \left(\begin{bmatrix} \mathbf{a}_1 \\ \mathbf{C} \end{bmatrix} \mathbf{B} \right) + t_2 \det \left(\begin{bmatrix} \mathbf{a}_2 \\ \mathbf{C} \end{bmatrix} \mathbf{B} \right) \end{aligned}$$

Nel secondo passaggio abbiamo usato la linearità del determinante nella prima riga. Questa formula mostra che la funzione $\det(\mathbf{AB})$, e quindi anche $D(\mathbf{A})$, è lineare nella prima riga di \mathbf{A} . La dimostrazione nel caso $\det(\mathbf{B}) \neq 0$ è così completa.

Per finire, supponiamo che $\det(\mathbf{B}) = 0$. In questo caso \mathbf{B} non ha rango massimo e quindi il nucleo di \mathbf{B} contiene un vettore non nullo \mathbf{v} . Questo vettore appartiene anche al nucleo di \mathbf{AB} perché $\mathbf{ABv} = \mathbf{A}\mathbf{0} = \mathbf{0}$. Quindi \mathbf{AB} non ha rango massimo e $\det(\mathbf{AB}) = 0$. Questo conclude la dimostrazione perché $\det(\mathbf{A}) \det(\mathbf{B}) = 0$ quando $\det(\mathbf{B}) = 0$. ■

COROLLARIO 6.2 (Determinante della matrice inversa)

Sia \mathbf{A} una matrice quadrata invertibile. Allora

$$\det(\mathbf{A}^{-1}) = \frac{1}{\det(\mathbf{A})}$$

DIMOSTRAZIONE. Dall'uguaglianza $\mathbf{I} = \mathbf{A}^{-1}\mathbf{A}$ segue

$$1 = \det(\mathbf{I}) = \det(\mathbf{A}^{-1}\mathbf{A})$$

Per il teorema di Binet

$$\det(\mathbf{A}^{-1}\mathbf{A}) = \det(\mathbf{A}^{-1}) \det(\mathbf{A})$$

Quindi $\det(\mathbf{A}^{-1}) \det(\mathbf{A}) = 1$, per cui il numero $\det(\mathbf{A}^{-1})$ è l'inverso di $\det(\mathbf{A})$.

OSSERVAZIONE Si osservi che l'uguaglianza $\det(\mathbf{A}^{-1}) \det(\mathbf{A}) = 1$ mostra direttamente che, se \mathbf{A} è invertibile, allora $\det(\mathbf{A}) \neq 0$. Viceversa, se $\det(\mathbf{A}) \neq 0$, per il teorema di Laplace 5.3 la matrice \mathbf{A} è invertibile con inversa $\frac{1}{\det(\mathbf{A})}\mathbf{A}^*$. Si ottiene così una nuova dimostrazione del fatto fondamentale che una matrice \mathbf{A} è invertibile se e solo se $\det(\mathbf{A}) \neq 0$.

COROLLARIO 6.3 (Invarianza del determinante per similitudine)

Siano \mathbf{A} e \mathbf{S} due matrici quadrate di ordine n . Se \mathbf{S} è invertibile, allora

$$\det(\mathbf{S}^{-1}\mathbf{AS}) = \det(\mathbf{A})$$

DIMOSTRAZIONE. Usiamo il teorema di Binet e il fatto che il determinante di \mathbf{S}^{-1} è l'inverso del determinante di \mathbf{S} :

$$\det(\mathbf{S}^{-1}\mathbf{AS}) = \det(\mathbf{S}^{-1}) \det(\mathbf{A}) \det(\mathbf{S}) = \det(\mathbf{S})^{-1} \det(\mathbf{S}) \det(\mathbf{A}) = \det(\mathbf{A})$$

Si osservi che nel secondo passaggio abbiamo potuto cambiare l'ordine dei fattori perché il prodotto di scalari è commutativo.

L'importante conseguenza del corollario precedente è che possiamo definire il *determinante di un'applicazione lineare*. Più precisamente:

DEFINIZIONE 6.4 (Determinante di un'applicazione lineare)

Sia \mathbf{V} uno spazio vettoriale di dimensione finita e sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{V}$ un'applicazione lineare (o come si dice in algebra un *endomorfismo*, per indicare il fatto che dominio e codominio coincidono). Data una base \mathcal{B} di \mathbf{V} , sia \mathbf{A} la matrice che rappresenta \mathcal{L} rispetto a \mathcal{B} . Il determinante di \mathcal{L} è lo scalare

$$\det(\mathcal{L}) = \det(\mathbf{A})$$

Tale scalare dipende solo da \mathcal{L} e non dalla base \mathcal{B} utilizzata per rappresentare \mathcal{L} .

Dobbiamo spiegare perché $\det(\mathcal{L})$ non dipende dalla base \mathcal{B} . Supponiamo che \mathbf{A} rappresenti \mathcal{L} rispetto a \mathcal{B} e supponiamo che \mathbf{B} rappresenti \mathcal{L} rispetto a un'altra base. Se \mathbf{S} è la matrice di passaggio dalla base \mathcal{B} all'altra base, allora, per la proposizione 5.5 del capitolo sulle applicazioni lineari,

$$\mathbf{B} = \mathbf{S}^{-1} \mathbf{A} \mathbf{S}$$

(si dice che \mathbf{A} e \mathbf{B} sono simili). Per il corollario precedente $\det(\mathbf{B}) = \det(\mathbf{A})$. Quindi le matrici che rappresentano \mathcal{L} hanno tutte lo stesso determinante e questo numero si dice determinante di \mathcal{L} .

Quando $\mathbf{V} = \mathbb{R}^n$, il modulo del determinante di \mathcal{L} ha un chiaro significato geometrico: è il *fattore di cambiamento di volume* (di area per $n = 2$). Per vedere perché, consideriamo in \mathbb{R}^n il parallelepipedo P che ha un vertice nell'origine e, come spigoli uscenti dall'origine, i vettori $\mathbf{v}_1, \dots, \mathbf{v}_n$ di \mathbb{R}^n . Sia \mathbf{M} la matrice che ha come colonne i vettori \mathbf{v}_k . Il volume $\text{Vol}(P)$ è il modulo del determinante di \mathbf{M} . Se $\mathcal{L} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ è lineare, l'immagine $\mathcal{L}(P)$ del parallelepipedo P è il parallelepipedo che ha un vertice nell'origine e, come spigoli uscenti dall'origine, i vettori $\mathcal{L}(\mathbf{v}_1), \dots, \mathcal{L}(\mathbf{v}_n)$. Il volume del parallelepipedo $\mathcal{L}(P)$ è quindi il modulo del determinante della matrice \mathbf{N} che ha come colonne i vettori $\mathcal{L}(\mathbf{v}_k)$. Se \mathbf{A} è la matrice che rappresenta \mathcal{L} rispetto alla base canonica, $\mathcal{L}(\mathbf{v}_k) = \mathbf{A}\mathbf{v}_k$ per cui $\mathbf{N} = \mathbf{AM}$. Quindi

$$\text{Vol}(\mathcal{L}(P)) = |\det(\mathbf{AM})| = |\det(\mathbf{A})||\det(\mathbf{M})| = |\det(\mathcal{L})|\text{Vol}(P)$$

Questo mostra che, come anticipato, il modulo del determinante di \mathcal{L} è il rapporto tra il volume dell'immagine $\mathcal{L}(P)$ e il volume di P . Approssimando una regione misurabile Ω di \mathbb{R}^n con l'unione disgiunta di parallelepipedi si trova

$$\text{Vol}(\mathcal{L}(\Omega)) = \int_{\mathcal{L}(\Omega)} d\mathbf{y} = |\det(\mathcal{L})| \int_{\Omega} d\mathbf{x} = |\det(\mathcal{L})| \text{Vol}(\Omega)$$

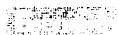
Non è sorprendente allora che nella formula di cambiamento di variabili degli integrali multipli il fattore di cambiamento dell'elemento infinitesimale di volume sia il modulo del determinante della matrice Jacobiana: la versione infinitesimale della formula precedente è

$$d\mathbf{y} = \left| \det \left(\begin{bmatrix} \frac{\partial y_i}{\partial x_j} \end{bmatrix} \right) \right| d\mathbf{x}$$

In forma integrale si ha

$$\int_{\Omega_y} f(\mathbf{y}) d\mathbf{y} = \int_{\Omega_x} f(\mathbf{y}(\mathbf{x})) \left| \det \left(\begin{bmatrix} \frac{\partial y_i}{\partial x_j} \end{bmatrix} \right) \right| d\mathbf{x}$$

che vale, per esempio, se Ω_x è una regione misurabile di \mathbb{R}^n , $\mathbf{y} = \mathbf{y}(\mathbf{x})$ è un cambiamento di variabili sufficientemente regolare, Ω_y è la stessa regione espressa però rispetto alle coordinate \mathbf{y} , mentre $f(\mathbf{y})$ è una funzione integrabile in Ω_y . Questa formula rappresenta una delle applicazioni principali del determinante.



Sia $\mathcal{L} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ una rotazione. Chiaramente \mathcal{L} preserva l'area, cioè l'area di $\mathcal{L}(\Omega)$ coincide con l'area di Ω per ogni regione misurabile Ω del piano cartesiano. Perciò il determinante di \mathcal{L} deve avere modulo 1. In effetti, rappresentando \mathcal{L} rispetto alla base canonica, si trova

$$\det(\mathcal{L}) = \det \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} = 1$$

Mostrare che, se \mathbf{A} è una matrice quadrata, allora $\det(\mathbf{A}^p) = \det(\mathbf{A})^p$ per ogni intero $p \geq 0$ (anche per gli interi p negativi se \mathbf{A} è invertibile).

Sia \mathbf{Q} una matrice quadrata reale ortogonale, cioè tale che $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$. Mostrare che $\det(\mathbf{Q}) = 1$ oppure $\det(\mathbf{Q}) = -1$.

Suggerimento: usando Binet mostrare che $\det(\mathbf{Q})^2 = 1$.

7 DETERMINANTE E RANGO

In questo paragrafo mostriamo che è possibile calcolare il rango di una matrice arbitraria \mathbf{M} (anche rettangolare) esaminando i determinanti delle sottomatrici quadrate di \mathbf{M} ; una sottomatrice di \mathbf{M} è una matrice ottenuta cancellando alcune righe o colonne di \mathbf{M} .

DEFINIZIONE 7.1 (Minori di una matrice)

Un minore di ordine p di una matrice \mathbf{M} è il determinante di una sottomatrice $p \times p$ di \mathbf{M} . Un minore $\det(\mathbf{B})$ di ordine $p+1$ *orla* un minore $\det(\mathbf{A})$ di ordine p se \mathbf{A} è una sottomatrice di \mathbf{B} .

TEOREMA 7.2 (Teorema di Kronecker)

Sia \mathbf{M} una matrice.

- a) se \mathbf{A} è una sottomatrice $p \times p$ di \mathbf{M} e $\det(\mathbf{A}) \neq 0$, le righe (rispettivamente le colonne) di \mathbf{M} che contengono le righe (rispettivamente le colonne) di \mathbf{A} sono linearmente indipendenti;
- b) se \mathbf{M} ha rango r , allora \mathbf{M} ha almeno un minore di ordine r non nullo, e tutti i minori di \mathbf{M} di ordine $p > r$ sono nulli;
- c) il rango di \mathbf{M} è r se e solo se \mathbf{M} contiene una sottomatrice quadrata \mathbf{A} di ordine r con $\det(\mathbf{A})$ diverso da zero e ogni sottomatrice quadrata di \mathbf{M} di ordine $r+1$ contenente \mathbf{A} ha determinante uguale a zero. Brevemente: $r(\mathbf{M}) = r$ se e solo se \mathbf{M} ha un minore δ di ordine r non nullo ed è nullo ogni minore che orla δ .

DIMOSTRAZIONE.

- a) Supponiamo per assurdo che le p righe di \mathbf{M} che contengono le p righe di \mathbf{A} siano linearmente dipendenti. La stessa relazione di dipendenza lineare che vale per queste p righe di \mathbf{M} vale anche per le p righe di \mathbf{A} . Quindi le righe di \mathbf{A} sono linearmente dipendenti e $\det(\mathbf{A}) = 0$, contraddicendo l'ipotesi, assurdo! Lo stesso ragionamento vale anche per le colonne.
- b) Supponiamo che \mathbf{M} sia una matrice $m \times n$. Siccome il rango di \mathbf{M} è r , la matrice ha r righe linearmente indipendenti; cancellando le altre righe otteniamo una sottomatrice \mathbf{M}_1 di tipo (r, n) , che ha ancora rango r perché le righe di \mathbf{M}_1 sono per costruzione indipendenti. Siccome il rango è anche il massimo numero di colonne linearmente indipendenti, possiamo trovare r colonne di \mathbf{M}_1 linearmente indipendenti: cancellando le altre colonne ottieniamo una matrice \mathbf{M}_2 quadrata di ordine r . Il rango di \mathbf{M}_2 è ancora r perché le colonne di \mathbf{M}_2 sono per costruzione indipendenti. Siccome \mathbf{M}_2 è quadrata di rango massimo, il suo determinante è diverso da zero. Abbiamo così trovato un minore di ordine r di \mathbf{M} non nullo. Ogni minore di ordine $p > r$ è invece nullo: altrimenti per a) la matrice avrebbe p righe linearmente indipendenti e quindi il rango sarebbe almeno p , il che è impossibile perché $p > r$.
- c) Da b) segue che, se \mathbf{M} ha rango r , allora esiste una sottomatrice quadrata \mathbf{A} di ordine r con determinante non nullo, e tutti i minori di ordine $p > r$ sono nulli; in particolare, sono nulli i minori di ordine $r + 1$.
- Viceversa, supponiamo che \mathbf{M} abbia una sottomatrice quadrata \mathbf{A} di ordine p con determinante non nullo e che tutte le sottomatrici \mathbf{B} di ordine $p + 1$ che contengono \mathbf{A} abbiano determinante nullo. Dobbiamo mostrare che $p = r(\mathbf{A})$. Per a) il rango di \mathbf{M} è $\geq p$ e le p righe di \mathbf{M} che contengono le p righe di \mathbf{A} sono linearmente indipendenti. Se il rango di \mathbf{M} fosse $> p$, potremmo aggiungere a queste p righe una riga di \mathbf{M} ottenendo un insieme linearmente indipendente di $p + 1$ righe di \mathbf{M} . La sottomatrice \mathbf{M}_1 formata da queste $p + 1$ righe contiene \mathbf{A} e ha rango $p + 1$. Le p colonne di \mathbf{M}_1 che contengono le colonne di \mathbf{A} sono linearmente indipendenti (per a)); siccome lo spazio colonna di \mathbf{M}_1 ha dimensione $r(\mathbf{M}_1) = p + 1$, possiamo aggiungere a queste colonne di \mathbf{M}_1 un'ulteriore colonna, ottenendo una sottomatrice quadrata \mathbf{B} di ordine $p + 1$ di \mathbf{M}_1 . Per costruzione \mathbf{B} contiene \mathbf{A} e ha rango $p + 1$, quindi $\det(\mathbf{B}) \neq 0$. Questo contraddice l'ipotesi, quindi $r(\mathbf{M}) = p$ come volevasi dimostrare. 

Il teorema di Kronecker consente di calcolare il rango di \mathbf{M} mediante la seguente procedura:

- \mathbf{M} è la matrice nulla se e solo se tutti i minori di ordine 1 sono nulli; in tal caso $r(\mathbf{M}) = 0$.
- Altrimenti \mathbf{M} ha un elemento (minore di ordine 1) $b = a_{ij}$ diverso da zero. Se tutti i determinanti delle sottomatrici quadrate di ordine due di \mathbf{M} contenenti b sono nulli, $r(\mathbf{A}) = 1$.
- Altrimenti \mathbf{M} contiene una sottomatrice quadrata \mathbf{B} di ordine 2 con $\det(\mathbf{B})$ non nullo. Se tutti i determinanti delle sottomatrici quadrate di ordine 3 di \mathbf{M} contenenti \mathbf{B} sono nulli, $r(\mathbf{M}) = 2$.
- Altrimenti \mathbf{M} contiene una sottomatrice quadrata \mathbf{B} di ordine 3 con $\det(\mathbf{B})$ non nullo. E cosivvia.



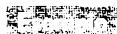
Sia $\mathbf{A} = \begin{bmatrix} 1 & 2 & 5 & 3 \\ 2 & 0 & 1 & 3 \\ -1 & 6 & 13 & 3 \end{bmatrix}$. Si tratta di una matrice di tipo $(3, 4)$, quindi il rango di \mathbf{A} è al massimo 3. La matrice ha minori di ordine 2 non nulli: per esempio, la sottomatrice $\mathbf{B} = \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix}$ ottenuta cancellando la terza riga e la terza e quarta colonna ha determinante 1. Quindi il rango di \mathbf{A} è almeno 2. Per vedere se è 2 o 3, dobbiamo calcolare i determinanti delle sottomatrici quadrate di ordine 3 di \mathbf{A} che contengono \mathbf{B} . Queste sono solo due: la sottomatrice \mathbf{C} ottenuta cancellando la terza colonna e la sottomatrice \mathbf{D} ottenuta cancellando la quarta colonna (si noti invece che \mathbf{A} ha 4 sottomatrici quadrate di ordine 3). Ora

$$\det(\mathbf{C}) = \det \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 3 \\ -1 & 6 & 3 \end{bmatrix} = 1(0 - 18) - 2(6 + 3) + 3(12 + 0) = 0$$

e

$$\det(\mathbf{D}) = \det \begin{bmatrix} 1 & 2 & 5 \\ 2 & 0 & 1 \\ -1 & 6 & 13 \end{bmatrix} = 1(0 - 6) - 2(26 + 1) + 5(12 + 0) = -6 - 54 + 60 = 0$$

Quindi il rango di \mathbf{A} è 2.



Vogliamo determinare, al variare di $k \in \mathbb{R}$, il rango della matrice

$$\mathbf{A}_k = \begin{bmatrix} k & 0 & k & k \\ 0 & k & 2 & 2k \\ 1 & k & k & k \end{bmatrix}$$

Il rango è almeno 2 perché (cancellando la riga 1 e le colonne 2 e 4)

$$\delta_2 = \begin{vmatrix} 0 & 2 \\ 1 & k \end{vmatrix} = -2 \neq 0$$

e al massimo 3 perché \mathbf{A}_k ha 3 righe. Orliamo δ_2 aggiungendo prima riga e seconda colonna

$$\delta_3 = \begin{vmatrix} k & 0 & k \\ 0 & k & 2 \\ 1 & k & k \end{vmatrix} = k^3 - 3k^2 = k^2(k - 3)$$

Se $k \neq 0$ e $k \neq 3$, il rango è 3 perché $\delta_3 \neq 0$ è un minore di ordine 3 non nullo.

Se $k = 0$,

$$\mathbf{A}_0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 2 \\ 1 & 0 & 0 \end{bmatrix}$$

È evidente che il rango di \mathbf{A}_0 è 2, ma lo vogliamo verificare utilizzando il teorema di Kronecker. I minori che orlano δ_2 sono $\delta_3 = 0$ e il minore δ'_3 che si ottiene aggiungendo la quarta colonna alla prima e alla terza:

$$\delta'_3 = \begin{vmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 0 & 0 \end{vmatrix} = 0$$

I due minori di ordine 3 che orlano δ_2 sono nulli per cui $r(\mathbf{A}_0) = 2$.

Infine se $k = 3$,

$$\mathbf{A}_3 = \begin{bmatrix} 3 & 0 & 3 & 3 \\ 0 & 3 & 2 & 6 \\ 1 & 3 & 3 & 3 \end{bmatrix}$$

In questo caso

$$\delta'_3 = \begin{vmatrix} 3 & 3 & 3 \\ 0 & 2 & 6 \\ 1 & 3 & 3 \end{vmatrix} = 3 \times 2 \begin{vmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \\ 1 & 3 & 3 \end{vmatrix} = 6((3 - 9) + (3 - 1)) = -24 \neq 0$$

e quindi $r(\mathbf{A}_3) = 3$.

Siano \mathbf{v} e \mathbf{w} due vettori di \mathbb{R}^3 e sia \mathbf{A} la matrice 3×2 che ha \mathbf{v} e \mathbf{w} come colonne. Mostrare che le tre componenti del prodotto vettoriale $\mathbf{v} \times \mathbf{w}$ sono, a meno del segno, i tre minori di ordine 2 della matrice \mathbf{A} . Concludere che i due vettori sono linearmente dipendenti se e solo se il loro prodotto vettoriale è il vettore nullo (geometricamente, questo significa che i due vettori sono paralleli). Se prendessimo due vettori di \mathbb{R}^4 , quanti minori avrebbe \mathbf{A} ? (*Risposta:* 6; c'è un analogo del prodotto vettoriale anche per vettori di \mathbb{R}^4 , ma è un vettore con 6 componenti).

Determinare al variare di k il rango della matrice

$$\begin{bmatrix} -5 & k & 0 & 5 \\ -3 & -3 & 1 & -1 - k \\ 0 & k + 5 & 3 & 3 \end{bmatrix}$$

8 COMPLEMENTI

Determinante di una matrice triangolare a blocchi

Una matrice quadrata si dice *triangolare (alta) a blocchi* se ha la forma

$$(8.1) \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{B}_{12} & \mathbf{B}_{13} & \cdots & \cdots \\ \mathbf{O} & \mathbf{A}_2 & \mathbf{B}_{23} & \cdots & \cdots \\ \mathbf{O} & \mathbf{O} & \mathbf{A}_3 & \cdots & \cdots \\ \vdots & \vdots & & \ddots & \vdots \\ \mathbf{O} & \mathbf{O} & & \mathbf{O} & \mathbf{A}_d \end{bmatrix}$$

dove i blocchi \mathbf{A}_k lungo la diagonale principale sono matrici quadrate e \mathbf{O} denota una matrice nulla. Quando i blocchi sulla diagonale principale hanno ordine uno si tratta di una ordinaria matrice triangolare alta. Generalizzando il fatto che il determinante di una matrice triangolare è il prodotto degli elementi sulla diagonale principale, il determinante di una matrice triangolare a blocchi è il prodotto dei determinanti dei blocchi sulla diagonale principale:

PROPOSIZIONE 8.1 (Determinante di una matrice triangolare a blocchi) Sia \mathbf{A} una matrice triangolare a blocchi come nella (8.1). Allora

$$(8.2) \quad \det(\mathbf{A}) = \det(\mathbf{A}_1) \cdots \det(\mathbf{A}_d) = \prod_{i=1}^d \det(\mathbf{A}_i)$$

DIMOSTRAZIONE. La dimostrazione è immediata: l'insieme dei pivots di \mathbf{A} è l'unione degli insiemi dei pivots dei blocchi \mathbf{A}_i e il determinante (a meno del segno) è il prodotto dei pivots. Più precisamente: se $\det(\mathbf{A}_1) = 0$, allora le colonne di \mathbf{A}_1 sono linearmente dipendenti; la stessa relazione di dipendenza vale per le corrispondenti colonne di \mathbf{A} e quindi $\det(\mathbf{A}) = 0$. Se invece $\det(\mathbf{A}_1) \neq 0$, l'algoritmo di eliminazione di Gauss per le prime n_1 righe di \mathbf{A} coincide con quello per \mathbf{A}_1 , per cui il prodotto dei primi n_1 pivots di \mathbf{A} (col segno appropriato se ci sono scambi di righe) è uguale a $\det(\mathbf{A}_1)$. Procedendo per induzione si trova la formula (8.2).

La matrice

$$\mathbf{A} = \left[\begin{array}{cc|cc} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \\ \hline 0 & 0 & c_3 & c_4 \\ 0 & 0 & d_3 & d_4 \end{array} \right]$$

è triangolare a blocchi, i blocchi lungo la diagonale principale sono

$$\mathbf{A}_1 = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} c_3 & c_4 \\ d_3 & d_4 \end{bmatrix}$$

Quindi

$$\det(\mathbf{A}) = (a_1 b_2 - a_2 b_1)(c_3 d_4 - c_4 d_3)$$

Per esempio

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 7 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 3 & 5 \end{vmatrix} = (3 - 8)(15 - 12) = -15$$

Minori principali e pivots

Data una matrice \mathbf{A} quadrata di ordine n , per ogni $k = 1, 2, \dots, n$ sia $\mathbf{A}_{(k)}$ la sottomatrice quadrata $k \times k$ ottenuta cancellando le ultime $n-k$ righe e le ultime $n-k$ colonne di \mathbf{A} . I minori $\delta_k = \det(\mathbf{A}_{(k)})$ si dicono *minori principali di nord ovest* di \mathbf{A} . Nel paragrafo sulla fattorizzazione LU abbiamo mostrato che, se \mathbf{A} è invertibile, l'algoritmo di eliminazione per la matrice \mathbf{A} non richiede scambi di righe, o equivalentemente \mathbf{A} ammette la fattorizzazione LU, se e solo se il rango di $\mathbf{A}_{(k)}$ è k , ovvero $\delta_k = \det(\mathbf{A}_{(k)}) \neq 0$. Otteniamo così

PROPOSIZIONE 8.2 (Minori principali di nord ovest e pivots)

Supponiamo che \mathbf{A} sia una matrice quadrata di ordine n invertibile. L'algoritmo di eliminazione per la matrice \mathbf{A} non richiede scambi di righe (cioè \mathbf{A} ammette la decomposizione \mathbf{LU}) se e solo se $\delta_k = \det(\mathbf{A}_{(k)}) \neq 0$ per ogni $k = 1, 2, \dots, n$. Se questo è il caso, δ_k è il prodotto dei primi k pivots di \mathbf{A} .

DIMOSTRAZIONE. Abbiamo già spiegato che l'algoritmo non richiede scambi righe se e solo se $\delta_k \neq 0$ per ogni k . Se questo è il caso, il processo di eliminazione per le prime k righe di \mathbf{A} consiste delle stesse operazioni dell'algoritmo di eliminazione per $\mathbf{A}_{(k)}$, quindi i primi k pivots di \mathbf{A} sono i pivots di $\mathbf{A}_{(k)}$ il cui prodotto è $\det(\mathbf{A}_{(k)}) = \delta_k$.

Derivata del determinante

La regola di Leibniz per la derivata di un prodotto di funzioni è la formula

$$\begin{aligned} \frac{d}{dx}(f_1(x)f_2(x) \cdots f_n(x)) &= \\ &= f'_1(x)f_2(x) \cdots f_n(x) + f_1(x)f'_2(x) \cdots f_n(x) + \cdots + f_1(x)f_2(x) \cdots f'_n(x) = \\ &= \sum_{i=1}^n f_1(x) \cdots f'_i(x) \cdots f_n(x) \end{aligned}$$

Una formula simile è valida per la derivata del determinante di una matrice i cui elementi siano funzioni derivabili $a_{ij}(x)$ della variabile x . Il motivo è che il determinante è una funzione multilineare delle sue righe, e come tale da molti punti di vista si comporta come se fosse un prodotto delle righe; per mettere in risalto la dipendenza dalle righe, scriviamo come in precedenza

$$\det([a_{ij}](x)) = D(\mathbf{v}_1(x), \mathbf{v}_2(x), \dots, \mathbf{v}_n(x))$$

dove $\mathbf{v}_1^T(x), \mathbf{v}_2^T(x), \dots, \mathbf{v}_n^T(x)$ sono le righe della matrice. Se queste funzioni vettoriali sono derivabili, per la derivata del determinante vale un regola alla regola di Leibniz:

$$\begin{aligned} \frac{d}{dx} D(\mathbf{v}_1(x), \mathbf{v}_2(x), \dots, \mathbf{v}_n(x)) &= D(\mathbf{v}'_1(x), \mathbf{v}_2(x), \dots, \mathbf{v}_n(x)) + \\ &+ D(\mathbf{v}_1(x), \mathbf{v}'_2(x), \dots, \mathbf{v}_n(x)) + \cdots + D(\mathbf{v}_1(x), \mathbf{v}_2(x), \dots, \mathbf{v}'_n(x)) = \\ &= \sum_{i=1}^n D(\mathbf{v}_1(x), \dots, \mathbf{v}'_i(x), \dots, \mathbf{v}_n(x)) \end{aligned}$$

Riscrivendo la formula in termini di matrici otteniamo:

PROPOSIZIONE 8.3 (Regola di Leibniz per la derivata del determinante) Supponiamo che le funzioni $a_{ij}(x)$ siano derivabili in x ($1 \leq i, j \leq n$). Sia $\mathbf{A}(x)$ la matrice che ha come elementi $a_{ij}(x)$, e per ogni $1 \leq i \leq n$ sia $\mathbf{B}_i(x)$ la matrice ottenuta sostituendo la riga i di $\mathbf{A}(x)$ con la sua derivata.

Allora la derivata del determinante $\det(\mathbf{A}(x))$ è la somma dei determinanti delle matrici $\mathbf{B}_i(x)$:

$$(8.3) \quad \frac{d}{dx} \begin{vmatrix} a_{11}(x) & a_{12}(x) & \cdots & a_{1n}(x) \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1}(x) & a_{n2}(x) & \cdots & a_{nn}(x) \end{vmatrix} = \sum_{i=1}^n \begin{vmatrix} a_{11}(x) & a_{12}(x) & \cdots & a_{1n}(x) \\ \vdots & \vdots & \vdots & \vdots \\ a'_{i1}(x) & a'_{i2}(x) & \cdots & a'_{in}(x) \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1}(x) & a_{n2}(x) & \cdots & a_{nn}(x) \end{vmatrix}$$

DIMOSTRAZIONE. Usiamo il fatto che la derivata di una somma è la somma delle derivate (o meglio la linearità della derivata) e la regola di Leibniz per la derivata di un prodotto di funzioni:

$$\begin{aligned} \frac{d}{dx} \det([a_{ij}(x)]) &= \frac{d}{dx} \sum_{\sigma \in S_n} (-1)^\sigma a_{1\sigma(1)}(x) \cdots a_{n\sigma(n)}(x) = \\ &= \sum_{\sigma \in S_n} (-1)^\sigma \frac{d}{dx} a_{1\sigma(1)}(x) \cdots a_{n\sigma(n)}(x) = \\ &= \sum_{\sigma \in S_n} (-1)^\sigma \sum_{i=1}^n a_{1\sigma(1)}(x) \cdots a'_{i\sigma(i)}(x) \cdots a_{n\sigma(n)}(x) = \\ &= \sum_{i=1}^n \left(\sum_{\sigma \in S_n} (-1)^\sigma a_{1\sigma(1)}(x) \cdots a'_{i\sigma(i)}(x) \cdots a_{n\sigma(n)}(x) \right) = \sum_{i=1}^n \det(\mathbf{B}_i). \end{aligned}$$

Supponiamo che $z_1(x), \dots, z_n(x)$ siano n funzioni derivabili n volte in un intervallo I . Come in Bramanti Pagani Salsa, *Analisi Matematica 2*, Zanichelli 2009, Capitolo 14.2, definiamo il determinante wronskiano $W(z_1, \dots, z_n)$ per mezzo della formula:

$$W(z_1, \dots, z_n)(x) = \begin{bmatrix} z_1(x) & z_2(x) & \cdots & z_n(x) \\ z'_1(x) & z'_2(x) & \cdots & z'_n(x) \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{(n-1)}(x) & z_2^{(n-1)}(x) & \cdots & z_n^{(n-1)}(x) \end{bmatrix}$$

Si tratta quindi del determinante della matrice, detta matrice wronskiana, che ha come elementi della prima riga le funzioni date valutate in x , come elementi della seconda riga le derivate prime di tali funzioni valutate in x e così via aumentando con le righe l'ordine di derivazione.

Se per un certo valore x_0 si ha $W(x_0) \neq 0$, le funzioni $z_i(x)$ sono linearmente indipendenti: infatti quando le funzioni sono dipendenti, le loro derivate di ogni ordine soddisfano la stessa relazione lineare, e quindi le colonne della matrice wronskiana sono linearmente dipendenti per un qualunque x fissato, per cui $W(x) = 0$ per ogni x .

Quando $z_k(x) = e^{\lambda_k x}$, la matrice wronskiana è

$$\begin{bmatrix} e^{\lambda_1 x} & e^{\lambda_2 x} & \cdots & e^{\lambda_n x} \\ \lambda_1 e^{\lambda_1 x} & \lambda_2 e^{\lambda_2 x} & \cdots & \lambda_n e^{\lambda_n x} \\ \vdots & \vdots & & \vdots \\ \lambda_1^{n-1} e^{\lambda_1 x} & \lambda_2^{n-1} e^{\lambda_2 x} & \cdots & \lambda_n^{n-1} e^{\lambda_n x} \end{bmatrix}$$

per cui

$$W(x) = e^{\sum_{i=1}^n \lambda_i x} V(\lambda_1, \dots, \lambda_n)$$

Siccome il determinante di Vandermonde $V(\lambda_1, \dots, \lambda_n)$ è diverso da 0 se gli scalari λ_k sono tutti distinti, ne concludiamo che le funzioni $e^{\lambda_k x}$ sono linearmente indipendenti se gli scalari λ_k sono tutti distinti.

Dalla regola di Leibniz per la derivata del determinante si ottiene (tenendo conto che il determinante è nullo se due righe sono uguali):

$$W'(x) = \begin{vmatrix} z_1(x) & z_2(x) & \cdots & z_n(x) \\ z'_1(x) & z'_2(x) & \cdots & z'_n(x) \\ \vdots & \vdots & & \vdots \\ z_1^{(n-2)}(x) & z_2^{(n-2)}(x) & \cdots & z_n^{(n-2)}(x) \\ z_1^{(n)}(x) & z_2^{(n)}(x) & \cdots & z_n^{(n)}(x) \end{vmatrix}$$

Supponiamo ora che tutte le funzioni $z_i(x)$ siano soluzioni dell'equazione differenziale lineare omogenea:

$$z^{(n)}(x) + a_{n-1}(x)z^{(n-1)}(x) + \cdots + a_1(x)z(x) + a_0(x) = 0$$

Possiamo allora ricavare $z_i^{(n)}$ da questa espressione e sostituirlo nell'ultima riga della formula precedente per $W'(x)$; tenendo conto dell'invarianza per scorrimento del determinante troviamo

$$W'(x) = \begin{vmatrix} z_1(x) & z_2(x) & \cdots & z_n(x) \\ z'_1(x) & z'_2(x) & \cdots & z'_n(x) \\ \vdots & \vdots & & \vdots \\ z_1^{(n-2)}(x) & z_2^{(n-2)}(x) & \cdots & z_n^{(n-2)}(x) \\ -a_{n-1}(x)z_1^{(n-1)}(x) & -a_{n-1}(x)z_2^{(n-1)}(x) & \cdots & -a_{n-1}(x)z_n^{(n-1)}(x) \end{vmatrix} = \\ = -a_{n-1}(x)W(x)$$

Da questa equazione si ricava per ogni $x_0, x \in I$

$$W(x) = W(x_0)e^{-\int_{x_0}^x a_{n-1}(t) dt}$$

La conseguenza importante è che per mostrare che $W(x) \neq 0$ per ogni $x \in I$ è sufficiente mostrare che $W(x_0) \neq 0$ per un solo $x_0 \in I$. Per apprezzare la rilevanza del risultato si veda il teorema 1.7 a pag. 24 del succitato volume di Bramanti, Pagani e Salsa.

Esercizi

 Trovare quattro matrici **A**, **B**, **C** e **D** quadrate di ordine 2 tali che

$$\det \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \neq \det(\mathbf{A}) \det(\mathbf{D}) - \det(\mathbf{B}) \det(\mathbf{C})$$

Mostrare che non è possibile prendere $\mathbf{B} = \mathbf{O}$ o $\mathbf{C} = \mathbf{O}$, perché se \mathbf{B} o \mathbf{C} è la matrice nulla l'uguaglianza vale certamente.

Autovalori e autovettori

1 INTRODUZIONE

In questo capitolo introduciamo le nozioni di *autovettore* (in inglese *eigenvector*) e di *autovalore* (in inglese *eigenvalue*) che sono, certamente, tra gli strumenti di algebra lineare più ricchi di applicazioni. Autovalori e autovettori servono, inoltre, a risolvere questo problema: data un'applicazione lineare \mathcal{L} di uno spazio vettoriale \mathbf{V} in se stesso, trovare la base di \mathbf{V} che semplifichi il più possibile l'espressione in coordinate di \mathcal{L} , cioè la matrice che rappresenta \mathcal{L} ; in altre parole il problema è quello di trovare le coordinate che meglio si adattino a trattare la funzione \mathcal{L} . Per esempio, se $\mathcal{L} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ è la proiezione ortogonale su una retta \mathbf{r} , la scelta naturale per le coordinate è di prendere \mathbf{r} come asse delle x , e una retta perpendicolare a \mathbf{r} come asse delle y : ciascuna di queste due rette ha la proprietà di essere lasciata invariata dall'applicazione \mathcal{L} , nel senso che \mathcal{L} manda un vettore di \mathbf{r} in un altro vettore di \mathbf{r} , mantenendone quindi invariata la direzione. In generale, vedremo che la chiave è di scegliere una base \mathcal{B} i cui assi coordinati siano rette \mathbf{r} che l'applicazione \mathcal{L} lascia invariate nel senso che $\mathcal{L}(\mathbf{r}) \subseteq \mathbf{r}$. Questo conduce alla nozione di *autovettore*: un autovettore è un vettore non nullo \mathbf{v} la cui direzione è lasciata invariata da \mathcal{L} . Le rette invarianti sono esattamente quelle che sono generate da un autovettore. Purtroppo, non sempre ci sono direzioni che un'applicazione lascia invarianti: per esempio una rotazione di un angolo retto del piano non lascia invariata alcuna direzione. Questo problema si risolve passando al campo complesso, dove invece direzioni invarianti esistono sempre: è per questo che la teoria degli autovettori richiede di lavorare nel campo complesso anche quando il problema di interesse richiede solo scalari reali. La nozione di autovettore è estremamente interessante anche nel caso di spazi funzionali. Il fatto che la funzione esponenziale sia un autovettore dell'operatore derivata, perché $\mathcal{D}(e^x) = e^x$, è alla base del metodo di soluzione delle equazioni lineari omogenee a coefficienti costanti.

2 AUTOVETTORI E AUTOVALORI DI UN'APPLICAZIONE LINEARE

Sia \mathbf{V} uno spazio vettoriale e sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{V}$ un'applicazione lineare. Ci poniamo il problema di trovare una base \mathcal{B} di \mathbf{V} tale che la matrice \mathbf{B} che rappresenta \mathcal{L} rispetto a \mathcal{B} sia la più semplice possibile. Cominciamo a vedere quale proprietà debba

avere la base \mathcal{B} affinché la matrice rappresentativa \mathbf{B} sia diagonale. Questo costituisce una drastica semplificazione: se la base non è scelta in maniera oculata, la matrice rappresentativa è una generica matrice quadrata di ordine $n = \dim \mathbf{V}$, che ha n^2 elementi; una matrice diagonale invece ha soltanto n elementi significativi, quelli sulla diagonale principale, tutti gli altri sono nulli; è evidente che poter fare i conti con n parametri anziché con n^2 costituisce una drastica semplificazione.

Come nel capitolo sulle matrici, denotiamo con il simbolo $\text{diag}(\lambda_1, \dots, \lambda_n)$ la matrice diagonale che ha λ_i come elemento di posto (i, i) :

$$\text{diag}(\lambda_1, \dots, \lambda_n) = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

Mostriamo ora come la ricerca di una base rispetto alla quale la matrice rappresentativa di \mathcal{L} sia diagonale conduca all'introduzione delle nozioni fondamentali di *autovettore* e *autovalore*.

PROPOSIZIONE 2.1 Sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{V}$ un'applicazione lineare, e sia $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ una base di \mathbf{V} . La matrice che rappresenta \mathcal{L} rispetto alla base \mathcal{B} è la matrice diagonale $\text{diag}(\lambda_1, \dots, \lambda_n)$ se e solo se

$$(2.1) \quad \mathcal{L}(\mathbf{v}_k) = \lambda_k \mathbf{v}_k \quad \text{per ogni } k = 1, 2, \dots, n$$

DIMOSTRAZIONE. Sia \mathbf{B} la matrice che rappresenta \mathcal{L} rispetto alla base \mathcal{B} . Per definizione la prima colonna di \mathbf{B}

$$\begin{bmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{n1} \end{bmatrix}$$

è il vettore delle coordinate di $\mathcal{L}(\mathbf{v}_1)$ rispetto alla base \mathcal{B} . Questo significa che:

$$\mathcal{L}(\mathbf{v}_1) = b_{11}\mathbf{v}_1 + b_{21}\mathbf{v}_2 + \dots + b_{n1}\mathbf{v}_n$$

Quindi la prima colonna di \mathbf{B} è

$$\begin{bmatrix} \lambda_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

se e solo se $\mathcal{L}(\mathbf{v}_1) = \lambda_1 \mathbf{v}_1$.

Analogamente, l'uguaglianza $\mathcal{L}(\mathbf{v}_k) = \lambda_k \mathbf{v}_k$ vale se e solo se la k -esima colonna di \mathbf{B} coincide con la k -esima colonna della matrice diagonale $\text{diag}(\lambda_1, \dots, \lambda_n)$ per ogni $k = 1, 2, \dots, n$.

Per poter trovare una base rispetto alla quale la matrice di \mathcal{L} sia diagonale occorre, quindi, studiare i vettori \mathbf{v} non nulli (altrimenti non possono fare parte di una base)

per i quali esiste uno scalare λ per cui $\mathcal{L}(\mathbf{v}) = \lambda\mathbf{v}$. Questi vettori si dicono *autovettori* di \mathcal{L} :

DEFINIZIONE 2.2 (Autovalori e autovettori di un'applicazione lineare)

Sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{V}$ un'applicazione lineare. Un vettore $\mathbf{v} \in \mathbf{V}$ si dice *autovettore* di \mathcal{L} se:

- a) \mathbf{v} non è il vettore nullo;
- b) esiste $\lambda \in \mathbb{K}$ tale che $\mathcal{L}(\mathbf{v}) = \lambda\mathbf{v}$.

Lo scalare λ è univocamente determinato da \mathbf{v} e si dice *autovalore* di \mathcal{L} relativo all'autovettore \mathbf{v} .

Un *autovalore* di \mathcal{L} è quindi uno scalare per cui esiste un vettore non nullo \mathbf{v} tale che l'equazione

$$(2.2) \quad \mathcal{L}(\mathbf{v}) = \lambda\mathbf{v}$$

sia soddisfatta. L'equazione (2.2) si dice *equazione degli autovettori e autovalori dell'applicazione* \mathcal{L} . Una soluzione dell'equazione è una coppia (\mathbf{v}, λ) formata da un autovettore e da un autovalore.

L'autovalore λ è univocamente determinato dal corrispondente autovettore \mathbf{v} : se $\mathcal{L}(\mathbf{v}) = \lambda\mathbf{v}$ ed $\mathcal{L}(\mathbf{v}) = \mu\mathbf{v}$, allora $\lambda = \mu$. Infatti

$$\lambda\mathbf{v} = \mu\mathbf{v} \Rightarrow (\lambda - \mu)\mathbf{v} = \mathbf{0}$$

da cui segue $\lambda - \mu = 0$ perché $\mathbf{v} \neq \mathbf{0}$.

Qual è il significato geometrico di questa definizione? Un vettore non nullo \mathbf{v} genera una retta $\mathbf{r} = \mathcal{L}(\mathbf{v})$, che è l'insieme dei vettori della forma $t\mathbf{v}$, al variare dello scalare t . Quindi \mathbf{v} è un autovettore di \mathcal{L} se e solo se \mathcal{L} manda \mathbf{v} in un vettore che appartiene ancora alla retta $\mathbf{r} = \mathcal{L}(\mathbf{v})$, in altre parole se e solo se $\mathcal{L}(\mathbf{v})$ è parallelo a \mathbf{v} . Se \mathbf{v} è un autovettore di \mathcal{L} relativo all'autovalore λ , allora ogni vettore non nullo di $\mathbf{r} = \mathcal{L}(\mathbf{v})$ è ancora un autovettore di \mathcal{L} relativo a λ poiché:

$$\mathcal{L}(t\mathbf{v}) = t\mathcal{L}(\mathbf{v}) = \lambda t\mathbf{v}$$

Quindi \mathcal{L} manda la retta \mathbf{r} in se stessa, e la restrizione di \mathcal{L} ai vettori della retta \mathbf{r} è la moltiplicazione per l'autovalore λ : l'autovalore λ è il numero (la matrice 1×1) che rappresenta l'applicazione $\mathcal{L} : \mathbf{r} \rightarrow \mathbf{r}$ rispetto alla base $\{\mathbf{v}\}$ di \mathbf{r} (e quindi rispetto a ogni altra base, perché \mathbf{r} ha dimensione 1). Geometricamente, la ricerca degli autovettori di \mathcal{L} coincide perciò con la ricerca delle rette (sottospazi di dimensione 1) mandate in se stesse da \mathcal{L} .

Nell'esempio a pagina 245 abbiamo determinato l'espressione in coordinate della riflessione ortogonale \mathcal{S} del piano cartesiano che ha per asse la retta \mathbf{r} generata dal vettore $[\cos(\alpha), \sin(\alpha)]^T$. La riflessione lascia fisso ogni vettore della retta \mathbf{r} : $\mathcal{S}(\mathbf{v}) = \mathbf{v}$ per ogni $\mathbf{v} \in \mathbf{r}$.

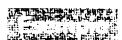
Questo significa che ogni vettore di \mathbf{r} è un autovettore di \mathcal{S} relativo all'autovalore 1. Inoltre, \mathcal{S} mappa ogni vettore \mathbf{w} della retta \mathbf{r}^\perp ortogonale a \mathbf{r} nel suo opposto: $\mathcal{S}(\mathbf{w}) = -\mathbf{w}$ per ogni $\mathbf{w} \in \mathbf{r}^\perp$. Ciò vuol dire che ogni vettore di \mathbf{r}^\perp è un autovettore di \mathcal{S} relativo all'autovalore -1. La base $\{\mathbf{b}_1 = [\cos(\alpha), \sin(\alpha)]^T, \mathbf{b}_2 = [-\sin(\alpha), \cos(\alpha)]^T\}$ è formata da autovettori di \mathcal{S} perché \mathbf{b}_1 appartiene a \mathbf{r} e \mathbf{b}_2 appartiene a \mathbf{r}^\perp ; per la proposizione 2.1 la matrice che rappresenta \mathcal{S} rispetto a tale base è

$$\mathbf{B} = \text{diag}(1, -1) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

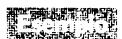
Rispetto alla base canonica la matrice di \mathcal{S} è (si veda pagina 245)

$$\mathbf{A} = \begin{bmatrix} \cos(2\alpha) & \sin(2\alpha) \\ \sin(2\alpha) & -\cos(2\alpha) \end{bmatrix}$$

La matrice *diagonale* \mathbf{B} è molto più semplice della matrice \mathbf{A} che rappresenta la stessa applicazione rispetto alla base canonica.



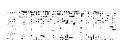
Se $\mathcal{L} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ denota la rotazione di un angolo di 90° , \mathcal{L} non ha nessun autovettore perché non lascia invariata alcuna direzione: nessun vettore viene mandato in un multiplo di se stesso. Quindi non esiste alcuna base di \mathbb{R}^2 rispetto alla quale la rotazione possa essere rappresentata da una matrice diagonale.



Sia $\mathcal{I} : \mathbf{V} \rightarrow \mathbf{V}$ l'applicazione identica:

$$\mathcal{I}(\mathbf{v}) = \mathbf{v} \quad \text{per ogni } \mathbf{v} \in \mathbf{V}$$

Ogni vettore non nullo di \mathbf{V} è un autovettore di \mathcal{I} relativo all'autovalore 1. Per la proposizione 2.1 la matrice che rappresenta \mathcal{I} rispetto a una qualunque base \mathbf{B} di \mathbf{V} è la matrice identità \mathbf{I} .



Gli autovettori di \mathcal{L} relativi all'autovalore 0 sono i vettori non nulli del nucleo $\text{Ker}(\mathcal{L})$ perché:

$$\mathcal{L}(\mathbf{v}) = 0\mathbf{v} \iff \mathcal{L}(\mathbf{v}) = 0\mathbf{v}$$



Sia \mathbf{V} uno spazio vettoriale e sia $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ una sua base. Consideriamo la proiezione sul primo asse coordinato: $\mathcal{P}(x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \dots + x_n\mathbf{b}_n) = x_1\mathbf{b}_1$. Il vettore \mathbf{b}_1 è un autovettore di \mathcal{P} relativo all'autovalore 1, i vettori $\mathbf{b}_2, \dots, \mathbf{b}_n$ sono autovettori relativi all'autovalore nullo. La matrice di \mathcal{P} rispetto alla base $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ è quindi $\text{diag}(1, 0, \dots, 0)$.

Il fatto che il nucleo di \mathcal{L} sia formato dagli autovettori relativi all'autovalore nullo, segue dalla semplice, ma fondamentale, osservazione che l'equazione degli autovettori e autovalori $\mathcal{L}(\mathbf{v}) = \lambda\mathbf{v}$ è lineare in \mathbf{v} una volta fissato λ (si osservi invece che l'equazione

non è lineare nel complesso (\mathbf{v}, λ) delle sue incognite). Per mettere in evidenza questo fatto, riscriviamo l'equazione nella forma

$$\mathcal{L}(\mathbf{v}) = \lambda\mathbf{v} = \lambda\mathcal{I}(\mathbf{v})$$

dove \mathcal{I} è l'applicazione identità. Il primo membro dell'equazione ha la stessa forma del secondo: un'applicazione lineare applicata a \mathbf{v} . Possiamo, quindi, portare a primo membro $\lambda\mathcal{I}(\mathbf{v})$ e ottenere:

$$(2.3) \quad (\mathcal{L} - \lambda\mathcal{I})(\mathbf{v}) = \mathbf{0}$$

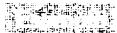
Abbiamo così separato i ruoli delle incognite λ e \mathbf{v} . Il grosso vantaggio è che, fissato λ , l'equazione (2.3) è un'equazione lineare omogenea in \mathbf{v} : un vettore \mathbf{v} è un autovettore di \mathcal{L} se e solo se è un vettore non nullo nel nucleo di $\mathcal{L} - \lambda\mathcal{I}$.

DEFINIZIONE 2.3 (Autospazi di un'applicazione lineare)

Per ogni autovalore $\lambda \in \mathbb{K}$ poniamo

$$\mathbf{V}_\lambda = \text{Ker}(\mathcal{L} - \lambda\mathcal{I}) = \{\mathbf{v} \in \mathbf{V} : \mathcal{L}(\mathbf{v}) = \lambda\mathbf{v}\}$$

Il sottospazio \mathbf{V}_λ di \mathbf{V} si dice *autospazio* di \mathcal{L} relativo all'autovalore λ . I suoi elementi non nulli sono precisamente gli autovettori di \mathcal{L} relativi all'autovalore fissato λ .



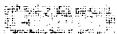
In questo esempio ricaviamo gli autospazi dell'operatore derivata. Quindi $\mathcal{L} = \mathcal{D}$ è la funzione che associa a una funzione $y(x)$ la sua derivata $y'(x)$. L'equazione degli autovettori e autovalori diviene

$$y'(x) = \lambda y(x) \quad \text{per ogni } x \in \mathbb{R}$$

Non è difficile mostrare (per esempio, dividendo per $y(x)$) che le soluzioni di questa equazione differenziale, con λ fissato, sono le funzioni esponenziali

$$y(x) = Ce^{\lambda x}$$

dove $C \in \mathbb{R}$ è una costante arbitraria. Quindi l'autospazio \mathbf{V}_λ ha dimensione uno e ha come base la funzione esponenziale $e^{\lambda x}$.



Illustriamo ora come la conoscenza di autovettori e autovalori consenta di trovare soluzioni di sistemi di equazioni differenziali lineari a coefficienti costanti. Questa è una delle principali applicazioni degli autovettori e autovalori. Consideriamo il sistema di equazioni differenziali:

$$(2.4) \quad \mathbf{x}' = \mathbf{A}\mathbf{x}$$

dove \mathbf{A} è una matrice $n \times n$ e l'incognita $\mathbf{x} = \mathbf{x}(t)$ è una n -upla di funzioni differenziabili:

$\mathbf{x} = [x_1(t), \dots, x_n(t)]^T$. Per esempio, se $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ e $\mathbf{x} = [x, y]^T$, il sistema (2.4) è

$$\begin{cases} x'(t) = ax(t) + by(t) \\ y'(t) = cx(t) + dy(t) \end{cases}$$

Si dice che un vettore $\mathbf{w} = [w_1, \dots, w_n]^T \in \mathbb{K}^n$ è un autovettore della matrice \mathbf{A} se \mathbf{w} è un autovettore dell'applicazione $\mathfrak{L}_\mathbf{A}$, quindi se esiste $\lambda \in \mathbb{K}$ tale che $\mathbf{Aw} = \lambda\mathbf{w}$. Ebbene, se $\mathbf{Aw} = \lambda\mathbf{w}$, allora la funzione vettoriale

$$\mathbf{u}(t) = e^{\lambda t} \mathbf{w} = [w_1 e^{\lambda t}, w_2 e^{\lambda t}, \dots, w_n e^{\lambda t}]^T$$

è una soluzione del sistema (2.4). Per mostrare questo, basta ricordare come si deriva la funzione esponenziale:

$$\mathbf{u}'(t) = \lambda e^{\lambda t} \mathbf{w}$$

Quindi

$$\mathbf{u}'(t) = e^{\lambda t} \lambda \mathbf{w} = e^{\lambda t} \mathbf{Aw} = \mathbf{A}(e^{\lambda t} \mathbf{w}) = \mathbf{A}\mathbf{u}(t)$$

Questo vale anche per λ e \mathbf{w} complessi e vedremo tra breve perché questo è utile anche se si è interessati solo a soluzioni reali.

Per un esempio specifico, consideriamo il sistema

$$(2.5) \quad \begin{cases} x' = x + y \\ y' = x + y \end{cases}$$

In questo caso la matrice del sistema è $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Il vettore $\mathbf{w}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ è un autovettore relativo a $\lambda_1 = 0$, e il vettore $\mathbf{w}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ è un autovettore relativo a $\lambda_2 = 2$. Il sistema ha perciò le due soluzioni

$$\mathbf{u}_1(t) = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \text{e} \quad \mathbf{u}_2(t) = e^{2t} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Si può mostrare che ogni altra soluzione è combinazione lineare di queste due soluzioni.

Per un altro esempio, consideriamo il sistema

$$(2.6) \quad \begin{cases} x' = x + y \\ y' = -x + y \end{cases}$$

In questo caso la matrice del sistema è $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$. Vedremo più avanti come determinare gli autovettori di \mathbf{A} : la matrice in questione non ha autovettori reali, ma ha l'autovettore complesso $\mathbf{w} = \begin{bmatrix} 1 \\ i \end{bmatrix}$ relativo all'autovalore $\lambda = 1 + i$ perché:

$$\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = \begin{bmatrix} 1+i \\ -1+i \end{bmatrix} = (1+i) \begin{bmatrix} 1 \\ i \end{bmatrix}$$

Quindi il sistema ha la soluzione complessa

$$\mathbf{u}(t) = e^{(1+i)t} \begin{bmatrix} 1 \\ i \end{bmatrix}$$

Scomponiamo tale soluzione come somma di una parte reale e di una parte immaginaria:

$$\mathbf{u}(t) = e^{(1+i)t} \begin{bmatrix} 1 \\ i \end{bmatrix} = e^t e^{it} \begin{bmatrix} 1 \\ i \end{bmatrix} = e^t (\cos(t) + i \sin(t)) \begin{bmatrix} 1 \\ i \end{bmatrix} = \begin{bmatrix} e^t \cos(t) \\ -e^t \sin(t) \end{bmatrix} + i \begin{bmatrix} e^t \sin(t) \\ e^t \cos(t) \end{bmatrix}$$

Si mostra che la parte reale e la parte immaginaria

$$\mathbf{u}_1(t) = \begin{bmatrix} e^t \cos(t) \\ -e^t \sin(t) \end{bmatrix} \quad \text{e} \quad \mathbf{u}_2(t) = \begin{bmatrix} e^t \sin(t) \\ e^t \cos(t) \end{bmatrix}$$

sono due soluzioni reali del sistema (2.6) (il lettore verifichi!). E si può anche mostrare che ogni soluzione del sistema è combinazione lineare di queste due e ha quindi la forma

$$\begin{bmatrix} c_1 e^t \cos(t) + c_2 e^t \sin(t) \\ -c_1 e^t \sin(t) + c_2 e^t \cos(t) \end{bmatrix}$$

Determinare per quali valori di m il vettore $[1, m]^T$ è un autovettore della riflessione ortogonale $\mathcal{L}([x, y]^T) = [y, x]^T$. Per tali valori, qual è l'autovalore corrispondente all'autovettore $[1, m]^T$?

Verificare che $\mathbf{e}_3 = [0, 0, 1]^T$ è un autovettore relativo all'autovalore 1 dell'applicazione $\mathcal{L}([x, y, z]^T) = [-y, x, z]^T$. Mostrare che l'autospazio di \mathcal{L} relativo all'autovalore 1 è l'asse z , e convincersi che \mathcal{L} è la rotazione dello spazio attorno all'asse z di un angolo retto in senso antiorario. Concludere che \mathcal{L} non ha altri autovettori reali (non ci sono altre direzioni, a parte l'asse, lasciate fisse dalla rotazione). Scrivere la matrice che rappresenta \mathcal{L} rispetto alla base canonica.

Mostrare che, se \mathbf{v} è un autovettore di \mathcal{L} relativo all'autovalore 3, allora è anche un autovettore di \mathcal{L}^2 relativo all'autovalore 9.

Suggerimento: $\mathcal{L}^2(\mathbf{v}) = \mathcal{L}(\mathcal{L}(\mathbf{v}))$.

Sia $\mathcal{L} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ l'applicazione lineare definita da

$$\mathcal{L} \begin{pmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \end{pmatrix} = \begin{bmatrix} \frac{1}{3}x + \frac{1}{3}y - \frac{1}{3}z \\ \frac{1}{3}x + \frac{5}{6}y + \frac{1}{6}z \\ -\frac{1}{3}x + \frac{1}{6}y + \frac{5}{6}z \end{bmatrix}$$

- Determinare una base del nucleo di \mathcal{L} .
- Determinare una base dell'immagine di \mathcal{L} .
- Mostrare che i vettori non nulli del nucleo e dell'immagine sono autovettori di \mathcal{L} .
- Determinare una base di \mathbb{R}^3 formata da autovettori di \mathcal{L} . Scrivere le matrici che rappresentano \mathcal{L} rispetto a tale base e rispetto alla base canonica.

3 AUTOVETTORI E AUTOVALORI DI UNA MATRICE

Per risolvere l'equazione degli autovettori e autovalori di un'applicazione lineare $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{V}$, il primo passo da fare è fissare delle coordinate, cioè una base di \mathbf{V} , in modo da poter svolgere i conti sulla matrice \mathbf{A} che rappresenta \mathcal{L} . La matrice \mathbf{A} è quadrata di ordine $n = \dim \mathbf{V}$. Supponiamo che $\mathcal{L}(\mathbf{v}) = \lambda \mathbf{v}$, cioè che \mathbf{v} sia un autovettore di \mathcal{L} relativo all'autovalore λ . Se \mathbf{x} è il vettore delle coordinate di \mathbf{v} rispetto alla base fissata, l'equazione $\mathcal{L}(\mathbf{v}) = \lambda \mathbf{v}$ è equivalente ad $\mathbf{Ax} = \lambda \mathbf{x}$. Possiamo quindi riformulare la definizione di autovettore e autovalore in termini di matrici e vettori colonna:

DEFINIZIONE 3.1 (Autovalori e autovettori di una matrice)

Sia \mathbf{A} una matrice quadrata di ordine n (a elementi reali o complessi). Un vettore non nullo $\mathbf{x} \in \mathbb{C}^n$ si dice *autovettore* di \mathbf{A} se esiste uno scalare $\lambda \in \mathbb{C}$ tale che

$$(3.1) \quad \mathbf{Ax} = \lambda \mathbf{x}$$

Lo scalare λ è univocamente determinato da \mathbf{x} e si dice *autovalore* di \mathbf{A} relativo all'autovettore \mathbf{x} .

Un autovalore di \mathbf{A} è quindi un numero complesso λ per cui esiste un vettore non nullo $\mathbf{x} \in \mathbb{C}^n$ tale che l'equazione (3.1) sia soddisfatta. L'equazione si dice *equazione degli autovettori e autovalori della matrice \mathbf{A}* . Una soluzione dell'equazione è una coppia (\mathbf{v}, λ) formata da un autovettore e da un autovalore. È utile considerare autovettori e autovalori complessi anche nel caso di matrici reali, perché come vedremo negli esempi esistono matrici reali che non hanno autovalori reali.

OSSERVAZIONE Il fatto che $\mathbf{x} \in \mathbb{C}^n$ sia un autovettore della matrice \mathbf{A} relativo a λ è equivalente al fatto che \mathbf{x} sia un autovettore relativo a λ dell'applicazione lineare $\mathcal{L}_{\mathbf{A}} : \mathbb{C}^n \rightarrow \mathbb{C}^n$, dove $\mathcal{L}_{\mathbf{A}} : \mathbb{C}^n \rightarrow \mathbb{C}^n$ denota come al solito moltiplicazione a sinistra per \mathbf{A} : $\mathcal{L}_{\mathbf{A}}(\mathbf{x}) = \mathbf{Ax}$. Quanto visto nel paragrafo precedente si applica quindi anche al caso di autovettori e autovalori di una matrice.

Un vettore non nullo del nucleo di \mathbf{A} è un autovettore di \mathbf{A} relativo all'autovalore $\lambda = 0$.

Il vettore $\mathbf{v} = [1, 1]^T$ è un autovettore relativo all'autovalore $\lambda = 3$ della matrice $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$ perché

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Il vettore $\mathbf{e}_2 = [0, 1]^T$ non è un autovettore della matrice \mathbf{A} perché

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \neq \lambda \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{per ogni } \lambda$$

Il vettore $\mathbf{e}_1 = [1, 0]^T$ è un autovettore della matrice \mathbf{A} relativo all'autovalore $\lambda = 2$ perché

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Denotiamo con \mathbf{e}_k il k -esimo vettore della base canonica di \mathbb{K}^n . Se \mathbf{A} è una matrice quadrata di ordine n , il prodotto $\mathbf{A}\mathbf{e}_k$ è la k -esima colonna di \mathbf{A} . Da questo segue:

a) se \mathbf{D} è la matrice diagonale $\text{diag}(\lambda_1, \dots, \lambda_n)$, allora

$$\mathbf{D}\mathbf{e}_k = \lambda_k \mathbf{e}_k$$

quindi, \mathbf{e}_k è un autovettore di \mathbf{D} relativo all'autovalore λ_k .

b) Se $\mathbf{U} = [u_{ij}]$ è triangolare alta, allora

$$\mathbf{U}\mathbf{e}_1 = u_{11}\mathbf{e}_1$$

quindi \mathbf{e}_1 è un autovettore di \mathbf{U} relativo all'autovalore $\lambda = u_{11}$.

c) Se $\mathbf{L} = [l_{ij}]$ è triangolare bassa, allora

$$\mathbf{L}\mathbf{e}_n = l_{nn}\mathbf{e}_n$$

quindi \mathbf{e}_n è un autovettore di \mathbf{L} relativo all'autovalore $\lambda = l_{nn}$.

In termini di matrici, il problema che ci siamo posti nel paragrafo precedente diviene: data una matrice quadrata \mathbf{A} di ordine n , trovare una base di \mathbb{K}^n che semplifichi il più possibile l'espressione in coordinate dell'applicazione $\mathcal{L}_{\mathbf{A}}$. Ricordiamo che la matrice \mathbf{A} rappresenta l'applicazione $\mathcal{L}_{\mathbf{A}}$ rispetto alla base canonica di \mathbb{K}^n ; nel capitolo sulle applicazioni lineari (proposizione 5.5 e formula (5.13) a pagina 250) abbiamo visto come cambi la matrice rappresentativa al variare della base: la matrice \mathbf{B} che rappresenta $\mathcal{L}_{\mathbf{A}}$ rispetto a una nuova base $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ di \mathbb{K}^n è

$$(3.2) \quad \mathbf{B} = \mathbf{S}^{-1}\mathbf{AS}$$

dove \mathbf{S} è la matrice di passaggio dalla base canonica alla base \mathcal{B} , le cui colonne sono i vettori della nuova base:

$$\mathbf{S} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_n]$$

Traduciamo quindi la proposizione 2.1 in termini di matrici

PROPOSIZIONE 3.2 Sia \mathbf{A} una matrice quadrata di ordine n a elementi in \mathbb{K} , e supponiamo che

$$\mathbf{S} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_n]$$

sia una matrice invertibile in $\mathbb{M}_{\mathbb{K}}(n, n)$. Allora

$$(3.3) \quad \mathbf{S}^{-1}\mathbf{AS} = \text{diag}(\lambda_1, \dots, \lambda_n)$$

se e solo se

$$(3.4) \quad \mathbf{Ab}_k = \lambda_k \mathbf{b}_k \quad \text{per ogni } k = 1, 2, \dots, n$$

DIMOSTRAZIONE. L'enunciato segue dalla proposizione 2.1 scegliendo $\mathcal{L} = \mathcal{L}_{\mathbf{A}}$ e tenendo conto che \mathbf{B} rappresenta $\mathcal{L}_{\mathbf{A}}$ rispetto alla base $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$. Ne diamo però una dimostrazione diretta. Poniamo $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_n)$. L'uguaglianza $\mathbf{S}^{-1}\mathbf{AS} = \mathbf{D}$ è equivalente a

$$\mathbf{AS} = \mathbf{SD}$$

La k -esima colonna di \mathbf{AS} è il prodotto \mathbf{Ab}_k di \mathbf{A} per la k -esima colonna di \mathbf{S} . Per concludere basta osservare che la k -esima colonna di \mathbf{SD} è $\lambda_k \mathbf{b}_k$:

$$\mathbf{SD} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_n] \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} = [\lambda_1 \mathbf{b}_1 \ \lambda_2 \mathbf{b}_2 \ \cdots \ \lambda_n \mathbf{b}_n]$$

DEFINIZIONE 3.3 Una matrice quadrata reale \mathbf{A} di ordine n si dice *diagonalizzabile su \mathbb{R}* se esiste una matrice quadrata reale e invertibile \mathbf{S} di ordine n tale che $\mathbf{S}^{-1}\mathbf{AS}$ sia una matrice diagonale. Una matrice quadrata reale o complessa \mathbf{A} di ordine n si dice *diagonalizzabile su \mathbb{C}* se esiste una matrice quadrata complessa e invertibile \mathbf{S} di ordine n tale che $\mathbf{S}^{-1}\mathbf{AS}$ sia una matrice diagonale.

OSSERVAZIONE Per una matrice quadrata reale si ottengono così due *diverse* nozioni di diagonalizzabilità. Poiché una matrice reale \mathbf{S} è anche una matrice complessa, una matrice diagonalizzabile su \mathbb{R} è anche diagonalizzabile su \mathbb{C} , ma il viceversa non è vero. Per esempio, come vedremo più avanti, la matrice

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

è diagonalizzabile su \mathbb{C} , ma non su \mathbb{R} (il motivo è che la matrice rappresenta la rotazione del piano di un angolo retto, che non lascia invariata alcuna retta e quindi non ammette autovettori e autovalori *reali*).

La proposizione 3.2 fornisce il seguente criterio di diagonalizzabilità:

PROPOSIZIONE 3.4 (Primo criterio di diagonalizzabilità)

Una matrice quadrata \mathbf{A} di ordine n è diagonalizzabile su \mathbb{K} se e solo se \mathbb{K}^n ha una base formata da autovettori di \mathbf{A} .

DIMOSTRAZIONE. Nelle notazioni della proposizione 3.2, la matrice \mathbf{S} è invertibile se e solo se le sue colonne formano una base di \mathbb{K}^n e $\mathbf{S}^{-1}\mathbf{AS}$ è diagonale se e solo se le colonne di \mathbf{S} sono autovettori di \mathbf{A} .

In altre parole, una matrice quadrata \mathbf{A} di ordine n è diagonalizzabile su \mathbb{K} se e solo se possiamo trovare n autovettori di \mathbf{A} in \mathbb{K}^n che siano linearmente indipendenti. Discuteremo nel paragrafo seguente l'esistenza di tali autovettori.

Una matrice diagonale è diagonalizzabile (prendere $\mathbf{S} = \mathbf{I}$); una base di \mathbb{K}^n formata da autovettori della matrice è la base canonica.

Per quanto visto a pagina 251 la matrice

$$\mathbf{A} = \begin{bmatrix} \cos(2\alpha) & \sin(2\alpha) \\ \sin(2\alpha) & -\cos(2\alpha) \end{bmatrix}$$

è diagonalizzabile mediante la matrice di passaggio

$$\mathbf{S} = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix}$$

Questo si può vedere anche osservando che le colonne \mathbf{b}_1 e \mathbf{b}_2 di \mathbf{S} sono autovettori di \mathbf{A} relativi agli autovalori $\lambda_1 = 1$ e $\lambda_2 = -1$, per cui

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

Determinare per quali valori di m il vettore $[1, m]^T$ è un autovettore della matrice $\begin{bmatrix} 3 & 7 \\ 0 & 2 \end{bmatrix}$. Per tali valori determinare l'autovalore corrispondente all'autovettore $[1, m]^T$.

Determinare per quali valori di k il vettore $[0, 1, 1]^T$ (rispettivamente $[0, 1, -1]^T$) è un autovettore della matrice

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & k & 0 \\ 3 & 0 & 1 \end{bmatrix}$$

Quali sono gli autovalori corrispondenti?

Mostrare che, se \mathbf{v} è un autovettore di \mathbf{A} relativo all'autovalore 2, allora è anche un autovettore di \mathbf{A}^2 relativo all'autovalore 4.

Suggerimento: $\mathbf{A}^2 \mathbf{v} = \mathbf{A}(\mathbf{Av})$.

4 RICERCA DI AUTOVALORI E AUTOVETTORI

Per trovare gli autovettori e gli autovalori di un matrice \mathbf{A} , occorre risolvere l'equazione

$$(4.1) \quad \mathbf{Ax} = \lambda \mathbf{x}$$

La difficoltà nel risolvere l'equazione degli autovettori e autovalori (4.1) è che a secondo membro compaiono entrambe le incognite \mathbf{x} e λ e che comunque l'equazione non è lineare nel complesso delle sue incognite (\mathbf{x}, λ). Il primo passo nella soluzione è di disaccoppiare i ruoli di \mathbf{x} e λ . Per questo riscriviamo il secondo membro nella stessa forma del primo: matrice per vettore, moltiplicando il vettore \mathbf{x} per la matrice identità:

$$\mathbf{Ax} = \lambda \mathbf{Ix}$$

Ora portiamo tutto a primo membro, raccogliamo \mathbf{x} e riscriviamo l'equazione nella forma equivalente

$$(4.2) \quad (\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0}$$

Abbiamo così disaccoppiato i ruoli di \mathbf{x} e λ : affinché λ sia un autovalore è necessario e sufficiente che esista una soluzione non nulla \mathbf{x} del sistema lineare omogeneo (4.2). Una soluzione non nulla esiste se e solo se il determinante della matrice $\mathbf{A} - \lambda \mathbf{I}$ è nullo. In conclusione, abbiamo dimostrato:

PROPOSIZIONE 4.1 Uno scalare $\lambda_0 \in \mathbb{C}$ è un autovalore di \mathbf{A} se e solo se λ_0 è una soluzione dell'equazione

$$(4.3) \quad \det(\mathbf{A} - \lambda \mathbf{I}) = 0$$

che si dice *equazione caratteristica di \mathbf{A}* .

Si noti che abbiamo così ottenuto un'equazione $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$ per i soli autovalori. La ricerca degli autovettori e autovalori di \mathbf{A} si effettua quindi in due passi:

- si trovano gli autovalori di \mathbf{A} risolvendo l'equazione caratteristica $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$.
- per ogni autovalore λ_0 trovato nel primo passo si risolve il sistema omogeneo:

$$(\mathbf{A} - \lambda_0 \mathbf{I})\mathbf{x} = \mathbf{0}$$

Le soluzioni non nulle sono gli autovettori di \mathbf{A} relativi all'autovalore λ_0 .

Polinomio caratteristico

Analizziamo ora l'equazione caratteristica (4.3) le cui soluzioni sono gli autovalori della matrice. Faremo vedere che si tratta di un'equazione algebrica di grado n in λ . La conseguenza è che una matrice quadrata di ordine n ha esattamente n autovalori complessi (se contati con l'opportuna molteplicità).

PROPOSIZIONE 4.2 (Polinomio caratteristico)

Sia \mathbf{A} una matrice quadrata di ordine n . La funzione $\det(\mathbf{A} - \lambda \mathbf{I})$ è un polinomio di grado n nella variabile λ :

$$\det(\mathbf{A} - \lambda \mathbf{I}) = (-1)^n \lambda^n + c_1 \lambda^{n-1} + \cdots + c_{n-1} \lambda + c_n$$

che si dice *polinomio caratteristico* di \mathbf{A} . Il secondo e l'ultimo coefficiente sono:

$$\begin{aligned} c_1 &= (-1)^{n-1} \operatorname{tr}(\mathbf{A}) \\ c_n &= \det(\mathbf{A}) \end{aligned}$$

OSSERVAZIONE Si ricordi che la traccia $\operatorname{tr}(\mathbf{A})$ è la somma degli elementi di \mathbf{A} sulla diagonale principale:

$$\operatorname{tr}(\mathbf{A}) = a_{11} + \cdots + a_{nn}$$

Per la proposizione la traccia è, a meno del segno, il coefficiente di λ^{n-1} nel polinomio caratteristico, mentre il determinante è il termine noto di tale polinomio.

In particolare, se \mathbf{A} ha ordine due, il polinomio caratteristico di \mathbf{A} è

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \lambda^2 - \text{tr}(\mathbf{A})\lambda + \det(\mathbf{A})$$

mentre se \mathbf{A} ha ordine tre, il polinomio caratteristico è

$$\det(\mathbf{A} - \lambda\mathbf{I}) = -\lambda^3 + \text{tr}(\mathbf{A})\lambda^2 + c_2\lambda + \det(\mathbf{A})$$

Si può mostrare che il coefficiente c_k di λ^{n-k} è, a meno del segno, la somma dei minori principali di ordine k di \mathbf{A} .

DIMOSTRAZIONE (della proposizione 4.2). Usiamo la formula per il determinante:

$$\det(\mathbf{A}) = \sum_{\sigma \in S_n} (-1)^\sigma a_{1\sigma(1)} \cdots a_{n\sigma(n)}$$

Tale formula mostra che il determinante è una somma di monomi di grado n nelle variabili a_{ij} . Il polinomio caratteristico si ottiene sostituendo a_{ii} con $a_{ii} - \lambda$; perciò un polinomio di grado al più n in λ .

Denotiamo con il simbolo $a_{ij}(\lambda)$ gli elementi della matrice $\mathbf{A} - \lambda\mathbf{I}$: $a_{ii}(\lambda) = a_{ii} - \lambda$ e $a_{ij}(\lambda) = a_{ij}$ se $i \neq j$. Nel prodotto

$$a_{1\sigma(1)}(\lambda) \cdots a_{n\sigma(n)}(\lambda)$$

λ appare solo nei fattori in cui $\sigma(i) = i$. Tale prodotto ha perciò grado al massimo $n - 2$ in λ a meno che $\sigma(i) = i$ per almeno $n - 1$ indici. Ma se σ soddisfa $\sigma(i) = i$ per $n - 1$ indici i , allora σ deve fissare anche l'ultimo indice e quindi σ è l'identità. Ne segue che i termini di grado n e $n - 1$ del polinomio caratteristico coincidono coi termini di grado n e $n - 1$ dell'addendo nella formula del determinante che corrisponde alla permutazione identità; tale addendo è

$$\begin{aligned} a_{11}(\lambda) \cdots a_{nn}(\lambda) &= (a_{11} - \lambda) \cdots (a_{nn} - \lambda) = \\ &= (-1)^n \lambda^n + (-1)^{n-1} (a_{11} + \cdots + a_{nn}) \lambda^{n-1} + \text{termini di grado minore in } \lambda = \\ &= (-1)^n \lambda^n + (-1)^{n-1} \text{tr}(\mathbf{A}) \lambda^{n-1} + \text{termini di grado minore in } \lambda \end{aligned}$$

Quindi i primi due coefficienti del polinomio caratteristico sono $(-1)^n$ e $(-1)^{n-1} \text{tr}(\mathbf{A})$. Per calcolare il termine noto c_n del polinomio basta porre $\lambda = 0$:

$$c_n = \det(\mathbf{A} - 0\mathbf{I}) = \det(\mathbf{A})$$



Richiamiamo ora alcune proprietà delle radici di un polinomio qualsiasi; ne ricaveremo informazioni sugli autovalori di \mathbf{A} , che sono le radici del polinomio caratteristico. Se $P(\lambda)$ è un polinomio di grado n a coefficienti reali o complessi, un numero complesso λ_1 si dice *radice* di $P(\lambda)$ se $P(\lambda_1) = 0$. Per esempio il polinomio reale $\lambda^2 + 1$ non ha radici reali, ma ha due radici complesse $\lambda_1 = i$ e $\lambda_2 = -i$. Se λ_1 è una radice di $P(\lambda)$, allora per il teorema di Ruffini $P(\lambda)$ è divisibile per $(\lambda - \lambda_1)$, il che significa che esiste un polinomio $Q(\lambda)$ di grado $n - 1$ tale che

$$P(\lambda) = (\lambda - \lambda_1)Q(\lambda)$$

Il quoziente $Q(\lambda)$ ha coefficienti reali se $P(\lambda)$ ha coefficienti reali e $\lambda_1 \in \mathbb{R}$, ma avrà coefficienti complessi in generale: per esempio, se $P(\lambda)$ ha coefficienti reali, ma λ_1 non è reale, $Q(\lambda)$ non ha tutti i coefficienti reali (per esempio, $\lambda^2 + 1 = (\lambda - i)(\lambda + i)$).

Se $Q(\lambda_1) \neq 0$, si dice che λ_1 è una radice semplice di $P(\lambda)$. In ogni caso esiste un unico intero $m \geq 1$ tale che

$$P(\lambda) = (\lambda - \lambda_1)^m Q(\lambda)$$

con $Q(\lambda_1) \neq 0$. Il numero m si dice molteplicità della radice λ_1 : intuitivamente, λ_1 è m volte radice di $P(\lambda)$. Per esempio, il polinomio

$$P(\lambda) = \lambda^3 - 2\lambda^2 + \lambda = \lambda(\lambda - 1)^2$$

ha la radice semplice $\lambda_1 = 0$ e la radice $\lambda_2 = 1$ di molteplicità 2 (radice doppia).

Il teorema fondamentale dell'algebra afferma che un polinomio a coefficienti complessi si decomponga completamente in \mathbb{C} : questo significa che un polinomio $P(\lambda)$ di grado n si scomponga nel prodotto di n fattori lineari:

$$P(\lambda) = c_0(\lambda - \lambda_1) \cdots (\lambda - \lambda_n)$$

con $c_0, \lambda_1, \dots, \lambda_n \in \mathbb{C}$ (c_0 è il coefficiente di λ^n , che è non nullo perché $P(\lambda)$ ha grado n ed è irrilevante in questa discussione). In altre parole, un polinomio complesso di grado n ha esattamente n radici $\lambda_1, \lambda_2, \dots, \lambda_n$, che possono essere però ripetute: se λ_1 ha molteplicità m , di tali radici m saranno uguali a λ_1 . Quindi il grado del polinomio è uguale al numero delle radici complesse contate con la loro molteplicità.

Questo non è vero in \mathbb{R} perché un polinomio reale potrebbe avere radici non reali, ma comunque un polinomio a coefficienti reali si decomponga completamente in \mathbb{C} e ha un numero di radici complesse pari al suo grado.

Tornando al caso del polinomio caratteristico di una matrice, sappiamo che le sue radici sono gli autovalori della matrice. La molteplicità algebrica a_{λ_0} di un autovalore λ_0 è la sua molteplicità come radice del polinomio caratteristico.

PROPOSIZIONE 4.3 Sia \mathbf{A} una matrice quadrata reale o complessa di ordine n . Allora \mathbf{A} ha esattamente n autovalori complessi $\lambda_1, \lambda_2, \dots, \lambda_n$ (ripetuti secondo la loro molteplicità algebrica). Inoltre la traccia e il determinante di \mathbf{A} sono, rispettivamente, la somma e il prodotto degli autovalori:

$$\begin{aligned} \text{tr}(\mathbf{A}) &= \lambda_1 + \lambda_2 + \cdots + \lambda_n \\ \det(\mathbf{A}) &= \lambda_1 \lambda_2 \cdots \lambda_n \end{aligned}$$

DIMOSTRAZIONE. Il polinomio caratteristico ha esattamente n radici complesse $\lambda_1, \dots, \lambda_n$, quindi

$$\det(\mathbf{A} - \lambda \mathbf{I}) = (-1)^n (\lambda - \lambda_1) \cdots (\lambda - \lambda_n)$$

Siccome gli autovalori sono le radici del polinomio caratteristico, $\lambda_1, \lambda_2, \dots, \lambda_n$ sono gli autovalori, contati con la loro molteplicità algebrica. D'altra parte, abbiamo sviluppato il polinomio caratteristico come

$$\det(\mathbf{A} - \lambda \mathbf{I}) = (-1)^n \lambda^n + (-1)^{n-1} \text{tr}(\mathbf{A}) \lambda^{n-1} + \cdots + c_{n-1} \lambda + \det(\mathbf{A})$$

Uguagliando i coefficienti di grado $n - 1$ e di grado 0 nelle due espressioni del polinomio caratteristico troviamo che la traccia e il determinante di \mathbf{A} sono, rispettivamente, la somma e il prodotto degli autovalori.

Si considerino le tre matrici

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{A}_3 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

I polinomi caratteristici delle tre matrici sono

$$\det(\mathbf{A}_1 - \lambda\mathbf{I}) = \lambda^2 - 1, \quad \det(\mathbf{A}_2 - \lambda\mathbf{I}) = \lambda^2 + 1, \quad \det(\mathbf{A}_3 - \lambda\mathbf{I}) = \lambda^2$$

La matrice \mathbf{A}_1 ha, quindi, due autovalori reali semplici (cioè di molteplicità algebrica 1) $\lambda_1 = 1$ e $\lambda_2 = -1$. La matrice \mathbf{A}_2 ha due autovalori complessi semplici $\lambda_1 = i$ e $\lambda_2 = -i$. La matrice \mathbf{A}_3 ha un unico autovalore doppio $\lambda = 0$.

Il determinante di una matrice triangolare è il prodotto degli elementi sulla diagonale principale. Se \mathbf{T} è triangolare, anche $\mathbf{T} - \lambda\mathbf{I}$ è triangolare, per cui:

Polinomio caratteristico di una matrice triangolare

Il polinomio caratteristico di una matrice *triangolare* $\mathbf{T} = [a_{ij}]$ è:

$$\det(\mathbf{T} - \lambda\mathbf{I}) = (a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda).$$

Gli autovalori di \mathbf{T} sono pertanto gli elementi sulla diagonale principale.

In particolare, gli autovalori di una matrice diagonale $\text{diag}(\lambda_1, \dots, \lambda_n)$ sono proprio $\lambda_1, \dots, \lambda_n$.

Per esempio:

- a) la matrice identità e la matrice nulla di ordine n hanno un unico autovalore di molteplicità algebrica n ; più in generale la matrice

$$a\mathbf{I} = \text{diag}(a, a, \dots, a)$$

ha solo l'autovalore $\lambda = a$ con molteplicità algebrica n ;

- b) la matrice triangolare alta

$$\mathbf{U} = \begin{bmatrix} 1 & 2 & \pi \\ 0 & 3 & \sqrt{2} \\ 0 & 0 & 1 \end{bmatrix}$$

ha un autovalore doppio $\lambda_1 = 1$ e un autovalore semplice $\lambda_2 = 3$;

- c) la matrice triangolare bassa

$$\mathbf{U} = \begin{bmatrix} 2 & 0 & 0 \\ 2 & 2 & 0 \\ 1 & 3 & 2 \end{bmatrix}$$

ha un unico autovalore $\lambda = 2$ di molteplicità algebrica 3. In particolare, ha lo stesso polinomio caratteristico della matrice \mathbf{I} .



La matrice trasposta \mathbf{A}^T ha lo stesso polinomio caratteristico di \mathbf{A} . Questo segue immediatamente dal fatto che il determinante di una matrice coincide con il determinante della trasposta e dall'uguaglianza:

$$\mathbf{A}^T - \lambda \mathbf{I} = (\mathbf{A} - \lambda \mathbf{I})^T$$

Quindi, \mathbf{A} e \mathbf{A}^T hanno gli stessi autovalori, con le stesse molteplicità algebriche.

Molteplicità geometrica di un autovalore

Abbiamo visto che una matrice è diagonalizzabile sul campo \mathbb{K} se e solo se esiste una base di \mathbb{K}^n formata da autovettori di \mathbf{A} , quindi se possiamo trovare n autovettori di \mathbf{A} che siano linearmente indipendenti. Cominciamo a chiederci quale sia *il massimo numero di autovettori linearmente indipendenti relativi a uno stesso autovalore fissato*. Poi faremo variare l'autovalore. Fissato un autovalore λ di una matrice \mathbf{A} , gli autovettori di \mathbf{A} relativi a λ sono le soluzioni non nulle del sistema lineare omogeneo $(\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0}$: se aggiungiamo il vettore nullo, otteniamo un sottospazio di \mathbf{V} :

DEFINIZIONE 4.4 Sia \mathbf{A} una matrice quadrata di ordine n con elementi nel campo \mathbb{K} . Fissato un autovalore $\lambda \in \mathbb{K}$ di \mathbf{A} , l'*autospazio* di \mathbf{A} relativo a λ è il sottospazio di \mathbb{K}^n :

$$\mathbf{V}_\lambda = \text{Ker}(\mathbf{A} - \lambda \mathbf{I}) = \{\mathbf{v} \in \mathbb{K}^n : \mathbf{A}\mathbf{v} = \lambda\mathbf{v}\}$$

L'autospazio \mathbf{V}_λ consiste quindi degli autovettori di \mathbf{A} relativi all'autovalore λ , più il vettore nullo. Si tratta di un sottospazio di \mathbb{K}^n perché è il nucleo di una matrice: questo naturalmente significa che, fissato λ , la somma di autovettori relativi a λ è ancora un autovettore relativo a λ oppure è il vettore nullo: il prodotto di un autovettore relativo a λ per uno scalare non nullo t è ancora un autovettore relativo a λ .

Proprio per definizione di dimensione di uno spazio vettoriale, la dimensione di \mathbf{V}_λ è il massimo numero di autovettori linearmente indipendenti relativi a λ che si possano trovare. Tale numero si dice molteplicità geometrica di λ :

DEFINIZIONE 4.5 La *molteplicità geometrica* g_λ di un autovalore λ di \mathbf{A} è la dimensione dell'autospazio \mathbf{V}_λ :

$$g_\lambda = \dim\{\mathbf{v} \in \mathbb{K}^n : \mathbf{A}\mathbf{v} = \lambda\mathbf{v}\}$$

Poiché $\mathbf{V}_\lambda = \text{Ker}(\mathbf{A} - \lambda \mathbf{I})$, per il teorema di nullità più rango la molteplicità geometrica di λ si calcola in funzione del rango di $\mathbf{A} - \lambda \mathbf{I}$: se \mathbf{A} ha ordine n , abbiamo

$$(4.4) \quad g_\lambda = n - r(\mathbf{A} - \lambda \mathbf{I})$$

OSSERVAZIONE Se \mathbf{A} è una matrice reale e λ è un autovalore reale, è possibile considerare l'autospazio reale $\{\mathbf{v} \in \mathbb{R}^n : \mathbf{A}\mathbf{v} = \lambda\mathbf{v}\}$ oppure quello complesso $\{\mathbf{v} \in \mathbb{C}^n : \mathbf{A}\mathbf{v} = \lambda\mathbf{v}\}$. Questo non deve preoccupare il lettore. Infatti ciò che conta è la

molteplicità geometrica (cioè il massimo numero di autovettori indipendenti)

$$g_\lambda = n - r(\mathbf{A} - \lambda \mathbf{I})$$

che non dipende dalla scelta del campo degli scalari. La molteplicità geometrica è la dimensione sia dell'autospazio reale (come spazio vettoriale reale) sia dell'autospazio complesso (come spazio vettoriale complesso). Relativamente a λ , esistono g_λ autovettori *reali* che sono lincarmente indipendenti tanto in \mathbb{R} quanto in \mathbb{C} e che quindi formano una base sia dell'autospazio reale sia di quello complesso. Nelle applicazioni quello che occorre determinare è la molteplicità geometrica e, se necessario, una base dell'autospazio costituita da autovettori reali.

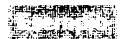


Sia \mathbf{I} la matrice identità. La molteplicità geometrica di $\lambda = 1$ come autovalore di \mathbf{I} è

$$g_1 = n - r(\mathbf{I} - \mathbf{I}) = n$$

In effetti, i vettori della base canonica sono n autovettori indipendenti di \mathbf{I} tutti relativi all'autovalore 1.

Facciamo ora tre esempi che illustrano le tre possibili risposte alla questione della diagonalizzabilità di una matrice reale: (1) la matrice è diagonalizzabile su \mathbb{R} ; (2) la matrice non è diagonalizzabile su \mathbb{R} perché non ha abbastanza autovalori reali, ma è diagonalizzabile su \mathbb{C} ; (3) la matrice non è diagonalizzabile nemmeno su \mathbb{C} , perché non ha abbastanza autovettori da formare una base di \mathbb{C}^n .



Consideriamo ancora una volta la matrice di una riflessione ortogonale (l'asse è la bisettrice $y = x$), ma questa volta fingeremo di non saperlo: ricaveremo gli autovettori esclusivamente per via algebrica. La matrice è $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, per cui

$$\mathbf{A} - \lambda \mathbf{I} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -\lambda & 1 \\ 1 & -\lambda \end{bmatrix}$$

Si noti che, in generale, la matrice $\mathbf{A} - \lambda \mathbf{I}$ si ottiene da \mathbf{A} sottraendo λ agli elementi della diagonale principale. Per la matrice in questione l'equazione caratteristica è

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \lambda^2 - 1 = 0$$

quindi \mathbf{A} ha due autovalori reali $\lambda_1 = 1$ e $\lambda_2 = -1$. Per trovare gli autovettori relativi a λ_1 , occorre risolvere il sistema omogeneo

$$(\mathbf{A} - \lambda_1 \mathbf{I}) \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

La matrice

$$\mathbf{A} - \lambda_1 \mathbf{I} = \mathbf{A} - \mathbf{I} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

ha rango 1 e quindi la molteplicità geometrica g_{λ_1} , ovvero la dimensione dell'autospazio \mathbf{V}_{λ_1} , è 1. La seconda equazione del sistema da risolvere è un multiplo della prima, e il sistema è equivalente alla singola equazione:

$$-x + y = 0$$

Una base dell'autospazio \mathbf{V}_{λ_1} è pertanto il vettore $\mathbf{b}_1 = [1, 1]^T$. Conti analoghi mostrano che l'autovalore -1 ha molteplicità geometrica 1, e una base del relativo autospazio è l'autovettore $\mathbf{b}_2 = [1, -1]^T$. I due autovettori \mathbf{b}_1 e \mathbf{b}_2 sono linearmente indipendenti e quindi formano una base di \mathbb{R}^2 . La matrice \mathbf{A} è diagonalizzabile perché esiste una base di \mathbb{R}^2 formata da autovettori di \mathbf{A} . Più precisamente, posto $\mathbf{S} = [\mathbf{b}_1 \quad \mathbf{b}_2]$, per la proposizione 3.2:

$$\mathbf{S}^{-1} \mathbf{A} \mathbf{S} = \text{diag}(\lambda_1, \lambda_2) = \text{diag}(1, -1) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$



Diamo ora un esempio di diagonalizzazione su \mathbb{C} di una matrice che non è diagonalizzabile su \mathbb{R} . Consideriamo la matrice $\mathbf{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Tale matrice rappresenta rispetto alla base canonica la rotazione di un angolo retto in senso antiorario. Siccome la rotazione non lascia invariata alcuna direzione, la matrice non può avere autovettori e autovalori reali. Verifichiamolo algebricamente. La matrice $\mathbf{A} - \lambda \mathbf{I}$ è:

$$\begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix}$$

che ha determinante $\lambda^2 + 1$. Pertanto l'equazione caratteristica $\lambda^2 + 1 = 0$ non ha radici reali: la matrice non ha autovalori reali, e quindi non è diagonalizzabile su \mathbb{R} . Non vi è però alcun problema a trovare autovalori e autovettori complessi della matrice. Gli autovalori complessi sono le radici $\lambda_1 = i$ e $\lambda_2 = -i$ dell'equazione caratteristica. Si noti che $i = e^{i\pi/2}$: l'autovalore complesso ricorda l'angolo di rotazione. Per trovare gli autovettori relativi a $\lambda_1 = i$, occorre risolvere il sistema omogeneo

$$(\mathbf{A} - i\mathbf{I}) \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

La matrice

$$\mathbf{A} - i\mathbf{I} = \begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix}$$

ha rango 1: la seconda riga si ottiene moltiplicando la prima per i . Il sistema da risolvere è quindi equivalente alla singola equazione:

$$-ix - y = 0$$

Una base per l'autospazio relativo all'autovalore $\lambda_1 = i$ è formata dall'autovettore $\mathbf{b}_1 = [1, -i]^T$. In maniera analoga si mostra che gli autovettori di \mathbf{A} relativi all'autovalore $-i$ sono i multipli non nulli dell'autovettore $\mathbf{b}_2 = [1, i]^T$. Possiamo verificare quest'affermazione mostrando che effettivamente $\mathbf{A}\mathbf{b}_2 = -i\mathbf{b}_2$:

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = \begin{bmatrix} -i \\ 1 \end{bmatrix} = -i \begin{bmatrix} 1 \\ i \end{bmatrix}$$

Infine i due autovettori \mathbf{b}_1 e \mathbf{b}_2 sono linearmente indipendenti, e quindi formano una base di \mathbb{C}^2 . Posto:

$$\mathbf{S} = \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix}$$

si ha perciò

$$\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$$

In conclusione abbiamo trovato un esempio di una matrice reale che è diagonalizzabile su \mathbb{C} , ma non su \mathbb{R} .

