APPUNTI DEL CORSO DI ALGEBRA LINEARE
per il corso di Laurea in Matematica [1]

L'algebra lineare svolge un ruolo cruciale in tutti i campi della matematica e, più in generale, delle discipline scientifiche. Sia nelle scienze pure che applicate molti problemi si traducono in problemi di algebra lineare.
Nella prima parte del corso ci occuperemo dello studio dei sistemi lineari. Tale studio ci condurrà in modo naturale allo studio più approfondito delle matrici tramite la teoria degli spazi vettoriali.

Vedremo che risolvere un sistema lineare significa determinare le soluzioni comuni a una o più equazioni di primo grado. Per prima cosa osserviamo che quando si richiede di risolvere un'equazione è indispensabile conoscere l'insieme dove si ricercano le soluzioni. In molti casi risolvere un'equazione significa individuare un sottoinsieme di $\mathbf{R}$ (insieme dei numeri reali) o di $\mathbf{R^n}$.

Per avere un'idea del tipo di situazioni che si possono incontrare, cominciamo con l'esaminare un caso semplicissimo.

Consideriamo il caso di un'equazione di primo grado in una sola incognita con coefficienti a e $b \in \mathbf{R}$:

$$ax = b$$

Sappiamo che se $a \neq 0$, l'equazione ammette un'unica soluzione reale (potrebbe invece non avere soluzioni intere!) $x = b/a$; se $a = 0$, ma $b \neq 0$, l'equazione non ha soluzioni; se $a = 0$ e $b = 0$, l'equazione ha infinite soluzioni, in quanto qualunque $x \in \mathbf{R}$ soddisfa l'equazione.
Dunque in questo caso le soluzioni o non esistono o sono infinite oppure ce n'è una sola; in nessun caso sono in numero finito maggiore di uno. Ci chiederemo se questo sarà vero per ogni sistema lineare indipendentemente dal numero delle equazioni e dal numero delle incognite. La risposta dipenderà dall'insieme nel quale ricercheremo le soluzioni.

Osserviamo che determinare le soluzioni reali dell'equazione $ax = b$ significa determinare l'insieme $f^{-1}(\{b\})$ dove $f : \mathbf{R} \to \mathbf{R}$ è l'applicazione definita da $f(x) = ax$.

---

[1] Con tali appunti si vuole fornire un sussidio didattico per lo studente. Essi raccolgono gli argomenti principali affrontati nel corso, spesso senza dimostrazione. Per una trattazione completa consultare i testi consigliati.

Consideriamo nel seguente esempio il caso di un'equazione di primo grado in due incognite.

Esempio. Determinare le soluzioni reali dell'equazione

$$2x - 2y = 1$$

o equivalentemente studiare l'insieme $f^{-1}(\{1\})$ dove $f : \mathbf{R^2} \to \mathbf{R}$ è l'applicazione definita da $f(x, y) = 2x - 2y$.
Geometricamente si richiede di individuare i punti della retta del piano di equazione $2x - 2y = 1$. Le soluzioni sono quindi infinite e possono anche essere descritte come l'insieme

$$\{(x, x - \frac{1}{2}) \in \mathbf{R^2} : \mathbf{x} \in \mathbf{R}\}.$$

Osserviamo invece che l'equazione $2x - 2y = 1$ non ha soluzioni intere (basta osservare che 1 non è pari).

Analogamente un'equazione del tipo $ax + by = c$ con $a, b, c \in \mathbf{R}$ ha sempre infinite soluzioni reali (eccetto il caso $a = b = 0$ e $c \neq 0$) ed è possibile determinare condizioni su $a, b, c \in \mathbf{Z}$ affinché l'equazione ammetta soluzioni intere (una risposta completa è data usando l'algoritmo euclideo).

Vediamo di esaminare altri esempi già noti allo studente al fine di evidenziare alcuni fatti che affronteremo nel seguito.

Esempio. Determinare le soluzioni reali dei seguenti sistemi di equazioni:

$$\text{a)} \begin{cases} 2x + y = 1 \\ x - y = 2 \end{cases} \quad \text{b)} \begin{cases} x + 2y = -1 \\ x - y = 2 \end{cases} \quad \text{c)} \begin{cases} x - y = 1 \\ 2x - 2y = 2 \end{cases} \quad \text{d)} \begin{cases} x - y = 1 \\ 2x - 2y = 1 \end{cases}$$

Osserviamo che il problema di determinare le soluzioni dei precedenti sistemi si può tradurre in un problema di geometria piana in quanto le equazioni in questione rappresentano rette del piano (x, y). Graficamente possiamo osservare che a) e b) ammettono una unica soluzione rappresentata dal punto di intersezione delle due rette non parellele, d) non ammette soluzioni in quanto il sistema rappresenta l'intersezione di due rette parallele non coincidenti, c) ammette infinite soluzioni essendo le rette coincidenti. In particolare il sistema a) ha come unica soluzione $(1, -1)$ ed è equivalente a b) (hanno le stesse soluzioni) in quanto la prima equazione in b) è ottenuta sottraendo le due equazioni del sistema a). Come prima detto, c) ammette infinite soluzioni e precisamente tutti i punti della retta $x - y = 1$, ossia $\{(x, x - 1) \in \mathbf{R^2} : \mathbf{x} \in \mathbf{R}\}$.

Chiaramente l'interpretazione geometrica è più complicata in presenza di più incognite e più equazioni. Dapprima affronteremo il problema algebricamente e a tal fine risulterà utile scrivere la "tabella" (nel seguito detta matrice) dei coefficienti delle incognite del sistema e dei termini noti. Riferendoci all'esempio precedente questo significa considerare

$$\text{a)} \begin{pmatrix} 2 & 1 & |1 \\ 1 & -1 & |2 \end{pmatrix} \quad \text{b)} \begin{pmatrix} 1 & 2 & |-1 \\ 1 & -1 & |\ 2 \end{pmatrix} \quad \text{c)} \begin{pmatrix} 1 & -1 & |1 \\ 2 & -2 & |2 \end{pmatrix} \quad \text{d)} \begin{pmatrix} 1 & -1 & |1 \\ 2 & -2 & |1 \end{pmatrix}$$

Osserviamo che in a) e b), sia nella tabella dei coefficienti delle incognite, che in quella completa con i termini noti, le righe non sono una multipla dell'altra (altro modo di dire che le rette non sono parallele). In c) le righe della matrice sono una multipla dell'altra (le rette sono coincidenti), in d) le righe della matrice completa non sono proporzionali, mentre nella tabella dei coefficienti continua a sussistere la proporzionalità (le rette sono parallele non coincidenti).

Vedremo che tale decodifica troverà esauriente spiegazione nella teoria che svolgeremo. A questo scopo sarà utile introdurre il calcolo matriciale.

1. Matrici e operazioni

Nel seguito quando scriveremo k intenderemo l'insieme dei numeri razionali o dei numeri reali o dei complessi. Dato che le proprietà algebriche che useremo in questi insiemi numerici saranno solo quelle che riguardano la loro struttura di campo, la teoria che svolgeremo continuerà a valere più in generale in un campo k.

Siano m, n interi positivi. Una *matrice* A di formato $m \times n$ è una collezione di mn elementi disposti in forma di tabella con m righe e n colonne delimitata da parentesi tonde:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

Per esempio, $\begin{pmatrix} 2 & -1 & 3 \\ 1 & 0 & 4 \end{pmatrix}$ è una matrice 2 x 3.

Gli elementi a_{ij} dove i, j sono indici (interi) con $1 \leq i \leq m$ e $1 \leq j \leq n$ sono chiamati *entrate della matrice*. Gli indici i e j sono chiamati, rispettivamente, indice riga e indice colonna. Così a_{ij} è l'elemento che compare nella $i-$esima riga e $j-$esima colonna. Nell'esempio precedente $a_{12} = -1, a_{23} = 4$.

L'insieme di tutte le matrici $m \times n$ a entrate in k si indica con $M_{mn}(k)$. Denoteremo di solito una matrice con A e scriveremo $A = (a_{ij})$.

• Una matrice con una sola riga, cioè una matrice $1 \times n$, viene detta *matrice riga.* Una matrice con una sola colonna, cioè una matrice $m \times 1$, viene detta *matrice colonna.*

• Indicheremo con

$$R_i = (a_{i1} \dots a_{in})$$

la matrice riga corrispondente alla $i-$esima riga di una matrice. Analogamente indicheremo con

$$C_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \dots \\ a_{mj} \end{pmatrix}$$

la matrice colonna corrispondente alla $j-$esima colonna.

• Se nella matrice A si ha $m \leq n$, allora gli elementi $a_{11}, a_{22}, \dots, a_{mm}$ vengono detti elementi della *diagonale principale*.

• La *trasposta* di una matrice $A \in M_{mn}(k)$ è la matrice ${}^tA = (b_{ij}) \in M_{nm}(k)$ dove $b_{ij} = a_{ji}$. Ossia ${}^tA$ è la matrice ottenuta da A scambiando ogni riga con la corrispondente colonna.

• Una matrice $n \times n$ è detta *quadrata* di ordine n.

Sia $A = (a_{ij})$ una matrice quadrata.

• A è *triangolare superiore* se $a_{ij} = 0$ per ogni $i > j$. Ad esempio

$$\begin{pmatrix} 2 & 10 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

• A è *triangolare inferiore* se $a_{ij} = 0$ per ogni $i < j$. Ad esempio

$$\begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 5 & 3 \end{pmatrix}.$$

• A è *diagonale* se A è sia triangolare superiore che inferiore, ossia se $a_{ij} = 0$ per ogni $i \neq j$. Ad esempio

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -3 \end{pmatrix}.$$

• A si dice *simmetrica* se $A = {}^tA$. In particolare le matrici diagonali sono matrici simmetriche.

• Si dice *matrice identica* di ordine n una matrice diagonale tale $a_{ii} = 1$ per ogni $1 \leq i \leq n$. Si denota

$$I_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & & & \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

e se non ci sarà ambiguità scriveremo I invece di I_n.

Definiamo ora le più importanti operazioni tra matrici:

1.1. Somma di matrici

Date $A = (a_{ij})$ e $B = (b_{ij}) \in M_{mn}(k)$ si definisce ancora in $M_{mn}(k)$ una matrice detta *matrice somma* e denotata con $A + B$ nel seguente modo:

$$A + B := \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \dots & & & \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{pmatrix}.$$

La somma di matrici gode delle seguenti proprietà:

1. Associatività. Se $A, B, C \in M_{mn}(k)$, allora $(A + B) + C = A + (B + C)$.
2. Commutatività. Se $A, B \in M_{mn}(k)$, allora $A + B = B + A$.
3. Esistenza dell'elemento neutro. Sia $0 \in M_{mn}(k)$ la matrice con tutte le entrate nulle. Allora $A + 0 = A$ per ogni $A \in M_{mn}(k)$.
4. Esistenza dell'opposto. Data $A \in M_{mn}(k)$ denotiamo con $-A$ la matrice $(-a_{ij})$, allora $A + (-A) = 0$.

Si prova che $M_{mn}(k)$ con l'operazione di somma è un gruppo commutativo.

1.2. Prodotto di una matrice per uno scalare

Dati $A = (a_{ij}) \in M_{mn}(k)$ e uno scalare $\lambda \in k$, si definisce una matrice $\lambda A \in M_{mn}(k)$ nel modo seguente:

$$\lambda A := \begin{pmatrix} \lambda a_{11} & \lambda a_{12} & \dots & \lambda a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \dots & \lambda a_{2n} \\ \dots & & & \\ \lambda a_{m1} & \lambda a_{m2} & \dots & \lambda a_{mn} \end{pmatrix}$$

Ricordiamo alcune proprietà di tale operazione "esterna":

1. Associatività. Se $A \in M_{mn}(k)$ e $\lambda, \mu \in k$, allora $(\lambda\mu)A = \lambda(\mu A)$.
2. Distributivitá. Se $A, B \in M_{mn}(k)$ e $\lambda \in k$, allora $\lambda(A+B) = \lambda A + \lambda B$. Inoltre se $\mu \in k$, allora $(\lambda + \mu)A = \lambda A + \mu A$.
3. Legge di annullamento del prodotto. Se $\lambda A = 0$, allora $A = 0$ oppure $\lambda = 0$.

1.3. Prodotto di matrici

Il prodotto tra matrici è più complesso e a prima vista poco naturale coinvolgendo matrici non sempre dello stesso formato. Tale prodotto è anche detto "righe per colonne" in quanto è definito iterativamente a partire dal caso particolare del prodotto di una matrice riga per una matrice colonna. In tal caso il prodotto è definito se le matrici hanno lo stesso numero di entrate. Si ha

$$(a_{11} \dots a_{1n}) \begin{pmatrix} b_{11} \\ b_{21} \\ \dots \\ b_{n1} \end{pmatrix} := (a_{11}b_{11} + a_{12}b_{21} + \dots + a_{1n}b_{n1})$$

Una possibile applicazione si ha se consideriamo un dolce contenente n ingredienti. Denotiamo con a_{1i} i grammi dell'ingrediente $i-$esimo e con b_{j1} il costo al grammo dell'ingrediente $j-$esimo. Allora il prodotto AB fornisce il costo del dolce. Si dovrà ricorrere al caso di matrici con più righe o colonne se si considerassero più oggetti, in questo caso più dolci, o si dovesse considerare la variazione del costo in diversi anni.

Se il *numero delle colonne* di una matrice A coincide con il *numero delle righe* di B possiamo definire il prodotto di A e B iterando il caso precedente.

Date le matrici $A = (a_{ij}) \in M_{m\mathbf{n}}(k)$ e $B = (b_{ij}) \in M_{\mathbf{n}p}(k)$, si definisce la matrice prodotto $AB \in M_{mp}(k)$ nel seguente modo:

$$AB := (c_{rs})$$

dove

$$c_{rs} = a_{r1}b_{1s} + a_{r2}b_{2s} + \dots + a_{rn}b_{ns} = \sum_{k=1}^{n} a_{rk}b_{ks}$$

è ottenuto moltiplicando la r-esima riga di A per la s-esima colonna di B.

