

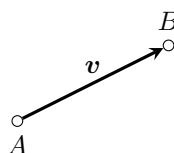
CAPITOLO 1

Elementi di algebra lineare

REFERENZE BIBLIOGRAFICHE. Edoardo Sernesi, *Geometria 1*, Bollati Boringhieri [Capitolo 1].
Gilbert Strang, *Introduction to linear algebra*, Wellesley–Cambridge press.

1. Introduzione agli spazi vettoriali

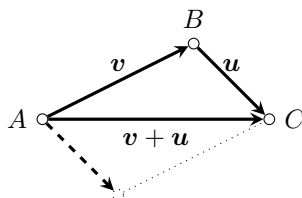
1.1. Vettori applicati e vettori geometrici. Per introdurre il concetto di *spazio vettoriale* iniziamo da un contesto più concreto e forse più semplice. Consideriamo due punti A e B nello spazio *ordinario*. Un *vettore applicato* $\mathbf{v} = \overrightarrow{AB}$ in A con punto finale B può essere immaginato come una freccia che collega A e B ,



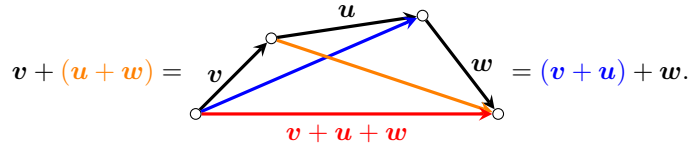
Un vettore applicato è caratterizzato dal suo punto di applicazione, A , da una direzione, un verso e una lunghezza. Se $A = B$, allora il corrispondente vettore è detto *nullo* e viene indicato con $\mathbf{0}$: la sua lunghezza è zero e la sua direzione e il suo verso non sono definiti. Due vettori $\mathbf{v} = \overrightarrow{AB}$ e $\mathbf{u} = \overrightarrow{CD}$ sono *equipollenti* se hanno stessa direzione, stessa lunghezza e stesso verso. L'equipollenza è una relazione di equivalenza e un *vettore geometrico* corrisponde ad una classe di equivalenza secondo equipollenza. Uno specifico vettore applicato è perciò un *rappresentante* di questa classe di equivalenza, mentre tutti i vettori con stessa direzione, stessa lunghezza e stesso verso corrispondono allo stesso oggetto di tipo “vettore geometrico”.

1.1.1. *Operazioni tra vettori geometrici.* I vettori geometrici sono oggetti molto diversi dai numeri usuali, ma è ugualmente possibile introdurre una serie di operazioni tra di essi.

Dati due vettori geometrici \mathbf{v} e \mathbf{u} , la loro *somma* $\mathbf{u} + \mathbf{v}$ può essere definita concatenando due loro rappresentanti. Per esempio, scelto un punto A , si applica prima \mathbf{v} ad A si ottiene il vettore \overrightarrow{AB} con coda in un punto B , e, consecutivamente, applicando \mathbf{u} a B si ottiene il vettore \overrightarrow{BC} con coda in C . Possiamo quindi *definire* \overrightarrow{AC} come rappresentante del vettore geometrico somma $\mathbf{v} + \mathbf{u}$. La costruzione ora descritta corrisponde alla cosiddetta *regola del parallelogramma*, ed è equivalente a considerare \mathbf{u} e \mathbf{v} applicati allo stesso punto A e costruire su di essi un parallelogramma, considerando quindi come somma la diagonale del quadrilatero ottenuto:

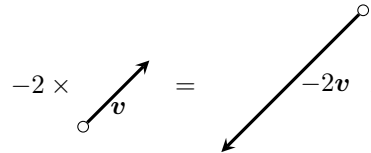


È evidente dalla costruzione stessa che l'operazione è commutativa, ovvero $\mathbf{v} + \mathbf{u} = \mathbf{u} + \mathbf{v}$, e tale per cui $\mathbf{v} + \mathbf{0} = \mathbf{v}$. L'operazione di somma che abbiamo definito è anche *associativa*. Se abbiamo tre vettori \mathbf{v} , \mathbf{u} e \mathbf{w}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



Infine, dato un vettore applicato \overrightarrow{AB} , il vettore applicato \overrightarrow{BA} è tale per cui $\overrightarrow{AB} + \overrightarrow{BA} = \mathbf{0}$: se il vettore geometrico associato a \overrightarrow{AB} è \mathbf{v} , denotiamo quindi $-\mathbf{v}$ il vettore geometrico associato a \overrightarrow{BA} e scriviamo $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$: $-\mathbf{v}$ è l'opposto di \mathbf{v} .

È possibile anche introdurre l'operazione *moltiplicazione per uno scalare* $c \in \mathbb{R}$: il vettore $c\mathbf{v}$ ha stessa direzione di \mathbf{v} , stesso verso se $c > 0$ e verso opposto se $c < 0$, e lunghezza uguale a $|c|$ volte quella di \mathbf{v} . Per esempio,



Se $c = 0$, allora $c\mathbf{v} = \mathbf{0}$. È facile verificare che, se $a, b \in \mathbb{R}$, allora $(a+b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$, $(ab)\mathbf{v} = a(b\mathbf{v})$ e, dati due vettori \mathbf{v} e \mathbf{u} , $a(\mathbf{v} + \mathbf{u}) = a\mathbf{v} + a\mathbf{u}$.

1.2. Spazi vettoriali. Lo spazio dei vettori geometrici può avere una interpretazione intuitiva abbastanza semplice. Può rappresentare, per esempio, lo spazio degli *spostamenti* nel piano. Tuttavia la sua struttura può essere generalizzata in forma astratta: essa è caratterizzata dalla presenza di due operazioni binarie (somma tra vettori e prodotto per scalare) e lo *spazio vettoriale* dei vettori geometrici è stato introdotto proprio per mezzo di queste operazioni.

L'idea ora è tentare di caratterizzare cosa sia uno spazio vettoriale senza necessariamente avere in mente una rappresentazione pittorica. In generale, la costruzione di uno spazio vettoriale richiede l'esistenza di un insieme \mathbb{V} di "vettori" e di un altro insieme di "scalari", \mathbb{K} . Ci limiteremo ad assumere che \mathbb{K} sia, a seconda dei casi \mathbb{Q} , \mathbb{R} o \mathbb{C} , ma utilizzeremo \mathbb{K} per lavorare in maggiore generalità.

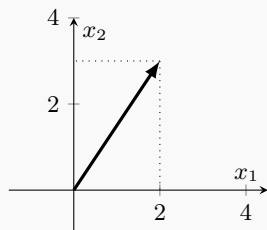
Il concetto di spazio vettoriale può quindi essere formalizzato come segue.

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

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

La definizione precedente si applica in effetti ai vettori geometrici, ed è anzi ispirata da essi. Come potete notare, è richiesto che \mathbb{K} sia un *campo*: come anticipato, per semplicità assumeremo che \mathbb{K} sia un campo numerico come \mathbb{R} , \mathbb{Q} e \mathbb{C} .