Sia $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. Il polinomio caratteristico di \mathbf{A} è λ^2 , per cui \mathbf{A} ha un unico autovalore $\lambda = 0$, con molteplicità algebrica 2. La molteplicità geometrica di $\lambda = 0$ è:

$$g_0 = 2 - r(\mathbf{A}) = 1$$

In questo caso esiste un unico autospazio \mathbf{V}_0 , che ha dimensione $g_0 = 1$. Quindi il massimo numero di autovettori indipendenti della matrice \mathbf{A} è 1 e la matrice non è diagonalizzabile, nemmeno su \mathbb{C} : se fosse diagonalizzabile, vi sarebbe una base di \mathbb{C}^2 formata da autovettori di \mathbf{A} e \mathbf{A} avrebbe perciò due autovettori linearmente indipendenti.

Diamo un esempio che aiuti a capire l'utilità della diagonalizzazione di una matrice: il fatto che la matrice sia diagonalizzabile consente di calcolare tutte le potenze della matrice. Il calcolo delle potenze di una matrice ha molte applicazioni (per esempio processi di Markov, sistemi dinamici discreti ecc.) come spiegato nell'introduzione al capitolo sull'algebra delle matrici, da cui riprendiamo l'esempio della dinamica della popolazione delle classi di Caronte e Mefistofele. Lo stato iniziale è descritto dal vettore $\mathbf{x}_0 = [c_0, m_0]^T$ dove c_0 è la percentuale degli studenti che hanno come docente Caronte, m_0 la percentuale degli studenti di Mefistofele, per cui $c_0 + m_0 = 1$. Dopo un intervallo di tempo il 40% degli studenti di Caronte rimane nella sua aula, il resto si trasferisce nell'aula di Mefistofele; mentre il 50% degli studenti di Mefistofele rimane e il 50% cambia aula. La legge di evoluzione del sistema è quindi

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} \quad \mathbf{A} = \begin{bmatrix} 2/5 & 1/2 \\ 3/5 & 1/2 \end{bmatrix}$$

Il polinomio caratteristico di \mathbf{A} è

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \lambda^2 - \frac{9}{10}\lambda - \frac{1}{10} = (\lambda - 1)\left(\lambda + \frac{1}{10}\right)$$

per cui gli autovalori di \mathbf{A} sono $\lambda_1 = 1$ e $\lambda_2 = -\frac{1}{10}$. Gli autovettori relativi a $\lambda_1 = 1$ sono i multipli scalari non nulli del vettore $[5, 6]^T$. Tra questi determiniamo quello che ha somma delle componenti uguali a 1 e quindi descrive un possibile stato del sistema (percentuale studenti di Caronte e percentuale studenti di Mefistofele):

$$\mathbf{e} = \frac{1}{11}[5, 6]^T = \left[\frac{5}{11}, \frac{6}{11} \right]^T = [c_e, m_e]^T$$

Osserviamo che \mathbf{e} è lo stato di equilibrio del sistema: poiché \mathbf{e} è un autovettore relativo all'autovalore 1, il sistema una volta che si trova nello stato \mathbf{e} ci rimane perché:

$$\mathbf{A}\mathbf{e} = \mathbf{e}$$

Un autovettore relativo a $\lambda_2 = -\frac{1}{10}$ è $[-1, 1]^T$, per cui la matrice $\mathbf{S} = \begin{bmatrix} 5 & -1 \\ 6 & 1 \end{bmatrix}$ diagonalizza \mathbf{A} :

$$\mathbf{S}^{-1}\mathbf{AS} = \begin{bmatrix} 1 & 0 \\ 0 & -\frac{1}{10} \end{bmatrix}$$

Moltiplicando a sinistra per \mathbf{S} e a destra per \mathbf{S}^{-1} otteniamo per \mathbf{A} la decomposizione

$$\mathbf{A} = \mathbf{SDS}^{-1}$$

dove $\mathbf{D} = \text{diag}(1, -\frac{1}{10})$. Sfruttiamo questa decomposizione per calcolare tutte le potenze \mathbf{A}^n di \mathbf{A} . Notiamo che le potenze della matrice diagonale si calcolano immediatamente:

$$\mathbf{D}^n = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}^n = \begin{bmatrix} 1 & 0 \\ 0 & \frac{(-1)^n}{10^n} \end{bmatrix}$$

Osserviamo poi che $\mathbf{A}^n = \mathbf{SD}^n\mathbf{S}^{-1}$ poiché

$$\mathbf{A}^n = (\mathbf{SDS}^{-1})^n = \mathbf{SD}(\mathbf{S}^{-1}\mathbf{S})\mathbf{D}(\mathbf{S}^{-1}\mathbf{S})\mathbf{D} \cdots \mathbf{D}(\mathbf{S}^{-1}\mathbf{S})\mathbf{D}(\mathbf{S}^{-1}\mathbf{S})\mathbf{D}\mathbf{S}^{-1} = \mathbf{SD}^n\mathbf{S}^{-1}$$

Quindi

$$\mathbf{A}^n = \begin{bmatrix} 5 & -1 \\ 6 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (-1)^n/10^n \end{bmatrix} \begin{bmatrix} 1/11 & 1/11 \\ -6/11 & 5/11 \end{bmatrix}$$

e

$$\lim_{n \rightarrow +\infty} \mathbf{A}^n = \begin{bmatrix} 5 & -1 \\ 6 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/11 & 1/11 \\ -6/11 & 5/11 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 6 & 0 \end{bmatrix} \begin{bmatrix} 1/11 & 1/11 \\ -6/11 & 5/11 \end{bmatrix} = \begin{bmatrix} 5/11 & 5/11 \\ 6/11 & 6/11 \end{bmatrix}$$

Possiamo ora dimostrare che, qualunque sia lo stato iniziale $\mathbf{x}_0 = [c_0, m_0]^T$ con $c_0 + m_0 = 1$, il sistema tende ad assumere lo stato di equilibrio:

$$\lim_{n \rightarrow +\infty} \mathbf{A}^n \mathbf{x}_0 = \begin{bmatrix} 5/11 & 5/11 \\ 6/11 & 6/11 \end{bmatrix} \begin{bmatrix} c_0 \\ m_0 \end{bmatrix} = \begin{bmatrix} 5(c_0 + m_0)/11 \\ 6(c_0 + m_0)/11 \end{bmatrix} = \begin{bmatrix} 5/11 \\ 6/11 \end{bmatrix} = \mathbf{e}$$

Numero degli autovettori linearmente indipendenti

Una matrice quadrata di ordine n è diagonalizzabile su \mathbb{K} se possiamo trovare n autovettori della matrice linearmente indipendenti. Occorre quindi determinare il massimo numero di autovettori linearmente indipendenti della matrice. Sappiamo già che, fissato un autovalore λ_0 , il massimo numero di autovettori linearmente indipendenti relativi a λ_0 è uguale alla molteplicità geometrica dell'autovalore. Faremo ora vedere che il massimo numero di autovettori indipendenti è la somma delle molteplicità geometriche degli autovalori: questo significa che facendo l'unione delle basi di ciascun autospazio della matrice si ottiene ancora un insieme linearmente indipendente. Il principio fondamentale è:

PROPOSIZIONE 4.6 (Autovettori relativi ad autovalori distinti sono linearmente indipendenti) Siano $\mathbf{v}_1, \dots, \mathbf{v}_s$ autovettori di \mathbf{A} relativi agli autovalori $\lambda_1, \dots, \lambda_s$. Se gli autovalori $\lambda_1, \dots, \lambda_s$ sono *distinti*, allora $\mathbf{v}_1, \dots, \mathbf{v}_s$ sono *linearmente indipendenti*.

DIMOSTRAZIONE. Mostriamo per semplicità dapprima il caso $s = 2$: supponiamo che \mathbf{v} e \mathbf{w} siano autovettori di una matrice \mathbf{A} e che i corrispondenti autovalori λ e μ siano distinti. Le ipotesi sono quindi: $\mathbf{Av} = \lambda\mathbf{v}$ con $\mathbf{v} \neq \mathbf{0}$, $\mathbf{Aw} = \mu\mathbf{w}$ con $\mathbf{w} \neq \mathbf{0}$, e $\lambda \neq \mu$. La tesi da dimostrare è che \mathbf{v} e \mathbf{w} sono linearmente indipendenti. Supponiamo che ci sia una relazione lineare tra \mathbf{v} e \mathbf{w} :

$$c_1\mathbf{v} + c_2\mathbf{w} = \mathbf{0}$$

Moltiplichiamo entrambi i membri dell'uguaglianza a sinistra per \mathbf{A} troviamo una seconda relazione:

$$c_1\lambda\mathbf{v} + c_2\mu\mathbf{w} = \mathbf{0}$$

Ora moltiplichiamo la prima relazione per μ e sottraiamo il risultato dalla seconda. Ottieniamo:

$$c_1(\lambda - \mu)\mathbf{v} = \mathbf{0}$$

Siccome $\mathbf{v} \neq \mathbf{0}$ per ipotesi, dev'essere nullo lo scalare $c_1(\lambda - \mu)$. Ma per ipotesi $\lambda \neq \mu$, quindi $c_1 = 0$. Sostituendo nella prima relazione troviamo $c_2\mathbf{w} = \mathbf{0}$, e quindi anche $c_2 = 0$ perché $\mathbf{w} \neq \mathbf{0}$. Abbiamo così mostrato che $c_1 = c_2 = 0$, il che significa che \mathbf{v} e \mathbf{w} sono linearmente indipendenti.

Generalizziamo ora il ragionamento a un numero qualsiasi di autovettori: supponiamo che $\mathbf{v}_1, \dots, \mathbf{v}_s$ siano autovettori di \mathbf{A} relativi ad autovalori distinti $\lambda_1, \dots, \lambda_s$. Vogliamo mostrare che $\mathbf{v}_1, \dots, \mathbf{v}_s$ sono linearmente indipendenti. Procediamo per induzione su s . Il caso $s = 1$ è ovvio perché \mathbf{v}_1 è per definizione di autovettore non nullo, e quindi è linearmente indipendente. Possiamo allora assumere che $s \geq 2$ e $\mathbf{v}_1, \dots, \mathbf{v}_{s-1}$ siano linearmente indipendenti.

Supponiamo che ci sia una relazione lineare tra i vettori $\mathbf{v}_1, \dots, \mathbf{v}_s$:

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_s\mathbf{v}_s = \mathbf{0}$$

Moltiplicando entrambi i membri dell'uguaglianza a sinistra per \mathbf{A} troviamo una seconda relazione:

$$c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \cdots + c_s\lambda_s\mathbf{v}_s = \mathbf{0}$$

Ora moltiplichiamo la prima relazione per λ_s e sottraiamo il risultato dalla seconda. Ottieniamo:

$$c_1(\lambda_1 - \lambda_s)\mathbf{v}_1 + c_2(\lambda_2 - \lambda_s)\mathbf{v}_2 + \cdots + c_{s-1}(\lambda_{s-1} - \lambda_s)\mathbf{v}_{s-1} = \mathbf{0}$$

Siccome $\mathbf{v}_1, \dots, \mathbf{v}_{s-1}$ sono linearmente indipendenti per ipotesi di induzione, abbiamo

$$c_1(\lambda_1 - \lambda_s) = c_2(\lambda_2 - \lambda_s) = \cdots = c_{s-1}(\lambda_{s-1} - \lambda_s) = 0$$

Ma per ipotesi gli autovalori $\lambda_1, \dots, \lambda_s$ sono distinti, quindi devono essere nulli i coefficienti c_1, \dots, c_{s-1} . Sostituendo nella prima relazione troviamo $c_s\mathbf{v}_s = \mathbf{0}$ e quindi anche $c_s = 0$ perché $\mathbf{v}_s \neq \mathbf{0}$.

Abbiamo così mostrato che $c_1 = c_2 = \cdots = c_s = 0$, il che significa che $\mathbf{v}_1, \dots, \mathbf{v}_s$ sono linearmente indipendenti. ■

Come conseguenza importante abbiamo:

TEOREMA 4.7 (Condizione sufficiente di diagonalizzabilità)

Sia \mathbf{A} una matrice quadrata di ordine n a elementi nel campo \mathbb{K} . Se \mathbf{A} ha n autovalori distinti in \mathbb{K} , allora \mathbf{A} è diagonalizzabile su \mathbb{K} .

DIMOSTRAZIONE. Supponiamo che \mathbf{A} abbia n autovalori distinti. Siano $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{K}^n$ autovettori di \mathbf{A} relativi a $\lambda_1, \dots, \lambda_n$, rispettivamente. Abbiamo mostrato che $\mathbf{v}_1, \dots, \mathbf{v}_n$ sono linearmente indipendenti, quindi formano una base di \mathbb{K}^n . La matrice è diagonalizzabile perché \mathbb{K}^n ha una base formata da autovettori della matrice.

OSSERVAZIONE Se $\mathbb{K} = \mathbb{C}$, una matrice quadrata \mathbf{A} di ordine n ha esattamente n autovalori, se questi sono contati con le loro molteplicità algebriche. L'ipotesi che \mathbf{A} abbia n autovalori distinti equivale quindi all'ipotesi che tutti gli autovalori abbiano molteplicità algebrica 1. Se $\mathbb{K} = \mathbb{R}$, l'ipotesi è che \mathbf{A} abbia tutti gli autovalori reali e di molteplicità algebrica 1.



La matrice $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ ha polinomio caratteristico $\lambda^2 - 1$, quindi due autovalori reali distinti $\lambda_1 = 1$ e $\lambda_2 = -1$. Per la proposizione precedente \mathbf{A} è diagonalizzabile sul campo reale: esiste una matrice invertibile reale \mathbf{S} tale che

$$\mathbf{S}^{-1}\mathbf{AS} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

La matrice $\mathbf{B} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ ha polinomio caratteristico $\lambda^2 + 1$, quindi non ha autovalori reali e non è diagonalizzabile sul campo reale. Però ha due autovalori complessi distinti $\lambda_1 = i$ e $\lambda_2 = -i$. Per la proposizione precedente \mathbf{B} è diagonalizzabile sul campo complesso: esiste una matrice invertibile complessa \mathbf{S} tale che:

$$\mathbf{S}^{-1}\mathbf{BS} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$$

Avevamo già ottenuto questi risultati calcolando \mathbf{S} esplicitamente. La proposizione ci consente di dire che la base di autovettori esiste senza bisogno di calcolare gli autovettori.



Consideriamo la matrice

$$\mathbf{A} = \begin{bmatrix} 5 & 1 & 1 \\ 0 & -3 & 0 \\ 1 & 0 & 5 \end{bmatrix}$$

Calcoliamo il polinomio caratteristico di \mathbf{A} sviluppando il determinante secondo Laplace rispetto alla prima colonna:

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \det \begin{bmatrix} 5 - \lambda & 0 & 1 \\ 0 & -3 - \lambda & 0 \\ 1 & 0 & 5 - \lambda \end{bmatrix} = (5 - \lambda)(-3 - \lambda)(5 - \lambda) + (3 + \lambda)$$

Raccogliamo il termine $3 + \lambda$:

$$\det(\mathbf{A} - \lambda\mathbf{I}) = (3 + \lambda)(1 - (5 - \lambda)^2) = (3 + \lambda)(1 - 5 + \lambda)(1 + 5 - \lambda) = (3 + \lambda)(\lambda - 4)(6 - \lambda)$$

La matrice \mathbf{A} ha perciò i tre autovalori distinti $\lambda_1 = -3$, $\lambda_2 = 4$, $\lambda_3 = 6$ ed è diagonalizzabile. Controlliamo che la traccia della matrice sia uguale alla somma degli autovalori:

$$\text{tr}(\mathbf{A}) = 5 - 3 + 5 = 7, \quad \lambda_1 + \lambda_2 + \lambda_3 = -3 + 4 + 6 = 7$$

OSSERVAZIONE Si può mostrare che una matrice *generica* ha autovalori distinti e quindi sul campo complesso è diagonalizzabile: intuitivamente, se per caso il polinomio caratteristico di una matrice ha radici multiple, cambiando di poco i coefficienti si ottiene quasi certamente una matrice il cui polinomio caratteristico non ha radici multiple. Quindi una matrice *generica* è diagonalizzabile sul campo complesso; per questo il teorema precedente è importante e di vasta applicabilità.

Per chiarire questo punto, consideriamo un'arbitraria matrice 2×2 :

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Il polinomio caratteristico di \mathbf{A} è

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \lambda^2 - (a+d)\lambda + ad - bc$$

La matrice ha autovalori distinti tranne nel caso

$$\Delta = (a+d)^2 - 4(ad - bc) = 0$$

Una matrice generica ha coefficienti che non soddisfano questa equazione, e quindi ha autovalori distinti.

Passiamo ora a occuparci del problema di determinare il massimo numero di autovettori linearmente indipendenti di una matrice data, senza escludere il caso di autovalori ripetuti. Abbiamo visto che, fissato un autovalore λ , il massimo numero di autovettori linearmente indipendenti relativi a λ è la molteplicità geometrica

$$g_\lambda = n - r(\mathbf{A} - \lambda \mathbf{I})$$

dell'autovalore.

LEMMA 4.8 Siano $\lambda_1, \dots, \lambda_s$ gli autovalori distinti di \mathbf{A} in \mathbb{K} , e siano g_1, \dots, g_s le relative molteplicità geometriche. Il massimo numero N di autovettori linearmente indipendenti di \mathbf{A} in \mathbb{K}^n è la somma delle molteplicità geometriche degli autovalori:

$$N = g_1 + \dots + g_s$$

DIMOSTRAZIONE. Sia \mathbf{V}_k l'autospazio relativo a λ_k . Per definizione di molteplicità geometrica \mathbf{V}_k ha dimensione g_k . In particolare, in ogni insieme \mathcal{S} di autovettori indipendenti ci sono al massimo g_k autovettori indipendenti relativi all'autovalore λ_k , e quindi \mathcal{S} ha al massimo $g_1 + \dots + g_s$ elementi. Questo mostra che $N \leq g_1 + \dots + g_s$. Per concludere, occorre ora costruire un insieme linearmente indipendente \mathcal{S} formato da $g_1 + \dots + g_s$ autovettori. Per costruire tale insieme, per ogni k fissiamo una base $\mathcal{B}_k = \{\mathbf{w}_1^{(k)}, \dots, \mathbf{w}_{g_k}^{(k)}\}$ di \mathbf{V}_k , e consideriamo l'insieme \mathcal{S} costituito da tutti questi vettori al variare di k :

$$\mathcal{S} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \dots \cup \mathcal{B}_s = \left\{ \mathbf{w}_i^{(k)} : 1 \leq k \leq s, 1 \leq i \leq g_k \right\}$$

Per costruzione l'insieme \mathcal{S} è formato da $g_1 + \dots + g_s$ autovettori. Rimane da mostrare che questi vettori sono linearmente indipendenti. Supponiamo che ci sia una relazione lineare tra i vettori di \mathcal{S} :

$$\sum_{k=1}^s \sum_{i=1}^{g_k} c_{ik} \mathbf{w}_i^{(k)} = \mathbf{0}$$

Occorre mostrare che tutti i coefficienti c_{ik} sono nulli. Poniamo $\mathbf{v}_k = \sum_{i=1}^{g_k} c_{ik} \mathbf{w}_i^{(k)}$ per ogni $k = 1, \dots, s$, in modo tale che la precedente uguaglianza si riscriva nella forma

$$\sum_{k=1}^s \mathbf{v}_k = \mathbf{0}$$

Il vettore \mathbf{v}_k è combinazione lineare dei vettori $\mathbf{w}_1^{(k)}, \dots, \mathbf{w}_{g_k}^{(k)}$ che formano una base dell'auto spazio \mathbf{V}_k , quindi appartiene a \mathbf{V}_k . Questo significa che \mathbf{v}_k è un autovettore relativo a λ_k , oppure è il vettore nullo. La relazione lineare, con coefficienti tutti uguali a uno, $\sum_{k=1}^s \mathbf{v}_k = \mathbf{0}$ mostra che gli eventuali \mathbf{v}_k non nulli sono linearmente *dipendenti*. Ma questo è impossibile perché autovettori relativi ad autovalori distinti sono indipendenti. Concludiamo che tutti i \mathbf{v}_k sono nulli:

$$\mathbf{v}_k = \sum_{i=1}^{g_k} c_{ik} \mathbf{w}_i^{(k)} = \mathbf{0} \quad \text{per ogni } k = 1, 2, \dots, s$$

Siccome $\mathbf{w}_1^{(k)}, \dots, \mathbf{w}_{g_k}^{(k)}$ sono linearmente indipendenti, i coefficienti c_{ik} sono nulli per ogni $i = 1, 2, \dots, g_k$. Questo vale per ogni k , quindi tutti i coefficienti c_{ik} sono nulli, come volevasi dimostrare. \square

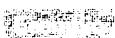
Dal lemma segue immediatamente che una matrice di ordine n è diagonalizzabile, cioè esiste una base di \mathbb{K}^n formata da autovettori della matrice, se e solo se la somma delle molteplicità geometriche degli autovalori è uguale a n . Si può rendere più efficace questo criterio di diagonalizzabilità mettendo a confronto la molteplicità geometrica di un autovalore con la sua molteplicità algebrica. Il risultato di base è:

PROPOSIZIONE 4.9 (Confronto molteplicità geometrica e algebrica)

Sia λ un autovalore di una matrice \mathbf{A} e siano g_λ e a_λ , rispettivamente, la molteplicità geometrica e la molteplicità algebrica di λ . Allora

$$1 \leq g_\lambda \leq a_\lambda$$

DIMOSTRAZIONE. Se λ è un autovalore, allora $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$, per cui $\mathbf{A} - \lambda \mathbf{I}$ non ha rango n , cioè $g_\lambda \geq 1$. Rimandiamo la dimostrazione dell'altra diseguaglianza al paragrafo sulla similitudine di matrici. \square



Sia $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. Il polinomio caratteristico di \mathbf{A} è λ^2 , per cui \mathbf{A} ha un unico autovalore $\lambda = 0$, con molteplicità algebrica 2. La molteplicità geometrica di $\lambda = 0$ è:

$$g_0 = 2 - r(\mathbf{A}) = 1$$

L'esempio mostra che è possibile che un autovalore abbia molteplicità geometrica strettamente minore di quella algebrica. Gli autovalori per cui questo non succede si dicono *regolari*. Gli autovalori con molteplicità algebrica 1 si dicono *semplici* perché

sono radici semplici del polinomio caratteristico. Per la proposizione 4.9 un autovalore semplice ha molteplicità geometrica 1 ed è quindi regolare. Riassumendo:

DEFINIZIONE 4.10 (Autovalori semplici. Autovalori regolari)

- Un autovalore si dice *regolare* se le sue molteplicità algebrica e geometrica coincidono.
- Un autovalore si dice *semplice* se ha molteplicità algebrica 1.

Per la proposizione 4.9 un autovalore semplice è regolare.

TEOREMA 4.11 (Secondo criterio di diagonalizzabilità)

Una matrice quadrata \mathbf{A} è diagonalizzabile su \mathbb{K} se e solo se sono verificate le seguenti condizioni:

- a) il polinomio caratteristico di \mathbf{A} ha tutte le sue radici in \mathbb{K} (ipotesi automaticamente soddisfatta se $\mathbb{K} = \mathbb{C}$);
- b) ogni autovalore λ di \mathbf{A} è regolare.

DIMOSTRAZIONE. Siano $\lambda_1, \dots, \lambda_s$ gli autovalori distinti di \mathbf{A} nel campo \mathbb{K} e siano g_1, \dots, g_s le relative molteplicità geometriche. Per il primo criterio di diagonalizzabilità \mathbf{A} è diagonalizzabile su \mathbb{K} se esiste in \mathbb{K}^n un insieme di n autovettori linearmente indipendenti di \mathbf{A} ; per il lemma 4.8 il massimo numero di autovettori linearmente indipendenti di \mathbf{A} in \mathbb{K}^n è esattamente la somma delle molteplicità geometriche $g_1 + \dots + g_s$. Quindi \mathbf{A} è diagonalizzabile se e solo se

$$n = g_1 + \dots + g_s$$

Siano ora a_1, \dots, a_s le molteplicità algebriche degli autovalori $\lambda_1, \dots, \lambda_s$. Allora il polinomio caratteristico di \mathbf{A} è divisibile per

$$(\lambda - \lambda_1)^{a_1} (\lambda - \lambda_2)^{a_2} \cdots (\lambda - \lambda_s)^{a_s}$$

Quindi la somma delle molteplicità algebriche è minore o uguale a n che è il grado del polinomio caratteristico. Ma $g_k \leq a_k$ per ogni k , per cui

$$g_1 + \dots + g_s \leq a_1 + \dots + a_s \leq n$$

Pertanto \mathbf{A} è diagonalizzabile se e solo se:

$$n = g_1 + \dots + g_s \leq a_1 + \dots + a_s \leq n$$

Questo è possibile se e solo se:

$$g_1 + \dots + g_s = a_1 + \dots + a_s = n$$

il che accade se e solo se $a_1 + \dots + a_s = n$ e $g_k = a_k$ per ogni k . La prima condizione indica che il polinomio caratteristico ha tutte le sue radici in \mathbb{K} , la seconda che gli autovalori sono regolari. 

Per quali valori di $a, b, c, d \in \mathbb{R}$ la matrice

$$\mathbf{A} = \begin{bmatrix} 1 & a & b \\ 0 & c & d \\ 0 & 0 & 4 \end{bmatrix}$$

è diagonalizzabile? La matrice è triangolare, quindi i suoi autovalori sono $\lambda_1 = 1$, $\lambda_2 = c$ e $\lambda_3 = 4$. Se $c \neq 1$ e $c \neq 4$, gli autovalori sono distinti, e quindi la matrice è diagonalizzabile su \mathbb{R} .

Supponiamo ora $c = 1$. Allora la matrice ha un autovalore doppio $\lambda_1 = \lambda_2 = 1$ e un autovalore semplice $\lambda_3 = 4$. L'autovalore semplice è regolare, per cui per il secondo criterio di diagonalizzabilità \mathbf{A} è diagonalizzabile se e solo se la molteplicità geometrica dell'autovalore doppio è 2. Ora

$$g_1 = 3 - r(\mathbf{A} - \mathbf{I}) = r \left(\begin{bmatrix} 0 & a & b \\ 0 & 0 & d \\ 0 & 0 & 3 \end{bmatrix} \right)$$

Quindi $g_1 = 2$ se e solo se $a = 0$. Se $c = 4$, un ragionamento analogo mostra che la matrice è diagonalizzabile se e solo se $d = 0$.

In conclusione, la matrice è diagonalizzabile se e solo se $c \neq 1$ oppure $c \neq 4$ oppure $c = 1$ e $a = 0$ oppure $c = 4$ e $d = 0$.

Consideriamo la matrice

$$\mathbf{A} = \begin{bmatrix} 3 & 0 & -7 \\ k & 3 & 4 \\ 0 & 0 & 2 \end{bmatrix} \quad (k \in \mathbb{R})$$

Vogliamo stabilire per quali valori di k la matrice è diagonalizzabile (su \mathbb{R}) e per tali valori determinare una base di \mathbb{R}^3 formata da autovettori di \mathbf{A} . Il polinomio caratteristico di \mathbf{A} è $(3-\lambda)^2(2-\lambda)$, quindi la matrice ha due autovalori reali, $\lambda_1 = 3$ con molteplicità algebrica 2, e $\lambda_2 = 2$ con molteplicità algebrica 1. L'autovalore λ_2 è semplice, quindi regolare, e la matrice è diagonalizzabile se e solo se anche λ_1 è regolare. Calcoliamo la molteplicità geometrica di λ_1 :

$$g_{\lambda_1} = 3 - r(\mathbf{A} - 3\mathbf{I}) = 3 - r \left(\begin{bmatrix} 0 & 0 & -7 \\ k & 0 & 4 \\ 0 & 0 & -1 \end{bmatrix} \right) = \begin{cases} 1 & \text{se } k \neq 0 \\ 2 & \text{se } k = 0 \end{cases}$$

Concludiamo che la matrice è diagonalizzabile se e solo se $k = 0$.

Se $k = 0$, la matrice è $\mathbf{A} = \begin{bmatrix} 3 & 0 & -7 \\ 0 & 3 & 4 \\ 0 & 0 & 2 \end{bmatrix}$. Calcoliamone gli autovettori. L'autospazio relativo a $\lambda_1 = 3$ ha equazioni

$$(\mathbf{A} - 3\mathbf{I}) \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Tutte e tre le equazioni di questo sistema sono equivalenti all'equazione $z = 0$, quindi l'autospazio relativo a λ_1 è il piano $z = 0$. Come autovettori linearmente indipendenti relativi a λ_1 possiamo scegliere i primi due vettori della base canonica \mathbf{e}_1 e \mathbf{e}_2 . L'autospazio relativo a $\lambda_2 = 2$ ha equazioni

$$(\mathbf{A} - 2\mathbf{I}) \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

che sono equivalenti a

$$\begin{cases} x - 7z = 0 \\ y - 4z = 0 \end{cases}$$

Scegliendo $z = 1$ troviamo l'autovettore $\mathbf{v} = [7, 4, 1]^T$. In conclusione $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{v}\}$ è una base di \mathbb{R}^3 formata da autovettori di \mathbf{A} . Sia \mathbf{S} la matrice che ha per colonne gli autovettori $\mathbf{e}_1, \mathbf{e}_2, \mathbf{v}$:

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 7 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$$

Per la proposizione 3.2

$$\mathbf{S}^{-1} \mathbf{A} \mathbf{S} = \text{diag}(3, 3, 2)$$

Un'importante classe di matrici non diagonalizzabili è fornita dai blocchi di Jordan:

DEFINIZIONE 4.12 (Blocco di Jordan)

Il *blocco di Jordan* $\mathbf{J} = \mathbf{J}_n(a)$ di ordine n associato allo scalare a è la matrice quadrata di ordine n

$$\mathbf{J}_n(a) = \begin{bmatrix} a & 1 & 0 & \dots & 0 & 0 \\ 0 & a & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & a & 1 \\ 0 & 0 & 0 & \dots & 0 & a \end{bmatrix}$$

che ha gli elementi sulla diagonale principale uguali ad a , gli elementi sulla diagonale immediatamente sopra alla diagonale principale uguali a 1 e tutti gli altri elementi nulli.

Per esempio:

$$\mathbf{J}_3(4) = \begin{bmatrix} 4 & 1 & 0 \\ 0 & 4 & 1 \\ 0 & 0 & 4 \end{bmatrix}, \quad \mathbf{J}_4(7) = \begin{bmatrix} 7 & 1 & 0 & 0 \\ 0 & 7 & 1 & 0 \\ 0 & 0 & 7 & 1 \\ 0 & 0 & 0 & 7 \end{bmatrix}$$

Un blocco di Jordan è una matrice *bidiagonale* perché gli elementi non nulli si trovano su due sole diagonali della matrice.

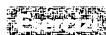
PROPOSIZIONE 4.13 (Autovalori e autovettori di un blocco di Jordan)

Il blocco di Jordan $\mathbf{J}_n(a)$ ha un unico autovalore $\lambda = a$. La molteplicità geometrica di a è 1, quella algebrica è n . L'autospazio \mathbf{V}_a è la retta generata dal primo vettore \mathbf{e}_1 della base canonica. Se $n > 1$, il blocco di Jordan $\mathbf{J}_n(a)$ non è diagonalizzabile.

DIMOSTRAZIONE. La dimostrazione è immediata: siccome \mathbf{J} è una matrice triangolare, a è l'unico autovalore e ha molteplicità algebrica pari all'ordine n di \mathbf{J} . La molteplicità geometrica è

$$g_a = n - r(\mathbf{J}_n(a) - a\mathbf{I}) = n - r(\mathbf{J}_n(0)) = n - (n - 1) = 1$$

Quindi l'autospazio \mathbf{V}_a ha dimensione 1 e una sua base è formata da un qualunque autovettore della matrice. Siccome $\mathbf{J}\mathbf{e}_1 = a \mathbf{e}_1$, il primo vettore della base canonica \mathbf{e}_1 è un autovettore di \mathbf{J} e tutti gli altri autovettori sono multipli di \mathbf{e}_1 . Se $n > 1$, l'autovalore a non è regolare e, quindi, la matrice \mathbf{J} non è diagonalizzabile.



Q Stabilire se la matrice $\mathbf{A} = \begin{bmatrix} 3 & 4 \\ 1 & 3 \end{bmatrix}$ è diagonalizzabile su \mathbb{R} . Se lo è, trovare una base di \mathbb{R}^2 formata da autovettori di \mathbf{A} .

Q Sia \mathbf{A} una matrice quadrata con polinomio caratteristico $-\lambda^3 - \lambda + 2$. Qual è l'ordine di \mathbf{A} ? Quanto valgono il determinante e la traccia di \mathbf{A} ? Mostrare che \mathbf{A} ha un autovalore reale e due autovalori complessi e verificare che la somma degli autovalori è la traccia, mentre il prodotto degli autovalori è il determinante.

Q Sia \mathbf{A} una matrice con polinomio caratteristico $\lambda^2 + 3\lambda + 2$. Qual è il polinomio caratteristico di $\mathbf{A} + 7\mathbf{I}$? In generale, che relazione c'è tra il polinomio caratteristico di una matrice \mathbf{A} e quello di $\mathbf{A} + c\mathbf{I}$ (c è uno scalare fissato)?

Q Mostrare che la matrice $\mathbf{A} = \begin{bmatrix} 3 & 0 \\ 2 & 3 \end{bmatrix}$ ha un solo autovalore e che tale autovalore non è regolare. Determinare gli autovettori di \mathbf{A} , e verificare che non esiste una base di \mathbb{K}^2 formata da autovettori di \mathbf{A} (quindi \mathbf{A} non è diagonalizzabile, né su \mathbb{R} né su \mathbb{C}).

Q Trovare una base di \mathbb{R}^2 formata da autovettori della matrice simmetrica $\mathbf{A} = \begin{bmatrix} 2 & -2 \\ 2 & -2 \end{bmatrix}$. Verificare che i vettori di tale base sono tra loro perpendicolari.

Q Trovare il polinomio caratteristico, gli autovalori e gli autovettori della matrice

$$\mathbf{A} = \begin{bmatrix} 3 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 2 & 1 \end{bmatrix}$$

Spiegare perché \mathbf{A} non è diagonalizzabile. Scrivere una matrice diagonalizzabile con lo stesso polinomio caratteristico di \mathbf{A} .

Q Mostrare che la matrice

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

ha un solo autovalore reale con molteplicità algebrica uguale a 1 (quindi è un autovalore regolare). Perché \mathbf{A} non è diagonalizzabile su \mathbb{R} ? La matrice è diagonalizzabile su \mathbb{C} ? Quali sono gli autovettori complessi di \mathbf{A} ?

Q Data la matrice

$$\mathbf{A} = \begin{bmatrix} 2 & -2 & 0 \\ -2 & 9 & 2 \\ 0 & 2 & 2 \end{bmatrix}$$

si trovino, se possibile, una matrice invertibile \mathbf{S} e una matrice diagonale \mathbf{D} tali che $\mathbf{D} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$.

Suggerimento: gli autovalori della matrice sono 1, 2 e 10.

 Per quali valori di a, b, d la matrice $\mathbf{A} = \begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$ è diagonalizzabile?

Si determinino gli autovalori e gli autovettori della matrice $\mathbf{A} = \begin{bmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{bmatrix}$ e si calcoli il limite $\mathbf{A}^\infty = \lim_{n \rightarrow +\infty} \mathbf{A}^n$.

 Si consideri la matrice

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 3 & 0 \\ k & 0 & 2 \end{bmatrix}$$

Per quali valori di k la matrice \mathbf{A} è diagonalizzabile? Per tali valori si determini una base di \mathbb{R}^3 costituita da autovettori di \mathbf{A} .

 Sia \mathbf{A} una matrice diagonalizzabile. Mostrare che $\text{Ker}(\mathbf{A}) \cap \text{Col}(\mathbf{A}) = \{\mathbf{0}\}$ e $\text{Col}(\mathbf{A}) + \text{Ker}(\mathbf{A}) = \mathbb{K}^n$. Mostrare anche che lo spazio colonna è generato dagli autovettori di \mathbf{A} relativi agli autovalori non nulli di \mathbf{A} e che il rango di \mathbf{A} coincide con il numero di autovalori non nulli di \mathbf{A} contati con la loro molteplicità.

Suggerimento: considerare prima il caso in cui \mathbf{A} è diagonale.

 Si consideri la matrice

$$\mathbf{A} = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 9 & 6 \\ 0 & 6 & 4 \end{bmatrix}$$

Si determinino delle basi di $\text{Ker}(\mathbf{A})$ e $\text{Col}(\mathbf{A})$ formate da autovettori di \mathbf{A} .

 Sia \mathbf{A} una matrice diagonalizzabile. Mostrare che $\text{Ker}(\mathbf{A}) = \text{Ker}(\mathbf{A}^2)$, $\text{Col}(\mathbf{A}) = \text{Col}(\mathbf{A}^2)$.

Suggerimento: fare prima il caso in cui \mathbf{A} è diagonale; osservare che, se $\mathbf{A} = \mathbf{SDS}^{-1}$, allora $\mathbf{A}^2 = \mathbf{SD}^2\mathbf{S}^{-1}$.

 Sia \mathbf{A} una matrice quadrata di ordine n e si supponga che esista un vettore $\mathbf{v} \in \mathbb{R}^n$ tale che $\mathbf{Av} \neq \mathbf{0}$ e $\mathbf{A}^2\mathbf{v} = \mathbf{0}$ (questo significa che il nucleo di \mathbf{A}^2 contiene propriamente il nucleo di \mathbf{A}). Mostrare che \mathbf{A} non è diagonalizzabile.

 Per quali valori del parametro reale a la matrice

$$\mathbf{A} = \begin{bmatrix} 10 & 2a & a \\ 0 & 9 & 2 \\ 0 & 2 & 6 \end{bmatrix}$$

è diagonalizzabile? Per tali valori si determini, se possibile, una base di \mathbb{R}^3 formata da autovettori di \mathbf{A} di lunghezza 1 e a due a due perpendicolari.

 Per quali valori del parametro reale k la matrice

$$\mathbf{A} = \begin{bmatrix} 3-k & -k & 1 \\ -1+k & 2+k & -1 \\ 1+k & k & 3 \end{bmatrix}$$

è diagonalizzabile? Per tali valori si determini una base di \mathbb{R}^3 formata da autovettori di \mathbf{A} .

Sia \mathbf{A} una matrice quadrata. Si supponga che la somma dei coefficienti di ciascuna colonna sia uguale a 18. Mostrare che 18 è un autovalore di \mathbf{A} .

Suggerimento: per ipotesi $[1, \dots, 1]^T$ è un autovettore di \mathbf{A}^T e \mathbf{A}^T ha lo stesso polinomio caratteristico di \mathbf{A} .

Si consideri il sistema di equazioni differenziali

$$\begin{cases} x'(t) = -x(t) + 3y(t) \\ y'(t) = x(t) + y(t) \end{cases}$$

Trovare due soluzioni linearmente indipendenti della forma $\mathbf{u}(t) = e^{\lambda t} \mathbf{w}$.

Data la matrice

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 2 \\ 2 & 1 & -1 \end{bmatrix}$$

si calcoli il prodotto $\mathbf{B} = \mathbf{A}\mathbf{A}^T$ e si trovino una matrice invertibile \mathbf{S} e una matrice diagonale \mathbf{D} tali che $\mathbf{S}^{-1}\mathbf{B}\mathbf{S} = \mathbf{D}$. Si ripeta l'esercizio con $\mathbf{B} = (\mathbf{A}^T)\mathbf{A}$.

Sia λ un autovalore di \mathbf{A} . Nel testo si mostra che λ è anche un autovalore di \mathbf{A}^T , con la stessa molteplicità algebrica. Mostrare che anche la molteplicità geometrica di λ , come autovalore di \mathbf{A}^T , coincide con la molteplicità geometrica di λ come autovalore di \mathbf{A} .

Si consideri la matrice

$$\mathbf{A} = \frac{1}{2} \begin{bmatrix} 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \\ -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \end{bmatrix}$$

- a) posto $\mathbf{V}_1 = \text{Ker}(\mathbf{A} - \mathbf{I})$ e $\mathbf{V}_{-1} = \text{Ker}(\mathbf{A} + \mathbf{I})$, determinare una base di \mathbf{V}_1 e una base di \mathbf{V}_{-1} , e verificare che $\dim \mathbf{V}_1 + \dim \mathbf{V}_{-1} = 4$;
- b) qual è il polinomio caratteristico di \mathbf{A} ?

Suggerimento: gli autovalori e le loro molteplicità sono determinati dal punto a).

Sia $\mathfrak{L} : \mathbb{M}(2, 2) \rightarrow \mathbb{M}(2, 2)$ l'applicazione lineare che a una matrice \mathbf{A} associa la sua trasposta. Determinare autovalori e autovettori di \mathfrak{L} .

Suggerimento: (1) cos'è $\mathfrak{L}(\mathbf{A})$ per una matrice simmetrica (rispettivamente antisimmetrica)? (2) in alternativa, scrivere la matrice rappresentativa di \mathfrak{L} , che è una matrice 4×4 e calcolarne gli autovalori e autovettori (naturalmente, il secondo metodo è sconsigliato).

Siano $\mathbf{v}_1, \dots, \mathbf{v}_d$ vettori di \mathbb{R}^n . Possiamo considerare i vettori \mathbf{v}_k come elementi di \mathbb{C}^n , perché un numero reale è anche un numero complesso. Mostrare che $\mathbf{v}_1, \dots, \mathbf{v}_d$ sono linearmente indipendenti in \mathbb{R}^n se e solo se sono linearmente indipendenti in \mathbb{C}^n .

Suggerimento: scrivere una combinazione lineare a coefficienti complessi come somma di una parte reale e di una parte immaginaria.

5 MATRICI SIMILI

In questo paragrafo tutte le matrici sono quadrate di ordine n a coefficienti nel campo \mathbb{K} .

DEFINIZIONE 5.1 (Matrici simili)

Una matrice \mathbf{B} si dice simile a una matrice \mathbf{A} se esiste una matrice invertibile \mathbf{S} tale che

$$\mathbf{B} = \mathbf{S}^{-1}\mathbf{AS}$$

La sostanza della definizione è che \mathbf{B} è simile ad \mathbf{A} se le due matrici rappresentano la stessa applicazione lineare; infatti, se $\mathbf{B} = \mathbf{S}^{-1}\mathbf{AS}$, la matrice \mathbf{B} rispetta l'applicazione $\mathcal{L}_\mathbf{A}$ rispetto alla base di \mathbb{K}^n formata dalle colonne di \mathbf{S} , e \mathbf{A} rappresenta la stessa applicazione rispetto alla base canonica. La similitudine tra matrici è una *relazione di equivalenza*:

PROPOSIZIONE 5.2 La similitudine è una relazione di equivalenza sull'insieme delle matrici quadrate di ordine n . Questo significa che valgono le seguenti proprietà:

- a) proprietà riflessiva: una matrice \mathbf{A} è simile a se stessa;
- b) proprietà simmetrica: se \mathbf{B} è simile \mathbf{A} , allora \mathbf{A} è simile a \mathbf{B} . Ha senso quindi dire che \mathbf{A} e \mathbf{B} sono simili;
- c) proprietà transitiva: se \mathbf{B} è simile ad \mathbf{A} e \mathbf{C} è simile a \mathbf{B} , allora \mathbf{C} è simile ad \mathbf{A} .

DIMOSTRAZIONE.

- a) Una matrice \mathbf{A} è simile a se stessa perché la matrice identità è invertibile e $\mathbf{A} = \mathbf{I}^{-1}\mathbf{AI}$.
- b) Proprietà simmetrica: se $\mathbf{B} = \mathbf{S}^{-1}\mathbf{AS}$, allora $\mathbf{A} = \mathbf{T}^{-1}\mathbf{BT}$ dove \mathbf{T} è la matrice invertibile \mathbf{S}^{-1} .
- c) Proprietà transitiva: supponiamo $\mathbf{B} = \mathbf{S}^{-1}\mathbf{AS}$ e $\mathbf{C} = \mathbf{T}^{-1}\mathbf{BT}$. Il prodotto \mathbf{ST} di due matrici invertibili è invertibile ed $(\mathbf{ST})^{-1} = \mathbf{T}^{-1}\mathbf{S}^{-1}$. Allora

$$\mathbf{C} = \mathbf{T}^{-1}\mathbf{S}^{-1}\mathbf{AST} = (\mathbf{ST})^{-1}\mathbf{A}(\mathbf{ST})$$

e quindi \mathbf{C} è simile ad \mathbf{A} come volevasi dimostrare.

La similitudine, come ogni relazione di equivalenza, ripartisce le matrici in *classi di similitudine* a due a due disgiunte: due matrici sono nella stessa classe se e solo se sono simili. Il problema di trovare la matrice più semplice che rappresenta un'applicazione data diventa il problema di trovare la matrice più semplice all'interno di una classe di similitudine.

Le matrici $\text{diag}(1, 0, 2, 2)$ e $\text{diag}(2, 1, 0, 2)$ sono simili. Le matrici $\text{diag}(2, 3)$, $\text{diag}(3, 2)$ e $\begin{bmatrix} 2 & 7 \\ 0 & 3 \end{bmatrix}$ sono simili. Le matrici $\text{diag}(2, 3)$ e $\text{diag}(2, 4)$ non sono simili.

Due matrici diagonali $\text{diag}(\lambda_1, \dots, \lambda_n)$ e $\text{diag}(\mu_1, \dots, \mu_n)$ sono simili se e solo se gli elementi sulla diagonale principale coincidono a meno dell'ordine, cioè (μ_1, \dots, μ_n) sono una permutazione di $(\lambda_1, \dots, \lambda_n)$.

Non è detto, però, che due matrici con lo stesso polinomio caratteristico siano simili, se non si verifica l'ipotesi che le due matrici siano entrambe diagonalizzabili. Per esempio la matrice nulla $\mathbf{O} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ e la matrice $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ hanno lo stesso polinomio caratteristico λ^2 , ma non sono simili, perché una matrice simile alla matrice nulla dev'essere la matrice nulla:

$$\mathbf{S}^{-1}\mathbf{OS} = \mathbf{O}$$

o anche perché la molteplicità geometrica di $\lambda = 0$ come autovalore della matrice nulla è 2, mentre come autovalore di \mathbf{A} ha molteplicità geometrica 1. Si noti che in questo caso la matrice nulla è diagonalizzabile, mentre la matrice \mathbf{A} non lo è.

Non è nemmeno sufficiente che due matrici abbiano gli stessi autovalori con le stesse molteplicità algebriche e geometriche perché siano simili. Si considerino per esempio le due matrici

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{e} \quad \mathbf{B} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Entrambe le matrici hanno polinomio caratteristico λ^4 , quindi un unico autovalore $\lambda = 0$ con molteplicità algebrica 4. La molteplicità geometrica di $\lambda = 0$ è 4 meno il rango della matrice, quindi è 2 per entrambe le matrici. Tuttavia \mathbf{A} e \mathbf{B} non sono simili: un semplice calcolo mostra che $\mathbf{A}^2 = \mathbf{O}$, mentre \mathbf{B}^2 non è la matrice nulla (l'elemento di posto (1, 3), cioè il prodotto della prima riga di \mathbf{B} con la terza colonna, è uguale a 1); se \mathbf{A} e \mathbf{B} fossero simili, allora esisterebbe \mathbf{S} invertibile tale che $\mathbf{B} = \mathbf{S}^{-1}\mathbf{AS}$ e si otterebbe la contraddizione

$$\mathbf{B}^2 = \mathbf{S}^{-1}\mathbf{A}^2\mathbf{S} = \mathbf{O}$$

Si può dimostrare che, per $n = 2$ e $n = 3$, due matrici che abbiano gli stessi autovalori con le stesse molteplicità algebriche e geometriche sono simili, ma, come l'esempio precedente mostra, questo è falso per matrici di ordine $n \geq 4$.

Completiamo ora la dimostrazione della proposizione 4.9: dobbiamo dimostrare che la molteplicità geometrica di un autovalore è minore o uguale di quella algebrica.

CONCLUSIONE DELLA DIMOSTRAZIONE della proposizione 4.9. Sia \mathbf{A} una matrice quadrata di ordine n e sia μ un autovalore di \mathbf{A} : usiamo μ al posto di λ per non far confusione con la variabile λ del polinomio caratteristico. Dobbiamo dimostrare che la molteplicità geometrica $g = g_\mu$ dell'autovalore μ è minore o uguale della sua molteplicità algebrica $a = a_\mu$. Fissiamo una base $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_g\}$ dell'autospazio \mathbf{V}_μ relativo a μ . Possiamo aggiungere

$n - g$ vettori a questo insieme in modo da ottenere una base \mathcal{B} di \mathbb{K}^n . Sia \mathbf{M} la matrice che rappresenta l'applicazione $\mathfrak{L}_{\mathbf{A}}(\mathbf{x}) = \mathbf{Ax}$ rispetto alla base \mathcal{B} : si ricordi che la colonna k di \mathbf{M} è il vettore delle componenti di \mathbf{Av}_k rispetto alla base \mathcal{B} . Poiché:

$$\mathbf{Av}_k = \mu \mathbf{v}_k \quad \text{per } k = 1, 2, \dots, g$$

le prime g colonne della matrice \mathbf{M} sono i primi g vettori della base canonica di \mathbb{K}^n . Questo significa che \mathbf{M} ha la forma

$$\mathbf{M} = \begin{bmatrix} \mu \mathbf{I}_g & \mathbf{B} \\ \mathbf{O} & \mathbf{C} \end{bmatrix}$$

dove \mathbf{I}_g è la matrice identità di ordine g e \mathbf{O} è la matrice nulla con $n - g$ righe e g colonne. Siccome il determinante di una matrice triangolare a blocchi è uguale al prodotto dei blocchi sulla diagonale,

$$\det(\mathbf{M} - \lambda \mathbf{I}) = \det(\mu \mathbf{I}_g - \lambda \mathbf{I}_g) \det(\mathbf{C} - \lambda \mathbf{I}_{n-g}) = (\mu - \lambda)^g \det(\mathbf{C} - \lambda \mathbf{I}_{n-g})$$

Quindi $(\lambda - \mu)^g$ divide il polinomio caratteristico di \mathbf{M} . Ma \mathbf{M} è simile ad \mathbf{A} , perché le due matrici rappresentano la stessa applicazione rispetto a basi distinte, per cui il polinomio caratteristico di \mathbf{M} coincide con quello di \mathbf{A} . Concludiamo che $(\lambda - \mu)^g$ divide il polinomio caratteristico di \mathbf{A} e questo significa che la molteplicità algebrica di μ è almeno g , come volevasi dimostrare.

Autovettori e autovalori di polinomi di matrici

A partire da una matrice quadrata \mathbf{A} , se ne possono costruire molte altre con questo procedimento: per ogni polinomio $P(x) = a_d x^d + \dots + a_1 x + a_0$, si ottiene una nuova matrice quadrata $P(\mathbf{A})$ sostituendo la variabile x con \mathbf{A} :

$$P(\mathbf{A}) = a_d \mathbf{A}^d + \dots + a_1 \mathbf{A} + a_0 \mathbf{I}$$

Il caso più semplice e importante è $P(x) = x^d$: in questo caso $P(\mathbf{A})$ è la potenza \mathbf{A}^d di \mathbf{A} . Osserviamo che, se \mathbf{D} è una matrice diagonale, $P(\mathbf{D})$ è ancora una matrice diagonale:

$$P(\text{diag}(\lambda_1, \dots, \lambda_n)) = \text{diag}(P(\lambda_1), \dots, P(\lambda_n)).$$

Gli autovalori e autovettori di \mathbf{A} e $P(\mathbf{A})$ sono legati tra loro:

PROPOSIZIONE 5.5 Supponiamo che \mathbf{v} sia un autovettore di \mathbf{A} relativo all'autovalore λ . Allora:

- a) \mathbf{v} è un autovettore di \mathbf{A}^m relativo all'autovalore λ^m , per ogni $m \geq 1$;
- b) se \mathbf{A} è invertibile, allora $\lambda \neq 0$ e \mathbf{v} è un autovettore di \mathbf{A}^{-1} relativo all'autovalore $1/\lambda$;
- c) se $P(x)$ è un polinomio, allora \mathbf{v} è un autovettore di $P(\mathbf{A})$ relativo all'autovalore $P(\lambda)$; inoltre, se $\mathbf{B} = \mathbf{S}^{-1} \mathbf{A} \mathbf{S}$, allora

$$P(\mathbf{B}) = \mathbf{S}^{-1} P(\mathbf{A}) \mathbf{S}$$

In particolare, se \mathbf{A} e \mathbf{B} sono simili, allora $P(\mathbf{A})$ e $P(\mathbf{B})$ sono simili; se \mathbf{A} è diagonalizzabile, allora $P(\mathbf{A})$ è diagonalizzabile.

DIMOSTRAZIONE. Supponiamo $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Mostriamo il punto a): per induzione possiamo supporre che $\mathbf{A}^{m-1}\mathbf{v} = \lambda^{m-1}\mathbf{v}$, e poi calcolare

$$\mathbf{A}^m\mathbf{v} = \mathbf{A}(\mathbf{A}^{m-1}\mathbf{v}) = \mathbf{A}(\lambda^{m-1}\mathbf{v}) = \lambda^{m-1}\mathbf{A}\mathbf{v} = \lambda^{m-1}(\lambda\mathbf{v}) = \lambda^m\mathbf{v}$$

Mostriamo il punto b): se \mathbf{A} è invertibile, il determinante di \mathbf{A} è non nullo e quindi \mathbf{A} non ha un autovalore nullo: $\lambda \neq 0$. Moltiplicando l'uguaglianza $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ a sinistra per $\frac{1}{\lambda}\mathbf{A}^{-1}$ otteniamo

$$\frac{1}{\lambda}\mathbf{v} = \mathbf{A}^{-1}\mathbf{v}$$

Questo mostra il punto b).

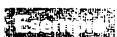
Il punto c) segue concettualmente dal fatto che \mathbf{B} rappresenta $\mathcal{L}_{\mathbf{A}}$ rispetto alla base di \mathbb{K}^n formata dalle colonne di \mathbf{S} e che $P(\mathbf{B})$ rappresenta $P(\mathcal{L}_{\mathbf{A}}) = \mathcal{L}_{P(\mathbf{A})}$ rispetto alla stessa base. Per una dimostrazione più concreta, osserviamo che

$$(\mathbf{S}^{-1}\mathbf{A}\mathbf{S})^m = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}\mathbf{S}^{-1}\mathbf{A}\mathbf{S} \cdots \mathbf{S}^{-1}\mathbf{A}\mathbf{S}\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \mathbf{S}^{-1}\mathbf{A}^m\mathbf{S}$$

Inoltre, per le proprietà del prodotto di matrici, l'operazione che manda \mathbf{A} in $\mathbf{S}^{-1}\mathbf{A}\mathbf{S}$ è lineare:

$$\mathbf{S}^{-1}(t_1\mathbf{A}_1 + t_2\mathbf{A}_2)\mathbf{S} = t_1\mathbf{S}^{-1}\mathbf{A}_1\mathbf{S} + t_2\mathbf{S}^{-1}\mathbf{A}_2\mathbf{S}$$

Siccome un polinomio è una combinazione lineare di monomi x^m , dalle due precedenti uguaglianze segue che $P(\mathbf{S}^{-1}\mathbf{A}\mathbf{S}) = \mathbf{S}^{-1}P(\mathbf{A})\mathbf{S}$ per ogni polinomio $P(x)$.



Lo stesso argomento mostra che, se $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{V}$ è un'applicazione lineare e $\mathcal{L}(\mathbf{v}) = \lambda\mathbf{v}$, allora $P(\mathcal{L})(\mathbf{v}) = P(\lambda)\mathbf{v}$. Questa semplice osservazione ha un'applicazione spettacolare, che trasforma le fondamentali equazioni differenziali lineari a coefficienti costanti in equazioni algebriche. Per semplicità ci limitiamo al caso delle equazioni omogenee del secondo ordine:

$$(5.1) \quad ay''(t) + by'(t) + cy(t) = 0$$

Il primo membro di questa equazione è il polinomio differenziale $P(\mathcal{D})$, dove

$$P(x) = ax^2 + bx + c$$

e $\mathcal{D}(y) = y'$ è l'operatore che a una funzione associa la sua derivata. Il punto è che, per ogni λ reale o complesso, la funzione esponenziale $y_{\lambda}(t) = e^{\lambda t}$ è un autovalore di \mathcal{D} relativo a λ perché:

$$\mathcal{D}(y_{\lambda}) = \lambda y_{\lambda}$$

Quindi $y_{\lambda}(t)$ è anche un autovettore di $P(\mathcal{D})$ relativo all'autovalore $P(\lambda)$: questo esplicitamente significa

$$ay_{\lambda}''(t) + by_{\lambda}'(t) + cy_{\lambda}(t) = P(\lambda)y_{\lambda}(t)$$

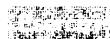
La conseguenza importante è questa: se $P(\lambda) = 0$, allora la funzione $y_{\lambda}(t)$ è soluzione dell'equazione differenziale (5.1). Abbiamo così sostanzialmente ridotto l'equazione differenziale all'equazione algebrica $P(\lambda) = 0$, che si dice equazione caratteristica di (5.1). Per un esempio specifico, consideriamo l'equazione

$$(5.2) \quad y''(t) - 4y(t) = 0$$

che corrisponde al polinomio $x^2 - 4$. L'equazione caratteristica $P(\lambda) = \lambda^2 - 4 = 0$ ha due radici $\lambda_1 = 2$ e $\lambda_2 = -2$. In corrispondenza troviamo due soluzioni dell'equazione differenziale (5.2)

$$y_1(t) = e^{2t} \quad \text{e} \quad y_2(t) = e^{-2t}$$

Siccome l'equazione è del secondo ordine, non è poi difficile mostrare che ogni soluzione di (5.2) è una combinazione lineare $c_1 e^{2t} + c_2 e^{-2t}$ di y_1 e y_2 . Abbiamo così trovato tutte le soluzioni dell'equazione differenziale.



Sia \mathbf{A} una matrice simile alla matrice diagonale $\text{diag}(1, 2, 3)$. Si tratta di una matrice diagonalizzabile? Qual è il suo polinomio caratteristico? La matrice \mathbf{A}^3 è diagonalizzabile? Qual è il suo polinomio caratteristico?

Stabilire per quali valori del parametro reale k le due matrici

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 2 & 3 \end{bmatrix} \quad \text{e} \quad \mathbf{B} = \begin{bmatrix} k & 3 \\ 0 & 4 \end{bmatrix}$$

sono simili.

Stabilire per quali valori del parametro complesso t le due matrici

$$\mathbf{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad \text{e} \quad \mathbf{B} = \begin{bmatrix} i & t \\ 1 & -i \end{bmatrix}$$

sono simili (come matrici complesse).

Stabilire per quali valori del parametro reale k le due matrici

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 0 \\ k-2 & 3 & 0 \\ 0 & k-1 & 2 \end{bmatrix} \quad \text{e} \quad \mathbf{B} = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

sono simili.

Stabilire per quali valori del parametro reale k la matrice

$$\mathbf{A} = \begin{bmatrix} k-1 & k-2 & 2-k \\ 0 & 1 & 0 \\ k-2 & k-2 & 3-k \end{bmatrix}$$

è diagonalizzabile. In corrispondenza a tali valori scrivere una matrice diagonale simile ad \mathbf{A} e la relativa matrice di passaggio.

Si consideri l'equazione matriciale $\mathbf{B} = \mathbf{X}^{-1} \mathbf{A} \mathbf{X}$, con

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 1 \\ 2 & 2 & 1 \\ 1 & -1 & 0 \end{bmatrix} \quad \text{e} \quad \mathbf{B} = \begin{bmatrix} 4 & 0 & 0 \\ -3 & 1 & 0 \\ -1 & -1 & 0 \end{bmatrix}$$

e \mathbf{X} incognita. L'equazione ammette soluzioni? perché? La soluzione, se esiste, è unica? Si determini, se possibile, una soluzione.

Supponiamo che \mathbf{A} sia diagonalizzabile e che \mathbf{B} non lo sia. Mostrare che \mathbf{A} e \mathbf{B} non sono simili (il motivo è che la diagonalizzabilità dipende solo dall'applicazione lineare rappresentata dalla matrice, e non dalla particolare matrice usata per rappresentarla).

Per quali valori di k le matrici

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & k & 4 \\ 0 & 0 & 5 \end{bmatrix} \quad \text{e} \quad \mathbf{B} = \begin{bmatrix} 1 & 4 & 0 \\ 0 & 5 & 0 \\ 2 & 3 & 1 \end{bmatrix}$$

sono simili?

Suggerimento: per i valori di k plausibili che relazione c'è tra le colonne di \mathbf{A} e quelle di \mathbf{B} ? come si esprime tale relazione in termini delle applicazioni rappresentate dalle due matrici?

Sia \mathbf{A} una matrice quadrata di ordine n e sia $\mathbf{B} = \mathbf{A}^2 - 4\mathbf{A} + \mathbf{I}$. Se \mathbf{v} è un autovettore di \mathbf{A} relativo all'autovalore 5, allora \mathbf{v} è un autovettore anche di \mathbf{B} ? qual è l'autovalore corrispondente?

Sia \mathbf{A} una matrice quadrata. Mostrare che, se $\mathbf{A}^2 = \mathbf{O}$, allora $\mathbf{I} + \mathbf{A}$ è invertibile.

Suggerimento: la condizione $\mathbf{A}^2 = \mathbf{O}$ determina gli autovalori di \mathbf{A} .

Come nel testo, poniamo $\mathcal{D}(y(t)) = y'(t)$ e $y_\lambda(t) = e^{\lambda t}$.

a) Mostrare per induzione che

$$\mathcal{D}^m(ty_\lambda(t)) = \lambda^m ty_\lambda(t) + m\lambda^{m-1} y_\lambda(t)$$

b) Dedurre che per ogni polinomio $P(x)$:

$$P(\mathcal{D})(ty_\lambda(t)) = P(\lambda)ty_\lambda(t) + P'(\lambda)y_\lambda(t)$$

c) Mostrare che $ty_\lambda(t)$ risolve l'equazione differenziale

$$y'' - 2\lambda y'(t) + \lambda^2 y(t) = 0$$

Una matrice quadrata \mathbf{N} si dice *nilpotente* se esiste un intero positivo m tale che \mathbf{N}^m sia la matrice nulla.

- a) Sia \mathbf{N} una matrice nilpotente. Si mostri che $\lambda = 0$ è l'unico autovalore di \mathbf{N} .
- b) Si mostri che una matrice simile a una matrice nilpotente è nilpotente.
- c) Sia \mathbf{N} una matrice nilpotente di ordine 2. Si mostri che \mathbf{N} è simile a una matrice \mathbf{U} della forma

$$\mathbf{U} = \begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix}$$

Si conclude che \mathbf{N}^2 è la matrice nulla.

Suggerimento: si rappresenti l'applicazione lineare $\mathbf{N}\mathbf{x}$ rispetto a una base $\{\mathbf{v}_1, \mathbf{v}_2\}$ di \mathbb{R}^2 in cui \mathbf{v}_1 appartiene al nucleo (cioè un autovettore di \mathbf{N}).

- d) Sia $\mathbf{U} = [u_{ij}]$ una matrice quadrata di ordine n . Si supponga che \mathbf{U} sia triangolare superiore con gli elementi sulla diagonale principale nulli:

$$u_{ij} = 0 \quad \text{se } i \geq j$$

Si mostri che \mathbf{U} è nilpotente.

- e) Si dia un esempio di una matrice nilpotente 2×2 che non sia triangolare.

6 IL PROBLEMA DELLA FORMA CANONICA

La proposizione 5.4 fornisce, sotto l'ipotesi di diagonalizzabilità, la risposta ai due problemi:

- i) determinare condizioni necessarie e sufficienti affinché due matrici siano simili;
- ii) all'interno di ogni classe di similitudine, individuare un *rappresentante canonico*, che sia unicamente determinato ed esprima nel modo più semplice possibile le proprietà della classe, in modo che si possa dire che due matrici sono simili se e solo se hanno lo stesso rappresentante canonico.

La risposta per le matrici diagonalizzabili è:

- i) due matrici *diagonalizzabili* sono simili se e solo se hanno lo stesso polinomio caratteristico (equivalentemente, gli stessi autovalori contati con la loro molteplicità algebrica);
- ii) la forma canonica di una matrice diagonalizzabile è la matrice diagonale che ha sulla diagonale gli autovalori della matrice (ripetuti con la loro molteplicità algebrica).

Una questione analoga si pone ogni qualvolta un insieme sia ripartito in classi di equivalenza. A questo proposito, è utile tenere presente l'esempio dei numeri razionali positivi, che sono definiti come classi di equivalenza di frazioni. Una frazione è una coppia ordinata (a, b) di numeri interi positivi, che viene rappresentata con il simbolo $\frac{a}{b}$; due frazioni $\frac{a}{b}$ e $\frac{c}{d}$ sono equivalenti e, quindi, definiscono lo stesso numero razionale, se $ad - bc = 0$; si verifica facilmente che si tratta di una relazione di equivalenza, per la quale il punto i) è già dato per definizione, ma il punto ii) è interessante: il rappresentante canonico è la frazione ridotta ai minimi termini, e due frazioni definiscono lo stesso numero razionale se e solo se sono uguali una volta ridotte ai minimi termini. Un altro esempio è fornito dai *vettori liberi*: nello spazio euclideo, due vettori applicati si dicono equivalenti se uno si ottiene dall'altro per traslazione e un vettore libero è una classe di equivalenza di vettori applicati; in questo caso

- i) due vettori applicati sono equivalenti se e solo se hanno ugual modulo, direzione e verso;
- ii) fissata un'origine O , il rappresentante canonico di un vettore libero \mathbf{v} si ottiene applicando \mathbf{v} in O , e due vettori liberi sono equivalenti se e solo se, una volta traslati nell'origine, coincidono.

In questo paragrafo illustriamo la soluzione di i) e ii) per la relazione di similitudine tra matrici. Introduciamo subito la classe delle *matrici di Jordan*: vedremo poi che ogni matrice \mathbf{A} è simile, mediante di una matrice di passaggio complessa, a una matrice di Jordan \mathbf{B} ; la matrice di Jordan \mathbf{B} è determinata essenzialmente da \mathbf{A} , e si dice *forma canonica di Jordan* di \mathbf{A} . Quando \mathbf{A} è diagonalizzabile, la forma canonica di Jordan di \mathbf{A} è la matrice diagonale che ha gli autovalori di \mathbf{A} come elementi della diagonale principale.

Ricordiamo la definizione di blocco di Jordan $\mathbf{J}_m(\lambda)$:

$$\mathbf{J}_m(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda \end{bmatrix}$$

è la matrice $m \times m$ che ha gli elementi sulla diagonale principale uguali a λ , e gli elementi sulla diagonale sopra a quella principale uguali a 1.

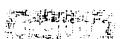
DEFINIZIONE 6.1 (Matrice di Jordan)

Una matrice quadrata \mathbf{C} si dice *matrice di Jordan* o *matrice a blocchi di Jordan* se è una matrice diagonale a blocchi

$$(6.1) \quad \mathbf{C} = \begin{bmatrix} \mathbf{J}_1 & \mathbf{O} & \dots & \mathbf{O} \\ \mathbf{O} & \mathbf{J}_2 & \dots & \mathbf{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O} & \mathbf{O} & \dots & \mathbf{J}_t \end{bmatrix}$$

in cui ciascun blocco \mathbf{J}_h è un blocco di Jordan.

OSSERVAZIONE Una matrice di Jordan è *bidiagonale*, nel senso che tutti gli elementi della matrice sono nulli tranne quelli che si trovano sulla diagonale principale o sulla diagonale immediatamente sopra alla diagonale principale. In particolare, una matrice di Jordan è triangolare alta, e quindi i suoi autovalori sono esattamente gli elementi della diagonale principale. Gli elementi sulla diagonale sopra alla diagonale principale sono quelli di posto $(i, i + 1)$ e sono uguali a 0 oppure a 1.



Le due matrici

$$\mathbf{C}_1 = \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array} \right] \quad \text{e} \quad \mathbf{C}_2 = \left[\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array} \right]$$

sono matrici di Jordan, e hanno un unico autovalore $\lambda_1 = 0$ con molteplicità algebrica 4. La matrice \mathbf{C}_1 ha due blocchi di Jordan di ordine due $\mathbf{J}_2(0) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, mentre la matrice \mathbf{C}_2 ha un blocco di ordine tre $\mathbf{J}_3(0) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ e un blocco di ordine uno $[0] = \mathbf{J}_1(0)$.

La dimostrazione del fatto che ogni matrice è simile a una matrice di Jordan è piuttosto lunga, e richiede ragionamenti per induzione per i quali è necessario pensare alle matrici come applicazioni lineari. Cercheremo di evitare notazioni pesanti in questo modo: il simbolo \mathbf{V} denoterà uno spazio vettoriale complesso di dimensione finita, e il simbolo \mathbf{T} un'applicazione lineare $\mathbf{V} \rightarrow \mathbf{V}$; scriveremo $\mathbf{T}\mathbf{v}$ al posto di $\mathbf{T}(\mathbf{v})$ e \mathbf{T}^m

per indicare il prodotto di composizione di m fattori uguali a \mathbf{T} . Quindi

$$\mathbf{T}^2\mathbf{v} = \mathbf{T}(\mathbf{T}\mathbf{v}), \quad \mathbf{T}^3\mathbf{v} = \mathbf{T}(\mathbf{T}^2\mathbf{v}), \quad \mathbf{T}^m\mathbf{v} = \mathbf{T}(\mathbf{T}^{m-1}\mathbf{v})$$

Diremo anche che \mathbf{T} è un *operatore*. In sostanza sostituiamo la notazione $\mathfrak{L}(\mathbf{v})$, che abbiamo fin qui usato per le applicazioni lineari, con la più snella $\mathbf{T}\mathbf{v}$; quando $\mathbf{V} = \mathbb{C}^n$, l'operatore \mathbf{T} si identifica con la matrice $n \times n$ che lo rappresenta rispetto alla base canonica e $\mathbf{T}\mathbf{v}$ è proprio il prodotto della matrice \mathbf{T} per il vettore colonna \mathbf{v} . Denoteremo con il simbolo \mathbf{I} tanto l'operatore identità (definito da $\mathbf{I}\mathbf{v} = \mathbf{v}$ per ogni $\mathbf{v} \in \mathbf{V}$) quanto la matrice identità, che è la matrice rappresentativa dell'operatore identità rispetto a una base arbitraria.

Abbiamo visto che una matrice è diagonalizzabile su \mathbb{C} se e solo se esiste una base di \mathbb{C}^n formata da autovettori della matrice. Per arrivare a dimostrare il teorema sulla forma canonica di Jordan occorre introdurre la nozione di *autovettore generalizzato*: mentre non è sempre possibile trovare una base di \mathbb{C}^n formata da autovettori, è invece sempre possibile trovare una base di \mathbb{C}^n formata da autovettori generalizzati, e questa base produce una matrice di Jordan simile alla matrice di partenza.

DEFINIZIONE 6.2 (Autovettore generalizzato)

Sia λ un autovalore di \mathbf{T} . Un vettore non nullo $\mathbf{v} \in \mathbf{V}$ si dice *autovettore generalizzato* relativo a λ se esiste un intero $d \geq 1$ tale che $(\mathbf{T} - \lambda\mathbf{I})^d\mathbf{v} = \mathbf{0}$. Il più piccolo intero d tale che $(\mathbf{T} - \lambda\mathbf{I})^d\mathbf{v} = \mathbf{0}$ si dice *indice* dell'autovettore generalizzato.

Dalla definizione segue che \mathbf{v} è un autovettore generalizzato relativo a λ di indice d se $(\mathbf{T} - \lambda\mathbf{I})^{d-1}\mathbf{v} \neq \mathbf{0}$ e $(\mathbf{T} - \lambda\mathbf{I})^d\mathbf{v} = \mathbf{0}$. In particolare, gli autovettori generalizzati di indice 1 sono precisamente gli autovettori: quando $d = 1$, le condizioni precedenti diventano $\mathbf{v} \neq \mathbf{0}$ e $\mathbf{T}\mathbf{v} = \lambda\mathbf{v}$. Dato un autovettore generalizzato \mathbf{v} di indice d , definiamo induttivamente una successione di d autovettori generalizzati \mathbf{w}_k ponendo:

$$(6.2) \quad \mathbf{w}_d = \mathbf{v} \quad \text{e} \quad \mathbf{w}_{k-1} = (\mathbf{T} - \lambda\mathbf{I})\mathbf{w}_k \quad \text{per } k = 2, \dots, d$$

Otteniamo così d vettori $\mathbf{w}_1, \dots, \mathbf{w}_d$: l'ultimo è l'autovettore generalizzato \mathbf{v} da cui siamo partiti. Per definizione $\mathbf{w}_k = (\mathbf{T} - \lambda\mathbf{I})^{d-k}\mathbf{v}$ per $k = 1, \dots, d$, quindi \mathbf{w}_k è un autovettore generalizzato di indice k ; in particolare \mathbf{w}_1 è un autovettore di \mathbf{T} . La relazione $\mathbf{w}_{k-1} = (\mathbf{T} - \lambda\mathbf{I})\mathbf{w}_k$ che definisce \mathbf{w}_{k-1} a partire da \mathbf{w}_k si può riscrivere nella forma

$$(6.3) \quad \mathbf{T}\mathbf{w}_k = \lambda\mathbf{w}_k + \mathbf{w}_{k-1} \quad \text{per } k = 2, \dots, d$$

Si vede così che \mathbf{w}_{k-1} dà una misura di quanto \mathbf{w}_k è lontano dall'essere un vero autovettore. Per il vettore \mathbf{w}_1 , che come abbiamo osservato è un vero autovettore, vale invece l'equazione:

$$(6.4) \quad \mathbf{T}\mathbf{w}_1 = \lambda\mathbf{w}_1$$

DEFINIZIONE 6.3 (Catena di autovettori generalizzati)

Una *catena di autovettori generalizzati di lunghezza d* ≥ 1 relativa all'autovettore λ di T è una successione finita $\mathbf{w}_1, \dots, \mathbf{w}_d$ di vettori di V con la proprietà che

$$(6.5) \quad \mathbf{T}\mathbf{w}_1 = \lambda\mathbf{w}_1 \quad \text{e} \quad \mathbf{T}\mathbf{w}_k = \lambda\mathbf{w}_k + \mathbf{w}_{k-1} \quad \text{per } k = 2, \dots, d.$$

OSSERVAZIONE La discussione precedente si riassume così: se $\mathbf{w}_1, \dots, \mathbf{w}_d$ è una catena di autovettori generalizzati, il primo vettore \mathbf{w}_1 è un vero autovettore, mentre per $k \geq 2$ il vettore \mathbf{w}_k è un autovettore generalizzato di indice k , e la catena è determinata dall'ultimo vettore \mathbf{w}_d perché $\mathbf{w}_k = (\mathbf{T} - \lambda\mathbf{I})^{d-k}\mathbf{w}$ per $k = 1, \dots, d$.

Consideriamo il blocco di Jordan $\mathbf{T} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, che è l'esempio più semplice di matrice non diagonalizzabile. La matrice ha un unico autovalore $\lambda = 0$. Il primo vettore della base canonica \mathbf{e}_1 è un autovettore: $\mathbf{T}\mathbf{e}_1 = \mathbf{0} = 0\mathbf{e}_1$. D'altra parte $\mathbf{T}\mathbf{e}_2 = 0\mathbf{e}_2 + \mathbf{e}_1$. Questo mostra che $\mathbf{e}_1, \mathbf{e}_2$ è una catena di autovettori generalizzati di lunghezza 2 per la matrice T.

Dato un operatore $\mathbf{T} : \mathbf{V} \rightarrow \mathbf{V}$, una base \mathcal{B} di V si dice *base di Jordan* per T se \mathcal{B} è l'unione di un certo numero di catene di autovettori generalizzati a due a due disgiunte. Quindi una base di Jordan è un base della forma

$$\{\mathbf{w}_{\ell,1}, \dots, \mathbf{w}_{\ell,d_\ell} : \ell = 1, \dots, t\}$$

dove, per ogni ℓ fissato, $\mathbf{w}_{\ell,1}, \dots, \mathbf{w}_{\ell,d_\ell}$ è una catena di autovettori generalizzati. Si noti che l'indice d_ℓ dell'autovettore generalizzato \mathbf{w}_{ℓ,d_ℓ} dipende da ℓ , in modo che la lunghezza delle catene possa variare. La seguente proposizione generalizza il fatto che una matrice quadrata è diagonalizzabile se e solo se esiste una base di autovettori:

PROPOSIZIONE 6.4 Sia $\mathbf{T} : \mathbf{V} \rightarrow \mathbf{V}$ un operatore lineare e sia \mathcal{B} una base di V. La matrice che rappresenta T rispetto a \mathcal{B} è una *matrice di Jordan* se e solo se \mathcal{B} è una *base di Jordan* per T.

DIMOSTRAZIONE. Siano $\mathbf{b}_1, \dots, \mathbf{b}_n$ i vettori della base \mathcal{B} , e sia \mathbf{A} la matrice che rappresenta T rispetto alla base \mathcal{B} . Questo significa che gli elementi a_{ik} della colonna k di \mathbf{A} sono le componenti di $\mathbf{T}\mathbf{b}_k$ rispetto a \mathcal{B} :

$$(6.6) \quad \mathbf{T}\mathbf{b}_k = a_{1k}\mathbf{b}_1 + \cdots + a_{nk}\mathbf{b}_n$$

Supponiamo che \mathbf{A} sia una matrice di Jordan, composta da t blocchi di Jordan $\mathbf{J}_1, \dots, \mathbf{J}_t$. Sia λ_i l'autovalore del blocco \mathbf{J}_i , e supponiamo che il primo elemento di \mathbf{J}_i si trovi sulla colonna k_i . Allora la colonna k_i di \mathbf{A} ha l'elemento sulla diagonale principale uguale a λ_i , e tutti gli altri elementi nulli; mentre le altre colonne del blocco \mathbf{J}_i , che sono quelle di indice k con $k_i < k < k_{i+1}$, hanno l'elemento sulla diagonale principale uguale a λ_i , quello immediatamente sopra uguale a 1, e gli altri elementi nulli. Da (6.6) segue allora

$$(6.7) \quad \mathbf{T}\mathbf{b}_{k_i} = \lambda_i\mathbf{b}_{k_i} \quad \text{e} \quad \mathbf{T}\mathbf{b}_k = \lambda_i\mathbf{b}_k + \mathbf{b}_{k-1} \quad \text{per } k_i < k < k_{i+1}$$

Questo significa che $\mathbf{b}_{k_i}, \mathbf{b}_{k_i+1}, \dots, \mathbf{b}_{k_{i+1}-1}$ è una catena di Jordan relativa all'autovalore λ_i , e quindi \mathcal{B} è una base di Jordan. L'argomento si può leggere al contrario: se \mathcal{B} è una base di Jordan, allora vale la (6.7) e la matrice \mathbf{A} è di Jordan.

Introduciamo ora i *sottospazi radicali* che sono i sottospazi formati dagli autovettori generalizzati relativi a uno stesso autovalore: questi sottospazi da un lato sono interessanti di per sé, e dall'altro sono utili nella dimostrazione dell'esistenza di una base di Jordan.

DEFINIZIONE 6.5 (Sottospazio radicale)

Sia λ un autovalore di \mathbf{T} . L'insieme

$$\mathbf{R}_\lambda(\mathbf{T}) = \{\mathbf{v} \in \mathbf{V} : \text{esiste } d \geq 1 \text{ tale che } (\mathbf{T} - \lambda \mathbf{I})^d \mathbf{v} = \mathbf{0}\}$$

si dice *sottospazio radicale* di \mathbf{T} relativo all'autovalore λ .

OSSERVAZIONE I vettori non nulli del sottospazio radicale \mathbf{R}_λ sono gli autovettori generalizzati relativi all'autovalore λ . In particolare, il sottospazio radicale contiene gli autovettori relativi a λ e quindi l'autospazio \mathbf{V}_λ . Come il nome suggerisce, si tratta di un sottospazio vettoriale di \mathbf{V} : questo si può verificare direttamente, ma è anche una conseguenza della proposizione che segue.

PROPOSIZIONE 6.6 (Indice di un autovalore)

Sia \mathbf{V} uno spazio vettoriale di dimensione finita, sia $\mathbf{T} : \mathbf{V} \rightarrow \mathbf{V}$ un operatore lineare, e sia λ un autovalore di \mathbf{T} . Esiste un unico intero $e = e_\lambda \geq 1$ tale che

$$\begin{aligned} \text{Ker}(\mathbf{T} - \lambda \mathbf{I})^d &\subsetneq \text{Ker}(\mathbf{T} - \lambda \mathbf{I})^{d+1} & \text{se } 0 \leq d < e, \\ \text{Ker}(\mathbf{T} - \lambda \mathbf{I})^d &= \text{Ker}(\mathbf{T} - \lambda \mathbf{I})^{d+1} & \text{se } d \geq e. \end{aligned}$$

In particolare, il sottospazio radicale \mathbf{R}_λ coincide con $\text{Ker}(\mathbf{T} - \lambda \mathbf{I})^e$ ed è un sottospazio vettoriale di \mathbf{V} . L'intero $e = e_\lambda$ si dice *indice* dell'autovalore λ , ed è minore o uguale a $\dim \mathbf{V}$.

DIMOSTRAZIONE. Il simbolo $\text{Ker}(\mathbf{T} - \lambda \mathbf{I})^d \subsetneq \text{Ker}(\mathbf{T} - \lambda \mathbf{I})^{d+1}$ significa che $\text{Ker}(\mathbf{T} - \lambda \mathbf{I})^d$ è contenuto in $\text{Ker}(\mathbf{T} - \lambda \mathbf{I})^{d+1}$ e non è uguale a $\text{Ker}(\mathbf{T} - \lambda \mathbf{I})^{d+1}$.

Mostriamo che per ogni operatore lineare $\mathbf{B} : \mathbf{V} \rightarrow \mathbf{V}$ esiste un unico intero $e \geq 0$, minore o uguale a $n = \dim \mathbf{V}$, tale che

$$(6.8) \quad \text{Ker } \mathbf{B}^d \subsetneq \text{Ker } \mathbf{B}^{d+1} \quad \text{se } 0 \leq d < e, \quad \text{Ker } \mathbf{B}^d = \text{Ker } \mathbf{B}^{d+1} \quad \text{se } d \geq e$$

Se $d \geq 1$ e \mathbf{v} appartiene al nucleo di \mathbf{B}^d , allora

$$\mathbf{B}^{d+1} \mathbf{v} = \mathbf{B}(\mathbf{B}^d \mathbf{v}) = \mathbf{B}\mathbf{0} = \mathbf{0}$$

Se $d = 0$, allora $\mathbf{B}^0 = \mathbf{I}$ e $\text{Ker } \mathbf{B}^0 = \{\mathbf{0}\}$. Quindi il nucleo di \mathbf{B}^d è contenuto nel nucleo di \mathbf{B}^{d+1} per ogni $d \geq 0$, e possiamo considerare la catena di inclusioni

$$\{\mathbf{0}\} = \text{Ker } \mathbf{B}^0 \subseteq \text{Ker } \mathbf{B} \subseteq \text{Ker } \mathbf{B}^2 \subseteq \cdots \subseteq \text{Ker } \mathbf{B}^n \subseteq \text{Ker } \mathbf{B}^{n+1}$$

Le dimensioni di questi sottospazi formano una successione non decrescente di interi

$$0 \leq \dim \text{Ker } \mathbf{B} \leq \dim \text{Ker } \mathbf{B}^2 \leq \cdots \leq \dim \text{Ker } \mathbf{B}^n \leq \dim \text{Ker } \mathbf{B}^{n+1}$$

Ora $\dim \text{Ker } \mathbf{B}^{n+1} \leq \dim \mathbf{V} = n$ quindi questa successione non può essere strettamente crescente, deve esistere cioè un intero d compreso tra 0 e n per cui $\dim \text{Ker } \mathbf{B}^d = \dim \text{Ker } \mathbf{B}^{d+1}$. Sia e il più piccolo intero per cui ciò succede, in modo che

$$(6.9) \quad \dim \text{Ker } \mathbf{B}^d < \dim \text{Ker } \mathbf{B}^{d+1} \text{ se } 0 \leq d < e, \quad \dim \text{Ker } \mathbf{B}^e = \dim \text{Ker } \mathbf{B}^{e+1}$$

Osserviamo che, se $\mathbf{H}_1 \subseteq \mathbf{H}_2$ sono due sottospazi di \mathbf{V} , allora $\mathbf{H}_1 = \mathbf{H}_2$ è equivalente a $\dim \mathbf{H}_1 = \dim \mathbf{H}_2$. Quindi la (6.9) è equivalente a

$$(6.10) \quad \text{Ker } \mathbf{B}^d \subsetneq \text{Ker } \mathbf{B}^{d+1} \text{ se } 0 \leq d < e, \quad \text{Ker } \mathbf{B}^e = \text{Ker } \mathbf{B}^{e+1}$$

Per terminare la dimostrazione della (6.8) dobbiamo ancora mostrare che $\text{Ker } \mathbf{B}^d = \text{Ker } \mathbf{B}^{d+1}$ per ogni $d \geq e + 1$. Supponiamo quindi che $d \geq e + 1$. Dato $\mathbf{v} \in \text{Ker } \mathbf{B}^{d+1}$, sia $\mathbf{w} = \mathbf{B}^{d-e}\mathbf{v}$. Allora

$$\mathbf{0} = \mathbf{B}^{d+1}\mathbf{v} = \mathbf{B}^{e+1}\mathbf{B}^{d-e}\mathbf{v} = \mathbf{B}^{e+1}\mathbf{w}$$

Questo significa che \mathbf{w} appartiene al nucleo di \mathbf{B}^{e+1} , che però coincide con il nucleo di \mathbf{B}^e . Quindi:

$$\mathbf{0} = \mathbf{B}^e\mathbf{w} = \mathbf{B}^e\mathbf{B}^{d-e}\mathbf{v} = \mathbf{B}^d\mathbf{v}$$

Questo mostra che $\mathbf{v} \in \text{Ker } \mathbf{B}^d$ per ogni $\mathbf{v} \in \text{Ker } \mathbf{B}^{d+1}$, cioè che $\text{Ker } \mathbf{B}^{d+1} \subseteq \text{Ker } \mathbf{B}^d$. Abbiamo già mostrato che l'inclusione opposta vale per ogni d , quindi $\text{Ker } \mathbf{B}^d = \text{Ker } \mathbf{B}^{d+1}$ per ogni $d \geq e + 1$. La dimostrazione della (6.8) è così completa.

Poniamo ora $\mathbf{B} = \mathbf{T} - \lambda\mathbf{I}$: siccome λ è un autovalore di \mathbf{T} , il nucleo di $\mathbf{T} - \lambda\mathbf{I}$ non è ridotto al vettore nullo, per cui e dev'essere almeno 1. Otteniamo così la (6.8). Il sottospazio radicale \mathbf{R}_λ consiste di quei vettori \mathbf{v} per cui esiste $d \geq 1$ tale che $\mathbf{v} \in \text{Ker}(\mathbf{T} - \lambda\mathbf{I})^d$. Abbiamo appena mostrato che $\text{Ker}(\mathbf{T} - \lambda\mathbf{I})^d \subseteq \text{Ker}(\mathbf{T} - \lambda\mathbf{I})^e$ per ogni d , quindi $\mathbf{R}_\lambda = \text{Ker}(\mathbf{T} - \lambda\mathbf{I})^e$.

OSSERVAZIONE L'indice e_λ è quindi il massimo degli indici degli autovettori generalizzati relativi a λ .

OSSERVAZIONE L'indice e_λ di un autovalore è 1, il minimo possibile, se e solo se ogni autovettore generalizzato relativo a λ è un vero autovettore. Infatti l'indice è 1 se e solo se il sottospazio radicale \mathbf{R}_λ coincide con l'autospazio $\mathbf{V}_\lambda = \text{Ker}(\mathbf{T} - \lambda\mathbf{I})$. Più avanti mostreremo che la dimensione del sottospazio radicale è uguale alla molteplicità algebrica di λ . Siccome la dimensione dell'autospazio \mathbf{V}_λ è la molteplicità geometrica, ne segue che un autovalore ha indice 1 se e solo se è un autovalore regolare. Quindi \mathbf{T} è diagonalizzabile se e solo se ogni suo autovalore ha indice 1.

Il prossimo lemma generalizza il fatto che autovettori relativi ad autovalori distinti sono linearmente indipendenti.

LEMMA 6.7 (Indipendenza degli autovettori generalizzati)

Supponiamo che $\lambda_1, \dots, \lambda_s$ siano autovalori distinti di \mathbf{T} e che $\mathbf{v}_k \in \mathbf{R}_{\lambda_k}$ per ogni $k = 1, \dots, s$. Se

$$\mathbf{v}_1 + \cdots + \mathbf{v}_s = \mathbf{0}$$

allora $\mathbf{v}_k = \mathbf{0}$ per ogni k .

DIMOSTRAZIONE. Supponiamo per assurdo $\mathbf{v}_1 \neq \mathbf{0}$. Allora \mathbf{v}_1 è un autovettore generalizzato. Sia d l'indice di \mathbf{v}_1 , cioè l'intero d tale che $(\mathbf{T} - \lambda_1 \mathbf{I})^{d-1} \mathbf{v} = \mathbf{0}$ e $(\mathbf{T} - \lambda_1 \mathbf{I})^{d-1} \mathbf{v} \neq \mathbf{0}$. Allora $\mathbf{w} = (\mathbf{T} - \lambda_1 \mathbf{I})^{d-1} \mathbf{v}$ è un autovettore di \mathbf{T} relativo a λ_1 : $\mathbf{w} \neq \mathbf{0}$ e $\mathbf{T}\mathbf{w} = \lambda_1 \mathbf{w}$.

Per ogni $k \geq 2$ sia e_k l'indice dell'autovalore λ_k , di modo che $\mathbf{R}_{\lambda_k} = \text{Ker}(\mathbf{T} - \lambda \mathbf{I})^{e_k}$. Consideriamo l'operatore

$$\mathbf{B} = (\mathbf{T} - \lambda_1 \mathbf{I})^{d-1} (\mathbf{T} - \lambda_2 \mathbf{I})^{e_2} \cdots (\mathbf{T} - \lambda_s \mathbf{I})^{e_s}$$

L'operatore \mathbf{B} si decomponе in fattori della forma $(\mathbf{T} - \lambda_j \mathbf{I})$, e questi commutano tra di loro:

$$(\mathbf{T} - \lambda_j \mathbf{I})(\mathbf{T} - \lambda_h \mathbf{I}) = \mathbf{T}^2 - (\lambda_j + \lambda_h) \mathbf{T} + \lambda_j \lambda_h \mathbf{I} = (\mathbf{T} - \lambda_h \mathbf{I})(\mathbf{T} - \lambda_j \mathbf{I})$$

Ne segue che possiamo permutare i fattori di \mathbf{B} senza cambiare il loro prodotto \mathbf{B} . In particolare possiamo calcolare $\mathbf{B}\mathbf{v}_2$ scrivendo $(\mathbf{T} - \lambda_2 \mathbf{I})^{e_2}$ come ultimo fattore di \mathbf{B} :

$$\mathbf{B}\mathbf{v}_2 = (\mathbf{T} - \lambda_1 \mathbf{I})^{d-1} (\mathbf{T} - \lambda_3 \mathbf{I})^{e_3} \cdots (\mathbf{T} - \lambda_s \mathbf{I})^{e_s} (\mathbf{T} - \lambda_2 \mathbf{I})^{e_2} \mathbf{v}_2 = \mathbf{0}$$

Nell'ultimo passaggio abbiamo usato il fatto che $\mathbf{v}_2 \in \mathbf{R}_{\lambda_2} = \text{Ker}(\mathbf{T} - \lambda_2 \mathbf{I})^{e_2}$. Nello stesso modo si mostra che $\mathbf{B}\mathbf{v}_k = \mathbf{0}$ per ogni $k \geq 2$. Calcoliamo ora $\mathbf{B}\mathbf{v}_1$:

$$\mathbf{B}\mathbf{v}_1 = (\mathbf{T} - \lambda_2 \mathbf{I})^{e_2} \cdots (\mathbf{T} - \lambda_s \mathbf{I})^{e_s} (\mathbf{T} - \lambda_1 \mathbf{I})^{d-1} \mathbf{v}_1 = (\mathbf{T} - \lambda_2 \mathbf{I})^{e_2} \cdots (\mathbf{T} - \lambda_s \mathbf{I})^{e_s} \mathbf{w}$$

Per costruzione \mathbf{w} è un autovettore di \mathbf{T} relativo all'autovalore λ_1 ; quindi $(\mathbf{T} - \lambda_k \mathbf{I})\mathbf{w} = (\lambda_1 - \lambda_k)\mathbf{w}$ e

$$\mathbf{B}\mathbf{v}_1 = (\mathbf{T} - \lambda_2 \mathbf{I})^{e_2} \cdots (\mathbf{T} - \lambda_s \mathbf{I})^{e_s} \mathbf{w} = (\lambda_1 - \lambda_2)^{e_2} \cdots (\lambda_1 - \lambda_s)^{e_s} \mathbf{w}$$

Per ipotesi, gli autovalori λ_k sono distinti, per cui $(\lambda_1 - \lambda_2)^{e_2} \cdots (\lambda_1 - \lambda_s)^{e_s} \neq 0$. Siccome \mathbf{w} è non nullo, dall'uguaglianza precedente deduciamo che $\mathbf{B}\mathbf{v}_1 \neq \mathbf{0}$.