Ricordiamo ancora che **per poter moltiplicare due matrici A e B occorre che il numero delle colonne di A coincida con il numero delle righe di B.**

Esercizio 1.3.1. Calcolare AB dove $A = \begin{pmatrix} 1 & 0 \\ 2 & -1 \\ -2 & 0 \end{pmatrix}$ e $B = \begin{pmatrix} 1 & 2 & 0 & 3 \\ -1 & 0 & -4 & -1 \end{pmatrix}$

Il prodotto di matrici gode delle seguenti proprietà:

1. Associatività. Siano $A \in M_{mn}(k),\ B \in M_{np}(k),\ C \in M_{pq}(k)$ allora $(AB)C = A(BC)$.
2. Distributività. Siano $A \in M_{mn}(k)$ e $B, C \in M_{np}(k)$, si ha $A(B + C) = AB + AC$.
Inoltre se $A, B \in M_{mn}(k)$ e $C \in M_{np}(k)$, allora $(A + B)C = AC + BC$.

Osservazione 1.3.2. E' importante osservare che **la commutatività** (ovviamente quando ha senso parlare sia di AB che di BA) **e la legge di annullamento del prodotto in generale non valgono**.

Ad esempio siano $A = \begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix}$ e $B = \begin{pmatrix} 1 & -2 \\ -1 & 2 \end{pmatrix}$. Allora $AB = \begin{pmatrix} -2 & 4 \\ -1 & 2 \end{pmatrix}$, mentre $BA = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. In questo caso $AB \neq BA$. Inoltre $BA = 0$ con A e B non nulle. Ci sono anche matrici $A \neq 0$ tali che $A^2 = 0$. Ad esempio $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$.

Diremo che una matrice A è *nilpotente* se esiste un intero positivo n tale che $A^n = 0$.

Questi esempi fanno capire che il prodotto righe per colonne che abbiamo definito non gode delle buone proprietà alle quali eravamo abituati con gli insiemi numerici. In particolare non vale la legge di cancellazione del prodotto, ossia data una matrice A non nulla tale $AB = AC$ non si ha necessariamente $B = C$.

Possiamo verificare che:
i) ${}^t(A + B) = {}^t A + {}^t B$
ii) ${}^t(\lambda A) = \lambda\, {}^t A$
iii) ${}^t(AB) = {}^t B\, {}^t A$
iv) ${}^t({}^t A) = A$

Osserviamo che l'operazione di prodotto appena definita permette di presentare un sistema lineare

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{cases}$$

nella forma più familiare

$$AX = B$$

dove $A = (a_{ij}) \in M_{mn}(k)$, $X = \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{pmatrix} \in M_{n1}(k)$ e $B = \begin{pmatrix} b_1 \\ b_2 \\ \ldots \\ b_m \end{pmatrix}$.

Moltiplicando infatti le singole righe della matrice A per la matrice colonna X, si ha

$$AX = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{pmatrix}$$

Diremo che A è la *matrice dei coefficienti*, B è la *matrice dei termini noti* e

$$M = (A|B) = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{pmatrix}$$

è la *matrice completa.*

Ad esempio il sistema $\begin{cases} 2x + y = 1 \\ x - y = 2 \end{cases}$ può essere riscritto come $AX = B$ dove

$$A = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}, \quad X = \begin{pmatrix} x \\ y \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

All'inizio di queste note abbiamo trattato il semplice caso dell'equazione:

$$ax = b$$

e, in quel caso, l'esistenza delle soluzioni si è giocata sull'invertibiltà di $a \in \mathbb{R}$. Nella sezione che seguirà vediamo di affrontare il problema nel caso $A \in M_n(k)$.

1.4. Matrici invertibili

Consideriamo la matrice identica $I_n \in M_n(k)$ e osserviamo che per ogni matrice quadrata $A \in M_n(k)$, sono verificate le seguenti uguaglianze

$$I_n A = A I_n = A.$$

Osserviamo quindi che in $M_n(k)$ sono definite le operazioni di somma e prodotto e la matrice identica è l'elemento neutro rispetto al prodotto.

Vedrete nel corso di algebra che $M_n(k)$ con le operazioni introdotte è un anello unitario non commutativo.

Definizione 1.4.1. Una matrice quadrata A si dice **invertibile** quando esiste una matrice B tale che

$$AB = BA = I.$$

Se A è invertibile, B si dice inversa di A ed è unica. Inoltre se A è invertibile, indicheremo la matrice inversa con A^{-1}.

Osserviamo inoltre che se A^{-1} è l'inversa di A, allora A^{-1} è invertibile e $(A^{-1})^{-1} = A$.

Esercizio 1.4.2. Verificare che l'inversa di $A = \begin{pmatrix} 3 & -1 \\ -5 & 2 \end{pmatrix}$ è $\begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix}$.

Osserviamo che la matrice nulla non è invertibile, ma ci sono anche matrici non nulle e non invertibili, ad esempio $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Più in generale le matrici nilpotenti non sono invertibili (provare questa affermazione).

E' naturale chiedere se, data una matrice quadrata A, può esistere B tale che $AB = I$ (B è detta inversa destra), ma $BA \neq I$. La domanda analoga si può porre per l'inversa sinistra (B take che $BA = I$). Proveremo nel seguito che A è invertibile se esiste B tale che $AB = I$ **oppure** $BA = I$. Siccome in generale non vale la legge commutativa, tale proprietà non è affatto ovvia.

Per il momento possiamo provare facilmente il seguente fatto:

Proposizione 1.4.3. *Se A ha un'inversa destra B e un'inversa sinistra C, allora $B = C$.*

Dalla precedente proposizione segue in particolare che l'inversa di una matrice, se esiste, è unica.

Proposizione 1.4.4. *Siano A e B matrici invertibili, allora AB è invertibile e l'inversa è $B^{-1}A^{-1}$.*

Il sottoinsieme di $M_n(k)$ delle matrici invertibili con l'operazione di prodotto (abbiamo visto che il prodotto di matrici invertibili è invertibile) è un gruppo, detto *gruppo lineare* e denotato con $Gl_n(k)$. Nel seguito caratterizzeremo gli elementi di $Gl_n(k)$.

Le matrici invertibili più semplici sono le **matrici elementari** che, come vedremo, troveranno una interessante applicazione nella teoria dei sistemi lineari.

Fissato un intero positivo n, oltre alla matrice identica, esistono tre tipi di matrici elementari:

1) Siano i, j interi tali $1 \leq i, j \leq n, i \neq j$. Definiamo $\mathbf{E_{ij}}$ la matrice ottenuta da I scambiando R_i con R_j.
Ad esempio se $n = 4$,

$$E_{23} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Osserviamo inoltre che $E_{ij}^2 = I$, quindi $E_{ij}^{-1} = E_{ij}$.

2) Siano i un intero, $1 \leq i \leq n$ e $\lambda \in k^*$. Definiamo $\mathbf{E_i}(\lambda)$ la matrice ottenuta da I sostituendo R_i con λR_i.
Ad esempio se $n = 4$,

$$E_2(-3) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Osserviamo che $E_i(\lambda)^{-1} = E_i(1/\lambda)$.

3) Siano i, j interi tali $1 \leq i, j \leq n, i \neq j$ e sia $\lambda \in k$. Definiamo $\mathbf{E_{ij}}(\lambda)$ la matrice ottenuta da da I sostituendo R_i con $R_i + \lambda R_j$ (la matrice identica con $a_{ij} = \lambda$).
Ad esempio se $n = 4$,

$$E_{21}(-3) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Osserviamo che $E_{ij}(\lambda)^{-1} = E_{ij}(-\lambda)$.

In particolare quindi abbiamo visto che ogni matrice elementare è invertibile e l'inversa è ancora una matrice elementare dello stesso tipo.

Data $A \in M_{mn}(k)$, non è difficile verificare che:

1) $E_{ij}A$ è la matrice ottenuta da A scambiando R_i con R_j.

2) $E_i(\lambda)A$ è la matrice ottenuta da A moltiplicando ogni elemento di R_i per λ.

3) $E_{ij}(\lambda)A$ è la matrice ottenuta da A sostituendo a R_i la riga $R_i + \lambda R_j$.

Con il seguente esempio vediamo come operando sulle righe di una matrice A o, equivalentemente, moltiplicando A per matrici elementari, è possibile ottenere matrici con più entrate nulle.

Esempio 1.4.5. Data la matrice

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

moltiplicando opportunamente per matrici elementari, otteniamo una matrice triangolare superiore.

$$(R_2 \to R_2 + 3R_1) \to \begin{pmatrix} 1 & 2 & 3 & -1 \\ 0 & 7 & 9 & -3 \\ 0 & -1 & 1 & 2 \\ 1 & 2 & -1 & 1 \end{pmatrix} (R_4 \to R_4 - R_1) \begin{pmatrix} 1 & 2 & 3 & -1 \\ 0 & 7 & 9 & -3 \\ 0 & -1 & 1 & 2 \\ 0 & 0 & -4 & 2 \end{pmatrix}$$

$$(R_3 \to 7R_3) \to \begin{pmatrix} 1 & 2 & 3 & -1 \\ 0 & 7 & 9 & -3 \\ 0 & -7 & 7 & 14 \\ 0 & 0 & -4 & 2 \end{pmatrix} \quad (R_3 \to R_3 + R_2) \to \begin{pmatrix} 1 & 2 & 3 & -1 \\ 0 & 7 & 9 & -3 \\ 0 & 0 & 16 & 11 \\ 0 & 0 & -4 & 2 \end{pmatrix}$$

$$(R_4 \to 4R_4) \to \begin{pmatrix} 1 & 2 & 3 & -1 \\ 0 & 7 & 9 & -3 \\ 0 & 0 & 16 & 11 \\ 0 & 0 & -16 & 8 \end{pmatrix} (R_4 \to R_4 + R_3) \to \begin{pmatrix} 1 & 2 & 3 & -1 \\ 0 & 7 & 9 & -3 \\ 0 & 0 & 16 & 11 \\ 0 & 0 & 0 & 19 \end{pmatrix} = A'$$

Osserviamo che $A' = E_{43}(1)E_4(4)E_{32}(1)E_3(7)E_{41}(-1)E_{21}(3)A$.

Il procedimento dell'esempio precedente è chiamato *riduzione per righe* oppure *eliminazione di Gauss.*

2. Matrici ridotte e eliminazione di Gauss per la risoluzione di sistemi lineari

E' facile vedere che con un procedimento algoritmico basato su successive moltiplicazioni per matrici elementari, ogni matrice non nulla può essere ridotta ad una matrice “a scalini” della forma

$$\begin{pmatrix}
0 & \dots & 0 & a_1 & * & \dots & & & * & \dots & * & & * & \dots & * & & \dots \\
0 & \dots & 0 & 0 & & \dots & 0 & a_2 & * & \dots & * & & * & \dots & * & & \dots \\
0 & \dots & 0 & & & \dots & & 0 & \dots & \dots & 0 & a_3 & * & \dots & * & & \dots \\
0 & \dots & & & & \dots & & & \dots & \dots & & & & \dots & 0 & a_4 & \dots \\
0 & \dots & \dots & & & \dots & & & \dots & \dots & \dots & & & \dots & & & \dots
\end{pmatrix}$$

dove gli a_i sono non nulli e $*$ denota un elemento di k (non necessariamente non nullo).

Più precisamente una matrice della forma precedente è detta *matrice ridotta per righe* o *matrice a scala* ed è caratterizzata dalla seguente proprietà:

Il primo elemento non nullo della $(i+1)-$esima riga si trova alla destra rispetto al primo elemento non nullo della $i-$esima riga.

Cerchiamo di capire come l'eliminazione di Gauss può essere applicata allo studio dei sistemi lineari. Ad esempio consideriamo il sistema

$$\begin{cases} x_1 + 2x_2 + 3x_3 = -1 \\ -3x_1 + x_2 = 0 \\ -x_2 + x_3 = 2 \\ \ \ x_1 + 2x_2 - x_3 = 1 \end{cases}$$

Osserviamo che la matrice completa del sistema è la matrice dell'esempio 1.4.5.. La matrice A' ottenuta corrisponde alla matrice completa del seguente sistema:

$$\begin{cases} x_1 + 2x_2 + 3x_3 = -1 \\ \qquad +7x_2 + 9x_3 = -3 \\ \qquad\qquad\quad 16x_3 = 11 \\ \qquad\qquad\qquad 0 = 19 \end{cases}$$

I due sistemi hanno una parentela? Osserviamo che le operazioni elementari operate sulla matrice A per ottenere A' corrispondono ad “operazioni lecite” sulle equazioni del sistema. Questo concetto verrà presto precisato per provare che i due sistemi hanno le stesse soluzioni. Guardando la nuova riformulazione del sistema, è facile vedere che non esistono soluzioni. Vediamo di formalizzare il procedimento.

Dati $A \in M_{mn}(k)$ e $B \in M_{m1}(k)$, si consideri il sistema lineare

$$AX = B$$

Siano $M = (A|B)$ la matrice completa $m \times (n+1)$ associata al sistema e E una matrice elementare o prodotto di matrici elementari $m \times m$.
Poniamo $M' = EM,\ A' = EA,\ B' = EB$.
Osserviamo che

$$M' = EM = (EA|EB) = (A'|B').$$

Proviamo che

Proposizione 2.1. *Le soluzioni del sistema $A'X = B'$ sono le stesse del sistema $AX = B$, ossia i sistemi sono equivalenti.*

Esercizio 2.2. Risolvere il seguente sistema lineare:

$$\begin{cases} x_1 + 2x_3 + x_4 = 5 \\ x_1 + x_2 + 5x_3 + 2x_4 = 7 \\ x_1 + 2x_2 + 8x_3 + 4x_4 = 12 \end{cases}$$

Si consiglia di determinare le soluzioni del sistema dopo aver ridotto per righe la matrice completa, usando la Proposizione 2.1.

Operando ancora con matrici elementari, possiamo verificare che ogni matrice non nulla può essere ridotta per righe ad una matrice della forma:

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

dove $*$ denota un numero qualsiasi e lo spazio vuoto è costituito da zeri.

La matrice A così ottenuta si dice *matrice totalmente ridotta per righe.* Una matrice totalmente ridotta per righe si può definire come segue:

a) *Il primo elemento non nullo di ogni riga è 1. Tale elemento è chiamato pivot.*

b) *Il pivot della $(i+1)$-esima riga si trova alla destra rispetto al pivot della i-esima riga.*

c) *Gli elementi al di sopra di un pivot sono nulli.*

Si può dimostrare che la matrice totalmente ridotta per righe ottenuta a partire da una matrice assegnata A, è unica, ossia non dipende dalla particolare sequenza di operazioni eseguita.