Un esempio di spazio vettoriale è dato dai vettori nel piano $\mathbb{V} = \mathbb{R}^2$. Un vettore nel piano \mathbf{x} è rappresentato da una freccia applicata nell'origine che punta in una posizione del piano



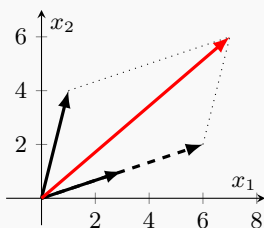
e può rappresentarsi come una coppia di numeri come

$$(1.1) \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2,$$

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

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

Per esempio, nella figura sotto $\mathbf{x}_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ e $\mathbf{x}_2 = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$, ed in rosso è rappresentata la loro somma $2\mathbf{x}_1 + \mathbf{x}_2 = \begin{pmatrix} 7 \\ 6 \end{pmatrix}$.



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

$$(1.3) \quad c_1 \mathbf{v}_1 + \dots + c_k \mathbf{v}_k =: \sum_{i=1}^k c_i \mathbf{v}_i.$$

Chiamiamo l'insieme di tutte le combinazioni lineari dei vettori $\mathcal{V} := \{\mathbf{v}_i\}_{i=1}^k$ lo *span* di $\{\mathbf{v}_i\}_{i=1}^k$ e scriviamo $\text{span}(\mathcal{V})$.

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

$$(1.4) \quad \sum_{i=1}^k c_i \mathbf{v}_i = \mathbf{0};$$

diversamente si dicono linearmente indipendenti.

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

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

$$(1.5) \quad \sum_{i=1}^k a_i \mathbf{v}_i = \sum_{i=1}^k b_i \mathbf{v}_i \Rightarrow a_i = b_i \quad \forall i = 1, \dots, k.$$

Infine, possiamo introdurre il concetto di *base*

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

Questo significa che ogni vettore $\mathbf{v} \in \mathbb{V}$ può scriversi in maniera unica come $\mathbf{v} = \sum_{i=1}^n c_i \mathbf{v}_i$, con coefficienti c_i che sono detti *coordinate* di \mathbf{v} secondo la base \mathcal{B} . Ciò comporta anche che non è possibile avere un set di vettori indipendenti con più di n elementi. Vale infatti il seguente

TEOREMA 1.2 (Dimensione). Sia $\mathcal{B} := \{\mathbf{v}_i\}_{i=1}^n$ una base di \mathbb{V} . Allora ogni set $\{\mathbf{u}_i\}_{i=1}^m$ di $m > n$ vettori di \mathbb{V} è costituito da elementi linearmente dipendenti. Di conseguenza, ogni base di \mathbb{V} ha n elementi: il numero n prende il nome di *dimensione* di \mathbb{V} .

Lo spazio vettoriale \mathbb{R}^n su \mathbb{R} è costituito da vettori n -dimensionali, rappresentabili come vettori colonna,

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

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

$$(1.6) \quad \mathbf{e}_1 := \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \mathbf{e}_2 := \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} \quad \dots \quad \mathbf{e}_n := \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

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

$$\mathbf{x} = x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \dots + x_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

1.2.1. *Sottospazi vettoriali.* Dato uno spazio vettoriale \mathbb{V} , è possibile identificare in alcuni casi un *sottospazio* vettoriale \mathbb{W} , ovvero un sottoinsieme di \mathbb{V} che è chiuso sotto le operazioni di somma e prodotto per uno scalare, ovvero

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

Per esempio, se $\mathcal{V} := \{\mathbf{v}_i\}_{i=1}^k \subset \mathbb{V}$, $\text{span}[\mathcal{V}]$ è un sottospazio di \mathbb{V} (in particolare, lo è anche per $k = 1$).

DEFINIZIONE 1.5 (Rango). Dato un insieme finito di vettori $\{\mathbf{v}_i\}_{i=1}^k$ di uno spazio vettoriale \mathbb{V} , il *rango* dell'insieme è la *dimensione* di $\text{span}[\{\mathbf{v}_i\}_{i=1}^k]$.

Inoltre la dimensione di un sottospazio è sempre limitata superiormente dalla dimensione dello spazio in cui vive. È interessante notare che se \mathbb{U} e \mathbb{W} sono sottospazi vettoriali di \mathbb{V} , allora $\mathbb{U} \cap \mathbb{W}$ è anche sottospazio vettoriale di \mathbb{V} (mentre la loro unione, in generale, non lo è).

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

Dati due sottospazi vettoriali U e W , se $U \cap W = \{\mathbf{0}\}$ è possibile costruire un nuovo spazio, detto *somma diretta di U e W* e indicato con $U \oplus W$. Questo spazio consiste di tutti i vettori \mathbf{v} nella forma $\mathbf{v} = \mathbf{u} + \mathbf{w}$ con $\mathbf{u} \in U$, $\mathbf{w} \in W$. Inoltre questa decomposizione è unica: se infatti assumessimo che esiste un'altra coppia tale per cui $\mathbf{v} = \mathbf{u}' + \mathbf{w}'$, con $\mathbf{u}' \in U$, $\mathbf{w}' \in W$, allora dovremmo avere $\mathbf{u} + \mathbf{w} = \mathbf{u}' + \mathbf{w}'$ e quindi $\mathbf{u} - \mathbf{u}' = \mathbf{w} - \mathbf{w}'$. Ma questa quantità può essere solo $\mathbf{0}$, unico elemento nell'intersezione dei due spazi.

2. Matrici

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

$$(1.7) \quad \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \quad a_{\mu\nu} \in \mathbb{K} \quad \forall \mu, \nu.$$

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

$$(1.8) \quad \mathbf{A} + \mathbf{B} := (a_{\mu\nu} + b_{\mu\nu})_{\mu\nu} \in \mathbb{M}_{m,n}(\mathbb{K}), \quad c\mathbf{A} := (ca_{\mu\nu})_{\mu\nu} \in \mathbb{M}_{m,n}(\mathbb{K}).$$

Queste operazioni soddisfano le proprietà richieste per rendere $\mathbb{M}_{m,n}(\mathbb{K})$ uno spazio vettoriale secondo la Definizione 1.1. Questo spazio ha dimensione mn . Una base è costituita dalle matrici $\mathbf{E}_{ij} = (\delta_{i\mu}\delta_{j\nu})_{\mu\nu}$ con soli elementi nulli eccezion fatta per la posizione (i, j) , dove compare 1.