Per concludere moltiplichiamo l'uguaglianza $\mathbf{v}_1 + \cdots + \mathbf{v}_s = \mathbf{0}$ per \mathbf{B} . Otteniamo $\mathbf{B}\mathbf{v}_1 = \mathbf{0}$ perché $\mathbf{B}\mathbf{v}_k = \mathbf{0}$ per $k \geq 2$. Questo contraddice il fatto che $\mathbf{B}\mathbf{v}_1 \neq \mathbf{0}$ che abbiamo dimostrato prima a partire dall'ipotesi che \mathbf{v}_1 fosse non nullo. Quindi \mathbf{v}_1 dev'essere nullo. Possiamo infine scambiare λ_1 con λ_k e dedurre che $\mathbf{v}_k = \mathbf{0}$ per ogni $k = 2, \dots, s$.

PAG
5/32

TEOREMA 6.8 (Esistenza base di Jordan)

Sia \mathbf{V} un spazio vettoriale complesso di dimensione finita, e sia $\mathbf{T} : \mathbf{V} \rightarrow \mathbf{V}$ un operatore lineare. Esiste una base di \mathbf{V} che è una base di Jordan per \mathbf{T} . In particolare, esiste una base di \mathbf{V} formata da *autovettori generalizzati* di \mathbf{T} .

DIMOSTRAZIONE. La dimostrazione è per induzione su $n = \dim(\mathbf{V})$. Il caso iniziale $n = 1$ è immediato. Supponiamo quindi $n > 1$. Siccome \mathbf{V} è uno spazio complesso, l'operatore \mathbf{T} ha almeno un autovalore complesso λ e un relativo autovettore $\mathbf{w} \in \text{Ker}(\mathbf{T} - \lambda \mathbf{I})$. Per semplificare le notazioni, possiamo supporre $\lambda = 0$: per questo basta osservare che una base è di Jordan per \mathbf{T} se e solo se lo è per $\mathbf{T} - \lambda \mathbf{I}$ e quindi sostituire \mathbf{T} con $\mathbf{T} - \lambda \mathbf{I}$. Allora il nucleo $\text{Ker}(\mathbf{T})$ ha dimensione almeno uno, e per il teorema di nullità più rango l'immagine $\text{Im}(\mathbf{T})$ ha dimensione strettamente minore di n . Allo spazio $\text{Im}(\mathbf{T})$ possiamo quindi applicare la nostra ipotesi di induzione: ogni operatore di $\text{Im}(\mathbf{T})$ in se stesso ha una base di Jordan. Come operatore usiamo ancora \mathbf{T} : quando applichiamo \mathbf{T} a un vettore $\mathbf{T}\mathbf{v}$ dell'immagine otteniamo $\mathbf{T}(\mathbf{T}\mathbf{v})$ che è ancora un vettore dell'immagine, quindi \mathbf{T} è un operatore di $\text{Im}(\mathbf{T})$ in se stesso. Per l'ipotesi di induzione esiste una base di Jordan di $\text{Im}(\mathbf{T})$: supponiamo che questa base sia

$$\{\mathbf{w}_{\ell,1}, \dots, \mathbf{w}_{\ell,d_\ell} : \ell = 1, \dots, m\}$$

Questa base consiste di r vettori, dove r è la dimensione di $\text{Im}(\mathbf{T})$ cioè il rango di \mathbf{T} . Possiamo supporre, riordinando se necessario le catene che compongono la base, che le prime p catene siano formate da autovettori generalizzati relativi all'autovalore nullo, e le altre $m - p$ catene siano invece relative ad autovalori non nulli. Le prime p catene sono quindi della forma

$$\mathbf{w}_{\ell,1}, \dots, \mathbf{w}_{\ell,d_\ell} : \quad \mathbf{w}_{\ell,k} = \mathbf{T}^{d_\ell - k} \mathbf{w}_{\ell,d_\ell}$$

Scegliamo altri p vettori \mathbf{u}_ℓ in questo modo: per ogni $1 \leq \ell \leq p$, l'ultimo vettore della catena \mathbf{w}_{ℓ,d_ℓ} è per costruzione un vettore dell'immagine $\text{Im}(\mathbf{T})$. Quindi esiste \mathbf{u}_ℓ tale che $\mathbf{T}\mathbf{u}_\ell = \mathbf{w}_{\ell,d_\ell}$. Si osservi che aggiungendo il vettore \mathbf{u}_ℓ alla catena ℓ otteniamo una nuova catena

$$(6.11) \quad \mathbf{w}_{\ell,1}, \dots, \mathbf{w}_{\ell,d_\ell}, \mathbf{u}_\ell$$

di autovettori generalizzati, di lunghezza maggiore di uno rispetto a quella di partenza; infatti per costruzione:

$$\mathbf{w}_{\ell,k} = \mathbf{T}^{d_\ell - k} \mathbf{w}_{\ell,d_\ell} = \mathbf{T}^{d_\ell + 1 - k} \mathbf{u}_\ell$$

Per arrivare a ottenere una base di \mathbf{V} dobbiamo aggiungere ancora un insieme di vettori: per ogni $1 \leq \ell \leq p$, il primo vettore $\mathbf{w}_{\ell,1}$ della catena ℓ è un autovettore relativo a $\lambda = 0$, cioè è un vettore del nucleo. Questi vettori formano un insieme

$$\{\mathbf{w}_{1,1}, \mathbf{w}_{2,1}, \dots, \mathbf{w}_{p,1}\}$$

che è linearmente indipendente in quanto parte di una base. Per il teorema del completamento della base esistono dei vettori $\mathbf{z}_1, \dots, \mathbf{z}_q$ che aggiunti ai vettori $\{\mathbf{w}_{\ell,1} : 1 \leq \ell \leq p\}$ formano una base di $\text{Ker}(\mathbf{T})$. Per il teorema di nullità più rango $q = n - r - p$: infatti abbiamo costruito una base del nucleo formata da $p + q$ vettori, quindi

$$p + q = \dim \text{Ker}(\mathbf{T}) = n - r$$

Riuniamo ora in un unico insieme \mathcal{B} i vettori $\mathbf{w}_{\ell,i}$ che formano la base dell'immagine, i vettori \mathbf{u}_ℓ che completano le catene relative all'autovalore nullo e i vettori \mathbf{z}_j . Come abbiamo osservato i vettori $\mathbf{w}_{\ell,i}$ sono r , i vettori \mathbf{u}_ℓ sono p , e i vettori \mathbf{z}_j sono $n - r - p$, quindi \mathcal{B} contiene esattamente n vettori, il numero giusto per una base di \mathbf{V} . Inoltre l'insieme \mathcal{B} è formato da catene di autovettori generalizzati: le p catene (6.11) generate dai vettori \mathbf{u}_ℓ , le $m - p$ catene relative agli autovalori diversi da zero, e le catene \mathbf{z}_j che consistono di un unico autovettore: i vettori \mathbf{z}_j appartengono al nucleo, e quindi sono autovettori relativi a $\lambda = 0$, e formano ciascuno una catena di lunghezza 1. Quindi, se \mathcal{B} è una base di \mathbf{V} , è anche una base di Jordan per \mathbf{T} , e il teorema è dimostrato.

Per mostrare che \mathcal{B} è una base è sufficiente far vedere che è un insieme linearmente indipendente perché \mathcal{B} consiste di $n = \dim \mathbf{V}$ vettori. Supponiamo dunque che una combinazione lineare dei vettori di \mathcal{B} sia il vettore nullo:

$$(6.12) \quad \sum_{j=1}^{n-r-p} a_j \mathbf{z}_j + \sum_{\ell=1}^p \left(\sum_{i=1}^{d_\ell} b_{\ell,i} \mathbf{w}_{\ell,i} + b_\ell \mathbf{u}_\ell \right) + \sum_{\ell=p+1}^m \sum_{i=1}^{d_\ell} c_{\ell,i} \mathbf{w}_{\ell,i} = \mathbf{0}$$

Il vettore

$$\mathbf{v}_1 = \sum_{j=1}^{n-r-p} a_j \mathbf{z}_j + \sum_{\ell=1}^p \left(\sum_{i=1}^{d_\ell} b_{\ell,i} \mathbf{w}_{\ell,i} + b_\ell \mathbf{u}_\ell \right)$$

è una combinazione lineare di autovettori generalizzati relativa all'autovalore nullo, e appartiene perciò al sottospazio radicale \mathbf{R}_0 . Il vettore $\mathbf{w} = \sum_{\ell=p+1}^m c_{\ell,i} \mathbf{w}_{\ell,i}$ è invece una combinazione lineare di autovettori generalizzati relativi agli altri autovalori di \mathbf{T} ; raggruppando gli

autovettori generalizzati relativi a uno stesso autovalore possiamo scrivere $\mathbf{w} = \mathbf{v}_2 + \cdots + \mathbf{v}_s$ con $\mathbf{v}_k \in \mathbf{R}_{\lambda_k}$ e $\lambda_k \neq 0$. La (6.12) in queste notazioni diviene

$$(6.13) \quad \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_s = \mathbf{0}$$

Dal lemma 6.7 segue $\mathbf{v}_1 = \mathbf{v}_2 = \cdots = \mathbf{v}_s = \mathbf{0}$. In particolare

$$\sum_{\ell=p+1}^m \sum_{i=1}^{d_\ell} c_{\ell,i} \mathbf{w}_{\ell,i} = \mathbf{v}_2 + \cdots + \mathbf{v}_s = \mathbf{0}$$

Per costruzione i vettori $\mathbf{w}_{\ell,i}$ sono lincarmente indipendenti, e quindi i coefficienti $c_{\ell,i}$ sono tutti nulli. Adesso analizziamo l'altro termine:

$$(6.14) \quad \sum_{j=1}^{n-r-p} a_j \mathbf{z}_j + \sum_{\ell=1}^p \left(\sum_{i=1}^{d_\ell} b_{\ell,i} \mathbf{w}_{\ell,i} + b_\ell \mathbf{u}_\ell \right) = \mathbf{v}_1 = \mathbf{0}$$

Moltiplichiamo questa equazione a sinistra per \mathbf{T} . Siccome i vettori \mathbf{z}_j e i vettori $\mathbf{w}_{\ell,i}$ appartengono al nucleo di \mathbf{T} , otteniamo

$$(6.15) \quad \sum_{\ell=1}^p \left(\sum_{i=2}^{d_\ell} b_{\ell,i} \mathbf{T} \mathbf{w}_{\ell,i} + b_\ell \mathbf{T} \mathbf{u}_\ell \right) = \mathbf{0}$$

Ora $\mathbf{T} \mathbf{u}_\ell = \mathbf{w}_{\ell,d_\ell}$ e $\mathbf{T} \mathbf{w}_{\ell,i} = \mathbf{w}_{\ell,i-1}$ per $i = 2, \dots, d_\ell$, quindi:

$$(6.16) \quad \sum_{\ell=1}^p \left(\sum_{i=2}^{d_\ell} b_{\ell,i} \mathbf{w}_{\ell,i-1} + b_\ell \mathbf{w}_{\ell,d_\ell} \right) = \mathbf{0}$$

Usando il fatto che i vettori $\mathbf{w}_{\ell,i}$ sono linearmente indipendenti deduciamo $b_{\ell,i} = 0$ per ogni ℓ e ogni $i \geq 2$, e $b_\ell = 0$. Sostituendo in (6.17) otteniamo

$$(6.17) \quad \sum_{j=1}^{n-r-p} a_j \mathbf{z}_j + \sum_{\ell=1}^p b_{\ell,1} \mathbf{w}_{\ell,1} = \mathbf{0}$$

Per costruzione i vettori \mathbf{z}_j e $\mathbf{w}_{\ell,1}$ formano una base del nucleo di \mathbf{T} , e sono perciò linearmente indipendenti. Possiamo allora concludere che anche i coefficienti a_j e $b_{\ell,1}$ sono nulli. Raccogliamo le fila: abbiamo mostrato che tutti i coefficienti a membro di sinistra di (6.12) sono nulli. Questo significa che \mathcal{B} è un insieme linearmente indipendente, e la dimostrazione è così completata.

Dal teorema precedente dedurremo facilmente l'esistenza della forma canonica di Jordan; per l'unicità della forma canonica avremo bisogno di conoscere i ranghi delle potenze dei blocchi di Jordan.

LEMMA 6.9 Sia $\mathbf{J} = \mathbf{J}_m(0)$ il blocco di Jordan di ordine m con autovalore $\lambda = 0$. Il rango di \mathbf{J}^d è $m - d$ se $0 \leq d \leq m$ ed è 0 se $d \geq m$. In particolare:

$$(6.18) \quad r(\mathbf{J}^{d+1}) - 2r(\mathbf{J}^d) + r(\mathbf{J}^{d-1}) = \begin{cases} 1 & \text{se } d = m \\ 0 & \text{se } d \geq 1 \text{ e } d \neq m \end{cases} .$$

DIMOSTRAZIONE. Sia \mathbf{T} l'operatore lineare $\mathbb{C}^m \rightarrow \mathbb{C}^m$ rappresentato da \mathbf{J} . La base canonica è una catena di autovettori generalizzati di \mathbf{T} relativi all'autovalore nullo:

$$\mathbf{T}\mathbf{e}_1 = \mathbf{0} = 0\mathbf{e}_1, \quad \mathbf{T}\mathbf{e}_k = \mathbf{e}_{k-1} \quad \text{per } k = 2, \dots, m.$$

L'immagine di \mathbf{T} è generata dai vettori immagine dei vettori della base canonica \mathbf{e} , quindi, una base di $\text{Im}(\mathbf{T})$ è $\{\mathbf{e}_1, \dots, \mathbf{e}_{m-1}\}$. Questo mostra che:

$$r(\mathbf{J}) = r(\mathbf{T}) = \dim \text{Im}(\mathbf{T}) = m - 1$$

Calcoliamo ora l'effetto di \mathbf{T}^2 sui vettori della base canonica:

$$\mathbf{T}^2\mathbf{e}_1 = \mathbf{T}\mathbf{0} = \mathbf{0}, \quad \mathbf{T}^2\mathbf{e}_2 = \mathbf{T}\mathbf{e}_1 = \mathbf{0}, \quad \mathbf{T}^2\mathbf{e}_k = \mathbf{e}_{k-2} \quad \text{per } k = 3, \dots, m$$

Una base di $\text{Im}(\mathbf{T}^2)$ è quindi $\{\mathbf{e}_1, \dots, \mathbf{e}_{m-2}\}$, e $r(\mathbf{T}^2) = m - 2$. Proseguendo così si vede che, per $d < m$, una base dell'immagine di \mathbf{T}^d è $\{\mathbf{e}_1, \dots, \mathbf{e}_{m-d}\}$ e quindi il rango di \mathbf{T}^d è $m - d$. Per $d \geq m$ si ottiene $\mathbf{T}^d\mathbf{e}_k = \mathbf{0}$ per ogni k , e quindi \mathbf{T}^d è l'operatore nullo e il suo rango è zero. Siccome \mathbf{J}^d è la matrice che rappresenta \mathbf{T}^d rispetto alla base canonica, il rango di \mathbf{J}^d coincide con quello di \mathbf{T}^d ed è perciò il massimo tra $m - d$ e 0, come volevasi dimostrare.

Si osservi che quanto abbiamo dimostrato si traduce così in termini di matrici: se $d < m$, la matrice \mathbf{J}^d ha gli elementi di posto $(i, i+d)$ uguali a 1, e tutti gli altri elementi di \mathbf{J}^d sono nulli; per $d \geq m$ la matrice \mathbf{J}^d è la matrice nulla. Per esempio, per $m = 4$, abbiamo:

$$\mathbf{J} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{J}^2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{J}^3 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{J}^4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Rimane da mostrare la (6.18). Se $1 \leq d \leq m - 1$, allora

$$r(\mathbf{J}^{d+1}) - 2r(\mathbf{J}^d) + r(\mathbf{J}^{d-1}) = (m - d - 1) - 2(m - d) + (m - d + 1) = 0$$

Se invece $d \geq m + 1$,

$$r(\mathbf{J}^{d+1}) - 2r(\mathbf{J}^d) + r(\mathbf{J}^{d-1}) = 0 - 0 + 0 = 0$$

Se infine $d = m$,

$$r(\mathbf{J}^{m+1}) - 2r(\mathbf{J}^m) + r(\mathbf{J}^{m-1}) = 0 - 0 + 1 = 1$$

Quindi la (6.18) è verificata, e la dimostrazione è completa. L'idea alla base della (6.18) è questa: la differenza prima di una funzione $r(d)$ è $\delta r(d) = r(d) - r(d - 1)$; la differenza seconda è la differenza prima della differenza prima:

$$\delta^2 r(d) = \delta(r(d) - r(d - 1)) = r(d) - 2r(d - 1) + r(d - 2)$$

Considerare la differenza seconda è l'analogo discreto di considerare la derivata seconda di una funzione. La funzione che compare nella (6.18) è la differenza seconda di $r(d) = r(\mathbf{J}^d)$, calcolata in $d + 1$. Siccome $r(d) = m - d$ è lineare a sinistra di m , ed è costante a destra di m , la sua differenza seconda si annulla tranne che nel punto di raccordo della funzione lineare con quella costante.

Data una matrice \mathbf{A} quadrata di ordine n con autovalori distinti $\lambda_1, \dots, \lambda_s$, definiamo ora degli invarianti che consentono di calcolare a partire da \mathbf{A} la sua forma canonica. Poniamo

$$r_{k,d}(\mathbf{A}) = r((\mathbf{A} - \lambda_k \mathbf{I})^d)$$

per ogni $k = 1, \dots, s$ e ogni $d = 1, \dots, n$. I numeri $r_{k,d}$ estendono le nozioni di molteplicità geometrica e algebrica dell'autovalore λ_k : in effetti, per il teorema di nullità più rango, $g_{\lambda_k} = n - r_{k,1}$, vedremo in seguito che $a_{\lambda_k} = n - r_{k,n}$; quindi $r_{k,1}$ e $r_{k,n}$ sono equivalenti alla molteplicità algebrica e alla molteplicità geometrica dell'autovalore λ_k . Siccome $\text{Ker}(\mathbf{A} - \lambda_k \mathbf{I})^d \subseteq \text{Ker}(\mathbf{A} - \lambda_k \mathbf{I})^{d+1}$, dal teorema di nullità più rango si ottiene $r_{k,d} \geq r_{k,d+1}$: questa disegualanza generalizza $g_\lambda \leq a_\lambda$.

TEOREMA 6.10 (Criterio di similitudine e forma canonica)

Siano \mathbf{A} e \mathbf{B} matrici quadrate di ordine n .

- i) Le matrici \mathbf{A} e \mathbf{B} sono simili sul campo dei numeri complessi se e solo se hanno gli stessi autovalori distinti $\lambda_1, \dots, \lambda_s$ e

$$r((\mathbf{A} - \lambda_k \mathbf{I})^d) = r((\mathbf{B} - \lambda_k \mathbf{I})^d)$$

per ogni $k = 1, \dots, s$ e ogni $d = 1, \dots, n$.

- ii) se \mathbf{A} ha autovalori distinti $\lambda_1, \dots, \lambda_s$ e invarianti $r_{k,d} = r((\mathbf{A} - \lambda_k \mathbf{I})^d)$, allora \mathbf{A} è simile alla matrice di Jordan:

$$(6.19) \quad \mathbf{C} = \begin{bmatrix} \mathbf{J}_1 & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \mathbf{J}_2 & \cdots & \mathbf{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O} & \mathbf{O} & \cdots & \mathbf{J}_p \end{bmatrix}$$

i cui blocchi di Jordan sono del tipo $\mathbf{J}_m(\lambda_k)$, e un blocco siffatto compare esattamente $r_{k,m+1} - 2r_{k,m} + r_{k,m-1}$ volte in \mathbf{C} . La matrice \mathbf{C} , che è unica a meno di un riordinamento dei blocchi diagonali, si dice *forma canonica di Jordan* di \mathbf{A} .

DIMOSTRAZIONE. Si ricordi che le due matrici sono simili su \mathbb{C} se esiste una matrice invertibile \mathbf{S} a elementi complessi tale che $\mathbf{S}^{-1}\mathbf{AS} = \mathbf{B}$. Questo equivale a dire che l'operatore lineare $\mathbf{T} = \mathcal{L}_{\mathbf{A}}$ è rappresentato da \mathbf{A} rispetto alla base canonica di \mathbb{C}^n e da \mathbf{B} rispetto alla base formata dalle colonne di \mathbf{S} . Da $\mathbf{S}^{-1}\mathbf{AS} = \mathbf{B}$ segue $\mathbf{S}^{-1}(\mathbf{A} - \lambda_k \mathbf{I})^d \mathbf{S} = (\mathbf{B} - \lambda_k \mathbf{I})^d$, quindi le due matrici $(\mathbf{A} - \lambda_k \mathbf{I})^d$ e $(\mathbf{B} - \lambda_k \mathbf{I})^d$ sono simili e hanno perciò lo stesso rango (intrinsecamente, le due matrici rappresentano entrambe l'operatore $(\mathbf{T} - \lambda_k \mathbf{I})^d$ e il loro rango coincide con la dimensione dell'immagine di questo operatore). Questo mostra che gli invarianti $r_{k,d}$ di \mathbf{A} e \mathbf{B} sono uguali se \mathbf{A} e \mathbf{B} simili, cioè il *solo se* del primo enunciato.

Mostriamo ora il secondo enunciato. Sia $\mathbf{T} = \mathcal{L}_{\mathbf{A}} : \mathbb{C}^n \rightarrow \mathbb{C}^n$ l'operatore lineare rappresentato dalla matrice \mathbf{A} rispetto alla base canonica (questo significa semplicemente che $\mathbf{T}\mathbf{v} = \mathbf{Av}$ per ogni $\mathbf{v} \in \mathbb{C}^n$). Per il teorema 6.8 esiste una base \mathcal{B} di \mathbb{C}^n che è di Jordan per \mathbf{T} , e per la proposizione 6.4 la matrice \mathbf{C} che rappresenta \mathbf{T} rispetto alla base \mathcal{B} è una matrice di Jordan; \mathbf{C} è simile ad \mathbf{A} perché rappresentano entrambe l'operatore \mathbf{T} ; esplicitamente, se \mathbf{S} è la matrice che ha i vettori della base \mathcal{B} come colonne, allora $\mathbf{C} = \mathbf{S}^{-1}\mathbf{AS}$.

Per quanto abbiamo già dimostrato, gli autovalori e gli invarianti $r_{k,d}$ di \mathbf{C} sono uguali a quelli di \mathbf{A} . Vediamo ora come questi invarianti determinino i blocchi di Jordan di \mathbf{C} . Per questo osserviamo che il rango di una matrice diagonale a blocchi $\text{diag}(\mathbf{J}_1, \dots, \mathbf{J}_p)$ è la

somma dei ranghi dei suoi blocchi \mathbf{J}_h , e che per ogni $d \geq 0$

$$(\text{diag}(\mathbf{J}_1, \dots, \mathbf{J}_p))^d = \text{diag}(\mathbf{J}_1^d, \dots, \mathbf{J}_p^d)$$

Da queste due osservazioni segue immediatamente che

$$r_{k,d}(\mathbf{A}) = r_{k,d}(\mathbf{C}) = \sum_{h=1}^p r_{k,d}(\mathbf{J}_h)$$

e quindi, per ogni $m \geq 1$,

$$(6.20) \quad r_{k,m+1}(\mathbf{A}) - 2r_{k,m}(\mathbf{A}) + r_{k,m-1}(\mathbf{A}) = \sum_{h=1}^p (r_{k,m+1}(\mathbf{J}_h) - 2r_{k,m}(\mathbf{J}_h) + r_{k,m-1}(\mathbf{J}_h))$$

Vogliamo ora determinare i blocchi di Jordan di \mathbf{C} . Innanzitutto, se $\mathbf{J}_m(\lambda)$ è uno dei blocchi di \mathbf{C} , dev'essere $m \leq n$ perché \mathbf{C} è $n \times n$; inoltre λ dev'essere uno degli autovalori di \mathbf{C} perché \mathbf{C} è triangolare alta e λ è uno degli elementi della diagonale principale di \mathbf{C} . Siccome \mathbf{C} ha gli stessi autovalori di \mathbf{A} , λ è uno degli autovalori λ_k di \mathbf{A} . Per calcolare il numero di blocchi di Jordan di \mathbf{C} che sono uguali a $\mathbf{J}_m(\lambda_k)$, determiniamo il contributo di ogni blocco \mathbf{J}_h di \mathbf{C} nella (6.20). Fissato h , sia λ l'autovalore del blocco \mathbf{J}_h . Se $\lambda \neq \lambda_k$, λ_k non è un autovalore di \mathbf{J}_h , e quindi la matrice

$$\mathbf{J}_h - \lambda_k \mathbf{I}_d$$

è invertibile e ha perciò rango d , così come tutte le sue potenze. Quindi

$$r_{k,m+1}(\mathbf{J}_h) - 2r_{k,m}(\mathbf{J}_h) + r_{k,m-1}(\mathbf{J}_h) = 0$$

e il contributo di \mathbf{J}_h nella (6.20) è nullo. Se invece $\lambda = \lambda_k$ e \mathbf{J}_h ha ordine m_h , la matrice $\mathbf{J}_h - \lambda_k \mathbf{I}_{m_h}$ è un blocco di Jordan $\mathbf{J}_{m_h}(0)$ di ordine m_h e autovalore nullo. Per il lemma 6.9 il contributo di \mathbf{J}_h nella (6.20) è 0 se $m_h \neq m$ ed è 1 se $m_h = m$. In conclusione, il contributo di \mathbf{J}_h nella (6.20) è 1 se $\mathbf{J}_h = \mathbf{J}_m(\lambda_k)$ ed è zero altrimenti. Quindi il numero di blocchi di Jordan di \mathbf{C} che sono uguali a $\mathbf{J}_m(\lambda_k)$ è $r_{k,m+1}(\mathbf{A}) - 2r_{k,m}(\mathbf{A}) + r_{k,m-1}(\mathbf{A})$ come volevasi dimostrare. Per quanto riguarda l'unicità, se \mathbf{C}' è un'altra matrice di Jordan simile a \mathbf{A} , per quanto abbiamo appena dimostrato \mathbf{C}' ha gli stessi blocchi di \mathbf{C} .

Rimane da dimostrare che due matrici \mathbf{A} e \mathbf{B} con gli stessi autovalori e gli stessi invarianti $r_{k,d}$ sono simili. Per ii) entrambe le matrici sono simili alla stessa matrice di Jordan \mathbf{C} . Siccome la similitudine è una relazione di equivalenza, \mathbf{A} e \mathbf{B} sono simili.

OSSERVAZIONE In breve il teorema 6.10 dice che due matrici sono simili se e solo se hanno la stessa forma canonica di Jordan, a meno dell'ordine dei blocchi \mathbf{J}_h . Una matrice \mathbf{C} di Jordan è bidiagonale e triangolare alta. Quindi ogni matrice è simile, sul campo complesso, a una matrice bidiagonale triangolare alta.

OSSERVAZIONE Se gli autovalori di \mathbf{A} sono tutti reali, la forma canonica di Jordan \mathbf{C} è reale, e anche la matrice di passaggio \mathbf{S} può essere scelta a elementi reali: se gli autovalori sono reali, nella dimostrazione si può sostituire \mathbb{C} con \mathbb{R} .

Per un k fissato, gli invarianti $r_{k,d}$ estendono le nozioni di molteplicità algebrica e geometrica dell'autovalore λ_k :

COROLLARIO 6.11 Consideriamo una matrice quadrata \mathbf{A} di ordine n , un autovalore λ di \mathbf{A} , e l'indice $e = e_\lambda$ dell'autovalore λ . La successione $r_d = r((\mathbf{A} - \lambda\mathbf{I})^d)$ è strettamente decrescente tra $d = 0$ e $d = e$ ed è poi costante:

$$(6.21) \quad n = r_0 > r_1 > \cdots > r_e = r_d \quad \text{per ogni } d \geq e$$

La molteplicità geometrica di λ è $n - r_1$, la sua molteplicità algebrica è $n - r_e$. L'indice e è il massimo ordine di un blocco di Jordan $\mathbf{J}_m(\lambda)$ nella forma canonica di Jordan di \mathbf{A} .

DIMOSTRAZIONE. Abbiamo definito l'indice e dell'autovalore λ nella proposizione 6.6 quando abbiamo mostrato che:

$$(6.22) \quad \begin{aligned} \text{Ker}(\mathbf{A} - \lambda\mathbf{I})^d &\subsetneq \text{Ker}(\mathbf{A} - \lambda\mathbf{I})^{d+1} && \text{se } 0 \leq d < e \\ \text{Ker}(\mathbf{A} - \lambda\mathbf{I})^d &= \text{Ker}(\mathbf{A} - \lambda\mathbf{I})^{d+1} && \text{se } d \geq e \end{aligned}$$

Per il teorema di nullità più rango $\dim \text{Ker}(\mathbf{A} - \lambda\mathbf{I})^d = n - r_d$, quindi la (6.21) segue dalla (6.22). La molteplicità geometrica è la dimensione dell'autospazio $\text{Ker}(\mathbf{A} - \lambda\mathbf{I})$ e quindi è uguale a $n - r_1$.

Sia \mathbf{C} la forma canonica di Jordan di una matrice che rappresenta \mathbf{A} . La molteplicità algebrica a_λ di λ è uguale al numero di volte che λ compare sulla diagonale principale di \mathbf{C} . Se q_m è il numero di blocchi $\mathbf{J}_m(\lambda)$ che compaiono in \mathbf{C} , il numero di elementi sulla diagonale principale di \mathbf{C} uguali a λ è $\sum_{m=1}^n mq_m$. Per il teorema 6.10

$$(6.23) \quad \sum_{m=1}^n mq_m = \sum_{m=1}^n m(r_{m+1} - 2r_m + r_{m-1})$$

Fissiamo un intero d con $2 \leq d \leq n - 1$. Nella somma (6.23) il termine r_d compare con coefficiente $d - 1$ quando $m + 1 = d$, con coefficiente $-2d$ quando $m = d$, e con coefficiente $d + 1$ quando $m - 1 = d$. La somma di questi coefficienti è zero. Il termine r_1 compare solo quando $m = 1$, con coefficiente -2 , e quando $m - 1 = 1$, con coefficiente 2 . Quindi gli unici termini che sopravvivono a secondo membro della (6.23) sono r_{n+1} , r_n e r_0 , e calcolando i loro coefficienti si ottiene

$$(6.24) \quad a_\lambda = nr_{n+1} - (n + 1)r_n + r_0$$

Siccome $e \leq n$, abbiamo $r_{n+1} = r_n = r_e$, e $r_0 = n$ perché per definizione $(\mathbf{A} - \lambda\mathbf{I})^0 = \mathbf{I}$. Sostituendo nella (6.24) troviamo

$$(6.25) \quad a_\lambda = n - r_e$$

Infine mostriamo che e è il massimo ordine di un blocco di Jordan $\mathbf{J}_m(\lambda)$ nella forma canonica \mathbf{C} . Se $m > e$, allora $r_{m+1} = r_m = r_{m-1} = r_e$, e quindi per il teorema 6.10 la forma canonica non ha blocchi della forma $\mathbf{J}_m(\lambda)$. Invece

$$(6.26) \quad r_{e+1} - 2r_e + r_{e-1} = r_{e-1} - r_e > 0$$

e quindi, sempre il teorema 6.10, \mathbf{C} ha almeno un blocco della forma $\mathbf{J}_e(\lambda)$.



COROLLARIO 6.12 (Significato geometrico molteplicità algebrica)

Sia \mathbf{A} una matrice quadrata e sia λ un autovalore di \mathbf{A} . La molteplicità algebrica di λ coincide con la dimensione del sottospazio radicale \mathbf{R}_λ relativo a λ . In particolare, l'indice e_λ dell'autovalore λ è minore o uguale alla molteplicità algebrica.

DIMOSTRAZIONE. Sia e l'indice dell'autovalore λ . Per la proposizione 6.6 il sottospazio radicale \mathbf{R}_λ è uguale a $\text{Ker}(\mathbf{A} - \lambda \mathbf{I})^e$ e, quindi, ha dimensione $n - r((\mathbf{A} - \lambda \mathbf{I})^e) = n - r_e$. Per il corollario 6.11 $n - r_e$ è la molteplicità algebrica dell'autovalore ed è maggiore o uguale a e perché la successione $n - r_d$ è strettamente crescente tra $d = 0$ e $d = e$.

OSSERVAZIONE Sia \mathbf{C} una matrice di Jordan. La molteplicità geometrica di un autovalore λ è il numero di blocchi di Jordan relativi a λ che compaiono in \mathbf{C} . Nelle notazioni della dimostrazione di 6.11, questo numero è $\sum_m q_m$, mentre la molteplicità algebrica è $\sum_m mq_m$: un blocco di ordine m relativo a λ contribuisce un autovettore vero e m autovettori generalizzati a una base di Jordan.

Dal corollario 6.11 segue che, se $n = 2$ o $n = 3$, due matrici che abbiano gli stessi autovalori con le stesse molteplicità algebriche e geometriche hanno tutti gli invarianti $r_{k,d}$ uguali, e quindi sono simili sul campo complesso.

Sia \mathbf{A} una matrice 3×3 con due autovalori distinti λ_1 e λ_2 , e supponiamo che λ_1 abbia molteplicità geometrica 1 e molteplicità algebrica 2. Allora la forma canonica di \mathbf{A} è

$$\mathbf{C} = \left[\begin{array}{cc|c} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ \hline 0 & 0 & \lambda_2 \end{array} \right]$$

Infatti dal corollario 6.11 segue che

$$r_{1,0} = 3 > r_{1,1} = 3 - g_{\lambda_1} = 2 > r_{1,2} = 3 - a_{\lambda_1} = 1 = r_{1,d} \quad \text{per ogni } d \geq 2$$

Quindi \mathbf{C} ha un unico blocco, di ordine $m = 2$, con autovalore λ_1 . L'altro blocco dev'essere di ordine 1 relativo all'altro autovalore.

Se $n \geq 4$ non basta conoscere gli autovalori e le loro molteplicità geometriche e algebriche per distinguere la classe di similitudine di una matrice. Per esempio consideriamo le due matrici

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{e} \quad \mathbf{B} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Entrambe le matrici sono matrici di Jordan, e hanno un unico autovalore $\lambda_1 = 0$ con molteplicità algebrica 4. La molteplicità geometrica di $\lambda_1 = 0$ è il numero di blocchi di Jordan, che è 2 in entrambi i casi. Infatti, la matrice \mathbf{A} ha due blocchi di Jordan di ordine due $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, mentre la matrice \mathbf{B} ha un blocco di ordine tre $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ e un blocco di ordine uno $[0]$. Tuttavia \mathbf{A} e \mathbf{B} non sono simili perché $r_{1,2}(\mathbf{A}) = r(\mathbf{A}^2) \neq r(\mathbf{B}^2) = r_{1,2}(\mathbf{B})$; oppure perché i blocchi di Jordan delle due matrici sono diversi.

Una matrice quadrata \mathbf{A} è sempre simile alla sua trasposta. Infatti \mathbf{A} e \mathbf{A}^T hanno lo stesso polinomio caratteristico, e quindi gli stessi autovalori distinti. Inoltre:

$$r((\mathbf{A} - \lambda_k \mathbf{I})^d) = r((\mathbf{A}^T - \lambda_k \mathbf{I})^d)$$

perché

$$(\mathbf{A}^T - \lambda_k \mathbf{I})^d = ((\mathbf{A} - \lambda_k \mathbf{I})^d)^T$$

Criteri di diagonalizzabilità

Possiamo ora riassumere i vari criteri che abbiamo a disposizione per stabilire se una matrice è diagonalizzabile:

TEOREMA 6.13 (Criteri di diagonalizzabilità)

Per una matrice quadrata \mathbf{A} le seguenti condizioni sono equivalenti:

- 1) \mathbf{A} è diagonalizzabile su \mathbb{C} ;
- 2) la forma canonica di Jordan di \mathbf{A} è una matrice diagonale;
- 3) gli autovalori di \mathbf{A} hanno tutti indice 1;

$$\text{Ker}((\mathbf{A} - \lambda \mathbf{I})) = \text{Ker}((\mathbf{A} - \lambda \mathbf{I})^2) \quad \text{per ogni autovalore } \lambda;$$

- 4) ogni autovalore di \mathbf{A} è regolare;
- 5) ogni autovettore generalizzato di \mathbf{A} è un vero autovettore.

DIMOSTRAZIONE. Se \mathbf{A} è diagonalizzabile, è simile a una matrice diagonale \mathbf{D} . Per l'unicità della forma canonica, la matrice di Jordan \mathbf{D} è la forma canonica di \mathbf{A} . Quindi 1) implica 2). Per il corollario 6.11 la forma canonica è diagonale, cioè non ha blocchi di ordine $m \geq 2$, se e solo se l'indice di ogni autovalore di \mathbf{A} è 1. Quindi 2) è equivalente a 3). Sempre per il corollario 6.11, l'indice di un autovalore è 1 se e solo se le molteplicità algebrica e geometrica dell'autovalore coincidono. Quindi 3) e 4) sono equivalenti. Se ogni autovalore ha indice 1, il sottospazio radicale \mathbf{R}_λ coincide con l'autospazio $\text{Ker}(\mathbf{A} - \lambda \mathbf{I})$ per ogni λ , e questo significa che ogni autovettore generalizzato è un autovettore vero. Quindi 4) implica 5). Infine, supponiamo che ogni autovettore generalizzato sia un autovettore vero. Per il teorema 6.8 possiamo trovare una base di \mathbb{C}^n formata da autovettori generalizzati, e quindi

di autovettori veri, di \mathbf{A} ; la matrice è perciò diagonalizzabile. Questo mostra che 5) implica 1) e conclude la dimostrazione.

Somme dirette

La nozione di somma diretta di sottospazi di \mathbf{V} chiarifica parte di quanto abbiamo dimostrato in questo paragrafo.

DEFINIZIONE 6.14 (Somma diretta di sottospazi)

Siano $\mathbf{V}_1, \dots, \mathbf{V}_s$ sottospazi di \mathbf{V} . Si dice che \mathbf{V} è la *somma diretta* dei sottospazi $\mathbf{V}_1, \dots, \mathbf{V}_s$, se ogni vettore $\mathbf{v} \in \mathbf{V}$ si scrive in uno e un solo modo nella forma

$$\mathbf{v} = \mathbf{v}_1 + \cdots + \mathbf{v}_s$$

con $\mathbf{v}_k \in \mathbf{V}_k$ per ogni $k = 1, \dots, s$. In tal caso si scrive:

$$\mathbf{V}_1 \oplus \cdots \oplus \mathbf{V}_s$$

Nel caso $s = 2$, $\mathbf{V} = \mathbf{V}_1 \oplus \mathbf{V}_2$ equivale alle due condizioni che $\mathbf{V} = \mathbf{V}_1 + \mathbf{V}_2$ e che $\mathbf{V}_1 \cap \mathbf{V}_2 = \{\mathbf{0}\}$ (cf. il capitolo sugli spazi vettoriali). In generale si dimostra senza difficoltà che:

PROPOSIZIONE 6.15 Uno spazio vettoriale \mathbf{V} è somma diretta dei suoi sottospazi $\mathbf{V}_1, \dots, \mathbf{V}_s$ se e solo se le due seguenti condizioni sono soddisfatte:

- a) per ogni vettore $\mathbf{v} \in \mathbf{V}$ esistono vettori $\mathbf{v}_k \in \mathbf{V}_k$ per $k = 1, \dots, s$ tali che

$$\mathbf{v} = \mathbf{v}_1 + \cdots + \mathbf{v}_s;$$
- b) se $\mathbf{v}_k \in \mathbf{V}_k$ per $k = 1, \dots, s$ e $\mathbf{v}_1 + \cdots + \mathbf{v}_s = \mathbf{0}$, allora $\mathbf{v}_k = \mathbf{0}$ per ogni $k = 1, \dots, s$.

OSSERVAZIONE La condizione a) nella proposizione 6.15 significa che \mathbf{V} è la somma dei suoi sottospazi \mathbf{V}_k ; la condizione b) è l'analogo per un insieme di sottospazi della condizione di indipendenza lineare per un insieme di vettori.

OSSERVAZIONE Uno spazio finito dimensionale \mathbf{V} è somma diretta dei suoi sottospazi $\mathbf{V}_1, \dots, \mathbf{V}_s$ se e solo se accostando basi di $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_s$ si ottiene una base di \mathbf{V} : fissate basi $\{\mathbf{b}_1^k, \dots, \mathbf{b}_{n_k}^k\}$ di \mathbf{V}_k per ogni k , l'insieme

$$\{\mathbf{b}_1^1, \dots, \mathbf{b}_{n_1}^1, \mathbf{b}_1^2, \dots, \mathbf{b}_{n_2}^2, \dots, \mathbf{b}_1^s, \dots, \mathbf{b}_{n_s}^s\}$$

è una base di \mathbf{V} . La dimostrazione di questo fatto è immediata e si trova nel capitolo sugli spazi vettoriali nel caso $s = 2$; viene perciò lasciata al lettore. In particolare

$$\dim(\mathbf{V}_1 \oplus \mathbf{V}_2 \oplus \cdots \oplus \mathbf{V}_s) = \sum_{k=1}^s \dim \mathbf{V}_k$$

TEOREMA 6.16 (Decomposizione in sottospazi radicali)

Sia \mathbf{A} una matrice quadrata di ordine n , e siano $\lambda_1, \dots, \lambda_s$ gli autovalori distinti di \mathbf{A} . Allora \mathbb{C}^n è somma diretta dei sottospazi radicali di \mathbf{A} :

$$\mathbb{C}^n = \mathbf{R}_{\lambda_1} \oplus \mathbf{R}_{\lambda_2} \oplus \cdots \oplus \mathbf{R}_{\lambda_s}$$

DIMOSTRAZIONE. Per il teorema 6.8 esiste una base di \mathbb{C}^n formata da autovettori generalizzati di \mathbf{A} . Perciò ogni vettore di \mathbb{C}^n è somma di autovettori generalizzati. Raggruppando gli autovettori generalizzati relativi a uno stesso autovalore si vede che \mathbb{C}^n è la somma dei sottospazi radicali. La somma è diretta per il lemma 6.7.

COROLLARIO 6.17 Una matrice quadrata \mathbf{A} di ordine n è diagonalizzabile su \mathbb{C} se e solo se \mathbb{C}^n è somma diretta degli autospazi di \mathbf{A} .

DIMOSTRAZIONE. Una matrice è diagonalizzabile se e solo se i suoi sottospazi radicali coincidono con gli autospazi.

TEOREMA 6.18 (Teorema di Hamilton-Cayley)

Ogni matrice quadrata \mathbf{A} è radice del proprio polinomio caratteristico: se $P(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I})$, allora $P(\mathbf{A})$ è la matrice nulla.

DIMOSTRAZIONE. Per il teorema 6.16 ogni vettore \mathbf{v} di \mathbb{C}^n si scrive nella forma

$$(6.27) \quad \mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_s$$

con $\mathbf{v}_k \in \mathbf{R}_{\lambda_k}$. Sappiamo che $\mathbf{R}_{\lambda_k} = \text{Ker}(\mathbf{A} - \lambda_k \mathbf{I})^{e_k}$ dove $e_k = e_{\lambda_k}$ è l'indice dell'autovalore λ_k . Per il corollario 6.11 la molteplicità algebrica a_k dell'autovalore λ_k è maggiore o uguale all'indice e_k . Quindi $(\mathbf{A} - \lambda_k \mathbf{I})^{a_k} \mathbf{v}_k = \mathbf{0}$. Ma allora moltiplicando la (6.27) a sinistra per la matrice

$$P(\mathbf{A}) = \pm (\mathbf{A} - \lambda_1 \mathbf{I})^{a_1} (\mathbf{A} - \lambda_2 \mathbf{I})^{a_2} \cdots (\mathbf{A} - \lambda_s \mathbf{I})^{a_s}$$

troviamo $P(\mathbf{A})\mathbf{v} = \mathbf{0}$. Siccome questo vale per ogni $\mathbf{v} \in \mathbb{C}^n$, $P(\mathbf{A})$ è la matrice nulla.

OSSERVAZIONE La dimostrazione precedente mostra anche che $Q(\mathbf{A}) = \mathbf{O}$ se $Q(\lambda) = \prod_{k=1}^s (\lambda - \lambda_k)^{e_k}$. Questo risultato è più fine del teorema di Hamilton-Cayley perché $Q(\lambda)$ divide il polinomio caratteristico (questo perché l'indice e_k è minore o uguale alla molteplicità algebrica a_k). Il polinomio $Q(\lambda)$ si dice *polinomio minimo* di \mathbf{A} : si può dimostrare che un polinomio $P_1(\lambda)$ soddisfa $P_1(\mathbf{A}) = \mathbf{O}$ se e solo se $Q(\lambda)$ divide $P_1(\lambda)$. Si osservi che una matrice è diagonalizzabile se e solo se il suo polinomio minimo è un prodotto di fattori lineari *distinti* (cioè tutti gli indici e_k sono uguali a 1).

Matrici nilpotenti e decomposizione di Jordan

DEFINIZIONE 6.19 (Matrice nilpotente)

Una matrice quadrata \mathbf{N} si dice *nilpotente* se esiste un intero $d \geq 1$ tale che \mathbf{N}^d è la matrice nulla.

Le matrici

$$\mathbf{N}_1 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{e} \quad \mathbf{N}_2 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

sono nilpotenti perché $\mathbf{N}_1^2 = \mathbf{N}_2^2 = \mathbf{O}$: la loro somma non è una matrice nilpotente.

PROPOSIZIONE 6.20 Una matrice quadrata \mathbf{N} di ordine n è nilpotente se e solo se è simile a una matrice triangolare alta che abbia tutti gli elementi sulla diagonale principale uguali a zero. Se \mathbf{N} è nilpotente, il suo polinomio caratteristico è $(-1)^n \lambda^n$ e $\mathbf{N}^n = \mathbf{O}$.

DIMOSTRAZIONE. Supponiamo che \mathbf{N} sia nilpotente, e sia $d \geq 1$ un intero tale che $\mathbf{N}^d = \mathbf{O}$. Se λ è un autovalore di \mathbf{N} , allora λ^d è un autovalore di $\mathbf{N}^d = \mathbf{O}$, e quindi $\lambda^d = 0$. Perciò $\lambda = 0$: una matrice nilpotente ha solo l'autovalore nullo. In particolare il polinomio caratteristico di \mathbf{N} è $(-1)^n \lambda^n$ e per il teorema di Hamilton-Cayley $\mathbf{N}^n = \mathbf{O}$. Siccome gli autovalori di \mathbf{N} sono tutti nulli, la forma canonica di Jordan \mathbf{C} di \mathbf{N} è una matrice triangolare alta con tutti gli elementi sulla diagonale principale uguali a zero.

Viceversa, una matrice triangolare alta \mathbf{U} con tutti gli elementi sulla diagonale principale uguali a zero ha polinomio caratteristico $(-1)^n \lambda^n$, e per il teorema di Hamilton-Cayley è nilpotente: $\mathbf{U}^n = \mathbf{O}$. Se $\mathbf{N} = \mathbf{S}^{-1} \mathbf{U} \mathbf{S}$ è simile a \mathbf{U} , allora $\mathbf{N}^n = \mathbf{S}^{-1} \mathbf{U}^n \mathbf{S} = \mathbf{O}$, e quindi \mathbf{N} è nilpotente.

In molte applicazioni, è utile osservare che la forma canonica di Jordan permette di scrivere una matrice come somma di una matrice diagonalizzabile e di una nilpotente che commutano tra loro (si può anche mostrare che tale decomposizione è unica):

TEOREMA 6.21 (Decomposizione di Jordan)

Sia \mathbf{A} una matrice quadrata. Allora \mathbf{A} si può scrivere nella forma

$$\mathbf{A} = \mathbf{B} + \mathbf{N}$$

dove \mathbf{B} è diagonalizzabile, \mathbf{N} è nilpotente, e $\mathbf{BN} = \mathbf{NB}$.

DIMOSTRAZIONE. La matrice \mathbf{A} è simile alla sua forma canonica di Jordan \mathbf{C} : esiste una matrice invertibile complessa \mathbf{S} tale che $\mathbf{S}^{-1} \mathbf{CS} = \mathbf{A}$. Scriviamo

$$\mathbf{C} = \mathbf{D} + \mathbf{U}$$

dove \mathbf{D} è la matrice diagonale con gli autovalori di \mathbf{C} sulla diagonale principale: allora gli unici elementi non nulli di \mathbf{U} sono sulla diagonale immediatamente sopra alla diagonale principale, e \mathbf{U} è nilpotente per la proposizione 6.20. Inoltre \mathbf{D} e \mathbf{U} commutano: basta verificare questo nel caso $\mathbf{C} = \mathbf{J}$ sia un blocco di Jordan; e in questo caso $\mathbf{D} = \lambda \mathbf{I}$ per cui $\mathbf{DU} = \lambda \mathbf{U} = \mathbf{UD}$. Poniamo $\mathbf{B} = \mathbf{S}^{-1}\mathbf{DS}$ e $\mathbf{N} = \mathbf{S}^{-1}\mathbf{US}$. Per costruzione \mathbf{B} è diagonalizzabile, \mathbf{N} è nilpotente, e

$$\mathbf{BN} = \mathbf{S}^{-1}\mathbf{DUS} = \mathbf{S}^{-1}\mathbf{UDS} = \mathbf{NB}$$

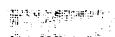
Diagonalizzazione simultanea di matrici che commutano

Vogliamo ora dimostrare un fatto che è importante in meccanica quantistica e in altre applicazioni: se \mathbf{A} e \mathbf{B} sono due matrici diagonalizzabili che commutano, allora \mathbf{A} e \mathbf{B} sono simultaneamente diagonalizzabili, cioè esiste una matrice invertibile \mathbf{S} tale che entrambe le matrici $\mathbf{S}^{-1}\mathbf{AS}$ e $\mathbf{S}^{-1}\mathbf{BS}$ siano diagonali. Per questo introduciamo la nozione di sottospazio \mathbf{T} -invariante, che generalizza ai sottospazi la nozione di autovettore, e che è utile in altri contesti.

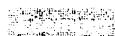
DEFINIZIONE 6.22 (Sottospazio invariante)

Sia $\mathbf{T} : \mathbf{V} \rightarrow \mathbf{V}$ un operatore lineare. Si dice che $\mathbf{H} \subset \mathbf{V}$ è un sottospazio \mathbf{T} -invariante se \mathbf{H} è un sottospazio e

$$\mathbf{Tv} \in \mathbf{H} \quad \text{per ogni } \mathbf{v} \in \mathbf{H}$$



Per ogni operatore \mathbf{T} , l'intero spazio \mathbf{V} e il sottospazio nullo $\{\mathbf{0}\}$ sono \mathbf{T} -invarianti. Sottospazi \mathbf{T} -invarianti più interessanti sono $\text{Ker}(\mathbf{T})$ e $\text{Im}(\mathbf{T})$.



Supponiamo che \mathbf{v} sia un autovettore di \mathbf{T} relativo all'autovalore λ . Sia

$$\mathbf{L} = \mathcal{L}(\mathbf{v}) = \{\mathbf{u} \in \mathbf{V} : \text{ esiste } t \in \mathbb{C} \text{ tale che } \mathbf{u} = t\mathbf{v}\}$$

il sottospazio generato da \mathbf{v} . Allora \mathbf{L} è un sottospazio \mathbf{T} -invariante di dimensione 1. Infatti \mathbf{v} è non nullo perché è un autovettore, e quindi genera un sottospazio di dimensione 1; inoltre se $\mathbf{u} = t\mathbf{v} \in \mathbf{L}$, allora anche $\mathbf{Tu} \in \mathbf{L}$ perché

$$\mathbf{Tu} = \mathbf{Ttv} = t\mathbf{Tv} = (t\lambda)\mathbf{v} \in \mathbf{L}$$

Viceversa, supponiamo che \mathbf{L} sia un sottospazio \mathbf{T} -invariante di dimensione 1 e che \mathbf{v} sia un generatore di \mathbf{L} . Allora $\mathbf{Tv} \in \mathbf{L}$ perché \mathbf{L} è \mathbf{T} -invariante e, quindi, esiste $\lambda \in \mathbb{C}$ tale che $\mathbf{Tv} = \lambda\mathbf{v}$ perché $\mathbf{L} = \mathcal{L}(\mathbf{v})$. Questo mostra che \mathbf{v} è un autovettore.

In conclusione, i sottospazi \mathbf{T} -invarianti di dimensione 1 sono i sottospazi di \mathbf{V} che sono generati da un autovettore di \mathbf{T} .

Sia λ un autovalore di \mathbf{T} . L'autospazio

$$\mathbf{V}_\lambda = \text{Ker}(\mathbf{T} - \lambda \mathbf{I}) = \{\mathbf{v} \in \mathbf{V} : \mathbf{T}\mathbf{v} = \lambda\mathbf{v}\}$$

è un sottospazio T -invariante. Infatti, per ogni $\mathbf{v} \in \mathbf{V}_\lambda$, $\mathbf{T}\mathbf{v} = \lambda\mathbf{v}$ appartiene a \mathbf{V}_λ perché:

$$\mathbf{T}(\mathbf{T}\mathbf{v}) = \mathbf{T}(\lambda\mathbf{v}) = \lambda(\mathbf{T}\mathbf{v})$$

Anche il sottospazio radicale \mathbf{R}_λ è \mathbf{T} -invariante.

Se \mathbf{H} è un sottospazio \mathbf{T} -invariante di \mathbf{V} , possiamo considerare \mathbf{T} come un operatore di \mathbf{H} in se stesso, perché $\mathbf{T}\mathbf{v} \in \mathbf{H}$ per ogni \mathbf{v} di \mathbf{H} . Questo operatore si dice *la restrizione di \mathbf{T} a \mathbf{H}* : agisce come \mathbf{T} sui vettori di \mathbf{H} , ha però \mathbf{H} come dominio e codominio al posto di \mathbf{V} . Abbiamo già utilizzato questa nozione nella dimostrazione del teorema 6.8, quando abbiamo considerato la restrizione di \mathbf{T} al sottospazio invariante $\text{Im}(\mathbf{T})$.

LEMMA 6.23 Sia $\mathbf{T} : \mathbf{V} \rightarrow \mathbf{V}$ un operatore diagonalizzabile, e sia \mathbf{H} un sottospazio \mathbf{T} -invariante. Allora la restrizione di \mathbf{T} a \mathbf{H} è diagonalizzabile.

DIMOSTRAZIONE. L'ipotesi è che esista un base di \mathbf{V} formata da autovettori di \mathbf{T} . La tesi è che esista un base di \mathbf{H} formata da autovettori di \mathbf{T} . Questo non è banale perché la base di \mathbf{V} potrebbe non contenere alcun vettore di \mathbf{H} . Sappiamo però che un operatore è diagonalizzabile se e solo se ogni autovettore generalizzato è un autovettore vero. Se questo è vero per \mathbf{T} , è vero banalmente anche per la sua restrizione a \mathbf{H} .

PROPOSIZIONE 6.24 Siano \mathbf{A} e \mathbf{B} due matrici quadrate dello stesso ordine. Se \mathbf{A} e \mathbf{B} sono diagonalizzabili e se $\mathbf{AB} = \mathbf{BA}$, allora \mathbf{A} e \mathbf{B} sono simultaneamente diagonalizzabili: esiste una matrice invertibile \mathbf{S} tale che $\mathbf{S}^{-1}\mathbf{AS}$ e $\mathbf{S}^{-1}\mathbf{BS}$ siano entrambe diagonali.

DIMOSTRAZIONE. Sia n l'ordine di \mathbf{A} . Per ipotesi \mathbb{C}^n è somma diretta degli autospazi $\mathbf{V}_\lambda(\mathbf{A})$ di \mathbf{A} . Mostriamo che questi autospazi sono \mathbf{B} -invarianti. Sia \mathbf{v} un vettore di $\mathbf{V}_\lambda(\mathbf{A})$: questo significa che $\mathbf{Av} = \lambda\mathbf{v}$. Calcoliamo $\mathbf{A}(\mathbf{Bv})$:

$$\mathbf{A}(\mathbf{Bv}) = \mathbf{ABv} = \mathbf{BAv} = \mathbf{B}(\lambda\mathbf{v}) = \lambda\mathbf{Bv}$$

Quindi anche \mathbf{Bv} appartiene a $\mathbf{V}_\lambda(\mathbf{A})$, cioè $\mathbf{V}_\lambda(\mathbf{A})$ è \mathbf{B} -invariante. Per il lemma 6.23 la restrizione di \mathbf{B} a $\mathbf{V}_\lambda(\mathbf{A})$ è diagonalizzabile: questo significa che esiste una base \mathcal{B}_λ di $\mathbf{V}_\lambda(\mathbf{A})$ formata da autovettori di \mathbf{B} ; i vettori della base \mathcal{B}_λ sono anche autovettori di \mathbf{A} perché appartengono all'autospazio $\mathbf{V}_\lambda(\mathbf{A})$. Facendo l'unione delle basi \mathcal{B}_λ al variare di λ tra gli autovalori di \mathbf{A} troviamo una base \mathcal{B} di \mathbb{C}^n , perché \mathbb{C}^n è la somma diretta degli autospazi di \mathbf{A} . Sia \mathbf{S} la matrice che ha come colonne i vettori della base \mathcal{B} . Per costruzione le colonne di \mathbf{S} sono autovettori sia di \mathbf{A} sia di \mathbf{B} , quindi \mathbf{S} diagonalizza tanto \mathbf{A} quanto \mathbf{B} .

Determinare la forma canonica di Jordan delle matrici

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{e} \quad \mathbf{B} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Le due matrici sono simili?

Quante possibilità ci sono per la forma canonica di Jordan di una matrice 3×3 nilpotente? Descrivelte tutte.

Suggerimento: in quanti modi si può scrivere 3 come somma di interi positivi? Le possibilità sono $3 = 3$, $3 = 2 + 1$, $3 = 1 + 1 + 1$.

Mostrare che due matrici quadrate di ordine $n = 2$ (oppure $n = 3$) che abbiano gli stessi autovalori con le stesse molteplicità algebriche e geometriche hanno necessariamente tutti gli invarianti $r_{k,d}$ uguali, e quindi sono simili.

Suggerimento: fissato k , la molteplicità algebrica e geometrica di λ_k determinano due degli $r_{k,d}$, e non c'è spazio perché ci siano tre valori distinti di $r_{k,d}$ (visto che $r_{k,1} < n$); la difficoltà è nel mostrare, nel caso $n = 3$, che, se $r_{k,1} = r_{k,2}$, allora $r_{k,2} = r_{k,3}$.

Mostrare che le due matrici

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 5 \end{bmatrix} \quad \text{e} \quad \mathbf{B} = \begin{bmatrix} 1 & 4 & 0 \\ 0 & 5 & 0 \\ 2 & 3 & 1 \end{bmatrix}$$

sono simili, e determinare la loro forma canonica di Jordan.

Determinare la forma canonica di Jordan di tutte le matrici non diagonalizzabili di questo capitolo.

Mostrare che una matrice nilpotente non è invertibile. Mostrare che la traccia di una matrice nilpotente è uguale a zero.

Mostrare che, se \mathbf{N} è nilpotente e $\mathbf{BN} = \mathbf{NB}$, allora \mathbf{BN} è nilpotente.

Mostrare che, se \mathbf{N}_1 e \mathbf{N}_2 sono nilpotenti e $\mathbf{N}_1\mathbf{N}_2 = \mathbf{N}_2\mathbf{N}_1$, allora $\mathbf{N}_1 + \mathbf{N}_2$ è nilpotente.

Suggerimento: binomio di Newton.

Mostrare che, se \mathbf{N} è nilpotente e diagonalizzabile, allora \mathbf{N} è la matrice nulla.

Mostrare che la restrizione di $\mathbf{T} - \lambda\mathbf{I}$ al sottospazio radicale $\mathbf{R}_\lambda(\mathbf{T})$ è nilpotente.

La decomposizione di Jordan è unica: si supponga che $\mathbf{A} = \mathbf{B} + \mathbf{N}$, con \mathbf{B} diagonalizzabile, \mathbf{N} nilpotente, e $\mathbf{NB} = \mathbf{BN}$; mostrare che i sottospazi radicali di \mathbf{A} coincidono con gli autospazi di \mathbf{B} . Perché questo determina \mathbf{B} (e quindi anche \mathbf{N})?

Suggerimento: $(\mathbf{B} - \lambda\mathbf{I})^d = (\mathbf{A} - \lambda\mathbf{I} - \mathbf{N})^d$.

Mostrare che, se \mathbf{A} è invertibile e $\mathbf{AB} = \mathbf{BA}$, allora $\mathbf{A}^{-1}\mathbf{B} = \mathbf{BA}^{-1}$.

Sia \mathbf{N} una matrice nilpotente. Mostrare che $\mathbf{I} + \mathbf{N}$ è invertibile e dare una formula per l'inversa. Mostrarre poi che, se \mathbf{B} è invertibile e \mathbf{N} è nilpotente e $\mathbf{BN} = \mathbf{NB}$, allora $\mathbf{B} + \mathbf{N}$ è invertibile.

Suggerimento: per l'invertibilità basta far vedere che 0 non è un autovalore di $\mathbf{I} + \mathbf{N}$; la formula per l'inversa si trova considerando l'identità $(1+x)(1-x+\dots\pm x^{d-1}) = 1-x^d$.

Sia \mathbf{A} una matrice tale che $\mathbf{A}^2 = \mathbf{I}$. Mostrare che \mathbf{A} è diagonalizzabile. Suggerimento: utilizzare la decomposizione di Jordan $\mathbf{A} = \mathbf{B} + \mathbf{N}$, mostrare che $\mathbf{B}^2 = \mathbf{I}$ e che \mathbf{N} dev'essere nulla usando l'esercizio precedente. Per i coraggiosi: dimostrare che, se $\mathbf{A}^d = \mathbf{I}$, allora \mathbf{A} è diagonalizzabile.

Mostrare che i seguenti sottospazi di \mathbf{V} sono \mathbf{T} -invarianti: il nucleo e l'immagine di \mathbf{T} , gli autospazi e i sottospazi radicali di \mathbf{T} .

Una difficoltà nella costruzione della forma canonica di Jordan è dovuta al seguente fatto: se $\mathbf{V}_1 \subset \mathbf{V}$ è un sottospazio \mathbf{T} -invariante, non è detto che \mathbf{V}_1 abbia un complemento \mathbf{T} -invariante, cioè che esista un sottospazio \mathbf{T} -invariante \mathbf{V}_2 tale che $\mathbf{V} = \mathbf{V}_1 \oplus \mathbf{V}_2$. Dimostare che $\text{Im}(\mathbf{T})$ ha un complemento \mathbf{T} -invariante \mathbf{V}_2 se e solo se $\text{Ker}(\mathbf{T}) \cap \text{Im}(\mathbf{T}) = \{0\}$, e in tal caso $\mathbf{V}_2 = \text{Ker}(\mathbf{T})$.

Dati due polinomi qualsiasi $P(x)$ e $Q(x)$, mostrare che $P(\mathbf{T})Q(\mathbf{T}) = Q(\mathbf{T})P(\mathbf{T})$. In particolare, $P(\mathbf{T})$ commuta con \mathbf{T} .

Supponiamo \mathbf{T} sia invertibile. Allora, dati due polinomi $P(x, y)$ e $Q(x, y)$ in due variabili, gli operatori $P(\mathbf{T}, \mathbf{T}^{-1})$ e $Q(\mathbf{T}, \mathbf{T}^{-1})$ commutano.

Supponiamo $\mathbf{B} : \mathbf{V} \rightarrow \mathbf{V}$ sia un operatore che commuta con \mathbf{T} . Mostrare che $\text{Ker}(\mathbf{B})$ e $\text{Im}(\mathbf{B})$ sono \mathbf{T} -invarianti. Dedurre che gli autospazi e i sottospazi radicali di \mathbf{T} sono \mathbf{T} -invarianti.

Mostrare che il sottospazio generato dai vettori $\mathbf{T}^m \mathbf{v}$ per $m \geq 1$ è un sottospazio \mathbf{T} -invariante, e che è contenuto in ogni sottospazio \mathbf{T} -invariante che contenga \mathbf{v} .

Sia $\mathbf{T} : \mathbf{V} \rightarrow \mathbf{V}$ un operatore lineare. Mostrare che le seguenti proprietà sono equivalenti per un intero $e \geq 0$:

1. $\text{Ker}(\mathbf{T}^e) = \text{Ker}(\mathbf{T}^m)$ per ogni $m \geq e$;
2. $\text{Im}(\mathbf{T}^e) = \text{Im}(\mathbf{T}^m)$ per ogni $m \geq e$;
3. $\text{Ker}(\mathbf{T}^e) \cap \text{Im}(\mathbf{T}^e) = \{0\}$;
4. $\mathbf{V} = \text{Ker}(\mathbf{T}^e) \oplus \text{Im}(\mathbf{T}^e)$.

Dimostrare che l'indice di un autovalore λ è 1 se e solo se $\text{Ker}(\mathbf{T} - \lambda \mathbf{I}) \cap \text{Im}(\mathbf{T} - \lambda \mathbf{I}) = \{0\}$. Enunciare un nuovo criterio di diagonalizzabilità.

Supponiamo:

$$\mathbf{V} = \mathbf{V}_1 \oplus \mathbf{V}_2 \oplus \dots \oplus \mathbf{V}_s$$

e che $\mathbf{V}_i \subset \mathbf{V}$ è un sottospazio \mathbf{T} -invariante, per ogni $i = 1, 2, \dots, s$. Fissate basi $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_s$ di $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_s$ rispettivamente, sia \mathcal{B} la base di \mathbf{V} ottenuta accostando i vettori di $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_s$. Mostrare che:

1. La matrice che rappresenta \mathbf{T} rispetto a \mathcal{B} è diagonale a blocchi, e i blocchi sulla diagonale sono le matrici che rappresentano le restrizioni \mathbf{T}_i di \mathbf{T} ai sottospazi \mathbf{V}_i rispetto a \mathcal{B}_i ;
2. il polinomio caratteristico di \mathbf{T} è il prodotto dei polinomi caratteristici delle restrizioni \mathbf{T}_i . Dare una nuova dimostrazione del fatto che la molteplicità algebrica di un autovalore è uguale alla dimensione del sottospazio radicale.

 Sia \mathbf{A} una matrice quadrata e sia $P(\lambda)$ un polinomio. Mostrare che $P(\mathbf{A}) = \mathbf{O}$ se e solo se il polinomio minimo di \mathbf{A} divide $P(\lambda)$.

Suggerimento: scomporre $\mathbf{P}(\lambda)$ in fattori della forma $(\lambda - \lambda_i)^{d_i}$ e analizzare la dimostrazione del teorema di Hamilton-Cayley.

 Mostrare che, se \mathbf{A} e \mathbf{B} sono simultaneamente diagonalizzabili, allora $\mathbf{AB} = \mathbf{BA}$.

Spazi euclidei

1 INTRODUZIONE

In questo capitolo introduciamo in \mathbb{R}^n il prodotto scalare di due vettori sul modello di quanto visto nel primo capitolo per lo spazio cartesiano. Questo consente di definire i concetti di *norma* (o *modulo* o *lunghezza*) di un vettore, di *angolo* formato da due vettori e di *distanza* tra due vettori. La distanza di due vettori \mathbf{v} e \mathbf{w} è per definizione la norma del vettore differenza $\mathbf{w} - \mathbf{v}$. Per capire questa definizione bisogna immaginare che $\mathbf{v} = \overrightarrow{OP}$ e $\mathbf{w} = \overrightarrow{OQ}$ siano i vettori posizionati di due punti P e Q ; il vettore differenza è allora $\mathbf{w} - \mathbf{v} = \overrightarrow{PQ}$ e, quindi, la distanza tra \mathbf{v} e \mathbf{w} è la lunghezza del segmento PQ . L'introduzione della distanza permette di parlare di approssimazioni e di limiti, insomma di fare dell'analisi matematica e questo apre la porta a infinite applicazioni. È utile estendere tutto questo al contesto astratto, in particolare per estendere le nozioni di distanza e approssimazione agli spazi di funzioni. Il pioniere di questa generalizzazione è stato Fourier: nel suo studio delle funzioni periodiche, Fourier ebbe l'idea di approssimare una funzione periodica qualsiasi con un polinomio trigonometrico, che è una combinazione lineare di semplici funzioni sinusoidali (armoniche fondamentali); si rese conto che il problema di trovare il polinomio trigonometrico che meglio approssima una data funzione è perfettamente analogo al problema di trovare la proiezione ortogonale di un vettore su un sottospazio di \mathbb{R}^n . Oggigiorno questo procedimento va sotto il nome di metodo dei minimi quadrati. Nel capitolo descriviamo un'applicazione importante del metodo dei minimi quadrati ai sistemi lineari sovradeterminati. A causa di errori sperimentali occorre spesso considerare sistemi $\mathbf{Ax} = \mathbf{b}$ che non ammettono soluzione: non esiste alcun \mathbf{x} per cui $\mathbf{Ax} = \mathbf{b}$. Ma si possono sempre determinare gli \mathbf{x} che minimizzano l'errore $\|\mathbf{Ax} - \mathbf{b}\|$, cioè la distanza tra \mathbf{Ax} e \mathbf{b} ; tali \mathbf{x} si dicono soluzioni ai minimi quadrati del sistema lineare. Nel capitolo descriviamo un esempio che viene dalla statistica: la determinazione della retta di regressione lineare consiste nel risolvere ai minimi quadrati un sistema lineare sovradeterminato.

Un'osservazione importante: quanto visto nei capitoli precedenti (escluso il primo) rimane valido se al posto di \mathbb{R} si prende come campo degli scalari l'insieme \mathbb{Q} dei numeri razionali. Questo sostanzialmente perché la risoluzione di un sistema lineare richiede soltanto le quattro operazioni: partendo da dati che sono numeri interi o

razionali, si ottengono soluzioni razionali. Ma nel momento in cui si vogliono misurare le lunghezze e gli angoli, il che comporta algebricamente risolvere equazioni almeno di secondo grado, diventa necessario introdurre un insieme numerico in cui si possano estrarre le radici quadrate, per cui l'uso dei numeri reali diventa imperativo. Questo era già stato osservato dagli antichi greci: la lunghezza della diagonale di un quadrato di lato unitario è $\sqrt{2}$, e questo è un numero irrazionale: la diagonale del quadrato non è commensurabile al lato.

2 SPAZI EUCLIDEI

Uno spazio euclideo è uno spazio vettoriale in cui è possibile misurare lunghezze e angoli in modo analogo al caso dello spazio euclideo tridimensionale studiato nel primo capitolo. Lo strumento tecnico fondamentale è il prodotto scalare di due vettori:

DEFINIZIONE 2.1 (Prodotto scalare - Spazio euclideo)

Sia \mathbf{V} uno spazio vettoriale sul campo dei numeri reali \mathbb{R} . Un *prodotto scalare* (o *prodotto interno*) in \mathbf{V} è una funzione che a ogni coppia di vettori \mathbf{v} e \mathbf{w} di \mathbf{V} associa un numero reale $\langle \mathbf{v}, \mathbf{w} \rangle$ in modo che le seguenti proprietà siano soddisfatte:

a) **Commutatività:**

$$\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle \quad \text{per ogni } \mathbf{v}, \mathbf{w} \in \mathbf{V}$$

b) **Linearità nel primo fattore:**

$$\langle \mathbf{v}_1 + \mathbf{v}_2, \mathbf{w} \rangle = \langle \mathbf{v}_1, \mathbf{w} \rangle + \langle \mathbf{v}_2, \mathbf{w} \rangle \quad \text{per ogni } \mathbf{v}_1, \mathbf{v}_2, \mathbf{w} \in \mathbf{V}$$

$$\langle t\mathbf{v}, \mathbf{w} \rangle = t \langle \mathbf{v}, \mathbf{w} \rangle \quad \text{per ogni } t \in \mathbb{R} \text{ e ogni } \mathbf{v}, \mathbf{w} \in \mathbf{V}$$

c) **Positività:**

$$\langle \mathbf{v}, \mathbf{v} \rangle \geq 0 \quad \text{per ogni } \mathbf{v} \in \mathbf{V}$$

e l'uguaglianza vale se e solo se $\mathbf{v} = \mathbf{0}$.

Uno *spazio euclideo* è uno spazio vettoriale su \mathbb{R} dotato di un prodotto scalare.

Molte altre notazioni vengono utilizzate per un prodotto scalare, tra queste le più comuni sono $\mathbf{v} \cdot \mathbf{w}$ e (\mathbf{v}, \mathbf{w}) .

OSSERVAZIONE Dalla commutatività e dalla linearità nel primo fattore, segue che un prodotto scalare è lineare anche nel secondo fattore, cioè

$$\langle \mathbf{v}, \mathbf{w}_1 + \mathbf{w}_2 \rangle = \langle \mathbf{v}, \mathbf{w}_1 \rangle + \langle \mathbf{v}, \mathbf{w}_2 \rangle \quad \text{e} \quad \langle \mathbf{v}, t\mathbf{w} \rangle = t \langle \mathbf{v}, \mathbf{w} \rangle$$

Il prodotto scalare è quindi *bilineare*, cioè lineare in ciascuno dei suoi fattori. La definizione può essere riformulata così: un prodotto scalare è una funzione bilineare $\mathbf{V} \times \mathbf{V} \rightarrow \mathbb{R}$ che sia *simmetrica* (proprietà a)) e *definita positiva* (proprietà c)).

Prodotto scalare standard in \mathbb{R}^n .

In \mathbb{R}^n il prodotto scalare standard è denotato con il simbolo $\mathbf{x} \cdot \mathbf{y}$ ed è definito dalla formula

$$(2.1) \quad \mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y}$$

Quindi, se $\mathbf{x} = [x_1, \dots, x_n]^T$ e $\mathbf{y} = [y_1, \dots, y_n]^T$,

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + \dots + x_n y_n$$

Se si pensa al primo vettore come a un vettore riga e al secondo come a un vettore colonna, il prodotto scalare non è altro che il prodotto riga per colonna. È semplice dimostrare che questo prodotto verifica le proprietà richieste dalla definizione di prodotto scalare. Mostriamo per esempio la positività: se $\mathbf{x} = [x_1, \dots, x_n]^T$, allora

$$\mathbf{x} \cdot \mathbf{x} = x_1^2 + \dots + x_n^2$$

per cui $\mathbf{x} \cdot \mathbf{x} \geq 0$ e l'uguaglianza vale se e solo se $\mathbf{x} = \mathbf{0}$. Si noti che stiamo utilizzando il fatto che il quadrato di un numero reale è un numero ≥ 0 : per questo per estendere la definizione al caso complesso occorre apportare delle modifiche.

Tecnicamente, uno spazio euclideo è una coppia $(\mathbf{V}, \langle \cdot, \cdot \rangle)$ costituita da uno spazio vettoriale reale \mathbf{V} insieme a un prodotto scalare in \mathbf{V} . D'ora innanzi quando parleremo di \mathbb{R}^n sottointenderemo che il prodotto scalare è quello standard. Il lettore che non ama l'astrazione può concentrarsi su questo esempio. Vedremo che in un senso preciso questo è l'unico esempio: uno spazio euclideo \mathbf{V} di dimensione n con prodotto scalare $\langle \cdot, \cdot \rangle$ può essere identificato, mediante la scelta di una base opportuna, con \mathbb{R}^n in modo che il prodotto scalare astratto $\langle \cdot, \cdot \rangle$ corrisponda al prodotto scalare standard di \mathbb{R}^n . Come al solito, la nozione astratta ha il vantaggio di non dipendere dalla scelta della base opportuna e di consentire, quindi, la scelta delle coordinate più convenienti a posteriori.

Prodotto scalare sui sottospazi.

Se \mathbf{V} è uno spazio euclideo e \mathbf{H} è un sottospazio vettoriale di \mathbf{V} , anche \mathbf{H} è evidentemente uno spazio euclideo: il prodotto scalare di due vettori di \mathbf{H} è definito come il loro prodotto scalare in \mathbf{V} .

Per esempio, consideriamo in \mathbb{R}^3 il piano \mathbf{H} di equazione $x + y + z = 0$ con la base $\mathbf{b} = [1, -1, 0]^T$ e $\mathbf{c} = [1, 0, -1]^T$. Dati due vettori $\mathbf{v} = x_1 \mathbf{b} + x_2 \mathbf{c} = [x_1 + x_2, -x_1, -x_2]^T$ e $\mathbf{v}_2 = y_1 \mathbf{b} + y_2 \mathbf{c} = [y_1 + y_2, -y_1, -y_2]^T$ di \mathbf{H} , il loro prodotto scalare è

$$\begin{aligned} \langle \mathbf{v}_1, \mathbf{v}_2 \rangle_{\mathbf{H}} &= \langle \mathbf{v}_1, \mathbf{v}_2 \rangle_{\mathbb{R}^3} = (x_1 + x_2)(y_1 + y_2) + x_1 y_1 + x_2 y_2 = \\ &= 2x_1 y_1 + 2x_2 y_2 + x_1 y_2 + x_2 y_1 \end{aligned}$$

Si noti che possiamo utilizzare questa formula per definire su \mathbb{R}^2 un prodotto scalare *diverso da quello standard*:

$$\langle [x_1, x_2]^T, [y_1, y_2]^T \rangle = 2x_1 y_1 + 2x_2 y_2 + x_1 y_2 + x_2 y_1$$

Lo spazio ℓ_2 .

Diamo ora il più semplice esempio di uno spazio euclideo di dimensione infinita, sostituendo i vettori con n componenti di \mathbb{R}^n con le successioni, che possiamo pensare come vettori con un'infinità numerabile di componenti. Dobbiamo però limitarci all'insieme ℓ_2 delle successioni $\mathbf{s} = \{a_n\}_{n \geq 0}$ di numeri reali che abbiano lunghezza finita nel senso seguente:

$$\sum_{n=0}^{+\infty} a_n^2 < +\infty$$

Date due successioni $\mathbf{s} = \{a_n\}_{n \geq 0}$ e $\mathbf{t} = \{b_n\}_{n \geq 0}$ in ℓ_2 , si può dimostrare (utilizzando la disuguaglianza di Schwarz che discuteremo più avanti) che la serie

$$\sum_{n=0}^{+\infty} a_n b_n$$

è convergente. Si può quindi porre $\mathbf{s} \cdot \mathbf{t} = \sum_{n=0}^{+\infty} a_n b_n$ e verificare che si ottiene così un prodotto scalare in ℓ_2 .

Possiamo pensare a una funzione $f : [a, b] \rightarrow \mathbb{R}$ come a un vettore con un'infinità di componenti: a ogni x di $[a, b]$ corrisponde la componente $f(x)$ del vettore f . Non si tratta più di un'infinità numerabile e se vogliamo sommare le componenti occorre sostituire le sommatorie con gli integrali. Ma per il resto le formule sono simili! Per essere precisi, sia \mathbf{V} lo spazio vettoriale delle funzioni $f : [a, b] \rightarrow \mathbb{R}$ che sono *continue* in $[a, b]$. In \mathbf{V} possiamo definire l'analogo del prodotto scalare standard di \mathbb{R}^n ponendo

$$\langle f, g \rangle = \int_a^b f(x)g(x) dx$$

Le proprietà richieste dalla definizione di prodotto scalare sono soddisfatte anche in questo caso. L'unica cosa non completamente evidente è che $\langle f, f \rangle = 0$ implica che f sia la funzione nulla. Per vedere questo, osserviamo che se $f(x)$ non è la funzione nulla, esistono $\epsilon > 0$ e un punto $x_0 \in [a, b]$ tali che $f(x_0)^2 > \epsilon$. Per la continuità di f , esiste un intorno $(c, d) \subset [a, b]$ di x_0 tale che $f(x)^2 > \epsilon$ per ogni $x \in (c, d)$. e quindi

$$\langle f, f \rangle = \int_a^b f(x)^2 dx \geq \int_c^d f(x)^2 dx > \int_c^d \epsilon dx = \epsilon(d - c) > 0$$

Perciò da $\langle f, f \rangle = 0$ segue $f(x) = 0$ per ogni x come volevasi dimostrare.

Prodotto scalare standard in $\mathbb{M}_{\mathbb{R}}(m, n)$.

Lo spazio vettoriale $\mathbb{M}_{\mathbb{R}}(m, n)$ delle matrici di tipo (m, n) si identifica con $\mathbb{R}^{m \times n}$ e, quindi, possiede un prodotto scalare standard. La formula compatta per questo prodotto è:

$$(2.2) \quad \langle \mathbf{A}, \mathbf{B} \rangle = \text{tr}(\mathbf{A}^T \mathbf{B})$$

Il simbolo tr denota la traccia, che è la somma degli elementi sulla diagonale principale di una matrice quadrata: $\text{tr}([c_{kl}]) = \sum c_{kk}$. Si osservi che $\mathbf{A}^T \mathbf{B}$ è una matrice quadrata di ordine n . Per vedere perché questo prodotto coincide con quello standard di $\mathbb{R}^{n \times n}$, scriviamo $\mathbf{A} = [a_{jk}]$ e $\mathbf{B} = [b_{jl}]$ e calcoliamo

$$\text{tr}(\mathbf{A}^T \mathbf{B}) = \text{tr} \left(\left[\sum_{j=1}^m a_{jk} b_{jl} \right] \right) = \sum_{k=1}^n \sum_{j=1}^m a_{jk} b_{jk}$$

Possiamo introdurre in uno spazio euclideo qualsiasi la nozione di *norma* di un vettore:

DEFINIZIONE 2.2 (Norma di un vettore)

Sia \mathbf{V} uno spazio euclideo con prodotto scalare $\langle \cdot, \cdot \rangle$. La *norma* (o *lunghezza* o *modulo*) di un vettore $\mathbf{v} \in \mathbf{V}$ è il numero reale ≥ 0 :

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

Nel caso di \mathbb{R}^n con il prodotto scalare standard la norma di un vettore è

$$\|[x_1, \dots, x_n]^T\| = \sqrt{x_1^2 + \dots + x_n^2}$$

Questa formula generalizza la formula per la lunghezza di un vettore in coordinate cartesiane nello spazio euclideo tridimensionale.

Nel caso del piano \mathbf{H} di equazione $x + y + z = 0$ in \mathbb{R}^3 , la norma del vettore $[x_1 + x_2, -x_1, -x_2]^T = x_1 \mathbf{b} + x_2 \mathbf{c}$ è

$$\|x_1 \mathbf{b} + x_2 \mathbf{c}\| = \sqrt{2x_1^2 + 2x_2^2 + 2x_1 x_2}$$

Nel caso dello spazio delle funzioni continue su $[a, b]$ con il prodotto scalare $\langle f, g \rangle = \int_a^b f(x)g(x) dx$, la norma di una funzione è

$$\|f\| = \left(\int_a^b (f(x))^2 dx \right)^{1/2}$$

Questa norma si dice *norma L^2* di f .

Due proprietà importanti della norma sono

- Omogeneità

$$(2.3) \quad \|t\mathbf{v}\| = |t| \|\mathbf{v}\| \text{ per ogni } t \in \mathbb{R} \text{ e ogni } \mathbf{v} \in \mathbf{V}$$

- Annullamento:

$$(2.4) \quad \|\mathbf{v}\| = 0 \quad \text{se e solo se } \mathbf{v} = 0$$

L'omogeneità della norma si verifica in questo modo:

$$\|t\mathbf{v}\| = \sqrt{\langle t\mathbf{v}, t\mathbf{v} \rangle} = \sqrt{t^2 \langle \mathbf{v}, \mathbf{v} \rangle} = |t| \|\mathbf{v}\|$$

La proprietà di annullamento è una riformulazione dell'assioma del prodotto scalare secondo cui $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ se e solo se $\mathbf{v} = \mathbf{0}$. Vi è una terza proprietà fondamentale della norma, la diseguaglianza triangolare, su cui torneremo più avanti.

DEFINIZIONE 2.3 (Versore)

Un *versore* è un vettore la cui norma è uguale a 1. In simboli: \mathbf{e} è un versore se $\|\mathbf{e}\| = 1$.

Dalla proprietà (3.11) di omogeneità della norma segue:

- i) se \mathbf{e} è un versore, il modulo di $t\mathbf{e}$ è precisamente $|t|$;
- ii) se \mathbf{v} è un vettore non nullo, esistono due e soltanto due versori sulla retta $\mathbb{R}\mathbf{v}$ generata da \mathbf{v} :

$$\mathbf{e}_1(\mathbf{v}) = \frac{\mathbf{v}}{\|\mathbf{v}\|} \quad \text{ed} \quad \mathbf{e}_2(\mathbf{v}) = -\frac{\mathbf{v}}{\|\mathbf{v}\|}$$

Il versore $\mathbf{e}_1(\mathbf{v})$ è ottenuto dividendo \mathbf{v} per la sua norma. Siccome $\frac{1}{\|\mathbf{v}\|} > 0$, il versore $\mathbf{e}_1(\mathbf{v})$ ha lo stesso verso di \mathbf{v} , mentre $\mathbf{e}_2(\mathbf{v})$ ha verso opposto.

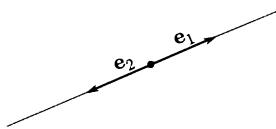


Figura 2.1. I due versori della retta r .

Consideriamo il vettore $\mathbf{v} = [2, 3]^T$ di \mathbb{R}^2 . La norma di \mathbf{v} è

$$\|\mathbf{v}\| = \sqrt{2^2 + 3^2} = \sqrt{13}$$

I due versori della retta $\mathbb{R}\mathbf{v}$ sono $\mathbf{e}_1 = [2/\sqrt{13}, 3/\sqrt{13}]^T$ ed $\mathbf{e}_2 = [-2/\sqrt{13}, -3/\sqrt{13}]^T$.

Il prodotto scalare consente di definire anche la nozione di ortogonalità (o perpendicolarità):

DEFINIZIONE 2.4 (Ortogonalità)

Sia \mathbf{V} uno spazio euclideo con prodotto scalare $\langle \cdot, \cdot \rangle$. I vettori $\mathbf{v}, \mathbf{w} \in \mathbf{V}$ si dicono *ortogonali* (o *perpendicolari*) e si scrive $\mathbf{v} \perp \mathbf{w}$, se

$$\langle \mathbf{v}, \mathbf{w} \rangle = 0$$

OSSERVAZIONE Dalla bilinearità del prodotto scalare segue che il prodotto scalare del vettore nullo con un vettore qualsiasi \mathbf{w} è nullo: il vettore nullo è perciò ortogonale a ogni vettore.



In \mathbb{R}^4 i vettori $\mathbf{v} = \frac{1}{2}[1, 1, 1, 1]^T$ e $\mathbf{w} = \frac{1}{2}[1, 1, -1, -1]^T$ sono versori e sono perpendicolari l'uno all'altro.

Sia \mathbf{V} lo spazio vettoriale delle funzioni continue $f : [0, 2\pi] \rightarrow \mathbb{R}$, con il prodotto scalare definito come prima per integrazione. Le funzioni $\cos(x)$ e $\sin(x)$ hanno norma $\sqrt{\pi}$ e sono ortogonali tra loro. Infatti

$$\|\cos(x)\|^2 = \int_0^{2\pi} \cos^2(x) dx = \frac{1}{2} \int_0^{2\pi} (1 + \cos(2x)) dx = \pi$$

Analogamente si verifica $\|\sin(x)\|^2 = \pi$. L'ortogonalità delle due funzioni segue dal fatto che la funzione prodotto $\cos(x)\sin(x)$ è dispari, per cui

$$\langle \cos(x), \sin(x) \rangle = \int_0^{2\pi} \cos(x) \sin(x) dx = \int_{-\pi}^{\pi} \cos(x) \sin(x) dx = 0$$

(il valore dell'integrale su $[0, 2\pi]$ e $[-\pi, \pi]$ è uguale perché le due funzioni sono periodiche di periodo 2π).



Determinare tutti i vettori di \mathbb{R}^4 ortogonali al vettore $\mathbf{n} = [1, 2, 3, 0]^T$ risolvendo un opportuno sistema lineare. Trovare tutti i versori ortogonali a \mathbf{n} .

Dato $\mathbf{v} = [1/3, -1/3, 2/3]$, spiegare perché $\frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{3\mathbf{v}}{\|3\mathbf{v}\|}$. Qual è la norma di \mathbf{v} ? E quella di $3\mathbf{v}$? Quale conviene calcolare? Scrivere esplicitamente le coordinate di $\frac{\mathbf{v}}{\|\mathbf{v}\|}$.

Verificare che la formula

$$\langle [x_1, x_2]^T, [y_1, y_2]^T \rangle = 2x_1y_1 + 2x_2y_2 + x_1y_2 + x_2y_1$$

definisce un prodotto scalare in \mathbb{R}^2 .

Siano $\lambda_1, \dots, \lambda_n$ dei numeri reali positivi. Dati $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, si definisca

$$\langle \mathbf{x}, \mathbf{y} \rangle = \lambda_1 x_1 y_1 + \dots + \lambda_n x_n y_n$$

Mostrare che $\langle \cdot, \cdot \rangle$ è un prodotto scalare in \mathbb{R}^n . Nel caso particolare $n = 2$, porre $\lambda_1 = 2$ e $\lambda_2 = 3$ e determinare tutti i vettori ortogonali, rispetto a questo prodotto scalare, al vettore $[1, 1]^T$.

Sia \mathbf{V} lo spazio vettoriale delle funzioni continue in $[0, 2\pi]$ con la norma L^2 . Mostrare che le funzioni $\cos(kx)$ (per $k \geq 0$) e $\sin(kx)$ (per $k \geq 1$), sono a due a due ortogonali. Mostrare che la norma L^2 di $1 = \cos(0x)$ è $\sqrt{2\pi}$, mentre la norma L^2 di $\cos(kx)$ e $\sin(kx)$ per $k \geq 1$ è $\sqrt{\pi}$.

3 IL TEOREMA DI PITAGORA E LA DISUGUAGLIANZA DI SCHWARZ

Il legame tra norma e ortogonalità è espresso dalla versione astratta del teorema più importante della geometria euclidea: il teorema di Pitagora.

PROPOSIZIONE 3.1

Sia \mathbf{V} uno spazio euclideo e siano $\mathbf{v}_1, \mathbf{v}_2$ vettori di \mathbf{V} .

- a) **Teorema di Pitagora:** se \mathbf{v}_1 e \mathbf{v}_2 sono ortogonali, allora

$$\|\mathbf{v}_1 + \mathbf{v}_2\|^2 = \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2$$

- b) **Teorema di Carnot:** per due vettori arbitrari vale la formula

$$\|\mathbf{v}_1 + \mathbf{v}_2\|^2 = \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 + 2 < \mathbf{v}_1, \mathbf{v}_2 >$$

DIMOSTRAZIONE. La dimostrazione è un esercizio sull'uso delle proprietà del prodotto scalare, perfettamente analogo allo sviluppo del quadrato in algebra elementare (solo la terminologia cambia: a scuola si parla di proprietà distributiva anziché di bilinearità del prodotto).

$$\begin{aligned} \|\mathbf{v}_1 + \mathbf{v}_2\|^2 &= < \mathbf{v}_1 + \mathbf{v}_2, \mathbf{v}_1 + \mathbf{v}_2 > = \text{(linearità rispetto primo fattore)} \\ &= < \mathbf{v}_1, \mathbf{v}_1 + \mathbf{v}_2 > + < \mathbf{v}_2, \mathbf{v}_1 + \mathbf{v}_2 > = \text{(linearità rispetto secondo fattore)} \\ &= < \mathbf{v}_1, \mathbf{v}_1 > + < \mathbf{v}_1, \mathbf{v}_2 > + < \mathbf{v}_2, \mathbf{v}_1 > + < \mathbf{v}_2, \mathbf{v}_2 > = \text{(commutatività)} \\ &= \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 + 2 < \mathbf{v}_1, \mathbf{v}_2 > \end{aligned}$$

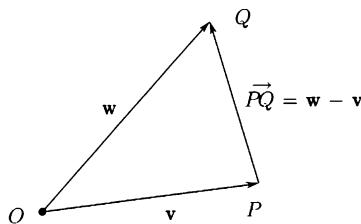
Se \mathbf{v}_1 e \mathbf{v}_2 sono perpendicolari, il termine $2 < \mathbf{v}_1, \mathbf{v}_2 >$ è nullo.

OSSERVAZIONE Dal teorema di Carnot segue che il prodotto scalare è determinato dalla funzione che a un vettore associa la sua norma:

$$(3.1) \quad < \mathbf{v}_1, \mathbf{v}_2 > = \frac{1}{2} (\|\mathbf{v}_1 + \mathbf{v}_2\|^2 - \|\mathbf{v}_1\|^2 - \|\mathbf{v}_2\|^2)$$

La formula (3.1) è nota come *formula di polarizzazione*.

In geometria elementare il teorema di Carnot è noto come teorema del coseno. Per spiegare il nesso con il teorema di Carnot astratto dobbiamo introdurre il coseno dell'angolo tra due vettori di uno spazio euclideo. Il coseno è legato alla proiezione ortogonale di un vettore nella direzione di un altro vettore. Cominciamo quindi con il descrivere tale proiezione ortogonale. Definiamo la *distanza di due vettori* \mathbf{v} e \mathbf{w} come la lunghezza $\|\mathbf{w} - \mathbf{v}\|$ della loro differenza. Intuitivamente questa definizione si giustifica così: se $\mathbf{v} = \overrightarrow{OP}$ e $\mathbf{w} = \overrightarrow{OQ}$ sono i vettori posizione di due punti P e Q , la distanza $\|\mathbf{w} - \mathbf{v}\|$ è la lunghezza del vettore $\mathbf{w} - \mathbf{v} = \overrightarrow{PQ}$ che congiunge P e Q , come nello spazio cartesiano.

