G. Romani

November 18, 2005

MATRICI E SISTEMI DI EQUAZIONI LINEARI

1. Preliminari.

Consideriamo l'equazione nell'*incognita* x, *coefficiente* a e *termine noto* b:

$$ax = b$$

Moltiplicando ambo i membri dell'equazione per l'inverso di $a : a^{-1}$, si ottiene $x = a^{-1}b$, che si verifica subito essere una soluzione dell'equazione. Evidentemente non vi possono essere soluzioni diverse da questa. Dunque l'equazione ammette una ed una sola soluzione: $x = a^{-1}b$. Questo però se esiste l'inverso di a, quindi per $a \neq 0$. Se $a = 0$ si possono presentare due casi: o $b = 0$ ed in tal caso ogni numero è soluzione dell'equazione, oppure $b \neq 0$ ed in tal caso nessun numero è soluzione. Riassumendo, in relazione alle soluzioni di un'equazione del tipo $ax = b$ si può presentare solo uno dei seguenti casi

i) $a \neq 0$.

La soluzione esiste, è unica ed è data da

$$x = a^{-1}b$$

ii) $a = 0, b \neq 0$

La soluzione non esiste

iii) $a, b = 0$

Le soluzioni sono infinite.

Vedremo come questa situazione si ripeterà per ogni sistema di equazioni lineari, qualunque sia il suo numero di incognite e qualunque sia il suo numero di equazioni: un sistema o ammetterà una ed una sola soluzione, tali *sistemi* verranno detti *compatibili e determinati*, o non ammetterà soluzioni, *sistemi incompatibili*, o ammetterà infinite soluzioni, *sistemi compatibili ed indeterminati.*

Consideriamo un altro caso particolare. Un'equazione lineare in due incognite

$$ax + by = c$$

Segnaliamo subito che nel seguito allorché si parlerà di *soluzione* non si intenderà un numero, ma *una coppia ordinata di numeri che sostituita rispettivamente al posto della x ed al posto della y verifica entrambe le equazioni.*

Ricavando la x da questa equazione si trova: $x = -a^{-1}by + a^{-1}c$ e si verifica subito che ogni coppia del tipo $(-a^{-1}by + a^{-1}c, y)$, al variare comunque di y, è soluzione dell'equazione. Le soluzioni dell'equazione data sono dunque infinite. Ma questo se a ammette inverso cioè se $a \neq 0$. Altrimenti si può seguire lo stesso ragionamento ricavando invece della x, la y. Si troveranno come soluzioni le coppie $(x, -b^{-1}ax + b^{-1}c)$ al variare comunque della x. Di nuovo le soluzioni sono infinite. Se tanto a quanto b sono diversi da zero, si possono considerare tanto le coppie ottenute ricavando la x quanto quelle ottenute ricavando la y, ma si potrebbe verificare facilmente che i due insiemi di coppie coincidono: una coppia ottenuta ricavando la x in corrispondenza ad un determinato valore di y, si può ottenere da quelle ottenute ricavando la y per un opportuno valore di x. Ma cosa succede se tanto a quanto b sono nulli? Anche ora sono possibili due casi: o anche c è nullo, ed in tal caso tutte le coppie ordinate di numeri reali soddisfano l'equazione: le soluzioni continuano ad essere infinite, oppure $c \neq 0$ ed in tal caso l'equazione non ammette soluzioni. Riassumendo possiamo dire che per l'equazione $ax + by = c$ si presentano due casi

i) $(a, b) \neq (0, 0)$ oppure $(a, b) = (0, 0)$, $c = 0$

Le soluzioni sono infinite e l'equazione è *compatibile ed indeterminata.*

ii) $(a, b) = (0, 0)$, $c \neq 0$

Non esistono soluzioni: l'equazione è *incompatibile.*

Facciamo come ultimo caso quello di un sistema di due equazioni in due incognite

$$\begin{cases} ax + by = c \\ a^{'}x + b^{'}y = c^{'} \end{cases}$$

Per un tale sistema una soluzione sarà una coppia ordinate di numeri che soddisfa entrambe le equazioni. Ricavando la x dalla prima equazione si trova $x = -a^{-1}by + a^{-1}c$, sostituendola nella seconda si ottiene

$$a^{'}(-a^{-1}by + a^{-1}c) + b^{'}y = c^{'}$$

che è un'equazione nella sola y ricavando la y da questa seconda equazione si trova

$$y = \frac{ac^{'} - a^{'}c}{ab' - a'b}$$

e sostituendo il valore della y così trovato nella precedente espressione della x si ottiene per la x il valore

$$x = \frac{cb^{'} - c^{'}b}{ab' - a'b}$$

In conclusione si è trovata la coppia ordinata

$$\left(\frac{cb^{'}-c^{'}b}{ab^{'}-a^{'}b}, \frac{ac^{'}-a^{'}c}{ab^{'}-a^{'}b}\right)$$

che si verifica essere una soluzione con una verifica diretta: sostituendola nelle due equazioni. Il ragionamento fatto è stato basato sull'ipotesi che sia possibile ricavare la x dalla prima equazione, ovvero sull'ipotesi che a sia diverso da zero. Si potrebbe però vedere che per poter essere fatto il ragionamento ha bisogno solo dell'ipotesi che sia $ab^{'}-a^{'}b \neq 0$. Infatti se questa seconda ipotesi è soddisfatta e, ad esempio $a = 0$, il ragionamento potrebbe essere fatto di nuovo o ricavando sempre la x, però dalla seconda equazione, oppure la y dalla prima equazione, se possibile, o dalla seconda. Inoltre si verificherebbe che in ognuno di questi casi, sempre ferma restando l'ipotesi che sia $ab^{'}-a^{'}b \neq 0$, come soluzione si troverebbe sempre la stessa coppia. Si può pertanto asserire che il sistema dato, di due equazioni in due incognite, ammette una ed una soluzione, ovvero è compatibile e determinato, se $ab^{'}-a^{'}b \neq 0$. Cosa succede se $ab^{'}-a^{'}b = 0$? Si hanno due possibilità. O anche $ac^{'}-a^{'}c = 0$ ed in tal caso si potrebbe vedere che il sistema ammette infinite soluzioni, ad esempio se $a \neq 0$ oppure $b \neq 0$, tutte le soluzioni della prima equazione sono soluzioni della seconda, si tratterebbe di un sistema compatibile ma indeterminato, oppure $ac^{'}-a^{'}c \neq 0$. In questo caso le soluzioni non esistono. Il sistema è incompatibile. Riassumendo

i) $ab^{'}-a^{'}b \neq 0$

Il sistema è *compatibile e determinato*

ii) $ab^{'}-a^{'}b$, $ac^{'}-a^{'}c = 0$

Le soluzioni sono infinite. *Sistema compatibile ed indeterminato*

iii) $ab^{'}-a^{'}b = 0$, *ed inoltre o* $ac^{'}-a^{'}c \neq 0$

La soluzione non esiste: *il sistema è incompatibile.*

Tutti i casi illustrati possono essere riassunti, così come tutti i casi possibili di sistemi lineari, in un teorema noto come *Teorema di Rouché-Capelli.* Per poter enunciare ed illustrare tale teorema, (Teorema che non dimostreremo), avremo bisogno delle nozioni di combinazioni lineari, di matrice, di determinante di una matrice, di rango di una matrice ed infine del Teorema (ovvero Regola) di Cramer. Nei paragrafi seguenti ci accingeremo a dare le suddette nozioni e le proprietà ad esse collegate.

2. Combinazioni lineari e dipendenza lineare. Rango.

Partiamo dalla considerazione di due equazioni lineari in due incognite: $ax+by=c$, $a^{'}x+b^{'}y=c^{'}$. Moltiplichiamo la prima equazione per un numero,

λ, e la seconda equazione per un altro numero, μ. Sommiamo poi le equazioni così ottenute. Troviamo ancora un'equazione lineare, la seguente:

$$(\lambda a + \mu a^{'})x + (\lambda b + \mu b^{'})y = \lambda c + \mu c^{'}$$

tale equazione viene detta *combinazione delle equazioni lineari assegnate con coefficienti λ e μ.*

Consideriamo ora il sistema dato dalle due equazioni lineari, $ax + by = c$, $a^{'}x + b^{'}y = c^{'}$ ovvero il sistema

$$\begin{cases} ax + by = c \\ a^{'}x + b^{'}y = c^{'} \end{cases}$$

a tale sistema aggiungiamo una qualunque combinazione lineare delle due equazioni e consideriamo il sistema sempre in due incognite ma di tre equazioni

$$\begin{cases} ax + by = c \\ a^{'}x + b^{'}y = c^{'} \\ (\lambda a + \mu a^{'})x + (\lambda b + \mu b^{'})y = \lambda c + \mu c^{'} \end{cases}$$

Se la coppia (x_0, y_0) è una soluzione del sistema delle sole prime due equazioni, ovvero se risulta

$$\begin{cases} ax_0 + by_0 = c \\ a^{'}x_0 + b^{'}y_0 = c^{'} \end{cases}$$

si controlla subito che la stessa coppia è anche soluzione della terza equazione, ovvero che risulta

$$(\lambda a + \mu a^{'})x_0 + (\lambda b + \mu b^{'})y_0 = \lambda c + \mu c^{'}$$

e quindi la coppia (x_0, y_0) è soluzione del sistema di tre equazioni. Naturalmente viceversa, se una coppia risolve il sistema costituito dalle tre equazioni, in particolare deve essere soluzione del sistema costituito dalle sole prime due. Dunque il sistema costituito dalle due equazioni ed uno qualunque dei sistemi di tre equazioni ottenuti aggiungendo alle due equazioni una loro combinazione lineare, ammettono sempre le stesse soluzioni: sono equivalenti. *Equivalenti sono sistemi le cui soluzioni coincidono.*

Date tre equazioni in due incognite: $ax+by = c$, $a^{'}x+b^{'}y = c^{'}$, $a^{''}x+b^{''}y = c^{''}$ è possibile stabilire se una di queste tre equazioni è combinazione lineare delle altre due? La risposta è sì. Lo si fa considerando i coefficienti delle incognite, nelle tre diverse equazioni, ed i termini noti. Per illustrare il metodo avremo bisogno di alcune nozioni. Per semplicità espositiva tratteremo il caso delle terne di numeri, ma il caso delle coppie di numeri si dedurrà immediatamente dimenticando la presenza del terzo numero ed il caso di quaterne o di più numeri, delle cosidette $n - ple$, si capirà immaginando di avere dopo il terzo numero altri numeri ancora.

Una terna ordinata di numeri è una terna di numeri di cui sia specificato quale sia il primo, quale il secondo, quale il terzo. Indicheremo una terna ordinata di numeri con (a, b, c) *o con simboli similari.*

Nel seguito, con un abuso di linguaggio, per terna intenderemo una terna ordinata.

Prodotto di una terna per uno scalare:

Data una terna (a, b, c) *ed un numero* λ, *detto in questo contesto scalare, si definisce prodotto della terna* (a, b, c) *per lo scalare* λ, *la terna* $(\lambda a, \lambda b, \lambda c)$. *In simboli*

$$\lambda(a, b, c) = (\lambda a, \lambda b, \lambda c)$$

Somma di due terne:

Date due terne (a, b, c), $(a^{'}, b^{'}, c^{'})$, *si definisce loro somma la terna* $(a + a^{'}, b + b^{'}, c + c^{'})$, *in simboli*

$$(a, b, c) + (a^{'}, b^{'}, c^{'}) = (a + a^{'}, b + b^{'}, c + c^{'})$$

Combinazioni lineari di due terne:

Date due terne (a, b, c), $(a^{'}, b^{'}, c^{'})$ *e due scalari,* λ, μ, *si definisce combinazione lineare delle due terne con coefficienti* λ, μ, *la terna* $(\lambda a + \mu a^{'}, \lambda b + \mu b^{'}, \lambda c + \mu c^{'})$, *in simboli*

$$\lambda(a, b, c) + \mu(a^{'}, b^{'}, c^{'}) = (\lambda a + \mu a^{'}, \lambda b + \mu b^{'}, \lambda c + \mu c^{'})$$

Dipendenza lineare di tre terne:

Tre terne sono dette linearmente dipendenti se una di esse si può esprimere come combinazione lineare delle altre due.

Sulla base delle definizioni appena date è immediato verificare la proposizione seguente:

Date due equazioni lineari $ax + by = c$, $a^{'}x + b^{'}y = c^{'}$, *la loro combinazione lineare con coefficienti* $\lambda, \mu : (\lambda a + \mu a^{'})x + (\lambda b + \mu b^{'})y = \lambda c + \mu c^{'}$, *ha come terna dei coefficienti delle incognite e termine noto, la combinazione delle terne dei coefficienti delle incognite e termini noti delle due equazioni. In formule risulta*

$$(\lambda a + \mu a^{'}, \lambda b + \mu b^{'}, \lambda c + \mu c^{'}) = \lambda(a, b, c) + \mu(a^{'}, b^{'}, c^{'})$$

In tal modo il problema di vedere quando di tre equazioni lineari una è combinazione lineare delle altre,è ricondotto a quello di vedere quando, date tre terne, le tre terne sono linearmente dipendenti. Tale dipendenza lineare si studia facendo uso della nozione di determinante e di rango di una matrice.

Entrambe, come vedremo, sono nozioni calcolabili ed a tale scopo elenchiamo qui appresso tutta una serie di proprietà delle due operazioni introdotte: della somma e del prodotto tra terne utili per effettuare i calcoli.