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

$$(1.9) \quad \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \Rightarrow \mathbf{A}^\top = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \dots & \dots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix}$$

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

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

$\mathbb{M}_{n,m}(\mathbb{K})$ tale che

$$(1.10) \quad c_{\mu\nu} = \sum_{\rho=1}^k a_{\mu\rho} b_{\rho\nu}.$$

In particolare, se \mathbf{u} è un n -vettore riga

$$(1.11) \quad \mathbf{u} = (u_1 \quad \cdots \quad u_n) \in \mathbb{M}_{1,n}(\mathbb{K})$$

e \mathbf{v} un n -vettore colonna,

$$(1.12) \quad \mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{M}_{n,1}(\mathbb{K})$$

è possibile definire il *prodotto scalare* tra i due come

$$(1.13) \quad \mathbf{u}\mathbf{v} = \sum_{\nu=1}^n u_\nu v_\nu.$$

Questo prodotto produce quindi un elemento in \mathbb{K} . In particolare, si indica con $\|\mathbf{v}\|^2 := \mathbf{v}^\top \mathbf{v} = \sum_{\nu=1}^n v_\nu^2 \geq 0$ la *norma euclidea* di \mathbf{v} .

Si noti che l'operazione di prodotto tra matrici *non è abeliana* e anzi se $n \neq m$ il prodotto \mathbf{BA} è non definito. Anche qualora l'operazione fosse possibile, ovvero se $n = m$, $\mathbf{AB} \neq \mathbf{BA}$ in generale. Le proprietà del prodotto tra matrici sono date dalla seguente Proposizione, la cui dimostrazione ometteremo dato che avviene controllando esplicitamente la validità di ogni affermazione.

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

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

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

$$(1.14) \quad \mathbf{A} = (a_{\mu\nu})_{\mu\nu} = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$$

ovvero se solo gli elementi sulla sua diagonale sono non nulli. Tra le matrici diagonali la *matrice unità*

$$(1.15) \quad \mathbf{I}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

ha un ruolo speciale dato che si comporta come l'identità rispetto al prodotto tra matrici, ovvero $\mathbf{A}\mathbf{I}_n = \mathbf{I}_n\mathbf{A} = \mathbf{A}$. Infine, una matrice quadrata $\mathbf{A} \in \mathbb{K}$ è *simmetrica* se $\mathbf{A} = \mathbf{A}^\top$ e *antisimmetrica* se $\mathbf{A} = -\mathbf{A}^\top$. Infine, la *traccia* di una matrice quadrata è uguale alla somma dei suoi elementi sulla diagonale.

Elenchiamo ora alcune proprietà delle matrici quadrate.

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

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

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

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

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

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

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

2.2.1. Determinante. Ad ogni matrice quadrata può essere associato scalare in \mathbb{K} che, come vedremo, è estremamente ricco di informazioni sulle sue proprietà. Questo scalare è il *determinante*. La sua definizione è relativamente complessa:

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

$$(1.16) \quad \det \mathbf{A} = \sum_{\sigma \in \mathcal{S}_n} (-1)^{n(\sigma)} \prod_{i=1}^n a_{i\sigma(i)}.$$

Il determinante di $\det \mathbf{A}$ si indica spesso anche come $|\mathbf{A}|$. La complicazione e stranezza della definizione sta nella comparsa della somma su *tutte le permutazioni di n elementi* \mathcal{S}_n e in un prefattore che fa riferimento al numero di trasposizioni $n(\sigma)$ che producono σ , ovvero alla *parità* della permutazione. La definizione sopra è piuttosto convoluta e difficile da applicare. Estremamente utile è il seguente teorema, che non dimostriamo, dovuto a Laplace.

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

$$(1.17) \quad \det \mathbf{A} = \sum_{\nu=1}^n a_{\mu\nu} c_{\mu\nu} = \sum_{\nu=1}^n a_{\nu\mu} c_{\nu\mu} = (\mathbf{AC}^\top)_{\mu\mu}, \quad c_{\mu\nu} := (-1)^{\mu+\nu} \det(\mathbf{A}_{\mu\nu})$$

indipendentemente da μ , dove $\mathbf{A}_{\mu\nu} \in \mathbb{K}^{(n-1) \times (n-1)}$ è la matrice $(n-1) \times (n-1)$ ottenuta da \mathbf{A} rimuovendo la μ -esima riga e la ν -esima colonna. Inoltre, per definizione dato uno scalare $c \in \mathbb{K}$, $\det(c) = c$.

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

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

$$(1.18) \quad \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \Rightarrow \det \mathbf{A} = a_{11}a_{22} - a_{12}a_{21}.$$

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

$$(1.19) \quad \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \Rightarrow \det \mathbf{A} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12}.$$

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

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

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

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

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

- se \mathbf{A} è una matrice diagonale, allora $\det \mathbf{A} = \prod_{i=1}^n a_{ii}$.
- $\det \mathbf{A} = \det \mathbf{A}^\top$ (e di conseguenza proprietà riferite alle colonne si applicano anche alle righe).
- Dato uno scalare $c \in \mathbb{K}$, $\det(c\mathbf{A}) = c^n \det \mathbf{A}$.
- Sia una colonna di \mathbf{A} tale che $\mathbf{a}^\nu = \lambda \mathbf{v} + \mathbf{u}$, con $\mathbf{v}, \mathbf{u} \in \mathbb{K}^n$ e $c \in \mathbb{K}$, i.e., $a_{\mu\nu} = \lambda v_\nu + u_\nu$. Allora

$$(1.20) \quad \det \mathbf{A} = \det \begin{pmatrix} a_{11} & \dots & a_{1\nu} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{n\nu} & \dots & a_{nn} \end{pmatrix} \\ = \lambda \det \begin{pmatrix} a_{11} & \dots & v_1 & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & v_n & \dots & a_{nn} \end{pmatrix} + \det \begin{pmatrix} a_{11} & \dots & u_1 & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & u_n & \dots & a_{nn} \end{pmatrix}$$

Si dice che di conseguenza il determinante è multilineare.

- Se \mathbf{A} ha due colonne identiche, $\det \mathbf{A} = 0$.
- Sia $\mathbf{B} \in \mathbb{M}_n(\mathbb{K})$; allora $\det \mathbf{AB} = \det \mathbf{A} \det \mathbf{B}$.

Una delle conseguenze delle proprietà suddette è che il determinante è che, se \mathbf{A} è una matrice invertibile, $\det \mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}}$. Inoltre, il determinante è *alternante*, ovvero guadagna un segno se due colonne vengono scambiate:

$$(1.21) \quad \det \begin{pmatrix} a_{11} & \dots & a_{1k} & \dots & a_{1k'} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nk} & \dots & a_{nk'} & \dots & a_{nn} \end{pmatrix} = -\det \begin{pmatrix} a_{11} & \dots & a_{1k'} & \dots & a_{1k} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nk'} & \dots & a_{nk} & \dots & a_{nn} \end{pmatrix}$$

Il ruolo cruciale del determinante è legato anche al seguente lemma che caratterizza l'invertibilità delle matrici quadrate.

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

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

Allora

$$(1.22) \quad \mathbf{A}^\top \mathbf{C} = \det \mathbf{A} \mathbf{I}_n.$$

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

$$(1.23) \quad \mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \mathbf{C}^\top.$$

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

Questo lemma implica il fatto che il determinante di \mathbf{A} ne *determina* in effetti la sua invertibilità: se $\det \mathbf{A} = 0$ la matrice quadrata \mathbf{A} non è invertibile. Inoltre, come conseguenza di quanto detto sul rango delle matrici quadrate, il rango di una matrice è massimo se e solo se il suo determinante è non nullo.