Figura 3.1. La distanza tra \mathbf{v} e \mathbf{w} è la norma di $\mathbf{w} - \mathbf{v}$.**PROPOSIZIONE 3.2 (Proiezione ortogonale su una retta)**

Sia \mathbf{V} uno spazio euclideo, sia $\mathbf{b} \in \mathbf{V}$ un vettore non nullo e sia $\mathbf{L} = \mathbb{R}\mathbf{b}$ la retta generata da \mathbf{b} . Il vettore

$$(3.2) \quad \mathbf{c} = \frac{\langle \mathbf{v}, \mathbf{b} \rangle}{\|\mathbf{b}\|^2} \mathbf{b}$$

gode delle seguenti proprietà:

- a) \mathbf{c} appartiene a \mathbf{L} ;
- b) $\mathbf{v} - \mathbf{c}$ è ortogonale a \mathbf{b} (e quindi a ogni vettore di \mathbf{L});
- c) \mathbf{c} è il vettore di \mathbf{L} a distanza minima da \mathbf{v} :

$$(3.3) \quad \|\mathbf{v} - \mathbf{c}\| < \|\mathbf{v} - \mathbf{w}\| \quad \text{per ogni } \mathbf{w} \in \mathbf{L}, \mathbf{w} \neq \mathbf{c}$$

Si dice che \mathbf{c} è la proiezione ortogonale di \mathbf{v} su \mathbf{L} e si denota con il simbolo $\mathbf{v}_{\mathbf{L}}$: dipende solo dalla retta \mathbf{L} e dal vettore \mathbf{v} , non dipende dalla scelta del vettore $\{\mathbf{b}\}$ di \mathbf{L} .

DIMOSTRAZIONE. Spieghiamo dapprima come si arriva alla formula (3.2) per la proiezione. Intuitivamente (identifichiamo i punti dello spazio coi loro vettori posizione) il vettore \mathbf{c} di \mathbf{L} a distanza minima da \mathbf{v} è il piede della perpendicolare tracciata da \mathbf{v} a \mathbf{L} ; questo significa che ci aspettiamo che $\mathbf{c} - \mathbf{b}$ sia perpendicolare a \mathbf{b} , come nel punto b) dell'enunciato. Cerchiamo quindi di determinare un vettore \mathbf{c} di \mathbf{L} tale che $\mathbf{v} - \mathbf{c}$ sia perpendicolare a \mathbf{b} . Siccome la retta \mathbf{L} consiste dei multipli scalari di \mathbf{b} , il vettore \mathbf{c} ha la forma $\hat{x}\mathbf{b}$ per uno scalare \hat{x} ; dobbiamo quindi cercare uno scalare \hat{x} tale che $\mathbf{v} - \hat{x}\mathbf{b}$ sia perpendicolare a \mathbf{b} . Ora

$$(3.4) \quad \langle \mathbf{v} - \hat{x}\mathbf{b}, \mathbf{b} \rangle = \langle \mathbf{v}, \mathbf{b} \rangle - \hat{x} \langle \mathbf{b}, \mathbf{b} \rangle$$

Ne segue che $\mathbf{v} - \hat{x}\mathbf{b}$ è ortogonale a \mathbf{b} se e solo se $\hat{x} = \frac{\langle \mathbf{v}, \mathbf{b} \rangle}{\|\mathbf{b}\|^2}$. Definiamo quindi \hat{x} mediante questa formula, e poniamo $\mathbf{c} = \hat{x}\mathbf{b}$. Per costruzione \mathbf{c} verifica a) e b).

Dobbiamo ancora mostrare la proprietà di minimo. Per questo sia \mathbf{w} un vettore di \mathbf{L} , cioè un multiplo scalare di \mathbf{v} . La differenza $\mathbf{c} - \mathbf{w}$ è ancora un multiplo scalare di \mathbf{b} . Per il punto b), il vettore $\mathbf{v} - \mathbf{c}$ è ortogonale a \mathbf{b} e quindi anche a $\mathbf{c} - \mathbf{w}$ che è un multiplo di \mathbf{b} . La diseguaglianza (3.3) ora segue dal fatto che $\mathbf{v} - \mathbf{w}$ è l'ipotenusa – si veda la figura 3.2 – del triangolo rettangolo che ha per cateti $\mathbf{v} - \mathbf{c}$ e $\mathbf{c} - \mathbf{w}$. Per essere rigorosi, per il teorema di

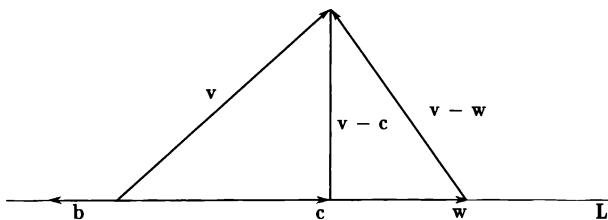


Figura 3.2. La proiezione ortogonale $c = \hat{c}b$ di v sulla retta generata da b .

Pitagora:

$$\|v - w\|^2 = \|(v - c) + (c - w)\|^2 = \|v - c\|^2 + \|c - w\|^2$$

Se $w \neq c$, allora $\|c - w\| > 0$, e quindi

$$\|v - w\|^2 = \|v - c\|^2 + \|c - w\|^2 > \|v - c\|^2$$

che è equivalente alla (3.3) perché la norma è una funzione non negativa.

Infine si osservi che la (3.3) ci dice che c è l'unico vettore di L a distanza minima da v ; pertanto c dipende solo da v e da L ed è univocamente determinato dalla diseguaglianza (3.3).

DEFINIZIONE 3.3 (Coefficienti di Fourier)

Dati due vettori v e $b \neq 0$ in uno spazio euclideo V , lo scalare

$$(3.5) \quad \frac{\langle v, b \rangle}{\|b\|^2}$$

si dice *coefficiente di Fourier di v rispetto a b*.



Nel caso particolare in cui $v = tb$ appartenga alla retta L generata da b , la proiezione ortogonale di v su L è v , e quindi $t = \frac{\langle v, b \rangle}{\|b\|^2}$: la coordinata t di v rispetto alla base b di L è il coefficiente di Fourier di v rispetto a b .

OSSERVAZIONE Se si usa un versore q come base di L , la formula per la proiezione ortogonale si semplifica perché $\|q\| = 1$:

$$(3.6) \quad v_L = \langle v, q \rangle q$$

Questo è in accordo con quanto visto nel primo capitolo, quando il termine $\langle v, q \rangle$ era interpretato come il prodotto della lunghezza di v per il coseno dell'angolo formato da v e q .

Supponiamo che \mathbf{V} sia \mathbb{R}^n con il prodotto interno standard $\mathbf{v}^T \mathbf{b}$. Allora la formula per la proiezione ortogonale sulla retta $\mathbf{L} = \mathbb{R}\mathbf{b}$ è:

$$(3.7) \quad \mathbf{v}_L = \frac{\mathbf{v}^T \mathbf{b}}{\mathbf{b}^T \mathbf{b}} \mathbf{b}$$

Per un esempio specifico, consideriamo in \mathbb{R}^3 il vettore $\mathbf{b} = [1, 2, 3]^T$. Il coefficiente di Fourier di $\mathbf{v} = [x, y, z]^T$ rispetto a \mathbf{b} è

$$\hat{x} = \frac{\mathbf{v}^T \mathbf{b}}{\mathbf{b}^T \mathbf{b}} = \frac{x + 2y + 3z}{1 + 4 + 9} = \frac{x + 2y + 3z}{14}$$

e la proiezione ortogonale di \mathbf{v} sulla retta \mathbf{L} è perciò:

$$\mathbf{v}_L = \hat{x} \mathbf{b} = \frac{1}{14} \begin{bmatrix} x + 2y + 3z \\ 2x + 4y + 6z \\ 3x + 6y + 9z \end{bmatrix}$$



Supponiamo che \mathbf{V} sia lo spazio vettoriale delle funzioni continue in $[0, 2\pi]$ con la norma L^2 . I classici coefficienti di Fourier a_k e b_k di una funzione $f \in \mathbf{V}$ sono i coefficienti di Fourier di $f(x)$ rispetto alle funzioni $\cos(kx)$ e $\sin(kx)$. Un semplice calcolo mostra che per $k \geq 1$

$$\| \cos(kx) \|^2 = \int_0^{2\pi} \cos^2(kx) dx = \pi = \| \sin(kx) \|^2$$

mentre la funzione costante $1 = \cos(0x)$ ha norma $\sqrt{2\pi}$. Perciò il coefficiente di Fourier di $f(x)$ rispetto alla funzione costante 1 è

$$\frac{1}{2\pi} \int_0^{2\pi} f(x) dx$$

mentre per $k \geq 1$ i coefficienti di Fourier a_k e b_k di $f(x)$ rispetto alle funzioni $\cos(kx)$ e $\sin(kx)$ sono rispettivamente

$$\begin{aligned} a_k &= \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(kx) dx \\ b_k &= \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(kx) dx \end{aligned}$$

La diseguaglianza di Schwarz e l'angolo tra due vettori

Vogliamo ora introdurre l'angolo tra due vettori di uno spazio euclideo. Nel contesto geometrico dello spazio cartesiano, il prodotto scalare è definito dalla formula

$$\mathbf{v} \cdot \mathbf{w} = \| \mathbf{v} \| \| \mathbf{w} \| \cos(\alpha)$$

dove α è l'angolo formato dai due vettori. Da questa uguaglianza possiamo ricavare una definizione di $\cos(\alpha)$ in termini di prodotto scalare e norma che è valida in ogni spazio euclideo:

$$\cos(\alpha) = \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\| \mathbf{v} \| \| \mathbf{w} \|}$$

Perché questo abbia senso occorre accertarsi che il numero a secondo membro sia compreso tra -1 e 1 :

PROPOSIZIONE 3.4 (Disuguaglianza di Schwarz)

Sia \mathbf{V} uno spazio euclideo. Allora

$$(3.8) \quad |\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\| \|\mathbf{w}\| \quad \text{per ogni } \mathbf{v}, \mathbf{w} \in \mathbf{V},$$

e l'uguaglianza vale se e solo se \mathbf{v} e \mathbf{w} sono linearmente dipendenti.

DIMOSTRAZIONE. Se $\mathbf{w} = \mathbf{0}$, entrambi i membri della disuguaglianza sono uguali a zero, e \mathbf{v} e \mathbf{w} sono linearmente dipendenti, per cui l'enunciato è vero in questo caso.

Supponiamo ora che $\mathbf{w} \neq \mathbf{0}$. Sia $\hat{x} = \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{w}\|^2}$ il coefficiente di Fourier di \mathbf{v} rispetto a \mathbf{w} , in modo che $\hat{x}\mathbf{w}$ sia la proiezione ortogonale di \mathbf{v} sulla retta $\mathbb{R}\mathbf{w}$. Allora $\hat{x}\mathbf{w}$ e $\mathbf{v} - \hat{x}\mathbf{w}$ sono ortogonali e per il teorema di Pitagora

$$(3.9) \quad \|\mathbf{v}\|^2 = \|\hat{x}\mathbf{w}\|^2 + \|\mathbf{v} - \hat{x}\mathbf{w}\|^2 \geq \|\hat{x}\mathbf{w}\|^2 = |\hat{x}|^2 \|\mathbf{w}\|^2 = \frac{|\langle \mathbf{v}, \mathbf{w} \rangle|^2}{\|\mathbf{w}\|^2}$$

Moltiplicando la (3.9) per $\|\mathbf{w}\|^2$ si ottiene la disuguaglianza di Schwarz (3.8). L'uguaglianza vale nella (3.8) e nella (3.9) se e solo se $\mathbf{v} = \hat{x}\mathbf{w}$; questo significa che \mathbf{v} appartiene alla retta $\mathbb{R}\mathbf{w}$ e, poiché \mathbf{w} è diverso da $\mathbf{0}$, equivale a dire che \mathbf{v} e \mathbf{w} sono linearmente dipendenti.

OSSERVAZIONE La disuguaglianza di Schwarz è tra le disuguaglianze più celebri e più ricche di applicazioni della matematica. Non deve perciò sorprendere che sia nota anche in altro modo: in altri testi è detta *disuguaglianza di Cauchy-Schwarz*, in altri ancora *disuguaglianza di Bunjakowski*.

Nel caso del prodotto scalare standard di \mathbb{R}^n la disuguaglianza di Schwarz in termini delle componenti è

$$|x_1 y_1 + \cdots + x_n y_n| \leq \sqrt{x_1^2 + \cdots + x_n^2} \sqrt{y_1^2 + \cdots + y_n^2}$$

Siano $\mathbf{s} = \{a_n\}_{n \geq 0}$ e $\mathbf{t} = \{b_n\}_{n \geq 0}$ due successioni in ℓ_2 : questo significa che

$$a = \sum_{n=0}^{+\infty} a_n^2 < +\infty \quad \text{e} \quad b = \sum_{n=0}^{+\infty} b_n^2 < +\infty$$

Per ogni $N \geq 0$, la disuguaglianza di Schwarz in \mathbb{R}^N ci dice che

$$\sum_{n=0}^N |a_n b_n| \leq \sqrt{\sum_{n=0}^N a_n^2} \sqrt{\sum_{n=0}^N b_n^2} \leq \sqrt{\sum_{n=0}^{+\infty} a_n^2} \sqrt{\sum_{n=0}^{+\infty} b_n^2} = \sqrt{a} \sqrt{b}$$

Questo mostra che la successione delle somme parziali della serie $\sum_{n=0}^{+\infty} |a_n b_n|$ è limitata e, quindi, che la serie $\sum_{n=0}^{+\infty} a_n b_n$ converge assolutamente. Il prodotto scalare in ℓ_2 è quindi ben definito.

Per la norma L^2 sullo spazio delle funzioni continue in $[a, b]$ la disuguaglianza di Schwarz è

$$\left| \int_a^b f(x)g(x) dx \right| \leq \sqrt{\int_a^b f(x)^2 dx} \sqrt{\int_a^b g(x)^2 dx}$$

Possiamo ora definire l'angolo tra due vettori non nulli:

DEFINIZIONE 3.5 (Angolo formato da due vettori)

Sia \mathbf{V} uno spazio euclideo con prodotto scalare $\langle \cdot, \cdot \rangle$. Dati due vettori non nulli $\mathbf{v}, \mathbf{w} \in \mathbf{V}$, l'angolo formato da \mathbf{v} e \mathbf{w} è

$$\widehat{\mathbf{vw}} = \arccos \left(\frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{v}\| \|\mathbf{w}\|} \right)$$

Due vettori non nulli di uno spazio euclideo sono ortogonali se e solo se l'angolo da essi formato è $\frac{\pi}{2}$.

Dal teorema di Carnot si ricava immediatamente il teorema del coseno: se \mathbf{v} e \mathbf{w} sono due vettori non nulli, allora

$$\|\mathbf{v} - \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - 2\|\mathbf{v}\| \|\mathbf{w}\| \cos(\widehat{\mathbf{vw}})$$

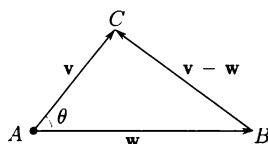


Figura 3.3. Il teorema del coseno: $\overline{BC}^2 = \overline{AB}^2 + \overline{AC}^2 - 2 \overline{AB} \overline{AC} \cos(\theta)$.

La disuguaglianza triangolare

Nel piano cartesiano, due vettori $\mathbf{v} \equiv \overrightarrow{OP}$ e $\mathbf{w} \equiv \overrightarrow{PQ}$ insieme al vettore somma $\mathbf{v} + \mathbf{w} \equiv \overrightarrow{OQ}$ costituiscono i tre lati del triangolo OPQ . La disuguaglianza triangolare afferma che la lunghezza del lato $\mathbf{v} + \mathbf{w}$ del triangolo è minore della somma delle lunghezze degli altri due lati. Tale disuguaglianza è valida in generale ed è in effetti equivalente alla disuguaglianza di Schwarz:

PROPOSIZIONE 3.6 (Proprietà della norma)

Sia \mathbf{V} uno spazio euclideo. La funzione norma gode delle seguenti proprietà:

a) **Disuguaglianza triangolare:**

$$(3.10) \quad \|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\| \quad \text{per ogni } \mathbf{v}, \mathbf{w} \in \mathbf{V}$$

e l'uguaglianza vale se e solo se \mathbf{v} e \mathbf{w} sono linearmente dipendenti e $\langle \mathbf{v}, \mathbf{w} \rangle \geq 0$ (intuitivamente questo significa che i due vettori sono paralleli e hanno lo stesso verso).

b) **Omogeneità**

$$(3.11) \quad \|t\mathbf{v}\| = |t| \|\mathbf{v}\| \text{ per ogni } t \in \mathbb{R} \text{ e ogni } \mathbf{v} \in \mathbf{V}$$

c) **Annulloamento:**

$$(3.12) \quad \|\mathbf{v}\| = 0 \quad \text{se e solo se } \mathbf{v} = 0$$

DIMOSTRAZIONE. La disuguaglianza triangolare segue dalla disuguaglianza di Schwarz:

$$\begin{aligned} \|\mathbf{v} + \mathbf{w}\|^2 &= \|\mathbf{v}\|^2 + 2 \langle \mathbf{v}, \mathbf{w} \rangle + \|\mathbf{w}\|^2 \leq \\ &= \|\mathbf{v}\|^2 + 2 \|\mathbf{v}\| \|\mathbf{w}\| + \|\mathbf{w}\|^2 = (\|\mathbf{v}\| + \|\mathbf{w}\|)^2 \end{aligned}$$

Questi stessi conti mostrano che l'uguaglianza vale se e solo se

$$\langle \mathbf{v}, \mathbf{w} \rangle = |\langle \mathbf{v}, \mathbf{w} \rangle| = \|\mathbf{v}\| \|\mathbf{w}\|$$

quindi se e solo se $\langle \mathbf{v}, \mathbf{w} \rangle \geq 0$ e vale l'uguaglianza in Schwarz, cioè \mathbf{v} e \mathbf{w} sono linearmente dipendenti. Abbiamo già dimostrato le proprietà di omogeneità e annulloamento a pagina 367.

OSSERVAZIONE Abbiamo definito la *distanza di due vettori* \mathbf{v} e \mathbf{w} come la lunghezza $\|\mathbf{w} - \mathbf{v}\|$ della loro differenza. Avendo a disposizione la nozione di distanza, è possibile definire il concetto di limite: si dice che

$$\lim_{n \rightarrow +\infty} \mathbf{v}_n = \mathbf{v}$$

se la distanza tra \mathbf{v}_n e \mathbf{v} tende a zero:

$$\lim_{n \rightarrow +\infty} \|\mathbf{v}_n - \mathbf{v}\| = 0$$

La disuguaglianza triangolare è quanto serve per l'algebra dei limiti: da essa segue per esempio che il limite di una somma è la somma dei limiti. Diventa così possibile fare dell'analisi in uno spazio euclideo anche di dimensione infinita.



Esercizio 1 Sia $\mathbf{v} = [2, 3]^T$. Per quali valori di m il vettore $\mathbf{w} = [1, m]^T$ è ortogonale a \mathbf{v} ? Verificare che $\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2$ solo per tali valori.

Esercizio 2 Calcolare (il coseno del) l'angolo formato dai vettori $[1, 0, 0, 0]^T$ e $[2, 2, 2, 2]^T$ (ripetere per i vettori $[0, 0, 0, 2]^T$ e $[1, 1, 1, 1]^T$).

Esercizio 3 In \mathbb{R}^4 trovare la proiezione ortogonale di $\mathbf{v} = [1, 2, 3, 4]^T$ sulla retta \mathbf{L} di \mathbb{R}^4 generata da $[1, 1, 1, 1]^T$, e scrivere \mathbf{v} come somma di un vettore di \mathbf{L} e di un vettore ortogonale a \mathbf{L} . Calcolare la distanza di \mathbf{v} da \mathbf{L} .

Dimostrare che in uno spazio euclideo \mathbf{V} vale la *legge del parallelogramma*:

$$\|\mathbf{v} + \mathbf{w}\|^2 + \|\mathbf{v} - \mathbf{w}\|^2 = 2\|\mathbf{v}\|^2 + 2\|\mathbf{w}\|^2 \quad \text{per ogni } \mathbf{v}, \mathbf{w} \in \mathbf{V}$$

Fare un disegno e spiegare perché si chiama legge del parallelogramma.

Se $\|\mathbf{v}\| = \|\mathbf{w}\| = 1$ e $\|\mathbf{v} - \mathbf{w}\| = \sqrt{3}$, qual è il coseno dell'angolo formato da \mathbf{v} e \mathbf{w} ?

Esercizio 4 È possibile che in uno spazio euclideo esistano due vettori \mathbf{v} e \mathbf{w} tali che $\|\mathbf{v}\| = \|\mathbf{w}\| = 1$ e $\|\mathbf{v} + \mathbf{w}\| = 3$?

Sia \mathbf{V} lo spazio vettoriale delle funzioni continue in $[0, 2\pi]$ con la norma L^2 . Calcolare la norma della funzione $\cos(x) + \sin(x)$ e verificare il teorema di Pitagora per le due funzioni $\cos(x)$ e $\sin(x)$, che sono tra loro ortogonali e hanno entrambe norma $\sqrt{\pi}$ per un esercizio precedente.

4 BASI ORTONORMALI E MATRICI ORTOGONALI

Avendo a disposizione le nozioni di lunghezza e ortogonalità, possiamo ora introdurre sistemi di coordinate sul modello cartesiano, richiedendo che gli assi coordinati siano tra loro ortogonali e che l'unità di misura sia la stessa su ciascun asse. Fissare un sistema di coordinate in uno spazio vettoriale significa fissare una base: la richiesta da fare è che i vettori della base siano dei versori a due a due perpendicolari: il fatto che siano dei versori, cioè vettori di lunghezza uno, intuitivamente pari all'unità di misura, garantisce che su ogni asse ci sia la stessa unità di misura. Una base siffatta si dice *ortonormale* ed è l'analogo della terna cartesiana **i, j, k**. Più in generale è utile considerare anche basi, dette *ortogonali*, in cui si richiede solamente l'ortogonalità, senza insistere sul fatto che gli elementi della base siano dei versori: questo perché per ottenere un versore da un vettore dato occorre dividere per la sua norma, il cui calcolo richiede l'estrazione di una radice quadrata e può quindi creare delle complicazioni non necessarie.

DEFINIZIONE 4.1 (Basi ortogonali e basi ortonormali)

Una base $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ di uno spazio euclideo \mathbf{V} si dice *ortogonale* se i suoi elementi sono a due a due ortogonali:

$$\langle \mathbf{b}_i, \mathbf{b}_j \rangle = 0 \quad \text{per ogni } i \neq j$$

Una base $\{\mathbf{q}_1, \dots, \mathbf{q}_n\}$ di uno spazio euclideo \mathbf{V} si dice *ortonormale* se è ortogonale e i suoi elementi sono *versori*; questo equivale a richiedere che

$$\langle \mathbf{q}_i, \mathbf{q}_j \rangle = \delta_{ij} \quad \text{per ogni } i, j = 1, \dots, n$$

dove δ_{ij} è il simbolo di Kronecker, che vale 1 se $i = j$, e 0 se $i \neq j$.

OSSERVAZIONE Spieghiamo l'equivalenza delle due condizioni date per la definizione di base ortonormale. La condizione $\langle \mathbf{q}_i, \mathbf{q}_j \rangle = \delta_{ij}$, quando $i = j$, significa che \mathbf{q}_i è un versore:

$$\|\mathbf{q}_i\|^2 = \langle \mathbf{q}_i, \mathbf{q}_i \rangle = \delta_{ii} = 1$$

Quando $i \neq j$, la condizione è

$$\langle \mathbf{q}_i, \mathbf{q}_j \rangle = \delta_{ij} = 0$$

e significa che \mathbf{q}_i e \mathbf{q}_j sono ortogonali.

La base canonica $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ di \mathbb{R}^n è una base ortonormale rispetto al prodotto scalare standard.

Se $\mathbf{i}, \mathbf{j}, \mathbf{k}$ sono tre versori di \mathbb{R}^3 a due a due perpendicolari, l'insieme $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ è una base ortonormale di \mathbb{R}^3 .

OSSERVAZIONE Da ogni base ortogonale $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ si può ottenere una base ortonormale sostituendo ogni \mathbf{b}_i con il corrispondente versore

$$\mathbf{q}_i = \frac{\mathbf{b}_i}{\|\mathbf{b}_i\|} \quad (i = 1, 2, \dots, n)$$

Dalla base ortogonale di \mathbb{R}^2 formata da $\mathbf{b}_1 = [1, 1]^T$ e $\mathbf{b}_2 = [-1, 1]^T$ ricaviamo la base ortonormale formata da

$$\mathbf{q}_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} \cos(\frac{\pi}{4}) \\ \sin(\frac{\pi}{4}) \end{bmatrix} \quad \text{e} \quad \mathbf{q}_2 = \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} -\sin(\frac{\pi}{4}) \\ \cos(\frac{\pi}{4}) \end{bmatrix}$$

Il teorema di Pitagora si estende al caso della somma di un numero arbitrario di vettori a due a due ortogonali; come conseguenza, un insieme di vettori non nulli che siano a due a due ortogonali è automaticamente linearmente indipendente:

PROPOSIZIONE 4.2 (Teorema di Pitagora generalizzato)

Sia \mathbf{V} uno spazio euclideo e siano $\mathbf{v}_1, \dots, \mathbf{v}_d$ vettori di \mathbf{V} .

- a) se i vettori $\mathbf{v}_1, \dots, \mathbf{v}_d$ sono a due a due ortogonali, allora

$$(4.1) \quad \|\mathbf{v}_1 + \cdots + \mathbf{v}_d\|^2 = \|\mathbf{v}_1\|^2 + \cdots + \|\mathbf{v}_d\|^2$$

- b) se i vettori $\mathbf{v}_1, \dots, \mathbf{v}_d$ sono non nulli e a due a due ortogonali, allora sono linearmente indipendenti.

DIMOSTRAZIONE. Con un conto analogo a quello del teorema di Carnot si ottiene l'uguaglianza

$$\|\mathbf{v}_1 + \cdots + \mathbf{v}_d\|^2 = \|\mathbf{v}_1\|^2 + \cdots + \|\mathbf{v}_d\|^2 + 2 \sum_{1 \leq i < j \leq d} \langle \mathbf{v}_i, \mathbf{v}_j \rangle$$

Quando i vettori sono a due a due ortogonali, quest'uguaglianza diviene la (4.1).

Mostriamo ora b): supponiamo che $\mathbf{v}_1, \dots, \mathbf{v}_d$ siano non nulli e a due a due ortogonali e supponiamo ci sia tra di loro una relazione di dipendenza lineare

$$(4.2) \quad t_1 \mathbf{v}_1 + t_2 \mathbf{v}_2 + \cdots + t_d \mathbf{v}_d = \mathbf{0}$$

I vettori $t_1 \mathbf{v}_1, t_2 \mathbf{v}_2, \dots, t_d \mathbf{v}_d$ sono a due a due ortogonali perché $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d$ lo sono. Quindi per il punto a):

$$\begin{aligned} 0 &= \|\mathbf{0}\|^2 = \|t_1 \mathbf{v}_1 + t_2 \mathbf{v}_2 + \cdots + t_d \mathbf{v}_d\|^2 = \\ &= \|t_1 \mathbf{v}_1\|^2 + \|t_2 \mathbf{v}_2\|^2 + \cdots \|t_d \mathbf{v}_d\|^2 = t_1^2 \|\mathbf{v}_1\|^2 + t_2^2 \|\mathbf{v}_2\|^2 + \cdots t_d^2 \|\mathbf{v}_d\|^2 \end{aligned}$$

Al membro di destra abbiamo una somma di numeri non negativi la cui somma è zero; tutti gli addendi $t_k^2 \|\mathbf{v}_k\|^2$ devono perciò essere nulli. Siccome \mathbf{v}_k per ipotesi è non nullo, concludiamo che $t_k = 0$ per ogni $k = 1, 2, \dots, d$. Questo mostra che i vettori $\mathbf{v}_1, \dots, \mathbf{v}_d$ sono linearmente indipendenti. □

OSSERVAZIONE Se \mathbf{V} ha dimensione n e i vettori $\mathbf{b}_1, \dots, \mathbf{b}_n$ sono a due a due ortogonali e non nulli (per esempio, dei versori), allora $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ è una base ortogonale di \mathbf{V} . Infatti, per la proposizione 4.2, i vettori sono linearmente indipendenti e, quindi, siccome sono $n = \dim(\mathbf{V})$, formano una base. Per controllare che dei vettori formino una base ortogonale occorre solo controllare che siano non nulli, a due a due ortogonali e in numero pari alla dimensione dello spazio.

Il grande vantaggio di una base ortogonale rispetto a una base qualsiasi è che le coordinate di un vettore possono essere ricavate facilmente in termini del prodotto scalare, esattamente come nel caso dello spazio cartesiano (capitolo primo, proposizione 6.4); è in accordo con l'intuizione geometrica che la componente di un vettore lungo un asse coordinato sia la proiezione ortogonale del vettore sull'asse.

PROPOSIZIONE 4.3 (Coordinate di un vettore rispetto a una base ortogonale) Sia $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ una base *ortogonale* di uno spazio euclideo \mathbf{V} . Per ogni vettore $\mathbf{v} \in \mathbf{V}$ vale l'uguaglianza:

$$(4.3) \quad \mathbf{v} = \hat{x}_1 \mathbf{b}_1 + \hat{x}_2 \mathbf{b}_2 + \cdots + \hat{x}_n \mathbf{b}_n$$

dove

$$(4.4) \quad \hat{x}_k = \frac{\langle \mathbf{v}, \mathbf{b}_k \rangle}{\|\mathbf{b}_k\|^2} \quad k = 1, 2, \dots, n$$

è il coefficiente di Fourier di \mathbf{v} rispetto a \mathbf{b}_k . In altre parole, le coordinate di \mathbf{v} rispetto alla base \mathcal{B} sono i coefficienti di Fourier $\hat{x}_1, \dots, \hat{x}_n$ e \mathbf{v} è la somma delle sue proiezioni ortogonali sugli assi coordinati $\mathbb{R}\mathbf{b}_k$.

DIMOSTRAZIONE. Siano x_1, \dots, x_n le coordinate di \mathbf{v} rispetto alla base \mathcal{B} : questo significa che $\mathbf{v} = x_1 \mathbf{b}_1 + \cdots + x_n \mathbf{b}_n = \sum_{i=1}^n x_i \mathbf{b}_i$. Per la linearità del prodotto scalare nel primo fattore

$$\langle \mathbf{v}, \mathbf{b}_k \rangle = \left\langle \sum_{i=1}^n x_i \mathbf{b}_i, \mathbf{b}_k \right\rangle = \sum_{i=1}^n x_i \langle \mathbf{b}_i, \mathbf{b}_k \rangle$$

Siccome \mathbf{b}_k è perpendicolare a \mathbf{b}_i se $i \neq k$, nella sommatoria a secondo membro tutti gli addendi sono nulli tranne quello che corrisponde a $i = k$. Quindi

$$\langle \mathbf{v}, \mathbf{b}_k \rangle = x_k \langle \mathbf{b}_k, \mathbf{b}_k \rangle = x_k \|\mathbf{b}_k\|^2$$

Per ipotesi \mathbf{b}_k non è nullo, quindi la sua norma $\|\mathbf{b}_k\|$ è diversa da zero, e possiamo perciò ricavare $x_k = \frac{\langle \mathbf{v}, \mathbf{b}_k \rangle}{\|\mathbf{b}_k\|^2} = \hat{x}_k$.

Consideriamo la base ortogonale $\{\mathbf{b}_1 = [1, 1]^T, \mathbf{b}_2 = [-1, 1]^T\}$ di \mathbb{R}^2 . I coefficienti di Fourier del vettore $\mathbf{v} = [3, 2]^T$ rispetto ai vettori di questa base sono

$$\hat{x}_1 = \frac{\langle \mathbf{v}, \mathbf{b}_1 \rangle}{\|\mathbf{b}_1\|^2} = \frac{5}{2} \quad \text{e} \quad \hat{x}_2 = \frac{\langle \mathbf{v}, \mathbf{b}_2 \rangle}{\|\mathbf{b}_2\|^2} = -\frac{1}{2}$$

Verifichiamo l'uguaglianza $\hat{x}_1 \mathbf{b}_1 + \hat{x}_2 \mathbf{b}_2 = \mathbf{v}$:

$$\frac{5}{2}[1, 1]^T - \frac{1}{2}[-1, 1]^T = \frac{1}{2}[6, 4]^T = [3, 2]^T$$

COROLLARIO 4.4 Sia $\mathcal{B} = \{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$ una base ortonormale di uno spazio euclideo \mathbf{V} . Allora:

- a) le coordinate di un vettore \mathbf{v} rispetto alla base \mathcal{B} sono i coefficienti di Fourier $\hat{x}_k = \langle \mathbf{v}, \mathbf{q}_k \rangle$ per $k = 1, 2, \dots, n$:

$$\mathbf{v} = \langle \mathbf{v}, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{v}, \mathbf{q}_2 \rangle \mathbf{q}_2 + \dots + \langle \mathbf{v}, \mathbf{q}_n \rangle \mathbf{q}_n$$

- b) la norma di un vettore \mathbf{v} è uguale alla norma \mathbb{R}^n del vettore delle sue coordinate:

$$\|\mathbf{v}\| = \sqrt{\hat{x}_1^2 + \hat{x}_2^2 + \dots + \hat{x}_n^2}$$

- c) dati due vettori $\mathbf{v}, \mathbf{w} \in \mathbf{V}$ di coordinate $\{\hat{x}_k\}$ e $\{\hat{y}_k\}$, il prodotto scalare di \mathbf{v} e \mathbf{w} in \mathbf{V} è uguale al prodotto scalare in \mathbb{R}^n dei vettori delle coordinate:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \hat{x}_1 \hat{y}_1 + \hat{x}_2 \hat{y}_2 + \dots + \hat{x}_n \hat{y}_n$$

DIMOSTRAZIONE. Il punto a) segue dalla proposizione precedente perché una base ortonormale è in particolare una base ortogonale. Il punto b) segue dal teorema di Pitagora perché i vettori $\hat{x}_k \mathbf{q}_k$ sono a due a due perpendicolari:

$$\|\mathbf{v}\|^2 = \|\hat{x}_1 \mathbf{q}_1 + \hat{x}_2 \mathbf{q}_2 + \dots + \hat{x}_n \mathbf{q}_n\|^2 = \hat{x}_1^2 \|\mathbf{q}_1\|^2 + \hat{x}_2^2 \|\mathbf{q}_2\|^2 + \dots + \hat{x}_n^2 \|\mathbf{q}_n\|^2 = \hat{x}_1^2 + \hat{x}_2^2 + \dots + \hat{x}_n^2$$

L'uguaglianza dei prodotti scalari segue dall'uguaglianza delle norme per la formula di polarizzazione (3.1). In alternativa, la si può dimostrare direttamente sfruttando la bilinearità del prodotto scalare:

$$\begin{aligned} \langle \mathbf{v}, \mathbf{w} \rangle &= \left\langle \sum_{i=1}^n \hat{x}_i \mathbf{q}_i, \sum_{j=1}^n \hat{y}_j \mathbf{q}_j \right\rangle = \sum_{i=1}^n \sum_{j=1}^n \hat{x}_i \hat{y}_j \langle \mathbf{q}_i, \mathbf{q}_j \rangle = \\ &= \sum_{i=1}^n \sum_{j=1}^n \hat{x}_i \hat{y}_j \delta_{ij} = \sum_{i=1}^n \hat{x}_i \hat{y}_i \end{aligned}$$

OSSERVAZIONE Supponiamo che $\mathcal{B} = \{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$ sia una base ortonormale di uno spazio euclideo \mathbf{V} e consideriamo il corrispondente isomorfismo $\mathbf{V} \rightarrow \mathbb{R}^n$ che a un vettore $\mathbf{v} \in \mathbf{V}$ associa il vettore $\hat{\mathbf{v}} = \mathbf{x}(\mathbf{v}) \in \mathbb{R}^n$ delle sue coordinate rispetto alla base \mathcal{B} . Per la proposizione precedente questa identificazione di \mathbf{V} e \mathbb{R}^n preserva anche il prodotto scalare, nel senso che

$$\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{V}} = \hat{\mathbf{v}}^T \hat{\mathbf{w}} = \langle \hat{\mathbf{v}}, \hat{\mathbf{w}} \rangle_{\mathbb{R}^n}$$

Quindi una base ortonormale consente di identificare \mathbf{V} con \mathbb{R}^n non solo come spazio vettoriale, ma anche come spazio euclideo.



Consideriamo la base ortonormale $\{\mathbf{q}_1, \mathbf{q}_2\}$ di \mathbb{R}^2 ottenuta ruotando la base canonica di 45° :

$$\mathbf{q}_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}, \quad \mathbf{q}_2 = \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$$

Le componenti del vettore $\mathbf{v} = [3, 2]^T$ rispetto a questa base sono

$$\hat{x}_1 = \langle \mathbf{v}, \mathbf{q}_1 \rangle = \frac{5\sqrt{2}}{2} \quad \text{e} \quad \hat{x}_2 = \langle \mathbf{v}, \mathbf{q}_2 \rangle = -\frac{\sqrt{2}}{2}$$

Verifichiamo che la norma del vettore coincida con la norma del vettore delle coordinate:

$$\| [3, 2]^T \| = 9 + 4 = 13, \quad \| \left[5\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2} \right]^T \| = \frac{25}{2} + \frac{1}{2} = 13$$



In questo esempio determiniamo una base ortonormale del piano \mathbf{H} di equazione $x+y+z=0$ in \mathbb{R}^3 . Il vettore $\mathbf{v}_1 = [1, -1, 0]^T$ appartiene ad \mathbf{H} . Un vettore $\mathbf{w} = [a, b, c]^T$ è ortogonale a \mathbf{v}_1 e appartiene ad \mathbf{H} se e solo se

$$0 = \mathbf{v}^T \mathbf{w} = a - b \quad \text{e} \quad a + b + c = 0$$

Le soluzioni di questo sistema sono i vettori della forma $[a, a, -2a]^T$. In particolare, $\mathbf{v}_2 = [1, 1, -2]^T$ appartiene ad \mathbf{H} ed è ortogonale a \mathbf{v}_1 : i vettori \mathbf{v}_1 e \mathbf{v}_2 formano una base ortogonale di \mathbf{H} . Per ottenere una base ortonormale prendiamo i versori con la stessa direzione e verso di \mathbf{v}_1 e \mathbf{v}_2 :

$$\mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \frac{1}{\sqrt{2}}[1, -1, 0]^T \quad \text{e} \quad \mathbf{q}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \frac{1}{\sqrt{6}}[1, 1, -2]^T$$

I vettori \mathbf{q}_1 e \mathbf{q}_2 formano una base ortonormale di \mathbf{H} . Questa base è comoda se è necessario calcolare le lunghezze dei vettori di \mathbf{H} : per il corollario 4.4.b si ha infatti

$$\|x_1 \mathbf{q}_1 + x_2 \mathbf{q}_2\|^2 = x_1^2 + x_2^2$$

Matrici ortogonali

Introduciamo ora la classe delle matrici \mathbf{Q} le cui colonne formano una base *ortonormale* di \mathbb{R}^n .

PROPOSIZIONE 4.5 Sia \mathbf{Q} una matrice quadrata di ordine n a coefficienti reali. Le colonne di \mathbf{Q} formano una base ortonormale di \mathbb{R}^n se e solo se

$$\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$$

DIMOSTRAZIONE. Siano $\mathbf{q}_1, \dots, \mathbf{q}_n$ le colonne di \mathbf{Q} . Calcoliamo il prodotto di \mathbf{Q}^T e \mathbf{Q} :

$$\mathbf{Q}^T \mathbf{Q} = \begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{q}_2^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1^T \mathbf{q}_1 & \mathbf{q}_1^T \mathbf{q}_2 & \cdots & \mathbf{q}_1^T \mathbf{q}_n \\ \mathbf{q}_2^T \mathbf{q}_1 & \mathbf{q}_2^T \mathbf{q}_2 & \cdots & \mathbf{q}_2^T \mathbf{q}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{q}_n^T \mathbf{q}_1 & \mathbf{q}_n^T \mathbf{q}_2 & \cdots & \mathbf{q}_n^T \mathbf{q}_n \end{bmatrix}$$

Gli elementi sulla diagonale principale di $\mathbf{Q}^T \mathbf{Q}$ sono le norme al quadrato $\mathbf{q}_i^T \mathbf{q}_i$ delle colonne \mathbf{q}_i , mentre gli elementi fuori dalla diagonale sono i prodotti scalari $\mathbf{q}_i^T \mathbf{q}_j$ di due colonne distinte. Quindi le colonne sono a due a due ortogonali e hanno norma uno se e solo se $\mathbf{Q}^T \mathbf{Q}$ è la matrice identità. □

DEFINIZIONE 4.6 (Matrice ortogonale)

Una matrice quadrata \mathbf{Q} di ordine n a coefficienti reali si dice *ortogonale* se $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$.

Quindi una matrice ortogonale è una matrice le cui colonne formano una base *ortonormale* di \mathbb{R}^n : sarebbe più logico chiamare una tale matrice ortonormale, ma la tradizione lo impedisce.

PROPOSIZIONE 4.7

- a) La matrice identità è ortogonale.
- b) Il prodotto di due matrici ortogonali di ordine n è una matrice ortogonale.
- c) Una matrice ortogonale \mathbf{Q} è invertibile; la sua inversa è \mathbf{Q}^T ed è anch'essa ortogonale.

DIMOSTRAZIONE. La matrice identità è ortogonale perché $\mathbf{I}^T \mathbf{I} = \mathbf{I} \mathbf{I} = \mathbf{I}$; in particolare la base canonica di \mathbb{R}^n , che è costituita dalle colonne di \mathbf{I} , è ortonormale.

Se \mathbf{Q}_1 e \mathbf{Q}_2 sono ortogonali, allora

$$(\mathbf{Q}_1 \mathbf{Q}_2)^T \mathbf{Q}_1 \mathbf{Q}_2 = \mathbf{Q}_2^T \mathbf{Q}_1^T \mathbf{Q}_1 \mathbf{Q}_2 = \mathbf{Q}_2^T \mathbf{I} \mathbf{Q}_2 = \mathbf{Q}_2^T \mathbf{Q}_2 = \mathbf{I}$$

e quindi anche la matrice prodotto $\mathbf{Q}_1 \mathbf{Q}_2$ è ortogonale.

Se \mathbf{Q} è ortogonale, l'uguaglianza $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$ mostra che \mathbf{Q} è invertibile e $\mathbf{Q}^{-1} = \mathbf{Q}^T$. Quindi

$$(\mathbf{Q}^T)^T \mathbf{Q}^T = \mathbf{Q} \mathbf{Q}^T = \mathbf{Q} \mathbf{Q}^{-1} = \mathbf{I}.$$

Questo mostra che, se \mathbf{Q} è ortogonale, anche \mathbf{Q}^T è ortogonale. □

OSSERVAZIONE Una conseguenza importante è che calcolare l'inversa di una matrice ortogonale è immediato: basta trasporre la matrice. Si ricordi che in generale il calcolo

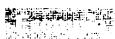
della matrice inversa è computazionalmente molto costoso. È notevole il fatto che, se \mathbf{Q} è ortogonale, anche \mathbf{Q}^T è ortogonale, e può apparire miracoloso: se le colonne di una matrice formano una base ortonormale di \mathbb{R}^n , allora automaticamente anche le colonne di \mathbf{Q}^T , cioè le righe di \mathbf{Q} , formano una base ortonormale.

OSSERVAZIONE Se \mathbf{Q} è una matrice ortogonale, il suo determinante è ± 1 . Infatti, ricordando che il determinante della trasposta di \mathbf{Q} è uguale al determinante di \mathbf{Q} e che il determinante di un prodotto è il prodotto dei determinanti, otteniamo:

$$1 = \det(\mathbf{I}) = \det(\mathbf{Q}^T \mathbf{Q}) = \det(\mathbf{Q})^2$$



Ogni matrice di permutazione \mathbf{P}_σ è una matrice ortogonale: le colonne di \mathbf{P}_σ sono, a meno dell'ordine, i vettori della base canonica e formano perciò una base ortonormale di \mathbb{R}^n , il che equivale a dire che \mathbf{P}_σ è ortogonale. Questo mostra nuovamente che \mathbf{P}_σ^T è l'inversa di \mathbf{P}_σ . Per lo stesso motivo, se \mathbf{Q} è una matrice ortogonale, ogni matrice ottenuta permutando le colonne (o le righe) di \mathbf{Q} è ancora ortogonale.



Matrici ortogonali di ordine 2

Supponiamo che \mathbf{Q} sia una matrice ortogonale di ordine 2. La prima colonna $[x, y]^T$ di \mathbf{Q} è un versore di \mathbb{R}^2 , cioè $x^2 + y^2 = 1$; geometricamente, il punto di coordinate (x, y) appartiene alla circonferenza di raggio 1 con centro nell'origine. I punti di questa circonferenza sono parametrizzati dall'angolo orientato θ che il semiasse positivo delle ascisse forma con il vettore $[x, y]^T$: esiste θ tale che $[x, y]^T = [\cos(\theta), \sin(\theta)]^T$. La seconda colonna di \mathbf{Q} è uno dei due versori della direzione perpendicolare alla prima colonna: questi versori sono $[-\sin(\theta), \cos(\theta)]^T$ e il suo opposto $[\sin(\theta), -\cos(\theta)]^T$. Quindi una matrice ortogonale è di una delle due seguenti forme

$$\mathbf{Q}_1 = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \quad \text{oppure} \quad \mathbf{Q}_2 = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix}$$

Abbiamo già incontrato queste matrici più volte: \mathbf{Q}_1 è la matrice che rappresenta la rotazione del piano attorno all'origine di un angolo θ in senso antiorario; la matrice \mathbf{Q}_2 rappresenta invece la riflessione ortogonale che ha per asse la retta \mathbf{r} generata dal vettore $[\cos(\theta/2), \sin(\theta/2)]^T$ (cf. pagina 245). Quindi una matrice ortogonale di ordine 2 è la matrice di una rotazione oppure di una riflessione ortogonale; i due casi sono distinti dal determinante: per le rotazioni il determinante è 1, per le riflessioni il determinante è -1 .



Costruire esempi di matrici ortogonali di ordine 3 (cioè basi ortonormali di \mathbb{R}^3) è più complesso; diventerà più semplice quando avremo a disposizione l'algoritmo di Gram-Schmidt. Supponiamo per esempio di voler determinare una base ortonormale di \mathbb{R}^3 i cui primi due elementi siano i vettori

$$\mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \frac{1}{\sqrt{2}}[1, -1, 0]^T \quad \text{e} \quad \mathbf{q}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \frac{1}{\sqrt{6}}[1, 1, -2]^T$$

che, come abbiamo visto a pagina 382, formano una base ortonormale del piano \mathbf{H} di equazione $x + y + z = 0$. Abbiamo bisogno di un versore \mathbf{q}_3 perpendicolare a \mathbf{q}_1 e \mathbf{q}_2 . Potremmo prendere il prodotto vettoriale $\mathbf{q}_1 \times \mathbf{q}_2$: questo comporta sostanzialmente calcolare un determinante 3×3 . Ma conosciamo già l'equazione $x + y + z = 0$ del piano che contiene \mathbf{q}_1 e \mathbf{q}_2 , e questo ci dice che il vettore $[1, 1, 1]^T$ è perpendicolare a \mathbf{q}_1 e \mathbf{q}_2 . Quindi definendo

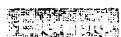
$$\mathbf{q}_3 = \frac{1}{\sqrt{3}}[1, 1, 1]^T$$

otteniamo la base ortonormale $\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$ di \mathbb{R}^3 . La corrispondente matrice ortogonale è:

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

I vettori riga di questa matrice formano un'altra base ortonormale di \mathbb{R}^3 :

$$\mathbf{q}'_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}, \quad \mathbf{q}'_2 = \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}, \quad \mathbf{q}'_3 = \begin{bmatrix} 0 \\ \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}.$$



La matrice

$$\frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ -1 & 1 & 1 & 1 \end{bmatrix}$$

è un esempio di una matrice ortogonale di ordine 4.



Abbiamo osservato che, se una matrice \mathbf{Q} è ortogonale, allora anche la matrice trasposta \mathbf{Q}^T è ortogonale. È facile, invece, trovare matrici che abbiano le colonne tra loro ortogonali, ma la cui trasposta non abbia le colonne tra loro ortogonali (le righe della matrici non sono ortogonali tra loro nonostante le colonne lo siano). In tali esempi necessariamente le colonne non sono versori. Un esempio è:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 1 & -2 \end{bmatrix}$$

/

Isometrie

Se pensiamo a una matrice quadrata \mathbf{A} in modo dinamico, come all'applicazione lineare $\mathcal{L}_{\mathbf{A}} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ che manda \mathbf{x} in \mathbf{Ax} , le matrici ortogonali corrispondono alle *isometrie*:

DEFINIZIONE 4.8 (Isometria)

Siano \mathbf{V}_1 e \mathbf{V}_2 due spazi euclidei. Si dice che una funzione $F : \mathbf{V}_1 \rightarrow \mathbf{V}_2$ è un'*isometria* se preserva le distanze:

$$\|F(\mathbf{v}) - F(\mathbf{w})\|_{\mathbf{V}_2} = \|\mathbf{v} - \mathbf{w}\|_{\mathbf{V}_1} \quad \text{per ogni } \mathbf{v}, \mathbf{w} \in \mathbf{V}_1$$

Si dice che $\mathfrak{T} : \mathbf{V}_1 \rightarrow \mathbf{V}_2$ è un'*isometria lineare* se è lineare ed è un'isometria. Si osservi che un'applicazione lineare $\mathfrak{T} : \mathbf{V}_1 \rightarrow \mathbf{V}_2$ è un'isometria se e solo se preserva la norma dei vettori:

$$\|\mathfrak{T}(\mathbf{v})\|_{\mathbf{V}_2} = \|\mathbf{v}\|_{\mathbf{V}_1} \quad \text{per ogni } \mathbf{v} \in \mathbf{V}_1$$

Le rototraslazioni del piano \mathbb{R}^2 (che sono i movimenti rigidi della geometria euclidea) sono delle isometrie. Tra queste solo le rotazioni sono lineari.

OSSERVAZIONE Un'isometria lineare preserva il prodotto scalare, nel senso che

$$(4.5) \quad \langle \mathfrak{T}(\mathbf{v}), \mathfrak{T}(\mathbf{w}) \rangle_{\mathbf{V}_2} = \langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{V}_1} \quad \text{per ogni } \mathbf{v}, \mathbf{w} \in \mathbf{V}_1$$

Questo segue dalla formula di polarizzazione (3.1), per la quale il prodotto scalare è determinato dalla norma e dalla linearità di \mathfrak{T} . Viceversa, un'applicazione lineare che preserva il prodotto scalare, per cui cioè vale la (4.5), è un'isometria: basta prendere $\mathbf{v} = \mathbf{w}$ nella (4.5).

OSSERVAZIONE Supponiamo che $\mathcal{B} = \{\mathbf{q}_1, \dots, \mathbf{q}_n\}$ sia una base ortonormale di \mathbf{V} , e sia $\mathfrak{T} : \mathbf{V} \rightarrow \mathbb{R}^n$ l'applicazione che a un vettore \mathbf{v} associa il vettore $\mathbf{x}(\mathbf{v})$ delle sue coordinate rispetto a \mathcal{B} . Dal corollario 4.4 segue che \mathfrak{T} è un'isometria:

$$\|\mathbf{v}\|_{\mathbf{V}} = \|\mathbf{x}(\mathbf{v})\|_{\mathbb{R}^n}$$

PROPOSIZIONE 4.9 (Matrici ortogonali rappresentano isometrie)

Per una matrice quadrata reale \mathbf{Q} di ordine n le seguenti condizioni sono equivalenti:

- a) \mathbf{Q} è ortogonale: $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$;
- b) \mathbf{Q} preserva la norma dei vettori:

$$\|\mathbf{Q}\mathbf{x}\| = \|\mathbf{x}\| \quad \text{per ogni } \mathbf{x} \in \mathbb{R}^n$$

(in altri termini, \mathbf{Q} rappresenta un'isometria di \mathbb{R}^n);

- c) \mathbf{Q} preserva il prodotto scalare:

$$\langle \mathbf{Q}\mathbf{x}, \mathbf{Q}\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle \quad \text{per ogni } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

DIMOSTRAZIONE. Se \mathbf{Q} è ortogonale, allora per ogni $\mathbf{x} \in \mathbb{R}^n$

$$\|\mathbf{Q}\mathbf{x}\|^2 = (\mathbf{Q}\mathbf{x})^T \mathbf{Q}\mathbf{x} = \mathbf{x}^T \mathbf{Q}^T \mathbf{Q}\mathbf{x} = \mathbf{x}^T \mathbf{I}\mathbf{x} = \mathbf{x}^T \mathbf{x} = \|\mathbf{x}\|^2$$

Siccome la norma è un numero ≥ 0 , dall'uguaglianza dei quadrati segue $\|\mathbf{Q}\mathbf{x}\| = \|\mathbf{x}\|$. Questo mostra che a) implica b).

Se \mathbf{Q} preserva la norma, allora preserva anche il prodotto scalare per la formula di polarizzazione (3.1). Quindi b) implica c).

Infine supponiamo che \mathbf{Q} preservi il prodotto scalare. Sia \mathbf{q}_k la colonna k di \mathbf{Q} : denotato come al solito con \mathbf{e}_k il k -esimo vettore della base canonica, si ha $\mathbf{q}_k = \mathbf{Q}\mathbf{e}_k$ e quindi

$$\langle \mathbf{q}_i, \mathbf{q}_j \rangle = \langle \mathbf{Q}\mathbf{e}_i, \mathbf{Q}\mathbf{e}_j \rangle = \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij}$$

Questo mostra che le colonne di \mathbf{Q} formano una base ortonormale di \mathbb{R}^n e quindi \mathbf{Q} è ortogonale. Abbiamo così fatto vedere che c) implica a), e questo conclude la dimostrazione.

Dalla classificazione delle matrici ortogonali di ordine 2 (si veda pagina 384) segue che le isometrie lineari del piano cartesiano sono le rotazioni e le riflessioni ortogonali. Siccome il prodotto di due matrici ortogonali è ancora ortogonale (la funzione composta di due isometrie è ancora un'isometria!), tenendo conto che le rotazioni hanno determinante 1 e le riflessioni -1 , vediamo che il prodotto di una rotazione e di una riflessione è una riflessione e che il prodotto di due riflessioni è una rotazione.

COROLLARIO 4.10 Sia \mathbf{Q} una matrice ortogonale e sia λ un autovalore reale di \mathbf{Q} . Allora $\lambda = 1$ oppure $\lambda = -1$.

DIMOSTRAZIONE. Siccome \mathbf{Q} è ortogonale, $\|\mathbf{x}\| = \|\mathbf{Q}\mathbf{x}\|$ per ogni $\mathbf{x} \in \mathbb{R}^n$. Se \mathbf{x} è un autovettore relativo a λ , allora

$$\|\mathbf{x}\| = \|\mathbf{Q}\mathbf{x}\| = \|\lambda\mathbf{x}\| = |\lambda| \|\mathbf{x}\|$$

Dividendo per $\|\mathbf{x}\|$ si trova $|\lambda| = 1$.

OSSERVAZIONE Dimostreremo più avanti che anche gli autovalori complessi di una matrice ortogonale hanno modulo 1.

Dato $\mathbf{b}_1 = [2, 3]^T$, trovare un vettore \mathbf{b}_2 di \mathbb{R}^2 ortogonale a \mathbf{b}_1 . Scrivere le componenti di $[x, y]^T$ rispetto alla base $\{\mathbf{b}_1, \mathbf{b}_2\}$.

Sotto quali condizioni su a, b, c la matrice $\mathbf{Q} = \frac{1}{c} \begin{bmatrix} a & b \\ b & -a \end{bmatrix}$ è ortogonale?

L'insieme delle matrici ortogonali è un sottospazio dello spazio vettoriale delle matrici $n \times n$?

Si considerino i vettori ortogonali $\mathbf{v}_1 = [1, 0, 1]^T$ e $\mathbf{v}_2 = [-1, 2, 1]^T$:

- a) Quante sono le matrici ortogonali di ordine 3 la cui prima colonna è un multiplo scalare di \mathbf{v}_1 e la cui seconda colonna è un multiplo di \mathbf{v}_2 ?

Suggerimento: ci sono due versori per ogni retta per l'origine; la direzione della terza colonna è determinata dalla direzione delle prime due colonne.

- b) Determinare una matrice ortogonale \mathbf{Q} di ordine 3 la cui prima colonna è un multiplo scalare di \mathbf{v}_1 e la cui seconda colonna è un multiplo di \mathbf{v}_2 . Le righe di \mathbf{Q} formano una base ortonormale di \mathbb{R}^3 ?

Dato un vettore non nullo \mathbf{w} in \mathbb{R}^n , si consideri la matrice $n \times n$

$$\mathbf{Q} = \mathbf{I} - \frac{2}{\mathbf{w}^T \mathbf{w}} \mathbf{w} \mathbf{w}^T$$

Scrivere \mathbf{Q} nel caso $n = 3$ e \mathbf{w} è un vettore perpendicolare al sottospazio di equazione $x_1 + x_2 - x_3 = 0$. In generale, mostrare che \mathbf{Q} è una matrice $n \times n$ ortogonale e simmetrica;

Sia \mathbf{E} una matrice antisimmetrica ($\mathbf{E}^T = -\mathbf{E}$) tale che $\mathbf{I} + \mathbf{E}$ sia invertibile. Mostrare che la matrice $\mathbf{Q} = (\mathbf{I} + \mathbf{E})^{-1}(\mathbf{I} - \mathbf{E})$ è ortogonale.

Suggerimento: le matrici $(\mathbf{I} - \mathbf{E})$ e $(\mathbf{I} + \mathbf{E})$ commutano. Per esempio, da $\mathbf{E} = \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix}$ si ottiene la matrice $\mathbf{Q} = \frac{-1}{5} \begin{bmatrix} 3 & -4 \\ 4 & 3 \end{bmatrix}$. E da $\mathbf{E} = \begin{bmatrix} 0 & -a \\ a & 0 \end{bmatrix}$ cosa si ottiene?

Sia \mathbf{Q} una matrice ortogonale di ordine n e sia \mathcal{B} la base ortonormale di \mathbb{R}^n formata dalle colonne di \mathbf{Q} . Mostrare che il vettore delle componenti di un vettore $\mathbf{x} \in \mathbb{R}^n$ rispetto a \mathcal{B} è $\mathbf{Q}^T \mathbf{x}$

Suggerimento: \mathbf{Q} è la matrice di passaggio dalla base canonica alla base \mathcal{B} . Come si confronta questa formula con quella data in termini di coefficienti di Fourier nel corollario 4.4?

Il prodotto di due riflessioni ortogonali del piano è una rotazione: calcolare l'angolo di rotazione in funzione dei coefficienti angolari degli assi delle due riflessioni (fare il disegno).

Il prodotto di una rotazione e di una riflessione ortogonale del piano è una riflessione: calcolare il coefficiente angolare dell'asse della riflessione in funzione dell'angolo di rotazione e del coefficiente angolare della riflessione originale (fare il disegno).

Sia $\mathfrak{T} : \mathbf{V} \rightarrow \mathbf{V}$ un'isometria lineare di uno spazio euclideo \mathbf{V} : \mathfrak{T} è lineare e $\|\mathfrak{T}(\mathbf{v})\| = \|\mathbf{v}\|$ per ogni $\mathbf{v} \in \mathbf{V}$. Mostrare che \mathfrak{T} è iniettiva. Concludere che, se \mathbf{V} ha dimensione finita, un'isometria lineare è un isomorfismo.

Siano \mathbf{V} e \mathbf{W} due spazi euclidei di dimensione n e sia $\mathfrak{T} : \mathbf{V} \rightarrow \mathbf{W}$ un'applicazione lineare. Mostrare che \mathfrak{T} è un'isometria se e solo se \mathfrak{T} manda una base ortonormale di \mathbf{V} in una base ortonormale di \mathbf{W} .

Sia $\mathfrak{T} : \mathbf{V} \rightarrow \mathbf{V}$ un'applicazione lineare di uno spazio euclideo \mathbf{V} in se stesso, sia \mathcal{B} una base ortonormale di \mathbf{V} e sia \mathbf{Q} la matrice che rappresenta \mathfrak{T} rispetto a \mathcal{B} . Mostrare che \mathfrak{T} è un'isometria se e solo se \mathbf{Q} è ortogonale.

Sia $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ un'isometria: $\|F(\mathbf{v}) - F(\mathbf{w})\| = \|\mathbf{v} - \mathbf{w}\|$ per ogni $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$. Mostrare che, se F preserva l'origine ($F(\mathbf{0}) = \mathbf{0}$), allora è lineare. Concludere che ogni isometria di \mathbb{R}^n si può ottenere come prodotto di composizione di una traslazione e di una isometria lineare. Se $n = 2$, ogni isometria è una rototraslazione oppure una traslazione seguita da una riflessione ortogonale.

5 PROIEZIONI ORTOGONALI E ALGORITMO DI GRAM-SCHMIDT

In questo paragrafo risolviamo due problemi fondamentali e ricchi di applicazioni: il problema della proiezione ortogonale e il problema della costruzione di una base ortonormale.

Problema della proiezione ortogonale *Sia \mathbf{V} uno spazio euclideo. Dati un vettore $\mathbf{v} \in \mathbf{V}$ e un sottospazio \mathbf{H} di \mathbf{v} , trovare il vettore $\mathbf{v}_\mathbf{H}$ di \mathbf{H} più vicino a \mathbf{v} , nel senso che*

$$\|\mathbf{v} - \mathbf{v}_\mathbf{H}\| < \|\mathbf{v} - \mathbf{w}\| \quad \text{per ogni } \mathbf{w} \in \mathbf{H}, \mathbf{w} \neq \mathbf{v}_\mathbf{H}$$

Non è ovvio che esista un vettore $\mathbf{v}_\mathbf{H}$ che minimizzi $\|\mathbf{v} - \mathbf{w}\|$ al variare di \mathbf{w} in \mathbf{H} : anzi, se \mathbf{H} non ha dimensione finita, è possibile che non esista. Dimostreremo però che la proiezione ortogonale esiste sempre se \mathbf{H} ha dimensione finita (anche se lo spazio \mathbf{V} ha dimensione infinita). Per analogia con quanto visto per le rette, è lecito attendersi che il problema sia equivalente a quello di *decomporre \mathbf{v} come somma di un vettore $\mathbf{v}_\mathbf{H}$ di \mathbf{H} e di un vettore \mathbf{v}_\perp ortogonale ad \mathbf{H}* .

Problema della costruzione di una base ortonormale *Dato un sottospazio \mathbf{H} di uno spazio euclideo \mathbf{V} , costruire una base ortonormale di \mathbf{H} .*

Già dall'esperienza dello spazio cartesiano tridimensionale possiamo intuire l'utilità della costruzione: supponiamo per esempio di voler studiare nello spazio una rotazione che abbia come asse la retta \mathbf{L} ; è naturale allora scegliere l'origine delle coordinate sull'asse \mathbf{L} e i versori fondamentali $\mathbf{i}, \mathbf{j}, \mathbf{k}$ in modo che \mathbf{i} sia diretto come l'asse e quindi \mathbf{j} e \mathbf{k} appartengano al piano ortogonale all'asse. In questo modo troviamo una base ortonormale di \mathbb{R}^3 il cui primo vettore è una base ortonormale dell'asse \mathbf{L} e gli altri due sono una base ortonormale del piano perpendicolare a \mathbf{L} .

L'algoritmo di Gram-Schmidt risolve il problema della costruzione di una base ortonormale, dopo di che è immediato risolvere il problema delle proiezioni ortogonali perché data una base ortonormale di \mathbf{H} è possibile scrivere una formula analitica per la proiezione ortogonale di un vettore su \mathbf{H} . D'altra parte le proiezioni ortogonali entrano nell'algoritmo in modo naturale: per costruire una base ortonormale, si parte da una base qualsiasi e si cerca, per così dire, di raddrizzarla mediante proiezioni ortogonali: per esempio se \mathbf{b} e \mathbf{c} formano una base qualsiasi di un piano \mathbf{W} , possiamo trovare una base ortogonale proiettando \mathbf{c} nella direzione perpendicolare a \mathbf{b} . Quindi le soluzioni dei due problemi sono interconnesse tra loro. Per procedere ordinatamente faremo così: prima tratteremo in dettaglio la nozione di proiezione ortogonale, poi faremo vedere che data una base ortonormale di \mathbf{H} esiste una formula analitica esplicita per la proiezione ortogonale $\mathbf{v}_\mathbf{H}$, infine per mezzo di questa formula esplicita descriveremo l'algoritmo di Gram-Schmidt di costruzione delle basi ortonormali.

Fissato un sottospazio \mathbf{H} di uno spazio euclideo \mathbf{V} , si dice che un vettore \mathbf{v} è *ortogonale ad \mathbf{H}* se è ortogonale a *ogni* vettore di \mathbf{H} . L'insieme dei vettori ortogonali ad \mathbf{H} si denota con il simbolo \mathbf{H}^\perp :

$$\mathbf{H}^\perp = \{\mathbf{v} \in \mathbf{V} : \langle \mathbf{v}, \mathbf{w} \rangle = 0 \text{ per ogni } \mathbf{w} \in \mathbf{H}\}$$

Supponiamo che \mathbf{V} sia \mathbb{R}^n con il prodotto scalare standard. Allora la condizione che un vettore $\mathbf{a} = [a_1, \dots, a_n]^T$ sia ortogonale a $\mathbf{x} = [x_1, \dots, x_n]^T$ è

$$(5.1) \quad \mathbf{a}^T \mathbf{x} = a_1 x_1 + \cdots + a_n x_n = 0$$

Il vettore \mathbf{a} appartiene ad \mathbf{H}^\perp se la (5.1) vale per ogni $\mathbf{x} \in \mathbf{H}$, cioè se la (5.1) è un'equazione cartesiana di \mathbf{H} . Quindi \mathbf{H}^\perp consiste dei vettori dei coefficienti delle equazioni di \mathbf{H} . Dalla simmetria della (5.1) si intuisce che $(\mathbf{H}^\perp)^\perp$ dev'essere \mathbf{H} . Vedremo che questo è il caso per ogni sottospazio \mathbf{H} di dimensione finita in uno spazio euclideo arbitrario.

- a) In \mathbb{R}^3 , con il prodotto scalare standard, l'insieme dei vettori ortogonali all'asse z è il piano xy ; e l'insieme dei vettori ortogonali al piano xy è l'asse z . Più in generale, l'insieme dei vettori ortogonali al piano generato da due vettori non paralleli \mathbf{v} e \mathbf{w} è la retta generata da $\mathbf{v} \times \mathbf{w}$ (il prodotto vettoriale di \mathbf{v} e \mathbf{w}).
- b) In \mathbb{R}^4 , con il prodotto scalare standard, l'insieme dei vettori perpendicolari al primo asse coordinato è l'iperpiano di equazione $x_1 = 0$. Infatti un vettore $\mathbf{v} = [x_1, x_2, x_3, x_4]^T$ è perpendicolare al primo vettore \mathbf{e}_1 della base canonica se e solo se il prodotto scalare

$$\mathbf{e}_1 \cdot \mathbf{v} = [1 \ 0 \ 0 \ 0] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_1$$

è uguale a zero.

- c) Sia \mathbf{V} lo spazio vettoriale delle funzioni reali continue su $[a, b]$, con il prodotto scalare definito per integrazione. Allora una funzione è ortogonale alla funzione costante $\mathbf{1}$ se e solo se ha media integrale nulla perché

$$\langle \mathbf{f}, \mathbf{1} \rangle = \int_a^b f(x) dx$$

Quindi $\mathcal{L}(1)^\perp$ è l'insieme delle funzioni con media integrale nulla.

- d) In ogni spazio euclideo \mathbf{V} , si ha $\mathbf{V}^\perp = \{\mathbf{0}\}$ e $\mathbf{0}^\perp = \mathbf{V}$. Per mostrare la prima uguaglianza, osserviamo che, se $\mathbf{v} \in \mathbf{V}^\perp$, allora \mathbf{v} è perpendicolare a se stesso, quindi

$$0 = \langle \mathbf{v}, \mathbf{v} \rangle = \|\mathbf{v}\|^2$$

e questo implica $\mathbf{v} = \mathbf{0}$ per la proprietà di positività del prodotto scalare. Questo mostra che $\mathbf{V}^\perp = \{\mathbf{0}\}$. L'uguaglianza $\mathbf{0}^\perp = \mathbf{V}$ è un altro modo di scrivere che $\mathbf{0}$ è ortogonale a ogni vettore.

PROPOSIZIONE 5.1

Sia \mathbf{V} uno spazio euclideo e sia \mathbf{H} un suo sottospazio. Allora:

- a) l'insieme \mathbf{H}^\perp è un sottospazio vettoriale di \mathbf{V} ;
- b) se \mathbf{v} è ortogonale a un insieme di generatori di \mathbf{H} (in particolare, ai vettori di una base di \mathbf{H}), allora $\mathbf{v} \in \mathbf{H}^\perp$.

DIMOSTRAZIONE. Il vettore nullo è ortogonale a ogni vettore e quindi appartiene ad \mathbf{H}^\perp . Supponiamo che $\mathbf{v}_1, \mathbf{v}_2 \in \mathbf{H}^\perp$. Allora per ogni $t_1, t_2 \in \mathbb{R}$ e per ogni $\mathbf{w} \in \mathbf{H}$:

$$\langle t_1 \mathbf{v}_1 + t_2 \mathbf{v}_2, \mathbf{w} \rangle = t_1 \langle \mathbf{v}_1, \mathbf{w} \rangle + t_2 \langle \mathbf{v}_2, \mathbf{w} \rangle = 0$$

Questo mostra che $t_1 \mathbf{v}_1 + t_2 \mathbf{v}_2$ appartiene ad \mathbf{H}^\perp , e quindi \mathbf{H}^\perp è un sottospazio.

Per quanto riguarda b), ricordiamo che $\mathbf{v}_1, \dots, \mathbf{v}_d$ generano \mathbf{H} se ogni vettore di \mathbf{H} è combinazione lineare di $\mathbf{v}_1, \dots, \mathbf{v}_d$. Quindi dobbiamo mostrare che, se \mathbf{v} è ortogonale a $\mathbf{v}_1, \dots, \mathbf{v}_d$, allora \mathbf{v} è ortogonale a ogni combinazione lineare $\mathbf{w} = t_1 \mathbf{v}_1 + t_2 \mathbf{v}_2 + \dots + t_d \mathbf{v}_d$ di tali vettori:

$$\begin{aligned} \langle \mathbf{v}, \mathbf{w} \rangle &= \langle \mathbf{v}, t_1 \mathbf{v}_1 + \dots + t_d \mathbf{v}_d \rangle = \\ &= t_1 \langle \mathbf{v}, \mathbf{v}_1 \rangle + \dots + t_d \langle \mathbf{v}, \mathbf{v}_d \rangle = \\ &= t_1 0 + \dots + t_d 0 = 0 \end{aligned}$$

Per dimostrare che \mathbf{v} è ortogonale ad \mathbf{H} è quindi sufficiente mostrare che \mathbf{v} è ortogonale a una base $\{\mathbf{v}_1, \dots, \mathbf{v}_d\}$ di \mathbf{H} . Nel caso più semplice, un vettore è ortogonale alla retta $\mathbb{R}\mathbf{v}_1$ se e solo se è perpendicolare a \mathbf{v}_1 . Per un caso con $d = 2$, si consideri il piano xy in \mathbb{R}^3 : un vettore è ortogonale al piano xy se e solo se è perpendicolare ai primi due versori \mathbf{e}_1 ed \mathbf{e}_2 della base canonica.

Dato un sottospazio \mathbf{H} di \mathbf{V} , ci chiediamo ora se è possibile decomporre un arbitrario vettore \mathbf{v} come somma di un vettore \mathbf{v}_H appartenente ad \mathbf{H} e di un vettore \mathbf{v}_\perp perpendicolare ad \mathbf{H} . Intuitivamente \mathbf{v}_H è la proiezione ortogonale di \mathbf{v} su \mathbf{H} , e dovrebbe quindi essere il vettore di \mathbf{H} a distanza minima da \mathbf{v} .

PROPOSIZIONE 5.2 (Proprietà di minimo della proiezione ortogonale)

Sia \mathbf{H} un sottospazio di uno spazio euclideo \mathbf{V} . Supponiamo che un vettore $\mathbf{v} \in \mathbf{V}$ si possa scrivere come somma

$$\mathbf{v} = \mathbf{v}_H + \mathbf{v}_\perp$$

di un vettore \mathbf{v}_H appartenente ad \mathbf{H} e di un vettore \mathbf{v}_\perp ortogonale ad \mathbf{H} . Allora \mathbf{v}_H è l'unico vettore di \mathbf{H} a distanza minima da \mathbf{v} :

$$(5.2) \quad \|\mathbf{v} - \mathbf{v}_H\| < \|\mathbf{v} - \mathbf{w}\| \quad \text{per ogni } \mathbf{w} \in \mathbf{H}, \mathbf{w} \neq \mathbf{v}_H$$

In particolare, \mathbf{v}_H è l'unico vettore di \mathbf{H} tale che $\mathbf{v} - \mathbf{v}_H$ sia ortogonale ad \mathbf{H} .

Si dice che \mathbf{v}_H è la *proiezione ortogonale* di \mathbf{v} su \mathbf{H} .

DIMOSTRAZIONE. La dimostrazione è la stessa che abbiamo già visto nel caso della proiezione su una retta. Sia \mathbf{w} un vettore di \mathbf{H} . La differenza $\mathbf{v}_H - \mathbf{w}$ è un vettore di \mathbf{H} perché \mathbf{H} è un sottospazio. Per ipotesi $\mathbf{v} - \mathbf{v}_H = \mathbf{v}_\perp \in \mathbf{H}^\perp$, quindi $\mathbf{v} - \mathbf{v}_H$ è ortogonale a $\mathbf{v}_H - \mathbf{w}$ che appartiene ad \mathbf{H} . La diseguaglianza (5.2) ora segue dal fatto che $\mathbf{v} - \mathbf{w}$ è l'ipotenusa – si veda la figura 5.1 – del triangolo rettangolo che ha per cateti $\mathbf{v} - \mathbf{v}_H$ e $\mathbf{v}_H - \mathbf{w}$. Infatti per il teorema di Pitagora:

$$\|\mathbf{v} - \mathbf{w}\|^2 = \|(\mathbf{v} - \mathbf{v}_H) + (\mathbf{v}_H - \mathbf{w})\|^2 = \|\mathbf{v} - \mathbf{v}_H\|^2 + \|\mathbf{v}_H - \mathbf{w}\|^2$$

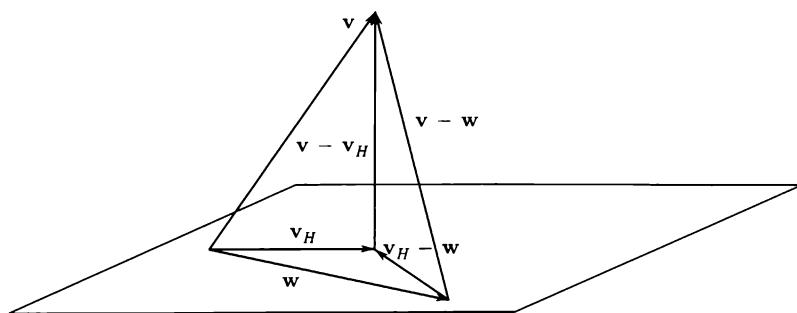


Figura 5.1. Il vettore $\mathbf{v} - \mathbf{v}_H$ è ortogonale a tutti i vettori di \mathbf{H} .

Se $\mathbf{w} \neq \mathbf{v}_H$, allora $\|\mathbf{v}_H - \mathbf{w}\| > 0$, e quindi

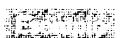
$$\|\mathbf{v} - \mathbf{w}\|^2 = \|\mathbf{v} - \mathbf{v}_H\|^2 + \|\mathbf{v}_H - \mathbf{w}\|^2 > \|\mathbf{v} - \mathbf{v}_H\|^2$$

che è equivalente alla (5.2).

OSSERVAZIONE È importante osservare che, se \mathbf{v}_H è la proiezione ortogonale di \mathbf{v} su \mathbf{H} , allora $\mathbf{v}_\perp = \mathbf{v} - \mathbf{v}_H$ è la proiezione di \mathbf{v} su \mathbf{H}^\perp . Infatti $\mathbf{v}_\perp \in \mathbf{H}^\perp$ e $\mathbf{v}_H = \mathbf{v} - \mathbf{v}_\perp$ appartiene ad \mathbf{H} e quindi è ortogonale ad \mathbf{H}^\perp . In particolare, \mathbf{v}_\perp è il vettore di \mathbf{H}^\perp a distanza minima da \mathbf{v} .



Facciamo un esempio che è banale, ma che è bene aver presente. Se \mathbf{v} appartiene ad \mathbf{H} , allora \mathbf{v} è la proiezione ortogonale di \mathbf{v} su \mathbf{H} ; in questo caso $\mathbf{v}_H = \mathbf{v}$ e $\mathbf{v}_\perp = \mathbf{0}$. Se invece $\mathbf{v} \in \mathbf{H}^\perp$ è perpendicolare ad \mathbf{H} , allora la proiezione ortogonale di \mathbf{v} su \mathbf{H} è il vettore nullo; in questo caso $\mathbf{v}_H = \mathbf{0}$ e $\mathbf{v}_\perp = \mathbf{v}$.



Sia ℓ_2 lo spazio euclideo delle successioni $\mathbf{a} = \{a_n\}$ tali che $\sum_{n=0}^{+\infty} a_n^2 < +\infty$. In ℓ_2 consideriamo il sottospazio \mathbf{H} delle successioni *finite*: $\{a_n\} \in \mathbf{H}$ se esiste N tale che $a_n = 0$ per ogni $n > N$. Allora $\mathbf{H}^\perp = \{\mathbf{0}\}$ e quindi le uniche successioni che si possono scrivere nella forma $\mathbf{a} = \mathbf{a}_H + \mathbf{a}_\perp$ con $\mathbf{a}_H \in \mathbf{H}$ e $\mathbf{a}_\perp \in \mathbf{H}^\perp$ sono le successioni finite. In particolare, la successione $\mathbf{v} = \left\{ \frac{1}{n+1} \right\}$ non ha proiezione ortogonale su \mathbf{H} . Non esiste una successione di \mathbf{H} a distanza minima da questa successione: per ogni k sia $\mathbf{v}_k \in \mathbf{H}$ la successione che è uguale a \mathbf{v} per $n \leq k$ e zero per tutti gli altri n . Allora

$$\|\mathbf{v} - \mathbf{v}_k\|^2 = \sum_{n=k+1}^{+\infty} \frac{1}{(n+1)^2} \rightarrow 0 \quad \text{per } k \rightarrow +\infty$$

Se esistesse in \mathbf{H} la proiezione ortogonale \mathbf{v}_H di \mathbf{v} , allora

$$\|\mathbf{v} - \mathbf{v}_H\| < \|\mathbf{v} - \mathbf{v}_k\| \rightarrow 0$$

implicherebbe $\|\mathbf{v} - \mathbf{v}_H\| = 0$ e quindi $\mathbf{v} = \mathbf{v}_H \in \mathbf{H}$, il che è assurdo.

OSSERVAZIONE La funzione che a \mathbf{v} associa la proiezione ortogonale \mathbf{v}_H , quando esiste, è lineare: se \mathbf{v}_H e \mathbf{w}_H sono le proiezioni ortogonali su H di \mathbf{v} e \mathbf{w} rispettivamente, allora per ogni coppia di scalari t_1 e t_2 il vettore $t_1\mathbf{v}_H + t_2\mathbf{w}_H$ è la proiezione ortogonale $(t_1\mathbf{v} + t_2\mathbf{w})_H$. Per vedere questo, basta osservare che

$$(t_1\mathbf{v} + t_2\mathbf{w}) - (t_1\mathbf{v}_H + t_2\mathbf{w}_H)$$

è perpendicolare ad H visto che $\mathbf{v} - \mathbf{v}_H$ e $\mathbf{w} - \mathbf{w}_H$ lo sono.

Mostriamo ora che la proiezione ortogonale \mathbf{v}_H esiste se H ha una base ortogonale.

PROPOSIZIONE 5.3 (Formula analitica della proiezione ortogonale)

Sia V uno spazio vettoriale euclideo e sia H un suo sottospazio. Supponiamo che H abbia una base *ortogonale* $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_d\}$. Allora per ogni vettore \mathbf{v} di V esiste la proiezione ortogonale \mathbf{v}_H di \mathbf{v} su H , e vale la formula:

$$(5.3) \quad \mathbf{v}_H = \hat{x}_1 \mathbf{b}_1 + \hat{x}_2 \mathbf{b}_2 + \cdots + \hat{x}_d \mathbf{b}_d$$

dove

$$(5.4) \quad \hat{x}_k = \frac{\langle \mathbf{v}, \mathbf{b}_k \rangle}{\|\mathbf{b}_k\|^2} \quad k = 1, 2, \dots, d$$

è il coefficiente di Fourier di \mathbf{v} rispetto a \mathbf{b}_k .

DIMOSTRAZIONE. Vogliamo mostrare che esiste un vettore di H , quindi una combinazione lineare

$$\mathbf{v}_H = t_1 \mathbf{b}_1 + t_2 \mathbf{b}_2 + \cdots + t_d \mathbf{b}_d = \sum_{i=1}^d t_i \mathbf{b}_i$$

tale che $\mathbf{v} - \mathbf{v}_H$ è perpendicolare ad H . Il vettore $\mathbf{v} - \mathbf{v}_H$ è perpendicolare a H se e solo se è perpendicolare ai vettori $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_d$, perché questi vettori formano una base di H . Quindi basta imporre che

$$\langle \mathbf{v} - \mathbf{v}_H, \mathbf{b}_k \rangle = 0 \quad \text{per ogni } k = 1, 2, \dots, d$$

Ora

$$\begin{aligned} \langle \mathbf{v} - \sum_{i=1}^d t_i \mathbf{b}_i, \mathbf{b}_k \rangle &= \langle \mathbf{v}, \mathbf{b}_k \rangle - \langle \sum_{i=1}^d t_i \mathbf{b}_i, \mathbf{b}_k \rangle = \\ &= \langle \mathbf{v}, \mathbf{b}_k \rangle - \sum_{i=1}^d t_i \langle \mathbf{b}_i, \mathbf{b}_k \rangle = \langle \mathbf{v}, \mathbf{b}_k \rangle - t_k \langle \mathbf{b}_k, \mathbf{b}_k \rangle \end{aligned}$$

dove nell'ultimo passaggio abbiamo usato il fatto che \mathbf{b}_i è ortogonale a \mathbf{b}_k se $i \neq k$. La conclusione è che $\mathbf{v} - \mathbf{v}_H$ è perpendicolare a \mathbf{b}_k se e solo se

$$t_k = \frac{\langle \mathbf{v}, \mathbf{b}_k \rangle}{\langle \mathbf{b}_k, \mathbf{b}_k \rangle} = \hat{x}_k$$

Quindi prendendo $t_k = \hat{x}_k$ per ogni k troviamo un vettore $\mathbf{v}_H \in H$ tale che $\mathbf{v} - \mathbf{v}_H$ è perpendicolare a H .

OSSERVAZIONE La formula (5.3) ci dice che, data una base *ortogonale* $\{\mathbf{b}_1, \dots, \mathbf{b}_d\}$ di \mathbf{H} , la proiezione di un vettore \mathbf{v} su \mathbf{H} è la somma delle proiezioni di \mathbf{v} sugli assi $\mathbb{R}\mathbf{b}_i$. Si faccia attenzione che questo è vero solo se la base è ortogonale.

Consideriamo il piano \mathbf{H} di \mathbb{R}^4 di equazioni $x_1 + x_2 = x_3 + x_4 = 0$. Una base ortogonale di \mathbf{H} è formata dai due vettori $\mathbf{b}_1 = [1, -1, 0, 0]^T$ e $\mathbf{b}_2 = [0, 0, 1, -1]^T$. I coefficienti di Fourier del vettore $\mathbf{v} = [x_1, x_2, x_3, x_4]^T$ sono

$$\hat{x}_1 = \frac{\mathbf{v}^T \mathbf{b}_1}{\mathbf{b}_1^T \mathbf{b}_1} = \frac{x_1 - x_2}{2} \quad \text{e} \quad \hat{x}_2 = \frac{\mathbf{v}^T \mathbf{b}_2}{\mathbf{b}_2^T \mathbf{b}_2} = \frac{x_3 - x_4}{2}$$

Quindi

$$\mathbf{v}_{\mathbf{H}} = \frac{x_1 - x_2}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} + \frac{x_3 - x_4}{2} \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} x_1 - x_2 \\ x_2 - x_1 \\ x_3 - x_4 \\ x_4 - x_3 \end{bmatrix}$$

In questo genere di esercizi, ci sono diversi modi di controllare di non avere commesso errori nei conti. Innanzitutto la proiezione $\mathbf{v}_{\mathbf{H}}$ dev'essere un vettore di \mathbf{H} ; nel caso di questo esempio $\mathbf{v}_{\mathbf{H}}$ deve quindi soddisfare le equazioni $x_1 + x_2 = x_3 + x_4 = 0$; quindi, la proiezione di un vettore di \mathbf{H} deve coincidere con il vettore stesso; basta controllare questo per i vettori della base: la proiezione di \mathbf{b}_i dev'essere \mathbf{b}_i .

Se \mathbf{V} ha dimensione finita, trovare la proiezione ortogonale $\mathbf{v}_{\mathbf{H}}$ di un vettore \mathbf{v} su \mathbf{H} è equivalente a trovare la proiezione $\mathbf{v}_{\perp} = \mathbf{v} - \mathbf{v}_{\mathbf{H}}$ di \mathbf{v} su \mathbf{H}^{\perp} . Per calcolare le due proiezioni conviene calcolare la proiezione sul più piccolo tra i due sottospazi \mathbf{H} e \mathbf{H}^{\perp} e ottenere l'altra per differenza.

Per esempio, consideriamo il piano \mathbf{H} di equazione $x + y + z = 0$ in \mathbb{R}^3 . In questo caso \mathbf{H}^{\perp} è la retta \mathbf{r} generata dal vettore $\mathbf{b} = [1, 1, 1]^T$. La proiezione ortogonale di un vettore $\mathbf{v} = [x, y, z]^T$ su \mathbf{r} è

$$\mathbf{v}_{\mathbf{r}} = \hat{x} \mathbf{b} = \frac{x + y + z}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} x + y + z \\ x + y + z \\ x + y + z \end{bmatrix}$$

Quindi la proiezione ortogonale sul piano \mathbf{H} è

$$\mathbf{v}_{\mathbf{H}} = \mathbf{v} - \mathbf{v}_{\mathbf{r}} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} - \frac{1}{3} \begin{bmatrix} x + y + z \\ x + y + z \\ x + y + z \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2x - y - z \\ 2y - x - z \\ 2z - x - y \end{bmatrix}$$

Se la base di \mathbf{H} non è ortogonale, non è vero che la proiezione su \mathbf{H} si ottiene come somma delle proiezioni sulle rette generate dai vettori della base. Per esempio, consideriamo il piano xy in \mathbb{R}^3 , ma con la base

$$\left\{ \mathbf{b}_1 = [1, 0, 0]^T, \quad \mathbf{b}_2 = [1, 1, 0]^T \right\}$$

La proiezione ortogonale di $[x, y, z]^T$ sull'asse $\mathbb{R}\mathbf{b}_2$ (che è la bisettrice di equazione $x = y$ nel piano xy) è, per la formula (5.3), il vettore

$$\frac{\mathbf{v}^T \mathbf{b}_2}{\mathbf{b}_2^T \mathbf{b}_2} \mathbf{b}_2 = \frac{x+y}{2} [1, 1, 0]^T = \left[\frac{x+y}{2}, \frac{x+y}{2}, 0 \right]^T$$

Quindi la somma delle due proiezioni sugli assi generati da \mathbf{b}_1 e \mathbf{b}_2 è

$$[x, 0, 0]^T + \left[\frac{x+y}{2}, \frac{x+y}{2}, 0 \right]^T = \left[\frac{3x+y}{2}, \frac{x+y}{2}, 0 \right]^T$$

che non ha nulla a che vedere con la proiezione ortogonale $[x, y, 0]^T$ di $[x, y, z]^T$ sul piano xy .

La proposizione 5.3 non richiede che \mathbf{V} abbia dimensione finita. L'applicazione più importante è al caso delle serie di Fourier. Fu proprio Fourier ad avere l'intuizione che una funzione periodica si potesse approssimare, con un errore piccolo a piacere, con una combinazione lineare di funzioni sinusoidali elementari e che i coefficienti di tale combinazione lineare andassero calcolati come nel caso delle proiezioni ortogonali di vettori. Per rendere preciso questo discorso, consideriamo lo spazio euclideo \mathbf{V} delle funzioni continue su $[0, 2\pi]$ con la norma L^2 . Un polinomio trigonometrico di grado $\leq n$ è una funzione della forma:

$$P(x) = \sum_{k=0}^n c_k \cos(kx) + \sum_{k=1}^n d_k \sin(kx)$$

L'insieme \mathbf{H}_n dei polinomi trigonometrici di grado $\leq n$ è, quindi, il sottospazio di \mathbf{V} generato dalle funzioni

$$1 = \cos(0x), \cos(x), \dots, \cos(nx), \sin(x), \dots, \sin(nx)$$

È semplice verificare che tali funzioni sono a due a due ortogonali, e quindi formano una base ortogonale di \mathbf{H}_n . Dalla proposizione 5.3 segue immediatamente che la proiezione ortogonale $S_n(f)$ di una funzione $f \in \mathbf{V}$ sul sottospazio \mathbf{H}_n è

$$S_n(f)(x) = \frac{a_0}{2} + \sum_{k=1}^n a_k \cos(kx) + \sum_{k=1}^n b_k \sin(kx)$$

dove

$$a_k = a_k(f) = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(kx) dx \quad k = 0, 1, 2, \dots$$

e

$$b_k = b_k(f) = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(kx) dx \quad k = 1, 2, \dots$$

Per la proprietà di minimo della proiezione ortogonale, $S_n(f)$ è il polinomio trigonometrico di grado $\leq n$ che meglio approssima f in norma L^2 . È interessante esplicitare il problema di minimo che abbiamo così risolto. Per ogni scelta di $(a_0, a_1, \dots, a_n, b_1, \dots, b_n)$ in \mathbb{R}^{2n+1} consideriamo il corrispondente polinomio trigonometrico:

$$P(a_0, a_1, \dots, b_n)(x) = \frac{a_0}{2} + \sum_{k=1}^n a_k \cos(kx) + \sum_{k=1}^n b_k \sin(kx)$$

La funzione da minimizzare è lo scarto quadratico medio da $f(x)$:

$$F(a_0, a_1, \dots, b_n) = \|f(x) - P(a_0, a_1, \dots, b_n)(x)\|^2 = \int_0^{2\pi} (f(x) - P(a_0, a_1, \dots, b_n)(x))^2 dx$$

Il risultato è che lo scarto quadratico medio è minimo quando $(a_0, a_1, \dots, a_n, b_1, \dots, b_n)$ sono i coefficienti di Fourier di $f(x)$:

$$\|f(x) - P(a_0, a_1, \dots, b_n)(x)\|^2 \geq \|f(x) - S_n(f)(x)\|^2$$

per ogni $(a_0, a_1, \dots, a_n, b_1, \dots, b_n)$, con uguaglianza se e solo se $(a_0, a_1, \dots, a_n, b_1, \dots, b_n)$ sono i coefficienti di Fourier di $f(x)$.

Si può dimostrare che facendo tendere $n \rightarrow +\infty$ l'errore di approssimazione $\|f(x) - S_n(f)(x)\|^2$ tende a zero e di solito molto velocemente. Questo è importante perché consente nelle applicazioni di sostituire una funzione periodica arbitraria con una funzione facilmente maneggevole quale è un polinomio trigonometrico.

L'algoritmo di Gram-Schmidt

Descriviamo ora l'algoritmo di Gram-Schmidt che permette di ricavare una *base ortogonale* (e ortonormale se necessario) di \mathbf{H} a partire da una base qualsiasi. In particolare, questo dimostra che ogni sottospazio \mathbf{H} di dimensione finita ammette una base ortonormale e, quindi, per la proposizione 5.3, la proiezione ortogonale su \mathbf{H} esiste per ogni vettore \mathbf{v} di \mathbf{V} .

PROPOSIZIONE 5.4 (Algoritmo di Gram-Schmidt)

Dato un insieme $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d \subseteq \mathbf{V}$ di vettori linearmente indipendenti, l'algoritmo di Gram-Schmidt costruisce induttivamente i vettori $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_d$ con la seguente procedura:

$$\begin{aligned}\mathbf{b}_1 &= \mathbf{v}_1 \\ \mathbf{b}_2 &= \mathbf{v}_2 - \hat{x}_2^1 \mathbf{b}_1 \\ \mathbf{b}_3 &= \mathbf{v}_3 - (\hat{x}_3^1 \mathbf{b}_1 + \hat{x}_3^2 \mathbf{b}_2)\end{aligned}$$

$$\mathbf{b}_d = \mathbf{v}_d - (\hat{x}_d^1 \mathbf{b}_1 + \dots + \hat{x}_d^{d-1} \mathbf{b}_{d-1})$$

dove

$$\hat{x}_k^j = \frac{\langle \mathbf{v}_k, \mathbf{b}_j \rangle}{\|\mathbf{b}_j\|^2} \quad \text{per ogni } 1 \leq k \leq d, 1 \leq j \leq k-1$$

è il coefficiente di Fourier di \mathbf{v}_k rispetto a \mathbf{b}_j .

I vettori \mathbf{b}_j sono a due a due ortogonali e, per ogni $k \leq d$, l'insieme $\{\mathbf{b}_1, \dots, \mathbf{b}_k\}$ è una base ortogonale del sottospazio di \mathbf{V} generato da $\mathbf{v}_1, \dots, \mathbf{v}_k$.