$(\lambda\mu)(a,b,c) = \lambda(\mu(a,b,c))$ *associatività rispetto al prodotto*
$(\lambda+\mu)(a,b,c) = \lambda(a,b,c) + \mu(a,b,c)$ *distributività del prodotto rispetto alla somma*
$1(a,b,c) = (a,b,c)$ *elemento neutro rispetto al prodotto*
$((a,b,c) + (a^{'},b^{'},c^{'})) + (a^{''},b^{''},c^{''}) = (a,b,c) + ((a^{'},b^{'},c^{'}) + (a^{''},b^{''},c^{''}))$ *associatività della somma*
$(a,b,c) + (a^{'},b^{'},c^{'}) = (a^{'},b^{'},c^{'}) + (a,b,c)$ *commutatività della somma*
$(a,b,c) + (0,0,0) = (a,b,c)$ *elemento neutro rispetto alla somma*
$(a,b,c) + (-a,-b,-c) = (0,0,0)$ *opposto nella somma*
$\lambda((a,b,c) + (a^{'},b^{'},c^{'})) = \lambda(a,b,c) + \lambda(a^{'},b^{'},c^{'})$ *distributività del prodotto rispetto alla somma*

Oltre alle proprietà suddette valgono anche le seguenti, anche esse utili nei calcoli

$$(-1)(a,b,c) = (-a,-b,-c)$$
$$0(a,b,c) = (0,0,0)$$
$$\lambda(0,0,0) = (0,0,0)$$

Il fatto che nell'insieme delle terne sia definita un'operazione di prodotto per scalari ed una operazione di somma tra terne, verificanti le prime 8 proprietà, si esprime dicendo che l'insieme delle terne costituisce uno spazio vettoriale sui reali.

Le prime 8 proprietà si possono verificare direttamente a partire dalla definizione stessa delle due operazioni. Anche le altre due proprietà si possono dimostrare a partire dalle definizioni di prodotto di una terna per uno scalare e di somma tra terne, ma possono anche dedursi dalle prime 8 proprietà in maniera indipendente dalla maniera in cui le due operazioni sono state definite. Per questo sono state indicate separatamente.

Per quanto riguarda le combinazioni lineari di due terne, (a,b,c), $(a^{'},b^{'},c^{'})$, è bene farsi un'idea di quali esse siano. Innanzitutto tra le combinazioni lineari, vi sono le terne stesse: la terna (a,b,c), che si può ottenere prendendo come coefficienti 1 e 0, la terna $(a^{'},b^{'},c^{'})$, che si può ottenere prendendo come coefficienti 0 ed 1. Inoltre tutte le terne del tipo $\lambda(a,b,c) = (\lambda a, \lambda b, \lambda c)$, le cosidette *terne proporzionali alla terna* (a,b,c), sono combinazioni lineari di tutte e due le terne: con coefficienti λ arbitrario, e $\mu = 0$. Analogamente tra le combinazioni lineari delle due terne, vi sono le terne proporzionali alla seconda terna, le terne del tipo $(\mu a, \mu b, \mu c)$: tali terne si ottengono prendendo $\lambda = 0$ e μ arbitrario. Infine la somma delle due terne, la differenza della prima meno la seconda, la differenza della seconda meno la prima, sono tutte combinazioni lineari ottenute prendendo per coefficienti rispettivamente, $(1,1)$, $(1,-1)$, $(-1,1)$.

In relazione alla dipendenza lineare, valgono le seguenti proprietà:

Tre terne sono linearmente dipendenti se e solo se si può trovare una combinazione lineare di tutte e tre le terne, i cui coefficienti non siano tutti nulli, uguale alla terna nulla.

Nell'enunciato precedente dovrebbe essere chiaro cosa debba intendersi per combinazione lineare di tre terne. Lo precisiamo comunque:

Date tre terne: (a,b,c), $(a^{'},b^{'},c^{'})$, $(a^{''},b^{''},c^{''})$,*ogni terna del tipo*

$$\lambda(a,b,c)+\mu(a^{'},b^{'},c^{'})+\nu(a^{''},b^{''},c^{''}) \qquad \lambda,\mu,\nu\in R$$

è detta combinazione lineare delle tre terne.

In seguito non preciseremo cosa debba intendersi per combinazione di un qualunque numero di terne, perché oramai il significato dovrebbe essere evidente.

Dimostriamo la proposizione precedente.

Se tre terne, (a,b,c), $(a^{'},b^{'},c^{'})$, $(a^{''},b^{''},c^{''})$, sono dipendenti, in base alla definizione, una delle tre è combinazione delle altre due. Poniamo che sia

$$\lambda(a,b,c)+\mu(a^{'},b^{'},c^{'})=(a^{''},b^{''},c^{''})$$

Ma allora sommando ad ambo i membri l'opposta della terna $(a^{''},b^{''},c^{''})$ e, successivamente scrivendo tale opposta come prodotto della terna $(a^{''},b^{''},c^{''})$ per -1, si ottiene

$$\lambda(a,b,c)+\mu(a^{'},b^{'},c^{'})+(-1)(a^{''},b^{''},c^{''})=(0,0,0)$$

che mostra come se le tre terne sono dipendenti esiste una combinazione lineare delle tre terne, con coefficienti non tutti nulli, uguale alla terna nulla.

Viceversa se esiste una combinazione lineare delle tre terne con coefficienti non tutti nulli, uguale alla terna nulla, poniamo che sia

$$\lambda(a,b,c)+\mu(a^{'},b^{'},c^{'})+\nu(a^{''},b^{''},c^{''})=(0,0,0) \qquad \nu\neq 0$$

con un ovvio uso delle regole sulla somma tra terne e sul prodotto tra una terna ed uno scalare, si ottiene

$$(-1)\frac{\lambda}{\nu}(a,b,c)+(-!)\frac{\mu}{\nu}(a^{'},b^{'},c^{'})=(a^{''},b^{''},c^{''})$$

ovvero la terza terna è combinazione delle prime due: le tre terne sono dipendenti.

Altra proprietà importante della dipendenza lineare è la seguente.

Se un certo numero di terne sono linearmente dipendenti, ovvero una tra le terne cosiderate è combinazione delle altre, comunque si aggiungano terne alle terne iniziali, si trova sempre un insieme di terne dipendenti.

La dimostrazione di tale fatto è immediata: se di certe terne una è combinazione lineare delle altre, tale terna sarà anche combinazione lineare delle terne considerate e di altre terne, quelunque siano le terne aggiunte, basterà prendere come coefficienti delle nuove terne sempre lo zero.

Conseguenze immediate della proposizione appena dimostrata sono le due seguenti proposizioni:

Se tra certe terne vi è la terna nulla, le terne sono dipendenti

Se tra certe terne ve ne sono due proporzionali, le terne sono dipendenti.

La prima proposizione si dimostra osservando che la terna nulla è combinazione lineare di qualunque insieme di terne con coefficienti tutti nulli, la seconda proposizione si dimostra osservando che se una terna è proporzionale ad un'altra, si otterrà moltiplicando la seconda per il fattore di proporzionalità delle due terne, e quindi sarà combinazione delle terne ottenute aggiungendo alla seconda altre terne, prendendo come coefficiente della seconda il fattore di proporzionalità che le lega e tutti gli altri coefficienti nulli.

E' giunto il momento di introdurre la nozione di rango di un certo sistema di terne.

Dato un insieme di terne, si chiama rango del sistema di terne assegnato, il massimo numero di terne indipendenti che possono essere trovate tra le terne assegnate.

Da ora in poi, per ottenere degli enunciati più omogenei, ovvero nei quali non si debbano fare troppi casi e sottocasi, converremo di dare le due definizioni seguenti:

La terna nulla è linearmente dipendente

Due terne sono dipendenti se sono proporzionali.

Inoltre converremo di dire che

Un sistema di terne ha rango zero se tutte le terne date coincidono con la terna nulla.

Nel caso che anche una sola delle terne di un dato sistema di terne sia diversa dalla terna nulla, il rango del sistema sarà diverso da zero.

Usando le definizioni appena introdotte, possiamo dire che

Il rango di un sistema di terne è uno se comunque prendo due terne del sistema queste sono proporzionali, è due se tra le terne date ve ne sono due non proporzionali (indipendenti), ma ogni altra terna è loro combinazione lineare, è tre se si possono trovare tre terne di cui nessuna sia combinazione lineare delle altre due.

Tra breve vedremo come

Qualunque sia il sistema di terne assegnato il suo rango non è mai maggiore di tre.

Per ora invitiamo lo studente ad estendere tutte le definizioni, proprietà, teoremi e proposizioni del presente paragrafo, eccezione fatta per l'ultima proposizione appena enunciata, al caso delle cosidette $n-ple$: insiemi di un numero arbitrario, ma fissato, n, di numeri, assegnati un un dato ordine.

L'ultima proposizione, come vedremo in seguito, nel caso delle n-ple andrà sostituita con la seguente:

Più di n $n-ple$ sono sempre linearmente dipendenti.

3. Matrici. Determinante. Regola di Cramer.

Una tabella (rettangolare) di numeri con un dato numero di righe ed un dato numero di colonne, viene detta matrice

Se il numero di righe è uguale al numero delle colonne, la matrice viene detta quadrata.

Alcuni esempi chiariranno la definizione. Cominciamo dalla seguente tabella di numeri:

$$\begin{pmatrix} 1 & \pi \\ e & -2 \end{pmatrix}$$

si tratta di una matrice quadrata a due rige e due colonne. Gli elementi della prima riga sono dati dalla coppia $(1, \pi)$, gli elementi della seconda riga sono dati dalla coppia $(e, -2)$. Gli elementi della prima colonna sono dati dalla coppia $(1, e)$ mentre quelli della seconda colonna sono $(\pi, -2)$.

Altro esempio è dato dalla tabella

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & -2 \end{pmatrix}$$

Si tratta di una matrice rettangolare: a tre righe e due colonne. Le righe sono fatte da coppie, le colonne da terne. Le righe sono rispettivamente date dalla coppie $(1, 1)$, $(1, 1)$, $(1, -2)$, le colonne dalle terne $(1, 1, 1)$, $(1, 1, -2)$.

La tabelle che segue

$$\begin{pmatrix} 0 & 1 & 4 \\ 2 & 1+\pi & 1 \end{pmatrix}$$

è una matrice rettangolare a due righe e tre colonne. Le righe sono terne, le colonne coppie.

Anche una tabella del tipo

$$\begin{pmatrix} 1 & 2 & 6 & 8 \end{pmatrix}$$

è una matrice: ad una riga e quattro colonne, così come la tabella

$$\begin{pmatrix} 1 \\ 3 \\ -e \\ -9 \end{pmatrix}$$

è una matrice ad una colonna e quattro righe. Le matrici fatte con una sola colonna sono dette *matrici colonna*, quelle fatte con una sola riga sono dette *matrici riga*. In generale una matrice verrà indicata con una lettera maiuscola, come ad esempio la lettera A, alla quale si apporranno due indici di cui il primo indicherà il numero delle righe ed il secondo quello delle colonne. se il numero di righe di una matrice è indicato con p, ed il numero delle colonne con q, la matrice sarà indicata con $A_{(p,q)}$. Se la matrice è quadrata, $p = q$, la si indicherà semplicemente con $A_{(p)}$. Il comune numero di righe e di colonne di una matrice quadrata sarà detto *ordine della matrice quadrata*.

Una prima operazione fondamentale che si definisce per le matrici, è il passaggio alla trasposta.

Data una matrice si chiama trasposta della matrice la matrice che ha per colonne le righe delle matrice data e per righe le colonne della matrice data.

Se $A_{(p,q)}$ *è una matrice a* p *righe e* q *colonne, la sua trasposta la si indicherà con* $A^t_{(p,q)}$. *Sarà una matrice a* q *righe e* p *colonne. In particolare se una matrice è quadrata di ordine* n *la sua trasposta sarà ancora una matrice quadrata di ordine* n.

Se $A_{(n)}$ *è una matrice quadrata tale matrice viene detta simmetrica se coincide con la sua trasposta*: $A^t_{(n)} = A_{(n)}$.

L'opposta di una matrice è la matrice che ha gli stessi elementi della matrice data, però ognuno cambiato di segno (questo anche se la matrice non è quadrata). Se la matrice viene indicata con A *l'opposta la si indicherà con* $-A$.

Una matrice quadrata è detta antisimmetrica, se coincide con l'opposta della sua trasposta: $A^t_{(n)} = -A_{(n)}$.

Per le matrici quadrate (*e solo per loro*) si introduce la nozione di *determinante.*

Diamo la nozione di determinante per le matrici quadrate di ordine due. In seguito la estenderemo ad una matrice quadrata di ordine arbitrario.

Sia

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

una matrice quadrata di ordine due. Definiamo suo determinante il numero

$$\det A = ad - cb$$

Indicheremo il determinante di A anche con $|A|$.

Passiamo alla definizione di determinante di una matrice quadrata di ordine tre. Questa è data dall'uguaglianza che segue:

$$\det\begin{pmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{pmatrix} = a\begin{vmatrix} b' & c' \\ b'' & c'' \end{vmatrix} - b\begin{vmatrix} a' & c' \\ a'' & c'' \end{vmatrix} + c\begin{vmatrix} a' & b' \\ a'' & b'' \end{vmatrix}$$

a parole

Il determinante di una matrice quadrata di ordine tre è dato dalla somma, a segni alterni (cominciando dal segno +), dei prodotti degli elementi della prima riga della matrice, per i determinanti delle matrici di ordine due ottenute dalla matrice data, cancellando la prima riga e la colonna alla quale appartiene l'elemento della prima riga che si sta moltiplicando.

Questa definizione permette una definizione, come si dice, per induzione, di determinannte. A tale scopo dovremo abituarci a cambiare simboli. Indichiamo con n l'ordine di una matrice quadrata e la matrice quadrata di ordine n con A_n. Porremo

$$A_n = \begin{pmatrix} a_1^1 & a_2^1 & . & . & a_n^1 \\ a_1^2 & a_2^2 & . & . & a_n^2 \\ . & . & & & . \\ . & . & & & . \\ a_1^n & a_2^n & . & . & a_n^n \end{pmatrix}$$

Il generico elemento della matrice sarà indicato con a_j^i. Si osservi come l'indice in alto indichi a quale riga appartiene l'elemento, mentre l'indice in basso indichi a quale colonna appartenga l'elemento: l'indice alto non varia quando ci si sposta su di una riga, quello basso non varia quando ci si sposta su di una colonna. La matrice verrà anche indicata con $A_n = \left((a_j^i)\right)$. Ciò posto definiamo per una qualunque matrice quadrata il determinante tramite la formula seguente

$$\det A_n = a_1^1 A^1 - a_2^1 A^2 + ... + (-1)^{1+n} a_n^1 A^n$$

dove con A^i si indica il determinante della matrice quadrata di ordine $n-1$ che si ottiene dalla matrice data cancellando la prima riga e la colonna alla quale appartiene l'elemento che si sta moltiplicando.