3. Sistemi di equazioni lineari

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

$$(1.24) \quad \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

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

$$(1.25) \quad \mathbf{Ax} = \mathbf{b}.$$

Il sistema di equazioni è detto *omogeneo* se $\mathbf{b} = \mathbf{0}$, mentre diversamente è detto *non omogeneo*. Ogni sistema è detto *compatibile* se ha almeno una soluzione: notare che un sistema omogeneo ha sempre la soluzione *banale* $\mathbf{x} = \mathbf{0}$. Se è dato un sistema non omogeneo $\mathbf{Ax} = \mathbf{b}$, $\mathbf{b} \neq \mathbf{0}$, allora il sistema omogeneo ad esso associato è $\mathbf{Ax} = \mathbf{0}$. Inoltre, se \mathbf{x}_b è soluzione del sistema non omogeneo e \mathbf{x}_0 è una soluzione non banale del sistema omogeneo, una nuova soluzione del sistema non omogeneo può essere ottenuta considerando $\mathbf{x}_b + c\mathbf{x}_0$ per qualunque $c \in \mathbb{K}$.

3.1. Metodo di eliminazione di Gauss. Risolvere un sistema tipo $\mathbf{Ax} = \mathbf{b}$ è un problema centrale dell'algebra lineare. L'idea generale è utilizzare delle semplici operazioni matriciali per poter infine ottenere una espressione per la soluzione \mathbf{x} , se essa esiste, o concludere che il sistema non ha soluzione se questo è il caso. Un approccio classico e istruttivo è il cosiddetto *metodo di eliminazione di Gauss*. Supponiamo di avere un sistema come in Eq. (1.24). Questo può essere

associato alla seguente matrice

$$(1.26) \quad \mathbf{H} = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \dots & \dots & \ddots & \vdots & \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array} \right)$$

ottenuta concatenando la matrice \mathbf{A} e la colonna \mathbf{b} e talvolta detta *matrice orlata*. Ogni matrice siffatta corrisponde ad un sistema come in Eq. (1.24). Possiamo eseguire una serie di operazioni di riga su questa matrice che lasciano inalterato il problema. In particolare

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

Questo tipo di operazioni può essere utilizzato per risolvere il sistema di equazioni lineari, dato che alterare la matrice orlata per via di queste operazioni non cambia, di fatto, il sistema di equazioni lineari associato. Il metodo consiste nel procedere come segue.

Si considera h_{11} , detto *primo pivot*, assumendolo diverso da zero (se $h_{11} = 0$, si scambia la prima riga con un'altra che ha primo elemento non nullo, cosa sempre possibile se la variabile x_1 appare nel sistema). A questo punto si sottrae a ogni riga \mathbf{H}_μ con $\mu > 1$ la prima riga con un adeguato moltiplicatore, ovvero

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

In questo modo, $h_{\mu 1} \mapsto 0$ in tutte le righe. A questo punto si ordinano le righe dalla seconda in giù in modo che la nuova seconda riga sia una delle righe con l'elemento non nullo più a sinistra. Questo elemento sarà il secondo pivot. Si ripete quindi l'operazione come sopra mettendo a zero tutti gli elementi sotto il secondo pivot. Si procede così fino all'ultima riga.

Vediamo come questa procedura funziona con un esempio.

Cerchiamo di risolvere il sistema

$$(1.27) \quad \begin{cases} 2x_1 + 4x_2 - 2x_3 = 2 \\ 4x_1 + 9x_2 - 3x_3 = 8 \\ -2x_1 - 3x_2 + 7x_3 = 10 \end{cases} \Leftrightarrow \begin{pmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \\ 10 \end{pmatrix}$$

così che la matrice orlata corrispondente è

$$(1.28) \quad \left(\begin{array}{ccc|c} 2 & 4 & -2 & 2 \\ 4 & 9 & -3 & 8 \\ -2 & -3 & 7 & 10 \end{array} \right)$$

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

$$(1.29) \quad \left(\begin{array}{ccc|c} 2 & 4 & -2 & 2 \\ 0 & 1 & 1 & 4 \\ 0 & 1 & 5 & 12 \end{array} \right)$$

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

$$(1.30) \quad \left(\begin{array}{ccc|c} 2 & 4 & -2 & 2 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 4 & 8 \end{array} \right)$$

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

all'insieme di equazioni

$$(1.31) \quad \begin{cases} 2x_1 + 4x_2 - 2x_3 = 2 \\ x_2 + x_3 = 4 \\ 4x_3 = 8 \end{cases}$$

che può essere risolto dal basso verso l'alto, ovvero risolvendo prima per x_3 , ottenendo $x_3 = 2$, la cui soluzione può essere sostituita nella seconda riga che dà $x_2 = 2$ e infine inserendo entrambe queste soluzioni nella prima, che fornisce $x_1 = -1$.

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

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

Torniamo ora alla nostra matrice orlata: a seguito della procedura di Gauss, se $m \leq n$, essa avrà la forma a gradini

$$(1.32) \quad \left(\begin{array}{cccccc|c} a_{11}^* & a_{12}^* & \dots & a_{1m}^* & \dots & a_{1n}^* & b_1^* \\ 0 & a_{22}^* & \dots & a_{2m}^* & \dots & a_{2n}^* & b_2^* \\ \vdots & \vdots & \ddots & \dots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & a_{mm}^* & \dots & a_{mn}^* & b_m^* \end{array} \right).$$

Supponiamo che $\prod_{i=1}^m a_{ii}^* \neq 0$: questa condizione equivale a dire che *ogni colonna ha il suo pivot*. L'esempio mostra che, se $m = n$ e il nostro algoritmo ci fornisce la matrice orlata a gradini, allora il sistema è risolubile: l'ultima riga fornisce l'equazione per x_n , la penultima per x_{n-1} e così via: in questo caso la soluzione del sistema è *unica*. Se $\prod_{i=1}^m a_{ii} = 0$ (ovvero, una qualche colonna difetta di pivot), il sistema può avere un sottospazio di possibili soluzioni o essere incompatibile, come mostrato dai due seguenti esempi.