Esercizio 2.3. Data

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

determinare la matrice totalmente ridotta per righe che si ottiene da A con operazioni elementari.

Diremo che una matrice è *ridotta per colonne* se ${}^tA$ è ridotta per righe.
Analogamente una matrice è *totalmente ridotta per colonne* se ${}^tA$ è totalmente ridotta per righe.

Come per la riduzione per righe, data una qualunque matrice A, possiamo ottenere la matrice ridotta (totalmente ridotta) per colonne tramite operazioni elementari sulle colonne o, equivalentemente, moltiplicando a destra la matrice A per matrici elementari.

Proviamo la seguente proposizione.

Proposizione Data una matrice $A \in M_n(k)$, allora
a) esiste una matrice $B \in Gl_n(K)$ tale che $BA = T$ con T matrice triangolare superiore.
b) esiste una matrice $C \in Gl_n(K)$ tale che $AC = T$ con T matrice triangolare inferiore.
c) esistono $B, C \in Gl_n(K)$ tali che $B\ A\ C = \Delta$ con Δ matrice diagonale.

Useremo la riduzione per righe per caratterizzare i sistemi lineari che ammettono soluzione anche se poi nel seguito ritorneremo sul problema affrontandolo da un altro punto di vista.

Proposizione 2.4. *Sia $M' = (A'|B')$ una matrice totalmente ridotta per righe. Allora il sistema di equazioni $A'X = B'$ ammette soluzioni se e soltanto se l'ultima colonna non contiene pivot.*

Chiaramente un sistema omogeneo $AX = 0$ ammette almeno la soluzione banale $X = 0$. Considerando la matrice totalmente ridotta, possiamo facilmente dedurre che se il sistema omogeneo ha m equazioni e n incognite con $m < n$, ammette infinite soluzioni. Basta infatti attribuire un valore arbitrario all'incognita x_i se la colonna $i-$esima non contiene pivot. Nel caso di un sistema lineare non omogeneo, la condizione $m < n$ non assicura neppure l'esistenza di soluzioni.

Esercizio 2.5. Determinare le soluzioni reali del seguente sistema lineare:

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

Utilizzeremo ora la riduzione per righe per caratterizzare le matrici quadrate invertibili.

Proposizione 2.6. *Sia A una matrice quadrata. Le seguenti condizioni sono equivalenti:*

a) A può essere ridotta alla matrice identica tramite una successione di operazioni elementari sulle righe.

b) A è prodotto di matrici elementari.

c) A è invertibile.

d) Il sistema lineare $AX = 0$ ammette solo la soluzione nulla.

Osservazione 2.7. Nella prova della Proposizione 2.6. abbiamo messo in luce i seguenti fatti:
1. se M è una matrice quadrata totalmente ridotta per righe, allora M è la matrice identica, oppure la sua ultima riga è nulla.
2. La riduzione per righe fornisce un metodo per calcolare l'inversa di una matrice. Possiamo facilmente osservare che se $E_1, \ldots, E_k$ sono matrici elementari tali

$$E_1 \cdots E_k A = I$$

allora A è invertibile e

$$A^{-1} = E_1 \cdots E_k.$$

Sia A una matrice invertibile. Per calcolare la sua inversa A^{-1}, basta quindi effettuare operazioni elementari sulle righe riducendola all'identità. La stessa successione di operazioni, applicata a I, fornisce A^{-1}.

Esercizio 2.8. Determinare l'inversa di

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

Terminiamo questa parte con un risultato che avevamo già citato e che ora può essere provato facilmente con i metodi introdotti.

Proposizione 2.9. *Sia A una matrice quadrata dotata di un'inversa sinistra B ($BA = I$), oppure di un'inversa destra ($AB = I$). Allora A è invertibile e B è la sua inversa.*

3. Determinanti e sistemi lineari.

Siano n un intero positivo e sia $J_n = \{1, 2, \dots, n\}$. Indichiamo con S_n l'insieme delle applicazioni iniettive di J_n in J_n. Osserviamo che essendo J_n un insieme finito, una applicazione $f : J_n \to J_n$ è iniettiva se e solo se è surgettiva, se e solo se è bigettiva. Gli elementi di S_n sono dette **permutazioni.** Una permutazione σ si può rappresentare con una tabella del tipo

$$\sigma = \begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix}.$$

Ad esempio la permutazione $\sigma : J_4 \to J_4$ definita da $\sigma(1) = 2$, $\sigma(2) = 3$, $\sigma(3) = 1$, $\sigma(4) = 4$ si indicherà con

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix}.$$

E' facile vedere che gli elementi di S_n sono $n!$ (ossia $n(n-1)(n-2)\cdots 2$).

Nell'insieme S_n abbiamo una operazione di composizione dovuta al fatto che la composta di due applicazioni bigettive è bigettiva. Tale operazione non è commutativa, ma è associativa; inoltre S_n possiede un elemento neutro rispetto a tale operazione: tale elemento è la permutazione identica. Inoltre ogni permutazione σ è invertibile e $\sigma^{-1} \in S_n$. E' anche chiaro che se $\sigma, \tau \in S_n$ allora

$$(\sigma\tau)^{-1} = \tau^{-1}\sigma^{-1}.$$

Tale operazione rende S_n un gruppo non commutativo che si chiama il **gruppo delle permutazioni** di I_n. Spesso parleremo di moltiplicazione in S_n invece che di composizione.

Un elemento $\sigma \in S_n$ si dirà una **trasposizione** o uno **scambio** o un $2-$**ciclo** se per qualche $i, j \in I_n$ risulta $\sigma(i) = j$, $\sigma(j) = i$ e $\sigma(k) = k, \forall k \neq i, j$. Denoteremo una tale permutazione semplicemente con (i, j). Analogamente diremo che una permutazione σ è un k ciclo se esistono $i_1, \ldots, i_k \in I_n$ tali che

$$\sigma(i_1) = i_2, \sigma(i_2) = i_3, \ldots, \sigma(i_{k-1}) = i_k, \sigma(i_k) = i_1, \sigma(j) = j, \forall j \neq i_1, \ldots, i_k.$$

Una tale permutazione la indicheremo semplicemente con $(i_1, \ldots, i_k)$.

Nell'analisi che stiamo per condurre, possiamo supporre che $\sigma \neq id$, ossia che la permutazione non sia quella identica.

Teorema 3.1. *Ogni permutazione è prodotto di cicli disgiunti.*

Ad esempio se

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 6 & 8 & 4 & 3 & 1 & 5 & 2 & 7 \end{pmatrix}.$$

si ha

$$\sigma = (1, 6, 5)(2, 8, 7)(3, 4).$$

Teorema 3.2. *Ogni ciclo è prodotto di 2-cicli.*

Infatti si ha la formula

$$(i_1, i_2, \ldots, i_d) = (i_1, i_d)(i_1, i_{d-1}) \cdots (i_1, i_2).$$

Si conclude quindi che ogni permutazione è prodotto di 2-cicli.

Una permutazione $\sigma \in S_n$ si dice di **classe pari** (risp. dispari) se si fattorizza nel prodotto di un numero pari (risp. dispari) di 2-cicli. La fattorizzazione di una permutazione nel prodotto di 2-cicli non è unica, ma si prova che se $\sigma = T_1 \cdots T_h$ e $\sigma = T_1' \cdots T_s'$ con T_i e T_i' 2-cicli, allora $h = s$ mod 2, ossia si preserva la parità (risp. disparità).

Se $\sigma \in S_n$ è una permutazione si definisce la **segnatura** di σ ponendo

$$sgn(\sigma) = \prod_{i<j} \frac{\sigma(j) - \sigma(i)}{j - i}.$$

Proposizione 3.3. *Se $\sigma, \tau \in S_n$ si ha:*

$$sgn(\tau\sigma) = sgn(\tau) sgn(\sigma).$$

Essendo $sgn((ij)) = -1$, risulta che per ogni $\sigma \in S_n$

$$sgn(\sigma) = \pm 1,$$

in particolare $sgn(\sigma) = 1$ se σ è di classe pari e $sgn(\sigma) = -1$ se di classe dispari.

Poiché $sgn(id) = 1$, si deduce che

$$sgn(\sigma) = sgn(\sigma^{-1}).$$

Da questa analisi si può provare anche che

$$sgn(\sigma) = (-1)^{\text{numero di inversioni di } \sigma}$$

ove per calcolare il numero di inversioni di σ, si calcola il numero delle coppie (i, j) con $i < j$ e $\sigma(i) > \sigma(j)$.

Sia ora $A = (a_{ij})$ una matrice quadrata ad elementi in k.
Diremo **determinante** di A l'elemento di k cosí definito:

$$det(A) := \sum_{\sigma \in S_n} sgn(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{n\sigma(n)}.$$

Il determinante di A si può indicare anche con $|A|$. Le seguenti proprietà del determinante sono di facile verifica:

1. $det(A) := \sum_{\tau \in S_n} sgn(\tau) a_{\tau(1)1} a_{\tau(2)2} \dots a_{\tau(n)n}$.

2. $det(A) = det({}^t A)$.

3. Se si scambiano in A due righe (o due colonne) il determinante cambia di segno.

4. Se A ha due righe (o due colonne) eguali , $det(A) = 0$.

5. Se

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{i1} + b_{i1} & \dots & a_{in} + b_{in} \\ \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}$$

allora

$$det(A) = det \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{1i} & \dots & a_{in} \\ \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} + det \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ b_{i1} & \dots & b_{in} \\ \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}$$

6. Se una riga (o una colonna) di A è nulla , allora $det(A) = 0$.

7. Se si sostiuisce ad una riga (o una colonna) di A la riga stessa moltiplicata per una costante $\lambda \in k$, allora si ottiene una matrice il cui determinante è $\lambda\ det(A)$.

8. Se due righe (o due colonne) di A sono proporzionali, allora $det(A) = 0$.

9. Se si aggiunge ad una riga (o colonna) di A un'altra riga (o colonna) moltiplicata per $\lambda \in k$, il determinante non cambia.

10. Se si ha:

$$A = \begin{pmatrix} a_{11} & \dots & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \\ 0 & \dots & 0 & a_{nn} \end{pmatrix}$$

allora $det(A) = a_{11} \cdots a_{nn}$.

11. Se si aggiunge ad una riga (o colonna) di A una combinazione lineare delle rimanenti righe (ossia a R_i si sostituisce $R_i + \sum_{j=1}^{n} \lambda_j R_j$ con $\lambda_j \in k$ e $j \neq i$)(ugualmente per le colonne), il determinante non cambia.

12. Si ha che $det(A) = 0$ se e solo se le righe di A sono linearmente dipendenti (ossia esistono $\lambda_1, \dots, \lambda_n \in k$ non tutti nulli tali che $\sum_{j=1}^{n} \lambda_j R_j = (0 \dots\dots 0)$). Vale l'eneunciato analogo con le colonne.

Osservazione 3.4. Da 3., 7. e 9. si verifica facilmente che

$$det(E_{ij}A) = -det(A)$$

$$det(E_i(\lambda)A) = \lambda det(A)$$

$$det(E_{ij}(\lambda)A) = det(A)$$

Poiché ogni matrice quadrata si può trasformare in una matrice triangolare superiore moltiplicando per matrici elementari del tipo E_{ij} e $E_{ij}(\lambda)$, la riduzione di Gauss dà un metodo effettivo per il calcolo del determinante.

Esercizio 3.5. Calcolare

$$det \begin{pmatrix} 0 & 1 & 2 & 4 \\ -1 & 1 & 2 & 0 \\ 3 & 1 & 0 & 4 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$

con il metodo di riduzione di Gauss.

Dall'osservazione precedente e dalla caratterizzazione data delle matrici invertibili, è facile dedurre il seguente importante risultato.

Teorema 3.6. *Una matrice quadrata A è invertibile se e solo se $det(A) \neq 0$.*

Infatti $det(A) \neq 0$ se e solo se $det(A') \neq 0$ dove A' denota la matrice totalmente ridotta ottenuta da A. Segue che $det(A') \neq 0$ se e solo se $A' = I$ se e solo se A è invertibile.

Osserviamo ora che

$$det(E_{ij}) = -1, \quad det(E_i(\lambda)) = \lambda, \quad det(E_{ij}(\lambda)) = 1.$$

Sia E una matrice elementare o prodotto di matrici elementari, da quanto detto sopra e dalle proprietà del determinante segue che

$$det(EA) = det(E)det(A).$$

Usando questo fatto si prova

Teorema di Binet. Se A e B sono due matrici quadrate $n \times n$, si ha

$$det(AB) = det(A)det(B).$$

Se A è una matrice quadrata ad entrate in k, diciamo **complemento algebrico** dell'elemento a_{ij}, l'elemento A_{ij} definito come $(-1)^{i+j}$ per il determinante della matrice che si ottiene da A eliminando la riga $i-ma$ e la colonna $j-ma$.

Primo Teorema di Laplace. Per ogni matrice quadrata $A = (a_{ij})$ e per ogni $r = 1, \dots, n$ si ha

$$det(A) = \sum_{j=1}^{n} a_{rj} A_{rj}.$$

Questo modo di esprimere il determinante di A si chiama **lo sviluppo del determinante** secondo la riga $r-ma$. Analogamente abbiamo una formula per lo sviluppo del determinante secondo la colonna $s-ma$. Precisamente si ha

$$det(A) = \sum_{j=1}^{n} a_{js} A_{js}.$$

Se A è una matrice quadrata diciamo **aggiunta** di A e la denotiamo con $A^\star$ la matrice

$$A^\star =^t ((A_{ij}))$$

e cioè la matrice che al posto ij ha il complemento algebrico dell'elemento di posto ji.

Secondo Teorema di Laplace. Per ogni matrice quadrata A si ha se r e s sono due interi distinti

$$0 = \sum_{j=1}^{n} a_{js} A_{jr}.$$

Usando questo teorema è facile provare che

$$AA^{\star} = det(A)I.$$

Se A è invertibile, allora $det(A) \neq 0$ e si ha quindi

$$A^{-1} = \frac{1}{det(A)} A^{\star}$$

Se A è una matrice $m \times n$ e t è un intero $1 \leq t \leq \min\{n, m\}$, diciamo **minore** di ordine t di A il determinante di una qualunque sottomatrice quadrata di A che si ottiene fissando t righe e t colonne. In particolare i minori di ordine 1×1 sono gli elementi di A e se A è una matrice quadrata $n \times n$ c'è un solo minore di ordine n di A ed è il $det(A)$.