In particolare, se $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d$ è una base di \mathbf{H} , allora $\{\mathbf{b}_1, \dots, \mathbf{b}_d\}$ è una base ortogonale di \mathbf{H} .

DIMOSTRAZIONE. Sia $\mathbf{H}_k = \mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ il sottospazio generato dai primi k dei vettori assegnati. L'idea alla base dell'algoritmo è semplice e intuitiva, come mostra la figura 5.2: al passo k dell'algoritmo, il vettore \mathbf{b}_k è la proiezione ortogonale di \mathbf{v}_k sulla retta ortogonale ad \mathbf{H}_{k-1} in \mathbf{H}_k ; quindi $\mathbf{v}_k - \mathbf{b}_k$ è la proiezione ortogonale di \mathbf{v}_k su \mathbf{H}_{k-1} , e questo è il significato della formula

$$\mathbf{b}_k = \mathbf{v}_k - (\hat{x}_k^1 \mathbf{b}_1 + \dots + \hat{x}_k^{k-1} \mathbf{b}_{k-1})$$

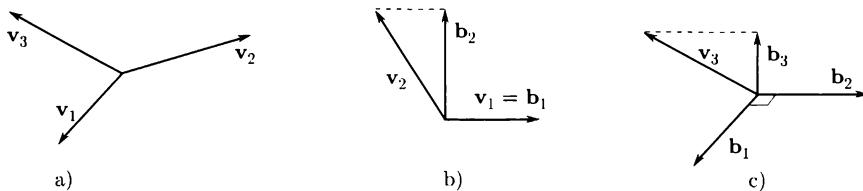


Figura 5.2. a) Tre vettori linearmente indipendenti; b) nel piano $\mathbf{H}_2 = \mathcal{L}(\mathbf{v}_1, \mathbf{v}_2)$, il vettore \mathbf{b}_2 è la proiezione ortogonale di \mathbf{v}_2 sulla retta ortogonale a \mathbf{b}_1 ; c) in $\mathbf{H}_3 = \mathcal{L}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$, il vettore \mathbf{b}_3 è la proiezione ortogonale di \mathbf{v}_3 sulla retta ortogonale a \mathbf{b}_1 e \mathbf{b}_2 .

Illustriamo i dettagli. Dobbiamo mostrare che i vettori $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_d$ sono due a due ortogonali e che

$$\mathbf{H}_k = \mathcal{L}(\mathbf{b}_1, \dots, \mathbf{b}_k) \quad \text{per ogni } k = 1, 2, \dots, d$$

Evidentemente

$$\mathcal{L}(\mathbf{b}_1, \dots, \mathbf{b}_k) = \mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_k) = \mathbf{H}_k$$

perché per costruzione il vettore \mathbf{b}_j è combinazione lineare dei vettori $\mathbf{v}_1, \dots, \mathbf{v}_k$, e viceversa \mathbf{v}_j è combinazione lineare dei vettori $\mathbf{b}_1, \dots, \mathbf{b}_k$. Possiamo supporre per induzione di aver verificato che $\mathbf{b}_1, \dots, \mathbf{b}_{k-1}$ siano a due a due ortogonali e procedere a mostrare che \mathbf{b}_k è ortogonale a $\mathbf{b}_1, \dots, \mathbf{b}_{k-1}$. Siccome $\mathbf{b}_1, \dots, \mathbf{b}_{k-1}$ è una base ortogonale di \mathbf{H}_{k-1} ,

$$\hat{x}_k^1 \mathbf{b}_1 + \dots + \hat{x}_k^{k-1} \mathbf{b}_{k-1}$$

è la proiezione ortogonale di \mathbf{v}_k su \mathbf{H}_{k-1} , e quindi

$$\mathbf{b}_k = \mathbf{v}_k - (\hat{x}_k^1 \mathbf{b}_1 + \dots + \hat{x}_k^{k-1} \mathbf{b}_{k-1})$$

è ortogonale a \mathbf{H}_{k-1} , cioè ai vettori $\mathbf{b}_1, \dots, \mathbf{b}_{k-1}$.

OSSERVAZIONE Per trovare una base ortonormale anziché ortogonale, è sufficiente normalizzare la base ortogonale $\{\mathbf{b}_k\}$ ponendo:

$$\mathbf{q}_k = \frac{\mathbf{b}_k}{\|\mathbf{b}_k\|}$$

OSSERVAZIONE L'algoritmo si può modificare in modo da poter essere applicato anche se i vettori \mathbf{v}_i non sono linearmente indipendenti: produce ancora dei vettori \mathbf{b}_i a due a due ortogonali con la proprietà che

$$\mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_k) = \mathcal{L}(\mathbf{b}_1, \dots, \mathbf{b}_k) \quad \text{per ogni } k = 1, 2, \dots, d$$

La differenza è che in questo caso, se \mathbf{v}_k è combinazione lineare dei precedenti \mathbf{v}_i , allora $\mathbf{b}_k = \mathbf{0}$. Questo \mathbf{b}_k va scartato. Alla fine i \mathbf{b}_k non nulli formano una base ortogonale del sottospazio generato dai vettori \mathbf{v}_i .

ESERCIZIO

Troviamo una base ortonormale dell'iperpiano \mathbf{H} di equazione $x_1 + x_2 + x_3 + x_4 = 0$ in \mathbb{R}^4 . Trattando x_2, x_3 e x_4 come variabili libere, si trova una base di \mathbf{H} formata dai tre vettori

$$\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Procediamo come prescritto dall'algoritmo di Gram-Schmidt: definiamo $\mathbf{b}_1 = \mathbf{v}_1 = [-1, 1, 0, 0]^T$ e

$$\mathbf{b}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{b}_1 \rangle}{\|\mathbf{b}_1\|^2} \mathbf{b}_1 =$$

$$= \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -1/2 \\ -1/2 \\ 1 \\ 0 \end{bmatrix}$$

$$\mathbf{b}_3 = \mathbf{v}_3 - \frac{\langle \mathbf{v}_3, \mathbf{b}_1 \rangle}{\|\mathbf{b}_1\|^2} \mathbf{b}_1 - \frac{\langle \mathbf{v}_3, \mathbf{b}_2 \rangle}{\|\mathbf{b}_2\|^2} \mathbf{b}_2 =$$

$$= \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1/2}{3/2} \begin{bmatrix} -1/2 \\ -1/2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1/3 \\ -1/3 \\ -1/3 \\ 1 \end{bmatrix}$$

Ricapitolando abbiamo così trovato una base ortogonale di \mathbf{H} costituita dai vettori:

$$\mathbf{b}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{b}_2 = \begin{bmatrix} -1/2 \\ -1/2 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{b}_3 = \begin{bmatrix} -1/3 \\ -1/3 \\ -1/3 \\ 1 \end{bmatrix}$$

Possiamo rendere questa base ortonormale dividendo i vettori \mathbf{b}_k per la loro norma. Si noti che è possibile e conveniente sostituire \mathbf{b}_2 con $2\mathbf{b}_2$ prima di calcolare la norma per evitare calcoli con le frazioni:

$$\mathbf{q}_2 = \frac{\mathbf{b}_2}{\|\mathbf{b}_2\|} = \frac{2\mathbf{b}_2}{\|2\mathbf{b}_2\|} = \frac{1}{\sqrt{1+1+4}}[-1, -1, 2, 0]^T = \frac{1}{\sqrt{6}}[-1, -1, 2, 0]^T$$

Analogamente.

$$\mathbf{q}_3 = \frac{\mathbf{b}_3}{\|\mathbf{b}_3\|} = \frac{3\mathbf{b}_3}{\|3\mathbf{b}_3\|} = \frac{1}{\sqrt{1+1+1+9}}[-1, -1, -1, 3]^T = \frac{1}{2\sqrt{3}}[-1, -1, -1, 3]^T$$

Una base ortonormale di \mathbf{H} è quindi costituita dai vettori:

$$\mathbf{q}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{q}_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ -1 \\ 2 \\ 0 \end{bmatrix} \quad \mathbf{q}_3 = \frac{1}{2\sqrt{3}} \begin{bmatrix} -1 \\ -1 \\ -1 \\ 3 \end{bmatrix}$$

presente che l'integrale di una funzione dispari su $[-1, 1]$ è nullo):

$$\begin{aligned}\mathbf{b}_1 &= \mathbf{v}_1 = 1, \quad \|\mathbf{b}_1\|^2 = \int_{-1}^1 1^2 dx = 2 \\ \mathbf{b}_2 &= \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{b}_1 \rangle}{\|\mathbf{b}_1\|^2} \mathbf{b}_1 = x - \frac{1}{2} \int_{-1}^1 x dx = x, \quad \|\mathbf{b}_2\|^2 = \int_{-1}^1 x^2 dx = \frac{2}{3} \\ \mathbf{b}_3 &= \mathbf{v}_3 - \frac{\langle \mathbf{v}_3, \mathbf{b}_1 \rangle}{\|\mathbf{b}_1\|^2} \mathbf{b}_1 - \frac{\langle \mathbf{v}_3, \mathbf{b}_2 \rangle}{\|\mathbf{b}_2\|^2} \mathbf{b}_2 = \\ &= x^2 - \frac{1}{2} \int_{-1}^1 x^2 dx - \frac{3}{2} \left(\int_{-1}^1 x^3 dx \right) x = x^2 - \frac{1}{3} \\ \mathbf{b}_4 &= \mathbf{v}_4 - \frac{\langle \mathbf{v}_4, \mathbf{b}_1 \rangle}{\|\mathbf{b}_1\|^2} \mathbf{b}_1 - \frac{\langle \mathbf{v}_4, \mathbf{b}_2 \rangle}{\|\mathbf{b}_2\|^2} \mathbf{b}_2 - \frac{\langle \mathbf{v}_4, \mathbf{b}_3 \rangle}{\|\mathbf{b}_3\|^2} \mathbf{b}_3 = \\ &= x^3 - 0 - \frac{3}{2} \left(\int_{-1}^1 x^4 dx \right) x - 0 = x^3 - \frac{3}{5}x\end{aligned}$$

I polinomi così ottenuti sono, a meno di una costante, i primi 4 polinomi di Legendre. Verifichiamo per esempio che \mathbf{b}_2 e \mathbf{b}_4 sono ortogonali:

$$\langle \mathbf{b}_2, \mathbf{b}_4 \rangle = \int_{-1}^1 x \left(x^3 - \frac{3}{5}x \right) dx = 2 \left[\frac{x^5}{5} - \frac{x^4}{5} \right]_0^1 = 0$$

La fattorizzazione QR

Applicando l'algoritmo di Gram-Schmidt alle colonne di una matrice invertibile \mathbf{A} si ottiene la cosiddetta fattorizzazione **QR**.

PROPOSIZIONE 5.5 (Fattorizzazione QR)

Sia \mathbf{A} una matrice quadrata invertibile a coefficienti reali. Allora esiste una fattorizzazione di \mathbf{A} come prodotto:

$$(5.5) \qquad \mathbf{A} = \mathbf{Q}\mathbf{R}$$

dove \mathbf{Q} è una matrice ortogonale ed \mathbf{R} è una matrice triangolare alta con tutti gli elementi sulla diagonale principale positivi.

DIMOSTRAZIONE. Sia d l'ordine di \mathbf{A} e siano $\mathbf{v}_1, \dots, \mathbf{v}_d \in \mathbb{R}^d$ le colonne di \mathbf{A} . Siccome \mathbf{A} è invertibile, le sue colonne formano una base di \mathbb{R}^d . L'algoritmo di Gram-Schmidt produce una base ortogonale $\{\mathbf{b}_1, \dots, \mathbf{b}_d\}$ di \mathbb{R}^d tale che

$$\begin{aligned}\mathbf{v}_1 &= \mathbf{b}_1 \\ \mathbf{v}_2 &= \hat{x}_2^1 \mathbf{b}_1 + \mathbf{b}_2 \\ \mathbf{v}_3 &= \hat{x}_3^1 \mathbf{b}_1 + \hat{x}_3^2 \mathbf{b}_2 + \mathbf{b}_3\end{aligned}$$

$$\mathbf{v}_d = \hat{x}_d^1 \mathbf{b}_1 + \cdots + \hat{x}_d^{d-1} \mathbf{b}_{d-1} + \mathbf{b}_d$$

In termini di matrici abbiamo:

$$[\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3 \ \cdots \ \mathbf{v}_d] = [\mathbf{b}_1 \ \mathbf{b}_2 \ \mathbf{b}_3 \ \cdots \ \mathbf{b}_d] \begin{bmatrix} 1 & \hat{x}_2^1 & \hat{x}_3^1 & \cdots & \hat{x}_{d-1}^1 & \hat{x}_d^1 \\ 0 & 1 & \hat{x}_3^2 & \cdots & \hat{x}_{d-1}^2 & \hat{x}_d^2 \\ 0 & 0 & 1 & \cdots & \hat{x}_{d-1}^3 & \hat{x}_d^3 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \hat{x}_d^{d-1} \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

ovvero

$$\mathbf{A} = \left[\frac{\mathbf{b}_1}{\|\mathbf{b}_1\|} \ \frac{\mathbf{b}_2}{\|\mathbf{b}_2\|} \ \cdots \ \frac{\mathbf{b}_d}{\|\mathbf{b}_d\|} \right] \begin{bmatrix} \|\mathbf{b}_1\| \ \hat{x}_2^1 \ \|\mathbf{b}_1\| & \cdots & \hat{x}_{d-1}^1 \ \|\mathbf{b}_1\| & \hat{x}_d^1 \ \|\mathbf{b}_1\| \\ 0 & \|\mathbf{b}_2\| & \cdots & \hat{x}_{d-1}^2 \ \|\mathbf{b}_2\| & \hat{x}_d^2 \ \|\mathbf{b}_2\| \\ 0 & 0 & \|\mathbf{b}_{d-1}\| & \hat{x}_d^{d-1} \ \|\mathbf{b}_{d-1}\| \\ 0 & 0 & 0 & \|\mathbf{b}_d\| \end{bmatrix}$$

Siccome i vettori \mathbf{b}_i sono a due a due ortogonali, il primo fattore a secondo membro dell'ugualianza è una matrice ortogonale \mathbf{Q} ed evidentemente il secondo fattore è una matrice triangolare alta con tutti gli elementi sulla diagonale principale positivi.

OSSERVAZIONE Più in generale, se \mathbf{A} è una matrice $m \times d$ le cui colonne sono linearmente indipendenti, la dimostrazione precedente produce una fattorizzazione \mathbf{QR} , dove \mathbf{Q} è una matrice $m \times d$ con colonne a due a due ortogonali e di norma uno, ed \mathbf{R} è una matrice $d \times d$ triangolare alta. Si osservi che $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$ (matrice identità $d \times d$) anche quando \mathbf{A} non è quadrata (e quindi \mathbf{Q} non è una matrice ortogonale perché non è quadrata).

ESERCIZIO 8.10

Consideriamo la matrice

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

le cui prime 3 colonne formano la base dell'iperpiano $x_1 + x_2 + x_3 = 0$ a cui abbiamo già applicato l'algoritmo di Gram-Schmidt a pagina 397; la quarta colonna è ortogonale alle prime 3, per cui

$$\begin{aligned} \begin{bmatrix} -1 & -1 & -1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} &= \begin{bmatrix} -1 & -1/2 & -1/3 & 1 \\ 1 & -1/2 & -1/3 & 1 \\ 0 & 1 & -1/3 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1/2 & 1/2 & 0 \\ 0 & 1 & 1/3 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} -\sqrt{2}/2 & -\sqrt{6}/6 & -\sqrt{3}/6 & 1/2 \\ \sqrt{2}/2 & -\sqrt{6}/6 & -\sqrt{3}/6 & 1/2 \\ 0 & \sqrt{6}/3 & -\sqrt{3}/6 & 1/2 \\ 0 & 0 & \sqrt{3}/2 & 1/2 \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & \sqrt{6}/2 & \sqrt{6}/6 & 0 \\ 0 & 0 & 2\sqrt{3}/3 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} = \mathbf{QR} \end{aligned}$$

OSSERVAZIONE Il caso più semplice dell'algoritmo di Gram-Schmidt è in realtà familiare al lettore fin da quando ha imparato a calcolare l'area di un parallelogramma con la formula base per altezza. Supponiamo infatti che il parallelogramma sia generato dai vettori \mathbf{v}_1 e \mathbf{v}_2 . Se prendiamo $\mathbf{b}_1 = \mathbf{v}_1$ come base, allora la proiezione ortogonale \mathbf{b}_2 di \mathbf{v}_2 nella direzione ortogonale alla base è l'altezza del parallelogramma. Quindi l'area del parallelogramma è $\|\mathbf{b}_1\| \|\mathbf{b}_2\|$. Analogamente, si può definire il *volume* del parallelepipedo generato dai vettori $\mathbf{v}_1, \dots, \mathbf{v}_d$ mediante la formula base per altezza: per far questo, prendiamo come base il parallelepipedo generato dai primi $d-1$ vettori e come altezza il vettore \mathbf{b}_d ottenuto (come nell'algoritmo di Gram-Schmidt) proiettando \mathbf{v}_d sulla retta ortogonale alla base nello spazio generato da $\mathbf{v}_1, \dots, \mathbf{v}_{d-1}$. Per induzione si vede immediatamente che il volume del parallelepipedo è il prodotto delle norme dei vettori \mathbf{b}_k costruiti con l'algoritmo di Gram-Schmidt. In termini della fattorizzazione **QR**, il volume è quindi il determinante della matrice triangolare **R**. Osserviamo poi che il determinante di **Q** è ± 1 perché **Q** è ortogonale. Dal teorema di Binet segue che il volume del parallelepipedo è uguale al modulo del determinante della matrice **A** che ha per colonne i vettori \mathbf{v}_k . In conclusione, se definiamo il volume di un parallelepipedo come volume della base per norma dell'altezza, tale volume è il modulo del determinante della matrice che ha come colonne gli spigoli del parallelepipedo uscenti dall'origine.

Riepilogo e complemento ortogonale

Riassumiamo nel seguente teorema quanto abbiamo visto per le proiezioni ortogonali.

TEOREMA 5.6 (Esistenza e unicità proiezione ortogonale)

Sia **H** un sottospazio di dimensione finita di uno spazio euclideo **V**. Allora:

- per ogni $\mathbf{v} \in \mathbf{V}$ esiste un unico vettore $\mathbf{v}_H \in \mathbf{H}$, detto proiezione ortogonale di \mathbf{v} su **H**, tale che $\mathbf{v} - \mathbf{v}_H \in \mathbf{H}^\perp$;
- la proiezione ortogonale \mathbf{v}_H è il vettore di **H** a distanza minima da \mathbf{v} :

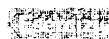
$$(5.6) \quad \|\mathbf{v} - \mathbf{v}_H\| < \|\mathbf{v} - \mathbf{w}\| \quad \text{per ogni } \mathbf{w} \in \mathbf{H}, \mathbf{w} \neq \mathbf{v}_H$$

- se $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_d\}$ è una base *ortogonale* di **H**, allora

$$(5.7) \quad \mathbf{v}_H = \sum_{k=1}^d \frac{\langle \mathbf{v}, \mathbf{b}_k \rangle}{\|\mathbf{b}_k\|^2} \mathbf{b}_k$$

DIMOSTRAZIONE. Per ipotesi **H** ha una base finita. Con l'algoritmo di Gram-Schmidt possiamo produrre una base ortogonale di **H**. Quindi per la proposizione 5.3 la proiezione ortogonale esiste ed è data dalla formula (5.7). Abbiamo già dimostrato la proprietà di minimo e l'unicità nella proposizione 5.3.

OSSERVAZIONE Si ricordi che, se **V** ha dimensione finita, ogni sottospazio **H** di **V** ha dimensione finita e, quindi, soddisfa le ipotesi del teorema precedente. Quando



Sia \mathbf{H} il piano di \mathbb{R}^3 di equazione $3x - y + z = 0$. Scrivere il vettore $\mathbf{v} = [1, 2, 3]^T$ come somma di un vettore di \mathbf{H} e di un vettore perpendicolare ad \mathbf{H} . Calcolare la distanza di \mathbf{v} da \mathbf{H} .

Trovare una base ortogonale $\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$ di \mathbb{R}^3 tale che il sottospazio generato da \mathbf{q}_1 e \mathbf{q}_2 coincida lo spazio colonna della matrice

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Sia \mathbf{H} il piano di \mathbb{R}^4 di equazioni $x_1 + x_2 + x_3 = x_2 + x_3 + x_4 = 0$. Trovare una base ortonormale di \mathbf{H} e una base ortonormale di \mathbf{H}^\perp . Determinare le proiezioni ortogonali del vettore $\mathbf{v} = [3, 2, 1, 0]^T$ su \mathbf{H} e su \mathbf{H}^\perp . Calcolare la distanza di \mathbf{v} da \mathbf{H} e da \mathbf{H}^\perp .

Sia \mathbf{V} lo spazio vettoriale delle funzioni continue $[0, 1] \rightarrow \mathbb{R}$ con la norma L^2 . Applicare l'algoritmo di Gram-Schmidt all'insieme $\{1, x, x^2\}$.

Sia \mathbf{V} lo spazio delle funzioni continue su $[-1, 1]$ con la norma L^2 . Sia \mathbf{H} il sottospazio di \mathbf{V} generato dai polinomi (a due a due ortogonali) $\{1, x, x^2 - \frac{1}{3}\}$. Trovare la proiezione ortogonale di $\sin(\pi x)$ su \mathbf{H} e calcolare la distanza di $\sin(\pi x)$ da \mathbf{H} . Quanto dista $\sin(x)$ da x ?

In \mathbb{R}^3 si consideri il prodotto scalare $\langle \mathbf{x}, \mathbf{y} \rangle = x_1y_1 + 2x_2y_2 + 3x_3y_3$. Applicare l'algoritmo di Gram-Schmidt ai vettori della base canonica per ottenere una base di \mathbb{R}^3 ortogonale rispetto a questo nuovo prodotto scalare.

6 EQUAZIONI NORMALI E IL METODO DEI MINIMI QUADRATI

Nelle applicazioni ci si imbatte spesso in sistemi lineari sovradeterminati. Il motivo è presto spiegato: si supponga che si debba determinare un vettore \mathbf{x} attraverso degli esperimenti e che ciascun esperimento fornisca un'equazione lineare che il vettore \mathbf{x} deve soddisfare. A causa degli errori sperimentali, non ci si può aspettare che queste equazioni siano soddisfatte esattamente, per cui per cercare di avere una stima accurata del risultato si fanno in genere molti più esperimenti del numero di incognite. Questo conduce a sistemi sovradeterminati con un grande numero di equazioni. In questo caso la soluzione non c'è, ma è interessante trovare il vettore $\bar{\mathbf{x}}$ che minimizza in un senso opportuno gli errori sperimentali. Per formalizzare questa situazione, abbiamo ora a disposizione la nozione di distanza $\|\mathbf{v} - \mathbf{w}\|$ tra due vettori dello spazio euclideo \mathbb{R}^n e possiamo risolvere questo problema di minimo col metodo delle proiezioni ortogonali. Siccome $\|\mathbf{v} - \mathbf{w}\|^2$ è una somma di quadrati, si parla di metodo dei minimi quadrati.

Facciamo subito l'esempio fondamentale della regressione lineare nel suo caso più semplice. Il problema da risolvere è il seguente. Si supponga che una variabile y sia funzione di una

variabile x e che la teoria preveda che, a meno di una costante, y dipenda linearmente da x e quindi abbia la forma $y = C + Dx$ dove C e D sono delle costanti da determinare. Per determinare C e D , si fanno degli esperimenti dando a x un certo numero dei valori x_1, \dots, x_N e misurando il corrispondente valore y_i di y . Se le misure fossero esatte, si avrebbe $y_i = C + Dx_i$ per ogni i e con sole due misurazioni si troverebbero i valori di C e D ; geometricamente, la retta $y = C + Dx$ è determinata se ne conosciamo due punti (x_1, y_1) e (x_2, y_2) . Ma nella realtà le misurazioni hanno sempre un errore. Facciamo l'ipotesi che i valori x_i assegnati alla variabile indipendente x siano esatti e che l'errore sia dovuto alla misurazione del valore di y : è naturale allora cercare di trovare C e D in modo da minimizzare l'errore tra il valore misurato y_i e il valore previsto dalla teoria che è $C + Dx_i$; se abbiamo N misurazioni, vogliamo minimizzare la differenza tra il vettore delle misurazioni $[y_1, \dots, y_N]^T$ e il vettore $[C + Dx_1, \dots, C + Dx_N]^T$ ed è naturale quantificare questo errore con la distanza in \mathbb{R}^N . Quindi il problema diviene: trovare C e D che minimizzano l'errore $e(C, D)$ dove

$$e(C, D)^2 = \sum_{i=1}^N (y_i - C - Dx_i)^2$$

L'interpretazione geometrica è che vogliamo determinare la retta $y = C + Dx$ che passa più vicina ai punti $P_i = (x_i, y_i)$ nel senso seguente: denotato con $Q_i = (x_i, C + Dx_i)$ il punto sulla retta che ha la stessa ascissa di P_i , la retta desiderata è quella che minimizza la norma euclidea del vettore degli errori $y(P_i) - y(Q_i)$. Tale retta si dice *retta di regressione lineare* ai minimi quadrati per i punti assegnati. Riformuliamo il problema in termini di sistemi lineari: se i punti (x_i, y_i) fossero allineati, il vettore $[C, D]^T$ risolverebbe il sistema lineare

$$(6.1) \quad \begin{cases} C + x_1 D = y_1 \\ C + x_2 D = y_2 \\ \vdots & \vdots \\ C + x_N D = y_N \end{cases}$$

Si tratta di un sistema di N equazioni nelle 2 incognite C e D . Per scrivere il sistema come d'abitudine nella forma $\mathbf{Ax} = \mathbf{b}$ dobbiamo porre

$$(6.2) \quad \mathbf{x} = \begin{bmatrix} C \\ D \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

In generale, per $N > 2$, il sistema non ha soluzioni. Quello che possiamo fare è cercare il vettore $\bar{\mathbf{x}}$ che minimizzi l'errore $e(C, D) = \|\mathbf{b} - \mathbf{Ax}\|$. In altre parole, tra tutti i vettori della forma \mathbf{Ax} , cerchiamo quello più vicino a \mathbf{b} :

$$\|\mathbf{A}\bar{\mathbf{x}} - \mathbf{b}\| \leq \|\mathbf{Ax} - \mathbf{b}\| \quad \text{per ogni } \mathbf{x} \in \mathbb{R}^2$$

Riconosciamo il legame con le proiezioni ortogonali: per la proprietà di minimo delle proiezioni, l'ultima disequazione mostra che $\mathbf{A}\bar{\mathbf{x}}$ è la proiezione ortogonale di \mathbf{b} sullo spazio dei vettori della forma \mathbf{Ax} , cioè sullo spazio colonna di \mathbf{A} .

Motivati dall'esempio della regressione, affrontiamo il problema più in generale: supponiamo di voler risolvere un sistema $\mathbf{Ax} = \mathbf{b}$ di m equazioni in n incognite, verosimilmente sovrardeterminato. Il sistema ha soluzione se e solo se il termine noto \mathbf{b} appartiene allo spazio colonna $\mathbf{H} = \text{Col}(\mathbf{A})$. Altrimenti il sistema è impossibile, ma possiamo sempre risolvere il sistema $\mathbf{Ax} = \mathbf{b}_\mathbf{H}$, dove $\mathbf{b}_\mathbf{H}$ è la proiezione ortogonale

di \mathbf{b} sullo spazio colonna \mathbf{H} . Rispetto alla distanza euclidea, questo è il meglio che possiamo fare perché $\mathbf{b}_\mathbf{H}$ è il vettore della forma \mathbf{Ax} , cioè dello spazio colonna, più vicino al termine noto: se $\bar{\mathbf{x}}$ risolve $\mathbf{Ax} = \mathbf{b}_\mathbf{H}$, allora

$$(6.3) \quad \|\mathbf{A}\bar{\mathbf{x}} - \mathbf{b}\| \leq \|\mathbf{Ax} - \mathbf{b}\| \quad \text{per ogni } x \in \mathbb{R}^n$$

per la proprietà di minimo della proiezione ortogonale. Se non si può risolvere $\mathbf{Ax} = \mathbf{b}$, si può almeno minimizzare l'errore $\|\mathbf{Ax} - \mathbf{b}\|$. Questo motiva la definizione seguente:

DEFINIZIONE 6.1 (Soluzione ai minimi quadrati)

Sia $\mathbf{Ax} = \mathbf{b}$ un sistema lineare di m equazioni a coefficienti reali in n incognite.

Si dice che $\bar{\mathbf{x}}$ è una *soluzione ai minimi quadrati* di $\mathbf{Ax} = \mathbf{b}$ se

$$(6.4) \quad \mathbf{A}\bar{\mathbf{x}} = \mathbf{b}_\mathbf{H}$$

dove $\mathbf{b}_\mathbf{H}$ è la proiezione ortogonale di \mathbf{b} sullo spazio colonna $\mathbf{H} = \text{Col}(\mathbf{A})$.

Equivalentemente, $\bar{\mathbf{x}}$ è una soluzione ai minimi quadrati se vale la (6.3), cioè $\bar{\mathbf{x}}$ minimizza la distanza euclidea (o scarto quadratico) di \mathbf{Ax} da \mathbf{b} .

Possiamo facilmente determinare il complemento ortogonale dello spazio colonna e degli altri spazi naturalmente associati a una matrice.

PROPOSIZIONE 6.2 (Complemento ortogonale degli spazi associati a una matrice) Per ogni matrice reale \mathbf{A} , il nucleo $\text{Ker}(\mathbf{A})$ è il complemento ortogonale dello spazio riga di \mathbf{A} :

$$(6.5) \quad \text{Ker}(\mathbf{A}) = \text{Row}(\mathbf{A})^\perp \quad \text{e} \quad \text{Row}(\mathbf{A}) = \text{Ker}(\mathbf{A})^\perp$$

e il nucleo $\text{Ker}(\mathbf{A}^T)$ della matrice trasposta è il complemento ortogonale dello spazio colonna di \mathbf{A} :

$$(6.6) \quad \text{Ker}(\mathbf{A}^T) = \text{Col}(\mathbf{A})^\perp \quad \text{e} \quad \text{Col}(\mathbf{A}) = \text{Ker}(\mathbf{A}^T)^\perp$$

DIMOSTRAZIONE. Supponiamo che \mathbf{A} sia una matrice $m \times n$ e che $\mathbf{v}_1, \dots, \mathbf{v}_m$ siano i vettori riga della matrice: si tratta di vettori di \mathbb{R}^n . Il prodotto di \mathbf{A} per un vettore $\mathbf{x} \in \mathbb{R}^n$ è

$$\mathbf{Ax} = \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_m^T \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{v}_1^T \mathbf{x} \\ \vdots \\ \mathbf{v}_m^T \mathbf{x} \end{bmatrix}$$

Quindi $\mathbf{Ax} = \mathbf{0}$, cioè \mathbf{x} è un vettore del nucleo $\text{Ker}(\mathbf{A})$, se e solo se

$$\mathbf{v}_1^T \mathbf{x} = \dots = \mathbf{v}_m^T \mathbf{x} = 0$$

cioè \mathbf{x} è ortogonale alle righe di \mathbf{A} . Siccome lo spazio riga di \mathbf{A} è il sottospazio di \mathbb{R}^n generato dai vettori riga, questo significa che un vettore appartiene al nucleo se e solo se è ortogonale allo spazio riga. Quindi il nucleo è il complemento ortogonale dello spazio riga:

$\text{Ker}(\mathbf{A}) = \text{Row}(\mathbf{A})^\perp$. Per le proprietà del complemento ortogonale (corollario 5.7) da questo segue $\text{Row}(\mathbf{A}) = \text{Ker}(\mathbf{A})^\perp$.

Tenendo conto che lo spazio colonna di \mathbf{A} coincide con lo spazio di riga di \mathbf{A}^T , per quanto abbiamo appena dimostrato il complemento ortogonale dello spazio colonna è il nucleo di \mathbf{A}^T .

Le soluzioni ai minimi quadrati di un sistema $\mathbf{Ax} = \mathbf{b}$ si possono trovare risolvendo un sistema lineare associato, che si dice *sistema delle equazioni normali*:

PROPOSIZIONE 6.3 (Equazioni normali)

Un vettore $\bar{\mathbf{x}}$ è una soluzione ai minimi quadrati di $\mathbf{Ax} = \mathbf{b}$ se e solo se è una soluzione del sistema delle equazioni normali:

$$(6.7) \quad \mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b}$$

DIMOSTRAZIONE. Supponiamo dapprima che $\bar{\mathbf{x}}$ sia una soluzione ai minimi quadrati di $\mathbf{Ax} = \mathbf{b}$, cioè che $\mathbf{A}\bar{\mathbf{x}} = \mathbf{b}_H$ dove \mathbf{b}_H è la proiezione ortogonale di \mathbf{b} sullo spazio colonna \mathbf{H} di \mathbf{A} . Per definizione di proiezione ortogonale, la differenza $\mathbf{b} - \mathbf{b}_H$ appartiene a \mathbf{H}^\perp , il complemento ortogonale dello spazio colonna, che per la proposizione 6.2 è il nucleo di \mathbf{A}^T : questo significa che

$$\mathbf{A}^T(\mathbf{b} - \mathbf{b}_H) = \mathbf{0}$$

Quindi

$$\mathbf{A}^T \mathbf{A} \bar{\mathbf{x}} = \mathbf{A}^T \mathbf{b}_H = \mathbf{A}^T \mathbf{b}$$

Questo mostra che $\bar{\mathbf{x}}$ soddisfa le equazioni normali (6.7).

Viceversa, supponiamo che $\mathbf{A}^T \mathbf{A} \bar{\mathbf{x}} = \mathbf{A}^T \mathbf{b}$. Allora $\mathbf{b} - \mathbf{A} \bar{\mathbf{x}}$ appartiene al nucleo di \mathbf{A}^T . Per la (6.6) il nucleo di \mathbf{A}^T è \mathbf{H}^\perp . Scriviamo

$$\mathbf{b} = \mathbf{A} \bar{\mathbf{x}} + (\mathbf{b} - \mathbf{A} \bar{\mathbf{x}})$$

Il primo addendo $\mathbf{A} \bar{\mathbf{x}}$ appartiene allo spazio colonna \mathbf{H} e per quanto abbiamo appena osservato il secondo addendo $\mathbf{b} - \mathbf{A} \bar{\mathbf{x}}$ appartiene ad \mathbf{H}^\perp . Quindi $\mathbf{A} \bar{\mathbf{x}}$ è la proiezione ortogonale \mathbf{b}_H di \mathbf{b} su \mathbf{H} : questo significa precisamente che $\bar{\mathbf{x}}$ è una soluzione ai minimi quadrati di $\mathbf{Ax} = \mathbf{b}$ e la dimostrazione è completa.

OSSERVAZIONE Abbiamo ora a disposizione un nuovo metodo per calcolare la proiezione ortogonale $\mathbf{p} = \mathbf{b}_H$ di un vettore \mathbf{b} su $\mathbf{H} = \text{Col}(\mathbf{A})$: data una soluzione ai minimi quadrati $\bar{\mathbf{x}}$ di $\mathbf{Ax} = \mathbf{b}$, trovata per esempio risolvendo il sistema delle equazioni normali, possiamo ricavare $\mathbf{p} = \mathbf{A} \bar{\mathbf{x}}$.

Consideriamo il sistema lineare

$$(6.8) \quad \begin{cases} 2x - y = 10 \\ 2x - 3y = -4 \\ x + y = 3 \end{cases}$$

La matrice dei coefficienti del sistema e la matrice completa sono rispettivamente

$$\mathbf{A} = \begin{bmatrix} 2 & -1 \\ 2 & -3 \\ 1 & 1 \end{bmatrix} \quad [\mathbf{A}|\mathbf{b}] = \begin{bmatrix} 2 & -1 & 10 \\ 2 & -3 & -4 \\ 1 & 1 & 3 \end{bmatrix}$$

Un semplice calcolo mostra che $\det(\mathbf{A}|\mathbf{b}) \neq 0$, quindi $r(\mathbf{A}|\mathbf{b}) = 3$, mentre $r(\mathbf{A}) = 2$: il sistema è sovradeterminato. Per scrivere il sistema delle equazioni normali, scriviamo

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 9 & -7 \\ -7 & 11 \end{bmatrix} \quad \mathbf{A}^T \mathbf{b} = \begin{bmatrix} 15 \\ 5 \end{bmatrix}$$

Risolvendo il sistema delle equazioni normali

$$(6.9) \quad \begin{cases} 9x - 7y = 15 \\ -7x + 11y = 5 \end{cases}$$

si trova l'unica soluzione ai minimi quadrati $\bar{\mathbf{x}} = [4, 3]^T$. La proiezione ortogonale \mathbf{p} di \mathbf{b} sul piano generato dalle colonne di \mathbf{H} è pertanto

$$\mathbf{p} = \mathbf{A}\bar{\mathbf{x}} = \begin{bmatrix} 2 & -1 \\ 2 & -3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \\ 7 \end{bmatrix}$$

Per verificare la correttezza di questi conti, possiamo controllare che $\mathbf{b} - \mathbf{p} = [5, -3, -4]^T$ sia perpendicolare alle colonne \mathbf{v}_1 e \mathbf{v}_2 della matrice:

$$(\mathbf{b} - \mathbf{p})^T \mathbf{v}_2 = [5 \ -3 \ -4] \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} = 0, \quad (\mathbf{b} - \mathbf{p})^T \mathbf{v}_2 = [5 \ -3 \ -4] \begin{bmatrix} -1 \\ -3 \\ 1 \end{bmatrix} = 0$$

OSSERVAZIONE Si osservi che il sistema delle equazioni normali $\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{b}$ è sempre quadrato: se \mathbf{A} è una matrice $m \times n$, la matrice $\mathbf{A}^T \mathbf{A}$ è $n \times n$. Inoltre $\mathbf{A}^T \mathbf{A}$ non è una matrice qualsiasi, ma è *semidefinita positiva* (su questo punto importante torneremo più avanti).

Inoltre, il sistema delle equazioni normali ha sempre soluzione, qualunque sia \mathbf{b} , poiché è equivalente al sistema $\mathbf{A} \mathbf{x} = \mathbf{b}_{\mathbf{H}}$, che ha soluzioni perché $\mathbf{b}_{\mathbf{H}}$ appartiene allo spazio colonna di \mathbf{A} . Quindi, per ogni \mathbf{b} , il termine noto $\mathbf{A}^T \mathbf{b}$ del sistema delle equazioni normali deve appartenere allo spazio colonna di $\mathbf{A}^T \mathbf{A}$: da questa osservazione segue facilmente che gli spazi colonna di \mathbf{A}^T e $\mathbf{A}^T \mathbf{A}$ devono coincidere. In effetti:

PROPOSIZIONE 6.4 Per ogni matrice reale \mathbf{A} valgono le uguaglianze:

$$\text{Ker}(\mathbf{A}) = \text{Ker}(\mathbf{A}^T \mathbf{A}), \quad \text{Col}(\mathbf{A}^T) = \text{Col}(\mathbf{A}^T \mathbf{A})$$

In particolare, $r(\mathbf{A}) = r(\mathbf{A}^T \mathbf{A})$.

DIMOSTRAZIONE. Osserviamo che

$$\|\mathbf{Ax}\|^2 = (\mathbf{Ax})^T \mathbf{Ax} = \mathbf{x}^T \mathbf{A}^T \mathbf{Ax}$$

Ne segue che, se \mathbf{x} appartiene al nucleo di $\mathbf{A}^T \mathbf{A}$, allora $\|\mathbf{Ax}\|^2 = 0$ e quindi $\mathbf{Ax} = \mathbf{0}$, cioè \mathbf{x} appartiene al nucleo di \mathbf{A} . Viceversa, se \mathbf{x} appartiene al nucleo di \mathbf{A} , evidentemente \mathbf{x} appartiene anche al nucleo di $\mathbf{A}^T \mathbf{A}$. Quindi i due nuclei sono uguali. Dall'uguaglianza dei nuclei, per il teorema di nullità più rango, segue l'uguaglianza dei ranghi. Per quanto riguarda gli spazi colonna, si può ragionare così: $\text{Col}(\mathbf{A}^T)$, che è lo spazio riga di \mathbf{A} , è il complemento ortogonale del nucleo $\text{Ker}(\mathbf{A})$. Ma $\text{Ker}(\mathbf{A})$ coincide con $\text{Ker}(\mathbf{A}^T \mathbf{A})$, il cui complemento ortogonale è lo spazio riga di $\mathbf{A}^T \mathbf{A}$. Quindi $\text{Col}(\mathbf{A}^T)$ coincide con lo spazio riga di $\mathbf{A}^T \mathbf{A}$. Ma $\mathbf{A}^T \mathbf{A}$ è una matrice simmetrica, per cui $\text{Col}(\mathbf{A}^T)$ coincide anche con lo spazio colonna di $\mathbf{A}^T \mathbf{A}$.



COROLLARIO 6.5 Il sistema $\mathbf{Ax} = \mathbf{b}$ di m equazioni in n incognite ammette una e una sola soluzione ai minimi quadrati se e solo se $r(\mathbf{A}) = n$. Se questo è il caso, la matrice $\mathbf{A}^T \mathbf{A}$ è invertibile, la soluzione ai minimi quadrati è

$$(6.10) \quad \bar{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$$

e la proiezione di \mathbf{b} sullo spazio colonna di \mathbf{A} è

$$(6.11) \quad \mathbf{b}_H = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$$

DIMOSTRAZIONE. Il sistema delle equazioni normali $\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b}$ è un sistema quadrato di n equazioni in n incognite, che ammette sempre soluzioni. La soluzione è unica se il numero di incognite n è uguale al rango $r(\mathbf{A}^T \mathbf{A})$ della matrice dei coefficienti e questo rango coincide con il rango di \mathbf{A} per la proposizione precedente. Quindi, se $r(\mathbf{A}) = n$, la matrice quadrata $\mathbf{A}^T \mathbf{A}$ ha rango massimo ed è perciò invertibile. Per il teorema di Cramer l'unica soluzione del sistema delle equazioni normali è $\bar{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$. La proiezione \mathbf{b}_H si trova come abbiamo osservato moltiplicando $\bar{\mathbf{x}}$ a sinistra per \mathbf{A} .



OSSERVAZIONE Naturalmente le formule (6.10) e la (6.11) non devono essere usate per trovare negli esercizi la soluzione ai minimi quadrati $\bar{\mathbf{x}}$ e la proiezione ortogonale \mathbf{b}_H : come sempre, il calcolo della matrice inversa è computazionalmente costoso e la regola di Cramer non è conveniente per risolvere un sistema (in questo caso quello delle equazioni normali). Conviene per esempio risolvere il sistema delle equazioni normali con il metodo di Gauss (o con qualche altro efficace algoritmo di soluzione).

Regressione lineare

Risolviamo ora con il metodo dei minimi quadrati il problema della regressione lineare: vogliamo determinare la soluzione ai minimi quadrati del sistema (6.1). Per questo

calcoliamo

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} = \begin{bmatrix} N & \sum_{i=1}^N x_i \\ \sum_{i=1}^N x_i & \sum_{i=1}^N x_i^2 \end{bmatrix}$$

$$\mathbf{A}^T \mathbf{b} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_i y_i \end{bmatrix}$$

per cui il sistema delle equazioni normali è

$$\begin{bmatrix} N & \sum_{i=1}^N x_i \\ \sum_{i=1}^N x_i & \sum_{i=1}^N x_i^2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_i y_i \end{bmatrix}$$

che (a condizione che almeno due degli x_i siano distinti) ha un'unica soluzione calcolabile con la formula di Cramer:

$$(6.12) \quad \begin{cases} C = \frac{\sum_i x_i^2 \sum_i y_i - \sum_i x_i \sum_i x_i y_i}{N \sum_i x_i^2 - (\sum_i x_i)^2} \\ D = \frac{N \sum_i x_i y_i - \sum_i x_i \sum_i y_i}{N \sum_i x_i^2 - (\sum_i x_i)^2} \end{cases}$$

Supponiamo di voler trovare la retta di regressione lineare ai minimi quadrati per i punti $P_1 = (-2, 4)$, $P_2 = (-1, 3)$, $P_3 = (0, 1)$ e $P_4 = (2, 0)$. Il sistema da risolvere ai minimi quadrati è

$$\begin{cases} C - 2D = 4 \\ C - D = 3 \\ C = 1 \\ C + 2D = 0 \end{cases}$$

Il corrispondente sistema di equazioni normali è

$$\begin{cases} 4C - D = 8 \\ -C + 9D = -11 \end{cases}$$

che ha come soluzione $[C, D] = [61/35, -36/35]$.

Dato il sistema lineare

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

calcolare il rango della matrice dei coefficienti \mathbf{A} e della matrice completa $[\mathbf{A}|\mathbf{b}]$ del sistema e dedurre che il sistema non ammette soluzioni. Determinare quindi le soluzioni ai minimi quadrati del sistema e la proiezione ortogonale \mathbf{p} del termine noto \mathbf{b} sullo spazio colonna di \mathbf{A} . Verificare infine che $\mathbf{b} - \mathbf{p}$ è perpendicolare alle colonne di \mathbf{A} .

Dati la matrice $\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -2 \\ 0 & 1 & 1 \end{bmatrix}$ e il vettore $\mathbf{b} = [1, 3, -2]^T$, si determini la proiezione ortogonale di \mathbf{b} sullo spazio colonna di \mathbf{A} e si scomponga il vettore \mathbf{b} nella forma $\mathbf{b} = \mathbf{v} + \mathbf{w}$ con \mathbf{v} appartenente a $\text{Col}(\mathbf{A})$ e \mathbf{w} appartenente al sottospazio ortogonale a $\text{Col}(\mathbf{A})$.

Trovare la soluzione ai minimi quadrati del sistema $\mathbf{Ax} = \mathbf{b}$ nel caso in cui $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ -1 & 1 \end{bmatrix}$ e $\mathbf{b} = \begin{bmatrix} 5 \\ 4 \\ -2 \end{bmatrix}$. Determinare inoltre la proiezione \mathbf{p} di \mathbf{b} sullo spazio colonna di \mathbf{A} e verificare che $\mathbf{b} - \mathbf{p}$ è perpendicolare alle colonne di \mathbf{A} .

Si consideri il sistema

$$\begin{cases} kx + y = k \\ x + ky = 1 \\ x + y = k \end{cases}$$

dove k è un parametro reale:

- (a) determinare per quali valori del parametro k il sistema ammette soluzioni;
- (b) in corrispondenza di tali valori del parametro determinare tutte le soluzioni del sistema;
- (c) posto $k = 2$ determinare la soluzione ai minimi quadrati.

Trovare la retta di regressione ai minimi quadrati per i punti $(1, 3)$, $(2, 4)$, $(3, 7)$, $(5, 8)$.

In questo esercizio si determina una parabola anziché la retta ai minimi quadrati. Dati i punti $(x_1, y_1) = (-1, 1)$, $(x_2, y_2) = (0, 1)$, $(x_3, y_3) = (1, 4)$ e $(x_4, y_4) = (2, 7)$, si determini la parabola $y(x) = ax^2 + bx + c$ che minimizza l'errore

$$\|\mathbf{b} - \mathbf{Ax}\|^2 = \sum_{k=1}^4 (y_k - y(x_k))^2$$

Spiegare perché, se $\mathbf{Ax} = \mathbf{b}$ ammette soluzioni, queste ultime coincidono con le soluzioni ai minimi quadrati.

Si supponga che \mathbf{A} abbia n colonne e rango r . Mostrare che le soluzioni ai minimi quadrati di $\mathbf{Ax} = \mathbf{b}$ dipendono da $n - r$ parametri.

Mostrare che, per ogni matrice reale \mathbf{A} , valgono le uguaglianze $\text{Ker}(\mathbf{A}^T) = \text{Ker}(\mathbf{AA}^T)$ e $\text{Col}(\mathbf{A}) = \text{Col}(\mathbf{AA}^T)$. In particolare, $r(\mathbf{AA}^T) = r(\mathbf{A}^T) = r(\mathbf{A})$. (Attenzione: \mathbf{AA}^T è una matrice simmetrica e non è la trasposta di $\mathbf{A}^T\mathbf{A}$, che è anch'essa simmetrica, e quindi coincide con la propria trasposta.)

 Sia \mathbf{A} una matrice $m \times n$ e sia $\mathcal{L} = \mathcal{L}_{\mathbf{A}} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ l'applicazione lineare rappresentata da \mathbf{A} : $\mathcal{L}(\mathbf{x}) = \mathbf{Ax}$. Il dominio di \mathcal{L} si decompona come somma diretta ortogonale $\mathbb{R}^n = \text{Row}(\mathbf{A}) \oplus \text{Ker}(\mathbf{A})$ e il codominio come $\mathbb{R}^m = \text{Col}(\mathbf{A}) \oplus \text{Ker}(\mathbf{A}^T)$. Mostrare che la restrizione di \mathcal{L} alla spazio riga è un isomorfismo dello spazio riga sulla spazio colonna. Questo spiega (nel caso reale) perché lo spazio riga e lo spazio colonna hanno la stessa dimensione.

7 MATRICI DI PROIEZIONI ORTOGONALI

Sia \mathbf{H} un sottospazio di \mathbb{R}^n . La funzione $\mathbb{R}^n \rightarrow \mathbb{R}^n$ che a un vettore \mathbf{v} associa la proiezione ortogonale $\mathbf{v}_{\mathbf{H}}$ è lineare. Quindi esiste una matrice \mathbf{P} di tipo $n \times n$ che rappresenta la proiezione:

$$\mathbf{P}\mathbf{v} = \mathbf{v}_{\mathbf{H}} \quad \text{per ogni } \mathbf{v} \in \mathbb{R}^n$$

Vediamo come determinare \mathbf{P} . Supponiamo che $\mathbf{b}_1, \dots, \mathbf{b}_d$ sia una base di \mathbf{H} e che \mathbf{A} sia la matrice che ha per colonne i vettori \mathbf{b}_k . Si tratta di una matrice $n \times d$. Per costruzione \mathbf{H} è lo spazio colonna di \mathbf{A} e l'equazione (6.11) del corollario 6.5:

$$\mathbf{v}_{\mathbf{H}} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{v} \quad \text{per ogni } \mathbf{v} \in \mathbb{R}^n$$

mostra che

$$(7.1) \quad \mathbf{P} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$$

Questa formula è troppo complicata per molte applicazioni, ma si può semplificare se invece di prendere una base arbitraria se ne sceglie una ortonormale. Il risultato è il seguente:

PROPOSIZIONE 7.1 (Matrice di una proiezione ortogonale)

Sia \mathbf{H} un sottospazio di \mathbb{R}^n e sia $\{\mathbf{q}_1, \dots, \mathbf{q}_d\}$ una base ortonormale di \mathbf{H} . Sia \mathbf{A} la matrice che ha per colonne i vettori \mathbf{q}_k . Allora

$$(7.2) \quad \mathbf{P} = \mathbf{AA}^T = \mathbf{q}_1 \mathbf{q}_1^T + \cdots + \mathbf{q}_d \mathbf{q}_d^T$$

è la matrice che rappresenta la proiezione ortogonale di \mathbb{R}^n sul sottospazio \mathbf{H} .

DIMOSTRAZIONE. Si osservi che, se \mathbf{q} è un vettore di \mathbb{R}^n , la matrice \mathbf{qq}^T è quadrata di ordine n , perché \mathbf{q} è una matrice $n \times 1$ e \mathbf{q}^T è una matrice $1 \times n$. Diamo tre dimostrazioni dell'importante formula (7.1). Le prime due dimostrazioni corrispondono ai due diversi metodi che abbiamo a disposizione per calcolare le proiezioni ortogonali: il metodo delle equazioni normali da una parte, e la formula per la proiezione in termini di una base ortogonale della proposizione 5.3.

Prima dimostrazione Per il corollario 6.5

$$\mathbf{P} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$$

Per ipotesi le colonne \mathbf{q}_k di \mathbf{A} sono d versori di \mathbb{R}^n a due a due perpendicolari: questo significa

$$\mathbf{q}_i^T \mathbf{q}_j = \delta_{ij} \quad \text{per ogni } 1 \leq i, j \leq d$$

Quindi $\mathbf{A}^T \mathbf{A}$ è la matrice identità $d \times d$ e la formula per \mathbf{P} si semplifica nella forma $\mathbf{P} = \mathbf{A}\mathbf{A}^T$. Infine

$$\mathbf{A}\mathbf{A}^T = [\mathbf{q}_1 \cdots \mathbf{q}_d] \begin{bmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_d^T \end{bmatrix} = \mathbf{q}_1\mathbf{q}_1^T + \cdots + \mathbf{q}_d\mathbf{q}_d^T$$

Seconda dimostrazione Per la proposizione 5.3 la proiezione ortogonale $\mathbf{v}_{\mathbf{H}}$ di \mathbf{v} su \mathbf{H} è

$$(7.3) \quad \mathbf{v}_{\mathbf{H}} = (\mathbf{q}_1^T \mathbf{v}) \mathbf{q}_1 + \cdots + (\mathbf{q}_d^T \mathbf{v}) \mathbf{q}_d$$

Ora osserviamo che per ogni coppia di vettori \mathbf{v}, \mathbf{q} di \mathbb{R}^n

$$(7.4) \quad (\mathbf{q}^T \mathbf{v}) \mathbf{q} = (\mathbf{q}\mathbf{q}^T)\mathbf{v}$$

Sostituendo nella (7.3) si trova

$$(7.5) \quad \mathbf{v}_{\mathbf{H}} = (\mathbf{q}_1\mathbf{q}_1^T + \cdots + \mathbf{q}_d\mathbf{q}_d^T) \mathbf{v}$$

che equivale alla (7.1).

Terza dimostrazione Estendiamo $\{\mathbf{q}_1, \dots, \mathbf{q}_d\}$ a una base ortonormale $\{\mathbf{q}_1, \dots, \mathbf{q}_n\}$ di \mathbb{R}^n e sia \mathbf{Q} la matrice ortogonale di ordine n corrispondente: quindi $\mathbf{Q} = [\mathbf{A}|\mathbf{B}]$ dove \mathbf{B} è una matrice $n \times (n - d)$ le cui colonne formano una base ortonormale di \mathbf{H}^\perp . Ora si osservi che, se $\mathbf{v} \in \mathbf{H}$, allora $\mathbf{P}\mathbf{v} = \mathbf{v}$ perché la proiezione su \mathbf{H} di un vettore di \mathbf{H} è il vettore stesso. Nel linguaggio degli autovettori e autovalori, questo significa che un vettore di \mathbf{H} è un autovettore di \mathbf{P} relativo all'autovalore 1. Se invece $\mathbf{v} \in \mathbf{H}^\perp$, allora $\mathbf{P}\mathbf{v} = \mathbf{0}$: perché la proiezione su \mathbf{H} di un vettore perpendicolare ad \mathbf{H} è il vettore nullo. Quindi un vettore di \mathbf{H}^\perp è un autovettore di \mathbf{P} relativo all'autovalore 0. Le colonne di \mathbf{Q} formano perciò una base di \mathbb{R}^n costituita da autovettori di \mathbf{P} , i primi d relativi all'autovalore 1, gli ultimi $n - d$ relativi all'autovalore 0. Ne segue che

$$\mathbf{Q}^{-1}\mathbf{P}\mathbf{Q} = \text{diag}(1, \dots, 1, 0, \dots, 0) = \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}$$

dove \mathbf{I} denota la matrice identità $d \times d$. Siccome \mathbf{Q} è ortogonale, $\mathbf{Q}^{-1} = \mathbf{Q}^T$ e

$$\mathbf{P} = \mathbf{Q} \text{diag}(1, \dots, 1, 0, \dots, 0) \mathbf{Q}^T = [\mathbf{A}|\mathbf{B}] \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{A}^T \\ \mathbf{B}^T \end{bmatrix} = \mathbf{A}\mathbf{A}^T$$

OSSERVAZIONE Prendendo \mathbf{H} di dimensione 1 nella proposizione precedente, si vede che, se \mathbf{q} è un versore, la matrice $\mathbf{q}\mathbf{q}^T$ è la matrice di rango 1 che rappresenta la proiezione sulla retta generata da \mathbf{q} . In generale, \mathbf{P} è la somma delle matrici $\mathbf{q}_k\mathbf{q}_k^T$ di proiezione sugli assi $\mathbb{R}\mathbf{q}_k$.

OSSERVAZIONE La terza dimostrazione mostra che la proiezione ortogonale su un sottospazio \mathbf{H} di dimensione d di \mathbb{R}^n è diagonalizzabile mediante una matrice ortogonale. Gli autovalori sono $\lambda_1 = 1$, che ha molteplicità algebrica e geometrica d , e $\lambda_2 = 0$ che ha molteplicità algebrica e geometrica $n - d$. L'autospazio relativo all'autovalore 1 è \mathbf{H} , l'autospazio relativo all'autovalore nullo è \mathbf{H}^\perp .

Consideriamo in \mathbb{R}^3 il vettore

$$\mathbf{w} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Il versore $\mathbf{q} = \frac{\mathbf{w}}{\|\mathbf{w}\|}$ forma una base ortonormale della retta generata da \mathbf{w} . Perciò la matrice della proiezione su questa retta è:

$$\mathbf{P} = \mathbf{q}\mathbf{q}^T = \frac{1}{\|\mathbf{w}\|^2} \mathbf{w}\mathbf{w}^T = \frac{1}{14} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = \frac{1}{14} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$$

e, dato $\mathbf{v} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, la proiezione di \mathbf{v} sulla retta generata da \mathbf{w} è

$$\mathbf{v}_H = \frac{1}{14} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{14} \begin{bmatrix} x_1 + 2x_2 + 3x_3 \\ 2x_1 + 4x_2 + 6x_3 \\ 3x_1 + 6x_2 + 9x_3 \end{bmatrix}$$

Se invece consideriamo in \mathbb{R}^4 il vettore

$$\mathbf{w} = [1, 1, 1, 1]^T$$

allora

$$\mathbf{P} = \frac{1}{\mathbf{w}^T \mathbf{w}} \mathbf{w}\mathbf{w}^T = \frac{1}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

e, dato $\mathbf{v} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$, la proiezione di \mathbf{v} sulla retta generata da \mathbf{w} è

$$\mathbf{v}_H = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} x_1 + x_2 + x_3 + x_4 \\ x_1 + x_2 + x_3 + x_4 \\ x_1 + x_2 + x_3 + x_4 \\ x_1 + x_2 + x_3 + x_4 \end{bmatrix}$$

Possiamo infine *caratterizzare* le matrici di proiezioni, cioè elencare quelle proprietà che identificano le matrici di proiezione tra tutte le matrici quadrate di ordine n :

PROPOSIZIONE 7.2 (Caratterizzazione delle matrici di proiezioni ortogonali) Una matrice quadrata reale \mathbf{P} di ordine n è la matrice della proiezione ortogonale su un sottospazio \mathbf{H} di \mathbb{R}^n se e solo se:

a) \mathbf{P} è idempotente: $\mathbf{P}^2 = \mathbf{P}$;

b) \mathbf{P} è simmetrica: $\mathbf{P}^T = \mathbf{P}$.

Se a) e b) sono verificate, allora \mathbf{H} è lo spazio colonna di \mathbf{P} e \mathbf{H}^\perp è il nucleo di \mathbf{P} .

DIMOSTRAZIONE. Supponiamo che \mathbf{P} rappresenti la proiezione ortogonale su un sottospazio \mathbf{H} . Per mostrare che \mathbf{P} è idempotente, basta ricordare che la proiezione ortogonale di un vettore di \mathbf{H} è il vettore stesso, per cui per ogni \mathbf{v} di \mathbb{R}^n :

$$\mathbf{P}^2\mathbf{v} = \mathbf{P}(\mathbf{P}\mathbf{v}) = \mathbf{P}\mathbf{v}_{\mathbf{H}} = \mathbf{v}_{\mathbf{H}} = \mathbf{P}\mathbf{v}$$

Siccome questo vale per ogni \mathbf{v} , concludiamo che $\mathbf{P}^2 = \mathbf{P}$. La simmetria di \mathbf{P} segue dalla proposizione precedente, per la quale esiste una matrice \mathbf{A} tale che $\mathbf{P} = \mathbf{AA}^T$. Infatti

$$\mathbf{P}^T = (\mathbf{AA}^T)^T = (\mathbf{A}^T)^T \mathbf{A}^T = \mathbf{AA}^T = \mathbf{P}$$

Una dimostrazione più diretta della simmetria di \mathbf{P} è la seguente. Per ogni \mathbf{v} la proiezione $\mathbf{P}\mathbf{v} = \mathbf{v}_{\mathbf{H}}$ appartiene ad \mathbf{H} , mentre la differenza $\mathbf{v}_{\perp} = \mathbf{v} - \mathbf{v}_{\mathbf{H}}$ è ortogonale ad \mathbf{H} . Quindi per l'additività del prodotto scalare

$$\langle \mathbf{P}\mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}_{\mathbf{H}}, \mathbf{w}_{\mathbf{H}} + \mathbf{w}_{\perp} \rangle = \langle \mathbf{v}_{\mathbf{H}}, \mathbf{w}_{\mathbf{H}} \rangle = \langle \mathbf{v}_{\mathbf{H}} + \mathbf{v}_{\perp}, \mathbf{w}_{\mathbf{H}} \rangle = \langle \mathbf{v}, \mathbf{P}\mathbf{w} \rangle$$

cioè

$$\mathbf{v}^T \mathbf{P}^T \mathbf{w} = \mathbf{v}^T \mathbf{P}\mathbf{w}$$

per ogni \mathbf{v} e \mathbf{w} . Prendendo $\mathbf{v} = \mathbf{e}_i$ (l' i -esimo vettore della base canonica) e $\mathbf{w} = \mathbf{e}_j$, si trova $\mathbf{P}_{ij}^T = \mathbf{P}_{ij}$ per ogni (i, j) , quindi $\mathbf{P} = \mathbf{P}^T$.

Viceversa, supponiamo che \mathbf{P} sia simmetrica e idempotente. Poniamo $\mathbf{H} = \text{Col}(\mathbf{P})$, e dimostriamo che per ogni $\mathbf{v} \in \mathbb{R}^n$ il vettore $\mathbf{P}\mathbf{v}$ è la proiezione ortogonale di \mathbf{v} su \mathbf{H} . Siccome $\mathbf{P}\mathbf{v}$ è la combinazione lineare delle colonne di \mathbf{P} che ha per coefficienti le componenti di \mathbf{v} , il vettore $\mathbf{P}\mathbf{v}$ appartiene ad \mathbf{H} . Dobbiamo ancora far vedere che il vettore $\mathbf{v} - \mathbf{P}\mathbf{v}$ è perpendicolare ad \mathbf{H} , cioè appartiene a $\text{Col}(\mathbf{P})^\perp = \text{Ker}(\mathbf{P}^T)$. Per ipotesi $\mathbf{P}^T = \mathbf{P}$, quindi dobbiamo far vedere che $\mathbf{v} - \mathbf{P}\mathbf{v} \in \text{Ker}(\mathbf{P})$. Usando l'ipotesi di idempotenza $\mathbf{P}^2 = \mathbf{P}$ otteniamo:

$$\mathbf{P}(\mathbf{v} - \mathbf{P}\mathbf{v}) = \mathbf{P}\mathbf{v} - \mathbf{P}^2\mathbf{v} = \mathbf{P}\mathbf{v} - \mathbf{P}\mathbf{v} = \mathbf{0}$$

Quindi $\mathbf{v} - \mathbf{P}\mathbf{v} \in \text{Ker}(\mathbf{P})$ e questo conclude la dimostrazione.

OSSERVAZIONE Si osservino i due casi banali della matrice identità, che è la matrice della proiezione su $\mathbf{H} = \mathbb{R}^n$ e della matrice nulla, che è la matrice della proiezione su $\mathbf{H} = \{\mathbf{0}\}$. Si noti anche che, se \mathbf{P} è la matrice della proiezione ortogonale su \mathbf{H} , allora $\mathbf{I} - \mathbf{P}$ è la matrice della proiezione ortogonale su \mathbf{H}^\perp .

Nel testo abbiamo calcolato la matrice \mathbf{P} della proiezione di \mathbb{R}^4 sulla retta generata dal vettore $[1, 1, 1, 1]^T$:

$$\mathbf{P} = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Spiegare perché $\mathbf{I} - \mathbf{P}$ è la matrice della proiezione ortogonale di \mathbb{R}^4 sull'iperpiano \mathbf{H} di equazione $x_1 + x_2 + x_3 + x_4$. Calcolare la proiezione ortogonale $\mathbf{v}_{\mathbf{H}}$ di $\mathbf{v} = [x_1, x_2, x_3, x_4]^T$ su \mathbf{H} . Verificare che $\mathbf{v}_{\mathbf{H}} = \mathbf{v}$ se $\mathbf{v} \in \mathbf{H}$ e che $\mathbf{v}_{\mathbf{H}} = \mathbf{0}$ se $\mathbf{v} \in \mathbf{H}^\perp$.

Sia \mathbf{V} uno spazio euclideo e sia $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{V}$ la proiezione ortogonale di \mathbf{V} su un suo sottospazio \mathbf{H} . Mostrare che $\mathbf{H} = \text{Im}(\mathcal{L})$ e $\mathbf{H}^\perp = \text{Ker}(\mathcal{L})$. Inoltre $\mathcal{L}^2 = \mathcal{L}$ ed \mathcal{L} è simmetrica nel senso che

$$\langle \mathcal{L}(\mathbf{v}), \mathbf{w} \rangle = \langle \mathbf{v}, \mathcal{L}(\mathbf{w}) \rangle \quad \text{per ogni } \mathbf{v}, \mathbf{w} \in \mathbf{V}$$

Viceversa, mostrare che, se $\mathcal{L} : \mathbf{V} \rightarrow \mathbf{V}$ è simmetrica e idempotente, allora \mathcal{L} è la proiezione ortogonale di \mathbf{V} sull'immagine di \mathcal{L} .

Sia \mathbf{E} una matrice quadrata di ordine n e si supponga che \mathbf{E} sia idempotente: $\mathbf{E}^2 = \mathbf{E}$. Mostrare che ogni vettore di \mathbb{R}^n si scrive in uno e un sol modo come somma di un vettore di $\text{Col}(\mathbf{E})$ e di un vettore di $\text{Ker}(\mathbf{E})$, cioè $\mathbb{R}^n = \text{Col}(\mathbf{E}) \oplus \text{Ker}(\mathbf{E})$. Mostrare inoltre che $\text{Col}(\mathbf{E})$ è l'autospazio di \mathbf{E} relativo all'autovalore 1, mentre $\text{Ker}(\mathbf{E})$ è l'autospazio di \mathbf{E} relativo all'autovalore 0. Concludere che la matrice \mathbf{E} è diagonalizzabile. Calcolare il polinomio caratteristico di \mathbf{E} (dipende dal rango di \mathbf{E}).

 Trovare una matrice quadrata \mathbf{A} di ordine 2 tale che $\text{Col}(\mathbf{A}) \cap \text{Ker}(\mathbf{A}) \neq \{0\}$ e $\text{Col}(\mathbf{A}) + \text{Ker}(\mathbf{A}) \neq \mathbb{R}^2$.

Suggerimento: cercare tra le matrici non diagonalizzabili.

 Sia \mathbf{A} una matrice quadrata di ordine n . Mostrare che, se \mathbf{A} è diagonalizzabile, allora $\text{Col}(\mathbf{A}) \cap \text{Ker}(\mathbf{A}) = \{0\}$ e $\text{Col}(\mathbf{A}) + \text{Ker}(\mathbf{A}) = \mathbb{R}^n$.

Suggerimento: fare prima il caso in cui \mathbf{A} è diagonale.

8 IL CASO COMPLESSO

I risultati dei paragrafi precedenti non si estendono in modo automatico a \mathbb{C}^n perché il quadrato di un numero complesso non è in generale un numero positivo. La conseguenza è che non possiamo definire la norma di un vettore come la radice quadrata dei quadrati delle componenti. Per esempio, per il vettore $[z_1, z_2]^T = [1, i]^T \in \mathbb{C}^2$ abbiamo

$$z_1^2 + z_2^2 = 1^2 + i^2 = 0$$

Ricordiamo le nozioni di base sui numeri complessi. Un numero complesso z si scrive in uno e un solo modo nella forma $z = x + iy$, dove x e y sono numeri reali, che si dicono rispettivamente la *parte reale* e la *parte immaginaria* di z . Quindi \mathbb{C} è uno spazio vettoriale reale di dimensione 2; una sua base su \mathbb{R} è formata dal numero 1 e dall'unità immaginaria i ; la parte reale e la parte immaginaria di un numero complesso sono le sue coordinate rispetto alla base $\{1, i\}$. Il *conjugato* di un numero complesso $z = x + iy$ è il numero complesso $\bar{z} = x - iy$ che ha la stessa parte reale di z , e parte immaginaria opposta. La parte reale e la parte immaginaria di z si esprimono in termini di z e \bar{z} mediante le formule

$$x = \text{Re}(z) = \frac{1}{2}(z + \bar{z}) \quad \text{e} \quad y = \text{Im}(z) = \frac{1}{2i}(z - \bar{z})$$

Il *modulo* di z è il numero reale non negativo:

$$|z| = \sqrt{z\bar{z}} = \sqrt{x^2 + y^2}$$

Quindi $|z|$ coincide con la norma del vettore $[x, y]^T$ che rappresenta z nel piano di Argand-Gauss e $|z| = 0$ se e solo se $z = 0$.

Passiamo a considerare vettori $\mathbf{z} = [z_1, \dots, z_n]^T \in \mathbb{C}^n$. La parte reale \mathbf{x} e la parte immaginaria \mathbf{y} di \mathbf{z} sono i vettori di \mathbb{R}^n le cui componenti sono rispettivamente le parti reali x_k e le parti immaginarie y_k delle componenti $z_k = x_k + iy_k$ di \mathbf{z} . Il coniugato $\bar{\mathbf{z}}$ di \mathbf{z} è il vettore che ha per componenti i coniugati delle componenti di \mathbf{z} :

$$\bar{\mathbf{z}} = [\bar{z}_1, \dots, \bar{z}_n]^T$$

Per esempio, per il vettore $\mathbf{z} = [1 + 2i, 3 - 4i]^T$ abbiamo $\mathbf{Re}(\mathbf{z}) = [1, 3]^T$, $\mathbf{Im}(\mathbf{z}) = [2, -4]^T$, $\bar{\mathbf{z}} = [1 - 2i, 3 + 4i]^T$. Valgono le uguaglianze:

$$\mathbf{x} = \mathbf{Re}(\mathbf{z}) = \frac{1}{2}(\mathbf{z} + \bar{\mathbf{z}}) \quad \text{e} \quad \mathbf{y} = \mathbf{Im}(\mathbf{z}) = \frac{1}{2i}(\mathbf{z} - \bar{\mathbf{z}})$$

Date la parte reale e la parte immaginaria, possiamo ricostruire \mathbf{z} come

$$\mathbf{z} = \mathbf{x} + i\mathbf{y}$$

quindi la corrispondenza

$$\mathbf{z} \mapsto \tilde{\mathbf{z}} = [\mathbf{Re}(\mathbf{z}), \mathbf{Im}(\mathbf{z})]^T$$

è biiettiva tra \mathbb{C}^n e \mathbb{R}^{2n} . Tale corrispondenza è anche lineare su \mathbb{R} e, quindi, identifica \mathbb{C}^n e \mathbb{R}^{2n} come spazi vettoriali reali, generalizzando l'usuale identificazione di \mathbb{C} con il piano di Argand-Gauss. Definiamo la *norma* o lunghezza di \mathbf{z} mediante la formula

$$\|\mathbf{z}\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2} = \sqrt{z_1\bar{z}_1 + \cdots + z_n\bar{z}_n} = \sqrt{\mathbf{z}^T \bar{\mathbf{z}}}$$

È immediato verificare che la norma $\|\mathbf{z}\|$ coincide con la norma del corrispondente vettore $\tilde{\mathbf{z}}$ in \mathbb{R}^{2n} : se $\mathbf{x} = \mathbf{Re}(\mathbf{z})$ e $\mathbf{y} = \mathbf{Im}(\mathbf{z})$, allora:

$$\|\mathbf{z}\| = \sqrt{\sum_{k=1}^n (x_k^2 + y_k^2)} = \|\tilde{\mathbf{z}}\|_{\mathbb{R}^{2n}}$$

Un *versore* di \mathbb{C}^n è un vettore di norma 1. Si faccia attenzione che, a differenza di quanto succede nel caso reale, su una retta per l'origine (sottospazio di dimensione 1) in \mathbb{C}^n ci sono infiniti versori, parametrizzati dalla circonferenza $|t| = 1$ del piano di Argand-Gauss (il luogo $|t| = 1$ è una circonferenza perché, scrivendo $t = a + ib$, l'equazione $|t| = 1$ equivale ad $a^2 + b^2 = 1$): se \mathbf{e} è un versore, per ogni $t \in \mathbb{C}$ di modulo 1 il vettore $t\mathbf{e}$ è un versore.

Introduciamo ora l'analogo del prodotto scalare standard di \mathbb{R}^n .

DEFINIZIONE 8.1 (Prodotto hermitiano standard)

Il prodotto hermitiano (o scalare o interno) di due vettori $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$ è il numero complesso

$$\mathbf{z}^T \bar{\mathbf{w}} = z_1\bar{w}_1 + \cdots + z_n\bar{w}_n$$

Altre notazioni comuni per il prodotto hermitiano sono $\langle \mathbf{z}, \mathbf{w} \rangle$, $\mathbf{z} \cdot \mathbf{w}$ e (\mathbf{z}, \mathbf{w}) .

Elenchiamo ora le principali proprietà del prodotto hermitiano lasciandone la semplice dimostrazione al lettore.

PROPOSIZIONE 8.2 (Proprietà del prodotto hermitiano)

Il prodotto hermitiano standard di \mathbb{C}^n gode delle seguenti proprietà:

a) è lineare nella prima variabile:

$$\langle \mathbf{z}_1 + \mathbf{z}_2, \mathbf{w} \rangle = \langle \mathbf{z}_1, \mathbf{w} \rangle + \langle \mathbf{z}_2, \mathbf{w} \rangle \text{ e } \langle t\mathbf{z}, \mathbf{w} \rangle = t \langle \mathbf{z}, \mathbf{w} \rangle$$

per ogni $\mathbf{z}_1, \mathbf{z}_2, \mathbf{w} \in \mathbb{C}^n$ e ogni $t \in \mathbb{C}$;

b) è *sequilineare* nella seconda variabile; questo significa:

$$\langle \mathbf{z}, \mathbf{w}_1 + \mathbf{w}_2 \rangle = \langle \mathbf{z}, \mathbf{w}_1 \rangle + \langle \mathbf{z}, \mathbf{w}_2 \rangle \text{ e } \langle \mathbf{z}, t\mathbf{w} \rangle = \bar{t} \langle \mathbf{z}, \mathbf{w} \rangle$$

per ogni $\mathbf{z}, \mathbf{w}_1, \mathbf{w}_2 \in \mathbb{C}^n$ e ogni $t \in \mathbb{C}$;

c) per ogni $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$

$$\langle \mathbf{w}, \mathbf{z} \rangle = \overline{\langle \mathbf{z}, \mathbf{w} \rangle}.$$

Da questo segue che $\langle \mathbf{z}, \mathbf{z} \rangle = \overline{\langle \mathbf{z}, \mathbf{z} \rangle}$, quindi $\langle \mathbf{z}, \mathbf{z} \rangle$ è un numero reale e ha pertanto senso chiedersi se è maggiore o minore di zero;

d) è definito positivo:

$$\langle \mathbf{z}, \mathbf{z} \rangle \geq 0 \quad \text{per ogni } \mathbf{z} \in \mathbb{C}^n,$$

con uguaglianza se e solo se $\mathbf{z} = \mathbf{0}$.

OSSERVAZIONE L'analogo complesso di uno spazio euclideo si dice *spazio hermitiano*: si tratta di uno spazio vettoriale \mathbf{V} sul campo dei numeri complessi insieme a un prodotto hermitiano, cioè a una funzione $\langle \cdot, \cdot \rangle: \mathbf{V} \times \mathbf{V} \rightarrow \mathbb{C}$ che soddisfi le proprietà della proposizione precedente. Si osservi che le proprietà non sono indipendenti: la proprietà b) segue infatti da a) e c). La proprietà d) consente di definire la norma di un vettore in ogni spazio hermitiano:

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

DEFINIZIONE 8.3 (Ortogonalità)

Due vettori di \mathbb{C}^n si dicono *ortogonali* (o *perpendicolari*) se il loro prodotto hermitiano è nullo.

Se \mathbf{z} e \mathbf{w} hanno parte immaginaria nulla, cioè in sostanza sono due vettori di \mathbb{R}^n , il loro prodotto hermitiano in \mathbb{C}^n coincide con il loro prodotto scalare in \mathbb{R}^n , per cui i due vettori sono perpendicolari in \mathbb{R}^n se e solo se lo sono in \mathbb{C}^n .

Cerchiamo i vettori $\mathbf{z} = [z_1, z_2]$ di \mathbb{C}^2 ortogonali a $\mathbf{w} = [1, i]^T$: si tratta dei vettori per cui

$$\mathbf{z}^T \overline{\mathbf{w}} = z_1 - iz_2 = 0$$

cioè dei vettori della forma $\mathbf{z} = [it, t]^T$ al variare di t in \mathbb{C} , o ancora della retta generata dal vettore $\mathbf{u} = [1, -i]^T$ che corrisponde a $t = -i$ (prendendo $t = 1$ si ottiene un altro generatore $\mathbf{v} = [i, 1]$).

Sia \mathbf{V} lo spazio delle funzioni continue $f : [a, b] \rightarrow \mathbb{C}$. Si tratta di uno spazio vettoriale complesso. In \mathbf{V} si introduce il prodotto hermitiano, detto prodotto L^2 , definito dalla formula:

$$\langle f, g \rangle = \int_a^b f(x) \overline{g(x)} dx$$

Supponiamo per esempio $[a, b] = [0, 2\pi]$. Sia n un numero intero (non necessariamente positivo) e consideriamo la funzione esponenziale

$$e^{inx} = \cos(nx) + i \sin(nx)$$

Allora, dato che $\overline{e^{inx}} = e^{-inx}$,

$$\|e^{inx}\|^2 = \int_0^{2\pi} e^{inx} e^{-inx} dx = 2\pi$$

Inoltre le funzioni e^{inx} sono a due a due ortogonali: infatti, se $m \neq n$,

$$\langle e^{imx}, e^{inx} \rangle = \int_0^{2\pi} e^{imx} e^{-inx} dx = \left[\frac{e^{i(m-n)x}}{i(m-n)} \right]_0^{2\pi} = 0$$

Il teorema di Carnot e la formula di polarizzazione vanno modificati nel caso complesso, anche se è ancora vero che il prodotto hermitiano è determinato dalla funzione norma. Riassumiamo il risultato:

PROPOSIZIONE 8.4 (Teorema di Carnot e formula di polarizzazione)
In \mathbb{C}^n (e più in generale in uno spazio hermitiano) vale il Teorema di Carnot nella forma:

$$(8.1) \quad \|\mathbf{z} + \mathbf{w}\|^2 = \|\mathbf{z}\|^2 + \|\mathbf{w}\|^2 + 2\operatorname{Re}(\langle \mathbf{z}, \mathbf{w} \rangle)$$

per ogni $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$. Il prodotto hermitiano è determinato dalla funzione norma grazie alla formula di polarizzazione

$$(8.2) \quad \operatorname{Re}(\langle \mathbf{z}, \mathbf{w} \rangle) = \frac{1}{2} (\|\mathbf{z} + \mathbf{w}\|^2 - \|\mathbf{z}\|^2 - \|\mathbf{w}\|^2)$$

e all'uguaglianza

$$(8.3) \quad \operatorname{Im}(\langle \mathbf{z}, \mathbf{w} \rangle) = \operatorname{Re}(\langle \mathbf{z}, i\mathbf{w} \rangle)$$

DIMOSTRAZIONE. La differenza con il caso reale è che un prodotto hermitiano non è commutativo, ma soddisfa $\langle \mathbf{w}, \mathbf{z} \rangle = \overline{\langle \mathbf{z}, \mathbf{w} \rangle}$, per cui:

$$\begin{aligned}\|\mathbf{z} + \mathbf{w}\|^2 &= \langle \mathbf{z} + \mathbf{w}, \mathbf{z} + \mathbf{w} \rangle = \\ &= \langle \mathbf{z}, \mathbf{z} \rangle + \langle \mathbf{z}, \mathbf{w} \rangle + \langle \mathbf{w}, \mathbf{z} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle = \\ &= \|\mathbf{z}\|^2 + \langle \mathbf{z}, \mathbf{w} \rangle + \overline{\langle \mathbf{z}, \mathbf{w} \rangle} + \|\mathbf{w}\|^2 = \\ &= \|\mathbf{z}\|^2 + 2\operatorname{Re}(\langle \mathbf{z}, \mathbf{w} \rangle) + \|\mathbf{w}\|^2\end{aligned}$$

la formula di polarizzazione si ottiene ricavando $\operatorname{Re}(\langle \mathbf{z}, \mathbf{w} \rangle)$ dalla (8.1). Per mostrare la (8.3), osserviamo che la parte immaginaria di un numero complesso $t = a + ib$ coincide con la parte reale del numero complesso $-it = b - ia$. Prendendo $t = \langle \mathbf{z}, \mathbf{w} \rangle$ e ricordando che $-i \langle \mathbf{z}, \mathbf{w} \rangle = \langle \mathbf{z}, i\mathbf{w} \rangle$ otteniamo

$$\operatorname{Im}(\langle \mathbf{z}, \mathbf{w} \rangle) = \operatorname{Re}(-i \langle \mathbf{z}, \mathbf{w} \rangle) = \operatorname{Re}(\langle \mathbf{z}, i\mathbf{w} \rangle)$$

Il prodotto hermitiano è determinato dalla funzione norma: infatti, se conosciamo $\|\mathbf{z}\|$ per ogni \mathbf{z} , dalla formula di polarizzazione (8.2) ricaviamo la parte reale $\operatorname{Re}(\langle \mathbf{z}, \mathbf{w} \rangle)$ del prodotto hermitiano, e la parte reale determina anche la parte immaginaria mediante la (8.3).

OSSERVAZIONE Torniamo alla mappa $\mathbb{C}^n \rightarrow \mathbb{R}^{2n}$ che a un vettore $\mathbf{z} = \mathbf{x} + iy$ di \mathbb{C}^n associa il vettore $\tilde{\mathbf{z}} = [\mathbf{x}, \mathbf{y}]^T \in \mathbb{R}^{2n}$, dove \mathbf{x} e \mathbf{y} sono rispettivamente la parte reale e la parte immaginaria di \mathbf{z} . Abbiamo già osservato che la norma di \mathbf{z} in \mathbb{C}^n coincide con la norma di $\tilde{\mathbf{z}}$ in \mathbb{R}^{2n} . Dalle formule di polarizzazione e dalla (8.3) segue che il prodotto hermitiano in \mathbb{C}^n è legato al prodotto scalare in \mathbb{R}^{2n} dalla formula:

$$(8.4) \quad \langle \mathbf{z}, \mathbf{w} \rangle_{\mathbb{C}^n} = \langle \tilde{\mathbf{z}}, \tilde{\mathbf{w}} \rangle_{\mathbb{R}^{2n}} + i \langle \tilde{\mathbf{z}}, i\tilde{\mathbf{w}} \rangle_{\mathbb{R}^{2n}}$$

La conseguenza è che la nozione di ortogonalità in \mathbb{C}^n è più forte di quella in \mathbb{R}^{2n} : il vettore \mathbf{z} è ortogonale a \mathbf{w} in \mathbb{C}^n se e solo se $\tilde{\mathbf{z}}$ è ortogonale a $\tilde{\mathbf{w}}$ e a $i\tilde{\mathbf{w}}$ in \mathbb{R}^{2n} .

Se $\mathbf{z} = \mathbf{x} + iy$ e $\mathbf{w} = \mathbf{a} + ib$ sono le decomposizioni in parte reale e immaginaria di \mathbf{z} e $\mathbf{w} \in \mathbb{C}^n$, la (8.4) si riscrive nella forma

$$\mathbf{z}^T \overline{\mathbf{w}} = \mathbf{x}^T \mathbf{a} + \mathbf{y}^T \mathbf{b} + i (\mathbf{y}^T \mathbf{a} - \mathbf{x}^T \mathbf{b})$$

Si noti che i due vettori $\tilde{\mathbf{w}} = [\mathbf{a}, \mathbf{b}]^T$ e $i\tilde{\mathbf{w}} = [-\mathbf{b}, \mathbf{a}]^T$ sono ortogonali in \mathbb{R}^{2n} : due vettori \mathbf{z} e \mathbf{w} ortogonali in \mathbb{C}^n definiscono quattro vettori $\tilde{\mathbf{z}}$, $i\tilde{\mathbf{z}}$, $\tilde{\mathbf{w}}$, $i\tilde{\mathbf{w}}$ a due a due ortogonali in \mathbb{R}^{2n} . Per esempio, i due vettori

$$\mathbf{z} = \begin{bmatrix} 1+2i \\ 3+4i \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} + i \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} 3-4i \\ -1+2i \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \end{bmatrix} + i \begin{bmatrix} -4 \\ 2 \end{bmatrix}$$

sono ortogonali in \mathbb{C}^2 perché

$$\mathbf{z}^T \overline{\mathbf{w}} = (1+2i)(3+4i) + (3+4i)(-1-2i) = 0$$

A \mathbf{z} e \mathbf{w} corrispondono i vettori di \mathbb{R}^4 :

$$\tilde{\mathbf{z}} = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 4 \end{bmatrix}, \quad i\tilde{\mathbf{z}} = \begin{bmatrix} -2 \\ -4 \\ 1 \\ 3 \end{bmatrix}, \quad \tilde{\mathbf{w}} = \begin{bmatrix} 3 \\ -1 \\ -4 \\ 2 \end{bmatrix}, \quad i\tilde{\mathbf{w}} = \begin{bmatrix} 4 \\ -2 \\ 3 \\ -1 \end{bmatrix}$$

che sono a due a due ortogonali.

Basi ortonormali e matrici unitarie

Sostanzialmente tutto quanto abbiamo visto nel caso reale su basi ortonormali e proiezioni ortogonali si estende al caso complesso. In uno spazio hermitiano ha senso parlare di basi ortogonali (costituite da vettori a due a due ortogonali) e ortonormali (costituite da versori a due a due ortogonali). L'algoritmo di Gram-Schmidt rimane valido per vettori di uno spazio hermitiano e produce una base ortogonale di un qualunque spazio hermitiano di dimensione finita. Anche i teoremi sulle proiezioni continuano a valere: se \mathbf{H} è un sottospazio finito dimensionale di uno spazio hermitiano \mathbf{V} , ogni vettore \mathbf{v} di \mathbf{V} si scrive in uno e un sol modo come somma $\mathbf{v} = \mathbf{v}_\mathbf{H} + \mathbf{v}_\perp$ di un vettore $\mathbf{v}_\mathbf{H} \in \mathbf{H}$ e di un vettore $\mathbf{v}_\perp \in \mathbf{H}^\perp$; inoltre per la proiezione vale la formula della proposizione 5.3:

$$(8.5) \quad \mathbf{v}_\mathbf{H} = \hat{x}_1 \mathbf{b}_1 + \hat{x}_2 \mathbf{b}_2 + \cdots + \hat{x}_d \mathbf{b}_d$$

dove

$$(8.6) \quad \hat{x}_k = \frac{\langle \mathbf{v}, \mathbf{b}_k \rangle}{\|\mathbf{b}_k\|^2} \quad k = 1, 2, \dots, d$$

è il coefficiente di Fourier di \mathbf{v} rispetto a \mathbf{b}_k . L'unica differenza rispetto al caso reale è che nelle formule $\langle \cdot, \cdot \rangle$ denota ora il prodotto hermitiano.

Supponiamo di voler determinare una base ortonormale $\{\mathbf{q}_1, \mathbf{q}_2\}$ di \mathbb{C}^2 in cui \mathbf{q}_1 sia un multiplo scalare di $\mathbf{v}_1 = [1, i]^T$. Possiamo estendere \mathbf{v}_1 a una base di \mathbb{C}^2 scegliendo $\mathbf{v}_2 = [1, 0]^T$. L'algoritmo di Gram-Schmidt dà $\mathbf{b}_1 = [1, i]^T$ e

$$\begin{aligned} \mathbf{b}_2 &= \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{b}_1 \rangle}{\|\mathbf{b}_1\|^2} \mathbf{b}_1 = [1, 0]^T - \frac{1}{2}[1, i]^T = \frac{1}{2}[1, -i]^T \\ \mathbf{q}_1 &= \frac{\mathbf{b}_1}{\|\mathbf{b}_1\|} = \frac{\sqrt{2}}{2}[1, i]^T \\ \mathbf{q}_2 &= \frac{\mathbf{b}_2}{\|\mathbf{b}_2\|} = \frac{\sqrt{2}}{2}[1, -i]^T \end{aligned}$$



Consideriamo nuovamente lo spazio \mathbf{V} delle funzioni continue $f : [0, 2\pi] \rightarrow \mathbb{C}$ con il prodotto L^2 :

$$\langle f, g \rangle = \int_0^{2\pi} f(x) \overline{g(x)} dx$$

Per ogni intero positivo n sia \mathbf{H}_n il sottospazio vettoriale di \mathbf{V} generato dalle funzioni e^{ikx} al variare di k tra $-n$ e n : una funzione di \mathbf{H}_n ha quindi la forma

$$P(x) = \sum_{k=-n}^n c_k e^{ikx} \quad (c_k \in \mathbb{C})$$

Queste funzioni si dicono polinomi trigonometrici complessi di grado $\leq n$. Nelle applicazioni è fondamentale approssimare, seguendo l'idea di Fourier, una funzione (periodica) qualsiasi con

un polinomio trigonometrico. La miglior approssimazione di una funzione $f \in \mathbf{V}$ mediante un polinomio $P(x) \in \mathbf{H}_n$ rispetto alla norma L^2 è la proiezione ortogonale di f su \mathbf{H}_n . Abbiamo già osservato che i polinomi e^{ikx} hanno tutti norma $\sqrt{2\pi}$ e sono a due a due ortogonali, pertanto per la (8.5) tale proiezione è il *polinomio di Fourier*

$$S_n(f)(x) = \sum_{k=-n}^n c_k e^{ikx}$$

dove i numeri c_k sono i coefficienti di Fourier:

$$c_k = \frac{\langle f, e^{ikx} \rangle}{\|e^{ikx}\|^2} = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-ikx} dx$$

L'analogo di una matrice ortogonale è una matrice $n \times n$ complessa le cui colonne formano una base ortonormale di \mathbb{C}^n . Per scrivere questa condizione in termini di matrici introduciamo l'analogo complesso della matrice trasposta.

DEFINIZIONE 8.5 (Matrice trasposta coniugata (o aggiunta))

Sia \mathbf{A} una matrice complessa. La matrice *trasposta coniugata* o *aggiunta* di \mathbf{A} è la matrice il cui elemento di posto (i, j) è il coniugato dell'elemento di posto (j, i) di \mathbf{A} . La si denota con il simbolo \mathbf{A}^H . Quindi $\mathbf{A}^H = \overline{\mathbf{A}}^T$.

Se $\mathbf{A} = \begin{bmatrix} 1 & -1 \\ i & -i \end{bmatrix}$, allora

$$\overline{\mathbf{A}} = \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \quad \text{e} \quad \mathbf{A}^H = \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix}$$

OSSERVAZIONE Se gli elementi di \mathbf{A} sono tutti reali, allora $\mathbf{A}^H = \mathbf{A}^T$. Il prodotto hermitiano standard di due vettori colonna $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$ è $\mathbf{w}^H \mathbf{z}$. Infatti:

$$\mathbf{w}^H \mathbf{z} = [\overline{w_1} \quad \cdots \quad \overline{w_n}] \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = \sum_{k=1}^n z_k \overline{w_k} = \mathbf{z}^T \overline{\mathbf{w}}$$

OSSERVAZIONE Il lettore può facilmente verificare le due proprietà fondamentali della matrice trasposta coniugata, che sono analoghe a quelle della matrice trasposta:

$$(8.7) \quad (\mathbf{A}^H)^H = \mathbf{A} \quad \text{e} \quad (\mathbf{AB})^H = \mathbf{B}^H \mathbf{A}^H$$

Il motivo per cui \mathbf{A}^H si dice anche matrice aggiunta è che vale l'uguaglianza:

$$(8.8) \quad \langle \mathbf{Az}, \mathbf{w} \rangle = \langle \mathbf{z}, \mathbf{A}^H \mathbf{w} \rangle$$

Infatti

$$\langle \mathbf{Az}, \mathbf{w} \rangle = \mathbf{w}^H \mathbf{Az} = (\mathbf{A}^H \mathbf{w})^H \mathbf{z} = \langle \mathbf{z}, \mathbf{A}^H \mathbf{w} \rangle$$

DEFINIZIONE 8.6 (Matrice unitaria)

Una matrice quadrata \mathbf{U} di ordine n a coefficienti complessi si dice *unitaria* se $\mathbf{U}^H \mathbf{U} = \mathbf{I}$.

Come nel caso reale si dimostra che una matrice è unitaria se e solo se le sue colonne (o le sue righe) formano una base ortonormale di \mathbb{C}^n . Si osservi che l'inversa di una matrice unitaria \mathbf{U} è la sua trasposta coniugata. da questo segue che $\mathbf{U} \mathbf{U}^H = \mathbf{I}$.