Cerchiamo di risolvere il sistema

$$(1.33) \quad \begin{cases} 2x_1 + 4x_2 - 2x_3 = 2 \\ 4x_1 + 8x_2 - 3x_3 = 8 \\ -2x_1 - 4x_2 + 7x_3 = 10 \end{cases} \Leftrightarrow \begin{pmatrix} 2 & 4 & -2 \\ 4 & 8 & -3 \\ -2 & -4 & 7 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \\ 10 \end{pmatrix}$$

così che la matrice orlata è

$$(1.34) \quad \mathbf{H} = \left(\begin{array}{ccc|c} 2 & 4 & -2 & 2 \\ 4 & 8 & -3 & 8 \\ -2 & -4 & 7 & 10 \end{array} \right)$$

Seguiamo la procedura di Gauss. Al primo step otteniamo

$$(1.35) \quad \left(\begin{array}{ccc|c} 2 & 4 & -2 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 5 & 12 \end{array} \right)$$

mentre al secondo otteniamo la nostra matrice a gradini

$$(1.36) \quad \mathbf{H}^* = \left(\begin{array}{ccc|c} 2 & 4 & -2 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & -8 \end{array} \right)$$

L'ultima riga in particolare corrisponde ad una equazione *falsa*, ovvero $0 = -8$. Questo significa che il sistema è *incompatibile*.

Consideriamo ora il sistema

$$(1.37) \quad \begin{cases} 2x_1 + 4x_2 - 2x_3 = 2 \\ 4x_1 + 8x_2 - 3x_3 = 8 \\ -8x_1 - 16x_2 + 6x_3 = 16 \end{cases} \Leftrightarrow \begin{pmatrix} 2 & 4 & -2 \\ 4 & 8 & -3 \\ -8 & -16 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \\ 16 \end{pmatrix}$$

così che la matrice orlata è

$$(1.38) \quad \mathbf{H} = \left(\begin{array}{ccc|c} 2 & 4 & -2 & 2 \\ 4 & 8 & -3 & 8 \\ -8 & -16 & 6 & 16 \end{array} \right)$$

Seguiamo la procedura di Gauss. Già al primo step otteniamo

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

mentre al secondo

$$(1.40) \quad \mathbf{H}^* = \left(\begin{array}{ccc|c} 2 & 4 & -2 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

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

Se $m > n$ la matrice orlata a gradini finale avrà una forma tipo

$$(1.41) \quad \left(\begin{array}{cccccc|c} a_{11} & a_{12} & \dots & a_{1m} & \dots & a_{1n} & b_1 \\ 0 & a_{22} & \dots & a_{2m} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \dots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \dots & a_{nn} & b_n \\ 0 & 0 & \dots & 0 & \dots & 0 & b_{n+1} \\ 0 & 0 & \dots & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 0 & 0 \end{array} \right).$$

Se, per via della procedura, $b_{n+1} = 0$, allora possiamo trascurare tutte le righe dalla $(n + 1)$ esima in poi, dato che non sono informative, e ci riduciamo ad un caso in cui di fatto $m = n$. Diversamente, se uno dei valori $b_{n+1} \neq 0$ allora la procedura ha prodotto una condizione falsa, $0 = b_{n+1} \neq 0$, che vuol dire che il nostro sistema è incompatibile e non esistono soluzioni.

3.2. Rango e compatibilità. I vari casi descritti sopra possono essere riassunti introducendo una nuova quantità relativa alla matrice a gradini ottenuta alla fine della procedura corrispondente.

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

Naturalmente il rango di una matrice è minore o uguale alla più piccola delle sue dimensioni.

Nel primo esempio, partendo dalla matrice orlata in Eq. (1.28), siamo arrivati alla matrice a gradini in Eq. (1.30). Osservando quest'ultima, il rango della matrice \mathbf{H} è uguale a 3. Se ci focalizziamo sulla sottomatrice corrispondente ad \mathbf{A} , abbiamo trasformato

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

La matrice a gradini \mathbf{A}^* contiene anch'essa tre pivot, quindi il rango di \mathbf{A} qui è 3. Viceversa, nell'esempio in Eq. (1.34), la matrice orlata produce una matrice a gradini con tre pivot, e quindi \mathbf{H} ha rango 3, mentre abbiamo ottenuto

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

La matrice a gradini contiene due pivot, quindi il rango di \mathbf{A} qui è 2. Si noti che se la matrice orlata avesse avuto solo elementi nulli nell'ultima riga, il rango della matrice orlata sarebbe stato 3.

Il rango di una matrice ha anche una diversa, interessante e importante interpretazione: è infatti la dimensione dello spazio generato dalle sue righe o sue colonne, come stabilito dal seguente teorema, che diamo senza dimostrazione.

TEOREMA 3.1 (Teorema del rango). *Sia $\mathbf{A} \in \mathbb{M}_{m,n}(\mathbb{K})$ una matrice di rango r . Allora, denotando $\{\mathbf{a}_\mu\}_{\mu=1}^m$ e $\{\mathbf{a}^\nu\}_{\nu=1}^n$ le sue colonne*

$$(1.42) \quad \text{rank}(\mathbf{A}) = \dim(\text{span}[\{\mathbf{a}_\mu\}_{\mu=1}^m]) = \dim(\text{span}[\{\mathbf{a}^\nu\}_{\nu=1}^n]).$$

Questa ulteriore informazione chiarifica ancor meglio il fatto che il rango di una matrice sia legato alla risolubilità di un sistema lineare è suggerito dal seguente fatto. Consideriamo il problema $\mathbf{A}\mathbf{x} = \mathbf{b}$: se indichiamo con \mathbf{a}^ν la ν -esima colonna di \mathbf{A} , questo problema può scriversi come

$$(1.43) \quad \sum_{\nu=1}^n x_\nu \mathbf{a}^\nu = \mathbf{b},$$

che esprime il fatto che stiamo cercando di scrivere \mathbf{b} come sovrapposizione lineare dei vettori colonna di \mathbf{A} , $\{\mathbf{a}^\nu\}_{\nu=1}^n$. Ciò sarà possibile se \mathbf{b} vive nello spazio generato dalle colonne di \mathbf{A} , $\text{span}[\{\mathbf{a}^\nu\}_{\nu=1}^n]$, la cui dimensione è proprio il rango di \mathbf{A} : in altre parole, perché il sistema abbia soluzione \mathbf{b} non deve vivere al di fuori dello spazio dei vettori colonna. Questa intuizione è espressa in termini rigorosi da un teorema fondamentale dell'algebra lineare, che diamo senza dimostrazione.