Sia $A \in M_{mn}(k)$, diciamo **caratteristica** di A e la denotiamo con $\rho(A)$, l'ordine massimo di un minore non nullo di A.

Si ha quindi

$$\rho(A) \leq min\{m, n\}.$$

In particolare se $A \in M_n(k)$ si ha $\rho(A) = n$ se e solo se $det(A) \neq 0$.

Osserviamo che *se A è una matrice ridotta per righe, $\rho(A)$ coincide con il numero di righe non nulle di A.*

Consideriamo infatti la sottomatrice $p \times p$ costituita dalle p righe non nulle e dalle p colonne contenenti i pivot: abbiamo una matrice triangolare superiore con gli elementi sulla diagonale non nulli, per cui il suo determinante è non nullo. Inoltre, dato che A ha solamente p righe non nulle, tutti i minori di ordine $p + 1$ sono nulli.

Visto che ogni matrice si può trasformare con operazioni elementari sulle righe (risp. colonne) in una matrice ridotta per righe (risp. colonne), è naturale chiedersi se la caratteristica di una matrice può variare per operazioni elementari.

Ricordando il comportamento del determinante rispetto ad operazioni elementari, è facile provare il seguente fatto

Proposizione 3.7. Sia A una matrice $m \times n$ e sia E una matrice elementare $m \times m$, allora

$$\rho(A) = \rho(EA)$$

Analogamente se E una matrice elementare $n \times n$, allora

$$\rho(A) = \rho(AE)$$

Possiamo concludere quindi che la caratteristica di una matrice coincide con il numero di righe non nulle di una (qualunque) matrice ridotta per righe ottenuta da A oppure con il numero di colonne non nulle di una (qualunque) matrice ridotta per colonne ottenuta da A.

Dai fatti precedenti possiamo dedurre il seguente risultato:

Corollario 3.9. Siano $A \in M_{mn}(k)$, $B \in Gl_m(k)$ e $C \in GL_m(k)$. Allora

$$\rho(A) = \rho(BAC)$$

Diremo che due matrici A_1 e A_2 sono *equivalenti* se esistono B e C invertibili tali

$$A_1 = BA_2C$$

L'equivalenza tra matrici è una relazione di equivalenza in $M_{mn}(k)$. Ogni matrice quadrata è, in particolare, equivalente ad una matrice diagonale.

Si può provare che se $\rho(A) = t$ allora A è equivalente a una matrice che ha sulla diagonale principale t elementi uguali a 1 e zeri altrove. Due matrici sono equivalenti se e solo se hanno la stessa caratteristica.

In particolare A e ${}^tA$ sono matrici equivalenti in quanto hanno la stessa caratteristica.

Osserviamo che per una matrice $m \times n$, le sottomatrici $p \times p$ sono $\binom{n}{p} \cdot \binom{m}{p}$, quindi anche al fine di calcolare la caratteristica di una matrice, il metodo di riduzione sembra ancora una volta una procedura estremamente utile. Vediamo ora un criterio che in molti casi semplificherà ancora il calcolo della caratteristica.

Teorema di Kronecker. *Sia A una matrice $m \times n$. Se esiste un minore di ordine t di A che è non nullo ma che orlato in tutti i modi possibili con l'aggiunta di una riga e una colonna di A è nullo, allora si ha :*

$$\rho(A) = t$$

Esempio 3.10. Si consideri la seguente matrice:

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

Osserviamo che $det \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \neq 0$. Il teorema di Kronecker afferma che per provare che $\rho(A) = 2$, non occorre controllare tutti i 16 minori di ordine 3 di A, ma basta controllare solo i quattro minori di ordine 4 che contengono il minore non nullo di ordine 2 che abbiamo selezionato.

La teoria introdotta ci permette ora di provare ulteriori condizioni di compatibilità di un sistema lineare. Sia dato il sistema lineare di m equazioni ed n incognite:

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

Abbiamo visto che se indichiamo con A la matrice $A = (a_{ij})$, e con

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

il sistema si può riscrivere come una equazione matriciale

$$AX = B.$$

Teorema di Cramer. *Se $m = n$ e A è invertibile allora il sistema dato ha una e una sola soluzione:*

$$X = A^{-1}B = \frac{1}{det(A)}A^{\star}B.$$

Notiamo che $A^{\star}B$ è una matrice $n \times 1$ che al posto i ha come elemento

$$A^{\star}_{i1}b_1 + A^{\star}_{i2}b_2 + \ldots + A^{\star}_{in}b_n = A_{1i}b_1 + A_{2i}b_2 + \ldots + A_{ni}b_n$$

Questo è il determinante della matrice che si ottiene da A sostituendo la colonna $i - ma$ con la colonna dei termini noti. Dunque si ha la seguente formula per la soluzione del sistema :

$$\forall i = 1, \ldots, n \qquad x_i = \frac{det \begin{pmatrix} a_{11} & a_{12} & \ldots & b_1 & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & b_2 & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \ldots & b_n & \ldots & a_{nn} \end{pmatrix}}{det(A)}$$

Dato il sistema $AX = B$ indichiamo con $M = (A|B)$ la **matrice completa** del sistema ossia la matrice che si ottiene da A aggiungendo come ultima colonna la colonna dei termini noti. E' chiaro che M è una matrice $m \times (n+1)$. Si prova il seguente risultato

Teorema di Rouchè-Capelli. *Il sistema $AX = B$ ha soluzione se e solo se*

$$\rho(A) = \rho(M).$$

Ora se il sistema $AX = B$ è risolubile, in accordo con il precedente teorema, sia $t = \rho(A) = \rho(M)$. Allora per determinare le soluzioni si consideri un minore $t \times t$ non nullo di A e vediamo come ridurre il sistema iniziale ad un sistema di Cramer. Si può provare che il sistema dato è equivalente al sistema che si ottiene trascurando le equazioni che non sono coinvolte nelle righe del minore scelto. Quindi si possono portare a termine noto i termini delle equazioni con le incognite che non sono coinvolte nelle colonne del minore scelto, ottenendo così un sistema $t \times t$ la cui matrice dei coefficienti ha per determinante il minore non nullo scelto. Otteniamo dunque un sistema di Cramer che sappiamo risolvere.

In tal modo si esprimeranno t delle incognite in funzione delle rimanenti $n - t$. Si dirà allora che il sistema ha ∞^{n-t} **soluzioni** , nel senso che le soluzioni del sistema si ottengono attribuendo ad arbitrio valori alle incognite "libere" che sono appunto $n - t$.

Se in particolare si deve studiare il sistema omogeneo

$$AX = 0$$

allora chiaramente $\rho(A) = \rho(M)$, questo corrisponde al fatto già visto che un tale sistema ha sempre almeno la soluzione **banale** ossia la soluzione

$$x_1 = 0, x_2 = 0, \ldots, x_n = 0.$$

E' importante osservare che dato un sistema lineare $AX = B$ che ammette soluzione, le soluzioni si ottengono da quelle del sistema lineare omogeneo associato $AX = 0$ sommando una soluzione particolare $\overline{X}$.

4. Spazi Vettoriali.

Un gruppo commutativo (abeliano) è un insieme G con una operazione binaria denotata usualmente con $+$, che gode delle seguenti proprietà:
a) Proprietà associativa.
b) Proprietà commutativa.
c) Esiste un elemento $g \in G$ tale che $a + g = g + a = a$ per ogni $a \in G$.
d) Per ogni elemento $a \in G$ esiste un elemento $b \in G$ tale che $a + b = b + a = g$.

Un elemento come in c) si chiamerà un elemento **neutro.** Un elemento come in d) si dirà un **opposto** di a.

Si hanno le seguenti proprietà:
a) Unicità dell'elemento neutro.
b) Unicità dell'opposto.
c) Legge di cancellazione, ossia

$$a + b = a + c \Rightarrow b = c.$$

Dopo aver verificato quanto sopra indicheremo con 0 l'unico elemento neutro del gruppo abeliano G e con $-g$ l'unico opposto dell'elemento $g \in G$. E' chiaro allora che la scrittura $g - h$ significa $g + (-h)$.

Nel seguito k indicherà un campo, ma per il momento ci limiteremo all'insieme dei numeri razionali, o l'insieme dei numeri reali, o l'insieme dei numeri complessi.

Uno spazio vettoriale su k è un gruppo abeliano V con una operazione esterna che ad ogni elemento (a, v) di $k \times V$ associa un elemento av di V verificante le seguenti proprietà:

a) $a(v_1 + v_2) = av_1 + av_2$, per ogni $a \in V$ e per ogni $v_1, v_2 \in V$.

b) $(a + b)v = av + bv$, per ogni $a, b \in k$ e per ogni $v \in V$.

c) $a(bv) = b(av) = (ab)v$ per ogni $a, b \in k$ e per ogni $v \in V$.

d) $1v = v$ per ogni $v \in V$.

Gli elementi di uno spazio vettoriale V saranno detti **vettori** mentre gli elementi di k saranno detti **scalari** .

k^n con le operazioni così definite è un k-spazio vettoriale:

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n)$$

e per ogni $a \in k$ si ha

$$a(x_1, \dots, x_n) = (ax_1, \dots, ax_n).$$

Si verificano le seguenti proprietà aritmetiche degli spazi vettoriali:

a) $0v = 0$, per ogni $v \in V$.

b) $a0 = 0$ per ogni $a \in k$.

c)$(-1)v = -v$ per ogni $v \in V$.

Un sottoinsieme W dello spazio vettoriale V si dirà un **sottospazio vettoriale** di V se W è uno spazio vettoriale rispetto alla restrizione a W delle operazioni di V.

Si può dedurre facilmente che un sottoinsieme non vuoto W dello spazio vettoriale V è un sottospazio vettoriale di V se e solo se W è chiuso rispetto alle operazioni di V ristrette al sottoinsieme, ossia se per ogni $a \in k$, $w_1, w_2 \in W$,

$$w_1 + w_2 \in W \qquad \text{e} \quad aw_1 \in W.$$

Ciò è equivalente alla condizione

$$\forall a, b \in k, \forall w_1, w_2 \in W \qquad aw_1 + bw_2 \in W.$$

Si verifica facilmente che $\{0_V\}$ e V sono sottospazi vettoriali di V.

Se W e Z sono due sottospazi di V, allora $W \cap Z$ è ancora un sottospazio di V. Lo stesso vale per una famiglia anche non finita di sottospazi di V.

Esempi 4.1. i) Tutti e soli i sottospazi vettoriali di $\mathbb{R}^2$ come $\mathbb{R}-$spazio vettoriale sono: $\{(0,0)\}$, le rette passanti per l'origine e $\mathbb{R}^2$ stesso.
ii) Sia $A \in M_{mn}(k)$, l'insieme delle soluzioni del sistema lineare omogeneo $AX = 0$ è un sottospazio vettoriale di k^n. Viceversa ogni sottospazio vettoriale di k^n è l'insieme delle soluzioni di un sistema lineare omogeneo.

Siano V un $k-$spazio vettoriale e $v \in V$, allora

$$< v >:= \{av \ / \ a \in k\}$$

è un sottospazio vettoriale di V detto *sottospazio generato da* v ed è il più piccolo (rispetto all'inclusione insiemistica) sottospazio di V contenente v.

Più in generale se S è un sottoinsieme (non necessariamente finito) dello spazio vettoriale V, definiamo $< S >$ come l'insieme dei vettori di V che si possono scrivere come somme finite

$$a_1v_1 + a_2v_2 + \ldots + a_nv_n$$

ove n è un intero positivo, $a_i \in k$ e $v_i \in S$.

E' facile vedere che $< S >$ è un sottospazio di V ed è il più piccolo sottospazio di V che contiene S. Diremo che $< S >$ è il **sottospazio generato da S.**

Osserviamo ora che l'unione di due sottospazi di V non sempre è un sottospazio di V. Ad esempio se consideriamo in $\mathbb{R}^2$ i sottospazi $W =< v_1 >$ e $Z =< v_2 >$ dove $v_1 = (1, 2)$ e $v_2 = (2, 2)$, allora $v_1 + v_2$ deve appartenere a un qualunque sottospazio vettoriale contenente $W \cup Z$, ma $v_1 + v_2 \notin W \cup Z$.

Dati i sottospazi W e Z dello spazio vettoriale V, indicheremo con $W + Z$ lo spazio $< W \cup Z >$. Tale notazione è giustificata dal fatto che risulta

$$W + Z = \{w + z | w \in W, z \in Z\}.$$

Tale sottospazio sarà detto la **somma** di W e di Z. Se $W \cap Z = \{0\}$ diremo che la somma $W + Z$ è **diretta** e scriveremo $W \oplus Z$.

Se $W_1, \ldots, W_n$ è un insieme finito di sottospazi di V possiamo definire induttivamente $W_1 + \ldots + W_n$; si ha che $W_1 + \ldots + W_n$ è l'insieme dei vettori che si possono scrivere $w_1 + \ldots + w_n$ con $w_i \in W_i$ per ogni $i = 1, \ldots, n$.

Inoltre

$$W_1 + \ldots + W_n = W_1 \oplus \ldots \oplus W_n$$

se per ogni $i = 1, \ldots, n$ si ha $W_i \cap (W_1 + \ldots +^i + \ldots W_n) = \{0\}$.

Se S è un insieme finito, sia

$$S = \{v_1, v_2, \ldots, v_n\}$$

scriveremo $< v_1, v_2, \dots, v_n >$ invece di $< \{v_1, v_2, \dots, v_n\} >$.
Lo spazio $< v_1, \dots, v_n >$ sarà detto **sottospazio generato** da $v_1, \dots, v_n$. Lo spazio $< v_1, \dots, v_n >$ è anche denotato $L(v_1, \dots, v_n)$.
Gli elementi di tale sottospazio sono i vettori del tipo

$$a_1 v_1 + \ldots + a_n v_n$$

ove $a_1, \dots, a_n \in k$. Un tale vettore sarà detto una **combinazione lineare** di $v_1, \dots, v_n$ e gli scalari $a_1, \dots, a_n$ sono detti i **coefficienti** della combinazione lineare.

Se i vettori $v_1, \dots, v_n \in V$ sono tali che

$$< v_1, \dots, v_n >= V$$

diremo che $v_1, \dots, v_n \in V$ sono **un sistema di generatori** per V. Ciò significa che ogni vettore $v \in V$ si può scrivere come una combinazione lineare di $v_1, \dots, v_n$.