Si ricordi che, dato un numero reale θ , il numero complesso

$$e^{i\theta} = \cos(\theta) + i \sin(\theta)$$

ha modulo 1 (nel piano di Argand-Gauss corrisponde al raggio della circonferenza unitaria che forma un angolo orientato θ con il semiasse positivo delle x). Dati due numeri reali α e β , i vettori $[e^{i\alpha}, 0]$ e $[0, e^{i\beta}]$ formano una base ortonormale di \mathbb{C}^2 . La corrispondente matrice unitaria è

$$\mathbf{U} = \begin{bmatrix} e^{i\alpha} & 0 \\ 0 & e^{i\beta} \end{bmatrix}$$

Più in generale, una matrice $n \times n$ diagonale è unitaria se e solo se i suoi elementi sulla diagonale principale sono numeri complessi di modulo 1, cioè della forma $e^{i\theta}$.

A partire dalla base ortonormale di \mathbb{C}^2 formata dai vettori $\mathbf{q}_1 = \frac{\sqrt{2}}{2}[1, i]^T$ e $\mathbf{q}_2 = \frac{\sqrt{2}}{2}[1, -i]^T$ costruiamo la matrice unitaria:

$$\mathbf{U} = \frac{\sqrt{2}}{2} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}$$

Possiamo controllare che \mathbf{U}^H sia l'inversa di \mathbf{U} :

$$\mathbf{U}^H \mathbf{U} = \left(\frac{\sqrt{2}}{2} \right)^2 \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 - i^2 & 1 + (-i)^2 \\ 1 + i^2 & 1 - i^2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \mathbf{I}$$

Se $\{\mathbf{q}_1, \dots, \mathbf{q}_n\}$ è una base ortonormale di \mathbb{C}^n , allora $\{\widetilde{\mathbf{q}}_1, i\widetilde{\mathbf{q}}_1, \dots, \widetilde{\mathbf{q}}_n, i\widetilde{\mathbf{q}}_n\}$ è una base ortonormale di \mathbb{R}^{2n} per l'osservazione a pagina 420. Quindi a ogni matrice unitaria \mathbf{U} di ordine n corrisponde una matrice ortogonale di ordine $2n$. Per esempio a

$$\mathbf{U} = [\mathbf{q}_1 \quad \mathbf{q}_2] = \frac{\sqrt{2}}{2} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}$$

corrisponde la matrice ortogonale

$$\mathbf{Q} = \begin{bmatrix} \widetilde{\mathbf{q}}_1 & i\widetilde{\mathbf{q}}_1 & \widetilde{\mathbf{q}}_2 & i\widetilde{\mathbf{q}}_2 \end{bmatrix} = \frac{\sqrt{2}}{2} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \end{bmatrix}$$

Le matrici unitarie rappresentano isometrie di \mathbb{C}^n :

PROPOSIZIONE 8.7 (Matrici unitarie rappresentano isometrie)

Per una matrice quadrata complessa \mathbf{U} di ordine n le seguenti condizioni sono equivalenti:

a) \mathbf{U} è unitaria: $\mathbf{U}^H \mathbf{U} = \mathbf{I}$;

b) \mathbf{U} preserva la norma dei vettori:

$$\|\mathbf{U}\mathbf{z}\| = \|\mathbf{z}\| \quad \text{per ogni } \mathbf{z} \in \mathbb{C}^n$$

(in altri termini, \mathbf{U} rappresenta un'isometria di \mathbb{C}^n);

c) \mathbf{U} preserva il prodotto hermitiano:

$$\langle \mathbf{U}\mathbf{z}, \mathbf{U}\mathbf{w} \rangle = \langle \mathbf{z}, \mathbf{w} \rangle \quad \text{per ogni } \mathbf{z}, \mathbf{w} \in \mathbb{C}^n$$

DIMOSTRAZIONE. La dimostrazione è sostanzialmente identica a quella della proposizione 4.9.

COROLLARIO 8.8 (Autovalori di una matrice unitaria)

Sia \mathbf{U} una matrice unitaria, e sia λ un autovalore di \mathbf{U} . Allora $|\lambda| = 1$. In particolare, gli autovalori di una matrice ortogonale reale hanno modulo 1.

DIMOSTRAZIONE. Siccome \mathbf{U} è unitaria, $\|\mathbf{z}\| = \|\mathbf{U}\mathbf{z}\|$ per ogni $\mathbf{z} \in \mathbb{C}^n$. Se \mathbf{z} è un autovettore relativo a λ , allora

$$\|\mathbf{z}\| = \|\mathbf{U}\mathbf{z}\| = \|\lambda\mathbf{z}\| = |\lambda| \|\mathbf{z}\|$$

Dividendo per $\|\mathbf{z}\|$ si trova $|\lambda| = 1$. Una matrice ortogonale è unitaria, per cui quanto appena dimostrato vale per le matrici ortogonali.

Consideriamo la matrice ortogonale che rappresenta la rotazione di un angolo θ in senso antiorario: $\mathbf{Q} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$. Il polinomio caratteristico di \mathbf{Q} è

$$\lambda^2 - \text{tr}(\mathbf{Q})\lambda + \det(\mathbf{Q}) = \lambda^2 - 2\cos(\theta)\lambda + 1 = \lambda^2 - 2\text{Re}(e^{i\theta})\lambda + e^{i\theta}e^{-i\theta} = (\lambda - e^{-i\theta})(\lambda - e^{i\theta})$$

Gli autovalori di \mathbf{Q} sono pertanto $\lambda_1 = e^{-i\theta}$ e $\lambda_2 = e^{i\theta}$ e hanno modulo 1 in accordo con il corollario precedente. I vettori $\mathbf{v}_1 = [1, i]^T$ e $\mathbf{v}_2 = [1, -i]^T$ sono autovettori di \mathbf{Q} relativi agli autovalori λ_1 e λ_2 :

$$\begin{aligned} \mathbf{Q} [\mathbf{v}_1 \quad \mathbf{v}_2] &= \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} = \begin{bmatrix} \cos(\theta) - i\sin(\theta) & \cos(\theta) + i\sin(\theta) \\ \sin(\theta) + i\cos(\theta) & \sin(\theta) - i\cos(\theta) \end{bmatrix} = \\ &= \begin{bmatrix} e^{-i\theta} & e^{i\theta} \\ ie^{-i\theta} & -ie^{i\theta} \end{bmatrix} = [e^{-i\theta} \mathbf{v}_1 \quad e^{i\theta} \mathbf{v}_2] \end{aligned}$$

Quindi le colonne di $\mathbf{U} = \frac{\sqrt{2}}{2} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}$ formano una base ortonormale di \mathbb{C}^2 costituita da autovettori di \mathbf{Q} . Quindi

$$\mathbf{U}^H \mathbf{Q} \mathbf{U} = \mathbf{U}^{-1} \mathbf{Q} \mathbf{U} = \text{diag}(e^{-i\theta}, e^{i\theta})$$

Abbiamo così verificato che \mathbf{Q} è *unitariamente diagonalizzabile*, cioè è diagonalizzabile mediante una matrice unitaria.

Calcolare il prodotto scalare in \mathbb{C}^2 dei due vettori $\mathbf{z} = [2+i, 3i]^T$ e $\mathbf{w} = [2, 4-i]^T$. Calcolare la norma dei due vettori. Scrivere i corrispondenti vettori $\tilde{\mathbf{z}}$ e $\tilde{\mathbf{w}}$ in \mathbb{R}^4 e calcolarne il prodotto scalare reale. Verificare la formula (8.4) in questo caso.

Data la matrice $\mathbf{A} = \begin{bmatrix} i & i & 3 \\ 2 & -i & i \end{bmatrix}$, calcolare \mathbf{A}^H , $\mathbf{A}^H \mathbf{A}$ e $\mathbf{A} \mathbf{A}^H$.

Trovare una base ortonormale di \mathbb{C}^2 il cui primo vettore sia un multiplo complesso di $[i, 2]^T$. Scrivere la corrispondente matrice unitaria di ordine 2, e verificare che $\mathbf{A}^H \mathbf{A} = \mathbf{I}$.

Mostrare che il prodotto di due matrici unitarie è una matrice unitaria. Mostrare che l'inversa di una matrice unitaria è unitaria.

Sia \mathbf{A} una matrice quadrata complessa. Che relazione c'è tra il determinante di \mathbf{A} e quello di \mathbf{A}^H ? E tra i polinomi caratteristici delle due matrici? E tra gli autovalori? Mostrare che $r(\mathbf{A}) = r(\mathbf{A}^H)$ calcolando il rango con il teorema di Kronecker.

Mostrare che, se \mathbf{A} è una matrice complessa $m \times n$, il nucleo di \mathbf{A} è il complemento ortogonale in \mathbb{C}^n dello spazio colonna di \mathbf{A}^H . Dedurre che $r(\mathbf{A}) = r(\mathbf{A}^H)$.

Una matrice *hermitiana* è una matrice complessa \mathbf{B} tale che $\mathbf{B}^H = \mathbf{B}$ (l'analogo complesso di una matrice simmetrica). Mostrare che, se \mathbf{A} è una matrice complessa $m \times n$, allora $\mathbf{A}^H \mathbf{A}$ è una matrice hermitiana $n \times n$, e $\mathbf{A} \mathbf{A}^H$ è una matrice hermitiana $m \times m$. Mostrare che

$$r(\mathbf{A}) = r(\mathbf{A}^H) = r(\mathbf{A}^H \mathbf{A}) = r(\mathbf{A} \mathbf{A}^H)$$

9 COMPLEMENTI

9.1 Il teorema di Eulero sulle rotazioni dello spazio

In geometria il teorema di Eulero afferma che nello spazio tridimensionale ogni movimento rigido che lasci fissa l'origine è una rotazione attorno a un asse di punti fissi. Un movimento rigido è un'isometria della spazio che preserva le terne destrose. Non è difficile mostrare che un'isometria che lascia fissa l'origine è lineare, e quindi corrisponde a una matrice ortogonale. Quindi il teorema di Eulero è equivalente a dire che ogni matrice ortogonale di ordine tre con determinante 1 rappresenta una rotazione dello spazio attorno a un asse. Una conseguenza importante è che la composta di due rotazioni, anche attorno ad assi distinti, è ancora una rotazione. La formulazione originale di Eulero era più o meno questa: se si muove una sfera lasciando fisso il suo centro, è sempre possibile trovare un diametro della sfera che è lasciato invariato

dal movimento. Tale diametro è un autovettore relativo all'autovalore $\lambda = 1$ della rotazione.

TEOREMA 9.1 (Teorema di Eulero)

Sia \mathbf{Q} una matrice ortogonale reale di ordine 3 con determinante uguale a 1.

Allora l'applicazione che a un vettore $\mathbf{x} \in \mathbb{R}^3$ associa $\mathbf{Q}\mathbf{x}$ è una rotazione attorno a un asse. Più precisamente \mathbf{Q} ha $\lambda = 1$ come autovalore e, se $\mathbf{Q} \neq \mathbf{I}$, l'asse di rotazione è l'autospazio relativo all'autovalore 1; l'angolo α di rotazione si può ricavare dalla formula

$$\text{tr}(\mathbf{Q}) = 1 + 2 \cos(\alpha)$$

DIMOSTRAZIONE. Il polinomio caratteristico di \mathbf{Q} è un polinomio di terzo grado a coefficienti reali, per cui ha almeno una radice reale μ . Siccome \mathbf{Q} è ortogonale, il modulo di μ è 1, quindi μ è $+1$ o -1 .

Ci sono due casi possibili:

- a) il polinomio caratteristico ha 3 radici reali: siccome le radici sono ± 1 e il loro prodotto è il determinante di \mathbf{Q} , che per ipotesi vale 1, almeno una delle radici è $+1$.
- b) il polinomio caratteristico ha 1 radice reale $\mu = \pm 1$ e 2 radici complesse coniugate λ e $\bar{\lambda}$.

Allora

$$1 = \det(\mathbf{Q}) = \mu \lambda \bar{\lambda}$$

Siccome $\lambda \bar{\lambda}$ è positivo, concludiamo che la radice reale è $\mu = +1$.

Quindi in entrambi i casi $+1$ è un autovalore di \mathbf{Q} e perciò esiste un versore $\mathbf{v} \in \mathbb{R}^3$ tale che $\mathbf{Q}\mathbf{v} = \mathbf{v}$. Possiamo scegliere altri due versori \mathbf{w} e \mathbf{z} in modo che $\mathbf{w}, \mathbf{z}, \mathbf{v}$ formino una base ortonormale di \mathbb{R}^3 . Sia \mathbf{P} la matrice che ha per colonne \mathbf{w}, \mathbf{z} e \mathbf{v} . Per costruzione \mathbf{P} è una matrice ortogonale.

Sia \mathbf{M} la matrice che rappresenta l'applicazione $\mathbf{x} \mapsto \mathbf{Q}\mathbf{x}$ rispetto alla base $\{\mathbf{w}, \mathbf{z}, \mathbf{v}\}$. Siccome $\mathbf{Q}\mathbf{v} = \mathbf{v}$, la matrice \mathbf{M} è della forma

$$\mathbf{M} = \begin{bmatrix} a & b & 0 \\ c & d & 0 \\ e & f & 1 \end{bmatrix}$$

D'altra parte,

$$\mathbf{M} = \mathbf{P}^{-1} \mathbf{Q} \mathbf{P} = \mathbf{P}^T \mathbf{Q} \mathbf{P}$$

è ortogonale in quanto prodotto di matrici ortogonali. In particolare, le colonne di \mathbf{M} sono ortogonali tra loro, per cui $e = f = 0$. Ma allora anche la sottomatrice $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ è ortogonale con determinante uguale a 1.

Per la classificazione delle matrici ortogonali di ordine 2 possiamo concludere che esiste un angolo α tale che:

$$\mathbf{M} = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

L'applicazione $\mathbf{X} \mapsto \mathbf{M}\mathbf{X}$ è quindi una rotazione di un angolo α attorno all'asse zeta delle coordinate \mathbf{X} . Tale applicazione coincide con l'applicazione $\mathbf{x} \mapsto \mathbf{Q}\mathbf{x}$ via il cambiamento di coordinate $\mathbf{x} = \mathbf{P}\mathbf{X}$. Siccome l'asse zeta delle coordinate \mathbf{X} è la retta generata da \mathbf{v} , concludiamo che \mathbf{Q} rappresenta una rotazione di un angolo α attorno alla retta generata da \mathbf{v} .

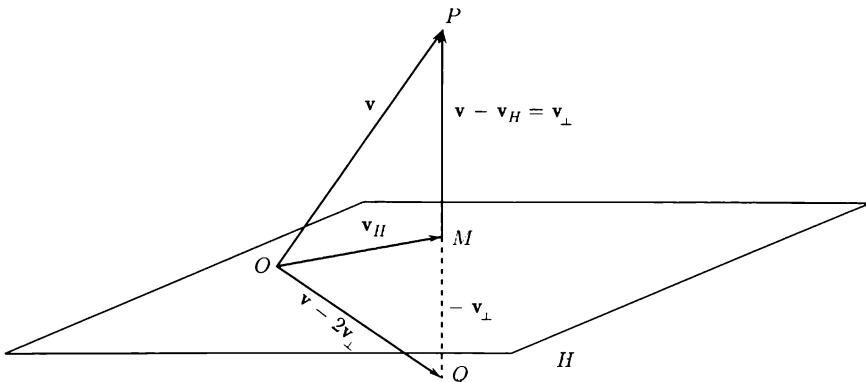


Figura 9.1. La riflessione ortogonale Q di P rispetto al sottospazio H .

Per trovare l'angolo di rotazione, notiamo che la traccia della matrice \mathbf{M} è

$$\text{tr}(\mathbf{M}) = 1 + 2 \cos(\alpha)$$

Ma la traccia, che è un coefficiente del polinomio caratteristico, è invariante per similitudine, per cui $\text{tr}(\mathbf{Q}) = \text{tr}(\mathbf{M})$, e quindi

$$\text{tr}(\mathbf{Q}) = 1 + 2 \cos(\alpha)$$

In conclusione, per trovare l'asse e il coseno dell'angolo di rotazione, non occorre determinare \mathbf{M} , ma è sufficiente calcolare la traccia di \mathbf{Q} e l'autovettore di \mathbf{Q} relativo all'autovalore $\lambda = 1$.

9.2 Riflessioni ortogonali

Nel piano cartesiano, la riflessione ortogonale rispetto a una retta \mathbf{r} , detta asse della riflessione, è l'applicazione \mathfrak{L} così definita: a un punto P la riflessione \mathfrak{L} associa il punto Q tale che la retta \mathbf{s} per P e Q sia ortogonale a \mathbf{r} e l'intersezione M di \mathbf{r} ed \mathbf{s} sia il punto medio del segmento PQ . In altre parole, la retta PQ è ortogonale a \mathbf{r} , i due punti P e Q hanno la stessa distanza da \mathbf{r} , ma giacciono da parti opposte rispetto a \mathbf{r} ; si dice che \mathbf{r} è l'asse del segmento PQ .

Per tradurre la costruzione in termini di vettori, fissiamo l'origine delle coordinate O sulla retta \mathbf{r} . Se $\mathbf{v} = \overrightarrow{OP}$ è il vettore posizione di P , la proiezione ortogonale di \mathbf{v} su \mathbf{r} è il vettore posizione \mathbf{v}_r di M , che è il piede della perpendicolare a \mathbf{r} che passa per P :

$$\overrightarrow{OP} = \mathbf{v} = \mathbf{v}_r + \mathbf{v}_{\perp} = \overrightarrow{OM} + \overrightarrow{MP}$$

Il vettore posizione di $Q = \mathfrak{L}(P)$ è

$$(9.1) \quad \overrightarrow{OQ} = \overrightarrow{OM} + \overrightarrow{MQ} = \overrightarrow{OM} - \overrightarrow{MP} = \mathbf{v}_r - \mathbf{v}_{\perp} = \mathbf{v} - 2\mathbf{v}_{\perp}$$

Più in generale, in \mathbb{R}^n (o in uno spazio euclideo finito dimensionale) possiamo considerare la riflessione ortogonale rispetto a un qualsiasi sottospazio \mathbf{H} : definiamo tale riflessione come l'unica funzione $\mathfrak{R} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ tale che, per ogni $\mathbf{v} \in \mathbb{R}^n$,

- a) la proiezione ortogonale di $\mathfrak{R}(\mathbf{v})$ su \mathbf{H} è uguale a quella di \mathbf{v} ;
- b) la proiezione ortogonale di $\mathfrak{R}(\mathbf{v})$ su \mathbf{H}^\perp è l'opposto di quella di \mathbf{v} .

Per vedere che esiste un'unica funzione con queste proprietà, scriviamo $\mathbf{v} = \mathbf{v}_\mathbf{H} + \mathbf{v}_\perp$ come somma delle sue proiezioni ortogonali su \mathbf{H} e \mathbf{H}^\perp . Da a) e da b) deduciamo

$$(9.2) \quad \mathfrak{R}(\mathbf{v}) = \mathbf{v}_\mathbf{H} - \mathbf{v}_\perp = \mathbf{v} - 2\mathbf{v}_\perp$$

Questo mostra l'esistenza e l'unicità di \mathfrak{R} e anche che \mathfrak{R} è un'applicazione lineare.

Un utile esercizio è verificare le seguenti proprietà di \mathfrak{R} :

- i) \mathfrak{R} è un'isometria;
- ii) se \mathbf{H} ha dimensione d , la riflessione \mathfrak{R} ha l'autovalore $\lambda = 1$ con molteplicità algebrica e geometrica uguali a d e l'autovalore $\lambda = -1$ con molteplicità algebrica e geometrica uguali a $n - d$;
- iii) esiste una base ortonormale di \mathbb{R}^n rispetto alla quale la matrice che rappresenta \mathfrak{R} è la matrice diagonale che ha sulla diagonale principale d elementi uguali a 1 ed $n - d$ elementi uguali a -1 ;
- iv) la matrice che rappresenta \mathfrak{R} rispetto alla base canonica è

$$\mathbf{R} = \mathbf{I} - 2\mathbf{P}$$

dove \mathbf{P} è la matrice della proiezione ortogonale su \mathbf{H}^\perp . La matrice \mathbf{R} è ortogonale e simmetrica.

- v) esiste una matrice ortogonale \mathbf{Q} tale che

$$\mathbf{Q}^T \mathbf{R} \mathbf{Q} = \mathbf{D}$$

dove \mathbf{D} è la matrice diagonale che ha sulla diagonale principale d elementi uguali a 1 ed $n - d$ elementi uguali a -1 . In particolare, $\det(\mathbf{R}) = (-1)^{n-d}$.

DIMOSTRAZIONE. Per il teorema di Pitagora

$$\|\mathfrak{R}(\mathbf{v})\|^2 = \|\mathbf{v}_\mathbf{H}\|^2 + \|\mathbf{v}_\perp\|^2 = \|\mathbf{v}_\mathbf{H}\|^2 + \|\mathbf{v}_\perp\|^2 = \|\mathbf{v}\|^2$$

quindi \mathfrak{R} è un'isometria.

Se $\mathbf{v} \in \mathbf{H}$, allora $\mathbf{v}_\perp = \mathbf{0}$ e quindi $\mathfrak{R}(\mathbf{v}) = \mathbf{v}$. Questo significa che \mathbf{H} è contenuto nell'autospazio relativo all'autovalore 1. Se $\mathbf{v} \in \mathbf{H}^\perp$, allora $\mathbf{v}_\mathbf{H} = \mathbf{0}$ e quindi $\mathfrak{R}(\mathbf{v}) = -\mathbf{v}$. Questo significa che \mathbf{H}^\perp è contenuto nell'autospazio relativo all'autovalore -1 . Esiste una base ortonormale di \mathbb{R}^n formata da vettori di \mathbf{H} e di \mathbf{H}^\perp . Questa è una base di \mathbb{R}^n formata da autovettori di \mathfrak{R} , e tra questi d sono relativi all'autovalore 1 ed $n - d$ all'autovalore -1 . Questo mostra ii) e iii). Per il punto iv) osserviamo che

$$\mathfrak{R}(\mathbf{v}) = \mathbf{v} - 2\mathbf{v}_\perp = \mathbf{I}\mathbf{v} - 2\mathbf{P}\mathbf{v} = (\mathbf{I} - 2\mathbf{P})\mathbf{v}$$

Quest'uguaglianza ci dice che $\mathbf{I} - 2\mathbf{P}$ è la matrice di \mathfrak{R} : Si tratta di una matrice ortogonale perché \mathfrak{R} è un'isometria ed è simmetrica perché \mathbf{I} e \mathbf{P} sono simmetriche.

Supponiamo che \mathbf{H} sia un iperpiano di \mathbb{R}^n , cioè abbia dimensione $n - 1$, e che \mathbf{n} sia un versore ortogonale ad \mathbf{H} . Allora $\mathbf{n}\mathbf{n}^T$ è la matrice della proiezione su \mathbf{H}^\perp , per cui la matrice della riflessione ortogonale rispetto ad \mathbf{H} è

$$\mathbf{R} = \mathbf{I} - 2\mathbf{n}\mathbf{n}^T$$

Per esempio, se \mathbf{H} è l'iperpiano di \mathbb{R}^4 di equazione $x_1 + x_2 + x_3 + x_4 = 0$, la direzione normale al piano è quella di $\mathbf{w} = [1, 1, 1, 1]^T$, e

$$\begin{aligned} \mathbf{R} &= \mathbf{I} - 2\frac{\mathbf{w}\mathbf{w}^T}{\|\mathbf{w}\|^2} = \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix} \end{aligned}$$

Si consideri la matrice

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

- Si spieghi perché \mathbf{A} è la matrice di una rotazione di \mathbb{R}^3 . Si determinino l'asse e l'angolo di rotazione.
- Si determinino gli autovalori di \mathbf{A} . La matrice è diagonalizzabile da una matrice complessa? E da una matrice reale?
- Si determini una matrice ortogonale reale \mathbf{Q} tale che

$$\mathbf{Q}^T \mathbf{A} \mathbf{Q} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

In \mathbb{R}^2 il versore $\mathbf{n} = [-\sin(\alpha), \cos(\alpha)]^T$ è ortogonale alla retta \mathbf{r} che ha coefficiente angolare $\operatorname{tg}(\alpha)$. Concludere che la matrice della riflessione ortogonale rispetto alla retta \mathbf{r} è

$$\mathbf{R} = \mathbf{I} - 2\mathbf{n}\mathbf{n}^T = \begin{bmatrix} \cos(2\alpha) & \sin(2\alpha) \\ \sin(2\alpha) & -\cos(2\alpha) \end{bmatrix}$$

(cf. pagina 245).

 Mostrare che la matrice $n \times n$

è la matrice della riflessione ortogonale di \mathbb{R}^n rispetto all'iperpiano di equazione $x_1 + \dots + x_n = 0$. Concludere che \mathbf{R} è una matrice ortogonale, con un autovalore semplice $\lambda = -1$ e un autovalore regolare $\lambda = 1$ di molteplicità $n - 1$. In particolare, il determinante di \mathbf{R} è -1 e il polinomio caratteristico di \mathbf{R} è $(-1)^n(\lambda + 1)(\lambda - 1)^{n-1}$. Per $n = 2$ si tratta della matrice

$$\mathbf{R} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$$

e per $n = 3$ della matrice

$$\mathbf{R} = \frac{1}{3} \begin{bmatrix} 1 & -2 & -2 \\ -2 & 1 & -2 \\ -2 & -2 & 1 \end{bmatrix}$$

Teoremi spettrali e forme quadratiche

1 INTRODUZIONE

Una matrice quadrata \mathbf{A} di ordine n si può diagonalizzare, e così semplificare drasticamente, se \mathbb{R}^n o \mathbb{C}^n ammette una base formata da autovettori di \mathbf{A} . Nel migliore dei casi è possibile scegliere una base *ortonormale* formata da autovettori della matrice; in questo caso il cambiamento di variabili che diagonalizza la matrice è un'isometria, preserva cioè distanze e angoli: la semplificazione del problema non comporta una deformazione dello spazio ambiente. Il primo importante teorema di questo capitolo, che è uno dei risultati più ricchi di applicazioni dell'algebra lineare, afferma che esiste una base ortonormale formata da autovettori di \mathbf{A} precisamente quando \mathbf{A} è una matrice simmetrica. Le matrici simmetriche rappresentano importanti applicazioni lineari, per esempio le proiezioni ortogonali e le riflessioni ortogonali, ma il più delle volte l'interesse per le matrici simmetriche nasce dal loro utilizzo per descrivere le *forme quadratiche*. Una forma quadratica è un polinomio *omogeneo di secondo grado*; per esempio una forma quadratica in due variabili x e y è un polinomio della forma

$$q(x, y) = ax^2 + 2bxy + cy^2 = [x \ y] \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

I coefficienti del polinomio sono gli elementi della matrice simmetrica $\mathbf{A} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$. Analogamente, una forma quadratica in n variabili $\mathbf{x} = [x_1, \dots, x_n]^T$ si può scrivere nella forma $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ dove \mathbf{A} è una matrice simmetrica di ordine n . Il più importante esempio di forma quadratica è il differenziale secondo di una funzione $f(x_1, \dots, x_n)$; in questo caso la matrice \mathbf{A} è la matrice hessiana di f , i cui elementi sono le derivate parziali seconde di f . Molte grandezze delle scienze naturali si esprimono mediante forme quadratiche. Per esempio, per un corpo rigido in rotazione, il *tensore di inerzia* è una forma quadratica nelle componenti della velocità angolare. Un altro esempio, che discutiamo brevemente in questo capitolo, è fornito dalla geometria differenziale: la prima e la seconda forma fondamentale di una superficie in \mathbb{R}^3 consentono di calcolare lunghezze, angoli e curvature della superficie. Nel contesto delle forme quadratiche, il teorema spettrale garantisce che è possibile trovare un'isometria $\mathbf{x} = \mathbf{Q}\mathbf{X}$ che trasforma la forma generica in una forma in cui compaiono soltanto i quadrati delle variabili.

Si vede così che gli assi del nuovo sistema di riferimento, sono assi di simmetria per gli insiemi di livello $q(\mathbf{x}) = c$ della forma quadratica: il teorema spettrale garantisce che tali assi sono a due a due ortogonali ed è perciò noto anche come *teorema degli assi principali*. Nel caso di due variabili, le curve di livello $q(x, y) = c$ sono delle coniche a centro (ellissi o iperboli) e gli assi di una conica a centro sono perciò ortogonali tra loro. Nel caso del tensore di inerzia di un corpo rigido, la superficie di livello 1 è l'ellissoide d'inerzia, i cui assi di simmetria sono gli assi principali d'inerzia del corpo rigido: di nuovo, per il teorema spettrale, tali assi sono a due a due ortogonali. Nelle applicazioni è fondamentale determinare il segno di una forma quadratica e a questo problema è dedicata gran parte del paragrafo sulle forme quadratiche: tra l'altro dimostriamo i criteri di positività forniti dagli autovalori e dai minori principali di nord-ovest della matrice. In particolare, studiamo le matrici definite positive di cui ricaviamo la fattorizzazione di Cholesky (o equivalentemente la fattorizzazione LU simmetrica).

Nel terzo paragrafo del capitolo descriviamo la *decomposizione ai valori singolari* di una matrice qualsiasi; questa decomposizione fornisce un utile surrogato del teorema spettrale per una matrice arbitraria e gioca un ruolo sempre più importante nelle applicazioni dell'algebra lineare. Nel quarto paragrafo illustriamo il teorema spettrale complesso, che caratterizza le matrici \mathbf{N} per cui esiste una base ortonormale di \mathbb{C}^n formata da autovettori di \mathbf{N} . Tali matrici si dicono *normali* e costituiscono una classe molto più ampia delle matrici simmetriche; per esempio le matrici ortogonali e le matrici antisimmetriche sono normali. Questo ci permette, nel paragrafo successivo, di fornire una forma canonica semplice, anche se non diagonale, per una matrice normale reale. Infine, nell'ultimo paragrafo usiamo il teorema spettrale per ricavare l'equazione canonica di una quadrica in \mathbb{R}^n . Una quadrica è un'ipersuperficie definita da una equazione di secondo grado; in particolare, ricaviamo le equazioni canoniche di una conica del piano.

2 TEOREMA SPETTRALE

In questo paragrafo tutte le matrici sono a elementi reali: una matrice simmetrica \mathbf{A} è una matrice quadrata a elementi reali tale che $\mathbf{A}^T = \mathbf{A}$; una matrice ortogonale \mathbf{Q} è una matrice quadrata a elementi reali tale che $\mathbf{Q}^T = \mathbf{Q}^{-1}$.

DEFINIZIONE 2.1 (Matrici ortogonalmente diagonalizzabili)

Sia \mathbf{A} una matrice quadrata reale. Si dice che \mathbf{A} è *ortogonalmente diagonalizzabile* se esiste una matrice ortogonale \mathbf{Q} tale che $\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$ sia diagonale.

OSSERVAZIONE Una matrice \mathbf{A} quadrata di ordine n è ortogonalmente diagonalizzabile se e solo se esiste una base ortonormale di \mathbb{R}^n i cui elementi sono autovettori di \mathbf{A} . Supponiamo infatti che \mathbf{A} e \mathbf{Q} siano matrici $n \times n$. La matrice \mathbf{Q} è ortogonale se e solo se le sue colonne $\mathbf{q}_1, \dots, \mathbf{q}_n$ formano una base ortonormale di \mathbb{R}^n . Se $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_n)$, allora l'uguaglianza $\mathbf{D} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$ equivale a

$$\mathbf{A}\mathbf{q}_k = \lambda_k \mathbf{q}_k \quad \text{per } k = 1, 2, \dots, n$$

Quindi \mathbf{A} è ortogonalmente diagonalizzabile se e solo se esiste una base ortonormale $\{\mathbf{q}_1, \dots, \mathbf{q}_n\}$ di \mathbb{R}^n i cui elementi sono autovettori di \mathbf{A} .

OSSERVAZIONE L'inversa di una matrice ortogonale \mathbf{Q} è la matrice trasposta \mathbf{Q}^T , per cui

$$\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q} = \mathbf{Q}^T\mathbf{A}\mathbf{Q}$$

Il risultato principale di questo paragrafo è il teorema spettrale secondo il quale una matrice è ortogonalmente diagonalizzabile se e solo se è simmetrica. Un'implicazione del teorema è semplice: se \mathbf{A} è ortogonalmente diagonalizzabile, allora \mathbf{A} è simmetrica. Infatti, se esiste una matrice ortogonale \mathbf{Q} tale che $\mathbf{D} = \mathbf{Q}^T\mathbf{A}\mathbf{Q}$ sia diagonale, allora $\mathbf{A} = \mathbf{Q}\mathbf{D}\mathbf{Q}^T$ e

$$\mathbf{A}^T = (\mathbf{Q}\mathbf{D}\mathbf{Q}^T)^T = (\mathbf{Q}^T)^T\mathbf{D}^T\mathbf{Q}^T = \mathbf{Q}\mathbf{D}\mathbf{Q}^T = \mathbf{A}$$

Si osservi che nel penultimo passaggio abbiamo usato il fatto che una matrice diagonale è simmetrica. La parte profonda del teorema spettrale è l'altra implicazione: se \mathbf{A} è simmetrica, allora \mathbf{A} è ortogonalmente diagonalizzabile. Cominciamo col mostrare che gli autovalori di una matrice simmetrica sono reali.

PROPOSIZIONE 2.2 Gli autovalori di una matrice simmetrica sono reali.

DIMOSTRAZIONE. Sia λ un autovalore di \mathbf{A} . A priori, λ è un numero complesso; possiamo comunque trovare un autovettore complesso relativo a λ , cioè un vettore non nullo $\mathbf{v} \in \mathbb{C}^n$ tale che $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Si ricordi che il prodotto hermitiano di due vettori di \mathbb{C}^n è definito dalla formula

$$\langle \mathbf{z}, \mathbf{w} \rangle = \mathbf{w}^H \mathbf{z} = \sum_{k=1}^n z_k \bar{w}_k$$

dove \mathbf{w}^H è il coniugato del vettore \mathbf{w}^T . Questo prodotto è lineare in \mathbf{z} e sequilineare in \mathbf{w} ; in particolare

$$\langle t\mathbf{z}, \mathbf{w} \rangle = t \langle \mathbf{z}, \mathbf{w} \rangle \quad \text{e} \quad \langle \mathbf{z}, t\mathbf{w} \rangle = \bar{t} \langle \mathbf{z}, \mathbf{w} \rangle \quad \text{per ogni } t \in \mathbb{C}$$

L'idea è di calcolare i prodotti hermitiani $\langle \mathbf{A}\mathbf{v}, \mathbf{v} \rangle$ e $\langle \mathbf{v}, \mathbf{A}\mathbf{v} \rangle$ e di sfruttare la simmetria di \mathbf{A} per mostrare che sono uguali. Il primo di questi prodotti è

$$\langle \mathbf{A}\mathbf{v}, \mathbf{v} \rangle = \langle \lambda\mathbf{v}, \mathbf{v} \rangle = \lambda \langle \mathbf{v}, \mathbf{v} \rangle = \lambda \|\mathbf{v}\|^2$$

Il secondo prodotto è

$$\langle \mathbf{v}, \mathbf{A}\mathbf{v} \rangle = \langle \mathbf{v}, \lambda\mathbf{v} \rangle = \bar{\lambda} \langle \mathbf{v}, \mathbf{v} \rangle = \bar{\lambda} \|\mathbf{v}\|^2$$

Siccome \mathbf{A} è una matrice simmetrica reale, \mathbf{A} è uguale alla sua matrice trasposta coniugata \mathbf{A}^H , e i due prodotti hermitiani che stiamo considerando sono uguali:

$$\langle \mathbf{v}, \mathbf{A}\mathbf{v} \rangle = (\mathbf{A}\mathbf{v})^H \mathbf{v} = \mathbf{v}^H \mathbf{A}^H \mathbf{v} = \mathbf{v}^H \mathbf{A}\mathbf{v} = \langle \mathbf{A}\mathbf{v}, \mathbf{v} \rangle$$

Perciò

$$\lambda \|\mathbf{v}\|^2 = \bar{\lambda} \|\mathbf{v}\|^2$$

Siccome \mathbf{v} è non nullo, la sua norma $\|\mathbf{v}\|$ è un numero diverso da zero; possiamo quindi dividere l'uguaglianza precedente per $\|\mathbf{v}\|^2$ ottenendo $\lambda = \bar{\lambda}$. Questo mostra che la parte immaginaria di λ è nulla e conclude, quindi, la dimostrazione.

PROPOSIZIONE 2.3 (Ortogonalità degli autovettori di una matrice simmetrica) Sia \mathbf{A} una matrice simmetrica. Supponiamo che λ e μ siano autovalori distinti di \mathbf{A} e che \mathbf{v} e \mathbf{w} siano autovettori di \mathbf{A} relativi a λ e μ rispettivamente. Allora \mathbf{v} e \mathbf{w} sono ortogonali.

DIMOSTRAZIONE. Se n è l'ordine di \mathbf{A} , gli autovettori appartengono a \mathbb{R}^n e dobbiamo mostrare che il loro prodotto scalare $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T \mathbf{w}$ è nullo. Siccome \mathbf{A} è simmetrica, i prodotti scalari $\langle \mathbf{A}\mathbf{v}, \mathbf{w} \rangle$ e $\langle \mathbf{v}, \mathbf{A}\mathbf{w} \rangle$ sono uguali:

$$\langle \mathbf{A}\mathbf{v}, \mathbf{w} \rangle = (\mathbf{A}\mathbf{v})^T \mathbf{w} = \mathbf{v}^T \mathbf{A}^T \mathbf{w} = \mathbf{v}^T \mathbf{A}\mathbf{w} = \langle \mathbf{v}, \mathbf{A}\mathbf{w} \rangle$$

Ora usiamo il fatto che \mathbf{v} è un autovettore di \mathbf{A} relativo a λ , cioè $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$:

$$\langle \mathbf{A}\mathbf{v}, \mathbf{w} \rangle = \langle \lambda\mathbf{v}, \mathbf{w} \rangle = \lambda \langle \mathbf{v}, \mathbf{w} \rangle$$

Analogamente, siccome $\mathbf{A}\mathbf{w} = \mu\mathbf{w}$

$$\langle \mathbf{v}, \mathbf{A}\mathbf{w} \rangle = \langle \mathbf{v}, \mu\mathbf{w} \rangle = \mu \langle \mathbf{v}, \mathbf{w} \rangle$$

Quindi $\lambda \langle \mathbf{v}, \mathbf{w} \rangle = \mu \langle \mathbf{v}, \mathbf{w} \rangle$. Poiché per ipotesi $\lambda \neq \mu$, dev'essere $\langle \mathbf{v}, \mathbf{w} \rangle = 0$.

Supponiamo ora che \mathbf{A} sia una matrice simmetrica di ordine n che abbia n autovalori distinti $\lambda_1, \lambda_2, \dots, \lambda_n$. Per la proposizione 2.2 gli autovalori sono reali; per ogni autovalore λ_k scegliamo un corrispondente autovettore $\mathbf{v}_k \in \mathbb{R}^n$. Per la proposizione 2.3 gli autovettori $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ sono a due a due ortogonali, e quindi formano una base ortogonale di \mathbb{R}^n . Normalizziamo gli autovettori \mathbf{v}_k definendo

$$\mathbf{q}_k = \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|} \quad \text{per ogni } k = 1, 2, \dots, n$$

Per costruzione \mathbf{q}_k è un versore ed è un autovettore di \mathbf{A} relativo a λ_k . Abbiamo così costruito una base ortonormale $\{\mathbf{q}_1, \dots, \mathbf{q}_n\}$ di \mathbb{R}^n formata da autovettori di \mathbf{A} , sotto le ipotesi che \mathbf{A} sia simmetrica e abbia autovalori distinti. Quest'ultima ipotesi in realtà non è necessaria:

TEOREMA 2.4 (Teorema spettrale o degli assi principali)

Una matrice simmetrica reale è ortogonalmente diagonalizzabile.

DIMOSTRAZIONE. Dimostriamo il teorema per induzione sull'ordine n della matrice. Il caso iniziale $n = 1$ è ovvio: ogni matrice 1×1 è diagonale, ammette 1 come autovettore, e $\{1\}$ è una base ortonormale di \mathbb{R} .

Sia ora $n \geq 2$ e supponiamo che il teorema sia vero per matrici simmetriche di ordine $n - 1$: dobbiamo dimostrarlo per una matrice simmetrica \mathbf{A} di ordine n . Per il teorema fondamentale dell'algebra il polinomio caratteristico $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ ha una radice $\lambda_1 \in \mathbb{C}$. In quanto radice del polinomio caratteristico, il numero complesso λ_1 è un autovalore di \mathbf{A} ;

dalla proposizione 2.2 segue che λ_1 è reale. Esiste perciò un autovettore $\mathbf{q}_1 \in \mathbb{R}^n$ relativo a λ_1 e di norma 1. Possiamo trovare una base di \mathbb{R}^n il cui primo elemento sia \mathbf{q}_1 e poi applicare l'algoritmo di Gram-Schmidt per ottenere una base ortonormale $\{\mathbf{q}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ di \mathbb{R}^n . Sia $\mathbf{P} = [\mathbf{q}_1 \ \mathbf{b}_2 \ \dots \ \mathbf{b}_n]$ la corrispondente matrice ortogonale. La matrice $\mathbf{C} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{P}^T\mathbf{A}\mathbf{P}$ rappresenta l'applicazione lineare $\mathcal{L}_{\mathbf{A}}$ rispetto alla base $\{\mathbf{q}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$; in particolare, la prima colonna di \mathbf{C} è il vettore delle coordinate di $\mathcal{L}_{\mathbf{A}}(\mathbf{q}_1)$ rispetto a tale base, cioè

$$\mathcal{L}_{\mathbf{A}}(\mathbf{q}_1) = c_{11}\mathbf{q}_1 + c_{21}\mathbf{b}_2 + \dots + c_{n1}\mathbf{b}_n$$

D'altra parte, $\mathcal{L}_{\mathbf{A}}(\mathbf{q}_1) = \mathbf{A}\mathbf{q}_1 = \lambda_1\mathbf{q}_1$. Confrontando queste due espressioni concludiamo che $c_{11} = \lambda_1$ e $c_{j1} = 0$ per $j = 2, \dots, n$. Osserviamo ora che la matrice \mathbf{C} è simmetrica perché è ortogonalmente simile alla matrice simmetrica \mathbf{A} :

$$\mathbf{C}^T = (\mathbf{P}^T\mathbf{A}\mathbf{P})^T = \mathbf{P}^T\mathbf{A}^T\mathbf{P} = \mathbf{P}^T\mathbf{A}\mathbf{P} = \mathbf{C}$$

Quindi $c_{1j} = c_{j1} = 0$ per $j = 2, \dots, n$. La matrice \mathbf{C} è perciò della forma

$$\mathbf{C} = \left[\begin{array}{c|c} \lambda_1 & \mathbf{0}^T \\ \hline \mathbf{0} & \mathbf{A}_1 \end{array} \right]$$

dove $\mathbf{0}$ denota il vettore nullo di \mathbb{R}^{n-1} , e \mathbf{A}_1 è una matrice quadrata di ordine $n-1$. Siccome \mathbf{C} è simmetrica, anche \mathbf{A}_1 è simmetrica. Per l'ipotesi di induzione \mathbf{A}_1 è ortogonalmente diagonalizzabile: esiste una matrice ortogonale \mathbf{Q}_1 di ordine $n-1$ tale che $\mathbf{Q}_1^T\mathbf{A}_1\mathbf{Q}_1 = \mathbf{D}_1$ è diagonale. La matrice

$$\mathbf{R} = \left[\begin{array}{c|c} 1 & \mathbf{0}^T \\ \hline \mathbf{0} & \mathbf{Q}_1 \end{array} \right]$$

è ortogonale perché \mathbf{Q}_1 lo è, e

$$\mathbf{R}^T\mathbf{C}\mathbf{R} = \left[\begin{array}{c|c} 1 & \mathbf{0}^T \\ \hline \mathbf{0} & \mathbf{Q}_1^T \end{array} \right] \left[\begin{array}{c|c} \lambda_1 & \mathbf{0}^T \\ \hline \mathbf{0} & \mathbf{A}_1 \end{array} \right] \left[\begin{array}{c|c} 1 & \mathbf{0}^T \\ \hline \mathbf{0} & \mathbf{Q}_1 \end{array} \right] = \left[\begin{array}{c|c} \lambda_1 & \mathbf{0}^T \\ \hline \mathbf{0} & \mathbf{Q}_1^T\mathbf{A}_1\mathbf{Q}_1 \end{array} \right] = \left[\begin{array}{c|c} \lambda_1 & \mathbf{0}^T \\ \hline \mathbf{0} & \mathbf{D}_1 \end{array} \right] = \mathbf{D}$$

La matrice \mathbf{D} è diagonale perché \mathbf{D}_1 lo è. Infine sia $\mathbf{Q} = \mathbf{P}\mathbf{R}$. La matrice \mathbf{Q} è ortogonale perché prodotto di matrici ortogonali, e

$$\mathbf{Q}^T\mathbf{A}\mathbf{Q} = \mathbf{R}^T\mathbf{P}^T\mathbf{A}\mathbf{P}\mathbf{R} = \mathbf{R}^T\mathbf{C}\mathbf{R} = \mathbf{D}$$

Siccome \mathbf{Q} è ortogonale e \mathbf{D} è diagonale, questo completa la dimostrazione.

COROLLARIO 2.5 (Autospazi di una matrice simmetrica)

Sia \mathbf{A} una matrice simmetrica di ordine n e siano $\lambda_1, \lambda_2, \dots, \lambda_s$ i suoi autovalori distinti. Sia \mathbf{V}_j l'autospazio relativo all'autovalore λ_j :

$$\mathbf{V}_j = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \lambda_j\mathbf{x}\}$$

e sia \mathcal{B}_j una base ortonormale di \mathbf{V}_j .

Allora

- a) l'unione $\mathcal{B}_1 \cup \dots \cup \mathcal{B}_s$ è una base ortonormale di \mathbb{R}^n ;
- b) $\mathbf{V}_1 + \mathbf{V}_2 + \dots + \mathbf{V}_s = \mathbb{R}^n$;
- c) gli autospazi \mathbf{V}_j sono a due a due ortogonali.

DIMOSTRAZIONE. Per il teorema spettrale ogni matrice simmetrica è diagonalizzabile. Nel capitolo sugli autovalori e autovettori abbiamo definito la molteplicità geometrica di λ_j come la dimensione di \mathbf{V}_j e abbiamo mostrato che per ogni matrice diagonalizzabile di ordine n , la somma delle molteplicità geometriche degli autovalori è n . Quindi l'unione $\mathcal{B} = \mathcal{B}_1 \cup \dots \cup \mathcal{B}_s$ consiste di n versori, e per mostrare che si tratta di una base ortonormale occorre solo far vedere che tali versori sono a due a due ortogonali. Siano \mathbf{q} e \mathbf{q}' due versori in \mathcal{B} . Se $\mathcal{B}_1 \cup \dots \cup \mathcal{B}_s$ sono autovettori relativi allo stesso autovalore λ_j , cioè appartengono entrambi a \mathcal{B}_j , allora sono ortogonali perché \mathcal{B}_j è per ipotesi ortonormale. Se invece \mathbf{q} e \mathbf{q}' sono autovettori relativi ad autovalori distinti, allora sono ortogonali per la proposizione 2.3. Questo mostra il primo punto. Il secondo punto è ora evidente. Il terzo punto significa che, se $i \neq j$, $\mathbf{v} \in \mathbf{V}_i$ e $\mathbf{w} \in \mathbf{V}_j$, allora \mathbf{v} è ortogonale a \mathbf{w} ; e questo è vero per la proposizione 2.3.

OSSERVAZIONE Per costruire una base ortonormale di \mathbb{R}^n formata da autovettori di una matrice simmetrica \mathbf{A} si procede come suggerito dal corollario precedente: per ogni autovalore λ_j di \mathbf{A} , si costruisce una base di $\mathbf{V}_j = \text{Ker}(\mathbf{A} - \lambda_j \mathbf{I})$ risolvendo il sistema $(\mathbf{A} - \lambda_j \mathbf{I})\mathbf{x} = \mathbf{0}$; dopodiché si trasforma tale base in una base ortonormale con l'algoritmo di Gram-Schmidt, infine, si prende l'unione delle basi ortonormali dei singoli autospazi.

OSSERVAZIONE Ogni base ortonormale di \mathbb{R}^n formata da autovettori di \mathbf{A} è l'unione di basi ortonormali dei singoli autospazi come nel corollario 2.5. Infatti, dato una autovalore λ , ogni base formata da autovettori contiene un numero di autovettori relativi a λ pari alla molteplicità geometrica di λ ; tali autovettori sono linearmente indipendenti e formano una base dell'autospazio \mathbf{V}_λ .

OSSERVAZIONE Se si hanno a disposizione i risultati del paragrafo sulla forma canonica di Jordan, la dimostrazione del teorema spettrale si semplifica. Infatti, per mostrare che la matrice \mathbf{A} è diagonalizzabile basta dimostrare che ogni autovalore ha indice 1, cioè $\text{Ker}(\mathbf{A} - \lambda \mathbf{I}) = \text{Ker}((\mathbf{A} - \lambda \mathbf{I})^2)$. Poniamo $\mathbf{B} = \mathbf{A} - \lambda \mathbf{I}$. La matrice \mathbf{B} è simmetrica, quindi $\mathbf{B} = \mathbf{B}^T$ e $\mathbf{B}^2 = \mathbf{B}^T \mathbf{B}$. Come abbiamo dimostrato nel paragrafo sui minimi quadrati, $\text{Ker}(\mathbf{B}) = \text{Ker}(\mathbf{B}^T \mathbf{B})$. Quindi ogni autovalore di \mathbf{A} ha indice 1 e \mathbf{A} è diagonalizzabile. Questa osservazione, insieme al fatto che gli autovalori di \mathbf{A} sono reali e che autovettori corrispondenti ad autovalori distinti sono ortogonali, dimostra il teorema spettrale. Il corollario 2.5 segue poi dal fatto che, se una matrice \mathbf{A} è diagonalizzabile su \mathbb{R} , allora \mathbb{R}^n è la somma diretta degli autospazi di \mathbf{A} .

Il corollario 2.5 si può riformulare in modo algebrico:

COROLLARIO 2.6 (Decomposizione spettrale)

Sia \mathbf{A} una matrice simmetrica e siano $\lambda_1, \lambda_2, \dots, \lambda_s$ i suoi autovalori *distinti*.

Sia \mathbf{P}_j la matrice della proiezione ortogonale sull'autospazio \mathbf{V}_j relativo a λ_j .

Allora:

- a) $\mathbf{A} = \lambda_1 \mathbf{P}_1 + \dots + \lambda_s \mathbf{P}_s$;
- b) $\mathbf{I} = \mathbf{P}_1 + \dots + \mathbf{P}_s$;
- c) le matrici \mathbf{P}_j sono simmetriche, $\mathbf{P}_j^2 = \mathbf{P}_j$, e $\mathbf{P}_i \mathbf{P}_j = \mathbf{O}$ se $i \neq j$.

DIMOSTRAZIONE. Per il teorema spettrale esiste una matrice ortogonale \mathbf{Q} tale che $\mathbf{Q}^T \mathbf{A} \mathbf{Q} = \text{diag}(\mu_1, \dots, \mu_n)$ (scriviamo μ_i e non λ_j perché non è detto che gli autovalori μ_i siano distinti). Detta \mathbf{q}_k la colonna k di \mathbf{Q} , valgono le uguaglianze:

$$\begin{aligned}\mathbf{I} &= \mathbf{Q} \mathbf{Q}^T = \mathbf{q}_1 \mathbf{q}_1^T + \mathbf{q}_2 \mathbf{q}_2^T + \cdots + \mathbf{q}_n \mathbf{q}_n^T, \\ \mathbf{A} &= \mathbf{Q} \mathbf{D} \mathbf{Q}^T = \mu_1 \mathbf{q}_1 \mathbf{q}_1^T + \mu_2 \mathbf{q}_2 \mathbf{q}_2^T + \cdots + \mu_n \mathbf{q}_n \mathbf{q}_n^T\end{aligned}$$

Possiamo raggruppare i termini corrispondenti a uno stesso autovalore: se $g(j)$ è la molteplicità di λ_j , allora esistono $g(j)$ indici $k_1, \dots, k_{g(j)}$ tali che $\mu_{k_i} = \lambda_j$, e $\{\mathbf{q}_{k_1}, \dots, \mathbf{q}_{k_{g(j)}}\}$ è una base ortonormale di \mathbf{V}_j . Nel capitolo sugli spazi euclidei, proposizione 7.1, abbiamo mostrato che la matrice che rappresenta la proiezione ortogonale di \mathbb{R}^n sul sottospazio \mathbf{V}_j è la matrice simmetrica

$$\mathbf{P}_j = \mathbf{q}_{k_1} \mathbf{q}_{k_1}^T + \cdots + \mathbf{q}_{k_{g(j)}} \mathbf{q}_{k_{g(j)}}^T$$

Otteniamo così:

$$\begin{aligned}\mathbf{I} &= \mathbf{P}_1 + \mathbf{P}_2 + \cdots + \mathbf{P}_s, \\ \mathbf{A} &= \lambda_1 \mathbf{P}_1 + \lambda_2 \mathbf{P}_2 + \cdots + \lambda_s \mathbf{P}_s\end{aligned}$$

Per terminare la dimostrazione fissiamo un vettore $\mathbf{x} \in \mathbb{R}^n$ e un indice j e poniamo $\mathbf{p} = \mathbf{P}_j(\mathbf{x}) \in \mathbf{V}_j$. Allora $\mathbf{P}_j(\mathbf{p}) = \mathbf{p}$ perché \mathbf{P}_j rappresenta la proiezione su \mathbf{V}_j , mentre, se $i \neq j$, $\mathbf{P}_i \mathbf{p} = \mathbf{0}$ perché $\mathbf{p} \in \mathbf{V}_j \subseteq \mathbf{V}_i^\perp$. Quindi $\mathbf{P}_j^2 \mathbf{x} = \mathbf{P}_j \mathbf{x}$ e $\mathbf{P}_i \mathbf{P}_j \mathbf{x} = \mathbf{0}$. Siccome \mathbf{x} è arbitrario, ne segue che $\mathbf{P}_j^2 = \mathbf{P}_j$, e $\mathbf{P}_i \mathbf{P}_j = \mathbf{0}$ se $i \neq j$.

L'insieme degli autovalori di \mathbf{A} si dice lo *spettro* di \mathbf{A} e la decomposizione

$$(2.1) \quad \mathbf{A} = \lambda_1 \mathbf{P}_1 + \cdots + \lambda_s \mathbf{P}_s$$

si dice *decomposizione spettrale* di \mathbf{A} , da cui il nome del teorema.

OSSERVAZIONE La decomposizione spettrale ha importanti conseguenze. Per esempio, dalla (2.1) segue che per ogni intero positivo p

$$(2.2) \quad \mathbf{A}^p = \lambda_1^p \mathbf{P}_1 + \cdots + \lambda_s^p \mathbf{P}_s$$

Questa formula vale anche per p negativi se \mathbf{A} è invertibile, cioè se ogni autovalore è diverso da zero. In particolare

$$(2.3) \quad \mathbf{A}^{-1} = \lambda_1^{-1} \mathbf{P}_1 + \cdots + \lambda_s^{-1} \mathbf{P}_s$$

Possiamo spingerci oltre: per ogni funzione $f(x)$ il cui dominio contenga gli autovalori di \mathbf{A} , possiamo definire la matrice $f(\mathbf{A})$ mediante la formula

$$(2.4) \quad f(\mathbf{A}) = f(\lambda_1) \mathbf{P}_1 + \cdots + f(\lambda_s) \mathbf{P}_s$$

Per esempio, prendendo $f(x) = \sqrt{x}$, possiamo definire $\sqrt{\mathbf{A}}$ a condizione che gli autovalori di \mathbf{A} siano maggiori o uguali a zero:

$$(2.5) \quad \sqrt{\mathbf{A}} = \sqrt{\lambda_1} \mathbf{P}_1 + \cdots + \sqrt{\lambda_s} \mathbf{P}_s$$

Lasciamo al lettore il compito di verificare che $\sqrt{\mathbf{A}}$ è l'unica matrice simmetrica con autovalori ≥ 0 il cui quadrato sia \mathbf{A} .

Consideriamo la matrice

$$\mathbf{A} = \begin{bmatrix} 5 & -4 & 0 \\ -4 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Il polinomio caratteristico di \mathbf{A} è $-(\lambda - 1)^2(\lambda - 9)$. Gli autovalori distinti di \mathbf{A} sono $\lambda_1 = 1$, che ha molteplicità 2 e $\lambda_2 = 9$, che è semplice. L'autospazio \mathbf{V}_1 relativo a $\lambda_1 = 1$ è il piano di equazione $x_1 - x_2 = 0$, una cui base ortonormale è formata dai vettori

$$\mathbf{q}_1 = \frac{1}{\sqrt{2}}[1, 1, 0]^T \quad \text{e} \quad \mathbf{q}_2 = [0, 0, 1]^T$$

L'autospazio \mathbf{V}_2 relativo a $\lambda_2 = 9$ è la retta di equazioni $x_1 + x_2 = x_3 = 0$, una cui base ortonormale è formata dal versore $\mathbf{q}_3 = \frac{1}{\sqrt{2}}[1, -1, 0]^T$. Una matrice ortogonale che diagonalizza \mathbf{A} è quindi

$$\mathbf{Q} = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \mathbf{q}_3] = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix}$$

La matrice diagonale a cui \mathbf{A} è ortogonalmente simile è

$$\mathbf{D} = \mathbf{Q}^T \mathbf{A} \mathbf{Q} = \text{diag}(1, 1, 9)$$

Per scrivere la decomposizione spettrale di \mathbf{A} calcoliamo le matrici delle proiezioni sugli autospazi:

$$\mathbf{P}_2 = \mathbf{q}_3 \mathbf{q}_3^T = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{e} \quad \mathbf{P}_1 = \mathbf{I} - \mathbf{P}_2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

La decomposizione spettrale di \mathbf{A} è perciò:

$$\mathbf{A} = \mathbf{P}_1 + 9\mathbf{P}_2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} + 9 \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Da questo segue per esempio che

$$\sqrt{\mathbf{A}} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} + 3 \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Il lettore dovrebbe verificare che $(\sqrt{\mathbf{A}})^2 = \mathbf{A}$ e che gli autovalori di $\sqrt{\mathbf{A}}$ sono $\lambda_1 = 1$ e $\lambda_2 = 3$.

Matrici ortogonali simmetriche

In questo esempio mostriamo che le matrici ortogonali simmetriche sono precisamente le matrici delle riflessioni ortogonali (che abbiamo già studiato nei complementi del capitolo sugli spazi euclidei). Supponiamo che una matrice reale \mathbf{R} sia ortogonale e simmetrica. Gli

autovalori di \mathbf{R} sono reali perché \mathbf{R} è simmetrica, e hanno modulo 1 perché \mathbf{R} è ortogonale. Pertanto una matrice ortogonale simmetrica ha solo gli autovalori $\lambda_1 = 1$ e $\lambda_2 = -1$. La decomposizione spettrale di \mathbf{R} è quindi:

$$\mathbf{R} = \mathbf{P}_1 - \mathbf{P}_2 = \mathbf{I} - 2\mathbf{P}_2$$

Si noti che \mathbf{P}_1 è la proiezione ortogonale sull'autospazio \mathbf{V}_1 relativo a $\lambda_1 = 1$; questo autospazio è l'asse della riflessione, cioè il luogo dei punti lasciati fissi dalla riflessione. Invece \mathbf{P}_2 è la proiezione sull'autospazio \mathbf{V}_2 relativo a $\lambda_2 = -1$, che è il complemento ortogonale di \mathbf{V}_1 e consiste dei vettori che sono mandati nel proprio opposto dalla riflessione. Il caso più comune nelle applicazioni è quello in cui \mathbf{V}_1 sia un iperpiano; in tal caso \mathbf{V}_2 è la retta normale a \mathbf{V}_1 e una sua base ortonormale non è altro che un versore \mathbf{n} normale a \mathbf{V}_1 ; la matrice della riflessione in questo caso è $\mathbf{R} = \mathbf{I} - 2\mathbf{n}\mathbf{n}^T$.

La matrice identità e la matrice nulla sono simmetriche? Sono matrici di proiezione? Su quale sottospazio? Una matrice diagonale è simmetrica? Quali matrici diagonali sono matrici di proiezione? Qual è la decomposizione spettrale di una matrice diagonale?

Siano \mathbf{A} e \mathbf{B} due matrici simmetriche di ordine n . Mostrare che, se $\mathbf{AB} = \mathbf{BA}$, allora \mathbf{AB} è simmetrica. L'ipotesi $\mathbf{AB} = \mathbf{BA}$ è necessaria?

Per ciascuna delle seguenti matrici simmetriche \mathbf{A}

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 2 \\ 2 & -2 \end{bmatrix}, \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix}, \begin{bmatrix} 2 & 2 & 0 \\ 2 & 4 & 2 \\ 0 & 2 & 2 \end{bmatrix}$$

si trovi una matrice ortogonale \mathbf{Q} che diagonalizza \mathbf{A} , e si scriva la decomposizione spettrale di \mathbf{A} . Si scriva poi la decomposizione spettrale di \mathbf{A}^3 e di \mathbf{A}^{-1} .

Si trovi una base ortonormale di \mathbb{R}^3 formata da autovettori della matrice $\mathbf{A} = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$. Si scriva la decomposizione spettrale di \mathbf{A} e si calcoli $\sqrt{\mathbf{A}}$. Si verifichi che $\sqrt{\mathbf{A}}$ è simmetrica con autovalori ≥ 0 e che $(\sqrt{\mathbf{A}})^2 = \mathbf{A}$.

Si consideri la matrice

$$\mathbf{A} = \begin{bmatrix} 9 & 2 & k \\ 2 & 2 & 0 \\ -2 & 0 & 2 \end{bmatrix}$$

Per quali valori di k la matrice è ortogonalmente diagonalizzabile? Per tali valori di k si trovino una matrice ortogonale \mathbf{Q} e una matrice diagonale \mathbf{D} tali che $\mathbf{D} = \mathbf{Q}^T \mathbf{A} \mathbf{Q}$ e si scriva il polinomio caratteristico di \mathbf{A}^3 .

Sia $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ la base canonica di \mathbb{R}^3 . Sia $\mathcal{L} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ l'applicazione lineare tale che

$$\mathcal{L}(\mathbf{e}_1) = 3\mathbf{e}_1 + \mathbf{e}_3, \quad \mathcal{L}(\mathbf{e}_2) = \mathbf{e}_2, \quad \mathcal{L}(\mathbf{e}_3) = \mathbf{e}_1 + 3\mathbf{e}_3.$$

Trovare la matrice \mathbf{A} che rappresenta \mathcal{L} rispetto alla base canonica. Trovare, se esiste, una matrice ortogonale \mathbf{Q} tale che $\mathbf{Q}^{-1} \mathbf{A} \mathbf{Q}$ sia diagonale. La matrice \mathbf{A}^4 ha una base ortonormale di autovettori?

 Si consideri la matrice

$$\mathbf{A} = \frac{1}{2} \begin{bmatrix} 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \\ -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \end{bmatrix}$$

Posto $\mathbf{V}_1 = \text{Ker}(\mathbf{A} - \mathbf{I})$ e $\mathbf{V}_{-1} = \text{Ker}(\mathbf{A} + \mathbf{I})$, determinare una base di \mathbf{V}_1 e una base di \mathbf{V}_{-1} , e verificare che $\dim \mathbf{V}_1 + \dim \mathbf{V}_{-1} = 4$. Qual è il polinomio caratteristico di \mathbf{A} ? Determinare se possibile una matrice ortogonale \mathbf{Q} e una matrice diagonale \mathbf{D} tali che $\mathbf{Q}^T \mathbf{A} \mathbf{Q} = \mathbf{D}$.

 Mostrare che, se \mathbf{w} è un vettore non nullo di \mathbb{R}^n , la matrice

$$\mathbf{Q} = \mathbf{I} - \frac{2}{\mathbf{w}^T \mathbf{w}} \mathbf{w} \mathbf{w}^T$$

è una matrice $n \times n$ ortogonale e simmetrica; scrivere tale matrice quando $n = 3$ e \mathbf{w} è un vettore perpendicolare al sottospazio di equazione $x_1 + x_2 - x_3 = 0$.

 Calcolare gli autovalori di

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & -2 \\ 2 & 5 & 1 \\ -2 & 1 & 5 \end{bmatrix}$$

e dedurre che \mathbf{A} è uguale a $6\mathbf{P}$ dove \mathbf{P} è una matrice di proiezione ortogonale. Trovare una base ortonormale di \mathbb{R}^3 formata da autovettori di \mathbf{A} .

 Per quali valori del parametro reale a la matrice

$$\mathbf{A} = \begin{bmatrix} 10 & 2a & a \\ 0 & 9 & 2 \\ 0 & 2 & 6 \end{bmatrix}$$

è diagonalizzabile? Per tali valori si determini, se possibile, una base ortonormale di \mathbb{R}^3 formata da autovettori di \mathbf{A} .

 Sia $\mathfrak{P} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ la proiezione ortogonale su un sottospazio \mathbf{H} di dimensione d . Qual è il polinomio caratteristico di \mathfrak{P} ?

 Un autovettore di una matrice simmetrica è necessariamente un versore? Quali matrici simmetriche hanno due autovettori linearmente indipendenti, ma non ortogonali tra loro?

 Sia \mathbf{A} una matrice diagonalizzabile e sia λ un autovalore di \mathbf{A} . Mostrare che λ^2 è un autovalore di \mathbf{A}^2 e che l'autospazio di \mathbf{A}^2 relativo a λ^2 coincide con l'autospazio di \mathbf{A} relativo a λ . Mostrare che questo non è necessariamente vero per una matrice non diagonalizzabile.

 Sia \mathbf{A} una matrice simmetrica con autovalori ≥ 0 . Mostrare che esiste un'unica matrice simmetrica \mathbf{B} con autovalori ≥ 0 tale che $\mathbf{B}^2 = \mathbf{A}$.

■ 3 FORME QUADRATICHE

Questo paragrafo è dedicato allo studio delle *forme quadratiche reali* e del loro segno. Si tratta di un argomento ricco di applicazioni, in primo luogo ai problemi di ottimizzazione per le funzioni di più variabili.

Il termine *forma* significa polinomio omogeneo, cioè un polinomio i cui monomi hanno tutti lo stesso grado. Una forma si dice *lineare* se i suoi monomi hanno grado 1: una forma lineare reale nelle variabili x_1, x_2, \dots, x_n è quindi una funzione del tipo

$$\mathfrak{L}(\mathbf{x}) = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$$

dove i coefficienti a_k sono numeri reali e $\mathbf{x} = [x_1, \dots, x_n]^T$. In altri termini, una forma lineare reale non è altro che un'applicazione lineare $\mathfrak{L} : \mathbb{R}^n \rightarrow \mathbb{R}$ e come sappiamo è rappresentata dal vettore riga $\mathbf{a}^T = [a_1, a_2, \dots, a_n]$.

Una forma si dice *quadratica* se i suoi monomi hanno grado 2 e si dice *reale* se i coefficienti dei monomi sono reali. Una forma quadratica reale nelle variabili x_1, \dots, x_n è quindi una combinazione lineare, a coefficienti reali, dei quadrati x_i^2 delle variabili e dei prodotti $x_i x_j$ di due variabili distinte (possiamo supporre $i < j$ perché $x_i x_j$ è uguale a $x_j x_i$). Per scrivere una forma quadratica $q(\mathbf{x})$, denotiamo il coefficiente di x_i^2 col simbolo a_{ii} e il coefficiente di $x_i x_j$ col simbolo $2a_{ij}$ (il fattore 2 è comodo per motivi che diverranno chiari tra poco), in modo che:

$$(3.1) \quad q(\mathbf{x}) = \sum_{i=1}^n a_{ii} x_i^2 + \sum_{1 \leq i < j \leq n} 2a_{ij} x_i x_j.$$

La seconda somma ha tanti addendi quanti sono i prodotti $x_i x_j$ di 2 variabili distinte, cioè $\frac{1}{2}n(n - 1)$, e i coefficienti di una forma quadratica in tutto sono

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



Nel caso $n = 2$ di due variabili, la generica forma quadratica ha 3 termini:

$$q(x_1, x_2) = a_{11} x_1^2 + 2a_{12} x_1 x_2 + a_{22} x_2^2$$

La forma quadratica $q_1(x_1, x_2) = x_1^2 - x_2^2$ ha coefficienti $a_{11} = 1$, $a_{22} = -1$, $a_{12} = 0$. La forma quadratica $q_2(x_1, x_2) = x_1 x_2$ ha coefficienti $a_{11} = a_{22} = 0$ e $a_{12} = 1/2$. Il più delle volte scriveremo x e y al posto di x_1 e x_2 , e a , b , e c al posto di a_{11} , a_{12} e a_{22} :

$$q(x, y) = ax^2 + 2bxy + cy^2$$



Nel caso $n = 3$ di tre variabili, la generica forma quadratica ha 6 termini:

$$q(x_1, x_2, x_3) = a_{11} x_1^2 + a_{22} x_2^2 + a_{33} x_3^2 + 2a_{12} x_1 x_2 + 2a_{13} x_1 x_3 + 2a_{23} x_2 x_3.$$

La forma quadratica $q(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$ ha coefficienti $a_{11} = a_{22} = a_{33} = 1$ e $a_{12} = a_{13} = a_{23} = 0$. I coefficienti non nulli della forma quadratica $q(x_1, x_2, x_3) = x_1 x_2 - 6x_2 x_3$ sono $a_{12} = 1/2$ e $a_{23} = 3$.

La norma al quadrato di un vettore di \mathbb{R}^n (rispetto al prodotto scalare standard) è

$$\|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x} = \sum_{i=1}^n x_i^2$$

Si tratta di una forma quadratica. I coefficienti in questo caso sono $a_{ii} = 1$ e $a_{ij} = 0$ se $i < j$.

Possiamo rappresentare la forma quadratica (3.1) mediante una matrice *simmetrica* \mathbf{A} di ordine n nel modo seguente. Per $i \leq j$, gli elementi a_{ij} di \mathbf{A} sono i coefficienti che compaiono nell'equazione (3.1); gli altri elementi sono definiti per simmetria: $a_{ij} = a_{ji}$ se $i > j$. In altri termini, gli elementi a_{ii} sulla diagonale principale di \mathbf{A} sono i coefficienti dei quadrati x_i^2 nella forma quadratica; mentre per $i \neq j$ gli elementi a_{ij} sono i coefficienti di $x_i x_j$ divisi per due. Con questa definizione vale l'uguaglianza fondamentale

$$(3.2) \quad q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} \quad \text{per ogni } \mathbf{x} \in \mathbb{R}^n$$

Infatti

$$\begin{aligned} \mathbf{x}^T \mathbf{A} \mathbf{x} &= \sum_{i=1}^n x_i (\mathbf{A} \mathbf{x})_i = \sum_{i=1}^n x_i \left(\sum_{j=1}^n a_{ij} x_j \right) = \\ &= \sum_{i,j=1}^n a_{ij} x_i x_j = \sum_{i=1}^n a_{ii} x_i^2 + \sum_{1 \leq i < j \leq n} 2a_{ij} x_i x_j \end{aligned}$$

Nell'ultimo passaggio abbiamo separato i termini coi quadrati delle variabili dai termini coi prodotti di due variabili distinte e usato il fatto che per $i < j$

$$a_{ij} x_i x_j + a_{ji} x_j x_i = 2 a_{ij} x_i x_j$$

Si dice che \mathbf{A} rappresenta la forma quadratica $q(\mathbf{x})$ e che $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ è la forma quadratica associata alla matrice simmetrica \mathbf{A} .

TEOREMA 2 (1860)

Nel caso di due variabili, la matrice associata alla forma quadratica $q(x, y) = ax^2 + 2bxy + cy^2$ è

$$\mathbf{A} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

La matrice $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ ha un unico elemento significativo non nullo $b = 1$, e la forma quadratica associata è perciò $q(x, y) = 2xy$. Possiamo ricavare questa espressione con più fatica dalla (3.2):

$$q(x, y) = [x \ y] \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = [x \ y] \begin{bmatrix} y \\ x \end{bmatrix} = xy + yx = 2xy$$

La matrice che rappresenta la forma quadratica $q(x_1, x_2, x_3) = x_1x_2 - 6x_2x_3$ è

$$\begin{bmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & -3 \\ 0 & -3 & 0 \end{bmatrix}$$

La matrice che rappresenta una forma quadratica $q(\mathbf{x})$ è diagonale se e solo se nell'espressione di $q(\mathbf{x})$ compaiono solo i monomi coi quadrati delle variabili:

$$q(\mathbf{x}) = a_{11}x_1^2 + a_{22}x_2^2 + \cdots + a_{nn}x_n^2$$

Per esempio, la matrice che rappresenta $\|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x}$ è la matrice identità.

L'esempio più importante di forma quadratica è il differenziale secondo di una funzione, cioè il termine di secondo grado nello sviluppo di Taylor. Per fissare le idee, supponiamo che $f(\mathbf{x})$ sia una funzione reale di classe C^2 (derivate parziali seconde continue) in un intorno U di $\mathbf{0}$ in \mathbb{R}^n . Il gradiente $\nabla f(\mathbf{0})$ di f in $\mathbf{0}$ è il vettore riga che ha come componenti le derivate parziali $\frac{\partial f}{\partial x_i}(\mathbf{0})$, e la matrice hessiana $\mathbf{H}_f(\mathbf{0})$ di f in $\mathbf{0}$ è la matrice simmetrica che ha come elemento di posto (i, j) la derivata parziale seconda $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{0})$. La funzione $f(\mathbf{x})$ ammette lo sviluppo di Taylor

$$f(\mathbf{x}) = f(\mathbf{0}) + \nabla f(\mathbf{0})\mathbf{x} + \frac{1}{2}\mathbf{x}^T \mathbf{H}_f(\mathbf{0})\mathbf{x} + o(\|\mathbf{x}\|^2)$$

dove il simbolo $o(\|\mathbf{x}\|^2)$ denota un'infinitesimo di ordine superiore a $\|\mathbf{x}\|^2$ per $\mathbf{x} \rightarrow \mathbf{0}$. Il differenziale secondo di $f(\mathbf{x})$ in $\mathbf{0}$ è la forma quadratica $\mathbf{x}^T \mathbf{H}_f(\mathbf{0})\mathbf{x}$ associata alla matrice hessiana.

Supponiamo ora di voler determinare se $\mathbf{0}$ è un punto di massimo (rispettivamente di minimo) locale della funzione, cioè se il segno di $f(\mathbf{x}) - f(\mathbf{0})$ è sempre ≤ 0 (rispettivamente ≥ 0) in un intorno dell'origine. Il termine lineare $\nabla f(\mathbf{0})\mathbf{x}$, se non è nullo, cambia di segno in un intorno dell'origine. Perciò (teorema di Fermat) il gradiente è nullo se $\mathbf{0}$ è un punto di massimo o di minimo. Lo sviluppo di Taylor ci dice allora che l'incremento della funzione

$$f(\mathbf{x}) - f(\mathbf{0}) = \frac{1}{2}\mathbf{x}^T \mathbf{H}_f(\mathbf{0})\mathbf{x} + o(\|\mathbf{x}\|^2)$$

è, a meno di infinitesimi di ordine superiore, il differenziale secondo $\mathbf{x}^T \mathbf{H}_f(\mathbf{0})\mathbf{x}$, e quindi il fatto che $\mathbf{0}$ sia o meno un punto di massimo o di minimo dipende dal segno del differenziale secondo.

Motivati dall'esempio precedente, affrontiamo lo studio del segno di una forma quadratica. Facciamo un'osservazione preliminare:

PROPOSIZIONE 3.1 Sia $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A}\mathbf{x}$ una forma quadratica in n variabili.

Allora:

- a) $q(\mathbf{0}) = 0$;

b) $q(t\mathbf{x}) = t^2 q(\mathbf{x}) \quad \text{per ogni } t \in \mathbb{R}.$

In particolare, fissato un vettore $\mathbf{x}_0 \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, la forma ha segno costante sulla retta $\mathbb{R}(\mathbf{x}_0)$ generata da \mathbf{x}_0 : se q è positiva (rispettivamente negativa) in \mathbf{x}_0 , allora q è positiva (rispettivamente negativa) in ogni punto della retta, con l'eccezione di $\mathbf{0}$, in cui si annulla; se invece $q(\mathbf{x}_0) = 0$, allora q si annulla in ogni punto della retta.

DIMOSTRAZIONE. Questo è evidente, perché $q(\mathbf{x})$ è un polinomio omogeneo di secondo grado nelle variabili x_i .

DEFINIZIONE 3.2 (Segno di una forma quadratica)

Si dice che una forma quadratica $q(\mathbf{x})$ in n variabili è:

- a) *definita positiva* se $q(\mathbf{x}) > 0$ per ogni $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$;
- b) *semidefinita positiva* se $q(\mathbf{x}) \geq 0$ per ogni $\mathbf{x} \in \mathbb{R}^n$ ed esiste $\mathbf{x} \neq \mathbf{0}$ tale che $q(\mathbf{x}) = 0$;
- c) *semidefinita negativa* se $q(\mathbf{x}) \leq 0$ per ogni $\mathbf{x} \in \mathbb{R}^n$ ed esiste $\mathbf{x} \neq \mathbf{0}$ tale che $q(\mathbf{x}) = 0$;
- d) *definita negativa* se $q(\mathbf{x}) < 0$ per ogni $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$;
- e) *indefinita* se esistono \mathbf{x}_1 e \mathbf{x}_2 in \mathbb{R}^n tali che $q(\mathbf{x}_1) < 0 < q(\mathbf{x}_2)$.

Una matrice simmetrica \mathbf{A} si dice (semi)definita positiva o (semi)definita negativa o indefinita a seconda che la forma quadratica $\mathbf{x}^T \mathbf{A} \mathbf{x}$ sia (semi)definita positiva o (semi)definita negativa o indefinita.

OSSERVAZIONE Tranne che per la forma quadratica nulla, che secondo la nostra definizione è sia semidefinita positiva sia semidefinita negativa, le condizioni a) – e) sono mutuamente esclusive ed esauriscono tutte le possibilità. La prima condizione equivale a richiedere che $\mathbf{x} = \mathbf{0}$ sia un punto di minimo (assoluto) forte di $q(\mathbf{x})$; la seconda che sia un punto di minimo, ma non l'unico punto di minimo; la terza che sia un punto di massimo, ma non l'unico; la quarta che sia un punto di minimo forte; l'ultima che non sia né un punto di massimo (locale) né un punto di minimo.

Altri autori preferiscono includere le forme definite tra quelle semidefinite e non richiedono quindi che una forma semidefinita si annulli in un $\mathbf{x} \neq \mathbf{0}$. In questo libro scriviamo che una forma quadratica (o una matrice) è *(semi)definita positiva* se è definita o semidefinita positiva. Molti scrivono $\mathbf{A} > 0$ (rispettivamente $\mathbf{A} \geq 0$) per indicare una matrice definita (rispettivamente (semi)definita) positiva.

Diamo un esempio per ogni possibilità nel caso di due variabili x_1 e x_2 :

- a) la forma quadratica $x_1^2 + x_2^2$ è definita positiva;

- b) la forma quadratica x_1^2 è semidefinita positiva (si annulla in $[0, 1]^T$);
- c) la forma quadratica $-x_1^2$ è semidefinita negativa;
- d) la forma quadratica $-x_1^2 - x_2^2$ è definita negativa;
- e) le forme quadratiche $x_1^2 - x_2^2$ e $x_1 x_2$ sono indefinite.

Prodotti scalari e matrici definite positive

Si ricordi che un prodotto scalare $\langle \mathbf{x}, \mathbf{y} \rangle$ in \mathbb{R}^n è una funzione bilineare simmetrica ($\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$) e definita positiva ($\langle \mathbf{x}, \mathbf{x} \rangle > 0$ se $\mathbf{x} \neq \mathbf{0}$). Possiamo associare a un dato prodotto scalare la matrice \mathbf{B} il cui elemento di posto (i, j) è il valore del prodotto dei vettori \mathbf{e}_i ed \mathbf{e}_j della base canonica:

$$\mathbf{B}_{ij} = \langle \mathbf{e}_i, \mathbf{e}_j \rangle$$

Dalla bilinearità del prodotto scalare segue

$$\begin{aligned} \langle \mathbf{x}, \mathbf{y} \rangle &= \left\langle \sum_{i=1}^n x_i \mathbf{e}_i, \sum_{j=1}^n y_j \mathbf{e}_j \right\rangle = \sum_{i=1}^n \sum_{j=1}^n x_i y_j \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \\ &= \sum_{i=1}^n x_i \left(\sum_{j=1}^n \mathbf{B}_{ij} y_j \right) = \sum_{i=1}^n x_i (\mathbf{By})_i = \mathbf{x}^T \mathbf{By} \end{aligned}$$

Quindi $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{By}$ per ogni $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ e la matrice \mathbf{B} determina così il prodotto scalare; si dice che \mathbf{B} rappresenta il prodotto scalare. Dalla commutatività e dalla positività del prodotto scalare segue che \mathbf{B} è una matrice simmetrica definita positiva.

Viceversa, supponiamo che \mathbf{B} sia una matrice simmetrica definita positiva. Definiamo

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{By} \quad \text{per ogni } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

È semplice verificare che la funzione $\langle \mathbf{x}, \mathbf{y} \rangle$ è un prodotto scalare in \mathbb{R}^n . Infatti $\mathbf{x}^T \mathbf{By}$ è lineare in \mathbf{x} , il prodotto è commutativo perché $\mathbf{x}^T \mathbf{By}$ è uno scalare e \mathbf{B} è simmetrica:

$$\mathbf{x}^T \mathbf{By} = (\mathbf{x}^T \mathbf{By})^T = \mathbf{y}^T \mathbf{B}^T \mathbf{x} = \mathbf{y}^T \mathbf{Bx}$$

infine la proprietà di positività è verificata perché \mathbf{B} è definita positiva:

$$\langle \mathbf{x}, \mathbf{x} \rangle = \mathbf{x}^T \mathbf{Bx} > 0 \quad \text{per ogni } \mathbf{x} \neq \mathbf{0}$$

In conclusione, i prodotti scalari in \mathbb{R}^n sono in corrispondenza biunivoca con le matrici simmetriche definite positive. Il prodotto scalare standard $\mathbf{x}^T \mathbf{y}$ corrisponde alla matrice identità. Si osservi che, se $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{By}$, la forma quadratica $q(\mathbf{x}) = \mathbf{x}^T \mathbf{Bx}$ associata a \mathbf{B} non è altro che la norma al quadrato.

Per studiare il segno di un forma quadratica $q(\mathbf{x})$, la cosa naturale da fare è cercare delle nuove coordinate \mathbf{X} rispetto alle quali l'espressione di $q(\mathbf{x})$ contenga solo i termini coi quadrati delle variabili:

$$(3.3) \quad q(\mathbf{x}) = d_1 X_1^2 + \cdots + d_n X_n^2$$

Diventa così immediato studiare il segno di $q(\mathbf{x})$. Per esempio $q(\mathbf{x})$ è definita positiva se e solo se tutti i coefficienti d_k sono positivi. Si dice che un tale cambiamento di

variabili *diagonalizza* la forma quadratica perché la matrice associata a $q(\mathbf{x})$ rispetto alle variabili \mathbf{X} è la matrice diagonale $\text{diag}(d_1, \dots, d_n)$.

Per procedere sistematicamente, cominciamo col determinare come cambia la matrice di una forma quadratica al variare delle coordinate. Supponiamo che \mathbf{A} sia una matrice simmetrica di ordine n e sia $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ la forma quadratica rappresentata da \mathbf{A} . Fissiamo una nuova base in \mathbb{R}^n e sia \mathbf{S} la matrice invertibile che ha come colonne i vettori della nuova base. Sia \mathbf{X} il vettore delle coordinate di \mathbf{x} rispetto alla nuova base, in maniera tale che $\mathbf{x} = \mathbf{S}\mathbf{X}$. Allora $\mathbf{S}^T \mathbf{A} \mathbf{S}$ è simmetrica e rappresenta la forma quadratica q nelle nuove coordinate \mathbf{X} perché

$$(3.4) \quad q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{X}^T \mathbf{S}^T \mathbf{A} \mathbf{S} \mathbf{X}$$

DEFINIZIONE 3.3 (Matrici congruenti)

Una matrice \mathbf{B} si dice *congruente* a una matrice simmetrica \mathbf{A} se esiste una matrice invertibile \mathbf{S} tale che $\mathbf{B} = \mathbf{S}^T \mathbf{A} \mathbf{S}$.

OSSERVAZIONE Se \mathbf{B} è congruente alla matrice simmetrica \mathbf{A} , allora \mathbf{B} è anch'essa simmetrica. La relazione di congruenza è una relazione di equivalenza sull'insieme delle matrici simmetriche; in particolare, \mathbf{B} è congruente ad \mathbf{A} se e solo se \mathbf{A} è congruente a \mathbf{B} e possiamo perciò dire semplicemente che \mathbf{A} e \mathbf{B} sono congruenti. Per la formula (3.4) due matrici sono congruenti se rappresentano la stessa forma quadratica rispetto a coordinate distinte.

OSSERVAZIONE La forma quadratica $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ si può diagonalizzare, cioè scrivere come in (3.3), se e solo se \mathbf{A} è congruente a una matrice diagonale. Vedremo che ogni forma quadratica si può diagonalizzare, anzi ci sono svariati modi di diagonalizzare una forma quadratica. Ci sono due metodi di diagonalizzare una forma quadratica che sono particolarmente importanti e che illustriamo in questo paragrafo. Il primo metodo utilizza il teorema spettrale, per il quale data una matrice simmetrica \mathbf{A} esiste una matrice ortogonale \mathbf{Q} tale che $\mathbf{Q}^T \mathbf{A} \mathbf{Q} = \text{diag}(\lambda_1, \dots, \lambda_n)$, dove gli elementi λ_k sono gli autovalori di \mathbf{A} . Questo significa che \mathbf{A} è congruente alla matrice diagonale che ha gli autovalori di \mathbf{A} sulla diagonale principale. Il secondo metodo si basa sulla fattorizzazione $\mathbf{A} = \mathbf{L} \mathbf{D} \mathbf{L}^T$ fornita dall'algoritmo di Gauss: questa fattorizzazione non è possibile per tutte le matrici simmetriche, ma esiste per quelle (semi)definite ed è nota come *fattorizzazione di Cholesky*. In questa fattorizzazione \mathbf{L} è una matrice triangolare bassa con gli elementi sulla diagonale principale uguali a 1 e \mathbf{D} è la matrice diagonale che ha i pivots di \mathbf{A} sulla diagonale principale; quindi \mathbf{A} è congruente alla matrice diagonale che ha i pivots di \mathbf{A} sulla diagonale principale.

OSSERVAZIONE Il problema della diagonalizzazione di una forma quadratica non va confuso col problema della diagonalizzazione di una matrice. Diagonalizzare una forma quadratica significa trovare una matrice diagonale *congruente* alla matrice rappresentativa della forma quadratica, mentre diagonalizzare una matrice significa trovare una matrice diagonale *simile* alla matrice data. La relazione di congruenza $\mathbf{B} = \mathbf{S}^T \mathbf{A} \mathbf{S}$ non va confusa con la relazione di similitudine $\mathbf{B} = \mathbf{S}^{-1} \mathbf{A} \mathbf{S}$, che identifica due matrici

che rappresentano la stessa applicazione lineare di \mathbb{R}^n in se stesso. Due matrici congruenti in generale non hanno gli stessi autovalori. Al contrario di quanto avviene per la similitudine, è piuttosto semplice stabilire se due matrici sono congruenti, come vedremo più avanti con la *legge di inerzia di Sylvester*.

Se \mathbf{Q} però è una matrice ortogonale, allora $\mathbf{Q}^T = \mathbf{Q}^{-1}$ e quindi la matrice

$$\mathbf{B} = \mathbf{Q}^T \mathbf{A} \mathbf{Q} = \mathbf{Q}^{-1} \mathbf{A} \mathbf{Q}$$

è al tempo stesso congruente e simile alla matrice \mathbf{A} . Il teorema spettrale mostra che una matrice simmetrica \mathbf{A} è al tempo stesso congruente e simile alla matrice diagonale che ha gli autovalori di \mathbf{A} sulla diagonale principale.

OSSERVAZIONE Supponiamo che \mathbf{A} e \mathbf{B} siano matrici congruenti. Allora \mathbf{A} è definita positiva se e solo se \mathbf{B} è definita positiva. Infatti

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{X}^T \mathbf{B} \mathbf{X} \quad (\text{dove } \mathbf{X} = \mathbf{S}^{-1} \mathbf{x})$$

per cui $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ per ogni \mathbf{x} non nullo equivale a $\mathbf{X}^T \mathbf{B} \mathbf{X} > 0$ per ogni \mathbf{X} non nullo. Naturalmente lo stesso vale per la proprietà di essere (semi)definita positiva, indefinita o (semi)definita negativa.

Segno di una forma quadratica e autovalori

Sia \mathbf{A} la matrice che rappresenta la forma quadratica $q(\mathbf{x})$. Per il teorema spettrale esiste una matrice ortogonale \mathbf{Q} tale che $\mathbf{Q}^T \mathbf{A} \mathbf{Q} = \text{diag}(\lambda_1, \dots, \lambda_n)$. Il cambiamento di variabili $\mathbf{X} = \mathbf{Q}^T \mathbf{x}$ diagonalizza la forma quadratica $q(\mathbf{x})$ e il segno della forma quadratica è determinato dagli autovalori della matrice. Più precisamente, siccome \mathbf{Q} è ortogonale e quindi il cambiamento di variabili preserva la norma, vale il seguente risultato che confronta i valori di una forma quadratica arbitraria con la norma euclidea al quadrato $\|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x}$.

PROPOSIZIONE 3.4 Sia \mathbf{A} una matrice simmetrica di ordine n sia $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ e sia $\|\mathbf{x}\|$ la norma euclidea di un vettore di \mathbb{R}^n .

a) Se $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, allora $q(\mathbf{v}) = \lambda \|\mathbf{v}\|^2$.

b) Se λ_{\min} e λ_{\max} sono il minimo e il massimo degli autovalori di \mathbf{A} , allora:

$$(3.5) \quad \lambda_{\min} \|\mathbf{x}\|^2 \leq q(\mathbf{x}) \leq \lambda_{\max} \|\mathbf{x}\|^2$$

DIMOSTRAZIONE. Il punto a) è immediato: se $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, allora

$$q(\mathbf{v}) = \mathbf{v}^T \mathbf{A} \mathbf{v} = \mathbf{v}^T (\lambda\mathbf{v}) = \lambda \|\mathbf{v}\|^2$$

Per il teorema spettrale possiamo decomporre la matrice nella forma $\mathbf{A} = \mathbf{Q} \mathbf{D} \mathbf{Q}^T$, dove \mathbf{Q} è una matrice ortogonale e \mathbf{D} è la matrice diagonale che ha sulla diagonale principale gli autovalori λ_k di \mathbf{A} . La forma quadratica si scrive

$$q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{x}^T \mathbf{Q} \mathbf{D} \mathbf{Q}^T \mathbf{x}$$

Posto $\mathbf{X} = \mathbf{Q}^T \mathbf{x}$, tenendo conto del fatto che $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_n)$ otteniamo

$$q(\mathbf{x}) = \mathbf{X}^T \mathbf{D} \mathbf{X} = \lambda_1 X_1^2 + \lambda_2 X_2^2 + \cdots + \lambda_n X_n^2$$

Ora, siccome $\lambda_i \leq \lambda_{\max}$ per ogni i proprio per definizione di λ_{\max} , concludiamo che

$$q(\mathbf{x}) = \lambda_1 X_1^2 + \lambda_2 X_2^2 + \cdots + \lambda_n X_n^2 \leq \lambda_{\max} X_1^2 + \lambda_{\max} X_2^2 + \cdots + \lambda_{\max} X_n^2 = \lambda_{\max} \|\mathbf{X}\|^2$$

Ma \mathbf{Q} è ortogonale, quindi il cambiamento di variabili $\mathbf{X} = \mathbf{Q}^T \mathbf{x}$ che abbiamo fatto è un'isometria:

$$\|\mathbf{X}\|^2 = \|\mathbf{Q}^T \mathbf{x}\|^2 = \mathbf{x}^T \mathbf{Q} \mathbf{Q}^T \mathbf{x} = \mathbf{x}^T \mathbf{x} = \|\mathbf{x}\|^2$$

Sostituendo nella precedente diseguaglianza troviamo

$$q(\mathbf{x}) \leq \lambda_{\max} \|\mathbf{x}\|^2$$

Questo dimostra la diseguaglianza di destra nella (3.5); la dimostrazione della diseguaglianza di sinistra è perfettamente analoga.

COROLLARIO 3.5 (Segno di una forma quadratica e autovalori)

Sia $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ la forma quadratica in n variabili associata alla matrice simmetrica \mathbf{A} . Allora:

- a) $q(\mathbf{x})$ è definita positiva (rispettivamente negativa) se e solo se tutti gli autovalori di \mathbf{A} sono strettamente maggiori (rispettivamente minori) di zero;
- b) $q(\mathbf{x})$ è semidefinita positiva (rispettivamente negativa) se e solo se gli autovalori di \mathbf{A} sono maggiori (rispettivamente minori) o uguali a zero, e almeno uno di essi è nullo;
- c) $q(\mathbf{x})$ è indefinita se e solo se \mathbf{A} ha almeno un autovalore positivo e almeno un autovalore negativo.

DIMOSTRAZIONE. Il corollario segue immediatamente dalla proposizione 3.4. Mostriamo per esempio che $q(\mathbf{x})$ è definita positiva se e solo se gli autovalori sono > 0 . Se gli autovalori sono strettamente maggiori di zero, allora per ogni $\mathbf{x} \in \mathbb{R}^n$ diverso da zero

$$q(\mathbf{x}) \geq \lambda_{\min} \|\mathbf{x}\|^2 > 0$$

perché λ_{\min} è un autovalore e quindi, per ipotesi, positivo. Questo mostra che $q(\mathbf{x})$ è definita positiva. Viceversa, se $q(\mathbf{x})$ è definita positiva, sia \mathbf{v} un autovettore relativo all'autovalore λ . Possiamo supporre che \mathbf{v} sia un versore. Allora per il punto a) della proposizione 3.4

$$\lambda = q(\mathbf{v}) > 0$$

perché $q(\mathbf{x})$ è definita positiva e $\mathbf{v} \neq \mathbf{0}$. Quindi ogni autovalore λ di \mathbf{A} è positivo.