TEOREMA 3.2 (Kronecker–Rouché–Capelli). *Un sistema lineare di m equazioni in n incognite $\mathbf{A}\mathbf{x} = \mathbf{b}$ è compatibile se e solo se la matrice orlata ha lo stesso rango della matrice \mathbf{A} . In tal caso, lo spazio delle soluzioni ha dimensione $n - \text{rank}(\mathbf{A})$.*

Consideriamo il seguente sistema:

$$(1.44) \quad \begin{cases} x_1 - 2x_3 = -2 \\ 2x_1 + 2x_2 - 3x_3 = -3 \\ 3x_1 + 4x_2 - 4x_3 = 4 \end{cases}$$

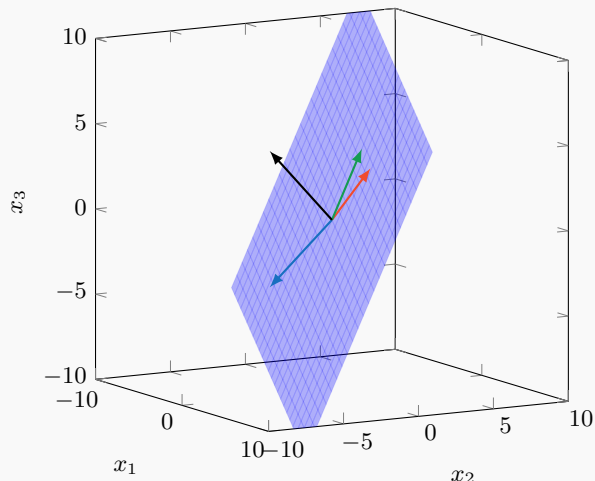
Applicando la procedura di Gauss otteniamo

$$(1.45) \quad \left(\begin{array}{ccc|c} 1 & 0 & -2 & -2 \\ 2 & 2 & -3 & -3 \\ 3 & 4 & -4 & 4 \end{array} \right) \mapsto \left(\begin{array}{ccc|c} 2 & 0 & -2 & -2 \\ 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & 8 \end{array} \right)$$

Il rango di \mathbf{A} e della matrice orlata sono diversi: la matrice orlata ha rango 3, mentre \mathbf{A} ha rango 2: il sistema è quindi incompatibile. Tentiamo di interpretare geometricamente questo risultato. Il fatto che il rango di \mathbf{A} sia 2 vuol dire che i suoi vettori colonna vivono in uno spazio *bidimensionale*. In effetti, se rappresentiamo graficamente i vettori

$$(1.46) \quad \mathbf{a}^1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \mathbf{a}^2 = \begin{pmatrix} 0 \\ 2 \\ 4 \end{pmatrix}, \quad \mathbf{a}^3 = \begin{pmatrix} -2 \\ -3 \\ -4 \end{pmatrix},$$

si osserva che i vettori colonna della matrice \mathbf{A} vivono tutti sullo stesso *piano*, che rappresenta $\text{span}\{\{\mathbf{a}^\nu\}_{\nu=1}^3\}$, mentre \mathbf{b} vive al di fuori di questo piano:



Di conseguenza non è possibile rappresentare \mathbf{b} usando i vettori colonna di \mathbf{A} e quindi trovare una soluzione al sistema $\mathbf{Ax} = \mathbf{b}$.

3.3. Metodo di Gauss–Jordan per l’inversa. Dato il problema $\mathbf{Ax} = \mathbf{b}$, se $\mathbf{A} \in \mathbb{A}_n(\mathbb{K})$, il metodo di Gauss descritto sopra permette di “invertire”, quando possibile, la matrice \mathbf{A} : se $\det \mathbf{A} \neq 0$ allora in effetti $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. In qualche modo, quindi, il metodo produce esattamente il risultato di questa operazione applicando l’inversa di \mathbf{A} a \mathbf{b} , quando questa esiste. Può quindi essere utilizzato per ottenere l’inversa di una matrice. Come sappiamo, la matrice inversa \mathbf{A}^{-1} è tale per cui $\mathbf{AA}^{-1} = \mathbf{I}_n$. Se denoto con \mathbf{x}^ν la i -esima colonna di \mathbf{A}^{-1}

$$(1.47) \quad \mathbf{Ax}^\nu = \mathbf{e}_\nu, \quad \text{dove } \mathbf{e}_\nu = (\delta_{\mu\nu})_\mu \text{ per } i = 1, \dots, n.$$

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

Consideriamo la matrice quadrata

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

Per trovare l’inversa

$$\mathbf{A}^{-1} = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix}$$

occorre risolvere le equazioni

$$\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{21} \\ x_{31} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{21} \\ x_{31} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{21} \\ x_{31} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

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

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

Questa matrice orlata ha la forma

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

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

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

A questo punto vorremmo mettere il sistema nella forma

$$(\mathbf{I} \mid \mathbf{B})$$

perché in questo caso $\mathbf{B} = \mathbf{A}^{-1}$! La matrice orlata così ottenuta corrisponderebbe infatti all'insieme di equazioni del tipo $x_{\mu\nu} = b_{\mu\nu}$ fornendoci la soluzione che cerchiamo. Dobbiamo quindi lavorare per rimuovere gli zeri *sopra* la diagonale procedendo stavolta dal basso verso l'alto.

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

Dividendo opportunamente ogni riga otteniamo

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

per cui

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

Il metodo descritto è detto *di Gauss-Jordan*. La procedura fallisce se l'inversa cercata non esiste: questo è il caso se, *dopo* aver ottenuto la forma a gradini \mathbf{A}^* , $\prod_{i=1}^n a_{ii} = 0$, ovvero qualche colonna difetta di pivot. Il fatto che nel caso di una matrice \mathbf{A} invertibile ogni colonna abbia il suo pivot equivale a dire che il rango della matrice è n :

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

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

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

3.4. Regola di Cramer. Il caso di un sistema lineare $\mathbf{A}\mathbf{x} = \mathbf{b}$ dove $\mathbf{b} \in \mathbb{M}_n(\mathbb{K})$ è particolarmente interessante dato che, come abbiamo appena visto, se $\det \mathbf{A} \neq 0$ allora $n = \text{rank}(\mathbf{A})$ e, per il teorema di Kronecker–Rouché–Capelli, il sistema è compatibile con unica soluzione. In questo caso, uesta soluzione si può facilmente mettere nella forma

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

Esplicitando, questo vuol dire che

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

Ma dalla forma per l'inversa data dall'Eq. (1.23) possiamo scrivere

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

dove $C = (c_{\mu\nu})_{\mu\nu}$ è la matrice dei cofattori di A . Cramer ha osservato l'interessante fatto che l'espressione nella somma ha la forma in Eq. (1.17) *come se* avessimo sostituito alla colonna μ -esima di A il vettore b : in altre parole, se indichiamo con A_μ la matrice in cui la μ -esima colonna è sostituita da b , allora

$$(1.48) \quad x_\mu = \frac{\det A_\mu}{\det A}.$$

4. Applicazioni lineari, endomorfismi e proprietà spettrali

4.1. Applicazioni lineari. Finora abbiamo studiato le proprietà di singoli spazi vettoriali, ma non abbiamo sistematicamente considerato delle funzioni tra essi. Ciò è naturalmente possibile (l'operazione di trasposizione ne è un esempio) e c'è una ampia libertà nel definire possibili funzioni tra spazi diversi. Tra esse vi è tuttavia una famiglia ha una particolare importanza.

DEFINIZIONE 4.1 (Applicazioni lineari). *Siano V e W due \mathbb{K} -spazi vettoriali. Una applicazione $F: V \rightarrow W$ è lineare se, per ogni $v_1, v_2 \in V$ e ogni $c_1, c_2 \in \mathbb{K}$, vale*

$$(1.49) \quad F(c_1 v_1 + c_2 v_2) = c_1 F(v_1) + c_2 F(v_2).$$

Se $V = W$ l'applicazione è detto endomorfismo.