Osserviamo che, ad esempio, $\mathbb{R}^2 =< (1,0), (0,1) >$ in quanto ogni vettore $(a, b) \in \mathbb{R}^2$ si scrive come $a(1,0) + b(0,1)$, ma è anche vero che $\mathbb{R}^2 =< (1,0), (0,1), (1,1) >$ in quanto ogni vettore $(a,b) \in \mathbb{R}^2$ si scrive anche come $(a+1)(1,0) + (b+1)(0,1) - (1,1)$.

Diciamo che i vettori $v_1, \dots, v_n \in V$ sono **linearmente indipendenti** se

$$a_1 v_1 + \ldots + a_n v_n = 0 \Rightarrow a_1 = a_2 = \ldots = a_n = 0.$$

Ciò significa che una combinazione lineare di $v_1, \dots, v_n$ è nulla solo quando tutti i coefficienti della combinazione lineare sono nulli.

Osserviamo che nell'esempio precedente i vettori $(1,0), (0,1)$ sono linearmente indipendenti, mentre $(1,0), (0,1), (1,1)$ sono linearmente dipendenti.

Osservazione 4.2. Se $v_1, \dots, v_n$ sono vettori linearmente indipendenti e $v \in < v_1, \dots, v_n >$, allora v si scrive in modo unico come combinazione lineare di $v_1, \dots, v_n$.

Diremo che un insieme di vettori $v_1, \dots, v_n \in V$ è una **base** di V se

a) $v_1, \dots, v_n$ sono linearmente indipendenti

b) $< v_1, \dots, v_n >= V$, ossia $v_1, \dots, v_n$ sono generatori per V.

Ad esempio una base per lo spazio vettoriale k^n è costituita dall'insieme degli n vettori

$$e_1 := (1, \dots, 0), \qquad e_2 := (0, 1, \dots, 0), \dots, e_n := (0, \dots, 1).$$

Tale base è detta **base canonica di** k^n

Usando il seguente **Lemma di scambio** si può dimostrare che tutte le basi di uno spazio vettoriale che ha un sistema finito di generatori sono equipotenti.

Lemma di scambio. Se $v_1, \dots, v_n$ sono una base di V, $w \in V$ e $w \notin < v_2, \dots, v_n >$, allora $w, v_2, \dots, v_n$ sono una base di V.

Teorema di equipotenza delle basi. Due basi di uno spazio vettoriale V sono formate dallo stesso numero di vettori.

Se V è uno spazio che ha una base, allora si è visto che tutte le basi hanno lo stesso numero di vettori. Tale numero intero si dirà la **dimensione** di V.

Ad esempio si ha

$$dim(k^n) = n.$$

Ci chiediamo ora quando uno spazio vettoriale ha una base.

Lemma di estrazione-completamento. Se $V =< v_1, \dots, v_n >$ e $v_1, \dots, v_k$ sono linearmente indipendenti, allora si può estrarre da $\{v_1, \dots, v_n\}$ una base che contiene $v_1, \dots, v_k$.

La strategia è la seguente: si guarda se $v_{k+1} \in < v_1, \dots, v_k >$. Se si, lo si cancella, se no, si mantiene e $v_1, \dots, v_k, v_{k+1}$ sono linearmente indipendenti. Procedendo in tale modo si arriva alla conclusione.

Come conseguenza si prova che :

Teorema 4.3. Ogni spazio vettoriale non nullo e finitamente generato ha una base.

Come ulteriore applicazione del precedente teorema, si prova il

Teorema di completamento di una base. Se $v_1, \dots, v_k$ sono vettori linearmente indipendenti nello spazio vettoriale V finitamente generato, allora si possono trovare vettori $v_{k+1}, \ldots, v_n \in V$ tali che $v_1, \ldots, v_n$ sia una base di V.

Alcune considerazioni dai fatti precedenti:

• Se $\dim(V) = n$ e $s > n$, allora s vettori in V sono sempre linearmente dipendenti.

• Dato il sistema lineare omogeneo

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = 0 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = 0 \end{cases}$$

usando il punto precedente, si può ottenere il risultato già noto che se il numero n di incognite è maggiore del numero m di equazioni, allora il sistema ha una soluzione

$$(s_1, \ldots, s_n) \neq (0, \ldots, 0).$$

Basta infatti osservare che le colonne della matrice A dei coefficienti individuano vettori di k^m che sono linearmente dipendenti ($n > m$).

• Se $\dim(V) = n$, ogni $n-$upla di vettori $v_1, \ldots, v_n$ linearmente indipendenti in V, forma una base di V.

• m vettori di k^n sono linearmente indipendenti se e solo se la caratteristica della matrice che ha per colonne (o per righe) le coordinate dei vettori, ha caratteristica m.

E' naturale chiedere la relazione che intercorre tra la dimensione di uno spazio vettoriale e quella di un suo sottospazio.

Teorema 4.4. Ogni sottospazio W di uno spazio finitamente generato V è finitamente generato e quindi ammette una base. Inoltre si ha

$$dim(W) \leq dim(V)$$

e vale l'uguale se e solo se $W = V$.

Se W e Z sono sottospazi dello spazio vettoriale V si prova la formula di Grassmann:

$$dim(W + Z) = dim(W) + dim(Z) - dim(W \cap Z).$$

In particolare
$$dim(W \oplus Z) = dim(W) + dim(Z).$$

Teorema di esistenza di un complemento diretto. Se V è finitamente generato, per ogni sottospazio W di V esiste un sottospazio T di V tale che $V = W \oplus T$.

E' facile vedere che per ogni sottospazio vettoriale W di V ci sono diversi sottospazi che sono complemento diretto di W in V. Ciò dipende dal fatto che un sistema di vettori linearmente indipendenti si possono completare ad una base in modi diversi. Vedremo che negli spazi euclidei ci sara' un complemento diretto privilegiato.

Lo studio legato alla determinazione di basi diverse di uno stesso spazio vettoriale può essere anche affrontato tramite l'uso delle matrici e delle applicazioni lineari. Prima di affrontare lo studio delle applicazioni lineari, diamo qualche accenno alla lettura dei fatti già visti in termini di matrici.

Supponiamo $V =< v_1, \ldots, v_n >$ e siano $w_1, \ldots, w_t$ vettori di V. Allora per ogni $i = 1, \ldots, t$ esistono $a_{ji} \in k$, $j = 1, \ldots, n$, tali che
$$w_i = \sum_{j=1}^{n} a_{ji} v_i.$$

Se si introduce la matrice $A = (a_{ij})$ che è di tipo $n \times t$, possiamo rappresentare questa scrittura con la notazione formale
$$(w_1, \ldots, w_t) = (v_1, \ldots, v_n)A.$$

Ad esempio la notazione
$$(w_1, w_2, w_3) = (e_1, e_2) \begin{pmatrix} 2 & 3 & 1 \\ -1 & 0 & -6 \end{pmatrix}$$

significa
$$w_1 = 2e_1 - e_2 = (2, -1), w_2 = 3e_1 + 0e_2 = (3, 0), w_3 = 1e_1 - 6e_2 = (1, -6)$$

Con la notazione precedente e assumendo che $\{v_1, \ldots, v_n\}$ sia una base di V, si possono dimostrare i seguenti fatti

1. $\{w_1, \ldots, w_t\}$ generano V se e solo se A è invertibile a destra (ossia esiste $B \in t \times n$ tale che $AB = I_n$). In particolare segue $t \geq n$.

2. $\{w_1, \ldots, w_t\}$ sono linearmente indipendenti se e solo se A è invertibile a sinistra se e solo se $\rho(A) = t$.

Da questi fatti si deduce una diversa prova dell'equipotenza delle basi che risulta interessante in quanto non usa in **1.** il fatto che gli elementi $a_{ij} \in K^*$ sono invertibili e quindi si presta a una generalizzazione nel caso di strutture algebriche che non richiedano tale proprietà, ad esempio i gruppi abeliani.

5. Applicazioni lineari.

D'ora in poi supporremo che tutti gli spazi vettoriali considerati siano finitamente generati.

Se V e W sono spazi vettoriali su k diciamo che una applicazione

$$f : V \to W$$

è **lineare** o che è un **omomorfismo** di k-spazi vettoriali se sono verificate le seguenti proprietà.

a) $f(v_1 + v_2) = f(v_1) + f(v_2)$, $\forall v_1, v_2 \in V$.

b) $f(av) = af(v)$, $\forall a \in k, \forall v \in V$.

Si vede facilmente che se f è una applicazione lineare si ha:

a) $f(0) = 0$

b) $f(-v) = -f(v), \forall v \in V$.

Ad esempio l'applicazione nulla è sicuramente lineare e se $W = V$ l'applicazione identica è sicuramente lineare.

Se

$$f : V \to W$$

è lineare definiamo $ker(f)$ e lo diciamo il **nucleo** di f, l'insieme di vettori così caratterizzati:

$$ker(f) := \{v \in V | f(v) = 0\}.$$

E' chiaro che $0 \in ker(f)$ e che $ker(f)$ è un sottospazio vettoriale di V. Si prova che

Teorema 5.1. L'applicazione lineare

$$f : V \to W$$

è iniettiva se e solo se $ker(f) = \{0\}$.

Se $f : V \to W$ è lineare definiamo $Im(f)$ e lo diciamo **Immagine** di f, l'insieme di vettori così caratterizzati:

$$Im(f) := \{w \in W | \exists v \in V, f(v) = w\}.$$

Si prova che $Im(f)$ è un sottospazio di W e che f è surgettiva se e solo se $Im(f) = W$ (se e solo se dim $Im(f)$ =dim W in quanto $Im(f) \subseteq W$).

Osserviamo che se $\{v_1, \ldots, v_n\}$ è un sistema di generatori di V, per ogni $v \in V$ possiamo scrivere $v = \sum_{i=1}^{n} \lambda_i v_i$, e quindi

$$f(v) = \sum_{i=1}^{n} \lambda_i f(v_i),$$

in particolare segue che $Im(f) =< f(v_1), \ldots, f(v_n) >$.

Ci si può chiedere se dati n vettori $w_1, \ldots, w_n$ di W esista sempre una applicazione lineare $f : V \to W$ tale che $f(v_i) = w_i$, $\forall i = 1, \ldots, n$. La risposta è negativa. Se però $\{v_1, \ldots, v_n\}$ è una base di V allora tale f esiste ed è unica. Infatti per ogni $v \in V$ potremo scrivere in modo unico $v = \sum_{i=1}^{n} \lambda_i v_i$ e quindi definire la f nel modo seguente:

$$f(v) := \sum_{i=1}^{n} \lambda_i w_i.$$

Un'applicazione lineare $f : V \to W$ è quindi univocamente determinata se si conoscono le immagini degli elementi di una base.

Una applicazione lineare $f : V \to W$ si dice un **isomorfismo** di spazi vettoriali se è contemporaneamente iniettiva e surgettiva.

Due spazi vettoriali V e W tra cui esista un isomorfismo si diranno **isomorfi**.

Teorema 5.2. Se V è spazio vettoriale su k di dimensione n e $\{v_1, \ldots, v_n\}$ è una sua base, allora l'applicazione

$$f : k^n \to W$$

definita con la formula

$$f(a_1, \ldots, a_n) = a_1 v_1 + \ldots + a_n v_n$$

è un isomorfismo di spazi vettoriali.

In particolare, dal teorema precedente, deduciamo che due spazi vettoriali di uguale dimensione sono isomorfi.

Osserviamo inoltre che il Teorema 5.2. afferma che ogni spazio vettoriale V di dimensione n "può essere pensato" come K^n identificando, in modo univoco, ogni vettore di V con la $n-$upla dei coefficienti che esprimono il vettore tramite una base dello spazio vettoriale.

Un importante risultato che lega il nucleo e l'immagine di una applicazione lineare è il seguente teorema.

Teorema di nullità. Se $f : V \to W$ è lineare allora vale:

$$dim(V) = dim(ker(f)) + dim(Im(f)).$$

Come immediato Corollario abbiamo

Corollario 5.3. Se $f : V \to V$ è lineare, allora
$$f \text{ è iniettiva} \iff f \text{ è surgettiva} \iff f \text{ è bigettiva.}$$

Matrici e omomorfismi.

Siano V e W due spazi vettoriali su k, $dim(V) = n$ e $dim(W) = m$. Sia poi $f : V \to W$ una applicazione lineare. Fissiamo una base $E = \{v_1, \ldots, v_n\}$ di V e una base $F = \{w_1, \ldots, w_m\}$ di W.

Allora per ogni $i = 1, \ldots, n$ il vettore $f(v_i)$ è un vettore di W e quindi si potrà scrivere

$$f(v_i) = \sum_{j=1}^{m} a_{ji} w_j.$$

In tal modo resta determinata la matrice (a_{ij}) di tipo $m \times n$ ad elementi in k che chiameremo la **matrice associata** a f mediante le basi E e F e che indicheremo con $M_{EF}(f)$.

Sappiamo che se $f : V \to W$ è una applicazione lineare e $\{v_1, \dots, v_n\}$ un sistema di generatori di V, per ogni $v \in V$ possiamo scrivere $v = \sum_{i=1}^n \lambda_i v_i$, e quindi

$$f(v) = \sum_{i=1}^n \lambda_i f(v_i).$$

Per le considerazioni già fatte si prova facilmente che

Proposizione 5.4. Dati gli spazi vettoriali V e W su k e una applicazione lineare $f : V \to W$, allora

$$\dim \operatorname{Im}(f) = \rho(M_{EF}(f))$$

comunque si fissino $E = \{v_1, \dots, v_n\}$ base di V e $F = \{w_1, \dots, w_m\}$ base di W.

Dati gli spazi vettoriali V e W su k, indichiamo con $Hom(V, W)$ l'insieme delle applicazioni lineari di V in W. Se $f, g \in Hom(V, W)$ e $\alpha \in k$ definiamo

$$(f + g) : V \to W \qquad (\alpha f) : V \to W$$

ponendo

$$(f + g)(v) := f(v) + g(v) \qquad (\alpha f)(v) := \alpha f(v)$$

per ogni $v \in V$.

E' facile dimostrare che $f + g$ e αf sono applicazioni lineari e che, con tali operazioni, $Hom(V, W)$ è un k-spazio vettoriale.