Facendo uso della definizione di determinante, il Teorema di Cramer precedentemente enunciato per sistemi di due equazioni in due incognite, assume la seguente formulazione.

Dato il sistema

$$\begin{cases} ax + by = c \\ a'x + b'y = c' \end{cases}$$

questo è compatibile e determinato se e solo se

$$\begin{vmatrix} a & b \\ a' & b' \end{vmatrix} \neq 0$$

In tal caso la soluzione è data dalla coppia

$$x = \frac{\begin{vmatrix} c & b \\ c' & b' \end{vmatrix}}{\begin{vmatrix} a & b \\ a' & b' \end{vmatrix}} \quad y = \frac{\begin{vmatrix} a & c \\ a' & c' \end{vmatrix}}{\begin{vmatrix} a & b \\ a' & b' \end{vmatrix}}$$

In relazione al sistema considerato, le matrici

$$\begin{vmatrix} a & b \\ a' & b' \end{vmatrix}, \begin{vmatrix} a & b & c \\ a' & b' & c' \end{vmatrix}$$

vengono rispettivamente dette, *matrice dei coefficienti e matrice dei coefficenti e termini noti, mentre la matrice colonna*

$$\begin{pmatrix} c \\ c' \end{pmatrix}$$

viene detta *colonna dei termini noti.* Con l'ulteriore uso di questa terminologia si arriva al seguente enunciato per il Teorema di Cramer per sistemi di due equazioni in due incognite:

Il sistema è compatibile e determinato se e solo se il determinante della matrice dei coefficienti è diverso da zero ed in tal caso la soluzione è data dalla coppia

$$x_1 = \frac{A_1}{A},\ x_2 = \frac{A_2}{A}$$

dove con A si indica il determinante della matrice dei coefficienti e con A_i $(i = 1, 2)$, il determinante della matrice ottenuta da quella dei coefficienti sostituendo alla $i - ma$ colonna, la colonna dei termini noti.

E' quest'ultimo enunciato che viene generalizzato al caso di un qualunque sistema quadrato. Precisamente dato un sistema di n equazioni in n incognite:

$$\begin{cases} a_1^1 x^1 + a_2^1 x^2 + ... + a_n^1 x^n = b^1 \\ a_1^2 x^1 + a_2^2 x^2 + ... + a_n^2 x^n = b^2 \\ ... \\ ... \\ a_1^n x^1 + a_2^n x^2 + ... + a_n^n x^n = b^n \end{cases}$$

detta *matrice dei coefficienti* la matrice

$$\begin{pmatrix} a_1^1 & a_2^1 & . & . & a_n^1 \\ a_1^2 & a_2^2 & . & . & a_n^2 \\ . & . & & & . \\ . & . & & & . \\ a_1^n & a_2^n & . & . & a_n^n \end{pmatrix}$$

e *colonna dei termini noti*, la matrice colonna

$$\begin{pmatrix} b^1 \\ b^2 \\ . \\ . \\ b^n \end{pmatrix}$$

il teorema di Cramer nel caso generale si enuncia dicendo che

Un sistema quadrato

$$\begin{cases} a_1^1 x^1 + a_2^1 x^2 + ... + a_n^1 x^n = b^1 \\ a_1^2 x^1 + a_2^2 x^2 + ... + a_n^2 x^n = b^2 \\ ... \\ ... \\ a_1^n x^1 + a_2^n x^2 + ... + a_n^n x^n = b^n \end{cases}$$

è compatibile e determinato se e solo se il determinante della matrice dei coefficienti è diverso da zero, ed in tal caso la soluzione è data dall'$n-pla$

$$x_1 = \frac{A_1}{A},\ x_2 = \frac{A_2}{A}, ..., x_n = \frac{A_n}{A}$$

dove con A *si indica il determinante della matrice dei coefficienti e con* $A_i, (1, ..., n)$, *il determinante della matrice ottenuta da quella dei coefficienti sostituendo alla* $i-$*ma colonna, la colonna dei termini noti.*(La dimostrazione del Teorema, viene omessa)

Già l'enunciato del teorema di Cramer, mostra l'interesse del calcolo del determinante di una matrice quadrata, calcolo che va ben oltre il caso specifico del teorema di Cramer. Per questo, in quello che segue, indicheremo una serie di proprietà dei determinanti utili al loro calcolo.

Il determinate di una matrice è nullo se e solo se una riga (oppure una colonna) è combinazione lineare delle altre righe (rispettivamente delle altre colonne).

Naturalmente il risultato può essere espresso equivalentemente dicendo che

Il determinante di una matrice è nullo se e solo se le righe (ovvero le colonne), sono linearmente dipendenti.

Osserviamo come in questo risultato vi sia una risposta parziale al problema sollevato in precedenza di vedere quando certe $n-ple$, sono dipendenti. Infatti nel caso di n, $n-ple$, la risposta è

n, n-ple sono dipendenti se e solo se il determinante della matrice quadrata di ordine n che ha le n n-ple come righe, è uguale a zero.

Il problema di stabilire quando $p, n-ple$, con $p \neq n$, sono dipendenti, rimane per ora aperto. Vedremo in seguito come tale problema trovi la sua soluzione nella nozione di rango di un sistema di $n-ple$, e di rango di una matrice (questa volta, non necessariamente quadrata).

Continuiamo con le proprietà dei determinanti. Innanzitutto, osservando che l'$n-pla$ nulla è linearmente dipendente con ogni altra $n-pla$, e che due $n-ple$ proporzionali, sono linearmente dipendenti, dalla proposizione enunciata, e dalla proprità a suo tempo segnalata che se certe $n-ple$ sono dipendenti sono dipendenti tutti gli insiemi che si ottengono aggiungendo $n-ple$ alle $n-ple$ dipendenti, segue che

Una matrice quadrata con una riga (od una colonna) nulla, così come una matrice quadrata con due righe (o due colonne), proporzionali, ha determinante zero.

Inoltre

Una matrice ottenuta da una data effettuando un numero pari di scambi tra le righe (le colonne) della matrice data, ha lo stesso determinante della matrice assegnata. Mentre se una matrice è ottenuta da un'altra effettuando un numero dispari di scambi tra le righe (ovvero tra le colonne), il determinante ha segno opposto al determinante dell'altra matrice.

Infine diamo una regola molto utile per il calcolo del determinante che dà la possibilità di calcolare il determinante di una matrice a partire anziché necessariamente dalla prima riga, da una riga, oppure anche, da una colonna qualunque. Premettiamo la nozione di complemento algebrico di un fissato elemento di una matrice quadrata. Partiamo da una matrice quadrata di ordine n: $A_{(n)} = ((a^i_j))$. Indichiamo con a^h_k un elemento fissato della matrice. Indichiamo con A^h_k il determinante della matrice (di ordine $n-1$) ottenuta da A cancellando la riga e la colonna alle quali appartiene a^h_k ovvero cancellando la riga $h-ma$, e la colonna $k-ma$.

Si chiama complemento algebrico dell'elemento a^h_k nella matrice A, e lo si indica con $\mathcal{A}^h_k$, il numero

$$\mathcal{A}^k_h = (-1)^{h+k} A^h_k$$

(si noti lo scambio della posizione degli indici che viene fatta per comodità di calcolo che appariranno in seguito). Ciò posto si hanno i due seguenti enunciati

(Sviluppo del determinante secondo una riga qualunque) Fissato i, si ha

$$A = \det A_{(n)} = \sum_{j=1}^{n} a_j^i \mathcal{A}_i^j$$

(*Sviluppo del determinante secondo una colonna qualunque) Fissato j, si ha*

$$A = \det A_{(n)} = \sum_{i=1}^{n} a_j^i \mathcal{A}_i^j$$

Da questa regola discende facilmente la seguente proprietà, che ci limitiamo ad enunciare nel caso delle righe, ma che vale tale e quale nel caso delle colonne. Consideriamo due $n - ple$: $(a_1, a_2, ..., a_n), (b_1, b_2, ..., b_n)$. Sia $(\alpha a_1 + \beta b_1, \alpha a_2 + \beta b_2, ..., \alpha a_n + \beta b_n)$, una loro combinazione lineare. Indichiamo con

$$\begin{pmatrix} . & . & . & . & . \\ . & . & . & . & . \\ \alpha a_1 + \beta b_1 & \alpha a_2 + \beta b_2 & . & . & \alpha a_n + \beta b_n \\ . & . & . & . & . \\ . & . & . & . & . \end{pmatrix}$$

una matrice quadrata di ordine n in cui una certa fissata riga, diciamola la $i - ma$, sia data dalla combinazione lineare suddetta. Indichiamo invece con

$$\begin{pmatrix} . & . & . & . & . \\ . & . & . & . & . \\ a_1 & a_2 & . & . & a_n \\ . & . & . & . & . \\ . & . & . & . & . \end{pmatrix}, \quad \begin{pmatrix} . & . & . & . & . \\ . & . & . & . & . \\ b_1 & b_2 & . & . & b_n \\ . & . & . & . & . \\ . & . & . & . & . \end{pmatrix}$$

le matrici quadrate di ordine n in cui la $i - ma$ riga sia data, rispettivamente, dalla prima $n - pla$, e dalla seconda $n - pla$.

Risulta

$$\det \begin{pmatrix} . & . & . & . & . \\ . & . & . & . & . \\ \alpha a_1 + \beta b_1 & \alpha a_2 + \beta b_2 & . & . & \alpha a_n + \beta b_n \\ . & . & . & . & . \\ . & . & . & . & . \end{pmatrix} =$$

$$= \alpha \det \begin{pmatrix} . & . & . & . & . \\ . & . & . & . & . \\ a_1 & a_2 & . & . & a_n \\ . & . & . & . & . \\ . & . & . & . & . \end{pmatrix} + \beta \det \begin{pmatrix} . & . & . & . & . \\ . & . & . & . & . \\ b_1 & b_2 & . & . & b_n \\ . & . & . & . & . \\ . & . & . & . & . \end{pmatrix}$$

Da questo risultato e dai risultati precedenti, discende senza difficoltà che,

Se una matrice è ottenuta da una data matrice sostituendo una riga (colonna) con la somma della riga e di una combinazione lineare di altre righe (colonne), le due matrici hanno lo stesso determinante.

In alcuni casi di matrici particolari, che appaiono con una certa frequenza, il determinante si calcola facilmente. Nel seguito sarà bene chiarire che

La diagonale principale di una matrice quadrata è la diagonale fatta dagli elementi del tipo a_h^h.

Inoltre

Una matrice è detta triangolare se tutti gli elementi della matrice che si trovano sotto (sopra) la diagonale principale sono nulli.

In relazione al determinante di matrici triangolari si ha il seguente teorema.

Il determinante di una matrice triangolare è dato dal prodotto degli elementi che si trovano sulla diagonale principale.

Ricordiamo le definizioni di matrici simmetriche ed antisimmetriche:

Una matrice A *è detta simmetrica se* $A^t = A$. *Una matrice* A *è detta antisimmetrica se* $A^t = -A$.

per le matrici antisimmetriche di ordine dispari si ha il seguente risultato:

Una matrice antisimmetrica di ordine dispari ha determinante uguale a zero.

Si è detto che le regole esposte sono utili al calcolo del determinante di una matrice. Diamo una breve illustrazione di tale fatto.

Per semplificare l'esposizione limitiamoci al caso di una matrice quadrata di ordine tre:

$$A = \begin{pmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{pmatrix}$$

Usando le regole esposte possiamo dire che la matrice ottenuta da A sottraendo alla seconda riga la prima moltiplicata per $\frac{a'}{a}$, e sottraendo alla terza la prima moltiplicata per $\frac{a''}{a}$, ovvero la matrice

$$B = \begin{pmatrix} a & b & c \\ 0 & b' - b\frac{a'}{a} & c' - c\frac{a'}{a} \\ 0 & b'' - b\frac{a''}{a} & c'' - c\frac{a''}{a} \end{pmatrix}$$

ha lo stesso determinante di A. Così lo stesso determinante di B e quindi di A avrà la matrice ottenuta da B sottraendo alla terza riga la seconda moltiplicata per $\frac{b^{''}-b\frac{a^{''}}{a}}{b^{'}-b\frac{a^{'}}{a}}$ ovvero la matrice

$$C = \begin{pmatrix} a & b & c \\ 0 & b^{'} - b\frac{a^{'}}{a} & c^{'} - c\frac{a^{'}}{a} \\ 0 & 0 & c^{''} - c\frac{a^{''}}{a} - (c^{'} - c\frac{a^{'}}{a})\frac{b^{''}-b\frac{a^{''}}{a}}{b^{'}-b\frac{a^{'}}{a}} \end{pmatrix}$$

Ma la matrice C è una matrice triangolare e pertanto il suo determinante si può calcolare moltiplicando gli elementi della sua diagonale principale. Il metodo illustrato non è l'unico uso che si può fare delle propietà dei determinanti per il loro calcolo. Altri usi di quelle regole appaiono solo nella pratica del calcolo dei determinanti.

Nel ragionamento precedente erano implicite delle ipotesi che nella pratica non sono sempre verificate. Ad esempio si supponeva $a \neq 0$, così successivamente si supponeva $b^{'} - b\frac{a^{'}}{a} \neq 0$. In tali casi il ragionamento va modificato. Accenniamo a come si debba modificare se $a = 0$. Si capirà poi come si debba procedere in casi analoghi. Se $a = 0$, basterà scambiare la prima riga della matrice con una qualunque riga il cui primo elemento non sia nullo. In tal modo si otterrà una matrice, con il determinante opposto al determinante di A, ma per il cui calcolo si potrà ripetere il procedimento illustrato. Se non vi sono righe con un primo elemento diverso da zero, vuol dire che la prima colonna della matrice assegnata, era tutta nulla e pertanto il determinante della matrice era zero. Siccome dopo il primo passo, la prima riga e la prima colonna della matrice da cui si parte non intervengono più, si capisce che se si presenta successivamente una situazione analoga, ad esempio se $b^{'} - b\frac{a^{'}}{a} = 0$, si ripete il procedimento appena indicato per la matrice, ottenuta dopo il primo passo, nella quale siano state cancellate la prima riga e la prima colonna.

L'estensione del metodo indicato al caso generale non dovrebbe presentare difficoltà.