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

$$(1.50) \quad A(c_1 v_1 + c_2 v_2) = c_1(Av_1) + c_2(Av_2), \quad c_1, c_2 \in \mathbb{K}, \quad v_1, v_2 \in \mathbb{K}^n.$$

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

$$(1.51) \quad F(v_i) = \sum_{j=1}^m w_j f_{ji}.$$

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

$$(1.52) \quad M_{wv}(F) = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{m1} & f_{m2} & \cdots & f_{mn} \end{pmatrix} \in \mathbb{M}_{m,n}(\mathbb{K}).$$

Ora, dato un qualunque $\mathbf{v} = \sum_{i=1}^n c_i \mathbf{v}_i$, abbiamo che

$$(1.53) \quad F(\mathbf{v}) = F\left(\sum_{i=1}^n c_i \mathbf{v}_i\right) = \sum_{i=1}^n c_i F(\mathbf{v}_i) = \sum_{j=1}^m \mathbf{w}_j \sum_{i=1}^n c_i f_{ji} \equiv \sum_{j=1}^m \mathbf{w}_j \hat{c}_j,$$

dove \hat{c}_j è il coefficiente relativo alla base dello spazio di arrivo. Ovvero

$$(1.54) \quad \begin{pmatrix} \hat{c}_1 \\ \vdots \\ \hat{c}_m \end{pmatrix} = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{m1} & f_{m2} & \cdots & f_{mn} \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}.$$

Notare che il vettore $(c_1, \dots, c_n)^\top$ fa riferimento alle componenti di \mathbf{v} secondo la base \mathcal{B}_v , mentre il vettore $(\hat{c}_1, \dots, \hat{c}_m)^\top$ fa riferimento alle componenti di $F(\mathbf{v})$ secondo la base \mathcal{B}_w . In altre parole, date due basi \mathcal{B}_v e \mathcal{B}_w di due spazi finito dimensionali \mathbb{V} e \mathbb{W} , ogni applicazione lineare $F: \mathbb{V} \rightarrow \mathbb{W}$ corrisponde univocamente ad una matrice $\mathbf{M}_{wv}(F) \in \mathbb{M}_{m,n}(\mathbb{K})$. Questa corrispondenza può essere portata oltre e verificata nel caso di composizione di applicazioni lineari. Per esempio, supponiamo di avere tre spazi vettoriali \mathbb{V} , \mathbb{W} e \mathbb{U} di dimensione n , m e k rispettivamente, e due applicazioni lineari $F: \mathbb{V} \rightarrow \mathbb{W}$ e $G: \mathbb{W} \rightarrow \mathbb{U}$. Come ottenere $G \circ F: \mathbb{V} \rightarrow \mathbb{U}$? Se abbiamo tre basi \mathcal{B}_v , \mathcal{B}_w e \mathcal{B}_u per i tre spazi rispettivamente, è possibile mostrare che

$$\mathbf{M}_{uv}(G \circ F) = \mathbf{M}_{uw}(G) \mathbf{M}_{wv}(F).$$

4.2. Endomorfismi. Un discorso speciale meritano gli *endomorfismi*, ovvero le applicazioni lineari da \mathbb{V} a se stesso. In questo caso la base di partenza è uguale a quella di arrivo, sia essa per esempio \mathcal{B}_v . Indichiamo con $\mathbf{M}_v(F)$ la matrice quadrata associata a F secondo questa base. Gli endomorfismi in \mathbb{V} possono essere raggruppati in classi sulla base della relazione di equivalenza di *similitudine*, definita rispetto alle matrici che li rappresentano.

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

Due matrici simili hanno lo stesso determinante. Inoltre, la rilevanza di questa definizione sta nella seguente

PROPOSIZIONE 4.1. Sia \mathbb{V} un \mathbb{K} -spazio n -dimensionale. Due matrici $\mathbf{A}, \mathbf{B} \in \mathbb{M}_n(\mathbb{K})$ sono simili se e solo se esiste un endomorfismo F su \mathbb{V} e due basi \mathcal{B}_e e \mathcal{B}_w tali per cui $\mathbf{A} = \mathbf{M}_e(F)$ e $\mathbf{B} = \mathbf{M}_w(F)$.

In altre parole, matrici simili rappresentano lo stesso endomorfismo in basi diverse: $\mathbf{M}_e(F) = \mathbf{S}^{-1} \mathbf{M}_f(F) \mathbf{S}$ e ciò che \mathbf{S} esegue è un semplice cambiamento di base. La matrice \mathbf{S} è costituita dalla rappresentazione di una in termini dell'altra. Supponiamo infatti di avere due basi, $\mathcal{B}_e = \{\mathbf{e}_i\}_{i=1}^n$ e $\mathcal{B}_w = \{\mathbf{w}_i\}_{i=1}^n$, e che valgano le seguenti decomposizioni:

$$(1.55) \quad \mathbf{e}_\mu = \sum_{\nu=1}^n s_{\mu\nu} \mathbf{w}_\nu \quad \mathbf{w}_\mu = \sum_{\nu=1}^n w_{\mu\nu} \mathbf{e}_\nu.$$

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

$$\mathbf{e}_\mu = \sum_{\nu=1}^n s_{\mu\nu} \mathbf{w}_\nu = \sum_{\nu=1}^n \sum_{\rho=1}^n s_{\mu\nu} w_{\nu\rho} \mathbf{e}_\rho = \sum_{\rho=1}^n \left[\sum_{\nu=1}^n s_{\mu\nu} w_{\nu\rho} \right] \mathbf{e}_\rho = \sum_{\rho=1}^n [\mathbf{S} \mathbf{W}]_{\nu\rho} \mathbf{e}_\rho.$$

Ora, \mathcal{B}_e è una base, e di conseguenza \mathbf{e}_μ è linearmente indipendente dai restanti componenti della base stessa: l'unico modo perché la relazione sopra sia consistente è che $[\mathbf{S} \mathbf{W}]_{\nu\rho} = \delta_{\nu\rho}$, che vuol dire $\mathbf{W} = \mathbf{S}^{-1}$.

Torniamo ora alla nostra applicazione lineare F . Per definizione,

$$F(\mathbf{w}_\mu) = \sum_{\nu=1}^n [\mathbf{M}_w]_{\mu\nu} \mathbf{w}_\nu.$$

Vediamo ora come si comporta F sulla base \mathcal{B}_e :

$$F(\mathbf{e}_\mu) = \sum_{\nu=1}^n e_{\mu\nu} F(\mathbf{w}_\nu) = \sum_{\nu\rho} s_{\mu\nu} [\mathbf{M}_w]_{\nu\rho} \mathbf{w}_\rho = \sum_{\kappa=1}^n \left[\sum_{\nu\rho\kappa} s_{\mu\nu} [\mathbf{M}_w]_{\nu\rho} w_{\rho\kappa} \right] \mathbf{e}_\kappa \equiv \sum_{\kappa=1}^n [\mathbf{S} \mathbf{M}_w \mathbf{W}]_{\mu\kappa} \mathbf{e}_\kappa$$

che per definizione significa che $\mathbf{M}_e = \mathbf{S} \mathbf{M}_w \mathbf{S}^{-1}$.