Fissate le basi $E = \{v_1, \dots, v_n\}$ di V e $F = \{w_1, \dots, w_m\}$ di W, definiamo allora una applicazione

$$\varphi : Hom(V, W) \to M_{mn}(k)$$

ponendo

$$\varphi(f) := M_{EF}(f).$$

E' facile provare, usando le considerazioni fatte precedentemente, che φ è lineare e che è un **isomorfismo di spazi vettoriali.**

In particolare ciò significa che valgono le seguenti relazioni:

$$M_{EF}(f + g) = M_{EF}(f) + M_{EF}(g) \qquad M_{EF}(\alpha f) = \alpha M_{EF}(f).$$

Siano dati ora tre spazi vettoriali V, W e Z su k e due applicazioni lineari

$$f : V \to W, \quad g : W \to Z.$$

Se E, F e G sono basi di V, W e Z ci possiamo chiedere come siano legate $M_{EG}(g \circ f)$ e $M_{FG}(g)$, $M_{EF}(f)$.

Supponiamo si abbia $E = \{v_1, \ldots, v_n\}$, $F = \{w_1, \ldots, w_m\}$ e $Z = \{z_1, \ldots, z_p\}$. Se poniamo $M_{FG}(g) = (a_{ij})$ e $M_{EF}(f) = (b_{ij})$ risulta

$$(g \circ f)(v_j) = g[f(v_j)] = g(\sum_{k=1}^{m} b_{kj} w_k) = \sum_{k=1}^{m} b_{kj} g(w_k) =$$

$$= \sum_{k=1}^{m} b_{kj} (\sum_{i=1}^{p} a_{ik} z_i) = \sum_{i=1}^{p} (\sum_{k=1}^{m} a_{ik} b_{kj}) z_i.$$

Questa formula ci dice che la matrice $M_{EG}(g \circ f)$ ha come elemento di posto ij l'elemento $\sum_{k=1}^{m} a_{ik} b_{kj}$ e quindi che

$$M_{EG}(g \circ f) = M_{FG}(g) M_{EF}(f).$$

Usando tale considerazione, si prova facilmente che:

• Date $A \in M_{mn}$ e $B \in M_{np}$, allora

$$\rho(AB) \leq min(\rho(A), \rho(B)).$$

Supponiamo ora di avere una applicazione lineare $f : V \to V$. Osserviamo che se f è un isomorfismo allora è iniettiva e surgettiva; quindi esiste una applicazione $g : V \to V$ tale che $f \circ g = g \circ f = id$. E' facile provare che tale g è una applicazione lineare. Si prova che

Teorema 5.5. Sia V uno spazio vettoriale di dimensione n su k e siano E e F due basi di V. Se $f : V \to V$ è una applicazione lineare e $A = M_{EF}(f)$, sono fatti equivalenti:

1. f è iniettiva.

2. f è surgettiva.

3. f è un isomorfismo.

4. A è invertibile.

5. Esiste una matrice $B \in M_n(k)$ tale che $AB = I$.

6. Esiste una matrice $C \in M_n(k)$ tale che $CA = I$.

7. $\rho(A) = n$.

8. $\det(A) \neq 0$.

Sia ora V uno spazio vettoriale e E e F due basi di V. Allora la matrice $M_{EF}(id)$ associata mediante le basi E e F alla applicazione identica $id : V \to V$ si dice **matrice di passaggio** dalla base F alla base E. Infatti se $M_{EF}(id) = (a_{ij})$ con $E = \{v_1, \ldots, v_n\}$ e $F = \{w_1, \ldots, w_n\}$, si ha per ogni $i = 1, \ldots, n$

$$v_i = id(v_i) = \sum_{k=1}^{n} a_{ki} w_k$$

e quindi tale matrice permette di scrivere i vettori della base E mediante i vettori della base F.

E' chiaro che si ha

$$M_{EF}(i) M_{FE}(i) = M_{FF}(i) = I$$

e quindi

$$(M_{EF}(i))^{-1} = M_{FE}(i).$$

Dunque $M_{EF}(i)$ e $M_{FE}(i)$ sono invertibili e anzi sono una l'inversa dell'altra. Inoltre ogni matrice $n \times n$ invertibile può essere pensata come una matrice di passaggio tra due basi qualunque di uno spazio vettoriale V di dimensione n.

Sia ora

$$f : V \to V$$

una applicazione lineare e E e F due basi di V. Allora risulta

$$M_{FF}(f) = M_{EF}(i) M_{EE}(f) M_{FE}(i) = M_{FE}(i)^{-1} M_{EE}(f) M_{FE}(i).$$

In generale due matrici quadrate A e B in $M_n(k)$ si diranno **simili** se si ha

$$B = P^{-1} A P$$

per qualche matrice $P \in Gl_n(k)$. Osserviamo che due matrici simili hanno lo stesso determinante, fatto che non si verifica in generale per matrici equivalenti.

Con questa definizione possiamo reinterpretare la formula precedente dicendo che matrici che rappresentano la stessa applicazione lineare $f : V \to V$ rispetto a basi diverse sono simili.

Anche il viceversa di tale risultato è vero. Ossia si prova che se A e B sono matrici simili in $M_n(k)$ e V è un k-spazio vettoriale di dimensione n, allora esistono due basi E e F di V e una applicazione lineare $f : V \to V$ tale che $M_{EE}(f) = A, \qquad M_{FF}(f) = B$. La relazione di similitudine è una **relazione di equivalenza** .

6. Omomorfismi diagonalizzabili.

Ci occuperemo dello studio degli endomorfismi di uno spazio vettoriale di dimensione finita. La nostra analisi consiste nel ricercare, per un dato endomorfismo, una “base speciale” dello spazio vettoriale in corrispondenza della quale la matrice associata sia la più semplice possibile. Abbiamo visto che matrici associate allo stesso endomorfismo tramite basi diverse sono simili, cerchiamo quindi un buon rappresentante della classe di equivalenza individuata da una qualsiasi matrice associata all'endomorfismo. Ad esempio ci chiediamo quando è simile ad una matrice diagonale.

Sia V uno spazio vettoriale e $f : V \to V$ una applicazione lineare. Diciamo che f è **diagonalizzabile** se esiste una base E di V tale che la matrice associata a f, $M_{EE}(f)$, sia diagonale.

Ciò significa che esiste una base $E = \{v_1, \dots, v_n\}$ di V e scalari $\lambda_1, \dots, \lambda_n \in k$ tali che

$$f(v_i) = \lambda_i v_i \qquad \forall i = 1, \dots, n$$

equivalentemente

$$M_E(f) = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \lambda_n \end{pmatrix}$$

Diremo pertanto che $\lambda \in k$ è un **autovalore** per f se esiste un vettore non nullo $v \in V$ tale che

$$f(v) = \lambda v.$$

Sia $\lambda \in k$ un autovalore per f. L'insieme

$$V_\lambda := \{v \in V | f(v) = \lambda v\}$$

è un sottospazio di V: infatti è il nucleo dell'omomorfismo

$$\lambda id - f : V \to V.$$

Tale sottospazio $V_\lambda = Ker(\lambda id - f)$ si dice **autospazio** associato a λ e i suoi vettori sono detti gli **autovettori** di λ. E' chiaro che per definizione di autovalore risulta $dim(V_\lambda) > 0$.

Osserviamo che $f(V_\lambda) \subseteq V_\lambda$. Più in generale diciamo che un sottospazio vettoriale W di V è $f-$invariante se $f(W) \subseteq W$. Poiché una base di W può essere completata a una base F di V, il fatto che W è $f-$invariante si può leggere direttamente dalla matrice associata a f che sarà della forma:

$$M_F(f) = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}.$$

Supponiamo che $V = W_1 \oplus W_2$ sia somma diretta di due sottospazi $f-$invarianti, allora esiste una base B di V (unione delle basi dei W_i) tale che la matrice associata a f mediante B sarà della forma:

$$M_B(f) = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}.$$

Osserviamo che se λ un autovalore di f, in particolare V_λ è invariante.

Vediamo ora come determinare gli autovalori di un endomorfismo.

Sia $f : V \to V$ un omomorfismo e λ un elemento di k. Se A è la matrice associata ad f rispetto ad una base E di V, allora la matrice $\lambda I - A$ è associata all'applicazione lineare $\lambda id - f$.

Ne segue che un elemento $\lambda \in k$ è un autovalore di f se e solo se $ker(\lambda i - f) \neq 0$ se e solo se $det(\lambda I - A) = 0$.

Dunque **gli autovalori di f sono gli elementi di k che sono radici del polinomio** $det(XI - A)$.
E' facile provare che se A e B sono matrici simili allora

$$det(XI - A) = det(XI - B).$$

Tale polinomio non dipende quindi dalla base scelta, ma soltanto dall'applicazione lineare f. Diremo che

$$p_f(X) := det(XI - A)$$

è il **polinomio caratteristico** di f.

E' chiaro che $p_f(X)$ è un polinomio monico di grado n, se $n = dim(V)$. Inoltre, se scriviamo

$$p_f(X) = X^n + c_1 X^{n-1} + \ldots + c_n,$$

allora si ha $c_n = (-1)^n det(A)$ e $c_1 = -Tr(A)$, ove $Tr(A)$ è la **traccia** di A ossia la somma degli elementi di A che sono sulla diagonale principale.

Ciò prova in particolare che matrici simili hanno lo stesso determinante e la stessa traccia. Tuttavia le due matrici

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

hanno lo stesso polinomio caratteristico, ma non sono simili.

Si prova facilmente che un omomorfismo f nilpotente ha come matrice associata una matrice nilpotente e ha come unico autovalore zero, quindi $p_f(X) = x^n$. Segue quindi che la traccia di una matrice nilpotente è nulla.

Le considerazioni precedenti rendono importante studiare come trovare le radici di un polinomio a coefficienti razionali, reali o complessi. Ora se $f(X)$ è un polinomio a coefficienti complessi di grado n, il teorema fondamentale dell'algebra ci assicura che $f(X)$ ha esattamente n radici complesse se si contano con la dovuta molteplicità.

Se invece $f(X)$ ha coefficienti reali sappiamo che può non avere alcuna radice reale. Esistono endomorfismi di $k-$spazi vettoriali di dimensione $n > 1$ o matrici $n \times n$ con $n > 1$, privi di autovalori e autovettori. Ad esempio $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in M_2(\mathbb{R})$.

Se $p(X) = a_0 + a_1 X + \ldots + a_m X^m$ è un qualunque polinomio a coefficienti in k e $A \in M_n(k)$, possiamo considerare

$$p(A) := a_0 I_n + a_1 A + \ldots + a_m A^m \ \in M_n(k).$$

Si può dimostrare facilmente che, data una matrice $A \in M_n(k)$, esistono polinomi non nulli $p(X) \in k[X]$ tali che $p(A) = 0$. Basta osservare che $M_n(k)$ come $k-$spazio vettoriale ha dimensione n^2 e quindi i vettori $I_n, A, A^2, \ldots, A^m$ sono linearmente dipendenti se $m \geq n^2$.
C'è un risultato più profondo (Teorema di Cayley-Hamilton) che non tratteremo in questo corso il quale afferma che se $A = M_E(f)$, allora

$$p_f(A) = 0.$$

Definizione. Sia $A \in M_n(k)$, diremo polinomio minimo di A il polinomio monico $m(X)$, di grado minimo, tale che $m(A) = 0$.

Chiaramente il polinomio caratteristico di A e' multiplo del polinomio minimo.

Ritorniamo alla ricerca di un criterio che caratterizzi gli endomorfismi diagonalizzabili.

Lemma 6.2. *Se $\lambda_1, \ldots, \lambda_r$ sono autovalori distinti di f e se $v_1, \ldots, v_r$ sono corrispondenti autovettori non nulli, allora $v_1, \ldots, v_r$ sono linearmente indipendenti.*

In particolare si prova facilmente che se $\lambda_1, \ldots, \lambda_r \in k$ autovalori distinti di f, allora

$$V_{\lambda_1} + \ldots + V_{\lambda_r} = V_{\lambda_1} \oplus \ldots \oplus V_{\lambda_r}.$$

Quindi possiamo provare che $f : V \to V$ è diagonalizzabile se e solo se esiste una base di V costituita da autovettori se solo se esistono $\lambda_1, \ldots, \lambda_r \in k$ auotvalori tali che $V = V_{\lambda_1} \oplus \ldots \oplus V_{\lambda_r}$.

Dal lemma precedente e dal fatto che f e' diagonalizzabile se e solo se esiste una base di V costituita da autovettori, segue immediatamente un primo risultato di diagonalizzabilità per un omomorfismo.

Teorema 6.3. *Se V ha dimensione n e f ha n autovalori distinti, allora f è diagonalizzabile.*

Il viceversa di questo risultato non vale: infatti l'applicazione identica di V in V è diagonalizzabile ma ha il solo autovalore 1.

Osserviamo che se λ è autovalore per f, allora

$$0 < dim V_\lambda \leq m(\lambda)$$

dove $m(\lambda)$ denota la molteplicità di λ in $p_f(X)$.
Infatti se $E = \{v_1, \ldots v_r\}$ è una base di V_λ, completando E a base di V, segue facilmente che $(X - \lambda)^r$ divide $p_f(X)$ e quindi $r \leq m(\lambda)$.

Teorema 6.4. *Se $f : V \to V$ è un omomorfismo, $\lambda_1, \ldots, \lambda_r$ sono le radici distinte del polinomio caratteristico di f, e $m_1, \ldots, m_r$ le corrispondenti molteplicità. Allora f è diagonalizzabile se e solo se per ogni $i = 1, \ldots, r$, $\lambda_i \in k$ e $dim(V_{\lambda_i}) = m_i$.*

Sia A una matrice quadrata ad elementi in k. Diremo che A è **diagonalizzabile** se A è simile ad una matrice diagonale. Allora:

Teorema 6.5. *Sia $A \in M_n(k)$ una matrice quadrata e sia V un k-spazio vettoriale di dimensione n. Se E è una base di V ed $f : V \to V$ l'omomorfismo tale che $M_{EE}(f) = A$, allora si ha che f è diagonalizzabile se e solo se A è diagonalizzabile.*

7. Prodotto scalare e matrici ortogonali

Lo scopo di questo capitolo e' di estendere il concetto di prodotto scalare gia' visto in $\mathbb{R}^2$ e $\mathbb{R}^3$ agli spazi vettoriali di dimensione qualsiasi e quindi di introdurre le nozioni di perpendicolarit e di angolo anche in questi casi, permettendo, di conseguenza, lo studio della geometria euclidea negli spazi di dimensione qualsiasi.