Il procedimento illustratato, che permette di passare da una matrice ad una matrice tiangolare con lo stesso determinante di quella di partenza, a meno di un eventuale cambiamento di segno, si presta, senza sostanziali modifiche, tanto al calcolo del rango di una matrice, ma di questo parleremo nel paragrafo successivo, quanto alla risoluzione dei sistemi di equazioni lineari, nel qual caso va sotto il nome di *metodo di soluzione di Gauss* o *metodo delle eliminazioni successive*.

La descrizione in generale del metodo delle eliminazioni successive, implica una serie di casi e sottocasi che oscurano un pò la sostanza del metodo. Preferiamo pertanto di darne ora una prima descrizione per di un sistema di tre e quazioni in tre incognite, e di impratichirci successivamente del metodo nel caso generale nel corso di esercizi.

Partiamo dal sistema

$$\begin{cases} ax + by + cz = d \\ a^{'}x + b^{'}y + c^{'}z = d^{'} \\ a^{''}x + b^{''}y + c^{''}z = d^{''} \end{cases}$$

Di importanza fndamentale in quello che segue, è il fatto di verifica immediata, che

Se si sostituisce in un sistema un'equazione con un'altra che è la somma dell'equazione stessa e di una combinazione lineare delle altre equazioni del sistema, si ottiene un sistema equivalente.

Ciò comporta che il sistema ottenuto da quello dato lasciando la prima equazione immutata e sostituendo la seconda con la differenza della seconda meno la prima moltiplicata per $\frac{a^{'}}{a}$ e sostituendo la terza con la differenza della terza meno la prima moltiplicata per $\frac{a^{''}}{a}$, ovvero il sistema

$$\begin{cases} ax + by + cz = d \\ (b^{'} - b\frac{a^{'}}{a})y + (c^{'} - c\frac{a^{'}}{a})z = d^{'} - d\frac{a^{'}}{a} \\ (b^{''} - b\frac{a^{''}}{a})y + (c^{''} - c\frac{a^{''}}{a})z = d^{''} - d\frac{a^{''}}{a} \end{cases}$$

è equivalente al sistema dato. Per lo stesso motivo sarà equivalente al sistema dato quello ottenuto dal secondo sistema lasciando immutate le prime due equazioni e sostituendo la terza con la differenza della terza meno la seconda moltiplicata per $\frac{b^{''}-b\frac{a^{''}}{a}}{b^{'}-b\frac{a^{'}}{a}}$:

$$\begin{cases} ax + by + cz = d \\ (b^{'} - b\frac{a^{'}}{a})y + (c^{'} - c\frac{a^{'}}{a})z = d^{'} - d\frac{a^{'}}{a} \\ \left[(c^{''} - c\frac{a^{''}}{a}) - (c^{'} - c\frac{a^{'}}{a})\left(\frac{b^{''}-b\frac{a^{''}}{a}}{b^{'}-b\frac{a^{'}}{a}}\right)\right] z = (d^{''} - d\frac{a^{''}}{a}) - (d^{'} - d\frac{a^{'}}{a})\frac{b^{''}-b\frac{a^{''}}{a}}{b^{'}-b\frac{a^{'}}{a}} \end{cases}$$

Si osservi come in tale sistema l'ultima equazione sia un'equazione nella sola z e come pertanto si possa da tale ultima equazione ricavare la z. Sostituendo il valore della z così trovato nella seconda equazione, si ottiene nuovamente un'equazione in una sola incognita, che in tal caso è la y. Ricavando il valore della y dalla seconda equazione nella quale sia stato sostituito il valore della z trovato dalla terza equazione e sostituendo poi nella prima equazione i valori ottenuti della y e della z si trova un'equazione nella sola x che può essere risolta rispetto alla x. In tal modo si ottiene la terna (x, y, z) soluzione del terzo sistema e quindi, essendo il terzo sistema equivalente al primo, si ottiene la terna soluzione del sistema assegnato.

Ci si sarà accorti che il ragionamento esposto ha fatto via via uso di ipotesi non sempre verificate. Nel primo passo si è sfruttata l'ipotesi che sia $a \neq 0$. Se ciò non fosse basterebbe cambiare l'ordine delle equazioni e mettere al posto della prima equazione una qualunque equazione con coefficiente della x non nullo. Qualora tutte le equazioni avessero avuto coefficiente della x nullo, si sarebbe potuto provvisoriament trattare il sistema di tre equazini in tre incognite come un sistema di tre equazioni in due incognite, ricordandosi però alla fine dei ragionamenti che una, eventuale, soluzione sarebbe comunque dovuta essere una terna, di cui il primo elemento, la x sarebbe stato arbitrario non dovendo soddisfare ad alcuna condizione. Nel secondo passo è stata sfruttata l'ipotesi che diverso da zero sia $b^{'} - b\frac{a^{'}}{a}$. Ma è chiaro che se ciò fosse stato, ci si sarebbe potuti comportare a questo punto analogamente al caso precedente, andando a vedere quale delle equazioni del secondo sistema, privato della prima equazione, avesse avuto un coefficiente della y non nullo e mettendo tale equazione per prima. Infine al terzo passo, per ricavare la z, è stata fatta l'ipotesi che risulti

$$(c^{''} - c\frac{a^{''}}{a}) - (c^{'} - c\frac{a^{'}}{a}) \left(\frac{b^{''} - b\frac{a^{''}}{a}}{b^{'} - b\frac{a^{'}}{a}} \right) \neq 0$$

qui si osservi che, qualora ciò non fosse stato, si sarebbero presentati due casi: o anche il termine noto della terza equazione del terzo sistema era nullo, ed in tal caso sulla z non ci sarebbero state condizioni ed il sistema si poteva riguardare come un sistema di due sole equazioni, le prime due, in tre incognite, oppure il termine noto era diverso da zero ed in tal caso la terza equazione sarebbe stata incompatibile comportando evidentemente ciò l'incompatibilità del sistema originario.

Notiamo come se le ipotesi, fatte implicitamente, sono soddisfatte, il sistema è un sistema compatibile e determinato. Qualora invece una qualunque delle ipotesi non sia soddisfatta, il sistema può tanto essere un sistema compatibile ma indeterminato, quanto un sistema incompatibile. E' bene ad avere chiaro sin da ora che la compatibilità o l'incompatibilità di un sistema, così come il fatto che il sistema sia, nel caso compatibile, determinato od indeterminato, non dipende dal numero delle equazoni in rapporto al numero delle incognite, come una prima impressione potrebbe suggerire. Inoltre il cenno dato sul metodo delle eliminazioni successive dovrebbe anche far capire che con tale metodo non si è in grado di sapere, senza tentare di calcolare la eventuale soluzione, di che tipo di sistema si tratti. La nozione che è alla base della suddivisione dei sistemi nei diversi tipi è la nozione di rango. Sulla nozione di rango è basato il Teorema di Rouché-Capelli. Tale teorema permette di stabilire il tipo del sistema senza avere bisogno di cercarne le soluzioni. Il Teorema di Rouché-Capelli fornisce però anche un metodo per risolvere il sistema. Parleremo della nozione di tango e del teorema di Rouché Capelli, nel paragrafo successivo.

4. Rango di una matrice. Teorema di Rouché-Capelli.

Il rango di un certo insieme (finito od infinito) di $n-ple$ è il massimo numero di $n-ple$ indipendenti che si possono trovare tra le $n-ple$ date.

Data una matrice, $A_{(m,n)}$, ad m righe ed n colonne, il rango della matrice $A_{(m,n)}$, indicato con $rgA_{(m,n)}$, è il rango dell'insieme di m $n-ple$, costituito dalle righe della matrice.

In relazione alla definizione di rango diamo subito un importante risultato di cui omettiamo la dimostrazione

Il rango di $A_{(m,n)}$, definito come il rango delle m $n-ple$ che costituiscono le righe della matrice, uguaglia il rango delle n $m-ple$ che costituiscono le colonne della matrice.

Lo stesso teorema si può enunciare anche dicendo che

Il massimo numero di righe indipendenti di una matrice uguaglia il massimo numero di colonne indipendenti della matrice.

E' chiaro come, date m $n-ple$, il problema del calcolo di quale sia il massimo numero di $n-ple$ indipendenti tra le $n-ple$ date, sia ricondotto al calcolo del rango della matrice ad m righe ed n colonne, avente come m righe le m $n-ple$ assegnate. Come vedremo il calcolo del rango di una matrice ci dirà anche qualcosa in più: ci permetterà di individuare tutti gli insiemi di $n-ple$ indipendenti tra le $n-ple$ date. Osserviamo inoltre come il teorema enunciato senza dimostrazione, abbia una conseguenza inaspettata:

Più di n $n-ple$ sono sempre dipendenti.

Infatti evidentemente in una matrice non si possono trovare più colonne indipendenti di quante non siano le colonne della matrice stessa. Se vengono assegnate m $n-ple$, con $m > n$, il rango della matrice non potrà superare il numero delle colonne, quindi non potrà superare n. Ma allora anche il massimo numero delle righe indipendenti, ovvero il massimo numero delle $n-ple$ indipendenti, non potrà superare n e pertanto le m $n-ple$ dovranno essere dipendenti.

Prima di fornire le regole per il calcolo del rango di una matrice, vogliamo enunciare la prima parte del teorema di Rouché-Capelli, dalla quale apparirà la stretta relazione tra la nozione di rango di una matrice, ed il problema, dal quale siamo partiti, della ricerca delle eventuali soluzioni di un sistema di equazioni lineari.

Innanzitutto da ora in poi i sistemi che considereremo non saranno necessariamente quadrati, ovvero con un numero di incognite ugualeal numero delle equazioni. Il numero delle incognite potrà essere tanto maggiore quanto minore o uguale del numero delle equazioni. Quindi un sistema arbitrario sarà del tipo

$$\begin{cases} a_1^1x^1 + a_2^1x^2 + ... + a_n^1x^n = b^1 \\ a_1^2x^1 + a_2^2x^2 + ... + a_n^2x^n = b^2 \\ ... \\ ... \\ a_1^mx^1 + a_2^mx^2 + ... + a_n^mx^n = b^m \end{cases}$$

dove con m si è indicato il numero delle equazioni e dove può essere $n \lesseqgtr m$. In relazione ad un tale sistema generale possiamo enunciare la prima parte del Teorema di Rouché-Capelli.

Teorema di Rouché-Capelli (I parte): Dato un sistema di m equazioni lineari in n incognite, tale sistema è compatibile se e solo se il rango della matrice dei coefficienti, detta matrice incompleta del sistema, è uguale al rango della matrice ottenuta da quella dei coefficienti aggiungendo alle colonne di tale matrice, l'ulteriore colonna dei termini noti, matrice che è detta matrice completa del sistema. Inoltre se il rango della matrice incompleta è uguale al rango della matrice completa e pertanto il sistema è compatibile, il sistema è anche determinato se e solo se il comune rango delle due matrici, uguaglia il numero delle incognite. Infine, sempre nel caso compatibile, se il comune rango delle due matrici è minore del numero delle incognite, le soluzioni sono infinite.

Vediamo dunque come sia possibile calcolare il rango di una matrice.

Immaginiamo di fissare una matrice arbitraria. Indichiamo con m il numero delle righe della matrice, e con n il numero delle sue colonne. Indichiamo la matrice stessa con $A_{(m,n)}$. Fissiamo ora un numero $p \leq min(m, n)$. Scegliamo p delle righe della matrice e p delle sue colonne. Andiamo poi a prendere tutti gli elementi della matrice che si trovano all'intersezione delle righe e delle colonne scelte disponendoli in una nuova matrice, quadrata a p righe e p colonne, nella stessa posizione che avevano nella matrice. Una tale matrice di ordine p viene detta *matrice di ordine p estratta dalla matrice data* ed il suo determinante viene detto *minore di ordine p*. Vi saranno minori di ordine uno, gli elementi della matrice $A_{(m,n)}$, minori di ordine due, i determinanti delle matrici quadrate estratte dalla matrice data, e così via, vi saranno i determinanti delle matrici estratte di ordine il minimo tra m ed n, che saranno i minori di ordine massimo che si possono estrarre da $A_{(m,n)}$. Ebbene si ha il seguente risultato

Il massimo ordine dei minori non nulli che si possono estrarre dalla ma-

trice $A_{(m,n)}$, uguaglia il rango di $A_{(m,n)}$, ovvero uguaglia il massimo numero di righe (colonne) indipendenti che si possono trovare nella matrice $A_{(m,n)}$.

Rendiamoci conto di come il risultato esposto fornisca una regola di calcolo per il rango di una matrice in un caso particolare, quello della seguente matrice

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

La matrice ha tre righe e quattro colonne. Il suo rango dorà essere minore od uguale al minimo tra tre e quattro quindi minore od uguale a tre. Le matrici quadrate di ordine massimo estraibili dalla matrice assegnata saranno di ordine tre. Esse si otterranno fissando tutte e tre le righe disponibili e tre delle quattro colonne della matrice. Quando si fissano tre colonne su quattro se ne esclude una, quindi fissare tre colonne equivale a togliere una colonna. Le maniere per togliere una colonna su quattro sono quattro: o si toglie la prima colonna, o la seconda, o la terza, o la quarta. Si otterranno in tal modo le seguenti matrici estratte di ordine tre

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

Siccome la terza riga della matrice dalla quale siamo partiti era la somma delle prime due righe, in ognuna delle matrici estratte la terza riga continuerà ad essere la somma delle prime due e quindi ogni matrice estratta, nel nostro caso, avrà determinante zero: tutti i quattro minori di ordine tre della matrice assegnata sono nulli, quindi l'ordine massimo dei minori non nulli della matrice data deve essere minore di tre e per il risultato esposto il rango della matrice dovrà essere minore di tre. Per vedere quanto il rango vale realmente dobbiamo continuare a calcolare i minori della matrice. Andiamo a vedere quelli di ordine due. Si ottengono calcolando i determinanti delle matrici estratte di ordine due. Tali matrici si ottengono fissando a piacere due righe e due colonne. Cominciano ad essere numerose. Allora si va a vedere ad occhio se tra le matrici estratte di ordine due non ve ne sia una non nulla. Si vede così che fissando la prima riga e la prima colonna, si ottiene la matrice estratta di ordine due

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

il cui determinante è -4 quindi non nullo. Abbiamo trovato un minore non nullo di ordine due. Abbiamo anche visto che non vi sono minori di ordine tre che non siano nulli. Quindi l'ordine massimo dei minori non nulli è due.