Ha senso quindi chiedersi se, tra tutte le basi, ne esiste una che fornisce una rappresentazione particolarmente semplice dell'endomorfismo.

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

$$(1.56) \quad \mathbf{M}_e(F) = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

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

Supponiamo di avere la rappresentazione matriciale $\mathbf{M}_w(F)$ di una applicazione in una base $\mathcal{B}_w = \{\mathbf{w}_i\}_{i=1}^n$. Dire che $\mathbf{M}_w(F)$ è diagonalizzabile significa che esiste una diversa base $\mathcal{B}_e = \{\mathbf{e}_i\}_{i=1}^n$ in cui $\mathbf{M}_e(F)$ è diagonale. Il passaggio dalla base \mathcal{B}_w alla base \mathcal{B}_e sarà dato da una matrice \mathbf{S} che, come nell'Eq. (1.55), semplicemente esprime gli elementi della base \mathcal{B}_e in termini degli elementi della base originaria \mathcal{B}_w .

La definizione di endomorfismo diagonalizzabile implica che l'endomorfismo F ha la proprietà $F(\mathbf{e}_i) = \lambda_i \mathbf{e}_i$. Il vettore \mathbf{e}_i si dice *autovettore* corrispondente all'autovalore λ_i : esso corrisponde ad una “direzione speciale” nello spazio vettoriale in cui l'azione di F si riduce semplicemente ad un cambio di scala e possibilmente un'inversione (se $\lambda_i < 0$) mentre la direzione rimane inalterata. È facile quindi vedere che, su un autovettore, $F^{\otimes 2}(\mathbf{e}_i) := F(F(\mathbf{e}_i)) = F(\lambda_i \mathbf{e}_i) = \lambda_i F(\mathbf{e}_i) = \lambda_i^2 \mathbf{e}_i$ e in generale applicando $F^{\otimes k}(\mathbf{e}_i) = \lambda_i^k \mathbf{e}_i$. L'azione di F sui suoi autovettori è quindi estremamente semplice. Rimane il problema di individuarli.

4.3. Polinomio caratteristico. Data una matrice quadrata $\mathbf{A} \in \mathbb{M}_n(\mathbb{K})$, per trovare un autovettore e il corrispondente autovalore dovremmo risolvere l'equazione $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, dove sia λ che \mathbf{x} sono incogniti. Se esistono un $\mathbf{x} \neq \mathbf{0}$ e un λ tali per cui l'equazione è valida, essi sono rispettivamente autovettore e autovalore di \mathbf{A} . Si noti che \mathbf{x} non è univocamente determinato: se \mathbf{x} è autovettore di \mathbf{A} con autovalore λ , anche $c\mathbf{x}$ è autovettore di \mathbf{A} con lo stesso autovalore per ogni c non-nullo.

Cercare un autovettore equivale a cercare di risolvere il sistema omogeneo

$$(1.57) \quad (\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0}.$$

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

$$(1.58) \quad \det(\mathbf{A} - \lambda \mathbf{I}) = 0.$$

L'espressione sopra è un *polinomio*, detto *polinomio caratteristico*, ed è di grado n : come si vede *non* coinvolge \mathbf{x} ma fornisce una equazione per i soli autovalori, fornendo al più n radici. La conseguenza delle considerazioni esposte finora è data dalla seguente

PROPOSIZIONE 4.2. *Sia \mathbb{V} uno spazio vettoriale di dimensione n e F un suo endomorfismo. Allora $\lambda \in \mathbb{K}$ è autovalore di F se e solo se λ è radice del polinomio caratteristico. Di conseguenza F ha al più n autovalori distinti.*

La condizione $\det(\mathbf{A} - \lambda_i \mathbf{I}) = 0$, verificata per ogni autovalore λ_i , fa sì che l'equazione $\mathbf{A}\mathbf{x} = \lambda_i \mathbf{x}$ abbia una soluzione non unica, come sappiamo: lo spazio delle soluzioni ha quindi una certa dimensione come specificato dal teorema di Kronecker–Rouché–Capelli (maggiore o uguale ad uno). In particolare, sia $\mathbb{V}_i(F)$ lo spazio di tutti gli autovettori relativi all'autovalore λ_i . Questo è uno spazio vettoriale: infatti, se \mathbf{v}_1 e \mathbf{v}_2 sono elementi di $\mathbb{V}_i(F)$, allora $c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 \in \mathbb{V}_i(F)$. La dimensione di questo spazio è detta *molteplicità geometrica*. Inoltre

PROPOSIZIONE 4.3. *Autovettori corrispondenti ad autovalori distinti sono linearmente indipendenti.*

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

Supponiamo quindi di aver trovato tutti gli zeri del polinomio caratteristico in \mathbb{K} , siano per esempio $\lambda_1, \dots, \lambda_k$ con $k \leq n$. La condizione di *diagonalizzabilità* di F è data dal seguente teorema.

TEOREMA 4.4 (Diagonalizzabilità). *Sia $\Lambda = \{\lambda_i\}_{i=1}^k$ lo spettro di F , costituito da $k \leq n$ autovalori distinti. Allora*

$$(1.59) \quad \sum_{i=1}^k \dim[\mathbb{V}_i(F)] \leq n.$$

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

Supponendo di riuscire a trovare n radici del polinomio in \mathbb{K} , siano esse λ_i per $i = 1, \dots, n$, allora gli elementi della base si possono trovare cercando tra le soluzioni non banali di $\mathbf{A}\mathbf{x} = \lambda_i \mathbf{x}$ per ciascuno di questi valori di \mathbf{x} .

Consideriamo la matrice

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

e cerchiamone autovettori e autovalori. Anzitutto calcoliamo il polinomio caratteristico come

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

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

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

Naturalmente questo sistema non ha un'unica soluzione. Otteniamo quindi che in generale la soluzione è nella forma $(x_1, x_2) = t(1, -1)$ per $t \in \mathbb{R}$. Usando l'autovalore λ_2 invece otteniamo che gli autovettori corrispondenti hanno la forma $(x_1, x_2) = t(1, 1)$.

Consideriamo ora la matrice

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

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

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

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

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

La seconda equazione fornisce l'identità (non informativa) $x_2 = x_2$ mentre la prima ci dà $x_1 + 2x_2 = x_1 \Rightarrow x_2 = 0$. Non abbiamo quindi alcuna informazione su x_1 e la nostra soluzione è nella forma $(x_1, x_2) = t(1, 0)$ con $t \in \mathbb{R}$ qualsivoglia. La dimensione di questo spazio è 1 e corrisponde all'unico autovalore che abbiamo trovato, quindi la matrice *non* è diagonalizzabile alla luce del Teorema 4.4.