Definizione. Sia V un $\mathbb{R}$-spazio vettoriale. Si definisce prodotto scalare la funzione

$$\cdot: \quad V \times V \to \mathbb{R}, \qquad (v_1, v_2) \to v_1 \cdot v_2, \qquad \forall v_1, v_2 \in V.$$

tale che se v, v_1, v_2, v_3 sono vettori di V, e $\lambda \in \mathbb{R}$, valgono le seguenti proprietà.

1. $v_1 \cdot v_2 = v_2 \cdot v_1$ per ogni $v_1, v_2 \in V$ *(commutatività)*
2. $(v_1 + v_2) \cdot v_3 = v_1 \cdot v_3 + v_2 \cdot v_3 \quad \lambda(v \cdot w) = (\lambda v) \cdot w$ *(linearità)*
3. $v \cdot v \geq 0$ e $v \cdot v = 0$ se e solo se $v = 0$.

Come immediata conseguenza delle definizioni si ha :

a) $\lambda(v_1 \cdot v_2) = (\lambda v_1) \cdot v_2 = v_1 \cdot (\lambda v_2)$

b) $0 \cdot v = v \cdot 0 = 0$

In $\mathbb{R}^n$ consideriamo i vettori $v = (x_1, \dots, x_n)$ e $w = (y_1, \dots, y_n)$, diremo **prodotto scalare standard** in $\mathbb{R}^n$ il numero reale

$$v \cdot w := x_1 y_1 + \ldots + x_n y_n$$

ottenuto eseguendo il prodotto righe per colonne del "vettore riga" v per il "vettore colonna" w. Si lascia per esercizio la verifica delle proprieta' 1., 2. e 3.

Esempio. Si consideri la funzione

$$\cdot: \quad \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$$

definita da $(x_1, x_2, x_3) \cdot (y_1, y_2, y_3) = 2x_1y_1 + 3x_2y_2 + 5x_3y_3$. Si verifichi che e' un prodotto scalare su $\mathbb{R}^3$, chiaramente diverso dal prodotto scalare standard. L'esempio appena incontrato permette di dedurre che, almeno su $\mathbb{R}^n$, e' possibile definire infiniti prodotti scalari, quindi infinite strutture euclidee.

Esempio. Si consideri la funzione

$$\cdot: \quad \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$$

definita da $(x_1, x_2, x_3) \cdot (y_1, y_2, y_3) = 2x_1y_1 + 3x_2y_2 - x_3y_3$. Si osserva che non e' un prodotto scalare su $\mathbb{R}^3$, perche' per esempio $(0, 0, 1) \cdot (0, 0, 1) < 0$ che contraddice il terzo assioma di definizione di prodotto scalare.

Esercizio. Si verifichi che in $M_{n,m}(\mathbb{R})$ la funzione

$$A \cdot B := tr({}^tA \cdot B)$$

e' un prodotto scalare.

Sia V uno spazio euclideo. Per ogni vettore $v \in V$ definiamo la **norma** o **modulo** di v come il numero reale

$$\|v\| := \sqrt{v \cdot v}.$$

In $\mathbb{R}^n$ con il prodotto scalare standard si ha

$$\|v\| = \sqrt{x_1^2 + \ldots + x_n^2}.$$

Si verifica che

1. $\|v\| \geq 0$ e $\|v\| = 0$ se e solo se $v = 0$.

2. Disuguaglianza di Cauchy-Schwartz: $v \cdot w \leq \|v\|\|w\|$.

3. Disuguaglianza triangolare: $\|v + w\| \leq \|v\| + \|w\|$.

Definizione. Dati $v, w \in V$ si definisce angolo tra i due vettori v e w, l'angolo $\alpha \in [0, \pi]$ tale che

$$cos\alpha = v \cdot w / \|v\|\|w\|.$$

Se $v, w \in V$ diciamo che v e w sono **ortogonali** e scriviamo $v \perp w$ se

$$v \cdot w = 0.$$

Se S è un sottoinsieme di V diciamo **ortogonale di** S l'insieme

$$S^\perp := \{v \in V | v \cdot w = 0, \quad \forall w \in S\}.$$

Ad esempio se $S = \{(1, 0, 1), (1, 1, 0)\}$, allora

$$S^\perp := \{(x, y, z) \in \mathbb{R}^n | x + z = x + y = 0\}.$$

Segue che $S^{\perp} =< (-1,1,1) >$.

E' facile dimostrare che $S^{\perp}$ e' un sottospazio di V, anche se S non lo è. E' chiaro che $\{0_V\}^{\perp} = V$ e $V^{\perp} = \{0_V\}$.

Se $W =< v_1, \dots, v_n >$, risulta $v \in W^{\perp}$ se e solo se $v \perp v_i$ per ogni $i = 1, \dots, n$. Quindi se $S = \{v_1, \dots, v_n\}$, allora

$$W^{\perp} = S^{\perp}.$$

E' chiaro che

$$W \cap W^{\perp} =< 0 >$$

Un sottoinsieme S di V non contenente il vettore nullo **si dice ortogonale** se è costituito da un solo vettore oppure se i suoi vettori sono a due a due ortogonali ($v_i \cdot v_j = 0$ per ogni $i \neq j$.)
S si dice **ortonormale** se è ortogonale e inoltre i suoi vettori sono di norma 1.

Proposizione 7.2. *Ogni insieme ortogonale è costituito da vettori linearmente indipendenti.*

Infatti se $S = \{v_1, \dots, v_r\}$ e $a_1 v_1 + \ldots + a_r v_r = 0$, moltiplicando scalarmente per v_i, $i = 1, \dots, r$, si ottiene

$$a_i v_i \cdot v_i = 0$$

da cui segue $a_i = 0$.

Quindi un insieme ortogonale non può contenere più di $n = dimV$ vettori. Un sottoinsieme S ortogonale costituito da n vettori è base di V e si dice **base ortogonale** di V. Si dice **base ortonormale** se i suoi vettori sono di norma uno.

Facciamo ora vedere come da una base qualsiasi di V si possa costruire una base ortonormale. Tale procedimento è noto come *metodo di ortonormalizzazione di Gram-Schmidt.*

Sia $B = \{v_1, \dots, v_n\}$ una base di V contenente il sottoinsieme $S = \{v_1, \dots, v_t\}$ ortonormale. Possiamo sempre ridurci a tale situazione, eventualmente sostituendo v_1 con $vers(v_1) = v_1/\|v_1\|$ e considerando $t = 1$. Possiamo supporre $t < n$ e consideriamo il vettore

$$w_{t+1} := v_{t+1} - \sum_{i=1}^{t} (v_{t+1} \cdot v_i) v_i.$$

Chiaramente $w_{t+1} \in V$ ed è ortogonale a ciascun vettore di S. Segue che

$$S \cup \{w_{t+1}/\|w_{t+1}\|\}$$

è ortonormale e quindi la base ortonormale di V si ottiene ripetendo il procedimento.

Sia W un sottospazio dello spazio euclideo V e $\{v_1, \ldots, v_n\}$ una base **ortogonale** di W. Per ogni $v \in V$ il vettore $v - \sum_{i=1}^{n}(v \cdot v_i)v_i$ è ortogonale a tutti i vettori di W, ossia

$$v - \sum_{i=1}^{n}(v \cdot v_i)v_i \in W^{\perp}.$$

Poiché $\sum_{i=1}^{n}(v \cdot v_i)v_i \in W$ e $W \cap W^{\perp} = \{0\}$, si ha

$$V = W \oplus W^{\perp}.$$

Da questo e' facile provare che

$$(W^{\perp})^{\perp} = W.$$

Teorema. *Sia $(V, \cdot)$ uno spazio euclideo reale.*

1. *Sia $B = \{e_1, \ldots, e_n\}$ una base ortonormale di V. Allora per ogni*

$$v = x_1e_1 + \ldots + x_ne_n, \quad w = y_1e_1 + \ldots + y_ne_n$$

si ha

$$v \cdot w = x_1y_1 + \ldots + x_ny_n.$$

2. *Sia $B = \{e_1, \ldots, e_n\}$ una base di V tale che per ogni*

$$v = x_1e_1 + \ldots + x_ne_n, \quad w = y_1e_1 + \ldots + y_ne_n$$

si ha

$$v \cdot w = x_1y_1 + \ldots + x_ny_n.$$

Allora B e' una base ortonormale.

In analogia con la definizione gia' vista in $\mathbb{R}^2$ e in $\mathbb{R}^3$:

Definizione. Sia $(V, \cdot)$ uno spazio euclideo reale e siano $v, w \in V$. Definiamo distanza tra v e w il numero reale

$$d(v, w) = \|v - w\|.$$

Sia $W \subseteq V$ un sottospazio vettoriale e sia $v \in V$. Si dimostra che esiste un unico vettore in W che abbia distanza minima da v. Tale vettore e' detto **proiezione ortogonale di v su W** e si denota con $pr_W(v)$.

Se $\{w_1, \ldots, w_n\}$ e' una base di W, allora

$$pr_W(v) = \sum_{i=1}^{n} (v \cdot w_i) w_i.$$

Si ricordi che ad esempio se $V = \mathbb{R}^2$, e $W =< w >$ con $\|w\| = 1$, allora per ogni $v \in V$

$$pr_w(v) = (v \cdot w)w.$$

Definizione. Una matrice quadrata A si dice **otogonale** se

$$A^t A = I.$$

Segue facilmente che se A e' una matrice ortogonale, allora
1. A e' invertibile e $A^{-1} =^t A$, segue che ${}^tAA = I$.
2. $det(A) = \pm 1$
3. se B e' ortogonale, allora AB e' ortogonale.

Proposizione *Sia $S = \{v_1, \ldots, v_n\}$ un insieme di vettori e sia $A \in M_n(\mathbb{R})$ una matrice che come righe (risp. colonne) i vettori di S. Allora*

$$S \text{ e' un insieme ortonormale} \iff A \text{ e' una matrice ortogonale}$$

Enunciamo la seguente proposizione che dà una motivazione geometrica allo studio delle matrici ortogonali:

Proposizione 7.1. *Sia $m : \mathbb{R}^n \to \mathbb{R}^n$ una applicazione lineare. Le seguenti condizioni sono equivalenti:*

1. m è un movimento rigido (conserva le distanze) che lascia fissa l'origine.

2. m conserva il prodotto scalare, ossia, per ogni X e Y in $\mathbb{R}^n$ si ha : $X \cdot Y = m(X) \cdot m(Y)$.

3. Esiste una matrice ortogonale A tale $m(X) = AX$ per ogni X in $\mathbb{R}^n$.

8. Diagonalizzazione delle matrici simmetriche reali

Le matrici ortogonali giocano un ruolo essenziale nella diagonalizzazione di matrici simmetriche reali. Vedremo che ogni matrice simmetrica reale A è simile ad una matrice diagonale tramite una matrice P ortogonale, ossia esiste P ortogonale e una matrice Δ diagonale tali che tale che

$$\Delta =^t PAP$$

Si prova in particolare che $\mathbb{R}^n$ ammette una base B ortonormale formata da autovettori per A.

Ricordiamo che $A = (a_{ij}) \in M_n(K)$ e' simmetrica se ${}^tA = A$, ossia se $a_{ij} = a_{ji}$ per ogni i, j.

Sia V un $\mathbb{R}$-spazio euclideo.

Definizione. Un endomorfismo f di V si dice **autoaggiunto** se per ogni $v, w \in V$ si ha

$$f(v) \cdot w = v \cdot f(w).$$

Si osservi che $B = \{v_1, \ldots, v_n\}$ e' una base di V, allora f e' autoaggiunto se e solo se per ogni $i, j = 1, \ldots, n$ si ha

$$f(v_i) \cdot v_j = v_i \cdot f(v_j).$$

Proposizione. Sia $B = \{v_1, \ldots, v_n\}$ e' una base ortonormale di V. Allora

$$f \text{ e' autoaggiunto} \quad \Longleftrightarrow \quad M_{BB}(f) \text{ e' simmetrica}$$

Per provare che le matrici simmetriche sono diagonalizzabili, cominciamo a provare la prima condizione, ossia che gli autovalori sono reali.

Teorema *Se $A \in M_n(\mathbb{R})$ è una matrice simmetrica reale, ogni radice del suo polinomio caratteristico $|XI - A|$ è un numero reale.*

Proposizione. *Sia A una matrice simmetrica reale e siano λ_1, λ_2 due autovalori distinti di A. Allora*

$$V_{\lambda_1} \perp V_{\lambda_2}.$$

Teorema 7.4. *Siano $f : \mathbb{R}^n \to \mathbb{R}^n$ un omomorfismo e E una base di $\mathbb{R}^n$ tale che $A = M_{EE}(f)$ è simmetrica, allora $\mathbb{R}^n$ possiede una base ortonormale di autovettori.*

Come corollario immediato si ottiene:

Teorema 7.5. *Se A è una matrice simmetrica reale, allora esiste una matrice ortogonale U tale che ${}^tUAU$ è diagonale.*

Per la precedente proposizione, si osserva infatti che l'unione di basi ortonormali degli autospazi e' una base ortonormale di $\mathbb{R}^n$. Allora la matrice U e' la matrice che ha per colonne la base ortonormale costituita da autovettori.

8. Matrici Hermitiane

La teoria delle matrici simmetriche reali si puo' estendere al caso di matrici complesse. Occorre dapprima presentare la definizione di prodotto scalare in un $\mathbb{C}$-spazio vettoriale V.

Definizione. Sia V un $\mathbb{C}$-spazio vettoriale. Si definisce prodotto scalare la funzione

$$\cdot : \quad V \times V \to \mathbb{C}, \qquad (v_1, v_2) \to v_1 \cdot v_2, \qquad \forall v_1, v_2 \in V.$$

tale che se v, v_1, v_2, v_3 sono vettori di V, e $\lambda \in \mathbb{C}$, valgono le seguenti proprietà.

1. $v_1 \cdot v_2 = \overline{v_2 \cdot v_1}$ per ogni $v_1, v_2 \in V$

2. $(v_1 + v_2) \cdot v_3 = v_1 \cdot v_3 + v_2 \cdot v_3 \qquad \lambda(v_1 \cdot v_2) = (\lambda v_1) \cdot v_2$ *(linearità)*

3. $v \cdot v \geq 0$ e $v \cdot v = 0$ se e solo se $v = 0$.