Il rango della nostra matrice è due. Il massimo numero di righe, o di colonne, indipendenti, che si possono trovare nella matrice, è due.

IL passaggio, "ad occhio", fa parte integrante del metodo esposto se si vogliono abbreviare i conti. Esiste comunque un risultato che permette a volte di abbreviare tali conti seguendo un diverso metodo per il calcolo dell'ordine massimo dei minori non nulli. Si tratta del risultato seguente.

Se un minore di ordine p è non nullo e sono nulli tutti i suoi orlati, allora sono nulli tutti i minori di ordine maggiore di p.

Per capire l'enunciato bisogna saper cosa è l'orlato di un minore. Una matrice estratta di ordine p di una matrice, è costruita fissando p righe e p colonne della matrice assegnata. Aggiungendo una riga ed una colonna alla matrice estratta, tra le righe e le colonne che non sono state considerate per la costruzione della matrice estratta, si ottiene una matrice estratta di ordine $p+1$ che viene detta *matrice orlata* della matrice estratta assegnata. Il determinante dell'orlata viene detto *minore orlato* del minore che è il determinante della matrice estratta dalla quale si è partiti. Fissata una matrice estratta di ordine p, questa in generale avrà diversi orlati di ordine $p+1$. Tali orlati non saranno tutti i minori di ordine $p+1$ della matrice, almeno nel caso generale. In ogni caso, sempre in generale, oltre a minori di ordine $p+1$ vi saranno minori di ordini maggiori. Quindi i minori orlati di un dato minore di ordine p sono meno dei minori di ordine più grande di p. Da questo si capisce che il teorema può costituire un criterio utile per abbreviare i conti del calcolo del rango. Facciamo come esempio il caso della matrice considerata in precedenza:

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

Si vede subito che la matrice estratta ottenuta fissando prima e seconda riga e prima e seconda colonna ha determinante diverso da zero. Quindi intanto il rango della matrice che stiamo considerando è almeno due giacché almeno un minore di ordine due non nullo, vi è. Le matrici orlate della matrice di ordine due estratta fissando prima e seconda riga e prima e seconda colonna, sono le seguenti matrici di ordine tre

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

la prima matrice è ottenuta aggiungendo la terza riga e la terza colonna, la seconda matrice è ottenuta aggiungendo la terza riga e la quarta colonna. Si noti come un'altra matrice estratta di ordine tre è la matrice

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

che però non è una matrice orlata della matrice estratta dalle quale siamo partiti. Andando a calcolare i determinanti delle due matrici orlate si trova in entrambi i casi zero e quindi tutti i minori orlati del minore non nullo di ordine due sono nulli, pertanto in base al teorema, ogni minore di ordine maggiore di due è nullo e il rango della matrice sarà due.

Passiamo all'enunciato della seconda parte del teorema di Rouché-Capelli. La prima parte riguardava l'esistenza o meno delle soluzioni di un sistema di equazion lineari. La seconda parte fornisce un metodo per il calcolo delle soluzioni nel caso queste esistano.

Teorema di Rouché-Capelli (II parte):

i) (Passaggio al sistema ridotto). Se un sistema di equazioni lineari è compatibile, detto p il comune rango della matrice completa e di quella incompleta, le soluzioni del sistema coincidono con le soluzioni del sistema ottenuto da quello dato, togliendo le equazioni che non sono relative alle righe di un (comunque) fissato minore non nullo di ordine massimo. Tale sistema si chiama sistema ridotto.

Detto n il numero delle incognite del sistema e quindi il numero delle colonne della matrice incompleta, risulterà $p \leq n$.

ii) Se $p = n$ allora il sistema ridotto è un sistema quadrato con matrice dei coefficienti proprio la matrice il cui determinante è il minore non nullo di ordine massimo scelto e pertanto tale sistema si può risolvere con la regola di Cramer. In tal caso il sistema ridotto ha una sola soluzione (per la regola di Cramer) e tale soluzione è anche l'unica soluzione del sistema assegnato.

iii) Se $p < n$ allora si possono portare a secondo membro del sistema ridotto le $n - p$ incognite che non sono relative alle colonne del minore fissato, e considerare il sistema così scritto come un sistema di p equazioni nelle sole p incognite lasciate a primo membro. Immaginando di fissare, genericamente, le incognite portate a secondo membro, il sistema ridotto pensato nelle sole incognite a primo membro sarà un sistema risolubile con Cramer dando così luogo ad una ben determinata soluzione per ogni scelta dei valori delle incognite a secondo membro. L'n-pla costituita dai valori della soluzione del sistema ridotto pensati in funzione delle variabili portate a secondo membro e dalle variabili portate a secondo membro, sono tutte soluzioni del sistema originario, comunque varino le variabili portate a secondo membro. Le soluzioni del sistema originario sono pertanto infinite. Siccome una soluzione *tra le infinite soluzioni rimane individuata da una scelta delle* $n-p$ *variabili a secondo membro, si dice che le soluzioni del sistema sono* ∞^{n-p}.

Illustriamo il Teorema di Rouché-Capelli, con un esempio. Facciamo il caso di un sistema di tre equazioni in quattro incognite con matrice dei coefficienti la matrice considerata in precedenza e con colonna dei termini noti la colonna

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

consideriamo quindi il sistema

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

le matrici completa ed incompleta sono rispettivamente

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

Si è visto che il rango della matrice incompleta è due. Quindi tutti i minori di ordine tre della matrice incompleta sono nulli. Questo non esclude che nella matrice completa vi sia qualche minore di ordine tre non nullo e che il rango della matrice completa sia tre. La matrice

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

è una matrice estratta dalla matrice completa: quella ottenuta fissando le tre righe della matrice completa e le sue ultime tre colonne. Il suo determinante non è nullo. Dunque la matrice completa ha rango tre. Matrice completa e matrice incompleta non hanno lo stesso rango. Il sistema non ammette soluzioni, è incompatibile.

Facciamo un altro caso. Prendiamo il sistema la cui matrice dei coefficienti è quella di prima, ma la colonna dei termini noti è data dalla colonna

$$\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

Dunque consideriamo il sistema

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

Questa volta una verifica diretta mostra che la matrice completa e la matrice completa hanno lo stesso rango e che tale rango comune è 2. Vediamo come si risolve il sistema.

Come primo passo bisogna fissare nella matrice incompleta una matrice estratta con determinante non nullo di ordine pari al rango della matrice incompleta. Immaginiamo di fissare la matrice estratta ottenuta fissando la prima e la seconda riga e la prima e la seconda colonna ovvero la matrice

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

Il secondo passo consiste nel considerare il sistema costituito solo dalle prime due equazioni:

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

Si tratta di un sistema in un numero di equazioni minore del numero delle incognite. Se equazioni ed incognite fossero state in ugual numero si sarebbe risolto il sistema con Cramer e la soluzione trovata sarebbe stata la soluzione del sistema assegnato. Ma in questo caso vi è un terzo passo da fare.

Il terzo passo consiste nello riscrivere il sistema portando a secondo membro la terza e la quarta incognita:

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

Se in questo sistema la z e la t non vengono riguardate come incognite ma come valori numerici non specificati e come incognite si riguardano solo x la e la y, il sistema appare come un sistema di due equazioni in due incognite, con matrice dei coefficienti la matrice estratta fissata e si può ricavare la soluzione con la regola di Cramer ottenendo in tal modo come soluzione la coppia

$$x = \frac{\begin{vmatrix} 1 - 3z - 2t & 2 \\ 1 - z & 0 \end{vmatrix}}{\begin{vmatrix} 1 & 2 \\ 2 & 0 \end{vmatrix}} \quad y = \frac{\begin{vmatrix} 1 & 1 - 3z - 2t \\ 2 & 1 - z \end{vmatrix}}{\begin{vmatrix} 1 & 2 \\ 2 & 0 \end{vmatrix}}$$

ovvero la coppia

$$x = \tfrac{2z-1}{-4} \quad y = \tfrac{5z+4t-1}{-4}$$

Le soluzioni del sistema assegnato sono tutte e sole le quaterne

$$\left(\frac{2z-1}{-4}, \frac{5z+4t-1}{-4}, z, t\right)$$

al variare di z e di t, che sono quelli che si chiamano i parametri che bisogna specificare per trovare una soluzione particolare attraverso la formula scritta.

Qualora anziché dalla matrice estratta ottenuta fissando la prima riga e la prima colonna, si fosse fissata un'altra matrice estratta, naturalmente sempre a determinante non nullo e dello stesso ordine, si sarebbe trovata una formula analoga nella quale però la soluzione sarebbe dipesa da altri parametri: dalle variabili portate a secondo membro in questo altro caso. Le soluzioni sarebbero state le stesse ma descritte parametricamente in maniera diversa.

5. Sistemi omogenei.

I sistemi omogenei sono i sistemi nei quali tutti i termini noti sono nulli, ossia i sistemi del tipo seguente

$$\begin{cases} a_1^1 x^1 + a_2^1 x^2 + ... + a_n^1 x^n = 0 \\ a_1^2 x^1 + a_2^2 x^2 + ... + a_n^2 x^n = 0 \\ ... \\ ... \\ a_1^m x^1 + a_2^m x^2 + ... + a_n^m x^n = 0 \end{cases}$$

Il teorema di Rouché-Capelli si applicherà anche a questo caso, che è un caso particolare di sistema di equazioni lineari. Si vede allora che, essendo la matrice completa ottenuta dalla matrice incompleta con l'aggiunta della colonna dei termini noti che in questo caso è la colonna nulla, le due matrici, quella completa e quella incompleta, avranno sempre, qualunque sia il sistema omogeneo considerato, lo stesso rango. Quindi i sistemi omogenei sono sempre compatibili, ammettono sempre soluzioni. Questo è peraltro un fatto ovvio giacché è evidente che, qualunque sia il sistema omogeneo, questo ammette sempre come soluzione l'$n-pla$ nulla. Per i sistemi omogenei non si tratta dunque di stabilire se le soluzioni esistono o meno, ma di stabilire se esistono altre soluzioni oltre all'$n-pla$ nulla. *Una soluzione di un sistema omogeneo diversa dalla soluzione nulla, viene detta autosoluzione* per un motivo che apparirà in seguito. L'$n-pla$ nulla viene detta *soluzione banale.*

La risposta al problema della ricerca delle eventuali autosoluzioni viene da un'applicazione immediata del teorema di Rouché-Capelli. Infatti sappiamo dal teorema di Rouché-Capelli, che un sistema compatibile, come sono i sistemi omogenei, è anche determinato, ovvero ammette un'unica soluzione, se e solo se il numero delle incognite, n, uguaglia il rango, p, della matrice incompleta. Ma nel caso dei sistemi omogenei la soluzione banale esiste sempre. Quindi un sistema omogeneo ammette solo la soluzione nulla se e solo se è determinato ovvero se e solo se il rango della matrice incompleta uguaglia il numero delle incognite. Tale teorema può essere enunciato al modo seguente

Un sistema omogeneo ammette autosoluzioni se e solo se $p < n$.

In particolare se il sistema omogeneo è quadrato, il numero delle equazioni uguaglia il numero delle incognite, $n = m$, e quindi quadrata sarà la matrice dei coefficienti, la condizione necessaria e sufficente affinché una tale matrice abbia rango minore di n sarà che sia nullo il suo determinante, essendo tale determinante l'unico minore di ordine n. Pertanto si conclude con il seguente risultato che tornerà utile molto spesso

Un sistema omogeneo quadrato ammette autosoluzioni se e solo se il determinante della matrice dei coefficenti è nullo.

Oltre ai risultati appena visti, per i sistemi omogenei ne valgono altri importanti che derivano dal fatto che le soluzioni di un sistema omogeneo soddisfano ad una proprietà che le soluzioni di un sistema non omogeneo non soddisfano. Si tratta della proprietà seguente:

Se $(x_1^1, x_1^2, ..., x_1^n), (x_2^1, x_2^2, ..., x_2^n)$ sono due soluzioni del sistema omogeneo

$$\begin{cases} a_1^1 x^1 + a_2^1 x^2 + ... + a_n^1 x^n = 0 \\ a_1^2 x^1 + a_2^2 x^2 + ... + a_n^2 x^n = 0 \\ ... \\ ... \\ a_1^m x^1 + a_2^m x^2 + ... + a_n^m x^n = 0 \end{cases}$$

e se λ, μ, sono due scalari, l'$n-pla$ combinazione lineare delle due soluzioni con coefficienti λ, μ, ossia l'n-pla

$$\lambda(x_1^1, x_1^2, ..., x_1^n) + \mu(x_2^1, x_2^2, ..., x_2^n) = (\lambda x_1^1 + \mu x_2^1, \lambda x_1^2 + \mu x_2^2, ..., \lambda x_1^n + \mu x_2^n)$$

è ancora soluzione dello stesso sistema omogeneo.

La dimostrazione consiste in una semplice verifica. Dire che le due $n-ple$ sono soluzioni del sistema significa che dette due $n-ple$ risolvono tutte le equazioni del sistema. In particolare la prima equazione. Dunque risulterà tanto

$$a_1^1 x_1^1 + a_2^1 x_1^2 + ... + a_n^1 x_1^n = 0$$

quanto

$$a_1^1 x_2^1 + a_2^1 x_2^2 + ... + a_n^1 x_2^n = 0$$

ma allora avremo

$$\begin{aligned} a_1^1(\lambda x_1^1 + \mu x_2^1) + a_2^1(\lambda x_1^2 + \mu x_2^2) + ... + a_n^1(\lambda x_1^n + \mu x_2^n) &= \\ \lambda(a_1^1 x_1^1 + a_2^1 x_1^2 + ... + a_n^1 x_1^n) + \mu(a_1^1 x_2^1 + a_2^1 x_2^2 + ... + a_n^1 x_2^n) &= 0 + 0 = 0 \end{aligned}$$

ovvero anche la combinazione lineare delle due $n-ple$ è soluzione della prima equazione. Ma il ragionamento fatto per la prima equazione si ripete tale e quale per ogni equazione del sistema e si conclude con la tesi.