Il corollario 3.5 in pratica è di difficile applicazione perché è computazionalmente costoso calcolare gli autovalori (o anche solo il polinomio caratteristico). Per questo è importante trovare criteri alternativi per determinare il segno di una forma quadratica. Nel caso di due variabili la situazione è molto semplice:

PROPOSIZIONE 3.6 (Segno di una forma quadratica in due variabili)

La matrice simmetrica $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$ è

- a) definita positiva se $a > 0$ e $ac - b^2 > 0$;
- b) semidefinita positiva se $a \geq 0$ e $ac - b^2 = 0$;
- c) semidefinita negativa se $a \leq 0$ e $ac - b^2 = 0$;
- d) definita negativa se $a < 0$ e $ac - b^2 > 0$;
- e) indefinita se $ac - b^2 < 0$.

DIMOSTRAZIONE. La matrice ha due autovalori reali λ_1 e λ_2 il cui prodotto è il determinante $ac - b^2$ e la cui somma è la traccia $a + c$. Se $ac - b^2 = \lambda_1\lambda_2 < 0$, i due autovalori hanno segno opposto e quindi la matrice è indefinita.

Supponiamo ora che il determinante $ac - b^2$ sia positivo: allora i due autovalori sono concordi e il loro segno coincide col segno della loro somma $a + c$; siccome $ac - b^2 > 0$, a e c sono concordi, e quindi il segno degli autovalori coincide col segno di a . Quindi, se $a > 0$, i due autovalori sono positivi e la matrice è definita positiva, se invece $a < 0$ la matrice è definita negativa.

Infine, supponiamo che il determinante sia nullo. Allora uno degli autovalori è nullo e l'altro è positivo, negativo o nullo a seconda che a sia positivo, negativo, o nullo. Quindi, se $a \geq 0$, la matrice è semidefinita positiva se $a \leq 0$, la matrice è semidefinita negativa.

Quoziente di Rayleigh. Massimo e minimo di $q(\mathbf{x})$ sulla sfera unitaria.

Il rapporto

$$\frac{q(\mathbf{x})}{\|\mathbf{x}\|^2} = \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|^2}$$

si dice *quoziente di Rayleigh*: esso confronta la forma quadratica $q(\mathbf{x})$ con la forma quadratica $\|\mathbf{x}\|^2$. L'enunciato della proposizione 3.4 si può riformulare così:

l'autovalore minimo λ_{\min} di \mathbf{A} è il minimo del quoziente di Rayleigh al variare di \mathbf{x} nel suo insieme di definizione $\mathbb{R}^n \setminus \{\mathbf{0}\}$:

$$(3.6) \quad \lambda_{\min} = \text{Min} \left\{ \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|^2} : \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\} \right\}$$

infatti, se dividiamo la diseguaglianza di sinistra nella (3.5) per $\|\mathbf{x}\|^2$, otteniamo:

$$\lambda_{\min} \leq \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|^2}$$

Questo mostra che λ_{\min} è un estremo inferiore per i valori del quoziente di Rayleigh. D'altra parte, se \mathbf{v} è un autovettore relativo a λ_{\min} , allora per la proposizione 3.4.a

$$\lambda_{\min} = \frac{\mathbf{v}^T \mathbf{A} \mathbf{v}}{\|\mathbf{v}\|^2}$$

quindi λ_{\min} è il valore minimo del quoziente di Rayleigh. Analogamente si mostra che

$$(3.7) \quad \lambda_{\max} = \operatorname{Max} \left\{ \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|^2} : \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\} \right\}$$

è il valore massimo del quoziente di Rayleigh.

Possiamo restringere queste considerazioni ai versori di \mathbb{R}^n , cioè ai punti della *sfera unitaria*

$$S^{n-1} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = 1\}$$

Se \mathbf{x} è un versore, il quoziente di Rayleigh in \mathbf{x} coincide con la forma quadratica:

$$\frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|^2} = \mathbf{x}^T \mathbf{A} \mathbf{x}$$

Si vede così che λ_{\min} e λ_{\max} sono il minimo valore e il massimo valore assunti dalla forma quadratica sulla sfera S^{n-1} :

$$\lambda_{\min} = \operatorname{Min} \{ \mathbf{x}^T \mathbf{A} \mathbf{x} : \mathbf{x} \in S^{n-1} \}$$

$$\lambda_{\max} = \operatorname{Max} \{ \mathbf{x}^T \mathbf{A} \mathbf{x} : \mathbf{x} \in S^{n-1} \}$$

Questa formula ha un risvolto applicativo importante. In generale, è computazionalmente costoso calcolare gli autovalori come radici del polinomio caratteristico. Nel caso delle matrici simmetriche, possiamo determinare λ_{\min} e λ_{\max} trovando il minimo e il massimo della forma quadratica $\mathbf{x}^T \mathbf{A} \mathbf{x}$ sulla sfera; questo metodo si può estendere anche alla ricerca degli altri autovalori di una matrice simmetrica e va sotto il nome di *minimax principle*.

Dal punto di vista teorico, queste osservazioni conducono a una nuova dimostrazione del teorema spettrale che non fa uso dei numeri complessi. Data una matrice simmetrica \mathbf{A} , la funzione $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ è continua sulla sfera S^{n-1} . Per il teorema di Weierstrass, dato che S^{n-1} è chiuso e limitato, la funzione $q(\mathbf{x})$ assume un valore minimo λ_m su S^{n-1} . Sia $\mathbf{v} \in S^{n-1}$ un punto in cui la funzione assume il suo valore minimo: $q(\mathbf{v}) = \lambda_m$. Per il teorema di Fermat sugli estremi vincolati, il gradiente di $q(\mathbf{x})$ in \mathbf{v} dev'essere perpendicolare al vincolo S^{n-1} , cioè dev'essere diretto come il raggio \mathbf{v} : deve quindi esistere $t \in \mathbb{R}$ tale che $\nabla q(\mathbf{v}) = t\mathbf{v}$. D'altra parte, un semplice conto mostra che

$$\nabla q(\mathbf{v}) = \mathbf{A} \mathbf{v}$$

Quindi $\mathbf{A} \mathbf{v} = t\mathbf{v}$; si noti che $t = \lambda_m$ perché da $\mathbf{A} \mathbf{v} = t\mathbf{v}$ segue $q(\mathbf{v}) = t$. Questo mostra che λ_m è un autovalore di \mathbf{A} relativo all'autovettore \mathbf{v} . La dimostrazione ora si conclude esattamente come nella dimostrazione del teorema 2.4. La differenza tra le due dimostrazioni è che per mostrare l'esistenza di un autovalore reale abbiamo usato qui il teorema di Weierstrass sull'esistenza dei massimi e minimi, mentre nel paragrafo precedente abbiamo usato il teorema fondamentale dell'algebra sull'esistenza delle radici (complesse) di un polinomio.

Matrici definite positive e minori di nord-ovest

Nel caso di più di due variabili, la proposizione 3.6 si generalizza in un criterio che consente di riconoscere le forme definite positive, e che conduce alla *fattorizzazione*

di Cholesky. Per questo occorre introdurre le *sottomatrici principali di nord-ovest* $\mathbf{A}_{(k)}$ di una matrice \mathbf{A} di ordine n . Per definizione $\mathbf{A}_{(k)}$ è la matrice $k \times k$ ottenuta cancellando le ultime $n-k$ righe e le ultime $n-k$ colonne di \mathbf{A} (per ogni $k = 1, 2, \dots, n$). I determinanti $\delta_k = \det(\mathbf{A}_{(k)})$ si dicono *minori principali di nord-ovest* di \mathbf{A} . Nel caso della matrice $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$ i minori principali di nord-ovest sono $\delta_1 = a$ e $\delta_2 = ac - b^2$; la proposizione 3.6 mostra che il segno di una forma quadratica in due variabili è determinato dal segno dei minori principali di nord-ovest. Nel caso $n \geq 3$ i minori principali di nord-ovest sono sufficienti a determinare se la matrice è definita positiva:

TEOREMA 3.7 (Matrici definite positive)

Per una matrice simmetrica \mathbf{A} le seguenti condizioni sono equivalenti:

- a) \mathbf{A} è definita positiva;
- b) i minori principali di nord-ovest di \mathbf{A} sono positivi;
- c) $\mathbf{A} = \mathbf{LDL}^T$ dove \mathbf{L} è triangolare bassa con elementi sulla diagonale principale uguali a 1 e $\mathbf{D} = \text{diag}(p_1, \dots, p_n)$ con $p_k > 0$ per ogni k .

DIMOSTRAZIONE. Supponiamo che \mathbf{A} sia definita positiva. Dato un arbitrario vettore non nullo \mathbf{y} di \mathbb{R}^k , sia \mathbf{x} il vettore di \mathbb{R}^n che ha le prime k componenti uguali a quelle di \mathbf{y} e le ultime $n-k$ componenti nulle. Allora

$$\mathbf{y}^T \mathbf{A}_{(k)} \mathbf{y} = \mathbf{x}^T \mathbf{A} \mathbf{x} > 0$$

perché \mathbf{A} è definita positiva. Questo mostra che anche $\mathbf{A}_{(k)}$ è definita positiva. Quindi i suoi autovalori sono positivi, e il suo determinante δ_k è positivo perché è il prodotto degli autovalori. Questo mostra che a) implica b).

Supponiamo che tutti i minori principali di nord-ovest siano positivi. In particolare $a_{11} = \delta_1 > 0$. Possiamo isolare i termini in $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ che contengono x_1 , e completare il quadrato:

$$\begin{aligned} q(\mathbf{x}) &= a_{11}x_1^2 + 2a_{12}x_1x_2 + \dots + 2a_{1n}x_1x_n + q_1(x_2, \dots, x_n) = \\ &= a_{11} \left(x_1^2 + 2\frac{a_{12}}{a_{11}}x_1x_2 + \dots + 2\frac{a_{1n}}{a_{11}}x_1x_n \right) + q_1(x_2, \dots, x_n) = \\ &= a_{11} \left(x_1 + \frac{a_{12}}{a_{11}}x_2 + \dots + \frac{a_{1n}}{a_{11}}x_n \right)^2 - \frac{a_{12}^2}{a_{11}}x_2^2 - \dots - \frac{a_{1n}^2}{a_{11}}x_n^2 + q_1(x_2, \dots, x_n) = \\ &= a_{11} \left(x_1 + \frac{a_{12}}{a_{11}}x_2 + \dots + \frac{a_{1n}}{a_{11}}x_n \right)^2 + q_2(x_2, \dots, x_n) \end{aligned}$$

Posto $X_1 = x_1 + \frac{a_{12}}{a_{11}}x_2 + \dots + \frac{a_{1n}}{a_{11}}x_n$ e $X_k = x_k$ per $k > 1$, nelle nuove variabili la forma quadratica è

$$a_{11}X_1^2 + q_2(X_2, \dots, X_n)$$

Conviene riscrivere il tutto in termini di matrici. Per questo scriviamo la matrice \mathbf{A} e la matrice \mathbf{R} del cambiamento di variabili $\mathbf{X} = \mathbf{Rx}$ in forma a blocchi:

$$\mathbf{A} = \left[\begin{array}{c|c} a_{11} & \mathbf{b}^T \\ \hline \mathbf{b} & \mathbf{C} \end{array} \right], \quad \mathbf{R} = \left[\begin{array}{c|c} 1 & \frac{1}{a_{11}}\mathbf{b}^T \\ \hline \mathbf{0} & \mathbf{I}_{n-1} \end{array} \right]$$

Si noti che \mathbf{C} è simmetrica. L'effetto del cambiamento di variabili $\mathbf{X} = \mathbf{Rx}$ sulla forma quadratica si traduce nella fattorizzazione

$$\mathbf{A} = \left[\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \frac{1}{a_{11}}\mathbf{b} & \mathbf{I}_{n-1} \end{array} \right] \left[\begin{array}{c|c} a_{11} & \mathbf{0}^T \\ \hline \mathbf{0} & \mathbf{C} - \frac{1}{a_{11}}\mathbf{bb}^T \end{array} \right] \left[\begin{array}{c|c} 1 & \frac{1}{a_{11}}\mathbf{b}^T \\ \hline \mathbf{0} & \mathbf{I}_{n-1} \end{array} \right] = \mathbf{R}^T \mathbf{BR}$$

La matrice \mathbf{B} è

$$\mathbf{B} = \left[\begin{array}{c|c} a_{11} & \mathbf{0}^T \\ \hline \mathbf{0} & \mathbf{A}_1 \end{array} \right]$$

dove \mathbf{A}_1 denota la matrice simmetrica $\mathbf{C} - \frac{1}{a_{11}}\mathbf{bb}^T$. Procediamo per induzione sull'ordine di \mathbf{A} . Per questo occorre mostrare che i minori di nord-ovest $\delta_k(\mathbf{A}_1)$ della matrice \mathbf{A}_1 sono positivi; osserviamo che la decomposizione $\mathbf{A} = \mathbf{R}^T(\mathbf{BR})$ corrisponde alle prime $n-1$ operazioni del MEG: la matrice

$$\mathbf{BR} = \left[\begin{array}{c|c} a_{11} & \mathbf{b}^T \\ \hline \mathbf{0} & \mathbf{C} - \frac{1}{a_{11}}\mathbf{bb}^T \end{array} \right]$$

è ottenuta da \mathbf{A} sommando a ciascuna delle righe successive alla prima un opportuno multiplo della prima. Questa operazione non cambia i minori principali di nord-ovest. Quindi i minori $\delta_k(\mathbf{BR})$ sono positivi. Lo sviluppo di Laplace lungo la prima colonna mostra $\delta_k(\mathbf{B}) = \delta_k(\mathbf{BR}) > 0$. Infine il minore di nord-ovest $\delta_k(\mathbf{A}_1)$ di \mathbf{A}_1 è uguale al rapporto $\delta_{k+1}(\mathbf{B})/a_{11}$, e quindi è anch'esso positivo. Questo mostra che i minori di nord ovest di \mathbf{A}_1 sono positivi. Procedendo per induzione possiamo supporre che $\mathbf{A}_1 = \mathbf{L}_1 \mathbf{D}_1 \mathbf{L}_1^T$, dove \mathbf{L}_1 è triangolare bassa con elementi sulla diagonale principale uguali a 1, e $\mathbf{D}_1 = \text{diag}(p_2, \dots, p_n)$ con $p_k > 0$ per ogni k . La matrice

$$\mathbf{L} = \mathbf{R}^T \left[\begin{array}{c|c} 1 & \mathbf{0}^T \\ \hline \mathbf{0} & \mathbf{L}_1 \end{array} \right] = \left[\begin{array}{c|c} 1 & \mathbf{0}^T \\ \hline \frac{1}{a_{11}}\mathbf{b} & \mathbf{L}_1 \end{array} \right]$$

è allora triangolare bassa con gli elementi sulla diagonale principale uguali a 1 e $\mathbf{A} = \mathbf{LDL}^T$ dove $\mathbf{D} = \text{diag}(a_{11}, p_2, \dots, p_n)$ ha tutti gli elementi sulla diagonale principale positivi. Quindi b) implica c) (Si osservi che si tratta della fattorizzazione \mathbf{LU} di \mathbf{A} , con $\mathbf{U} = \mathbf{DL}^T$. In particolare, a_{11}, p_2, \dots, p_n sono i pivots di \mathbf{A}).

Infine supponiamo che $\mathbf{A} = \mathbf{LDL}^T$ con $\mathbf{D} = \text{diag}(p_1, \dots, p_n)$ e $p_k > 0$ per ogni k . Allora \mathbf{D} è definita positiva perché ha tutti gli autovalori positivi, e quindi anche \mathbf{A} , che è congruente a \mathbf{D} , è definita positiva. Questo mostra che c) implica a) e conclude la dimostrazione. \blacksquare

OSSERVAZIONE La positività dei minori di nord-ovest è più veloce da verificare della positività degli autovalori. Per verificare che una matrice simmetrica $\mathbf{A} = [a_{ij}]$ di ordine 3 sia definita positiva basta verificare che $\delta_1 = a_{11}$, $\delta_2 = a_{11}a_{22} - a_{12}^2$ e $\delta_3 = \det(\mathbf{A})$ siano positivi.

OSSERVAZIONE Una matrice \mathbf{A} è definita negativa se e solo se $-\mathbf{A}$ è definita positiva. Dal teorema 3.7 si deduce quindi che i segni dei minori di nord-ovest δ_k di una matrice definita negativa sono negativi se k è dispari e positivi se k è pari. Inoltre una matrice definita negativa ammette la fattorizzazione $\mathbf{A} = \mathbf{LDL}^T$ dove \mathbf{D} è diagonale con gli elementi sulla diagonale principale negativi.

Fattorizzazione di Cholesky

Nel paragrafo sulla fattorizzazione **LU** del capitolo sulle matrici abbiamo definito una matrice definita positiva come una matrice simmetrica che ammette una fattorizzazione \mathbf{LDL}^T con $\mathbf{D} = \text{diag}(p_1, \dots, p_n)$ e $p_k > 0$. Per il teorema 3.7 quella definizione è equivalente a quella data in questo capitolo. La fattorizzazione \mathbf{LDL}^T si riscrive di solito in questo modo: si pone $\mathbf{C} = \mathbf{L}\sqrt{\mathbf{D}}$ dove $\sqrt{\mathbf{D}} = \text{diag}(\sqrt{p_1}, \dots, \sqrt{p_n})$. Allora \mathbf{C} è una matrice triangolare bassa con gli elementi sulla diagonale principale positivi, e

$$(3.8) \quad \mathbf{A} = \mathbf{CC}^T$$

La (3.8) è nota come *fattorizzazione di Cholesky* della matrice definita positiva \mathbf{A} .

Ci sono altri modi di dedurre l'esistenza della fattorizzazione di Choleski, e vogliamo citarne qui uno. Innanzitutto ogni matrice (semi)definita positiva si può scrivere nella forma $\mathbf{B}^T\mathbf{B}$ e in molti modi. Per vedere questo, osserviamo che per il teorema spettrale una matrice semidefinita positiva \mathbf{A} ha una radice quadrata (semi)definita positiva $\sqrt{\mathbf{A}}$: se $\mathbf{A} = \sum_i \lambda_i \mathbf{P}_i$ è la decomposizione spettrale di \mathbf{A} , gli autovalori λ_i sono maggiori o uguali a zero, e $\sqrt{\mathbf{A}} = \sum_i \sqrt{\lambda_i} \mathbf{P}_i$ è l'unica radice quadrata (semi)definita positiva di \mathbf{A} . Posto $\mathbf{E} = \sqrt{\mathbf{A}}$, troviamo $\mathbf{A} = \mathbf{E}^2 = \mathbf{E}^T\mathbf{E}$ perché \mathbf{E} è simmetrica. Possiamo ora trovare tante decomposizioni del tipo $\mathbf{A} = \mathbf{B}^T\mathbf{B}$: se \mathbf{Q} è una matrice ortogonale, da $\mathbf{A} = \mathbf{E}^2$ con \mathbf{E} simmetrica segue

$$\mathbf{A} = \mathbf{E}^T\mathbf{E} = \mathbf{E}^T\mathbf{Q}^T\mathbf{QE} = (\mathbf{QE})^T(\mathbf{QE})$$

Nel caso di una matrice definita positiva \mathbf{A} si può così ottenere la fattorizzazione di Cholesky: siccome tutti gli autovalori di \mathbf{A} sono positivi, anche quelli di \mathbf{E} sono positivi; in particolare \mathbf{E} è invertibile, e può essere decomposta come prodotto $\mathbf{E} = \mathbf{QR}$ con \mathbf{Q} ortogonale e \mathbf{R} triangolare alta con tutti gli elementi sulla diagonale principale positivi. Quindi

$$\mathbf{A} = \mathbf{E}^T\mathbf{E} = (\mathbf{QR})^T(\mathbf{QR}) = \mathbf{R}^T\mathbf{R} = \mathbf{CC}^T$$

dove $\mathbf{C} = \mathbf{R}^T$.

Matrici semidefinite positive e minori principali

Con un ulteriore sforzo si può estendere il teorema alle matrici (semi)definite positive. Il principale guadagno è di ottenere una fattorizzazione di Cholesky anche per le matrici semidefinite positive. Il principale svantaggio è che non basta più considerare i minori di nord-ovest, ma occorre prendere in considerazione tutti i minori principali. Una *sottomatrice principale* di una matrice quadrata \mathbf{A} è una sottomatrice quadrata \mathbf{B} di \mathbf{A} che è simmetrica rispetto alla diagonale principale; questo significa che esistono $1 \leq h_1 < h_2 < \dots < h_k \leq n$ tali che gli elementi di \mathbf{B} sono gli elementi di posto (i, j) con $i, j \in \{h_1, \dots, h_k\}$ (le sottomatrici principali di nord-ovest sono quelle per cui $\{h_1, \dots, h_k\} = \{1, \dots, k\}$). Un *minore principale* di \mathbf{A} è il determinante di una sottomatrice principale. Per esempio, i minori principali di ordine 1 di \mathbf{A} sono gli elementi a_{hh} della diagonale principale.

TEOREMA 3.8 Matrici (semi)definite positive

Per una matrice simmetrica \mathbf{A} le seguenti condizioni sono equivalenti:

- a) \mathbf{A} è (semi)definita positiva;
- b) i minori principali di \mathbf{A} sono maggiori o uguali a zero;
- c) $\mathbf{A} = \mathbf{LDL}^T$ dove \mathbf{L} è triangolare bassa con elementi sulla diagonale principale uguali a 1, e $\mathbf{D} = \text{diag}(p_1, \dots, p_n)$ con $p_k \geq 0$ per ogni k .

DIMOSTRAZIONE. La dimostrazione è analoga a quella del teorema 3.7, con una difficoltà in più che ora illustriamo. Il problema è nel mostrare che, se tutti i minori principali sono non negativi, allora \mathbf{A} ammette la fattorizzazione $\mathbf{A} = \mathbf{LDL}^T$ del punto c). Se $a_{11} > 0$ la dimostrazione procede sostanzialmente come nel teorema 3.7. Se $a_{11} = 0$, consideriamo per ogni $j \geq 2$, il minore principale di ordine 2

$$\begin{vmatrix} a_{11} & a_{1j} \\ a_{1j} & a_{jj} \end{vmatrix} = -a_{1j}^2$$

Siccome i minori principali sono non negativi, questo implica $a_{1j} = 0$ per ogni $j \geq 1$. La matrice \mathbf{A} ha quindi la forma

$$\mathbf{A} = \left[\begin{array}{c|c} 0 & \mathbf{0}^T \\ \hline \mathbf{0} & \mathbf{A}_1 \end{array} \right]$$

e di nuovo si può procedere per induzione. Si noti che la fattorizzazione $\mathbf{LDL}^T = \mathbf{LU}$ che si ottiene così non coincide col MEG perché la matrice \mathbf{U} ha la prima riga nulla; il MEG, almeno per come l'abbiamo definito in questo libro, imporrebbbe a questo punto uno scambio di righe.

Analogamente, una matrice simmetrica è (semi)definita negativa se e solo se i suoi minori principali di ordine k sono nulli o hanno segno $(-1)^k$ e in tal caso la matrice ammette la fattorizzazione $\mathbf{A} = \mathbf{LDL}^T$ con \mathbf{D} diagonale a elementi ≤ 0 .

Non è invece vero che una matrice simmetrica *indefinita* abbia sempre una fattorizzazione \mathbf{LDL}^T con \mathbf{D} diagonale. L'esempio più semplice di una matrice simmetrica che non ha tale fattorizzazione è

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

In effetti, se una matrice 2×2 ha la fattorizzazione \mathbf{LDL}^T , allora è del tipo

$$\begin{bmatrix} 1 & 0 \\ a & 1 \end{bmatrix} \begin{bmatrix} d & 0 \\ 0 & e \end{bmatrix} \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} d & 0 \\ ad & e \end{bmatrix} \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} d & ad \\ ad & a^2d + e \end{bmatrix}$$

I minori principali di nord-ovest della matrice

$$\mathbf{A}_h = \begin{bmatrix} 3 & -2 & 1 \\ -2 & 2 & 0 \\ 1 & 0 & h \end{bmatrix}$$

sono $\delta_1 = 3$, $\delta_2 = \begin{vmatrix} 3 & -2 \\ -2 & 2 \end{vmatrix} = 2$ e

$$\delta_3 = \det(\mathbf{A}) = \begin{vmatrix} -2 & 1 \\ 2 & 0 \end{vmatrix} + h \begin{vmatrix} 3 & -2 \\ -2 & 2 \end{vmatrix} = -2 + 2h = 2(h - 1)$$

Quindi \mathbf{A}_h è definita positiva se e solo se $h > 1$. I minori principali di \mathbf{A} di ordine 1, oltre a δ_1 , sono 2 e h , mentre quelli di ordine 2, oltre a δ_2 , sono

$$\begin{vmatrix} 3 & 1 \\ 1 & h \end{vmatrix} = 3h - 1 \quad \text{e} \quad \begin{vmatrix} 2 & 0 \\ 0 & h \end{vmatrix} = 2h$$

Quindi la matrice è semidefinita positiva se e solo se $h = 1$.

La legge di inerzia di Sylvester

Affrontiamo la questione di determinare quando due matrici \mathbf{A} e \mathbf{B} sono congruenti. Si ricordi che questo equivale a determinare quando le forme quadratiche $\mathbf{x}^T \mathbf{A} \mathbf{x}$ e $\mathbf{x}^T \mathbf{B} \mathbf{x}$ coincidono a meno di una cambiamento di variabili $\mathbf{x} = \mathbf{S} \mathbf{x}$.

Sia \mathbf{A} una matrice definita positiva. Allora \mathbf{A} ammette la fattorizzazione di Cholesky $\mathbf{A} = \mathbf{C} \mathbf{C}^T$, dove \mathbf{C} è una matrice triangolare bassa invertibile. Quindi \mathbf{A} è congruente alla matrice identità \mathbf{I} . Siccome la congruenza è una relazione di equivalenza, due matrici definite positive sono tra loro congruenti.

L'esempio delle matrici definite positive si generalizza così:

TEOREMA 3.9 (Legge di inerzia di Sylvester)

- a) Due matrici simmetriche di ordine n sono congruenti se e solo se hanno lo stesso numero di autovalori positivi e lo stesso numero di autovalori negativi (e quindi anche lo stesso numero di autovalori nulli).
- b) Sia \mathbf{A} una matrice simmetrica di ordine n con s autovalori positivi e t autovalori negativi. Allora \mathbf{A} è congruente alla matrice diagonale che ha s autovalori uguali a 1, t autovalori uguali a -1 , e $n - s - t$ autovalori nulli.
Quindi esiste un cambiamento di coordinate $\mathbf{x} = \mathbf{S}\mathbf{X}$ tale che

$$(3.9) \quad \mathbf{x}^T \mathbf{A} \mathbf{x} = X_1^2 + \cdots + X_s^2 - X_{s+1}^2 - \cdots - X_{s+t}^2.$$

DIMOSTRAZIONE. Il punto b) segue immediatamente dal punto a). Fissiamo una matrice simmetrica \mathbf{A} e la forma quadratica $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$. Sia p il massimo della dimensione di un sottospazio \mathbf{H} di \mathbb{R}^n su cui q è definita positiva:

$$q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} > 0 \quad \text{per ogni } \mathbf{x} \in \mathbf{H}$$

Mostriremo che p è il numero degli autovalori positivi di \mathbf{A} . Per definizione, il numero p dipende solo dalla forma quadratica q e non dalla matrice \mathbf{A} che la rappresenta. Quindi p è anche il numero di autovalori positivi di ogni altra matrice congruente ad \mathbf{A} . Analogamente, il numero di autovalori negativi di ogni matrice congruente ad \mathbf{A} è uguale al massimo delle dimensioni dei sottospazi su cui q è definita negativa. Questo dimostra il punto a). per completare la dimostrazione dobbiamo quindi far vedere che p è uguale al numero s di autovalori positivi di \mathbf{A} . Per il teorema spettrale esistono coordinate $\mathbf{X} = \mathbf{Q}^T \mathbf{x}$ tali che:

$$q(\mathbf{x}) = \lambda_1 X_1^2 + \cdots + \lambda_n X_n^2$$

dove $\lambda_1, \dots, \lambda_n$ sono gli autovalori di \mathbf{A} . Possiamo assumere che gli autovalori positivi di \mathbf{A} siano $\lambda_1, \dots, \lambda_s$. Sia \mathbf{H}_s il sottospazio di \mathbb{R}^n definito dalle equazioni $X_{s+1} = X_{s+2} = \cdots = X_n = 0$. La forma quadratica $q(\mathbf{x})$ su \mathbf{H}_s è definita positiva perché $\lambda_1, \dots, \lambda_s$ sono positivi. Quindi

$$s = \dim \mathbf{H}_s \leq p$$

Per finire occorre mostrare che $p \leq s$: questo significa che, se \mathbf{H} è un sottospazio di \mathbb{R}^n di dimensione $> s$, allora $q(\mathbf{x})$ non è definita positiva su \mathbf{H} . Sia quindi \mathbf{H} un sottospazio di dimensione $d > s$. Sia $\mathcal{L} : \mathbf{H} \rightarrow \mathbf{H}_s$ la proiezione che a un vettore $[X_1, \dots, X_n]^T$ di \mathbf{H} associa il vettore $[X_1, \dots, X_s, 0, \dots, 0]^T$ di \mathbf{H}_s . Per ipotesi $\dim \mathbf{H} > \dim \mathbf{H}_s$ e, quindi, per il teorema di nullità più rango esiste un vettore non nullo $\mathbf{v} \in \mathbf{H}$ tale che $\mathcal{L}(\mathbf{v}) = 0$: questo significa che $X_1(\mathbf{v}) = X_2(\mathbf{v}) = \cdots = X_s(\mathbf{v}) = 0$. Ma allora:

$$q(\mathbf{v}) = \lambda_1 X_1(\mathbf{v})^2 + \cdots + \lambda_n X_n(\mathbf{v})^2 = \lambda_{s+1} X_{s+1}(\mathbf{v})^2 + \cdots + \lambda_n X_n(\mathbf{v})^2$$

Per ogni $k \geq s+1$ gli autovalori λ_k sono minori o uguali a zero, quindi $q(\mathbf{v}) \leq 0$. Siccome \mathbf{v} è un vettore non nullo di \mathbf{H} , questo mostra che la forma quadratica $q(\mathbf{x})$ non è definita positiva su \mathbf{H} . La dimostrazione è così la completa.

OSSERVAZIONE Dalla legge di inerzia segue che, se una forma quadratica si può esprimere in forma diagonale rispetto sia alle variabili \mathbf{X} sia alle variabili \mathbf{Y} :

$$q(\mathbf{x}) = \lambda_1 X_1^2 + \cdots + \lambda_n X_n^2 = \mu_1 Y_1^2 + \cdots + \mu_n Y_n^2$$

allora il numero di λ_k positivi (rispettivamente negativi, rispettivamente nulli) coincide col numero di μ_k positivi (rispettivamente negativi, rispettivamente nulli). Infatti i λ_k (rispettivamente i μ_k) sono gli autovalori della matrice diagonale che rappresenta la forma quadratica rispetto alle coordinate \mathbf{X} (rispettivamente \mathbf{Y}).

OSSERVAZIONE La legge di inerzia classifica le forme quadratiche reali: ogni forma quadratica in n variabili coincide, a meno di un cambiamento di variabili, con una e una sola delle forme quadratiche del tipo

$$(3.10) \quad X_1^2 + \cdots + X_s^2 - X_{s+1}^2 - \cdots - X_{s+t}^2$$

al variare di s e t nell'insieme $0 \leq s, t \leq n$, $s + t \leq n$. Le forme definite positive sono quelle per cui $s = n$ e quindi $t = 0$. Quelle (semi)definite positive sono le forme per cui $t = 0$. Quelle indefinite sono le forme per cui s e t sono entrambi positivi.

La terna di numeri $(s, t, n - s - t)$, cioè il numero di autovalori positivi, negativi e nulli, si dice *segnatura* (di una matrice simmetrica o della forma quadratica rappresentata dalla matrice). La legge di inerzia afferma che due matrici sono congruenti se e solo se hanno la stessa segnatura. In questo senso i numeri s , t e $n - s - t$ sono un insieme completo di invarianti per le classi di congruenza delle matrici simmetriche.

Esempio

Supponiamo che la matrice simmetrica \mathbf{A} si possa fattorizzare nella forma $\mathbf{A} = \mathbf{LDL}^T$ con \mathbf{D} diagonale e \mathbf{L} invertibile. Allora \mathbf{A} è congruente a \mathbf{D} . Gli elementi sulla diagonale principale di \mathbf{D} sono gli autovalori di \mathbf{A} perché \mathbf{D} è diagonale, e dalla legge di inerzia segue che gli autovalori positivi (rispettivamente negativi) di \mathbf{A} sono tanti quanti gli elementi positivi (rispettivamente negativi) sulla diagonale principale di \mathbf{D} . In particolare, se \mathbf{A} ammette la fattorizzazione \mathbf{LU} , il segno degli autovalori di \mathbf{A} coincide col segno dei pivots di \mathbf{A} .

Esempio

In generale, non è unico il sottospazio di dimensione massima s su cui una forma quadratica è definita positiva. Per esempio la forma quadratica $q(x, y) = x^2 - y^2$, che ha $s = 1$ e $t = 1$, è definita positiva su ogni retta di equazione $y = mx$ con $|m| < 1$; la forma si annulla sulle due bisettrici $y = \pm x$, ed è definita negativa sulle rette $y = mx$ con $|m| > 1$ e sull'asse y . La forma $q(x, y) = x^2$, che ha $s = 1$ e $t = 0$, è definita positiva su ogni retta di equazione $y = mx$, ma è nulla sull'asse y .

Diagonalizzazione simultanea di forme quadratiche

Illustriamo ora una versione più generale del teorema spettrale ottenuta sostituendo il prodotto scalare standard di \mathbb{R}^n con un prodotto scalare arbitrario. Come abbiamo visto, ogni prodotto scalare su \mathbb{R}^n è del tipo

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{B}} = \mathbf{x}^T \mathbf{B} \mathbf{y}$$

dove \mathbf{B} è una matrice definita positiva. Data una forma quadratica $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$, ci poniamo ora il problema di trovare il massimo e il minimo del quoziente di Rayleigh

$$(3.11) \quad R(\mathbf{x}) = \frac{q(\mathbf{x})}{\|\mathbf{x}\|_{\mathbf{B}}^2} = \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{B} \mathbf{x}}$$

per $\mathbf{x} \neq \mathbf{0}$. Questa volta è necessario diagonalizzare simultaneamente le due forme quadratiche $q_0(\mathbf{x}) = \mathbf{x}^T \mathbf{B} \mathbf{x} = \|\mathbf{x}\|_{\mathbf{B}}^2$ e $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$. Anche questo si può fare utilizzando il teorema spettrale; allo stesso tempo si risolve l'equazione generalizzata degli autovalori e autovettori:

$$(3.12) \quad \mathbf{A} \mathbf{x} = \lambda \mathbf{B} \mathbf{x}$$

che si ottiene dall'equazione standard

$$(3.13) \quad \mathbf{A} \mathbf{x} = \lambda \mathbf{x} = \lambda \mathbf{I} \mathbf{x}$$

degli autovalori e autovalori sostituendo la matrice identità con la matrice definita positiva \mathbf{B} . Si osservi che la (3.13) equivale all'equazione degli autovalori e autovalori $\mathbf{B}^{-1} \mathbf{A} \mathbf{x} = \lambda \mathbf{x}$ per la matrice $\mathbf{B}^{-1} \mathbf{A}$. Il teorema spettrale fornisce una base di \mathbb{R}^n formata da autovettori di $\mathbf{B}^{-1} \mathbf{A}$ che è ortonormale *rispetto al prodotto scalare definito da \mathbf{B}* :

TEOREMA 3.10 (Diagonalizzazione simultanea di forme quadratiche)

Supponiamo che \mathbf{A} e \mathbf{B} siano due matrici simmetriche di ordine n , con \mathbf{B} definita positiva. Siano $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ e $q_0(\mathbf{x}) = \mathbf{x}^T \mathbf{B} \mathbf{x}$. Allora esiste un cambiamento di variabili $\mathbf{x} = \mathbf{P} \mathbf{X}$ tale che

$$\begin{aligned} q(\mathbf{x}) &= \lambda_1 X_1^2 + \cdots + \lambda_n X_n^2 \\ q_0(\mathbf{x}) &= X_1^2 + \cdots + X_n^2 \end{aligned}$$

Le colonne \mathbf{v}_k di \mathbf{P} formano una base ortonormale di \mathbb{R}^n rispetto al prodotto scalare definito da \mathbf{B} e le coppie $(\mathbf{v}_k, \lambda_k)$ risolvono l'equazione generalizzata degli autovalori e autovettori $\mathbf{A} \mathbf{x} = \lambda \mathbf{B} \mathbf{x}$:

$$(3.14) \quad \mathbf{A} \mathbf{v}_k = \lambda_k \mathbf{B} \mathbf{v}_k \quad \text{per ogni } k = 1, \dots, n$$

DIMOSTRAZIONE. Siccome \mathbf{B} è definita positiva, \mathbf{B} è congruente alla matrice identità: esiste una matrice invertibile \mathbf{S} tale che $\mathbf{S}^T \mathbf{B} \mathbf{S} = \mathbf{I}$. Poiché la matrice \mathbf{A} è simmetrica, la matrice $\mathbf{N} = \mathbf{S}^T \mathbf{A} \mathbf{S}$ è simmetrica, quindi esiste una matrice ortogonale \mathbf{Q} tale che

$$\mathbf{Q}^T \mathbf{N} \mathbf{Q} = \text{diag}(\lambda_1, \dots, \lambda_n)$$

dove i numeri λ_i sono gli autovalori di \mathbf{N} . Poniamo $\mathbf{P} = \mathbf{S} \mathbf{Q}$, allora:

$$\mathbf{P}^T \mathbf{B} \mathbf{P} = \mathbf{Q}^T \mathbf{S}^T \mathbf{B} \mathbf{S} \mathbf{Q} = \mathbf{Q}^T \mathbf{Q} = \mathbf{I} \quad \text{e} \quad \mathbf{P}^T \mathbf{A} \mathbf{P} = \mathbf{Q}^T \mathbf{N} \mathbf{Q} = \text{diag}(\lambda_1, \dots, \lambda_n)$$

Questo mostra che il cambiamento di variabili $\mathbf{X} = \mathbf{P}^{-1} \mathbf{x}$ diagonalizza $q(\mathbf{x})$ e $q_0(\mathbf{x})$ come nell'enunciato. L'uguaglianza $\mathbf{P}^T \mathbf{B} \mathbf{P} = \mathbf{I}$ mostra che le colonne $\mathbf{v}_k = \mathbf{P} \mathbf{e}_k$ di \mathbf{P} formano

una base ortonormale di \mathbb{R}^n rispetto al prodotto scalare definito da \mathbf{B} (qui \mathbf{e}_k denota come sempre la colonna \mathbf{k} della matrice identità):

$$\langle \mathbf{v}_h, \mathbf{v}_k \rangle_{\mathbf{B}} = \mathbf{e}_h \mathbf{P}^T \mathbf{B} \mathbf{P} \mathbf{e}_k = \mathbf{e}_h^T \mathbf{I} \mathbf{e}_k = \delta_{hk}$$

Infine, da $\mathbf{P}^T \mathbf{B} \mathbf{P} = \mathbf{I}$ segue $\mathbf{P}^T = \mathbf{P}^{-1} \mathbf{B}^{-1}$, da cui ricaviamo

$$\mathbf{P}^T \mathbf{A} \mathbf{P} = \mathbf{P}^{-1} \mathbf{B}^{-1} \mathbf{A} \mathbf{P} = \text{diag}(\lambda_1, \dots, \lambda_n)$$

Quest'ultima uguaglianza mostra che, per ogni k , la colonna \mathbf{v}_k della matrice \mathbf{P} è un autovettore di $\mathbf{B}^{-1} \mathbf{A}$ relativo a λ_k . Questo equivale alla (3.14) e conclude la dimostrazione.

OSSERVAZIONE La matrice $\mathbf{B}^{-1} \mathbf{A}$ non è necessariamente simmetrica (il prodotto di due matrici simmetriche non è in generale simmetrico), però è diagonalizzabile con autovalori reali, perché è simile alla matrice simmetrica $\mathbf{N} = \mathbf{S}^T \mathbf{A} \mathbf{S}$ della dimostrazione:

$$\mathbf{N} = \mathbf{S}^T \mathbf{A} \mathbf{S} = \mathbf{S}^{-1} \mathbf{B}^{-1} \mathbf{A} \mathbf{S}$$

Gli autovalori λ_k di $\mathbf{B}^{-1} \mathbf{A}$ sono precisamente le radici del polinomio $\det(\mathbf{A} - \lambda \mathbf{B})$ perché

$$\det(\mathbf{A} - \lambda \mathbf{B}) = \det(\mathbf{B}(\mathbf{B}^{-1} \mathbf{A} - \lambda \mathbf{I})) = \det(\mathbf{B}) \det(\mathbf{B}^{-1} \mathbf{A} - \lambda \mathbf{I})$$

COROLLARIO 3.11 (Massimo e minimo del quoziente di Rayleigh)

Siano \mathbf{A} e \mathbf{B} due matrici simmetriche, con \mathbf{B} definita positiva, e siano q e q_0 le forme quadratiche rappresentate da \mathbf{A} e \mathbf{B} rispettivamente. Siano λ_{\min} e λ_{\max} il minimo e il massimo autovalore di $\mathbf{B}^{-1} \mathbf{A}$. Allora λ_{\min} e λ_{\max} sono il minimo e il massimo del quoziente di Rayleigh $R(\mathbf{x}) = q(\mathbf{x})/q_0(\mathbf{x})$ al variare di \mathbf{x} in $\mathbb{R}^n \setminus \{0\}$.

Esplicitamente:

$$\lambda_{\min} \leq \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{B} \mathbf{x}} \leq \lambda_{\max} \quad \text{per ogni } \mathbf{x} \in \mathbb{R}^n \setminus \{0\}$$

Inoltre esistono \mathbf{v} e \mathbf{w} tali che:

$$R(\mathbf{v}) = \lambda_{\min} \quad \text{e} \quad R(\mathbf{w}) = \lambda_{\max}$$

Se $\lambda_{\min} \neq \lambda_{\max}$, i vettori \mathbf{v} e \mathbf{w} sono ortogonali rispetto al prodotto scalare definito da \mathbf{B} : $\mathbf{v}^T \mathbf{B} \mathbf{w} = 0$.

DIMOSTRAZIONE. Usiamo le notazioni del teorema precedente: facciamo il cambiamento di variabili $\mathbf{x} = \mathbf{P} \mathbf{X}$. Nelle nuove coordinate le forme quadratiche sono

$$q_0(\mathbf{X}) = X_1^2 + \cdots + X_n^2 \quad \text{e} \quad q(\mathbf{X}) = \lambda_1 X_1^2 + \cdots + \lambda_n X_n^2$$

e la tesi è a questo punto evidente. Come \mathbf{v} e \mathbf{w} si prendano le colonne di \mathbf{P} corrispondenti agli autovalori λ_{\min} e λ_{\max} .



Curvatura di una superficie

Illustriamo ora brevemente l'esempio classico della prima e della seconda forma fondamentali di una superficie introdotte da Gauss. Localmente una superficie regolare S in \mathbb{R}^3 è l'immagine di una funzione iniettiva $\mathbf{r}(u, v) = [x(u, v), y(u, v), z(u, v)]^T$ dove $[u, v]^T$ varia in un intorno U di $[u_0, v_0]^T$ in \mathbb{R}^2 ; si richiede che le funzioni $x(u, v)$, $y(u, v)$ e $z(u, v)$ siano di classe C^2 e che i vettori tangenti alle linee coordinate

$$\mathbf{r}_u(u, v) = [x_u(u, v), y_u(u, v), z_u(u, v)]^T \quad \text{e} \quad \mathbf{r}_v(u, v) = [x_v(u, v), y_v(u, v), z_v(u, v)]^T$$

siano linearmente indipendenti per ogni (u, v) . Poniamo $P = \mathbf{r}(u_0, v_0)$, $\mathbf{v}_1 = \mathbf{r}_u(u_0, v_0)$, $\mathbf{v}_2 = \mathbf{r}_v(u_0, v_0)$. Il piano tangente $T_P S$ a S in p è il piano di \mathbb{R}^3 generato da \mathbf{v}_1 e \mathbf{v}_2 . Il vettore

$$\mathbf{n} = \mathbf{v}_1 \times \mathbf{v}_2$$

è quindi normale al piano tangente. Il piano tangente $T_P S$ ha il prodotto scalare ottenuto per restrizione dal prodotto scalare standard di \mathbb{R}^3 . Per un vettore $x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2$ del piano tangente abbiamo

$$\|x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2\|^2 = Ex_1^2 + 2Fx_1x_2 + Gx_2^2$$

dove, utilizzando le notazioni di Gauss, abbiamo posto $E = \|\mathbf{v}_1\|^2 = \mathbf{v}_1^T \mathbf{v}_1$, $F = \mathbf{v}_1^T \mathbf{v}_2$, e $G = \|\mathbf{v}_2\|^2 = \mathbf{v}_2^T \mathbf{v}_2$. La forma quadratica $I(x_1, x_2) = Ex_1^2 + 2Fx_1x_2 + Gx_2^2$ si dice prima forma fondamentale della superficie; la matrice che la rappresenta è

$$\mathbf{B} = \begin{bmatrix} E & F \\ F & G \end{bmatrix}$$

La prima forma fondamentale, in quanto restrizione a $T_P S$ del prodotto scalare standard di \mathbb{R}^3 , è un prodotto scalare su $T_P S$ e quindi la matrice \mathbf{B} è definita positiva. La seconda forma fondamentale è

$$II(x_1, x_2) = ex_1^2 + 2fx_1x_2 + gx_2^2$$

dove $e = \mathbf{n}^T \mathbf{r}_{uu}(u_0, v_0)$, $f = \mathbf{n}^T \mathbf{r}_{uv}(u_0, v_0)$, $g = \mathbf{n}^T \mathbf{r}_{vv}(u_0, v_0)$. La matrice che rappresenta la seconda forma fondamentale è

$$\mathbf{A} = \begin{bmatrix} e & f \\ f & g \end{bmatrix}$$

La seconda forma fondamentale misura la curvatura della superficie in P : più precisamente, se $\mathbf{v} = x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2$ è un vettore tangente alla superficie in P e \mathbf{H}_v è il piano normale alla superficie generato da \mathbf{n} e da \mathbf{v} , allora il quoziente di Rayleigh

$$R(\mathbf{v}) = \frac{II(x_1, x_2)}{I(x_1, x_2)} = II\left(\frac{\mathbf{v}}{\|\mathbf{v}\|}\right)$$

è la curvatura in P della sezione $S \cap H_v$. Un famoso teorema di Eulero afferma che la curvatura $R(\mathbf{v})$ di $S \cap H_v$, quando non è costante, assume un valore minimo $k_1 = R(\mathbf{v}_1)$ e un valore massimo $k_2 = R(\mathbf{v}_2)$ in due direzioni perpendicolari: \mathbf{v}_1 e \mathbf{v}_2 sono ortogonali tra loro in \mathbb{R}^3 . I valori k_1 e k_2 si dicono curvature principali della superficie in P e le rette generate da \mathbf{v}_1 e \mathbf{v}_2 si dicono direzioni principali di curvatura. Il teorema di Eulero è una conseguenza immediata del corollario 3.41: $k_1 \leq k_2$ sono gli autovalori di $\mathbf{B}^{-1} \mathbf{A}$; quando $k_1 \neq k_2$, cioè $R(\mathbf{v})$ non è costante, gli autovettori relativi a k_1 e k_2 sono ortogonali rispetto al prodotto scalare definito da \mathbf{B} , cioè sono ortogonali in \mathbb{R}^3 . La curvatura Gaussiana K e la curvatura media H della superficie in P sono definite come il prodotto e la media aritmetica delle curvature principali:

$$K = k_1 k_2, \quad H = \frac{1}{2}(k_1 + k_2)$$

Ne segue che K è il prodotto degli autovalori di $\mathbf{B}^{-1}\mathbf{A}$, quindi

$$K = \det(\mathbf{B}^{-1}\mathbf{A}) = \frac{\det(\mathbf{A})}{\det(\mathbf{B})} = \frac{eg - f^2}{EG - F^2}$$

La curvatura media H è invece la media aritmetica degli autovalori, quindi la metà della traccia della matrice:

$$H = \frac{1}{2} \operatorname{tr}(\mathbf{B}^{-1}\mathbf{A}) = \frac{1}{2} \frac{eG - 2fF + gE}{EG - F^2}$$

Una volta calcolate H e K , possiamo ricavare le curvature principali:

$$k_{1,2} = H \pm \sqrt{H^2 - K}$$



Calcolare tutti i minori principali della matrice

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 0 \\ 3 & 0 & 4 \end{bmatrix}$$

Quali sono i minori principali di nord ovest della matrice? Qual è il segno della forma quadratica associata?

Sia \mathbf{A} una matrice qualsiasi. Mostrare che $\mathbf{A}^T\mathbf{A}$ è (semi)definita positiva. Quando è definita positiva?

Scrivere la matrice che rappresenta la forma quadratica

$$q(x, y) = 3x^2 + 2xy + y^2$$

e stabilire il segno della forma quadratica.

Scrivere la matrice che rappresenta la forma quadratica

$$q(x, y) = 3x^2 + 2xy + y^2$$

e stabilire il segno della forma quadratica. Determinare un cambiamento di variabili che diagonalizza la forma quadratica.

Stabilire il segno della forma quadratica

$$q(x, y, z) = x^2 + 8xy + 7y^2 - 2xz + 8yz + 5z^2$$

Sia \mathbf{A} una matrice simmetrica. Mostrare che \mathbf{A}^2 è simmetrica e (semi)definita positiva.

Per ciascuna delle matrici simmetriche degli esercizi del paragrafo precedente si scriva la forma quadratica associata, e si determini se è (semi)definita positiva/negativa oppure indefinita. Si calcolino i minori principali di nord-ovest di tali matrici e, quindi, si verifichi il criterio di positività (negatività) della forma in termini di tali minori.

La funzione

$$q(x, y, z) = 2x^2 - 4xy + 9y^2 + 4yz + 2z^2$$

ha nell'origine $(0, 0, 0)$ un punto di massimo o di minimo assoluto?

Si determinino il valore massimo e minimo assoluto della funzione

$$R(x, y, z) = \frac{10x^2 + 9y^2 + 4yz + 6z^2}{x^2 + y^2 + z^2}$$

al variare di (x, y, z) in $\mathbb{R}^3 \setminus \{(0, 0, 0)\}$.

Si determinino il valore massimo e minimo assoluto della funzione

$$R(x, y, z) = \frac{9x^2 + 4xy + 2y^2 - 8xz + 8z^2}{x^2 + y^2 + 4z^2}$$

al variare di (x, y, z) in $\mathbb{R}^3 \setminus \{(0, 0, 0)\}$.

Mostrare che gli elementi sulla diagonale principale di una matrice definita positiva sono positivi.

Si mostri che il quoziente di Rayleigh è costante sulle rette uscenti dall'origine, e che è invariante per trasformazioni di coordinate $\mathbf{x} = \mathbf{Q}\mathbf{x}'$ con \mathbf{Q} ortogonale.

Trovare una formula per la curvatura Gaussiana della superficie $z = f(x, y)$ (il grafico della funzione $f(x, y)$).

Calcolare le curvature principali di una sfera e di un cilindro.

4 LA DECOMPOSIZIONE AI VALORI SINGOLARI

Per una matrice simmetrica \mathbf{A} , il teorema spettrale fornisce la decomposizione

$$(4.1) \quad \mathbf{A} = \mathbf{Q}\mathbf{D}\mathbf{Q}^T$$

dove \mathbf{Q} è una matrice ortogonale e \mathbf{D} è una matrice diagonale. L'interpretazione geometrica è che il cambiamento di coordinate $\mathbf{X} = \mathbf{Q}\mathbf{x}$, che è un'isometria, trasforma la matrice rappresentativa dell'applicazione lineare $\mathcal{L}_{\mathbf{A}}$ in una matrice diagonale. Questa decomposizione è possibile solo per le matrici simmetriche perché dalla (4.1) segue che \mathbf{A} è simmetrica. In questo paragrafo descriviamo la *decomposizione ai valori singolari* di una matrice, che è un utile surrogato della (4.1) per una matrice qualsiasi. In inglese tale decomposizione è detta *singular value decomposition* o in forma abbreviata SVD.

Consideriamo un'arbitraria matrice reale \mathbf{A} di tipo (m, n) . A partire da \mathbf{A} si possono costruire le due matrici simmetriche $\mathbf{A}^T\mathbf{A}$ e $\mathbf{A}\mathbf{A}^T$. Si osservi che lo spazio riga e lo spazio colonna di una matrice simmetrica sono uguali; in particolare

$$\text{Row}(\mathbf{A}^T\mathbf{A}) = \text{Col}(\mathbf{A}^T\mathbf{A}) \quad \text{e} \quad \text{Row}(\mathbf{A}\mathbf{A}^T) = \text{Col}(\mathbf{A}\mathbf{A}^T)$$

Nel paragrafo sul metodo dei minimi quadrati abbiamo già considerato la matrice $\mathbf{A}^T \mathbf{A}$, e abbiamo dimostrato che ha lo stesso nucleo, e quindi lo stesso rango, di \mathbf{A} . Sostituendo \mathbf{A} con \mathbf{A}^T , si vede che il rango di $\mathbf{A} \mathbf{A}^T$ è uguale al rango di \mathbf{A}^T , che coincide anch'esso col rango di \mathbf{A} . In conclusione:

Le quattro matrici \mathbf{A} , \mathbf{A}^T , $\mathbf{A}^T \mathbf{A}$ e $\mathbf{A} \mathbf{A}^T$ hanno lo stesso rango.

Mostriamo ora che le matrici $\mathbf{A}^T \mathbf{A}$ e $\mathbf{A} \mathbf{A}^T$ hanno anche gli stessi autovalori non nulli. Questo implica nuovamente che hanno lo stesso rango, perché il rango di una matrice diagonalizzabile è il numero dei suoi autovalori non nulli.

LEMMA 4.1 Sia \mathbf{A} una matrice reale di tipo (m, n) . Allora le due matrici simmetriche $\mathbf{A}^T \mathbf{A}$ e $\mathbf{A} \mathbf{A}^T$ sono (semi)definite positive e hanno gli stessi autovalori non nulli con uguali molteplicità.

DIMOSTRAZIONE. La matrice $\mathbf{A}^T \mathbf{A}$ è (semi)definita positiva perché

$$\mathbf{x}^T (\mathbf{A}^T \mathbf{A}) \mathbf{x} = \|\mathbf{Ax}\|^2 \geq 0 \quad \text{per ogni } \mathbf{x} \in \mathbb{R}^n$$

Prendendo \mathbf{A}^T al posto di \mathbf{A} si ottiene che $\mathbf{A} \mathbf{A}^T$ è (semi)definita positiva. Sia λ un autovalore non nullo di $\mathbf{A}^T \mathbf{A}$ e sia $\mathbf{v} \in \mathbb{R}^n$ un autovettore di $\mathbf{A}^T \mathbf{A}$ relativo a λ : $\mathbf{A}^T \mathbf{A} \mathbf{v} = \lambda \mathbf{v}$. Moltiplicando questa uguaglianza a sinistra per \mathbf{A} otteniamo

$$(4.2) \quad \mathbf{A} \mathbf{A}^T (\mathbf{Av}) = \lambda (\mathbf{Av})$$

Osserviamo che $\mathbf{Av} \neq \mathbf{0}$: altrimenti avremmo

$$\lambda \mathbf{v} = \mathbf{A}^T \mathbf{A} \mathbf{v} = \mathbf{A}^T \mathbf{0} = \mathbf{0}$$

il che è impossibile perché $\lambda \neq 0$ e $\mathbf{v} \neq \mathbf{0}$. Quindi $\mathbf{Av} \neq \mathbf{0}$, e dalla (4.2) segue allora che \mathbf{Av} è un autovettore di $\mathbf{A} \mathbf{A}^T$ relativo a λ . Questo mostra che λ è un autovalore anche di $\mathbf{A} \mathbf{A}^T$. Per mostrare l'uguaglianza delle molteplicità dobbiamo approfondire il ragionamento fatto. Siccome la matrice $\mathbf{A}^T \mathbf{A}$ è simmetrica, i suoi autovalori sono regolari, per cui la molteplicità algebrica di λ coincide con la sua molteplicità geometrica, e quindi è la dimensione dell'autospazio

$$\mathbf{V}_\lambda = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{A}^T \mathbf{A} \mathbf{v} = \lambda \mathbf{v}\}$$

Per lo stesso motivo, la molteplicità di λ come autovalore di $\mathbf{A} \mathbf{A}^T$ è la dimensione dell'autospazio

$$\mathbf{W}_\lambda = \{\mathbf{w} \in \mathbb{R}^m : \mathbf{A} \mathbf{A}^T \mathbf{w} = \lambda \mathbf{w}\}$$

Abbiamo appena mostrato che, se \mathbf{v} appartiene a \mathbf{V}_λ , allora \mathbf{Av} appartiene a \mathbf{W}_λ , e che, se $\mathbf{v} \in \mathbf{V}_\lambda$ è non nullo, anche $\mathbf{Av} \neq \mathbf{0}$. Quindi l'applicazione lineare $\mathcal{L} : \mathbf{V}_\lambda \rightarrow \mathbf{W}_\lambda$ che manda \mathbf{v} in \mathbf{Av} ha nel nucleo solo il vettore nullo, per cui è iniettiva. In particolare $\dim(\mathbf{V}_\lambda) \leq \dim(\mathbf{W}_\lambda)$. Scambiando i ruoli di \mathbf{A} e \mathbf{A}^T , lo stesso argomento mostra che $\dim(\mathbf{W}_\lambda) \leq \dim(\mathbf{V}_\lambda)$. Quindi $\dim(\mathbf{W}_\lambda) = \dim(\mathbf{V}_\lambda)$, il che conclude la dimostrazione.

Si può riassumere la dimostrazione in questo modo: se $\lambda \neq 0$, la moltiplicazione a sinistra per \mathbf{A} mappa l'autospazio di λ come autovalore di $\mathbf{A}^T \mathbf{A}$ nell'autospazio di λ come autovalore di $\mathbf{A} \mathbf{A}^T$; tale applicazione lineare è un isomorfismo (esercizio: l'applicazione inversa è la moltiplicazione a sinistra per $\frac{\mathbf{A}^T}{\lambda}$).

Siccome $\mathbf{A}^T \mathbf{A}$ è (semi)definita positiva, i suoi autovalori non nulli sono positivi.

DEFINIZIONE 4.2 (Valori singolari di una matrice)

Sia \mathbf{A} una matrice qualsiasi. I valori singolari di \mathbf{A} sono le radici quadrate positive degli autovalori non nulli di $\mathbf{A}^T \mathbf{A}$, contati con la loro molteplicità come autovalori di $\mathbf{A}^T \mathbf{A}$.

OSSERVAZIONE Sia r il rango di \mathbf{A} . Siccome $\mathbf{A}^T \mathbf{A}$ è simmetrica e quindi diagonalizzabile, il numero dei suoi autovalori non nulli coincide col rango di $\mathbf{A}^T \mathbf{A}$, che è anch'esso uguale a r : per questo i valori singolari sono r . Di solito si ordinano in senso decrescente e li si denota col simbolo σ_k ; si scrive che i valori singolari di \mathbf{A} sono $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$. Per definizione, $\sigma_k = \sqrt{\lambda_k}$, dove $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$ sono gli autovalori non nulli di $\mathbf{A}^T \mathbf{A}$ in ordine decrescente; ogni autovalore (e quindi ogni valore singolare) si ripete un numero di volte uguale alla sua molteplicità algebrica.



Se \mathbf{A} è una matrice simmetrica, i valori singolari di \mathbf{A} sono i valori assoluti degli autovalori non nulli di \mathbf{A} . Infatti, se \mathbf{A} è simmetrica e i suoi autovalori sono $\lambda_1, \dots, \lambda_n$, allora $\mathbf{A}^T \mathbf{A} = \mathbf{A}^2$ ha autovalori $\lambda_1^2, \dots, \lambda_n^2$ (si osservi che la molteplicità di λ_k^2 come autovalore di \mathbf{A}^2 coincide con la molteplicità di λ_k come autovalore di \mathbf{A} perché \mathbf{A} è diagonalizzabile). I valori singolari di \mathbf{A} sono perciò $\sqrt{\lambda_k^2} = |\lambda_k|$ per i k per cui $\lambda_k \neq 0$. Per esempio, i valori singolari di $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, che ha autovalori $\lambda_1 = 1$ e $\lambda_2 = -1$, sono $\sigma_1 = \sigma_2 = 1$.

In particolare, i valori singolari di una matrice diagonale sono i valori assoluti degli elementi non nulli della matrice.



Dati due numeri reali a e b , consideriamo la matrice

$$\mathbf{A} = \begin{bmatrix} 0 & a & 0 \\ b & 0 & 0 \end{bmatrix}$$

I valori singolari di \mathbf{A} sono gli autovalori non nulli di $\mathbf{A}^T \mathbf{A}$, che è una matrice 3×3 , o anche di $\mathbf{A} \mathbf{A}^T$, che è una matrice 2×2 . Conviene calcolare la matrice di ordine minore:

$$\mathbf{A} \mathbf{A}^T = \begin{bmatrix} 0 & a & 0 \\ b & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & b \\ a & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} a^2 & 0 \\ 0 & b^2 \end{bmatrix}$$

Gli autovalori di $\mathbf{A} \mathbf{A}^T$ sono a^2 e b^2 (quelli di $\mathbf{A}^T \mathbf{A}$ sono perciò 0, a^2 e b^2 : spiegare e verificare). Se assumiamo che a e b siano diversi da zero, cioè che \mathbf{A} abbia rango 2, allora i valori singolari di \mathbf{A} sono $|a|$ e $|b|$. Per esempio i valori singolari di

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

sono $\sigma_1 = 2$ e $\sigma_2 = 1$. I valori singolari di

$$\mathbf{A} = \begin{bmatrix} 0 & -3 & 0 \\ -3 & 0 & 0 \end{bmatrix}$$

sono $\sigma_1 = \sigma_2 = 3$.

Nella decomposizione ai valori singolari di una matrice \mathbf{A} di tipo (m, n) , i valori singolari si dispongono in una matrice Σ di tipo (m, n) in questo modo:

$$(4.3) \quad \Sigma = \left[\begin{array}{cccc|ccc} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \sigma_r & 0 & \cdots & 0 \\ \hline 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{array} \right]$$

DEFINIZIONE 4.3 Una matrice $\Sigma = [b_{ij}]$ si dice *pseudodiagonale* se $b_{ij} = 0$ per ogni (i, j) con $i \neq j$.

Si osservi che una matrice pseudodiagonale quadrata è una matrice diagonale; una matrice pseudodiagonale però non è necessariamente quadrata. Per esempio le matrici

$$\Sigma_1 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 4 \\ 0 & 0 \end{bmatrix}$$

sono pseudodiagonali, ma non sono diagonali.

OSSERVAZIONE Il rango di una matrice pseudodiagonale coincide col numero dei suoi elementi diversi da zero.

TEOREMA 4.4 (Decomposizione ai valori singolari)

Sia \mathbf{A} una matrice reale di tipo (m, n) e di rango r . Allora esistono una matrice pseudodiagonale Σ di tipo (m, n) come nella (4.3) con $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, una matrice ortogonale \mathbf{U} di tipo (m, m) e una matrice ortogonale \mathbf{V} di tipo (n, n) tali che

$$(4.4) \quad \mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$$

Tale decomposizione si dice *decomposizione ai valori singolari di \mathbf{A}* ; non è unica, ma in ogni decomposizione siffatta gli elementi non nulli σ_k di Σ sono i valori singolari di \mathbf{A} .

DIMOSTRAZIONE. Verifichiamo prima che gli elementi non nulli di Σ sono i valori singolari della matrice. Supponiamo che $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ come nell'enunciato. Allora dall'ortogonalità di \mathbf{U} segue

$$\mathbf{A}^T\mathbf{A} = \mathbf{V}\Sigma^T\mathbf{U}^T\mathbf{U}\Sigma\mathbf{V}^T = \mathbf{V}\Sigma^T\Sigma\mathbf{V}^T$$

Siccome Σ è pseudodiagonale, la matrice $\mathbf{D} = \Sigma^T\Sigma$ è una matrice *diagonale* $n \times n$; inoltre per $k = 1, 2, \dots, r$ l'elemento λ_k di posto (k, k) in \mathbf{D} è il quadrato del corrispondente elemento σ_k in Σ . Per ipotesi $\sigma_k > 0$, quindi $\sigma_k = \sqrt{\lambda_k}$. D'altra parte

$$\mathbf{D} = \Sigma^T\Sigma = \mathbf{V}^T\mathbf{A}^T\mathbf{A}\mathbf{V} = \mathbf{V}^{-1}\mathbf{A}^T\mathbf{A}\mathbf{V}$$

Questo mostra che gli elementi λ_k sulla diagonale principale di \mathbf{D} sono gli autovalori di $\mathbf{A}^T \mathbf{A}$, e che la colonna k di \mathbf{V} è un autovettore di $\mathbf{A}^T \mathbf{A}$ relativo a λ_k (per $k > r$ si ha $\lambda_k = 0$ e le corrispondenti colonne sono vettori del nucleo di $\mathbf{A}^T \mathbf{A}$).

Queste considerazioni mostrano come costruire la decomposizione. Siano $\lambda_1, \lambda_2, \dots, \lambda_n$ gli autovalori di $\mathbf{A}^T \mathbf{A}$, ordinati in modo che $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$. Siccome $r(\mathbf{A}^T \mathbf{A}) = r(\mathbf{A}) = r$, gli ultimi $n - r$ autovalori sono nulli. Per il teorema spettrale, esiste una matrice ortogonale \mathbf{V} tale che

$$\mathbf{V}^T \mathbf{A}^T \mathbf{A} \mathbf{V} = \text{diag}(\lambda_1, \dots, \lambda_n)$$

La k -esima colonna \mathbf{v}_k di \mathbf{V} è un autovettore di $\mathbf{A}^T \mathbf{A}$ relativo all'autovalore λ_k . Le n colonne di \mathbf{V} formano una base ortonormale di \mathbb{R}^n e le ultime $n - r$ appartengono a $\text{Ker}(\mathbf{A}^T \mathbf{A})$, che coincide con $\text{Ker}(\mathbf{A})$ come abbiamo già osservato. Definiamo

$$\mathbf{u}_k = \frac{1}{\sigma_k} \mathbf{A} \mathbf{v}_k \quad \text{per } k = 1, 2, \dots, r$$

Per costruzione i vettori \mathbf{u}_k appartengono allo spazio colonna $\text{Col}(\mathbf{A})$: vogliamo mostrare che ne formano una base ortonormale. Siccome lo spazio colonna ha dimensione $r(\mathbf{A}) = r$ e i vettori \mathbf{u}_k sono esattamente r , basta mostrare che sono dei versori a due a due ortogonali. I vettori \mathbf{u}_k sono dei versori perché

$$\mathbf{u}_k^T \mathbf{u}_k = \frac{1}{\sigma_k^2} \mathbf{v}_k^T \mathbf{A}^T \mathbf{A} \mathbf{v}_k = \frac{1}{\sigma_k^2} \lambda_k \mathbf{v}_k^T \mathbf{v}_k = \mathbf{v}_k^T \mathbf{v}_k = 1$$

Nel penultimo passaggio abbiamo utilizzato l'uguaglianza $\lambda_k = \sigma_k^2$, nell'ultimo il fatto che \mathbf{v}_k è un versore in quanto colonna di una matrice ortogonale. I vettori \mathbf{u}_k sono a due a due ortogonali perché, se $h \neq k$, allora

$$\mathbf{u}_k^T \mathbf{u}_h = \frac{1}{\sigma_h \sigma_k} \mathbf{v}_k^T \mathbf{A}^T \mathbf{A} \mathbf{v}_h = \frac{\lambda_h}{\sigma_h \sigma_k} \mathbf{v}_k^T \mathbf{v}_h = 0$$

Nell'ultimo passaggio abbiamo usato il fatto che le colonne di \mathbf{V} sono a due a due ortogonali.

Abbiamo così dimostrato che $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ è una base ortonormale di $\text{Col}(\mathbf{A})$. Sia ora $\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$ una base ortonormale di $\text{Ker}(\mathbf{A}^T)$, che è il complemento ortogonale dello spazio colonna in \mathbb{R}^m , e sia \mathbf{U} la matrice ortogonale che ha per colonne i vettori \mathbf{u}_j per $j = 1, 2, \dots, m$. Per costruzione

$$(4.5) \quad \mathbf{A} \mathbf{v}_k = \sigma_k \mathbf{u}_k \quad \text{per } k = 1, 2, \dots, r \quad \text{e} \quad \mathbf{A} \mathbf{v}_k = \mathbf{0} \quad \text{per } k > r$$

Quindi

$$\begin{aligned} \mathbf{AV} &= [\mathbf{Av}_1 \cdots \mathbf{Av}_r | \mathbf{Av}_{r+1} \cdots \mathbf{Av}_n] = [\sigma_1 \mathbf{u}_1 \cdots \sigma_r \mathbf{u}_r | \mathbf{0} \cdots \mathbf{0}] = \\ &= [\mathbf{u}_1 \cdots \mathbf{u}_r | \mathbf{u}_{r+1} \cdots \mathbf{u}_m] \left[\begin{array}{cc|c} \sigma_1 & \cdots & 0 & \mathbf{O} \\ \vdots & & \vdots & \mathbf{O} \\ 0 & \cdots & \sigma_r & \mathbf{O} \\ \hline \mathbf{O} & & & \mathbf{O} \end{array} \right] = \mathbf{U} \Sigma \end{aligned}$$

Moltiplicando a destra per \mathbf{V}^T otteniamo la decomposizione $\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^T$.

(Alternativamente, la (4.5) mostra che Σ è la matrice che rappresenta l'applicazione lineare $\mathbf{x} \mapsto \mathbf{Ax}$ rispetto alla base di \mathbb{R}^n formata dalle colonne di \mathbf{V} e alla base di \mathbb{R}^m formata dalle colonne di \mathbf{U} . Per la formula del cambiamento della matrice rappresentativa $\Sigma = \mathbf{U}^{-1} \mathbf{AV} = \mathbf{U}^T \mathbf{AV}$, per cui $\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^T$).

OSSERVAZIONE Ripercorrendo con attenzione la dimostrazione si vede che nella decomposizione $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$:

- le colonne di \mathbf{V} sono autovettori di $\mathbf{A}^T\mathbf{A}$; le prime r formano una base ortonormale dello spazio riga $\text{Row}(\mathbf{A})$; le altre $n - r$ formano una base ortonormale di $\text{Ker}(\mathbf{A})$;
- le colonne di \mathbf{U} sono autovettori di $\mathbf{A}\mathbf{A}^T$; le prime r formano una base ortonormale dello spazio colonna $\text{Col}(\mathbf{A})$; le altre $m - r$ formano una base ortonormale di $\text{Ker}(\mathbf{A}^T)$.

OSSERVAZIONE Dal punto di vista geometrico, otteniamo una fattorizzazione di un'arbitraria applicazione lineare $\mathbf{x} \mapsto \mathbf{Ax}$ come prodotto dell'isometria $\mathbf{x} \mapsto \mathbf{X} = \mathbf{V}^T\mathbf{x}$, seguita dalla mappa $\mathbf{X} \mapsto \mathbf{Y} = \Sigma\mathbf{X}$ che, nel caso $m = n$, consiste semplicemente nel riscalare gli assi coordinati, seguita dall'isometria $\mathbf{Y} \mapsto \mathbf{y} = \mathbf{UY}$.

Se \mathbf{A} è una matrice simmetrica, gli autovettori di $\mathbf{A}^T\mathbf{A} = \mathbf{A}^2$ coincidono con gli autovettori di \mathbf{A} e Σ è la matrice diagonale che ha sulla diagonale principale i valori assoluti $\sigma_k = |\lambda_k|$ degli autovalori di \mathbf{A} ; inoltre:

$$\mathbf{u}_k = \frac{1}{\sigma_k} \mathbf{Av}_k = \frac{\lambda_k}{\sigma_k} \mathbf{v}_k = \pm \mathbf{v}_k \quad \text{per } k = 1, 2, \dots, r$$

dove \pm è il segno di λ_k . Se tutti gli autovalori di \mathbf{A} sono nonnegativi (cioè se \mathbf{A} è (semi)definita positiva), e scegliamo $\mathbf{u}_k = \mathbf{v}_k$ per ogni $k \geq r + 1$, allora $\mathbf{U} = \mathbf{V}$ e la decomposizione SVD di \mathbf{A} coincide con la decomposizione $\mathbf{A} = \mathbf{Q}\Sigma\mathbf{Q}^T$ del teorema spettrale. Se però la matrice ha degli autovalori negativi, le due decomposizioni sono diverse. Per esempio, la decomposizione $\mathbf{A} = \mathbf{Q}\Sigma\mathbf{Q}^T$ della matrice $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ è

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$$

mentre la decomposizione SVD è

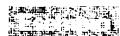
$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$$

Calcoliamo una SVD della matrice $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$. Il prodotto $\mathbf{A}^T\mathbf{A}$ è la matrice $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ che ha un unico autovalore non nullo $\lambda = 2$; l'autospazio relativo è la retta $x_1 = x_2$. Perciò $\Sigma = \text{diag}(\sqrt{2}, 0)$ e possiamo scegliere $\mathbf{V} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$. Allora

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}} \mathbf{Av}_1 = [1, 0]^T$$

Scegliendo $\mathbf{u}_2 = [0, 1]^T$ come base di \mathbf{u}_1^\perp , otteniamo $\mathbf{U} = \mathbf{I}$. Una decomposizione SVD di \mathbf{A} è perciò:

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$



Consideriamo la matrice $\mathbf{A} = \begin{bmatrix} 0 & -2 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}$. Per costruire una SVD di \mathbf{A} , calcoliamo

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

Un autovettore di norma uno relativo all'autovalore 2 di $\mathbf{A}^T \mathbf{A}$ è il secondo vettore della base canonica di \mathbb{R}^2 e un autovettore di norma uno relativo all'autovalore 1 è il primo vettore della base canonica. Quindi in questo caso:

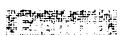
$$\mathbf{V} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Per costruire la matrice \mathbf{U} calcoliamo:

$$\mathbf{u}_1 = \frac{1}{\sigma_1} \mathbf{A} \mathbf{v}_1 = [-1, 0, 0]^T \quad \text{e} \quad \mathbf{u}_2 = \frac{1}{\sigma_2} \mathbf{A} \mathbf{v}_2 = [0, 1, 0]^T$$

La terza colonna \mathbf{u}_3 di \mathbf{U} dev'essere un versore ortogonale a \mathbf{u}_1 e \mathbf{u}_2 , per esempio $\mathbf{u}_3 = [0, 0, 1]^T$. La SVD di \mathbf{A} è perciò:

$$\begin{bmatrix} 0 & -2 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$



Se $\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^T$ è una SVD di \mathbf{A} , allora $\mathbf{A}^T = \mathbf{V} \Sigma^T \mathbf{U}^T$ è una SVD di \mathbf{A}^T . Per esempio, dall'esempio precedente deduciamo che la matrice $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ -2 & 0 & 0 \end{bmatrix}$ ha la decomposizione

$$\begin{bmatrix} 0 & 1 & 0 \\ -2 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



Calcoliamo una SVD della matrice $\mathbf{B} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}$. Siccome $\mathbf{B}^T \mathbf{B} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, gli autovalori di $\mathbf{B}^T \mathbf{B}$ sono $\lambda_1 = 3$ e $\lambda_2 = 1$. I valori singolari di \mathbf{B} sono perciò $\sigma_1 = \sqrt{3}$ e $\sigma_2 = 1$. L'autospazio di $\mathbf{B}^T \mathbf{B}$ relativo a $\lambda_1 = 3$ è la retta generata da $[1, 1]^T$, mentre quello relativo a $\lambda_2 = 1$ è la retta generata da $[-1, 1]^T$. Possiamo perciò prendere

$$\mathbf{V} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Passiamo al calcolo della matrice \mathbf{U} . La prima colonna è

$$\mathbf{u}_1 = \frac{1}{\sigma_1} \mathbf{Bv}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}.$$

mentre la seconda colonna è:

$$\mathbf{u}_2 = \frac{1}{\sigma_2} \mathbf{Bv}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

Un versore \mathbf{u}_3 perpendicolare a \mathbf{u}_1 e \mathbf{u}_2 è $\frac{1}{\sqrt{3}}[1, -1, 1]^T$. Quindi

$$\mathbf{U} = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

Posto $\Sigma = [\sqrt{3} \ 0 \ 0 \ 0 \ 1 \ 0]^T$, una SVD di \mathbf{B} è $\mathbf{U}\Sigma\mathbf{V}^T$.

Decomposizione polare

La seguente proposizione fornisce per le matrici quadrate una decomposizione analoga alla decomposizione polare di un numero complesso.

PROPOSIZIONE 4.5 (Decomposizione polare)

Sia \mathbf{A} una matrice quadrata reale. Allora esistono una matrice (semi)definita positiva \mathbf{H} e una matrice ortogonale \mathbf{Q} tali che

$$\mathbf{A} = \mathbf{HQ}$$

Se \mathbf{A} è invertibile, allora \mathbf{H} è definita positiva e la decomposizione è unica.

DIMOSTRAZIONE. Sia $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ una SVD di \mathbf{A} . Siccome \mathbf{A} è quadrata, le matrici \mathbf{U} , Σ e \mathbf{V} sono tutte quadrate dello stesso ordine e possiamo riscrivere la decomposizione nella forma

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T = (\mathbf{U}\Sigma\mathbf{U}^T)(\mathbf{U}\mathbf{V}^T)$$

Poniamo $\mathbf{H} = \mathbf{U}\Sigma\mathbf{U}^T$ e $\mathbf{Q} = \mathbf{U}\mathbf{V}^T$. Allora \mathbf{H} è (semi)definita positiva perché i suoi autovalori sono gli elementi sulla diagonale principale di Σ , che sono non negativi; e \mathbf{Q} è ortogonale perché \mathbf{U} e \mathbf{V} lo sono.

Se \mathbf{A} è invertibile, anche \mathbf{H} lo è, e quindi \mathbf{H} è definita positiva. Per quel che riguarda l'unicità, se $\mathbf{A} = \mathbf{H}_1\mathbf{Q}_1$ con \mathbf{H}_1 definita positiva e \mathbf{Q}_1 ortogonale, allora $\mathbf{A}\mathbf{A}^T = \mathbf{H}_1^2$, e quindi

$$\mathbf{H}_1 = \sqrt{\mathbf{A}\mathbf{A}^T} = \mathbf{H}$$

Da questa uguaglianza segue $\mathbf{Q}_1 = \mathbf{H}_1^{-1}\mathbf{A} = \mathbf{Q}$ il che prova l'unicità.

OSSERVAZIONE Analogamente è possibile decomporre una matrice quadrata reale \mathbf{A} nella forma

$$\mathbf{A} = \mathbf{PK}$$

dove \mathbf{P} è ortogonale e $\mathbf{K} = \sqrt{\mathbf{A}^T\mathbf{A}}$ è (semi)definita positiva.

La matrice

$$\mathbf{J} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

soddisfa l'equazione $\mathbf{J}^2 = -\mathbf{I}$. Questo suggerisce di costruire una copia di \mathbb{C} all'interno dell'insieme delle matrici reali 2×2 : a ogni numero complesso $z = a + ib$ associamo la matrice reale $\mathbf{M}(z)$

$$\mathbf{M}(z) = a\mathbf{I} + b\mathbf{J} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

Si verifica facilmente che $\mathbf{M}(z+w) = \mathbf{M}(z) + \mathbf{M}(w)$ e $\mathbf{M}(zw) = \mathbf{M}(z)\mathbf{M}(w)$. L'operazione di coniugio in \mathbb{C} corrisponde alla trasposizione di matrici: $\mathbf{M}(\bar{z}) = \mathbf{M}(z)^T$. Si osservi che $\det(\mathbf{M}(z)) = a^2 + b^2 = |z|^2$, quindi $\mathbf{M}(z)$ è invertibile se $z \neq 0$ e in tal caso $\mathbf{M}(z)^{-1} = \mathbf{M}(z^{-1})$. Chiaramente $\mathbf{M}(z)$ è simmetrica se e solo se $b = 0$ cioè z è reale, e $\mathbf{M}(z)$ è definita positiva se e solo se z è reale positivo ($a > 0$ e $b = 0$). Infine

$$\mathbf{M}(z)^T \mathbf{M}(z) = \mathbf{M}(\bar{z}z) = \begin{bmatrix} a^2 + b^2 & 0 \\ 0 & a^2 + b^2 \end{bmatrix}$$

quindi $\mathbf{M}(z)$ è ortogonale se e solo se $z = e^{i\theta}$ è un numero complesso di modulo uno. Fissiamo $z \neq 0$, e calcoliamo la decomposizione polare $\mathbf{M}(z) = \mathbf{H}\mathbf{Q}$ di $\mathbf{M}(z)$:

$$\mathbf{H} = \sqrt{\mathbf{M}(z)\mathbf{M}(z)^T} = \sqrt{\begin{bmatrix} a^2 + b^2 & 0 \\ 0 & a^2 + b^2 \end{bmatrix}} = \begin{bmatrix} \sqrt{a^2 + b^2} & 0 \\ 0 & \sqrt{a^2 + b^2} \end{bmatrix} = \mathbf{M}(|z|)$$

e

$$\mathbf{Q} = \mathbf{M}(z)\mathbf{H}^{-1} = \mathbf{M}(z)\mathbf{M}(|z|^{-1}) = \mathbf{M}\left(\frac{z}{|z|}\right) = \begin{bmatrix} \frac{a}{\sqrt{a^2+b^2}} & -\frac{b}{\sqrt{a^2+b^2}} \\ \frac{b}{\sqrt{a^2+b^2}} & \frac{a}{\sqrt{a^2+b^2}} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

Quindi la decomposizione polare $\mathbf{H}\mathbf{Q}$ di $\mathbf{M}(z)$ corrisponde esattamente alla decomposizione polare

$$z = |z| \frac{z}{|z|} = \rho e^{i\theta}$$

del numero complesso z nel prodotto del suo modulo $|z| = \rho$ per il numero complesso unitario $\frac{z}{|z|} = e^{i\theta}$.

La matrice pseudoinversa di Moore-Penrose

Sia \mathbf{A} una matrice reale $m \times n$. Moltiplicazione a sinistra per \mathbf{A} induce un'applicazione lineare invertibile $\mathfrak{M} : \text{Row}(\mathbf{A}) \rightarrow \text{Col}(\mathbf{A})$ dallo spazio riga di \mathbf{A} allo spazio colonna di \mathbf{A} . A partire dall'applicazione inversa $\mathfrak{M}^{-1} : \text{Col}(\mathbf{A}) \rightarrow \text{Row}(\mathbf{A})$ si costruisce una matrice \mathbf{A}^+ di tipo $n \times m$ che è il miglior surrogato possibile della matrice inversa di \mathbf{A} .

PROPOSIZIONE 4.6 (Matrice pseudoinversa)

Sia \mathbf{A} una matrice reale di tipo (m, n) . Esiste un'unica matrice reale \mathbf{A}^+ di tipo (n, m) , detta *matrice pseudoinversa o inversa di Moore-Penrose* di \mathbf{A} , tale che

- a) $\mathbf{A}^+ \mathbf{A} \mathbf{x}^+ = \mathbf{x}^+$ per ogni $\mathbf{x}^+ \in \text{Row}(\mathbf{A})$,
- b) $\mathbf{A}^+ \mathbf{y}_0 = \mathbf{0}$ per ogni $\mathbf{y}_0 \in \text{Col}(\mathbf{A})^\perp$.

Valgono le uguaglianze:

$$\text{Row}(\mathbf{A}^+) = \text{Col}(\mathbf{A}), \quad \text{Col}(\mathbf{A}^+) = \text{Row}(\mathbf{A}), \quad \text{Ker}(\mathbf{A}^+) = \text{Col}(\mathbf{A})^\perp$$

La matrice \mathbf{A} è la pseudoinversa di \mathbf{A}^+ : $\mathbf{A} = (\mathbf{A}^+)^+$. In particolare

- c) $\mathbf{A} \mathbf{A}^+ \mathbf{p} = \mathbf{p}$ per ogni $\mathbf{p} \in \text{Col}(\mathbf{A})$.

DIMOSTRAZIONE. Consideriamo l'applicazione lineare $\mathcal{L}_{\mathbf{A}} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ associata alla matrice \mathbf{A} , cioè l'applicazione $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$. Il nucleo e lo spazio riga di \mathbf{A} sono sottospazi del dominio \mathbb{R}^n , e sono uno il complemento ortogonale dell'altro. Lo spazio colonna $\text{Col}(\mathbf{A})$ coincide con l'immagine di $\mathcal{L}_{\mathbf{A}}$ ed è un sottospazio del codominio \mathbb{R}^m . Mostriamo più precisamente che ogni vettore \mathbf{p} dello spazio colonna $\text{Col}(\mathbf{A})$ è della forma $\mathbf{A}\mathbf{x}^+$ per un unico vettore \mathbf{x}^+ dello spazio riga $\text{Row}(\mathbf{A})$. Infatti, se \mathbf{p} appartiene allo spazio colonna di \mathbf{A} , allora esiste $\mathbf{x} \in \mathbb{R}^n$ tale che $\mathbf{p} = \mathbf{A}\mathbf{x}$. Ora $\mathbf{x} \in \mathbb{R}^n$ si decompone come somma $\mathbf{x} = \mathbf{x}^+ + \mathbf{x}_0$ di un vettore $\mathbf{x}^+ \in \text{Row}(\mathbf{A})$ e di un vettore $\mathbf{x}_0 \in \text{Row}(\mathbf{A})^\perp = \text{Ker}(\mathbf{A})$, quindi

$$\mathbf{p} = \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}^+ + \mathbf{A}\mathbf{x}_0 = \mathbf{A}\mathbf{x}^+$$

Il vettore \mathbf{x}^+ dello spazio riga è univocamente determinato: se $\mathbf{p} = \mathbf{A}\mathbf{x}_1$ con \mathbf{x}_1 appartenente allo spazio riga, allora

$$\mathbf{x}^+ - \mathbf{x}_1 \in \text{Row}(\mathbf{A}) \cap \text{Ker}(\mathbf{A}) = \text{Row}(\mathbf{A}) \cap \text{Row}(\mathbf{A})^\perp = \{\mathbf{0}\}$$

e quindi $\mathbf{x}_1 = \mathbf{x}^+$. In breve, abbiamo dimostrato che l'applicazione $\mathbf{x}^+ \mapsto \mathbf{A}\mathbf{x}^+$ dallo spazio riga allo spazio colonna di \mathbf{A} è biettiva.

Mostriamo l'unicità di \mathbf{A}^+ . Il prodotto $\mathbf{A}^+ \mathbf{p}$ è determinato da a) per ogni vettore \mathbf{p} della forma $\mathbf{A}\mathbf{x}^+$ con $\mathbf{x}^+ \in \text{Row}(\mathbf{A})$: per quanto abbiamo appena visto, tali vettori \mathbf{p} sono i vettori dello spazio colonna $\text{Col}(\mathbf{A})$. Per il punto b) il prodotto $\mathbf{A}^+ \mathbf{y}_0$ è determinato anche per ogni vettore \mathbf{y}_0 ortogonale a $\text{Col}(\mathbf{A})$. Siccome ogni vettore $\mathbf{y} \in \mathbb{R}^m$ si scrive nella forma $\mathbf{y} = \mathbf{p} + \mathbf{y}_0$ con $\mathbf{p} \in \text{Col}(\mathbf{A})$ e $\mathbf{y}_0 \in \text{Col}(\mathbf{A})^\perp$, il prodotto $\mathbf{A}^+ \mathbf{y}$ è determinato da a) e b) per ogni $\mathbf{y} \in \mathbb{R}^m$, e quindi la matrice \mathbf{A}^+ è determinata da a) e b).

Mostriamo l'esistenza di \mathbf{A}^+ . Definiamo un'applicazione $\mathcal{L}^+ : \mathbb{R}^m \rightarrow \mathbb{R}^n$ in questo modo: dato $\mathbf{y} \in \mathbb{R}^m$, scriviamo $\mathbf{y} = \mathbf{p} + \mathbf{y}_0$ con $\mathbf{p} \in \text{Col}(\mathbf{A})$ e $\mathbf{y}_0 \in \text{Col}(\mathbf{A})^\perp$. Sappiamo che $\mathbf{p} = \mathbf{A}\mathbf{x}^+$ per un unico $\mathbf{x}^+ \in \text{Row}(\mathbf{A})$. Poniamo $\mathcal{L}^+(\mathbf{p}) = \mathbf{x}^+$ e $\mathcal{L}^+(\mathbf{y}_0) = \mathbf{0}$, e quindi $\mathcal{L}^+(\mathbf{y}) = \mathbf{x}^+$. Si verifica che \mathcal{L}^+ è lineare; quindi esiste una matrice \mathbf{A}^+ tale che $\mathcal{L}^+(\mathbf{y}) = \mathbf{A}^+ \mathbf{y}$ per ogni $\mathbf{y} \in \mathbb{R}^m$. Per costruzione \mathbf{A}^+ soddisfa a) e b). Questo mostra l'esistenza.

Dall'uguaglianza

$$\mathbf{A}^+ \mathbf{y} = \mathbf{A}^+ \mathbf{p} + \mathbf{A}^+ \mathbf{y}_0 = \mathbf{A}^+ \mathbf{A} \mathbf{x}^+ = \mathbf{x}^+$$

segue che l'immagine di \mathcal{L}^- , cioè lo spazio colonna di \mathbf{A}^+ , è lo spazio riga di \mathbf{A} . Inoltre $\mathbf{A}^+ \mathbf{y} = \mathbf{0}$ se e solo se $\mathbf{p} = \mathbf{A}\mathbf{x}^+ = \mathbf{0}$, cioè $\mathbf{y} = \mathbf{y}_0$ appartiene al complemento ortogonale

$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ è $\mathbf{A}^+ = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$ come abbiamo visto a pagina 472. La matrice \mathbf{A} ha autovalori $\lambda_1 = 1$ e $\lambda_2 = 0$, quindi $\mathbf{A} = \mathbf{SDS}^{-1}$ è simile a $\mathbf{D} = \text{diag}(1, 0)$. Ma \mathbf{A}^+ ha l'autovalore $\frac{1}{2}$ e quindi non è simile alla matrice $\mathbf{D}^+ = \text{diag}(1, 0)$, cioè non è della forma $\mathbf{SD}^+\mathbf{S}^{-1}$.

Dalla proposizione 4.7 ricaviamo una formula per la pseudoinversa di una matrice arbitraria in termini di una sua SVD.

COROLLARIO 4.8 (Decomposizione ai valori singolari della pseudoinversa) Se $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ è una decomposizione ai valori singolari di \mathbf{A} , allora

$$(4.7) \quad \mathbf{A}^+ = \mathbf{V}\Sigma^+\mathbf{U}^T$$

è una decomposizione ai valori singolari di \mathbf{A}^+ . In particolare, i valori singolari di \mathbf{A}^+ sono gli inversi dei valori singolari di \mathbf{A} .

DIMOSTRAZIONE. La tesi segue dalla proposizione 4.7 perché in una SVD le matrici \mathbf{U} e \mathbf{V} sono ortogonali. Come abbiamo visto negli esempi, la matrice Σ^+ è la matrice pseudodiagonale i cui elementi non nulli sono gli inversi dei valori singolari di \mathbf{A} .

A pagina 467 abbiamo calcolato una SVD della matrice $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$:

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

La matrice pseudoinversa è perciò:

$$\mathbf{A}^+ = \mathbf{V}\Sigma^+\mathbf{U}^T = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

in accordo con quanto trovato a pagina 472.

Abbiamo calcolato in precedenza la SVD della matrice $\mathbf{A} = \begin{bmatrix} 0 & -2 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}$:

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

Quindi

$$\mathbf{A}^+ = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -1/2 & 0 & 0 \end{bmatrix}$$

Il lettore dovrebbe controllare che $\mathbf{A}^+\mathbf{A} = \mathbf{I}_2$, mentre \mathbf{AA}^+ è la matrice della proiezione ortogonale sul piano di equazione $x_3 = 0$, che è lo spazio colonna di \mathbf{A} .

PROPOSIZIONE 4.9 (Proprietà della matrice pseudoinversa)

Sia \mathbf{A} una matrice di tipo (m, n) , e sia \mathbf{A}^+ la sua matrice pseudoinversa. Allora

- la matrice $\mathbf{A}^+ \mathbf{A}$ rappresenta la proiezione ortogonale di \mathbb{R}^n su $\text{Row}(\mathbf{A})$;
- la matrice $\mathbf{A} \mathbf{A}^+$ rappresenta la proiezione ortogonale di \mathbb{R}^m su $\text{Col}(\mathbf{A})$;
- per ogni $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{x}^+ = \mathbf{A}^+ \mathbf{y}$ è l'unica soluzione ai minimi quadrati di $\mathbf{Ax} = \mathbf{y}$ che appartenga allo spazio riga di \mathbf{A} ;
- se \mathbf{A} ha rango n , allora $\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ è un'inversa sinistra di \mathbf{A} ;
- se \mathbf{A} ha rango m , allora $\mathbf{A}^+ = \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1}$ è un'inversa destra di \mathbf{A} ;
- se \mathbf{A} è quadrata e invertibile, allora $\mathbf{A}^+ = \mathbf{A}^{-1}$.