Come immediata conseguenza delle definizioni si ha :

a) $v_1 \cdot (\lambda v_2) = \overline{\lambda}(v_1 \cdot v_2)$

b) $0 \cdot v = v \cdot 0 = 0$

Osserviamo che per 1. si ha $v \cdot v = \overline{v \cdot v}$, quindi $v \cdot v \in \mathbb{R}$ e 3. ha significato.

In $\mathbb{C}^n$ consideriamo i vettori $v = (x_1, \ldots, x_n)$ e $w = (y_1, \ldots, y_n)$, diremo **prodotto scalare standard** in $\mathbb{C}^n$ il numero reale

$$v \cdot w := x_1\overline{y_1} + \ldots + x_n\overline{y_n}$$

Si lascia per esercizio la verifica delle proprieta' 1., 2. e 3.

Se V e' un $\mathbb{C}$-spazio euclideo con prodotto scalare, come nel caso reale, la norma di un vettore $v \in V$ e' il numero reale

$$\|v\| := \sqrt{v \cdot v}.$$

In $\mathbb{C}^n$ con il prodotto scalare standard si ha

$$\|v\| = \sqrt{x_1\overline{x_1} + \ldots + x_n\overline{x_n}}.$$

Si verifica che

1.$\|v\| \geq 0$ e $\|v\| = 0$ se e solo se $v = 0$.

La norma ha le stesse proprieta' come nel caso reale. Dal prodotto scalare, e' definita l'ortogonalita' di due vettori. Si estendono quindi le definizioni di insieme ortogonale, insieme ortonormale, di base ortogonale, base ortonormale, ecc....

Esempio. Ad esempio se $V = \mathbb{C}^2$, allora

$$B = \{1/2(1+i, -1+i), 1/2(1+i, 1-i)\}$$

e' una base ortonormale.

Definizione. Una matrice $P \in M_n(\mathbb{C})$ si dice **unitaria** se

$$P\ {}^t\overline{P} = I$$

In particolare una matrice unitaria e' invertibile e $P^{-1} =^t \overline{P}$. E' immediato verificare che le matrici ortogonali reali sono unitarie.

Analogamente al corrispondente caso reale, si prova che:

A e' unitaria $\iff$ I vettori riga (colonna) costituiscono un insieme ortonormale

Si prova tramite l'algoritmo di Gram-Schmidt che un $\mathbb{C}$-spazio vettoriale finitamente generato e non nullo, ammette sempre una base ortonormale.

Matrici Hermitiane

Definizione. Una matrice $A \in M_n(\mathbb{C})$ si dice **Hermitiana** se ${}^tA = \overline{A}$.

Quindi $A = (a_{ij})$ e' Hermitiana se e sole se $a_{ij} = \overline{a_{ji}}$ per ogni i, j. In particolare gli elementi sulla diagonale principale sono reali.

Ad esempio la matrice $A = \begin{pmatrix} 2 & 1+2i \\ 1-2i & -1 \end{pmatrix} \in M_2(\mathbb{C})$ e' Hermitiana.

Inoltre se $A \in M_n(\mathbb{R})$, la matrice A e' Hermitiana se e solo se A e' simmetrica.

La definizione di autoaggiunto si estende anche al caso complesso senza variazioni. Analogamente al caso reale e ricordando le proprieta' del prodotto scalare in $\mathbb{C}$ si provano i seguenti risultati (le dimostrazione si lasciano per esercizio estendendo in modo naturale le dimostrazioni viste nel caso reale).

Proposizione. Sia B una base ortonormale di un $\mathbb{C}$-spazio vettoriale V con prodotto scalare. Allora $f \in Hom_{\mathbb{C}}(V, V)$ e' autoaggiunto se e solo se $M_B(f)$ e' Hermitiana.

Proposizione. Sia $AA \in M_n(\mathbb{C})$ una matrice Hermitiana. Allora le radici di $p_A(x)$ sono reali.

Usando i fatti precedenti si prova l'estensione del Teorema sulla diagonalizzabilita' delle matrici simmetriche reali.

Teorema *Se A è una matrice Hermitiana, allora A e' diagonalizzabile ed esiste una matrice unitaria U tale che ${}^t\overline{U}AU$ è diagonale.*

8. Forme quadratiche

Sia $A \in M_n(\mathbb{R})$ e siano $x_1, \dots, x_n$ indeterminate. Se indichiamo con X la matrice $(x_1 \dots x_n)$, la matrice $XA\ {}^tX$ è un polinomio omogeneo di secondo grado in $x_1, \dots, x_n$ che si dice **forma quadratica** associata alla matrice A.

E' facile vedere che matrici diverse possono individuare la stessa forma quadratica. Ad esempio se

$$A = \begin{pmatrix} 1 & -1 \\ -3 & 5 \end{pmatrix} \quad B = \begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix}$$

allora

$$XA\ {}^tX = XB\ {}^tX = x_1^2 - 4x_1x_2 + 5x_2^2.$$

Se viceversa $Q(x_1, \ldots, x_n)$ è una forma quadratica reale, ossia un polinomio omogeneo di secondo grado in $x_1, \ldots, x_n$, è chiaro che ci possono essere tante matrici A tali che $Q(x_1, \ldots, x_n) = XA\ {}^tX$, ma una sola di esse è simmetrica. Infatti se

$$Q(x_1, \ldots, x_n) = \sum_{1 \leq i \leq j \leq n} b_{ij} x_i x_j$$

è chiaro che la matrice $A = (a_{ij})$ ove

$$a_{ij} = \begin{cases} \frac{b_{ij}}{2}, & \text{se i minore di j} \\ \frac{b_{ji}}{2}, & \text{se j minore di i} \\ b_{ii}, & \text{se i=j .} \end{cases}$$

è l'unica matrice simmetrica tale che $Q(x_1, \ldots, x_n) = XA\ {}^tX$.

Nel seguito diremo matrice associata ad una forma quadratica quella simmetrica.

Data la forma quadratica $Q(x) := Q(x_1, \ldots, x_n) = XA\ {}^tX$ con A matrice simmetrica reale , sia U la matrice unitaria che diagonalizza A, ossia tale che $U^{-1} =^t\ U$ e

$${}^tUAU = \Delta = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}$$

Allora si ha

$$A = U\Delta^tU$$

e se si considera la "trasformazione"

$$Y = XU$$

si ottiene

$${}^tY =^t U^tX$$

e quindi:

$$Q(X) = XA\ {}^tX = X(U\Delta^tU)\ {}^tX = Y\Delta^tY = \sum_{i=1}^{n} \lambda_i y_i^2.$$

Diremo che la trasformazione $Y = XU$ riduce la forma quadratica alla forma canonica.

Definizione. Una forma quadratica si dice in **forma canonica** se la matrice associata e' diagonale.

Esempio. Ridurre a forma canonica la seguente forma quadratica $Q(x_1, x_2, x_3) = 2x_1x_2 - x_3^2$.

Si ha $Q = XA^tX$ ove

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

Calcoliamo il polinomio caratteristico di A. Si ha

$$|XI - A| = det \begin{pmatrix} X & -1 & 0 \\ -1 & X & 0 \\ 0 & 0 & X+1 \end{pmatrix} = (X+1)(X^2-1) = (X+1)^2(X-1).$$

Quindi gli autovalori di A sono 1 con molteplicità 1 e -1 con molteplicità 2.

L'autospazio V_1 è costituito dai vettori soluzione del sistema

$$\begin{cases} x_1 - x_2 = 0 \\ 2x_3 = 0 \end{cases}$$

e quindi una sua base è $(1, 1, 0)$.

L'autospazio V_{-1} è costituito dai vettori soluzione del sistema

$$-x_1 - x_2 = 0$$

e quindi una sua base è $(1, -1, 0), (0, 0, 1)$. La base di V_1 si ortonormalizza immediatamente ottenendo $(\sqrt{2}/2, \sqrt{2}/2, 0)$.

Una base ortonormale di V_{-1} è data dai vettori $(\sqrt{2}/2, -\sqrt{2}/2, 0), (0, 0, 1)$.

Sia allora

$$P := \begin{pmatrix} \sqrt{2}2 & \sqrt{2}/2 & 0 \\ \sqrt{2}/2 & -\sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Se consideriamo la trasformazione

$$Y = XP$$

ossia la trasformazione

$$\begin{cases} y_1 = \sqrt{2}/2x_1 + \sqrt{2}/2x_2 \\ y_2 = \sqrt{2}/2x_1 - \sqrt{2}/2x_2 \\ y_3 = x_3. \end{cases}$$

si ottiene

$$Q(x_1, x_2, x_3) = y_1^2 - y_2^2 - y_3^2.$$

E infatti è facile verificare che

$$(\sqrt{2}/2x_1 + \sqrt{2}/2x_2)^2 - (\sqrt{2}/2x_1 - \sqrt{2}/2x_2)^2 - x_3^2 = 2x_1x_2 - x_3^2.$$

Segno di una forma quadratica

Definizione. Una forma quadratica $Q(X)$ si dice
1. *definita positiva* se $Q(X) > 0 \ \forall X \in \mathbb{R}^n, X \neq 0$
2. *semidefinita positiva* se $Q(X) \geq 0 \ \forall X \in \mathbb{R}^n$
3. *definita negativa* se $Q(X) < 0 \ \forall X \in \mathbb{R}^n, X \neq 0$.
4. *semidefinita negativa* se $Q(X) \leq 0 \ \forall X \in \mathbb{R}^n$,
5. *indefinita* altrimenti.

Si osservi che quindi X e' indefinita se esistono $X, Y \in \mathbb{R}^N$ tali che $Q(X) < 0$ e $Q(Y) > 0$.

Esempi. 1. $Q(x, y, z) = x^2 + 4xy + y^2 + z^2$ e' indefinita perche' $Q(1, 0, 0) > 0$ e $Q(-1, 1, 0) < 0$.

2. $Q(x, y, z) = x^2 + 2xy + y^2 + z^2 = (x + y)^2 + z^2 \geq 0$ e' semidefinita positiva.

Osservazione. Sia $Q(X) = XA^tX$ con $A = (a_{ij})$. Si osservi che

$$Q(1, 0, \ldots, 0) = a_{11}, Q(0, 1, \ldots, 0) = a_{22}, \ldots, Q(0, 0, \ldots, 1) = a_{nn},$$

quindi il segno degli elementi sulla diagonale principale da' condizioni necessarie per determinare il segno di una forma quadratica. Ad esempio se $Q(X)$ e semidefinita positiva, allora $a_{ii} \geq 0$. Non vale il viceversa come si evince dall'esempio 1.

Osservazione. Se $Q(X) = X\Delta^tX$ e' in forma canonica il segno della forma quadratica dipende dai segni dei coefficienti di x_i^2 o equivalentemente dai segni degli elementi sulla diagonale principale di Δ.

E' quindi naturale chiedersi se il segno di una forma quadratica possa essere valutato sulla forma canonica, ossia attraverso il segno degli autovalori della matrice associata.

Definizione. Siano Q e R due forme quadratiche con matrici associate rispettivamente A e B, si dice che Q e R sono equivalenti se esiste $P \in Gl_n(\mathbb{R})$ tale $B =^t PAP$.

In particolare una forma quadratica e una corrispondente forma quadratica canonica sono equivalenti.

Osservazione. Due matrici simmetriche A e B sono simili se esiste una matrice ortogonale P tale che

$$B =^t PAP.$$

E' facile da provare usando il fatto che A e B sono simili alla stessa matrice diagonale tramite matrici ortogonali. Quindi due forme quadratiche sono equivalenti se le corrispondenti matrici associate sono simili.

Ricordiamo che due matrici simili hanno la stessa caratteristica e stesso polinomio caratteristico. Quindi ad esempio $Q(x, y, z) = x^2 + 2xy + y^2 + z^2$ e $R(x, y, z) = x^2 + 4xy + y^2 + z^2$ non sono forme equivalenti perche' le matrici associate non hanno stessa caratteristica.

Osservazione. Provare che $Q(x_1, x_2) = x_1 x_2$ e $Q(y_1, y_2) = (1/2)y_1^2 - (1/2)y_2^2$ sono equivalenti.

Proposizione. Due forme quadratiche equivalenti hanno lo stesso segno.

Si prova infatti che le forme quadratiche $R(Y)$ e $Q(X)$ equivalenti assumono gli stessi valori. Sia infatti $Y_0 \in \mathbb{R}^n$ e sia P la matrice invertibile tramite la quale le corrispondenti matrici associate sono equivalenti. Se $X_0 = Y_0^t P$, allora

$$Q(X_0) = Q(Y_0^t P) = R(Y_0).$$

Con il cambiamento di coordinate ${}^tX = P\ {}^tY$ la forma quadratica $Q(X)$ viene trasformata nella forma quadratica equivalente $R(Y)$.

Abbiamo provato che ogni forma quadratica $Q(X) = XA^tX$ ha una forma canonica del tipo

$$R(Y) = \lambda_1 Y_1^2 + \ldots + \lambda_n Y_n^2$$

dove $\lambda_1, \ldots, \lambda_n$ sono gli autovalori di A. Quindi il segno di Q e' determinato dagli autovalori di $A \in M_n(\mathbb{R})$. Se denotiamo con:
$P =$ numero autovalori positivi
$N =$ numero autovalori negativi
$Z =$ numero autovalori nulli
allora $P + N + Z = n$ e $P + N = \rho(A)$.

Si ha che
A e' definita positiva se $P = n$

A e' definita negativa se $N = n$
A e' semidefinita positiva se $N = 0$
A e' semidefinita negativa se $P = 0$
A e' indefinita $N > 0$ e $P > 0$.

In generale non e' facile determinare esplicitamente gli autovalori, ma osserviamo che per determinare il segno di una forma quadratica e' sufficiente avere informazioni sul segno delle radici (tutte reali!) del polinomio caratteristico della matrice associata alla forma quadratica. A tal fine puo' essere utile ricordare il seguente risultato.

Teorema. *Regola di Cartesio.*
Sia $p(x) = \sum_{i=0}^{s} a_i x^i \in \mathbb{R}[x]$ un polinomio con tutte le radici reali. Allora le radici positive di $p(x)$ sono tante quanti i cambiamenti adiacenti di segno (variazioni) nella successione $a_0, a_1, \ldots, a_s$ dei coefficienti.