Il risultato esposto viene arricchito della considerazione seguente. Consideriamo il caso di un sistema omogeneo con autosoluzioni. Dunque per il quale risulti $p < n$. In tal caso sappiamo dal teorema di Riucé-Capelli che il sistema dato equivale ad un sistema ridotto con solo p equazioni. Per semplicità supponiamo che come p equazioni del sistema ridotto possano scegliersi le prime p equazioni del sistema omogeneo assegnato e che quindi il sistema ridotto sia il sistema

$$\begin{cases} a_1^1x^1 + a_2^1x^2 + ... + a_n^1x^n = 0 \\ a_1^2x^1 + a_2^2x^2 + ... + a_n^2x^n = 0 \\ ... \\ ... \\ a_1^px^1 + a_2^px^2 + ... + a_n^px^n = 0 \end{cases} \qquad p < n$$

Di nuovo supponiamo per semplicità espositiva che le variabili che possono essere portate a secondo membro in modo da lasciare a primo membro delle variabili con matrice dei coefficienti a determinante non nullo, siano le ultime $n-p$ variabili. Portiamo a secondo membro le ultime $n-p$ variabili e consideriamo il sistema

$$\begin{cases} a_1^1x^1 + a_2^1x^2 + ... + a_p^1x^p = -a_{p+1}^1x^{p+1} - a_{p+2}^1x^{p+2} - ... - a_n^1x^n \\ a_1^2x^1 + a_2^2x^2 + ... + a_p^2x^p = -a_{p+1}^2x^{p+1} - a_{p+2}^2x^{p+2} - ... - a_n^2x^n \\ ... \\ ... \\ a_1^px^1 + a_2^px^2 + ... + a_p^px^p = -a_{p+1}^px^1 - a_{p+2}^px^{p+2} - ... - a_n^px^n \end{cases}$$

Per ogni scelta delle $n-p$ variabili portate a secondo membro rimane determinato un sistema quadrato di ordine p risolubile con la regola di Cramer. Risolvendo il sistema si troveranno i valori delle prime p variabili che aggiunti ai valori assegnati alle $n-p$ variabili portate a secondo membro forniranno una soluzione particolare del sistema omogeneo che si sta considerando. Questo sempre per il teorema di Rouché-Capelli. Assegnamo allora alle variabili a secondo membro i valori

$$x^{p+1} = 1, x^{p+2} = 0, ..., x^n = 0$$

Indicando con

$$x_1^1, x_1^2, ..., x_1^p$$

i valori delle p variabili a primo membro, calcolate in conseguenza, avremo come soluzione del sistema omogeneo l'$n-pla$

$$(x_1^1, x_1^2, ..., x_1^p, 1, 0, ..., 0)$$

Ripetendo il ragionamento prendendo i valori delle variabili a secondo membro tutti nulli eccetto per una variabile, otterremo esattamente $n-p$ soluzioni:

$$\begin{array}{l}(x_1^1, x_1^2, ..., x_1^p, 1, 0, ..., 0)\\(x_2^1, x_2^2, ..., x_2^p, 01, , ..., 0)\\.\\.\\(x_{n-p}^1, x_{n-p}^2, ..., x_{n-p}^p, 0, 0, ..., 1)\end{array}$$

Tali $n-ple$ sono indipendenti poiché considerando la matrice che ha le $n-ple$ per righe, si vede che la sua matrice estratta fissando tutte le $n-p$ righe e le ultime $n-p$ colonne è la matrice

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

il cui determinante è uno e quindi diverso da zero. Quindi esiste una minore non nullo di ordine pari al numero delle $n-ple$ e pertanto le $(n-p)$, $n-ple$ sono indipendenti. A questo punto possiamo dire che, non solo come osservato, ogni combinazione lineare delle $n-p$ soluzioni trovate è ancora una soluzione, ma possiamo dire di più. Possiamo asserire che facendo tutte le possibili combinazioni lineari delle soluzioni trovate, troviamo tutte le soluzioni del sistema omogeneo. Infatti, sempre per il teorema di Rouché-Capelli, le soluzioni del sistema si trovano tutte scegliendo ad arbitrio le variabili a secondo membro. Quindi lasciando variare comunque l'$(n-p)-pla$ delle variabili a secondo membro. Ma tutte le combinazioni lineari delle $n-p$ scelte delle variabili a secondo membro forniscono proprio la generica $(n-p)-pla$. Infatti si ha

$$\lambda^1(1,0,...,0) + \lambda^2(0,1,0,...,0) + ... + \lambda^{n-p}(0,0,...,1) = (\lambda^1, \lambda^2, ..., \lambda^{n-p})$$

Le autosoluzioni trovate del sistema omogeneo

$$\begin{array}{l}(x_1^1, x_1^2, ..., x_1^p, 1, 0, ..., 0)\\(x_2^1, x_2^2, ..., x_2^p, 01, , ..., 0)\\.\\.\\(x_{n-p}^1, x_{n-p}^2, ..., x_{n-p}^p, 0, 0, ..., 1)\end{array}$$

vengono dette essere un *Sistema fondamentale di autosoluzioni.* In generale si dà la definizione seguente

Un sistema fondamentale di autosoluzioni di un sistema omogeneo è una $(n-p)-pla$ di autosoluzioni indipendenti del sistema.

Quello che si può vedere, ma che vedremo in seguito in relazione alla teoria degli spazi verttoriali, è che non solo partendo dal particolare sistema fondamentale di autosoluzioni dal quale noi siamo partiti, si possono trovare tutte le soluzioni del sistema con le combinazioni lineari, ma partendo da un qualunque sistema fondamentale di autosoluzioni del sistema si ottengono tutte le soluzioni del sistema per combinazioni lineari.

Facciamo un esempio relativo alle ultime cose dette. Consideriamo il seguente sistema omogeneo di tre equazioni in quattro incognite

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

Come si noterà si tratta del sistema considerato in precedenza in cui la colonna dei termini noti è stata sostituita con la colonna nulla.

Un sistema omogeneo ottenuto da un sistema non omogeneo sostituendo la colonna dei termini noti con la colonna nulla viene detto sistema omogeneo associato al sistema dato.

Chiaramente ad ogni sistema rimane associato un sistema omogeneo, ma un sistema omogeneo è associato ad infiniti sistemi non omogenei.

Il sistema da noi considerato è il sistema omogeneo associato al sistema non omogeneo che avevamo considerato in precedenza.

Torniamo al nostro sistema omogeneo. Già si era visto che il suo rango era due e che un minore di ordine massimo, due, non nullo era quello ottenuto considerando la prima riga e la prima colonna. Quindi il sistema ridotto sarà

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

Inoltre potremo portare a secondo membro la terza e la quarta variabile

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

Dando a tali variabili prima i valori $z=1, t=0$ poi i valori $z=0, t=1$, otteniamo rispettivamente le due soluzioni

$$(-\frac{1}{2}, -\frac{5}{4}, 1, 0), (0, -1, 0, 1)$$

Tale coppia di soluzioni costituisce un sistema fondamentale di soluzioni del sistema assegnato. La generica soluzione del sistema sarà

$$\lambda(-\frac{1}{2}, -\frac{5}{4}, 1, 0) + \mu(0, -1, 0, 1)$$

Esiste un importante legame tra le soluzioni di un sistema e le soluzioni del sistema omogeneo associato. Ognuno può verificare senza difficoltà che, prese due soluzioni di un sistema non omogeneo, la loro differenza è una soluzione del sistema omogeneo associato. Quindi

Una soluzione qualunque di un sistema omogeneo può ottenersi sommando ad una soluzione particolare del sistema non omogeneo, una qualunque soluzione del sistema omogeneo associato.

Un tipo di sistemi omogenei che ricorre frequentemente è quello dei sistemi omogenei in n incognite ed $n-1$ equazioni, di rango $n-1$. Si tratta di sistemi omogenei del tipo

$$\begin{cases} a_1^1x^1 + a_2^1x^2 + ... + a_n^1x^n = 0 \\ a_1^2x^1 + a_2^2x^2 + ... + a_n^2x^n = 0 \\ ... \\ ... \\ a_1^{n-1}x^1 + a_2^{n-1}x^2 + ... + a_n^{n-1}x^n = 0 \end{cases}$$

La matrice dei coefficienti è la matrice

$$\begin{pmatrix} a_1^1 & a_2^1 & . & . & a_n^1 \\ a_1^2 & a_2^2 & . & . & a_n^2 \\ . & . & & & . \\ . & . & & & . \\ a_1^{n-1} & a_2^{n-1} & . & . & a_n^{n-1} \end{pmatrix}$$

che ha $n-1$ righe ed n colonne. Per ogni colonna che viene cancella ta si ha un minore di ordine $n-1$. Cancellando successivamente la prima colonna, la seconda colonna,..., l'ultima colonna, si ottengono tutti gli n minori di ordine $n-1$ che indicheremo rispettivamente con

$$A_1, A_2, ..., A_n$$

Tali minori non sono tutti nulli visto che stiamo supponendo il rango della matrice $n-1$. Ebbene

Dato un sistema omogeneo in n incognite, $n-1$ equazioni di rango $n-1$, una soluzione del sistema è data dall'n-pla (non nulla)

$$(A_1, -A_2, ..., (-1)^{n-1}A_n)$$

Tutte le altre soluzioni sono proporzionali a tale n-pla.

Omettiamo la dimostrazione in generale di tale teorema limitandoci al caso particolare di un sistema omogeneo di due equazioni in tre incognite di rango due:

$$\begin{cases} ax + by + cz = 0 \\ a^{'}x + b^{'}y + c^{'}z = 0 \end{cases}$$

La matrice dei coefficienti è

$$\begin{pmatrix} a & b & c \\ a' & b' & c' \end{pmatrix}$$

I tre minori di ordine due sono

$$\begin{vmatrix} b & c \\ b' & c' \end{vmatrix}, \begin{vmatrix} a & c \\ a' & c' \end{vmatrix}, \begin{vmatrix} a & b \\ a' & b' \end{vmatrix}$$

Una soluzione sarà

$$\left(\begin{vmatrix} b & c \\ b' & c' \end{vmatrix}, - \begin{vmatrix} a & c \\ a' & c' \end{vmatrix}, \begin{vmatrix} a & b \\ a' & b' \end{vmatrix} \right)$$

La generica soluzione sarà

$$\lambda \left(\begin{vmatrix} b & c \\ b' & c' \end{vmatrix}, - \begin{vmatrix} a & c \\ a' & c' \end{vmatrix}, \begin{vmatrix} a & b \\ a' & b' \end{vmatrix} \right) = \left(\lambda \begin{vmatrix} b & c \\ b' & c' \end{vmatrix}, -\lambda \begin{vmatrix} a & c \\ a' & c' \end{vmatrix}, \lambda \begin{vmatrix} a & b \\ a' & b' \end{vmatrix} \right)$$

Per convincersi del risultato enunciato in questo caso particolare, basta osservare due cose. Primo che, in base a quanto prima visto, dato che nel nostro caso $3 - 2 = 1$, le soluzioni dovranno potersi ottenere tutte da una sola non nulla (che in questo caso è, in base alla definizione, un sistema fondamentale di soluzioni) facendone tutte le possibili combinazioni lineari. Ma le combinazioni lineari di una sola terna sono semplicemente le terne proporzionali. In secondo luogo basta verificare che la terna indicata è di fatto una soluzione del sistema omogeneo. A tale scopo si tratta di sostituire la terna e di vedere se soddisfa le due equazioni. Sostituendo la terna nella prima equazione si trova a primo membro

$$a \begin{vmatrix} b & c \\ b' & c' \end{vmatrix} - b \begin{vmatrix} a & c \\ a' & c' \end{vmatrix} + c \begin{vmatrix} a & b \\ a' & b' \end{vmatrix}$$

e si vede che si tratta dello sviluppo secondo la prima riga del determinante della matrice

$$\begin{pmatrix} a & b & c \\ a & b & c \\ a' & b' & c' \end{pmatrix}$$

Poiché tale matrice ha la prima riga uguale alla seconda, il suo determinante deve essere nullo e quindi deve essere

$$a \begin{vmatrix} b & c \\ b' & c' \end{vmatrix} - b \begin{vmatrix} a & c \\ a' & c' \end{vmatrix} + c \begin{vmatrix} a & b \\ a' & b' \end{vmatrix} = 0$$

ossia la terna data deve soddisfare la prima equazione. Per verificare che la stessa terna soddisfa anche la seconda equazione, si procede analogamente e si vede che sostituendo la terna nella seconda equazione, si trova lo sviluppo del determinante di una matrice con due righe uguali e pertanto nullo. Il caso generale, in un numero arbitrario di incognite è del tutto analogo.

TRASFORMAZIONI LINEARI

1. Definizioni e prime proprietà.

Una funzione lineare di una variabile reale, x, è una funzione del tipo

$$y = ax + b$$

con a, b, costanti. Se $b = 0$ la funzione lineare viene detta omogenea o anche forma lineare.

Una funzione lineare di due variabili reali x, y è una funzione del tipo

$$z = ax + by + c$$

con a, b, c, costanti. Anche ora se $c = 0$ la funzione lineare viene detta omgenea oppure forma lineare.

Il caso generale di una funzione lineare in n variabili lo scriveremo al modo seguente

$$y = a_1x^1 + a_2x^2 + ... + a_nx^n + b$$

dove i coefficienti $a_1, a_2, ..., a_n, b$, sono delle costanti. Al solito se $b = 0$ si tratterà di una funzione lineare omogenea ovvero di una forma lineare.

Oltre a singole funzioni lineari si può considerare un numero finito di forme lineari in n variabili, diciamolo m.

Un insieme di m funzioni lineari ognuna in uno stesso numero n di variabili, lo scriveremo

$$\begin{cases} y^1 = a_1^1 x^1 + a_2^1 x^2 + ... + a_n^1 x^n + b^1 \\ y^2 = a_1^2 x^1 + a_2^2 x^2 + ... + a_n^2 x^n + b^2 \\ . \\ . \\ y^m = a_1^m x^1 + a_2^m x^2 + ... + a_n^m x^n + b^m \end{cases}$$

e lo chiameremo trasformazione lineare. Se tutte le funzioni lineari sono omogenee, la trasformazione lineare sarà detta omogenea.