DIMOSTRAZIONE. Per definizione, per ogni vettore $\mathbf{x}^+ \in \text{Row}(\mathbf{A})$

$$\mathbf{A}^+ \mathbf{A} \mathbf{x}^+ = \mathbf{x}^+$$

mentre per ogni vettore $\mathbf{x}_0 \in \text{Ker}(\mathbf{A})$

$$\mathbf{A}^+ \mathbf{A} \mathbf{x}_0 = \mathbf{A}^+ \mathbf{0} = \mathbf{0}$$

Questo mostra che $\mathbf{A}^+ \mathbf{A}$ è la matrice che rappresenta la proiezione ortogonale di \mathbb{R}^n su $\text{Row}(\mathbf{A})$. D'altra parte, se $\mathbf{p} = \mathbf{Ax}^+$ appartiene allo spazio colonna

$$\mathbf{A} \mathbf{A}^+ \mathbf{p} = \mathbf{A} \mathbf{x}^+ = \mathbf{p}$$

mentre, se \mathbf{y}_0 appartiene al complemento ortogonale dello spazio colonna, cioè a $\text{Ker}(\mathbf{A}^+)$, allora

$$\mathbf{A} \mathbf{A}^+ \mathbf{y}_0 = \mathbf{A} \mathbf{0} = \mathbf{0}$$

Questo mostra che $\mathbf{A} \mathbf{A}^+$ è la matrice che rappresenta la proiezione ortogonale di \mathbb{R}^m su $\text{Col}(\mathbf{A})$.

Mostriamo c): fissato $\mathbf{y} \in \mathbb{R}^n$, sia \mathbf{p} la proiezione ortogonale di \mathbf{y} sullo spazio colonna e sia $\mathbf{x}^+ = \mathbf{A}^+ \mathbf{y}$. Allora $\mathbf{Ax}^+ = \mathbf{p}$, cioè \mathbf{x}^+ è una soluzione ai minimi quadrati di $\mathbf{Ax} = \mathbf{y}$. Siccome l'applicazione $\mathbf{x}^+ \mapsto \mathbf{Ax}^+$ è iniettiva sullo spazio riga, si tratta dell'unica soluzione ai minimi quadrati che appartenga allo spazio riga.

Se $r(\mathbf{A}) = n$, allora $\text{Row}(\mathbf{A}) = \mathbb{R}^n$, e quindi a) implica $\mathbf{A}^+ \mathbf{A} = \mathbf{I}_n$, cioè \mathbf{A}^+ è un'inversa sinistra di \mathbf{A} . La matrice $\mathbf{A}^T \mathbf{A}$ è quadrata di ordine n e ha rango

$$r(\mathbf{A}^T \mathbf{A}) = r(\mathbf{A}) = n$$

Perciò $\mathbf{A}^T \mathbf{A}$ è invertibile, e possiamo considerare la matrice $\mathbf{B} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$. Evidentemente anche \mathbf{B} è un'inversa sinistra di \mathbf{A} : $\mathbf{BA} = \mathbf{I}_n$. Questo implica che

$$\mathbf{B} \mathbf{A} \mathbf{x}^+ = \mathbf{x}^+$$

per ogni \mathbf{x}^+ appartenente allo spazio riga di \mathbf{A} . D'altra parte, se \mathbf{y}_0 è ortogonale allo spazio colonna di \mathbf{A} , allora $\mathbf{y}_0 \in \text{Ker}(\mathbf{A}^T)$ e quindi

$$\mathbf{B} \mathbf{y}_0 = \mathbf{0}$$

Quindi $\mathbf{B} = \mathbf{A}^+$ per l'unicità della matrice pseudoinversa. Analogamente si mostra e), ed f) segue da d).

OSSERVAZIONE Quando $r(\mathbf{A}) < n$, le soluzioni ai minimi quadrati di $\mathbf{Ax} = \mathbf{b}$ sono infinite, e precisamente sono i vettori \mathbf{x} della forma

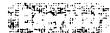
$$\mathbf{x} = \mathbf{A}^+ \mathbf{b} + \mathbf{x}_0 \quad \text{al variare di } \mathbf{x}_0 \text{ in } \text{Ker}(\mathbf{A})$$

Tra tutte queste la soluzione $\mathbf{x}^+ = \mathbf{A}^+ \mathbf{b}$ è quella che ha norma minima: infatti \mathbf{x}^+ appartiene allo spazio riga, ed è quindi ortogonale ai vettori \mathbf{x}_0 del nucleo; per il teorema di Pitagora

$$\|\mathbf{x}\|^2 = \|\mathbf{x}^+\|^2 + \|\mathbf{x}_0\|^2 \geq \|\mathbf{x}^+\|^2$$

In altri termini, \mathbf{x}^+ è tra le soluzioni ai minimi quadrati di $\mathbf{Ax} = \mathbf{y}$ quella a distanza minima dall'origine. Per questo a volte si dice che \mathbf{x}^+ è la *soluzione ottima* ai minimi quadrati di $\mathbf{Ax} = \mathbf{b}$.

OSSERVAZIONE Le due matrici $\mathbf{A}^+ \mathbf{A}$ e $\mathbf{A} \mathbf{A}^+$ sono simmetriche e idempotenti, come ogni matrice di una proiezione ortogonale.



💡 Sia $\mathbf{w} = [a, b, c]^T$ un vettore arbitrario di \mathbb{R}^3 . Si scrivano le matrici $\mathbf{w}^T \mathbf{w}$ e $\mathbf{w} \mathbf{w}^T$ e se ne calcolino gli autovalori. Quali sono i valori singolari di $\mathbf{A} = \mathbf{w}$? Mostrare che, se $\mathbf{w} \neq 0$, allora $\mathbf{w}^+ = \frac{1}{\|\mathbf{w}\|^2} \mathbf{w}^T$.

💡 Si consideri la matrice

$$\mathbf{A} = \begin{bmatrix} -2 & 2 & 1 \\ 1 & -1 & 2 \end{bmatrix}$$

Per determinare i valori singolari di \mathbf{A} conviene calcolare $\mathbf{A}^T \mathbf{A}$ o $\mathbf{A} \mathbf{A}^T$? Si determinino i valori singolari, una SVD e la matrice pseudoinversa di \mathbf{A} .

💡 Sia \mathbf{A} una matrice di tipo (m, n) e sia \mathbf{B} una matrice di tipo (n, m) . Mostrare che \mathbf{AB} e \mathbf{BA} hanno gli stessi autovalori non nulli. Dato un autovalore non nullo λ di \mathbf{AB} , siano \mathbf{V}_λ e \mathbf{W}_λ gli autospazi relativi a λ delle matrici \mathbf{AB} e \mathbf{BA} . Mostrare che l'applicazione lineare $\mathbf{v} \mapsto \mathbf{Bv}$ da \mathbf{V}_λ a \mathbf{W}_λ è invertibile e che la sua inversa è l'applicazione $\mathbf{w} \mapsto \frac{\mathbf{Aw}}{\lambda}$.

💡 Si calcoli la matrice pseudoinversa della matrice

$$\mathbf{A} = \begin{bmatrix} 5 & -2 & 1 \\ -2 & 2 & 2 \\ 1 & 2 & 5 \end{bmatrix}$$

Per ogni $\mathbf{b} \in \mathbb{R}^3$ qual è la soluzione ottimale ai minimi quadrati di $\mathbf{Ax} = \mathbf{b}$?

💡 Si definisca la *norma* di una matrice \mathbf{A} di tipo (m, n) nel modo seguente:

$$\|\mathbf{A}\| = \text{Sup} \{ \|\mathbf{Ax}\| : \mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\| = 1 \} = \text{Sup} \left\{ \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|} : \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq 0 \right\}$$

Mostrare che $\|\mathbf{A}\| = \sigma_1$: la norma di una matrice (non nulla) coincide col massimo dei suoi valori singolari. Mostrare anche che l'estremo superiore che definisce la norma è un massimo, cioè che esiste un versore \mathbf{x} tale che $\|\mathbf{Ax}\| = \|\mathbf{A}\|$. Dimostrare la diseguaglianza triangolare $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$.

Il numero di condizionamento di una matrice quadrata invertibile \mathbf{A} è

$$c = \|\mathbf{A}\| \|\mathbf{A}^{-1}\|$$

Questo numero è di fondamentale importanza nelle applicazioni numeriche perché controlla la propagazione degli errori dai dati alla soluzione di un sistema lineare. Mostrare che c è uguale al rapporto tra il massimo e il minimo valore singolare di \mathbf{A} .

5 IL CASO COMPLESSO

In questo paragrafo dimostriamo la versione complessa del teorema spettrale. Si ricordi che, data una matrice complessa \mathbf{A} , la matrice \mathbf{A}^H è la matrice trasposta coniugata di \mathbf{A} , e che l'analogo complesso di una matrice ortogonale è una matrice unitaria, cioè una matrice invertibile \mathbf{U} tale che $\mathbf{U}^H = \mathbf{U}^{-1}$. Una matrice è unitaria se e solo se le sue colonne formano una base ortonormale di \mathbb{C}^n . Ci domandiamo quali siano le matrici complesse che siano *unitariamente diagonalizzabili* o, equivalentemente, quali siano le matrici complesse per le quali esiste una base ortonormale di \mathbb{C}^n formata da autovettori della matrice. La risposta è che le matrici unitariamente diagonalizzabili sono le matrici \mathbf{N} che commutano con la propria trasposta coniugata: $\mathbf{N}^H \mathbf{N} = \mathbf{N} \mathbf{N}^H$. Tali matrici si dicono *normali* e formano una classe molto più ampia delle matrici simmetriche: per esempio, le matrici ortogonali reali e le matrici antisimmetriche reali sono normali, e il teorema spettrale complesso ha applicazioni anche allo studio di queste importanti matrici.

Dedurremo il teorema spettrale complesso da un teorema noto come *Lemma di Schur*, che è un risultato importante di per sé.

TEOREMA 5.1 (Lemma di Schur)

Sia \mathbf{A} una matrice quadrata complessa. Allora \mathbf{A} è *unitariamente simile* a una matrice triangolare alta. Questo significa che esiste una matrice unitaria \mathbf{U} tale che $\mathbf{T} = \mathbf{U}^H \mathbf{A} \mathbf{U}$ è una matrice triangolare alta.

OSSERVAZIONE Si noti che, se \mathbf{U} è unitaria, allora $\mathbf{U}^H = \mathbf{U}^{-1}$ e quindi $\mathbf{T} = \mathbf{U}^H \mathbf{A} \mathbf{U}$ è effettivamente *simile* ad \mathbf{A} . Nel Lemma di Schur non ci sono ipotesi su \mathbf{A} : ogni matrice quadrata complessa è simile a una matrice triangolare. Il teorema si applica anche a matrici quadrate \mathbf{A} reali (i cui elementi cioè sono numeri reali), perché una matrice reale è un caso particolare di una matrice complessa; però, anche se \mathbf{A} è reale, in generale \mathbf{U} e \mathbf{T} sono complesse. Con una ipotesi importante in più si può dimostrare una versione reale del Lemma di Schur: *se \mathbf{A} è una matrice reale e tutti gli autovalori di \mathbf{A} sono reali, allora esiste una matrice reale ortogonale \mathbf{Q} tale che $\mathbf{T} = \mathbf{Q}^T \mathbf{A} \mathbf{Q}$ è triangolare* (in questo caso anche la matrice \mathbf{T} è reale, perché \mathbf{A} e \mathbf{Q} sono reali).

DIMOSTRAZIONE DEL LEMMA DI SCHUR. La dimostrazione è del tutto simile a quella del teorema spettrale e procede per induzione sull'ordine n della matrice. Il caso iniziale $n = 1$ è ovvio.

Sia ora $n \geq 2$ e supponiamo che il teorema sia vero per matrici quadrate di ordine $n - 1$: dobbiamo dimostrarlo per una matrice quadrata \mathbf{A} di ordine n . Il polinomio caratteristico

$\det(\mathbf{A} - \lambda \mathbf{I}) = 0$ ha una radice $\lambda_1 \in \mathbb{C}$. Il numero complesso λ_1 è un autovalore di \mathbf{A} , ed esiste perciò un autovettore $\mathbf{v}_1 \in \mathbb{C}^n$ relativo a λ_1 . Il vettore $\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$ è anch'esso un autovettore di \mathbf{A} relativo a λ_1 e ha norma 1. Possiamo trovare una base di \mathbb{C}^n che abbia \mathbf{u}_1 come primo elemento: mediante l'algoritmo di Gram-Schmidt trasformiamo tale base in una base ortonormale $\{\mathbf{u}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ di \mathbb{C}^n . Sia $\mathbf{P} = [\mathbf{q}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_n]$ la corrispondente matrice unitaria. La matrice $\mathbf{C} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{P}^H\mathbf{A}\mathbf{P}$ rappresenta l'applicazione lineare $\mathcal{L}_{\mathbf{A}}$ rispetto alla base $\{\mathbf{u}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$; in particolare, la prima colonna di \mathbf{C} è il vettore delle coordinate di $\mathcal{L}_{\mathbf{A}}(\mathbf{u}_1)$ rispetto a tale base, cioè

$$\mathcal{L}_{\mathbf{A}}(\mathbf{u}_1) = c_{11}\mathbf{u}_1 + c_{21}\mathbf{b}_2 + \cdots + c_{n1}\mathbf{b}_n$$

D'altra parte, $\mathcal{L}_{\mathbf{A}}(\mathbf{u}_1) = \mathbf{A}\mathbf{u}_1 = \lambda_1\mathbf{u}_1$. Confrontando queste due espressioni concludiamo che $c_{11} = \lambda_1$ e $c_{j1} = 0$ per $j = 2, \dots, n$. La matrice \mathbf{C} è quindi della forma

$$\mathbf{C} = \left[\begin{array}{c|c} \lambda_1 & \mathbf{w}^T \\ \hline \mathbf{0} & \mathbf{A}_1 \end{array} \right]$$

dove \mathbf{w} è un vettore di \mathbb{C}^{n-1} che non ci interessa specificare, $\mathbf{0}$ denota il vettore nullo di \mathbb{C}^{n-1} , e \mathbf{A}_1 è una matrice quadrata di ordine $n-1$. Per l'ipotesi di induzione \mathbf{A}_1 è unitariamente simile a una matrice triangolare alta: esiste una matrice unitaria \mathbf{U}_1 di ordine $n-1$ tale che $\mathbf{U}_1^H \mathbf{A}_1 \mathbf{U}_1 = \mathbf{T}_1$ è triangolare alta. La matrice

$$\mathbf{R} = \left[\begin{array}{c|c} 1 & \mathbf{0}^T \\ \hline \mathbf{0} & \mathbf{U}_1 \end{array} \right]$$

è unitaria perché \mathbf{U}_1 lo è, e

$$\mathbf{R}^H \mathbf{C} \mathbf{R} = \left[\begin{array}{c|c} 1 & \mathbf{0}^T \\ \hline \mathbf{0} & \mathbf{U}_1^H \end{array} \right] \left[\begin{array}{c|c} \lambda_1 & \mathbf{w}^T \\ \hline \mathbf{0} & \mathbf{A}_1 \end{array} \right] \left[\begin{array}{c|c} 1 & \mathbf{0}^T \\ \hline \mathbf{0} & \mathbf{U}_1 \end{array} \right] = \left[\begin{array}{c|c} \lambda_1 & \mathbf{w}^T \mathbf{U}_1 \\ \hline \mathbf{0} & \mathbf{U}_1^H \mathbf{A}_1 \mathbf{U}_1 \end{array} \right] = \left[\begin{array}{c|c} \lambda_1 & \mathbf{w}^T \mathbf{U}_1 \\ \hline \mathbf{0} & \mathbf{T}_1 \end{array} \right] = \mathbf{T}$$

La matrice \mathbf{T} è triangolare alta perché \mathbf{T}_1 lo è. Infine sia $\mathbf{U} = \mathbf{P}\mathbf{R}$. La matrice \mathbf{U} è unitaria perché prodotto di matrici unitarie, e

$$\mathbf{U}^H \mathbf{A} \mathbf{U} = \mathbf{R}^H \mathbf{P}^H \mathbf{A} \mathbf{P} \mathbf{R} = \mathbf{R}^H \mathbf{C} \mathbf{R} = \mathbf{T}$$

Siccome \mathbf{U} è unitaria e \mathbf{T} è triangolare alta, questo completa la dimostrazione.

DEFINIZIONE 5.2 (Matrice normale)

Una matrice quadrata complessa \mathbf{N} si dice *normale* se commuta con la sua trasposta coniugata:

$$\mathbf{N}^H \mathbf{N} = \mathbf{N} \mathbf{N}^H$$

Una matrice simmetrica (rispettivamente antisimmetrica, rispettivamente ortogonale) reale \mathbf{A} è normale. Infatti in questo caso $\mathbf{A}^H = \mathbf{A}^T = \mathbf{A}$ (rispettivamente $\mathbf{A}^H = \mathbf{A}^T = -\mathbf{A}$, rispettivamente $\mathbf{A}^H = \mathbf{A}^T = \mathbf{A}^{-1}$) e quindi \mathbf{A}^H commuta con \mathbf{A} . Una matrice unitaria è normale perché $\mathbf{U}^H = \mathbf{U}^{-1}$ commuta con \mathbf{U} . Una matrice diagonale (anche se a coefficienti complessi) è normale.

TEOREMA 5.3 (Teorema spettrale complesso)

Una matrice quadrata \mathbf{N} è normale se e solo se è unitariamente diagonalizzabile, cioè esiste una matrice unitaria \mathbf{U} tale che $\mathbf{D} = \mathbf{U}^H \mathbf{N} \mathbf{U}$ sia diagonale.

DIMOSTRAZIONE. Cominciamo col mostrare che: *una matrice unitariamente simile a una matrice normale è normale*. Supponiamo infatti che \mathbf{A} sia unitariamente simile a una matrice normale \mathbf{N} : questo significa che esiste una matrice unitaria \mathbf{U} tale che

$$\mathbf{N} = \mathbf{U}^{-1} \mathbf{A} \mathbf{U} = \mathbf{U}^T \mathbf{A} \mathbf{U}$$

Allora $\mathbf{A} = \mathbf{U} \mathbf{N} \mathbf{U}^H$, e quindi:

$$\begin{aligned}\mathbf{A}^H \mathbf{A} &= (\mathbf{U} \mathbf{N} \mathbf{U}^H)^H (\mathbf{U} \mathbf{N} \mathbf{U}^H) = (\mathbf{U}^H)^H \mathbf{N}^H \mathbf{U}^H \mathbf{U} \mathbf{N} \mathbf{U}^H = \mathbf{U} \mathbf{N}^H \mathbf{N} \mathbf{U}^H \\ \mathbf{A} \mathbf{A}^H &= (\mathbf{U} \mathbf{N} \mathbf{U}^H) (\mathbf{U} \mathbf{N} \mathbf{U}^H)^H = \mathbf{U} \mathbf{N} \mathbf{U}^H \mathbf{U}^H \mathbf{N}^H \mathbf{U}^H = \mathbf{U} \mathbf{N}^H \mathbf{N} \mathbf{U}^H\end{aligned}$$

Ma $\mathbf{N}^H \mathbf{N} = \mathbf{N} \mathbf{N}^H$ perché \mathbf{N} è normale, e quindi $\mathbf{A}^H \mathbf{A} = \mathbf{A} \mathbf{A}^H$, cioè \mathbf{A} è normale.

Supponiamo ora che \mathbf{A} sia unitariamente diagonalizzabile: questo significa che \mathbf{A} è unitariamente simile alla matrice diagonale \mathbf{D} . Siccome una matrice diagonale è normale, anche \mathbf{A} è normale per quanto abbiamo appena mostrato.

Viceversa, supponiamo che \mathbf{N} sia una matrice normale. Per il Lemma di Schur \mathbf{N} è unitariamente simile a una matrice triangolare alta \mathbf{T} . Basta ora mostrare che \mathbf{T} è diagonale. Siccome \mathbf{N} è normale, anche \mathbf{T} è normale. La tesi ora segue dal fatto che:

Sia \mathbf{T} una matrice triangolare alta che sia anche normale. Allora \mathbf{T} è diagonale.

Dimostriamo questa affermazione per induzione sull'ordine di \mathbf{T} . Se la matrice è 1×1 l'enunciato è ovvio. Supponiamo che l'enunciato sia vero per matrici di ordine $n-1$, e dimostriamolo per matrici di ordine n . Sia quindi \mathbf{T} una matrice triangolare alta di ordine n che sia anche normale. Scriviamo \mathbf{T} nella forma a blocchi:

$$(5.1) \quad \mathbf{T} = \begin{bmatrix} a & \mathbf{w}^T \\ \mathbf{0} & \mathbf{T}_1 \end{bmatrix}$$

dove a è uno scalare, \mathbf{w} è un vettore di \mathbb{C}^{n-1} , $\mathbf{0}$ è il vettore nullo di \mathbb{C}^{n-1} , e \mathbf{T}_1 è una matrice triangolare alta di ordine $n-1$.

Osserviamo che, siccome \mathbf{T} è normale,

$$\|\mathbf{T}\mathbf{v}\|^2 = \mathbf{v}^H \mathbf{T}^H \mathbf{T} \mathbf{v} = \mathbf{v}^H \mathbf{T} \mathbf{T}^H \mathbf{v} = \|\mathbf{T}^H \mathbf{v}\|^2 \quad \text{per ogni } \mathbf{v} \in \mathbb{C}^n$$

Applichiamo questa uguaglianza a $\mathbf{v} = \mathbf{e}_1$, il primo vettore della base canonica. Allora $\mathbf{T}\mathbf{e}_1$ è la prima colonna di \mathbf{T} , mentre $\mathbf{T}^H \mathbf{e}_1$ è il coniugato della prima riga di \mathbf{T} , quindi

$$|a|^2 = \|\mathbf{T}\mathbf{e}_1\|^2 = \|\mathbf{T}^H \mathbf{e}_1\|^2 = |a|^2 + \|\mathbf{w}\|^2$$

Questo mostra che la norma di \mathbf{w} è zero, e perciò $\mathbf{w} = \mathbf{0}$. È ora immediato verificare che \mathbf{T}_1 dev'essere normale. Dall'ipotesi di induzione segue che \mathbf{T}_1 è diagonale. Siccome $\mathbf{w} = \mathbf{0}$, la (5.1) mostra ora che \mathbf{T} è diagonale e questo conclude la dimostrazione.

COROLLARIO 5.4 Autovettori di una matrice normale relativi ad autovalori distinti sono ortogonali.

DIMOSTRAZIONE. Supponiamo che \mathbf{N} sia una matrice normale di ordine n , e che $\mathbf{w}, \mathbf{z} \in \mathbb{C}^n$ siano autovettori di \mathbf{N} relativi ad autovalori distinti λ e μ . Per il teorema spettrale esiste una base ortonormale \mathcal{B} di \mathbb{C}^n formata da autovettori di \mathbf{N} . Siano g_λ e g_μ le molteplicità geometriche di λ e μ rispettivamente, cioè il massimo numero di autovettori indipendenti relativi a λ e a μ . La base \mathcal{B} , come ogni base formata da autovettori, contiene g_λ autovettori $\mathbf{u}_1, \dots, \mathbf{u}_{g_\lambda}$ relativi a λ , e g_μ autovettori $\mathbf{v}_1, \dots, \mathbf{v}_{g_\mu}$ relativi a μ . L'autovettore \mathbf{w} relativo a λ è combinazione lineare degli autovettori \mathbf{u}_j , l'autovettore \mathbf{z} relativo a μ è combinazione lineare degli autovettori \mathbf{v}_k . Siccome la base \mathcal{B} è ortonormale, i vettori \mathbf{u}_j sono ortogonali a tutti i vettori \mathbf{v}_k , e quindi \mathbf{w} è ortogonale a \mathbf{z} .

L'analogo complesso di una matrice simmetrica è una matrice hermitiana:

DEFINIZIONE 5.5 (Matrice hermitiana. Matrice antihermitiana)

Una matrice quadrata \mathbf{A} si dice *hermitiana* o *autoaggiunta* se coincide con la sua coniugata trasposta (o aggiunta): $\mathbf{A}^H = \mathbf{A}$. Una matrice \mathbf{A} si dice *antihermitiana* se $\mathbf{A}^H = -\mathbf{A}$.

OSSERVAZIONE Una matrice a coefficienti reali è (anti)hermitiana se e solo se è (anti)simmetrica. Una matrice (anti)hermitiana è evidentemente normale. Le matrici hermitiane (o autoaggiunte) giocano un ruolo fondamentale nella formulazione matematica della meccanica quantistica. Anche lo studio delle matrici antisimmetriche ha le sue radici in fisica, in particolare nella meccanica classica.

OSSERVAZIONE Un matrice \mathbf{A} è antihermitiana se e solo se $i\mathbf{A}$ è hermitiana, poiché

$$(i\mathbf{A})^H = -i\mathbf{A}^H$$

Ogni matrice quadrata è somma (in un unico modo) di una matrice hermitiana e di una matrice antihermitiana:

$$\mathbf{A} = \frac{\mathbf{A} + \mathbf{A}^H}{2} + \frac{\mathbf{A} - \mathbf{A}^H}{2}$$

La matrice \mathbf{A} è normale se e solo se la sua parte hermitiana $\frac{\mathbf{A} + \mathbf{A}^H}{2}$ e la sua parte antihermitiana $\frac{\mathbf{A} - \mathbf{A}^H}{2}$ commutano.

COROLLARIO 5.6 (Autovalori di matrici hermitiane e antihermitiane)

Sia \mathbf{A} una matrice quadrata di ordine n . Allora \mathbf{A} è hermitiana (rispettivamente antihermitiana) se e solo se esiste una matrice unitaria \mathbf{U} tale che

$$\mathbf{U}^H \mathbf{A} \mathbf{U} = \text{diag}(\lambda_1, \dots, \lambda_n)$$

dove $\lambda_1, \dots, \lambda_n$ sono numeri reali (rispettivamente immaginari puri). In particolare, gli autovalori di una matrice hermitiana sono reali, e gli autovalori di una matrice antihermitiana (a fortiori, di una matrice antisimmetrica reale) sono immaginari puri.

DIMOSTRAZIONE. Sia \mathbf{A} una matrice hermitiana. Allora \mathbf{A} è normale, e quindi per il teorema spettrale esiste una matrice unitaria \mathbf{U} tale che $\mathbf{U}^H \mathbf{A} \mathbf{U} = \mathbf{D}$ sia diagonale. Allora

$$\mathbf{D}^H = (\mathbf{U}^H \mathbf{A} \mathbf{U})^H = \mathbf{U}^H \mathbf{A}^H \mathbf{U} = \mathbf{U}^H \mathbf{A} \mathbf{U} = \mathbf{D}$$

Ma per una matrice diagonale

$$\mathbf{D}^H = (\text{diag}(\lambda_1, \dots, \lambda_n))^H = \text{diag}(\overline{\lambda_1}, \dots, \overline{\lambda_n})$$

Quindi da $\mathbf{D}^H = \mathbf{D}$ segue

$$\lambda_1 = \overline{\lambda_1}, \dots, \lambda_n = \overline{\lambda_n}$$

cioè gli autovalori λ_k sono numeri reali. Viceversa, se i λ_k sono reali, allora $\mathbf{D}^H = \mathbf{D}$ e quindi $\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{U}^H$ è hermitiana.

Il caso antihermitiano segue da quello hermitiano, perché \mathbf{A} è antihermitiana se e solo se $i\mathbf{A}$ è hermitiana. ■



La matrice $\mathbf{A} = \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}$ è hermitiana. Il suo polinomio caratteristico è $\lambda^2 - 1$, e i suoi autovalori sono $\lambda_1 = 1$ e $\lambda_2 = -1$. Un autovettore relativo a $\lambda_1 = 1$ è $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -i \end{bmatrix}$, un autovettore relativo a $\lambda_2 = -1$ è $\mathbf{v}_2 = \begin{bmatrix} 1 \\ i \end{bmatrix}$. I due autovalori sono reali, e i due autovettori corrispondenti sono perpendicolari in \mathbb{C}^2 :

$$\mathbf{v}_2^H \mathbf{v}_1 = [1 \ -i] \begin{bmatrix} 1 \\ -i \end{bmatrix} = 1 - 1 = 0$$



Moltiplicando per i la matrice hermitiana dell'esempio precedente si ottiene la matrice antisimmetrica reale

$$\mathbf{E} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

Il suo polinomio caratteristico è $\lambda^2 + 1$, e i suoi autovalori $\lambda_1 = i$ e $\lambda_2 = -i$ sono numeri immaginari puri.



Sia \mathbf{A} una matrice quadrata complessa. Mostrare che $\mathbf{A} + \mathbf{A}^H$ e $\mathbf{A}^H \mathbf{A}$ sono hermitiane.

Mostrare che una matrice quadrata \mathbf{A} è normale se e solo se la sua parte hermitiana $\frac{\mathbf{A} + \mathbf{A}^H}{2}$ e la sua parte antihermitiana $\frac{\mathbf{A} - \mathbf{A}^H}{2}$ commutano.

Sia \mathbf{Q} la matrice di una rotazione del piano. Si trovi una matrice unitaria che diagonalizza \mathbf{Q} .

Si trovi una matrice unitaria che diagonalizza la matrice $\mathbf{E} = \begin{bmatrix} 0 & 4 \\ -4 & 0 \end{bmatrix}$.

Si mostri che una matrice 2×2

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

è hermitiana se e solo se a e d sono reali e $c = \bar{b}$. Se questo è il caso, trovare gli autovalori di \mathbf{A} e verificare che sono reali.

Mostrare che una matrice quadrata \mathbf{N} di ordine n è normale se e solo se $\|\mathbf{N}\mathbf{v}\| = \|\mathbf{N}^H\mathbf{v}\|$ per ogni $\mathbf{v} \in \mathbb{C}^n$. Dedurre che, se \mathbf{N} è normale e \mathbf{v} è un autovettore di \mathbf{N} relativo all'autovalore λ , allora \mathbf{v} è un autovettore di \mathbf{N}^H relativo all'autovalore $\bar{\lambda}$.

 Dimostrare senza usare il teorema spettrale che autovettori relativi ad autovalori distinti di una matrice normale sono ortogonali.

Dare un'altra dimostrazione del teorema spettrale complesso facendo vedere che tutti gli autovalori di una matrice normale sono regolari, e mostrando che autovettori relativi ad autovalori distinti sono ortogonali in \mathbb{C}^n .

Suggerimento: se \mathbf{N} è normale, allora anche $\mathbf{N} - \lambda \mathbf{I}$ è normale. Per mostrare che gli autovalori sono regolari, cioè hanno indice uno, basta allora mostrare che $\text{Ker}(\mathbf{B}) \cap \text{Col}(\mathbf{B}) = \{\mathbf{0}\}$ per ogni matrice normale \mathbf{B} . Per questo si usa il trucco seguente: se $\mathbf{v} \in \text{Ker}(\mathbf{B})$ e \mathbf{B} è normale, allora \mathbf{v} appartiene anche a $\text{Ker}(\mathbf{B}^H)$. Invece se $\mathbf{v} \in \text{Col}(\mathbf{B})$, allora \mathbf{v} è ortogonale a ogni vettore in $\text{Ker}(\mathbf{B}^H)$ (questo è vero anche se \mathbf{B} non è normale).

6 MATRICI NORMALI REALI

Sia \mathbf{A} una matrice quadrata reale di ordine n . Il polinomio caratteristico di \mathbf{A} è un polinomio $P(\lambda)$ a coefficienti reali di grado n . Siccome i coefficienti di $P(\lambda)$ sono reali,

$$P(\bar{\lambda}) = \overline{P(\lambda)} \quad \text{per ogni } \lambda \in \mathbb{C}$$

Quindi, se $\lambda_0 \in \mathbb{C}$ è un autovalore di \mathbf{A} , cioè $P(\lambda_0) = 0$, allora anche $\bar{\lambda}_0$ è un autovalore di \mathbf{A} . Se $\lambda_0 = \bar{\lambda}_0$, allora λ_0 è un numero reale: altrimenti $\lambda_0 \neq \bar{\lambda}_0$, e abbiamo così una coppia di autovalori distinti non reali. Inoltre λ_0 e $\bar{\lambda}_0$ hanno la stessa molteplicità come radici di $P(\lambda)$. Quindi gli autovalori di \mathbf{A} si possono suddividere in s coppie di numeri complessi non reali $\lambda_1, \bar{\lambda}_1, \dots, \lambda_s, \bar{\lambda}_s$, e in un gruppo di autovalori reali μ_1, \dots, μ_t , e $n = 2s + t$ perché un polinomio di grado n ha esattamente n radici complesse. Per esempio, il polinomio

$$\lambda^3 + 1 = (\lambda + 1)(\lambda^2 - \lambda + 1)$$

ha le due radici complesse non reali $\lambda_1 = \frac{1}{2} + i\frac{\sqrt{3}}{2}$ e $\lambda_2 = \frac{1}{2} - i\frac{\sqrt{3}}{2}$, e una radice reale $\mu_1 = -1$.

TEOREMA 6.1 (Forma canonica di una matrice normale reale)

Sia \mathbf{N} una matrice normale reale. Siano

$$a_k \pm ib_k \quad (k = 1, 2, \dots, s)$$

i $2s$ autovalori di \mathbf{N} con parte immaginaria non nulla, e siano μ_1, \dots, μ_t gli autovalori reali di \mathbf{N} . Allora esiste una matrice ortogonale \mathbf{Q} tale che $\mathbf{B} = \mathbf{Q}^T \mathbf{N} \mathbf{Q}$ è la matrice diagonale a blocchi

$$(6.1) \quad \mathbf{B} = \text{diag}(\mathbf{D}_1, \dots, \mathbf{D}_s, \mu_1, \dots, \mu_t)$$

dove

$$(6.2) \quad \mathbf{D}_k = \begin{bmatrix} a_k & -b_k \\ b_k & a_k \end{bmatrix}$$

DIMOSTRAZIONE. Il punto fondamentale della dimostrazione è il seguente

Fatto 1 *Sia $\lambda = a - ib$ un autovalore non reale di \mathbf{N} e sia $\mathbf{v} = \mathbf{x} + i\mathbf{y}$ la decomposizione in parte reale e immaginaria di un autovettore \mathbf{v} relativo a λ . Allora \mathbf{x} e \mathbf{y} hanno la stessa norma e sono ortogonali tra loro. Inoltre*

$$(6.3) \quad \mathbf{N}\mathbf{x} = a\mathbf{x} + b\mathbf{y}, \quad \mathbf{N}\mathbf{y} = -b\mathbf{x} + a\mathbf{y}$$

Dimostriamo questa affermazione: fissiamo un autovalore $\lambda = a - ib$ con $b \neq 0$ (il segno meno non è un errore di stampa), e un autovettore $\mathbf{v} \in \mathbb{C}^n$ relativo a λ : \mathbf{v} è non nullo e $\mathbf{N}\mathbf{v} = \lambda\mathbf{v}$. Prendendo i coniugati troviamo

$$\overline{\mathbf{N}\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}$$

D'altra parte, siccome \mathbf{N} è reale,

$$\overline{\mathbf{N}\mathbf{v}} = \overline{\mathbf{N}\overline{\mathbf{v}}} = \mathbf{N}\overline{\mathbf{v}}$$

Quindi $\overline{\mathbf{v}}$ è un autovalore di \mathbf{N} relativo all'autovalore $\bar{\lambda}$. Siccome λ non è reale, $\lambda \neq \bar{\lambda}$ e perciò \mathbf{v} e $\overline{\mathbf{v}}$ sono ortogonali in quanto autovettori relativi ad autovalori distinti di una matrice normale.

Il prodotto hermitiano di \mathbf{v} e $\overline{\mathbf{v}}$ è

$$\mathbf{v}^H \mathbf{v} = (\mathbf{x} + i\mathbf{y})^T (\mathbf{x} + i\mathbf{y}) = \mathbf{x}^T \mathbf{x} - \mathbf{y}^T \mathbf{y} + i(\mathbf{y}^T \mathbf{x} + \mathbf{x}^T \mathbf{y}) = \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2 + 2i\mathbf{x}^T \mathbf{y}$$

Siccome \mathbf{v} e $\overline{\mathbf{v}}$ sono ortogonali, il loro prodotto hermitiano è nullo, quindi $\|\mathbf{x}\|^2 = \|\mathbf{y}\|^2$ e $\mathbf{x}^T \mathbf{y} = 0$. Questo mostra che \mathbf{x} e \mathbf{y} hanno la stessa norma e sono ortogonali tra loro.

Siccome \mathbf{N} è reale, la parte reale e la parte immaginaria di $\mathbf{N}\mathbf{v}$ sono $\mathbf{N}\mathbf{x}$ e $\mathbf{N}\mathbf{y}$ rispettivamente. D'altra parte

$$\lambda\mathbf{v} = (a - ib)(\mathbf{x} + i\mathbf{y}) = a\mathbf{x} + b\mathbf{y} + i(-b\mathbf{x} + a\mathbf{y})$$

Quindi $\mathbf{N}\mathbf{x} = a\mathbf{x} + b\mathbf{y}$ e $\mathbf{N}\mathbf{y} = -b\mathbf{x} + a\mathbf{y}$, e questo conclude la dimostrazione del fatto 1.

Fatto 2 *Siano \mathbf{v} e \mathbf{w} due vettori di \mathbb{C}^n tali che \mathbf{v} e $\overline{\mathbf{v}}$ siano entrambi ortogonali a \mathbf{w} . Allora la parte reale e la parte immaginaria di \mathbf{v} sono ortogonali sia alla parte reale sia alla parte immaginaria di \mathbf{w} .*

Dimostriamo questo fatto. Siano $\mathbf{v} = \mathbf{x} + i\mathbf{y}$ e $\mathbf{w} = \mathbf{p} + i\mathbf{q}$ le decomposizioni in parte reale e immaginaria di \mathbf{v} e \mathbf{w} . Per ipotesi il prodotto hermitiano di \mathbf{v} e \mathbf{w} è nullo:

$$0 = \mathbf{w}^H \mathbf{v} = (\mathbf{p} - i\mathbf{q})^T (\mathbf{x} + i\mathbf{y}) = \mathbf{p}^T \mathbf{x} + \mathbf{q}^T \mathbf{y} + i(\mathbf{p}^T \mathbf{y} - \mathbf{q}^T \mathbf{x})$$

Quindi

$$(6.4) \quad \mathbf{p}^T \mathbf{x} + \mathbf{q}^T \mathbf{y} = \mathbf{p}^T \mathbf{y} - \mathbf{q}^T \mathbf{x} = 0$$

Per ipotesi anche il prodotto hermitiano di $\bar{\mathbf{v}}$ e \mathbf{w} è nullo, e quindi

$$(6.5) \quad -\mathbf{p}^T \mathbf{x} + \mathbf{q}^T \mathbf{y} = \mathbf{p}^T \mathbf{y} + \mathbf{q}^T \mathbf{x} = 0$$

Mettendo a sistema le equazioni (6.4) e (6.5) otteniamo

$$(6.6) \quad \mathbf{p}^T \mathbf{x} = \mathbf{q}^T \mathbf{y} = \mathbf{p}^T \mathbf{y} = \mathbf{q}^T \mathbf{x} = 0$$

Questo mostra che la parte reale e la parte immaginaria di \mathbf{v} sono ortogonali sia alla parte reale sia alla parte immaginaria di \mathbf{w} .

Costruzione di Q Gli autovalori non reali di \mathbf{N} sono $a_k \pm ib_k$ per $k = 1, \dots, s$. Poniamo $\lambda_k = a_k - ib_k$. Per il teorema spettrale esiste una base ortonormale di \mathbb{C}^n formata da autovettori di \mathbf{N} . In particolare, possiamo trovare autovettori $\mathbf{v}_1, \dots, \mathbf{v}_s$ relativi agli autovalori $\lambda_1, \dots, \lambda_s$, e autovettori reali $\mathbf{w}_1, \dots, \mathbf{w}_t$ relativi agli autovalori reali μ_1, \dots, μ_t tali che l'insieme

$$\{\mathbf{v}_1, \dots, \mathbf{v}_s, \mathbf{w}_1, \dots, \mathbf{w}_t\}$$

sia ortonormale: i suoi elementi sono versori a due a due ortogonali in \mathbb{C}^n .

Come abbiamo visto nella dimostrazione del fatto 1, i vettori $\bar{\mathbf{v}}_1, \dots, \bar{\mathbf{v}}_s$ sono autovettori di \mathbf{N} relativi agli autovalori $\bar{\lambda}_1, \dots, \bar{\lambda}_s$; inoltre sono versori, perché $\|\bar{\mathbf{v}}_k\| = \|\mathbf{v}_k\| = 1$, e sono a due a due ortogonali perché

$$\bar{\mathbf{v}}_i^H \bar{\mathbf{v}}_j = \overline{\mathbf{v}_i^H \mathbf{v}_j} = \bar{0} = 0 \quad \text{se } i \neq j$$

Infine i $\bar{\mathbf{v}}_k$ sono ortogonali a ciascun \mathbf{v}_i e a ciascun \mathbf{w}_j perché autovettori di una matrice normale relativi ad autovalori distinti sono ortogonali. Quindi

$$\{\mathbf{v}_1, \bar{\mathbf{v}}_1, \dots, \mathbf{v}_s, \bar{\mathbf{v}}_s, \mathbf{w}_1, \dots, \mathbf{w}_t\}$$

è una base ortonormale di \mathbb{C}^n . Sia ora $\mathbf{v}_k = \mathbf{x}_k + i\mathbf{y}_k$ la decomposizione in parte reale e immaginaria di \mathbf{v}_k per $k = 1, \dots, s$. Per il fatto 2 gli n vettori

$$\{\mathbf{x}_1, \mathbf{y}_1, \dots, \mathbf{x}_s, \mathbf{y}_s, \mathbf{w}_1, \dots, \mathbf{w}_t\}$$

sono a due a due ortogonali in \mathbb{R}^n . Per costruzione i vettori \mathbf{w}_k sono versori, mentre per il fatto 1 i vettori \mathbf{x}_k e \mathbf{y}_k hanno la stessa norma. Ora:

$$1 = \|\mathbf{v}_k\|^2 = \|\mathbf{x}_k\|^2 + \|\mathbf{y}_k\|^2 = 2\|\mathbf{x}_k\|^2$$

Quindi:

$$\mathcal{B} = \{\sqrt{2}\mathbf{x}_1, \sqrt{2}\mathbf{y}_1, \dots, \sqrt{2}\mathbf{x}_s, \sqrt{2}\mathbf{y}_s, \mathbf{w}_1, \dots, \mathbf{w}_t\}$$

è una base ortonormale di \mathbb{R}^n . Sia \mathbf{Q} la matrice ortogonale che ha come colonne i vettori della base \mathcal{B} . Dal fatto 1 segue che:

$$(6.7) \quad \begin{aligned} \mathbf{N}(\sqrt{2}\mathbf{x}_k) &= a_k(\sqrt{2}\mathbf{x}_k) + b_k(\sqrt{2}\mathbf{y}_k), \\ \mathbf{N}(\sqrt{2}\mathbf{y}_k) &= -b_k(\sqrt{2}\mathbf{x}_k) + a_k(\sqrt{2}\mathbf{y}_k) \quad \text{per ogni } k = 1, \dots, s. \end{aligned}$$

D'altra parte

$$(6.8) \quad \mathbf{N}\mathbf{w}_h = \mu_h \mathbf{w}_h \quad \text{per ogni } h = 1, \dots, t$$

Questo dimostra il teorema perché le uguaglianze (6.7) e (6.8) equivalgono a $\mathbf{N}\mathbf{Q} = \mathbf{Q}\mathbf{B}$.

OSSERVAZIONE La matrice $\mathbf{B} = \text{diag}(\mathbf{D}_1, \dots, \mathbf{D}_s, \mu_1, \dots, \mu_t)$ del teorema ha l'aspetto

$$(6.9) \quad \mathbf{B} = \left[\begin{array}{cc|c} a_1 & -b_1 & & \\ b_1 & a_1 & & \\ \hline & & a_s & -b_s \\ & & b_s & a_s \\ \hline & & & \mu_1 \\ & & & \mu_t \end{array} \right]$$

COROLLARIO 6.2 (Forma canonica di una matrice ortogonale)

Sia \mathbf{P} una matrice ortogonale. Esistono $\theta_1, \dots, \theta_s \in [0, 2\pi)$ e una matrice ortogonale \mathbf{Q} tale che $\mathbf{B} = \mathbf{Q}^T \mathbf{P} \mathbf{Q}$ è la matrice diagonale a blocchi

$$(6.10) \quad \mathbf{B} = \text{diag}(\mathbf{D}_1, \dots, \mathbf{D}_s, \mu_1, \dots, \mu_t)$$

dove

$$(6.11) \quad \mathbf{D}_k = \begin{bmatrix} \cos(\theta_k) & -\sin(\theta_k) \\ \sin(\theta_k) & \cos(\theta_k) \end{bmatrix}, \quad \mu_h = \pm 1$$

DIMOSTRAZIONE. Gli autovalori di una matrice ortogonale hanno modulo 1. Quelli non reali sono della forma $e^{i\theta} = \cos(\theta) + i \sin(\theta)$ con $\sin(\theta) \neq 0$, quelli reali possono solo essere 1 o -1. La tesi segue ora dal teorema 6.1. □

COROLLARIO 6.3 (Forma canonica di una matrice antisimmetrica reale) Sia \mathbf{E} una matrice antisimmetrica reale. Esistono numeri reali positivi b_1, \dots, b_s e una matrice ortogonale \mathbf{Q} tale che $\mathbf{B} = \mathbf{Q}^T \mathbf{E} \mathbf{Q}$ è la matrice diagonale a blocchi

$$(6.12) \quad \mathbf{B} = \text{diag}(\mathbf{D}_1, \dots, \mathbf{D}_s, 0, \dots, 0)$$

dove

$$(6.13) \quad \mathbf{D}_k = \begin{bmatrix} 0 & -b_k \\ b_k & 0 \end{bmatrix}$$

In particolare, il rango di \mathbf{E} è pari.

DIMOSTRAZIONE. Gli autovalori di una matrice antisimmetrica sono immaginari puri, quindi o sono nulli o sono della forma $\pm b_k i$ con $b_k > 0$. L'esistenza di \mathbf{Q} e \mathbf{B} segue dal teorema 6.1. Il rango di \mathbf{E} è uguale al rango di \mathbf{B} perché le due matrici sono simili, e il rango di \mathbf{B} è $2s$. □

Si consideri la matrice

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

Si spieghi perché \mathbf{A} è la matrice di una rotazione di \mathbb{R}^3 . Si determinino l'asse e l'angolo di rotazione. Si determinino gli autovalori di \mathbf{A} . La matrice è diagonalizzabile da una matrice complessa? E da una matrice reale? Si determini una matrice ortogonale reale \mathbf{Q} tale che

$$\mathbf{Q}^T \mathbf{A} \mathbf{Q} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Si mostri che ogni isometria lineare di \mathbb{R}^n è il prodotto di composizione di isometrie di questo tipo: una rotazione di un piano $\mathbf{H} \subseteq \mathbb{R}^n$ che lascia fisso \mathbf{H}^\perp , una riflessione ortogonale con asse un iperpiano. Se l'isometria ha determinante 1, non c'è bisogno delle riflessioni; d'altra parte, mostrare che ogni rotazione è il prodotto di due riflessioni, quindi ogni isometria lineare è un prodotto di riflessioni ortogonali.

7 QUADRICHÉ

In questo paragrafo pensiamo gli elementi di \mathbb{R}^n come punti. Intuitivamente, identifichiamo un punto col vettore \mathbf{x} delle sue coordinate: il sistema di riferimento è fissato: l'origine è il punto che corrisponde a $\mathbf{x} = \mathbf{0}$, e gli assi coordinati sono le rette generate dai versori della base canonica. Una *quadrica* in \mathbb{R}^n è il luogo dei punti di \mathbb{R}^n che soddisfano un'equazione di secondo grado in x_1, \dots, x_n :

$$(7.1) \quad q(\mathbf{x}) = \sum_{i=1}^n a_{ii} x_i^2 + \sum_{1 \leq i < j \leq n} 2a_{ij} x_i x_j + \sum_{i=1}^n 2b_i x_i + c = 0$$

I coefficienti a_{ij} , b_i e c dell'equazione sono numeri reali. Perché l'equazione sia di secondo grado occorre richiedere che almeno uno dei coefficienti a_{ij} sia diverso da zero. Quando $n = 2$ si parla di *conica* anziché di quadrica; se si usano x e y al posto di x_1 e x_2 come coordinate nel piano, l'equazione di una conica ha la forma

$$(7.2) \quad q(x, y) = a_{11} x^2 + a_{22} y^2 + 2a_{12} xy + 2b_1 x + 2b_2 y + c = 0$$

L'equazione di una conica dipende da 6 coefficienti. L'equazione di una quadrica in \mathbb{R}^3 dipende 10 coefficienti: nelle coordinate x , y , z ha la forma

$$a_{11} x^2 + a_{22} y^2 + a_{33} z^2 + 2a_{12} xy + 2a_{13} xz + 2a_{23} yz + 2b_1 x + 2b_2 y + 2b_3 z + c = 0$$

Per riscrivere il polinomio $q(\mathbf{x})$ in una forma compatta e maneggevole, introduciamo la matrice simmetrica $\mathbf{A} = [a_{ij}]$ e il vettore riga $\mathbf{b}^T = [b_i]$. Allora

$$(7.3) \quad q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + 2\mathbf{b}^T \mathbf{x} + c$$

Abbiamo così messo in evidenza il fatto che il polinomio di secondo grado $q(\mathbf{x})$ è somma della forma quadratica $\mathbf{x}^T \mathbf{A} \mathbf{x}$, della forma lineare $2\mathbf{b}^T \mathbf{x}$ e della costante c .

Il nostro scopo in questo paragrafo è di trovare un cambiamento di coordinate che consenta di semplificare la (7.1) in modo da poter riconoscere la quadrica geometricamente. Per esempio, nel caso delle coniche, ci aspettiamo che l'equazione (7.2) definisca un'ellisse o un'iperbole o una parabola (eventualmente degeneri); per provare questo, troveremo una rototraslazione del piano che trasforma l'equazione (7.2) in un'equazione *canonica*, ovvero del tipo:

$$(7.4) \quad \frac{x^2}{a^2} \pm \frac{y^2}{b^2} = 1 \quad \text{per ellissi e iperbole}, \quad x^2 + 2py = 0 \quad \text{per una parabola}$$

Prima di cominciare, occorre fare una digressione per descrivere i cambiamenti di coordinate e il loro effetto sull'equazione di una quadrica.

Cambiamenti di coordinate e rototraslazioni

Finora abbiamo considerato solo cambiamenti di coordinate lineari, cioè della forma $\mathbf{x} = \mathbf{S}\mathbf{X}$, dove \mathbf{S} è una matrice invertibile. Nello studio delle quadriche è però necessario poter cambiare anche l'origine: se la quadrica ha un centro di simmetria, come succede per esempio per un'ellisse o un'iperbole o una sfera, l'equazione della quadrica si semplifica se si prende come origine delle coordinate il centro di simmetria.

DEFINIZIONE 7.1 (Affinità)

Una funzione $\mathbf{x} = F(\mathbf{X}) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ si dice *trasformazione di coordinate affine* o *affinità* se esistono una matrice $n \times n$ invertibile \mathbf{S} e un vettore $\mathbf{v} \in \mathbb{R}^n$ tali che $F(\mathbf{X}) = \mathbf{S}\mathbf{X} + \mathbf{v}$.

OSSERVAZIONE Un'applicazione lineare invertibile $\mathbf{x} = \mathbf{S}\mathbf{X}$ è un'affinità: basta prendere $\mathbf{v} = \mathbf{0}$; una traslazione $\mathbf{x} = \mathbf{X} + \mathbf{v}$ è un'affinità: basta prendere $\mathbf{S} = \mathbf{I}$. Un'affinità arbitraria $F(\mathbf{X}) = \mathbf{S}\mathbf{X} + \mathbf{v}$ è la funzione composta $F = H \circ G$ dell'applicazione lineare $G(\mathbf{X}) = \mathbf{S}\mathbf{X}$ con la traslazione $H(\mathbf{Y}) = \mathbf{Y} + \mathbf{v}$: un'affinità consiste quindi di un cambiamento lineare di coordinate seguito da una traslazione. L'affinità si può anche descrivere come una traslazione seguita da un cambiamento lineare di coordinate; infatti l'uguaglianza

$$F(\mathbf{X}) = \mathbf{S}\mathbf{X} + \mathbf{v} = \mathbf{S}(\mathbf{X} + \mathbf{w}) \quad \text{dove } \mathbf{w} = \mathbf{S}^{-1}\mathbf{v}$$

mostra che $F = G \circ H_1$ è uguale alla traslazione $H_1(\mathbf{X}) = \mathbf{X} + \mathbf{w}$ seguita da $G(\mathbf{Y}) = \mathbf{S}\mathbf{Y}$.

PROPOSIZIONE 7.2 (Le affinità formano un gruppo)

L'insieme delle affinità $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ soddisfa le seguenti proprietà:

- a) la funzione identità $F(\mathbf{X}) = \mathbf{X}$ è un'affinità;

- b) un'affinità è una funzione invertibile e l'inversa di un'affinità è un'affinità; più precisamente, l'inversa di $\mathbf{x} = F(\mathbf{X}) = \mathbf{S}\mathbf{X} + \mathbf{v}$ è

$$\mathbf{X} = F^{-1}(\mathbf{x}) = \mathbf{S}^{-1}\mathbf{x} + (-\mathbf{S}^{-1}\mathbf{v})$$

- c) il prodotto di composizione di due affinità è un'affinità; più precisamente, se $F(\mathbf{X}) = \mathbf{S}\mathbf{X} + \mathbf{v}$ e $G(\mathbf{Y}) = \mathbf{T}\mathbf{Y} + \mathbf{w}$, allora:

$$G \circ F(\mathbf{X}) = \mathbf{T}\mathbf{S}\mathbf{X} + (\mathbf{T}\mathbf{v} + \mathbf{w})$$

DIMOSTRAZIONE. La dimostrazione è semplice ed è lasciata come esercizio.

Suggerimento:

$$\mathbf{S}^{-1}(\mathbf{S}\mathbf{X} + \mathbf{v}) + (-\mathbf{S}^{-1}\mathbf{v}) = \mathbf{X} + \mathbf{S}^{-1}\mathbf{v} - \mathbf{S}^{-1}\mathbf{v} = \mathbf{X}$$

OSSERVAZIONE Se $\mathbf{x} = \mathbf{S}\mathbf{X} + \mathbf{v}$, l'origine delle coordinate \mathbf{x} , cioè il punto che ha le coordinate \mathbf{x} tutte nulle, ha coordinate \mathbf{X} uguali a $-\mathbf{S}^{-1}\mathbf{v}$. L'origine delle coordinate \mathbf{X} ha invece coordinate \mathbf{x} uguali a \mathbf{v} .

OSSERVAZIONE Iterando il punto c) della proposizione 7.2 si vede che componendo in ordine qualsiasi un numero arbitrario di traslazioni e di applicazioni lineari invertibili si ottiene comunque un'affinità, cioè un'applicazione lineare invertibile seguita da una traslazione.

DEFINIZIONE 7.3 (Rototraslazione)

Un'applicazione lineare invertibile $\mathbf{x} = \mathbf{Q}\mathbf{X}$ si dice *rotazione* se \mathbf{Q} è una matrice ortogonale con determinante uguale a 1. Un'affinità $F(\mathbf{X}) = \mathbf{Q}\mathbf{X} + \mathbf{v}$ si dice *rototraslazione* se $\mathbf{x} = \mathbf{Q}\mathbf{X}$ è una rotazione.

OSSERVAZIONE Cerchiamo di giustificare la terminologia. Se $n = 2$, una matrice ortogonale con determinante uguale a 1 rappresenta la rotazione del piano di un angolo θ attorno all'origine. Se $n = 3$, per il teorema di Eulero una matrice ortogonale con determinante uguale a 1 rappresenta una rotazione dello spazio attorno a un asse passante per l'origine. In \mathbb{R}^n , il teorema di classificazione delle matrici ortogonali garantisce che una matrice ortogonale con determinante uguale a 1 è il prodotto di applicazioni che ruotano un piano bidimensionale \mathbf{H} lasciando fisso \mathbf{H}^\perp . Questo giustifica la definizione di rotazione di \mathbb{R}^n . Una rototraslazione è per definizione una rotazione seguita da una traslazione.

OSSERVAZIONE Anche per le rototraslazioni valgono le proprietà della proposizione 7.2: l'applicazione identità è una rototraslazione, l'inversa di una rototraslazione è una rototraslazione, e la composta di due rototraslazioni è una rototraslazione. In particolare, componendo un numero arbitrario di traslazioni e di rotazioni in qualsiasi ordine si ottiene ancora una rototraslazione.

Vediamo ora quale sia l'effetto di un'affinità sull'equazione di una quadrica.

PROPOSIZIONE 7.4 (Trasformazione dell'equazione di una quadrica)

Sia

$$(7.5) \quad q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + 2\mathbf{b}^T \mathbf{x} + c$$

un polinomio di secondo grado nelle variabili \mathbf{x} , e sia $\mathbf{x} = \mathbf{S}\mathbf{X} + \mathbf{v}$ un'affinità.
Allora $\tilde{q}(\mathbf{X}) = q(\mathbf{S}\mathbf{X} + \mathbf{v})$ è il polinomio di secondo grado nelle variabili \mathbf{X} :

$$(7.6) \quad \tilde{q}(\mathbf{X}) = \mathbf{X}^T \tilde{\mathbf{A}} \mathbf{X} + 2\tilde{\mathbf{b}}^T \mathbf{X} + \tilde{c}$$

determinato da

$$(7.7) \quad \tilde{\mathbf{A}} = \mathbf{S}^T \mathbf{A} \mathbf{S}, \quad \tilde{\mathbf{b}} = \mathbf{S}^T (\mathbf{A}\mathbf{v} + \mathbf{b}), \quad \tilde{c} = \mathbf{v}^T \mathbf{A}\mathbf{v} + 2\mathbf{b}^T \mathbf{v} + c$$

DIMOSTRAZIONE. Sostituendo $\mathbf{x} = \mathbf{S}\mathbf{X} + \mathbf{v}$ nella (7.5) si ottiene

$$(7.8) \quad \tilde{q}(\mathbf{X}) = \mathbf{X}^T \mathbf{S}^T \mathbf{A} \mathbf{S} \mathbf{X} + \mathbf{v}^T \mathbf{A} \mathbf{S} \mathbf{X} + \mathbf{X}^T \mathbf{S}^T \mathbf{A} \mathbf{v} + \mathbf{v}^T \mathbf{A} \mathbf{v} + 2\mathbf{b}^T \mathbf{S} \mathbf{X} + 2\mathbf{b}^T \mathbf{v} + c$$

Considerando i termini di secondo grado si ottiene $\tilde{\mathbf{A}} = \mathbf{S}^T \mathbf{A} \mathbf{S}$, mentre il termine costante è $\tilde{c} = \mathbf{v}^T \mathbf{A} \mathbf{v} + 2\mathbf{b}^T \mathbf{v} + c$. Il termine lineare in \mathbf{X} contiene $\mathbf{X}^T \mathbf{S}^T \mathbf{A} \mathbf{v}$ che è uno scalare e pertanto coincide col suo trasposto:

$$(7.9) \quad \mathbf{X}^T \mathbf{S}^T \mathbf{A} \mathbf{v} = (\mathbf{X}^T \mathbf{S}^T \mathbf{A} \mathbf{v})^T = \mathbf{v}^T \mathbf{A}^T \mathbf{S} \mathbf{X} = \mathbf{v}^T \mathbf{A}^T \mathbf{S} \mathbf{X}$$

dove nell'ultimo passaggio abbiamo usato il fatto che $\mathbf{A}^T = \mathbf{A}$ perché \mathbf{A} è simmetrica. Il termine lineare in $\tilde{q}(\mathbf{X})$ è perciò

$$(7.10) \quad 2\mathbf{v}^T \mathbf{A} \mathbf{S} \mathbf{X} + 2\mathbf{b}^T \mathbf{S} \mathbf{X}$$

il che mostra $\tilde{\mathbf{b}} = \mathbf{S}^T (\mathbf{A}\mathbf{v} + \mathbf{b})$.

Vi è un modo ancora più compatto di scrivere un polinomio di secondo grado $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + 2\mathbf{b}^T \mathbf{x} + c$: per questo introduciamo la matrice simmetrica di ordine $n+1$

$$(7.11) \quad \mathbf{B} = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{b} \\ \hline \mathbf{b}^T & c \end{array} \right]$$

Allora

$$(7.12) \quad q(\mathbf{x}) = [\mathbf{x}^T \ 1] \ \mathbf{B} \ \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} = \mathbf{z}^T \ \mathbf{B} \ \mathbf{z}$$

dove \mathbf{z} è il vettore di \mathbb{R}^{n+1} ottenuto aggiungendo a \mathbf{x} un'ultima componente uguale a 1:

$$(7.13) \quad \mathbf{z}^T = [x_1, \dots, x_n, 1]$$

Diremo che \mathbf{B} è la *matrice associata* al polinomio $q(\mathbf{x})$. La matrice \mathbf{B} contiene tutti i coefficienti del polinomio, e assegnare \mathbf{B} è equivalente ad assegnare $q(\mathbf{x})$.



Consideriamo il polinomio in due variabili

$$q(x, y) = x^2 + 8xy + 7y^2 - 2x + 8y + 5$$

La matrice \mathbf{B} associata a $q(x, y)$ è

$$\mathbf{B} = \left[\begin{array}{cc|c} 1 & 4 & -1 \\ 4 & 7 & 4 \\ \hline -1 & 4 & 5 \end{array} \right]$$

Il lettore dovrebbe verificare la (7.12) in questo caso:

$$q(x, y) = [x \ y \ 1] \begin{bmatrix} 1 & 4 & -1 \\ 4 & 7 & 4 \\ -1 & 4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Analogamente, a un'affinità $\mathbf{x} = \mathbf{S}\mathbf{X} + \mathbf{v}$ possiamo associare la matrice quadrata di ordine $n + 1$

$$(7.14) \quad \mathbf{F} = \left[\begin{array}{c|c} \mathbf{S} & \mathbf{v} \\ \hline \mathbf{0}^T & 1 \end{array} \right]$$

Allora

$$(7.15) \quad \mathbf{z} = \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} = \left[\begin{array}{c|c} \mathbf{S} & \mathbf{v} \\ \hline \mathbf{0}^T & 1 \end{array} \right] \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix} = \mathbf{FZ}$$

Con queste notazioni la (7.7) diventa:

$$(7.16) \quad \hat{\mathbf{B}} = \mathbf{F}^T \mathbf{BF}$$

Siccome l'inversa di una rototraslazione è una rototraslazione e il prodotto di due rototraslazioni è una rototraslazione, possiamo definire una relazione di equivalenza sull'insieme dei polinomi dichiarando equivalenti due polinomi che differiscono per una rototraslazione: diciamo che due polinomi $q(\mathbf{x})$ e $\tilde{q}(\mathbf{x})$ sono *metricamente equivalenti* se esiste una rototraslazione $\mathbf{x} = \mathbf{Q}\mathbf{X} + \mathbf{v}$ tale che $\tilde{q}(\mathbf{X}) = q(\mathbf{Q}\mathbf{X} + \mathbf{v})$. In questo caso la rototraslazione porta la quadrica di equazione $q = 0$ nella quadrica di equazione $\tilde{q} = 0$. Il termine *metricamente* è giustificato dal fatto che le rototraslazioni sono delle isometrie e quindi preservano le proprietà *metriche* delle quadriche (per esempio, la lunghezza dei semiassi di un'ellisse o l'angolo formato dagli asymptoti di un'iperbole).

TEOREMA 7.5 (Forma canonica di un polinomio di secondo grado)

Sia

$$q(\mathbf{x}) = \mathbf{z}^T \mathbf{B} \mathbf{z} = \mathbf{x}^T \mathbf{A} \mathbf{x} + 2\mathbf{b}^T \mathbf{x} + c$$

un polinomio di secondo grado nelle variabili x_1, \dots, x_n . Allora gli autovalori di \mathbf{A} , il rango di \mathbf{A} , il rango e il determinante di \mathbf{B} sono invarianti per rototraslazioni. Sia r il rango di \mathbf{A} e siano $\lambda_1, \dots, \lambda_r$ gli autovalori non nulli di \mathbf{A} . Allora $r \leq r(\mathbf{B}) \leq r+2$ e il polinomio $q(\mathbf{x})$ è metricamente equivalente a

- a) $\tilde{q}(\mathbf{X}) = \lambda_1 X_1^2 + \dots + \lambda_r X_r^2$ se $r(\mathbf{B}) = r$;
- b) $\tilde{q}(\mathbf{X}) = \lambda_1 X_1^2 + \dots + \lambda_r X_r^2 + \tilde{c}$, con $\tilde{c} \neq 0$, se $r(\mathbf{B}) = r+1$;
- c) $\tilde{q}(\mathbf{X}) = \lambda_1 X_1^2 + \dots + \lambda_r X_r^2 + 2p X_{r+1}$, con $p \neq 0$, se $r(\mathbf{B}) = r+2$.

DIMOSTRAZIONE. La rototraslazione $\mathbf{x} = \mathbf{Q}\mathbf{X} + \mathbf{v}$ trasforma \mathbf{A} in $\tilde{\mathbf{A}} = \mathbf{Q}^T \mathbf{A} \mathbf{Q}$. Siccome \mathbf{Q} è ortogonale, $\tilde{\mathbf{A}}$ e \mathbf{A} sono simili e quindi hanno gli stessi autovalori e lo stesso rango; il rango r coincide col numero di autovalori non nulli $\lambda_1, \dots, \lambda_r$ di \mathbf{A} . La matrice \mathbf{B} si trasforma in $\tilde{\mathbf{B}} = \mathbf{F}^T \mathbf{B} \mathbf{F}$ per la (7.16). Occorre stare attenti al fatto che \mathbf{F} non è necessariamente ortogonale, quindi gli autovalori di $\tilde{\mathbf{B}}$ e \mathbf{B} possono non coincidere. La matrice \mathbf{F} è comunque invertibile con determinante uguale a 1:

$$\det(\mathbf{F}) = \det \left(\begin{bmatrix} \mathbf{Q} & \mathbf{v} \\ \mathbf{0}^T & 1 \end{bmatrix} \right) = \det(\mathbf{Q}) = 1$$

Perciò \mathbf{B} e $\tilde{\mathbf{B}}$ hanno lo stesso rango e, per il teorema di Binet, lo stesso determinante.

Vediamo ora come portare $q(\mathbf{x})$ in forma canonica. Per questo usiamo la proposizione 7.4: è sufficiente ricordare che una traslazione $\mathbf{x} = \mathbf{X} + \mathbf{v}$ lascia \mathbf{A} invariata e trasforma \mathbf{b} in $\mathbf{Av} + \mathbf{b}$, mentre una rotazione $\mathbf{x} = \mathbf{Q}\mathbf{X}$ trasforma \mathbf{A} in $\mathbf{Q}^T \mathbf{A} \mathbf{Q}$ e \mathbf{b} in $\mathbf{Q}^T \mathbf{b}$. Cerchiamo prima di eliminare \mathbf{b} , cioè il termine lineare di $q(\mathbf{x})$, mediante una traslazione: questo è possibile se e solo se $r(\mathbf{A}) = r([\mathbf{A}|\mathbf{b}])$.

Caso 1 Supponiamo $r(\mathbf{A}) = r([\mathbf{A}|\mathbf{b}])$.

Allora esiste \mathbf{w} tale che $\mathbf{Aw} = \mathbf{b}$. Facciamo la traslazione $\mathbf{x} = \mathbf{Y} - \mathbf{w}$. Il polinomio $q_1(\mathbf{Y}) = q(\mathbf{Y} - \mathbf{w})$ non ha più termini di primo grado perché $\mathbf{A}(-\mathbf{w}) + \mathbf{b} = \mathbf{0}$, mentre i termini di secondo grado sono rimasti invariati:

$$q_1(\mathbf{Y}) = \mathbf{Y}^T \mathbf{A} \mathbf{Y} + c_1$$

Ora usiamo il teorema spettrale per diagonalizzare la forma quadratica $\mathbf{Y}^T \mathbf{A} \mathbf{Y}$: esiste una matrice ortogonale \mathbf{Q} tale che

$$\tilde{\mathbf{A}} = \mathbf{Q}^T \mathbf{A} \mathbf{Q} = \text{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0)$$

Si ricordi che il determinante di una matrice ortogonale è ± 1 . Possiamo supporre, cambiando se necessario segno alla prima colonna di \mathbf{Q} , che $\det(\mathbf{Q}) = 1$. La rotazione $\mathbf{Y} = \mathbf{Q}\mathbf{X}$ trasforma \mathbf{A} in $\mathbf{Q}^T \mathbf{A} \mathbf{Q}$ e non reintroduce termini lineari perché $\mathbf{Q}^T \mathbf{0} = \mathbf{0}$. Quindi

$$\tilde{q}(\mathbf{X}) = q_1(\mathbf{Q}\mathbf{X}) = \mathbf{X}^T \tilde{\mathbf{A}} \mathbf{X} + \tilde{c} = \lambda_1 X_1^2 + \dots + \lambda_r X_r^2 + \tilde{c}$$

La rotazione successiva $\mathbf{X} = \mathbf{Q}\mathbf{Y}$ fa sì che i nuovi assi coordinati siano paralleli agli autovettori di \mathbf{A} . Geometricamente, si tratta di *assi di simmetria della quadrica*. Vediamo perché per il primo asse coordinato. La simmetria ortogonale rispetto al primo asse coordinato manda il punto $P = [X_1, X_2, \dots, X_n]$ in $R = [X_1, -X_2, \dots, -X_n]$; le coordinate di P soddisfano l'equazione $\tilde{q}(\mathbf{X})$ se e solo se le coordinate di R la soddisfano, quindi la quadrica è simmetrica rispetto al primo asse coordinato. Analogo discorso vale per gli altri assi. In conclusione, le rette passanti per il centro della quadrica e dirette come gli autovettori di \mathbf{A} sono assi di simmetria della quadrica. Il teorema spettrale mostra dunque che una quadrica a centro ha n assi di simmetria a due a due ortogonali; per questo il teorema spettrale è noto anche come *teorema degli assi principali*.

OSSERVAZIONE Si dice che due quadriche $\mathbf{q}_1(\mathbf{x}) = 0$ e $\mathbf{q}_2(\mathbf{x}) = 0$ di \mathbb{R}^n sono *congruenti* se esiste una costante $k \neq 0$ tale che \mathbf{q}_1 e $k\mathbf{q}_2$ siano metricamente equivalenti (si osservi che moltiplicare l'equazione di una quadrica per uno scalare non nullo non modifica l'insieme dei punti della quadrica). Intuitivamente, questo significa che esiste una rototraslazione che porta una quadrica nell'altra. Per dirla con Euclide, due quadriche sono congruenti se esiste un movimento rigido di \mathbb{R}^n che porta l'una nell'altra.

Classificazione delle coniche Vogliamo determinare le classi di congruenza delle coniche. Una conica ha un'equazione del tipo

$$(7.17) \quad q(x, y) = a_{11}x^2 + a_{22}y^2 + 2a_{12}xy + 2b_1x + 2b_2y + c = 0$$

Poniamo come sopra

$$\mathbf{B} = \begin{bmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ b_1 & b_2 & c \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

I classici invarianti ortogonali dell'equazione della conica sono

$$I_1 = \text{tr}(\mathbf{A}) = \lambda_1 + \lambda_2, \quad I_2 = \det(\mathbf{A}) = \lambda_1 \lambda_2, \quad I_3 = \det(\mathbf{B})$$

Qui λ_1 e λ_2 denotano gli autovalori di \mathbf{A} . Si parla di invarianti perché, per il teorema 7.5, sono lasciati invariati da una rototraslazione. L'invariante I_1 si dice invariante lineare perché $I_1(kq) = kI_1(q)$: se moltiplichiamo l'equazione per lo scalare k , l'invariante I_1 viene moltiplicato per k . L'invariante I_2 si dice quadratico perché $I_2(kq) = k^2 I_2(q)$, e I_3 cubico perché $I_3(kq) = k^3 I_3(q)$. Si osservi che moltiplicando l'equazione della conica per uno scalare non nullo non si modifica il fatto che I_1 , I_2 e I_3 siano o non siano nulli, e non si modifica il segno di I_2 né quello di $I_1 I_3$.

I casi possibili sono:

- a) $r(\mathbf{A}) = r(\mathbf{B}) = 1$ (in questo caso $I_2 = I_3 = 0$).

In questo caso la forma canonica di $q(x, y)$ è $\lambda_1 X^2 = 0$. Dividendo per λ_1 vediamo che la conica è congruente alla conica di equazione $X^2 = 0$: si dice che la conica è la retta $X = 0$ contata due volte (convincersi che $q(x, y)$ è λ_1 per il quadrato di una forma lineare).

- b) $r(\mathbf{A}) = 1, r(\mathbf{B}) = 2$ (anche in questo caso $I_2 = I_3 = 0$).

In questo caso la forma canonica di $q(x, y)$ è $\lambda_1 X^2 + \tilde{c}$ con $\tilde{c} \neq 0$. Dividendo per λ_1 vediamo che la conica è congruente alla conica di equazione $X^2 + d = 0$ dove $d = \tilde{c}/\lambda_1$. Se $d < 0$ la conica è l'unione delle due rette parallele di equazione $X = \sqrt{-d}$ e $X = -\sqrt{-d}$. Se $d > 0$, la conica non ha punti reali; nel piano complesso consiste delle due rette disgiunte $X = i\sqrt{d}$ e $X = -i\sqrt{d}$.

- c) $r(\mathbf{A}) = 2, r(\mathbf{B}) = 2$ (questo equivale a $I_2 \neq 0$ e $I_3 = 0$).

In questo caso la forma canonica di $q(x, y)$ è $\lambda_1 X^2 + \lambda_2 Y^2$ dove λ_1 e λ_2 sono gli autovalori di \mathbf{A} . Dividendo per λ_1 vediamo che la conica è congruente alla conica di equazione $X^2 + dY^2 = 0$ dove $d = \lambda_2/\lambda_1 \neq 0$. Se $I_2 = \lambda_1 \lambda_2 < 0$, gli autovalori sono discordi e d è negativo: la conica è allora l'unione delle due rette incidenti $X = \sqrt{-d}Y$ e $X = -\sqrt{-d}Y$. Se invece $I_2 > 0$, cioè $d > 0$, la conica ha solo un punto reale (l'origine, cioè il suo centro); nel piano complesso consiste delle due rette $X = i\sqrt{d}Y$ e $X = -i\sqrt{d}Y$.

- d) $r(\mathbf{A}) = 1, r(\mathbf{B}) = 3$ (questo equivale a $I_2 = 0$ e $I_3 \neq 0$).

In questo caso la forma canonica di $q(x, y)$ è $\lambda_1 X^2 + 2pY = 0$ dove λ_1 è l'autovalore non nullo di \mathbf{A} . Dividendo per λ_1 vediamo che la conica è congruente alla conica di equazione $X^2 + 2aY = 0$ con $a = p/\lambda_1 \neq 0$. Si tratta perciò di una parabola.

- d) $r(\mathbf{A}) = 2, r(\mathbf{B}) = 3$ (questo equivale a $I_2 \neq 0$ e $I_3 \neq 0$).

In questo caso la forma canonica di $q(x, y)$ è $\lambda_1 X^2 + \lambda_2 Y^2 + \tilde{c} = 0$ e $\tilde{c} \neq 0$. Si osservi che $I_3 = \det(\mathbf{B}) = \lambda_1 \lambda_2 \tilde{c} = \det(\mathbf{A}) \tilde{c} = I_2 \tilde{c}$ (per scrivere la forma canonica dell'equazione è pertanto sufficiente calcolare gli autovalori di \mathbf{A} e il determinante di \mathbf{B}). Dividendo per $-\tilde{c} = -I_3/I_2$ vediamo che la conica è congruente alla conica di equazione

$$eX^2 + fY^2 = 1$$

dove $e = -\lambda_1 I_2/I_3$ e $f = -\lambda_2 I_2/I_3$. Distinguiamo i seguenti sottocasi:

1. $I_2 > 0$ e $I_1 I_3 < 0$

In questo caso gli autovalori λ_1 e λ_2 sono concordi perché $I_2 = \lambda_1 \lambda_2 > 0$, e hanno segno opposto a I_3 perché $I_1 = \lambda_1 + \lambda_2$ ha segno opposto a I_3 . Quindi e ed f sono positivi: posto $a = 1/\sqrt{-e}$ e $b = 1/\sqrt{-f}$ l'equazione della conica diviene

$$\frac{X^2}{a^2} + \frac{Y^2}{b^2} = 1$$

Si tratta perciò di un'ellisse.

2. $I_2 > 0$ e $I_1 I_3 > 0$

In questo caso e e f sono negativi. Ragionando come sopra si porta l'equazione nella forma

$$\frac{X^2}{a^2} + \frac{Y^2}{b^2} = -1$$

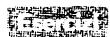
Non ci sono soluzioni $[X, Y]^T \in \mathbb{R}^2$ di questa equazione. Si dice che la conica è un'ellisse immaginaria.

3. $I_2 < 0$

In questo caso gli autovalori sono discordi, e l'equazione si può portare (ruotando l'asse X nell'asse Y se necessario) nella forma

$$\frac{X^2}{a^2} - \frac{Y^2}{b^2} = 1$$

La conica è perciò un'iperbole. Si noti che l'iperbole è equilatera quando $a = b$, e questo equivale a $I_1 = \lambda_1 + \lambda_2 = 0$.



Q Riconoscere la conica di equazione

$$x^2 + y^2 + 4x - 2y - 6 = 0$$

Q Riconoscere la conica di equazione

$$x^2 + 6xy - 2x - 8y = 0$$

Q Riconoscere la conica di equazione

$$x^2 + y^2 - 2xy + 2x - 4y = 0$$

Q Riconoscere la conica di equazione

$$x^2 - 4xy + y^2 - 2x + 4y - 1$$

e trovarne (se esistono) il centro e gli assi di simmetria.

Q Ridurre a forma canonica la quadrica di equazione

$$2x^2 + 2y^2 + 2z^2 + 2xy + 4xz + 2yz + 2x + 4y + 2z + 1 = 0$$

Q Ridurre a forma canonica la quadrica di equazione

$$2x^2 + 3z^2 - 4xy + 3yz - x + 2y = 0$$

Q Ridurre a forma canonica la quadrica di equazione

$$x^2 - 3y^2 + z^2 - 2xy + 2x + 2z - 1 = 0$$

Q Ridurre a forma canonica la quadrica di equazione

$$9x^2 - 5y^2 - 4z^2 - 4yz + 18x - 8y + c = 0$$

al variare di $c \in \mathbb{R}$.

Q Mostrare che ogni retta di \mathbb{R}^n che non sia contenuta nella quadratica $q(\mathbf{x}) = 0$ interseca la quadrica in al più due punti distinti.

Q Mostrare che le quadriche di equazione $x^2 - y^2 = 1 - z^2$ (iperboloide iperbolico) e $x^2 - y^2 = z$ (paraboloide iperbolico) contengono infinite rette.

Q In \mathbb{R}^n si consideri la quadrica Q di equazione:

$$X_1^2 + \dots + X_{n-1}^2 - X_n^2 = 0$$

Mostrare che Q è un cono con vertice nell'origine: questo significa che, se un punto P appartiene alla quadrica, allora l'intera retta che congiunge P all'origine è contenuta nella quadrica.

Q Mostrare che, se \mathbf{B} è definita positiva, la quadrica di equazione $\mathbf{z}^T \mathbf{B} \mathbf{z} = 0$ non ha punti reali.

Q Dedurre dalla (7.7) che la *segnatura* di \mathbf{B} , cioè il numero di autovalori positivi, negativi e nulli di \mathbf{B} , è invariante per rototraslazioni. Mostrare con un esempio che gli autovalori di \mathbf{B} possono invece variare.

Indice analitico

A

- affinità, 487
- algoritmo
 - di eliminazione di Gauss, 94
 - di Gauss-Jordan, 138
 - di Gram-Schmidt, 396
- alternante, 261
- angolo, 30, 363
 - in uno spazio euclideo, 375
- applicazione
 - duale, 259
 - identità, 235
 - lineare, 217, 221
 - – invertibile, 236
- approssimazione lineare, 218
- area, 42
- ascissa, 32
- assi coordinati, 23, 27
- assioma delle parallele, 6
- autospazio, 305, 316
- autovalore, 301–303, 308
 - regolare, 325
 - semplice, 325
- autovalori
 - di una matrice antihermitiana, 480
 - di una matrice antisimmetrica, 480
 - di una matrice hermitiana, 480
 - di una matrice ortogonale, 387, 424
 - di una matrice simmetrica, 433
 - di una matrice unitaria, 424
- autovalori e autovettori
 - di un blocco di Jordan, 327
 - di una matrice di proiezione, 413
- autovettore, 70, 103, 301–303, 308
 - generalizzato, 341

B

- baricentro, 53
- base, 23, 27, 153, 173
 - canonica, 79
 - di Jordan, 342
 - duale, 253
 - esistenza di una, 180
 - ortogonale, 378
 - ortonormale, 32, 378
- biiettiva, 228
- bilineare, 364
- blocco di Jordan, 327

C

- cambiamento della matrice rappresentativa, 250
- campo, 72
- codominio, 227
- coefficiente
 - angolare, 51
 - di un’equazione lineare, 71
 - di Fourier, 372
- cofattore, 280
- combinazione lineare, 24, 28, 78, 163

complemento
 – algebrico, 280
 ortogonale, 389, 402
 – degli spazi associati a una matrice,
 406
 componenti, 24, 28, 173
 conica, 486
 coniugato
 – di un vettore complesso, 416
 coordinate, 23, 27, 173, 184
 rispetto a una base ortogonale, 380
 coseni direttori, 48
 criterio
 – di inettività, 230
 – di isomorfismo, 237

D

decomposizione
 – ai valori singolari, 432, 465
 di Jordan, 356
 – polare, 469
 – spettrale, 437
 derivata
 – del determinante, 297
 determinante, 41
 è il prodotto degli autovalori, 314
 – della matrice trasposta, 278
 – di un'applicazione lineare, 290
 – di una matrice di permutazione, 270
 – di una matrice ortogonale, 292
 di una matrice triangolare a blocchi,
 295
 – di Vandermonde, 286
 – e matrici non singolari, 265
 – e operazioni elementari di riga, 262
 – esistenza, 275
 formula analitica, 271
 diagonale principale, 93
 diagonalizzabilità
 – condizione sufficiente, 321
 – criteri, 353
 primo criterio, 310
 secondo criterio, 325
 diagonalizzazione simultanea
 – di forme quadratiche, 458

– di matrici che commutano, 358
 differenza di vettori, 17
 differenziale
 – secondo, 443
 dimensione, 180
 dipendenza lineare, 168
 direzione, 4, 6, 10
 distanza, 363, 370
 disuguaglianza
 – di Schwarz, 39, 374, 376 39
 – triangolare, 39
 dominio, 227

E

Elementi di Euclide, 1
 elemento neutro, 16
 endomorfismo, 233, 290
 equazione
 – autovettori e autovalori, 303, 308
 – caratteristica, 312
 – lineare, 70
 omogenea, 71
 – omogenea associata, 71
 equazioni
 – cartesiane, 160
 – normali, 407
 – parametriche di una retta, 46
 espressione in coordinate cartesiane
 – del prodotto scalare, 36
 – del prodotto vettoriale, 41

F

fascio di piani, 66
 fattorizzazione
 – di Cholesky, 446, 450, 453
 – QR, 399
 fattorizzazione di Choleski, 146
 fibra, 227
 forma
 – bilineare
 – definita positiva, 364
 – canonica
 – di Jordan, 349

- di un polinomio di secondo grado, 491
- di una matrice normale, 482
- lineare, 238, 253, 441
- quadratica, 431, 441
- - segno, 444

formula

- di Cramer, 98, 262, 286
- di Grassmann, 208, 402
- di polarizzazione, 370, 419

funzione

- invertibile, 235
- lineare, 217, 221

G

gruppo abeliano, 15

I

identità di Lagrange, 42

immagine, 230

indice

- di un autovalore, 343
- di un autovettore generalizzato, 341

indipendenza lineare, 26, 168

iniettiva, 227

insieme di generatori, 166

intersezione, 205

inversa

- destra, 254

- sinistra, 254

inversione, 268

iperpiano, 104

isometria, 386

- lineare, 386

isomorfismo, 237

L

legge

- di annullamento del prodotto per scalare, 21
- di cancellazione della somma, 17
- di inerzia di Sylvester, 456

- del parallelogramma, 377
- lemma fondamentale, 178
- lineare, 24
- linearità delle componenti, 29
- linearizzazione, 218
- lunghezza, 10

M

matrice, 80

- a scala, 91
- aggiunta, 422
- antihermitiana, 480
- antisimmetrica, 134, 161
- associata a un polinomio di secondo grado, 490
- bidiagonale, 327, 340
 - completa di un sistema lineare, 84
- definita positiva, 146, 444
- dei coefficienti di un sistema, 83
- di Jordan, 340
- di passaggio, 188, 250
 - di permutazione, 147, 267, 269
- di una proiezione ortogonale, 412, 414
- di una riflessione ortogonale, 428
- di Vandermonde, 286
- diagonale, 125
- diagonalizzabile, 310
 - emisimmetrica, 134
- hermitiana, 425, 480
- hessiana, 443
- identità, 125
- inversa, 128
- - destra, 129
- - di Moore-Penrose, 471
- - formula di Laplace, 285
- - sinistra, 129
- nilpotente, 338, 356
- nonsingolare, 131
- normale, 432, 477, 478
- nulla, 125
- ordine di, 93
 - ortogonale, 383
- - come isometrie, 386
- orthogonalmente diagonalizzabile, 432
- pseudo inversa, 471

- quadrata, 93
- rappresentativa di un'applicazione lineare, 219, 242
- semidefinita positiva, 408
- simmetrica, 134, 161
- trasposta, 133
 - coniugata, 422
 - triangolare, 93, 141
 - a blocchi, 295
 - unitaria, 423
 - come isometria, 424
 - unitariamente
 - diagonalizzabile, 425, 477
 - simile, 477
- matrici
 - congruenti, 446
 - simili, 331
- MEG, 94
- metodo
 - dei minimi quadrati, 406
 - soluzione ottima, 476
- metodo di eliminazione di Gauss, 70
- metricamente equivalenti, 490
- minimax principle, 450
- minore, 292
 - principale, 454
 - - nord ovest, 296
- minori
 - di nord-ovest, 451
- modulo, 10
- molteplicità
 - algebrica, 314
 - - è dimensione sottospazio radicale, 352
 - - e geometrica, confronto, 324
 - di una radice, 314
 - geometrica, 316, 436
 - - di un autovalore, 316
- multilineare, 261, 273

N

- nilpotent, 162
- norma, 10, 363, 367
 - L^2 , 367
 - di un vettore complesso, 417

- di una matrice, 476
- proprietà, 376
- normale, 9
- nucleo, 160
 - di un'applicazione lineare, 228
 - di una matrice, 85
- nullità, 196
- numero di condizionamento, 477

O

- omomorfismo, 232
- operazioni elementari sulle righe, 90
- ordinata, 32
- origine, 7, 22, 27

P

- parallelepipedo, 43
- parallelogramma, 40
- parametri direttori, 47
- parte
 - immaginaria di un vettore complesso, 416
 - reale di un vettore complesso, 416
- permutazione, 147, 267
- pari o dispari, 269
- piano iperbolico, 9
- pivot, 91
- pivoting, 94
- polinomio
 - caratteristico, 284, 312
 - di Fourier complesso, 422
 - di Legendre, 399
 - differenziale, 233
 - minimo, 355
 - omogeneo di primo grado, 219
- posizione reciproca di due rette, 48
- potenze
 - di una matrice, 126
- principio di sovrapposizione, 116
- prodotto
 - cartesiano, 212
 - di composizione di due funzioni, 232
 - di matrici, 122

diagonali, 126
 hermitiano, 418
 – standard, 417
 – interno, 364
 – riga per colonna, 81
 – scalare, 364
 complesso, 417
 – standard, 365
 prodotto
 – interno, 34
 – misto o triplo, 42
 – per uno scalare, 19
 – scalare, 33, 34
 – vettoriale, 40
 proiezione
 – ortogonale, 389, 391
 – esistenza e unicità, 401
 – formula analitica, 393
 matrice di, 412, 414
 – minimizza distanza, 391
 – su una retta, 371
 proiezione
 – ortogonale, 33, 34
 – su un piano, 39
 proprietà
 – associativa, 15
 – commutativa, 15
 pseudodiagonale, 465
 punto
 – di applicazione, 10
 – medio, 52

Q

quadrica, 486
 – a centro, 493
 – assi di simmetria, 494
 quadriche
 congruenti, 494
 quota, 32
 quoziante
 – di Rayleigh, 449

R

radice, 313
 rango, 70, 96
 – per colonne, 193
 – per righe, 193
 regola
 – del parallelogramma, 14
 – di Sarrus, 271
 relazione
 – di equivalenza, 6, 331
 – lineare, 169
 restrizione
 – di un operatore a un sottospazio invariante, 358
 retta, 165
 – di regressione lineare, 405
 rette
 – complanari, 5
 – incidenti, 5
 ortogonali, 7
 – parallele, 5
 – sghembe, 5
 riflessione
 – ortogonale, 427, 438
 rotazione, 488
 rototraslazione, 488

S

scalare, 18
 scambio, 267
 Schur
 – lemma di, 477
 segmento, 52
 segnatura, 457, 496
 sequilineare, 418
 sfera, 59
 unitaria, 450
 similitudine, 331, 339
 – criterio per matrici diagonalizzabili, 333
 – invarianti, 332
 sistema
 – di riferimento, 7, 22, 27
 – cartesiano, 31

- lineare, 69
- - (sovra/sotto)determinato, 103
 - equivalente, 90
- - omogeneo associato, 84
- soluzione
 - ai minimi quadrati, 406
- somma
 - diretta, 212, 354
 - di vettori, 13
- sostituzione all'indietro, 74
- sottospazio
 - complementare, 214
 - generato da un insieme di vettori, 165
 - intersezione, 205
 - invariante, 357
 - radicale, 343
 - somma, 206
 - vettoriale, 159
- spazi isomorfi, 237
- spazio
 - colonna di una matrice, 192
 - duale, 238, 253
 - euclideo, 364
 - funzionale, 157
 - hermitiano, 418
 - nullo, 85
 - riga di una matrice, 192
 - vettoriale, 4, 78, 153, 154
 - di dimensione finita, 153, 179
- spettro, 437
- successione finita, 239
- suriettiva, 227
- sviluppo di Laplace, 280, 281

T

- tensor d'inerzia, 431
- teorema
 - degli assi principali, 432, 494
 - del completamento della base, 182
 - della funzione implicita, 219
 - della funzione inversa, 218

- di Binet, 289
- di Carnot, 370, 419
- di Cramer, 97
 - di Eulero, 426
- di Hamilton-Cayley, 355
- di Kronecker, 292
- di nullità più rango, 196, 255
- di Pitagora, 31, 370, 379
- di rappresentazione di un'applicazione lineare, 242
- di Rouché-Capelli, 99
 - di Sylvester sul segno di una forma quadratica, 456
- rango per righe coincide col rango per colonne, 198
- spettrale, 434
- spettrale complesso, 479
- termine noto di un'equazione lineare, 71
- terna destrorsa, 32
- traccia, 225, 257, 312
 - è la somma degli autovalori, 314
- traslazione, 13

V

- valori singolari, 464
- variabili libere, 101
- verso, 10
- versore, 30, 47, 368, 417
- vettore
 - applicato, 10
 - colonna, 77
 - direzione, 47
 - libero, 12
 - nullo, 16, 71
 - posizione, 13
- vettori
 - linearmente (in)dipendenti, 168
 - ortogonali, 31, 368
 - - in \mathbb{C}^n , 418
- volume, 43, 401

UNA SCOMMessa DI CIVILTÀ

La nuova legge italiana sulle fotocopie è chiara.

È possibile fotocopiare una parte di un libro (fino al 15%) pagando, tramite la STAE, all'autore e all'editore un prezzo proporzionato alla parte riprodotta.

In questo modo, chi ha bisogno di leggere alcuni capitoli può evitare di acquistare l'opera intera.

Ma la fotocopia di tutto o di gran parte di un libro è illecita: induce al mancato acquisto, rendendo così vano il lavoro di chi il libro lo ha scritto, redatto, composto, impaginato e illustrato.

La legge si propone lo scopo di tenere vivo l'interesse a scrivere libri.

Se questo interesse venisse a mancare, ben pochi libri nuovi sarebbero pubblicati: saremmo tutti costretti a leggere fotocopie, ormai illeggibili, di libri vecchi e non aggiornati.

Fotocopiare tutto un libro è un po' come lasciare un'auto in seconda fila: i più non lo fanno, non solo per paura della multa, ma soprattutto perché si rendono conto che, se tutti si comportassero così, ne deriverebbe un danno generale.

Sta quindi ai lettori far sì che la legge funzioni e produca effetti positivi.

È una scommessa di civiltà: se la si vince, il premio non andrà solo ad autori ed editori, ma a tutto il sistema culturale e scientifico italiano.

- Nel sito www.zanichelli.it/f_info_fotocopie.html la normativa.

Nello stesso sito si darà comunicazione del giorno in cui la nuova normativa acquiserà piena efficacia.

La piena efficacia della nuova normativa infatti è subordinata alla stipulazione di accordi fra le categorie interessate.

L'editore mette a disposizione degli studenti non vedenti o con particolari problemi di apprendimento una copia dei file, solitamente in formato pdf, in cui sono memorizzate le pagine di questo libro. Il formato dei file permette l'ingrandimento dei caratteri del testo. I docenti o i responsabili educativi possono richiedere i file scrivendo a:
Zanichelli - Direzione Generale - Via Irnerio 34 - 40126 Bologna