Siccome nel seguito considereremo solo trasformazioni lineari omogenee, ossia solo quello del tipo

$$\begin{cases} y^1 = a_1^1 x^1 + a_2^1 x^2 + ... + a_n^1 x^n \\ y^2 = a_1^2 x^1 + a_2^2 x^2 + ... + a_n^2 x^n \\ . \\ . \\ y^m = a_1^m x^1 + a_2^m x^2 + ... + a_n^m x^n \end{cases}$$

chiameremo queste trasformazioni lineari omettendo l'aggettivo omogenee.

Notiamo che i due numeri n, m, sono indipendenti uno dall'altro: possono essere uno maggiore, minore, uguale all'altro.

Data la trasformazione lineare di cui sopra, si chiama matrice della trasformazione lineare la matrice ad m righe ed n colonne

$$A_{(m,n)=} \begin{pmatrix} a_1^1 & a_2^1 & . & a_n^1 \\ a_1^2 & a_2^2 & . & a_n^2 \\ . & . & & . \\ . & . & & . \\ a_1^m & a_2^m & . & a_n^m \end{pmatrix}$$

Una trasformazione lineare (omogenea) quale quella scritta, dà luogo ad una applicazione tra lo spazio numerico reale $n-$dimensionale R^n, lo spazio delle $n-ple$ di numeri reali, e lo spazio numerico reale $m-$ dimensionale R^m, delle $m-ple$ di numero reali: la corrispondenza che all'$n-pl$a $(x^1, x^2, ..., x^n)$ associa l'$m-pla$ $(y^1, y^2, ..., y^n)$ che si calcola attraverso le formule scritte. Notiamo che

Fissati due numeri n ed m, esiste una corrispondenza biunivoca tra le trasformazioni lineari tra R^n ed R^m, e le matrici ad m righe ed n colonne.

Data una trasformazione lineare rimane chiaramente individuata la sua matrice ed a partire da una matrice ad m righe ed n colonne si può univocamente scrivere la trasformazione lineare tra R^n ed R^m che ha la matrice data come matrice associata. Questo fatto ovvio torna spesso utile. Indichiamo brevemente una trasformazione lineare con L. Indicando con $\mathbf{x} = (x^1, x^2, ..., x^n)$ una generica $n - pla$ e con $\mathbf{y} = (y^1, y^2, ..., y^m)$ l'$m - pla$ che corrisponde all'$n - pla$ $\mathbf{x}$ secondo la trasformazione lineare L, potremo scrivere sinteticamente la trasformazione L ponendo

$$L(\mathbf{x}) = \mathbf{y}$$

Ebbene le trasformazioni lineari godono delle due importanti proprietà

$$\begin{cases} L(\mathbf{x}_1+\mathbf{x}_2) = L(\mathbf{x}_1) + L(\mathbf{x}_2) \\ L(\lambda\mathbf{x}) = \lambda L(\mathbf{x}) \end{cases}$$

dove con $\mathbf{x}_1, \mathbf{x}_2$ si sono indicate due arbitrarie $n-ple$ e con $\mathbf{x}_1+\mathbf{x}_2$ si è indicata la loro somma. Mentre con $\lambda\mathbf{x}$ si è indicato il prodotto dello scalare λ per l'$n - pla$ $\mathbf{x}$ e con $\lambda L(\mathbf{x})$ il prodotto dello stesso scalare per l'$n - pla$ $L(\mathbf{x})$.

Si osservi che, come apparirà dalla dimostrazione delle due proprietà che ora daremo, le proprietà indicate non sarebbero soddisfatte se anche una sola delle funzioni lineari che costituiscono la trasformazione lineare, non fosse omogenea.

Ci limitiamo a dimostrare la prima proprietà essendo la dimostrazione della seconda analoga e più semplice.

Calcoliamo la prima funzione lineare delle m che costituiscono la trasformazione lineare, sull'$n - pla$ somma delle due $n - ple$. Otteniamo

$$\begin{aligned} &a_1^1(x_1^1 + x_2^1) + a_2^1(x_1^2 + x_2^2) + ... + a_n^1(x_1^n + x_2^n) = \\ &= (a_1^1x_1^1 + a_2^1x_1^2 + ... + a_n^1x_1^n) + (a_1^1x_2^1 + a_2^1x_2^2 + ... + a_n^1x_2^n) \end{aligned}$$

Dunque se il primo elemento dell'$n-pla$ $\mathbf{y}_1$=$L(\mathbf{x}_1)$ lo indichiamo con $L(\mathbf{x}_1)^1$, il primo elemento dell'$n-pla$ $\mathbf{y}_2$=$L(\mathbf{x}_2)$ lo indichiamo con $L(\mathbf{x}_2)^1$ ed il primo elemento dell'$n-pla$ $L(\mathbf{x}_1+\mathbf{x}_2)$ lo indichiamo con $L(\mathbf{x}_1+\mathbf{x}_2)^1$ potremo scrivere che

$$L(\mathbf{x}_1+\mathbf{x}_2)^1 = L(\mathbf{x}_1)^1 + L(\mathbf{x}_2)^1$$

Iterando il ragionamento su tutte le funzioni lineari che costituiscono la trasformazione lineare, si ottiene il risultato enunciato.

La proprietà indicata per le trasformazioni lineari di fatto caratterizza le trasformazioni lineari tra tutte le corrispondenze tra due spazi numerici, nel senso che

Se L è una corrispondenza tra R^n ed R^m *che verifica le due proprietà*

$$\begin{cases} L(\mathbf{x}_1+\mathbf{x}_2) = L(\mathbf{x}_1) + L(\mathbf{x}_2) \\ L(\lambda\mathbf{x}) = \lambda L(\mathbf{x}) \end{cases}$$

allora L è una trasformazione lineare tra R^n *ed* R^m*, ovvero l'*$m-pla$ $L(\mathbf{x}) = \mathbf{y} = (y^1, y^2, ..., y^m)$ *corrispondente secondo L all'*$n-pla$ $\mathbf{x} = (x^1, x^2, ..., x^n)$*, si può scrivere nella forma*

$$\begin{cases} y^1 = a_1^1 x^1 + a_2^1 x^2 + ... + a_n^1 x^n \\ y^2 = a_1^2 x^1 + a_2^2 x^2 + ... + a_n^2 x^n \\ . \\ . \\ y^m = a_1^m x^1 + a_2^m x^2 + ... + a_n^m x^n \end{cases}$$

in particolare rimane individuata la matrice di L.

Omettiamo la dimostrazione di tale fatto.

Altra proprietà importante delle trasformazioni lineari, di verifica immediata è che se con $\mathbf{0}$ si indica tanto l'$n-pla$ nulla quanto l'$m-pla$ nulla, risulta

$$L(\mathbf{0}) = \mathbf{0}$$

dove a primo membro appare l'$n-pla$ nulla ed a secondo membro l'$m-pla$ nulla.

Il legame tra le trasformazioni lineari che stiamo considerando ed i sistemi di equazioni lineari considerati in precedenza, è stretto. Infatti se immaginiamo di fissare una $m-pla$, diciamola $(b^1, b^2, ..., b^m)$, e poi ci domamniamo se esistono $n-ple$ che sono trasformate dalla trasformazione lineare L nell'$m-pla$ fissata, ci accorgiamo che la risposta a tale problema consiste nello stabilire se il sistema

$$\begin{cases} b^1 = a_1^1 x^1 + a_2^1 x^2 + ... + a_n^1 x^n \\ b^2 = a_1^2 x^1 + a_2^2 x^2 + ... + a_n^2 x^n \\ . \\ . \\ b^m = a_1^m x^1 + a_2^m x^2 + ... + a_n^m x^n \end{cases}$$

è compatibile o meno. Inoltre in caso si stabilisca l'esistenza di dette $n-ple$, ovvero si stabilisca che il sistema è compatibile, la ricerca delle $n-ple$ trasformate da L nell'$m-pla$ assegnata, equivale alla ricerca delle soluzioni del sistema scritto.

2. Immagine e nucleo.

In generale l'insieme delle $m-ple$ che sono l'immagine di una qualche $n-pla$, viene detto *immagine di* L ed indicato con $Im(L)$. Invece l'insieme delle $n-ple$ che hanno per immagine l'$m-pla$ nulla, viene detto *nucleo di* L ed indicato con $ker(L)$. Vedremo in seguito che $Im(L)$ e $ker(L)$ sono sottospazi vettoriali, il primo dello spazio vettoriale delle $m-ple$, R^m, il secondo dello spazio vettoriale delle $n-ple$, R^n. Osserviamo comunque subito che poiché per ogni trasformazione lineare $L(\mathbf{0})=\mathbf{0}$ per ogni trasformazione lineare avverrà che nel suo nucleo ci sarà sempre almeno l'$n-pla$ nulla. In particolare il nucleo di una trasformazione lineare non è mai vuoto.

L'immagine ed il nucleo svolgono un ruolo fondamentale nello studio delle proprietà della trasformazione lineare L. Le prime proprietà che si cerca normalmente di stabilire per una trasformazione lineare sono le seguenti. Primo vedere quando la trasformazione L è *surgettiva*, in altre parole quando avviene che ogni $m-pla$ è immagine di una qualche $n-pla$, ossia quando lo spazio immagine coincide con tutto R^m. La seconda consiste nel vedere quando avviene che due $n-ple$ distinte hanno sempre immagini distinte, ovvero quando la trasformazione L non identifica mai $n-ple$ diverse. Trasformazioni con tale proprietà vengono dette *iniettive*. Quindi il secondo problema è di vedere quando la trasformazione lineare L è iniettiva. Infine il terzo problema consiste nello stabilire quando la trasformazione lineare è simultaneamente iniettiva e surgettiva. Quando avviene che ogni $m-pla$ proviene da una $n-pla$ ma da non più di una $n-pla$. Tali trasformazioni lineari sono dette *biunivoche*. La risposta ai suddetti problemi vine data facendo uso della nozine di *rango della trasformazione lineare L:*

Si chiama rango di L, e lo si indica con $rg(L)$, *il massimo numero di vettori indipendenti che si possono trovare in* $Im(L)$

E' bene chiarire subito, quanto vedremo tra poco, che tale rango altri non è se non il rango della matrice della trasformazione lineare L, ossia

$$rg(L) = rg(A)$$

Ciò posto la risposta ai tre quesiti precedenti è

La trasformazione lineare L è surgettiva se e solo se $rg(L) = m$

La trasformazione lineare L è iniettiva se e solo se $rg(L) = n$

La trasformazione lineare L èbiunivoca se e solo se $rg(L) = m = n$.

Inoltre si ha un'altra caratterizzazione delle trasformazioni lineari iniettive. Precisamente

Una trasformazione lineare è iniettiva se e solo se $\ker(L) = \{\mathbf{0}\}$

Omettiamo la dimostrazione dei precedenti enunciati. Ci limitiamo a semplici considerazioni, pressocché immediate, allo scopo di dare parzialmente ragione degli enunciati. La prima. Se per una data trasformazione lineare

$\ker(L) \neq \{\mathbf{0}\}$ ciò significa che nel $\ker(L)$ vi sono altre $n-ple$ oltre all'$n-pla$ nulla. Quindi oltre all' $n-pla$ nulla anche un'altra $n-pla$ ha per immagine l'$m-pla$ nulla. Due $n-ple$ diverse hanno la stessa immagine e pertanto se $\ker(L) \neq \{\mathbf{0}\}$ la trasformazione lineare non può essere iniettiva. Anche il viceversa sarebbe semplice da dimostrare, ma ne omettiamo la dimostrazione. Passiamo alla seconda considerazione. Riscriviamo la trasformazione lineare

$$\begin{cases} y^1 = a_1^1 x^1 + a_2^1 x^2 + ... + a_n^1 x^n \\ y^2 = a_1^2 x^1 + a_2^2 x^2 + ... + a_n^2 x^n \\ . \\ . \\ y^m = a_1^m x^1 + a_2^m x^2 + ... + a_n^m x^n \end{cases}$$

ed osserviamo come se ci andiamo a calcolare l'$m-pla$ che corrisponde all'$n-pla$ $(1, 0, ..., 0)$, troviamo l'$m-pla$ $(a_1^1, a_1^2, ..., a_1^m)$ ossia la prima colonna della matrice della trasformazione lineare. Andandosi a calcolare l'$m-pla$ corrispondente all'$n-pla$ $(0, 1, ..., 0)$, triveremmo l'$m-pla$ data dalla seconda colonna della matrice della trasformazione lineare. Così via, via, troveremmo tutte le colonne della matrice della trasformazione lineare fino all'ultima ottenuta in corrispondenza dell'$n-pla$ $(0, 0, ..., 1)$. Quindi le $n-ple$ costituite dalle diverse colonne della trasformazione lineare, si trovano tutte nello spazio immagine di $L : Im(L)$. Se di nuovo torniamo a guardare le equazioni che danno la trasformazione lineare, vediamo come ogni $m-pla$ che si trova nell'immagine di L è combinazione lineare delle colonne della matrice con coefficienti che sono dati dalla generica $n-pla$ di R^n ovvero da $(x^1, x^2, ..., x^n)$. Nel capitolo sugli operatori lineari, torneremo su tali argomenti, precisandoli ed approfondendoli.

3. Prodotto di trasformazioni lineari.

Sia L una trasformazione lineare di R^n in R^m e K una trasformazione lineare di, R^m, in R^p. Ovvero L trasformi un'$n-pla$ in una $m-pla$ e K trasformi un'$m-pla$ in una $p-pla$. I numeri n, m, p essendo pensati fissati ma del tutto indipendenti tra di loro, potendo indifferentemente uno essere maggiore, minore o uguale all'altro. Se indichiamo con $\mathbf{x}_1, \mathbf{x}_2$ due $n-ple$ e con λ uno scalare, risulterà

$$\begin{aligned} L(\mathbf{x}_1+\mathbf{x}_2) &= L(\mathbf{x}_1)+L(\mathbf{x}_2) \\ L(\lambda\mathbf{x}_1) &= \lambda L(\mathbf{x}_2) \end{aligned}$$

Se indichiamo con $\mathbf{y}_1, \mathbf{y}_2$ due $m-ple$ e con λ uno scalare, risulterà

$$\begin{aligned} K(\mathbf{y}_1+\mathbf{y}_2) &= K(\mathbf{y}_1)+K(\mathbf{y}_2) \\ K(\lambda\mathbf{y}_1) &= \lambda K(\mathbf{y}_2) \end{aligned}$$

Assegnata una $n-pla$ $\mathbf{x}$, $L(\mathbf{x})$ sarà un'$m-pla$ e su tale $m-pla$ potremo calcolare la trasformazione lineare K. Otterremo in tal modo la $p-pla$ $K(L(\mathbf{x}))$. Ecco che rimane definita una corrispondenza, che indicheremo con $K \circ L$,(ancora non sappiamo se si tratta o meno di una trasformazione lineare) tra R^m ed R^p, quella che all'$n-pla$ $\mathbf{x}$ associa la $p-pla$, diciamola $\mathbf{z}$, data da $K(L(\mathbf{x}))$. Ebbene

L'applicazione $K \circ L$ è una trasformazione lineare tra R^n ed R^p in quanto verifica le

$$\begin{array}{l} (K \circ L)(\mathbf{x}_1+\mathbf{x}_2) = (K \circ L)(\mathbf{x}_1) + (K \circ L)(\mathbf{x}_2) \\ (K \circ L)(\lambda \mathbf{x}_1) = \lambda (K \circ L)(\mathbf{x}_2) \end{array}$$

Tale trasformazione lineare sarà detta prodotto delle trasformazioni lineari L e K e la matrice associata a tale trasformazioni lineare sarà detta matrice prodotto della matrice associata ad L per la matrice associata a K. Se la matrice associata ad L la si indica con $A_{(m,n)}$, quella associata a K con $B_{(p,m)}$ e con $C_{(n,p)}$ si indica la matrice associata a $K \circ L$, si porrà

$$C_{(n,p)} = B_{(p,m)} A_{(m,n)}$$

Infatti per quanto riguarda la prima delle due uguaglianze, utilizzando la linearità di K e di L, si ha

$$(K \circ L)(\mathbf{x}+\mathbf{y}) = K(L(\mathbf{x}+\mathbf{y})) = K(L(\mathbf{x})+L(\mathbf{y})) = K(L(\mathbf{x}))+K(L(\mathbf{y})) = (K \circ L)(\mathbf{x})+(K \circ L)(\mathbf{y})$$

Analogamente per la seconda proprietà. Dunque l'applicazione $K \circ L$ è una trasformazione lineare di R^n in R^p.

Ci poniamo il problema di vedere come si possa costruire la matrice della trasformazione lineare $K \circ L$, detta *prodotto delle due trasformazioni lineari L e K*, a partire dalle matrici delle trasformazioni lineari L e K.

Immaginiamo che la trasformazione lineare L si scriva

$$\begin{cases} y^1 = a^1_1 x^1 + a^1_2 x^2 + ... + a^1_n x^n \\ y^2 = a^2_1 x^1 + a^2_2 x^2 + ... + a^2_n x^n \\ . \\ . \\ y^m = a^m_1 x^1 + a^m_2 x^2 + ... + a^m_n x^n \end{cases}$$

e che la trasformazione lineare K si scriva

$$\begin{cases} z^1 = b^1_1 y^1 + b^1_2 y^2 + ... + b^1_m y^m \\ z^2 = b^2_1 y^1 + b^2_2 y^2 + ... + b^2_m y^m \\ . \\ . \\ z^p = b^p_1 y^1 + b^p_2 y^2 + ... + b^p_m y^m \end{cases}$$

La trasformazione lineare $K \circ L$ si scriverà

$$\begin{cases} z^1 = c^1_1 x^1 + c^1_2 x^2 + ... + c^1_n x^n \\ z^2 = c^2_1 x^1 + c^2_2 x^2 + ... + c^2_n x^n \\ . \\ . \\ z^p = c^p_1 x^1 + c^p_2 x^2 + ... + c^p_n x^n \end{cases}$$

Il nostro problema consiste nel vedere come possa essere calcolata la matrice

$$C_{(p,n)=} \begin{pmatrix} c^1_1 & c^1_2 & . & c^1_n \\ c^2_1 & c^2_2 & . & c^2_n \\ . & . & & . \\ . & . & & . \\ c^p_1 & c^p_2 & . & c^p_n \end{pmatrix}$$

a partire dalle matrici

$$A_{(m,n)=} \begin{pmatrix} a^1_1 & a^1_2 & . & a^1_n \\ a^2_1 & a^2_2 & . & a^2_n \\ . & . & & . \\ . & . & & . \\ a^m_1 & a^m_2 & . & a^m_n \end{pmatrix}$$

e

$$B_{(p,m)=} \begin{pmatrix} b^1_1 & b^1_2 & . & b^1_m \\ b^2_1 & b^2_2 & . & b^2_m \\ . & . & & . \\ . & . & & . \\ b^p_1 & b^p_2 & . & b^p_m \end{pmatrix}$$

Cominciamo ad illustrare la costruzione della matrice prodotto dal caso più semplice tra quelli significativi: quello di m arbitrario ma $n = p = 1$. Quindi la trasformazione lineare L trasformerà *un numero*, x, in una $m-pla$, $(y^1, y^2, ..., y^m)$, e si scriverà nella forma

$$\begin{cases} y^1 = a^1 x \\ y^2 = a^2 x \\ . \\ . \\ y^m = a^m x \end{cases}$$

la trastormazione lineare K trasformerà $m-ple$ in *numeri* e si scriverà nella forma

$$z = b^1 y^1 + b^2 y^2 + ... + b^m y^m$$

mentre la trasformazione lineare $K \circ L$ consisterà in una semplice funzione lineare omogenea che fa passare da un numero ad un numero. Si scriverà nella forma

$$z = cx$$

La matrice della trasformazione lineare L è la matrice colonna

$$A_{(m,1)} = \begin{pmatrix} a^1 \\ a^2 \\ . \\ . \\ a^m \end{pmatrix}$$

La matrice della trasformazione lineare K è la matrice riga

$$B_{(1,m)} = \begin{pmatrix} b^1 & b^2 & . & . & b^m \end{pmatrix}$$

infine la matrice del prodotto $K \circ L$ è una matrice con un'unico elemento

$$C_{(1)} = (c)$$

Il problema in questo caso semplice consiste nel vedere come il numero c è collegato agli elementi delle matrici $A_{(m,1)}$ e $B_{(1,m)}$. Basta andare a sostituire al posto delle variabili che appaiono a secondo membro delle equazioni di K i loro valori calcolati a partire da x attraverso le equazioni di L. Si trova

$$z = b^1 a^1 x + b^2 a^2 x + ... + b^m a^m x = (b^1 a^1 + b^2 a^2 + ... + b^m a^m)x$$

da cui si vede che

$$c = b^1 a^1 + b^2 a^2 + ... + b^m a^m$$

e come risposta al nostro problema, nel caso particolare considerato, si trova la seguente risposta

L'unico elemento, c, della matrice della trasformazione lineare prodotto, è dato dalla somma dei prodotti degli elementi corrispondenti della matrice riga di K per quelli della matrice colonna di L.

Il procedimento utilizzato nel caso particolare continua a valere sostanzialmente immutato nel caso generale e fornisce la regola seguente

L'elemento, c^i_j che nella matrice $C_{(n,p)}$ della trasformazione prodotto $K \circ L$ si trova sulla $i-$ma riga e sulla $j-$ma colonna, si ottiene facendo la somma dei prodotti degli elementi corrispondenti della $i-$ ma riga della matrice di K per quelli della $j-$ ma colonna della matrice di L. In formule

$$c^i_j = b^i_1 a^1_j + b^i_2 a^2_j + ... + b^i_n a^n_j$$

Ci limiteremo ad applicare la formula scritta in un solo altro caso: quello di due trasformazioni lineari di R^2 in sé . Si considerino dunque le due seguenti trasformazioni lineari di R^2 in R^2

$$L : \begin{cases} y^1 = a_1^1 x^1 + a_2^1 x^2 \\ y^2 = a_1^2 x^1 + a_2^2 x^2 \end{cases}$$

e

$$K : \begin{cases} z^1 = b_1^1 y^1 + b_2^1 y^2 \\ z^2 = b_1^2 y^1 + b_2^2 y^2 \end{cases}$$

In base alla regola enunciata, la trasformazione lineare prodotto $K \circ L$, avrà come matrice la matrice seguente

$$\begin{pmatrix} b_1^1 a_1^1 + b_2^1 a_1^2 & b_1^1 a_2^1 + b_2^1 a_2^2 \\ b_1^2 a_1^1 + b_2^2 a_1^2 & b_1^2 a_2^1 + b_2^2 a_2^2 \end{pmatrix}$$

e dunque si scriverà

$$\begin{cases} z^1 = (b_1^1 a_1^1 + b_2^1 a_1^2) x^1 + (b_1^1 a_2^1 + b_2^1 a_2^2) x^2 \\ z^2 = (b_1^2 a_1^1 + b_2^2 a_1^2) x^1 + (b_1^2 a_2^1 + b_2^2 a_2^2) x^2 \end{cases}$$

come può essere failmente verificato sostituendo al posto delle y che appaiono nelle equazioni della trasformazione lineare K l'espressione che hanno in funzione delle x nella trasformazione lineare L.

4. Prodotto tra matrici. Matrice inversa.

In quello che abbiamo esposto nel precedente paragrafo, vi è implicita una definizione di prodotto tra matrici indipendente dal fatto che le matrici che si moltiplicano siano o meno matrici di trasformazioni lineari.

Prendiamo due matrici legate dalla sola condizione che *il numero delle colonne della prima sia uguale al numero delle righe della seconda*: $A_{(p,m)} = \left((a_j^i)\right)$, $B_{(m,n)} = \left((b_k^h)\right)$. Diamo la definizione seguente

Si chiama matrice prodotto della matrice $A_{(p,m)} = \left((a_j^i)\right)$ *per la matrice* $B_{(m,n)} = \left((b_k^h)\right)$, *la matrice* $C_{(p,n)} = A_{(p,m)} B_{(m,n)} = ((c_s^r))$ *il cui elemento che si trova sulla* $r-ma$ *riga ed* $s-ma$ *colonna è dato dalla somma dei prodotti degli elementi della* $r-ma$ *riga della matrice* $A_{(p,m)}$ *per i corrispondenti elementi della* $s-ma$ *colonna della matrice* $B_{(m,n)}$. *In formule*

$$c_s^r = \sum_{t=1}^{m} a_t^r b_s^t$$

Sottolineamo che due qualunque matrici non possono essere moltiplicate tra di loro a meno che non sia soddisfatta la condizione enunciata: che la prima abbia lo stesso numero di colonnedel numero di righe della seconda. Questo comporta che il problema della commutatività del prodotto neanche si pone: due matrici possono soddisfare la condizione di moltiplicabilitàse prese in un dato ordine, mentre nell'ordine inverso possono anche non essere moltiplicabili. Se si fissa l'attenzione solo sulle matrici quadrate e tutte dello stesso ordine allora queste sono sempre moltiplicabili tra di loro qualunque sia l'ordine secondo il quale le si considera. Ciononostante non è sempre vero (anzi quasi mai), che la matrice prodotto non dipenda dall'ordine delle matrici che si moltiplicano. Dunque il prodotto tra matrici pensato limitatamente all'insieme delle matrici quadrate di ordine fissato, non è commutativo. Andiamo invece a vedere quali sono le proprietà che valgono per il prodotto tra matrici. Le scriviamo qui appresso con un ovvio significato dei simboli.

$$A_{(p,m)}(B_{(m,n)}C_{(n,q)}) = (A_{(p,m)}B_{(m,n)})C_{(n,q)} \textit{associatività}$$

A tali proprietà va aggiunta la seguente che ci accingiamo ad illustrare. Comunque si fissi un intero n rimane definita una matrice tutta nulla eccetto sulla diagonale principale, dove ha tutti 1. Tale matrice si chiama matrice unità di ordine n e viene indicata con Δ_n:

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

In relazione a tale matrice vale la seguente proprietà che viene chiamata

Esistenza di una unità destra e di una unità sinistra

$$\Delta_p A_{(p,m)} = A_{(p,m)}\Delta_m = A_{(p,m)}$$

Si noti che le due matrici unità sono in generale diverse tra di loro avendo in generale ordini diversi. A meno che la matrice A non sia quadrata. In tal caso coincideranno con la matrice unità di ordine pari all'ordine della matrice A.

Ed ora veniamo ad un altro argomento importante: la *matrice inversa.*

Nel paragrafo due di questo capitolo abbiamo visto che

Una trasformazione lineare L *è biunivoca se e solo se* $rg(L) = n = m$

Ricordiamo che una corrispondenza è detta biunivoca se ogni elemento dell'insieme in cui la corrispondenza assume valori è di fatto il valore della corrispondenza su di un qualche elemento dell'insieme in cui la corrispondenza è definita (la corrispondenza è surgettiva); inoltre non deve mai avvenire che per due diversi elementi dell'insieme in cui la corrispondenza è definita, la corrispondenza assuma lo stesso valore (la corrispondenza è iniettiva). Per le corrispondenze biunivoche si può definire una corrispondenza, detta inversa della corrispondenza data, che viene ad essere definita laddove la corrispondenza data assume valori, ed assume valori laddove la corrispondenza data è definita. Vediamo come si fa nel caso di una corrispondenza tra R^n ed R^m (anche se di fatto la regola che illustreremo vale per ogni corrispondenza biunivoca).

Sia L una corrispondenza biunivoca tra R^n ed R^m (per ora non è necessario supporre che L sia una trasformazione lineare). Definiamo l'inversa di L, che indicheremo con L^{-1}, al modo seguente. Detta $\mathbf{y}$ una qualunque $m-pla$ di R^m, per l'ipotesi di surgettività di L esisterà una ed una sola $n-pla$ $\mathbf{x}$ di R^n tale che $L(\mathbf{x}) = \mathbf{y}$. Ebbene porremo $L^{-1}(\mathbf{y}) = \mathbf{x}$.

In relazione alle due corrispondenze L e la sua inversa L^{-1}, valgono le due seguenti proprietà che, anche se ovvie, è bene avere presenti

$$\begin{aligned} L(L^{-1}(\mathbf{y}) &= \mathbf{y} \\ L^{-1}(L(\mathbf{x})) &= \mathbf{x} \\ (L^{-1})^{-1} &= L \end{aligned}$$

Di fatto si potrebbe facilmente mostrare che le suddette proprietà caratterizzano l'inversa dell'applicazione L.

Sin qui tutto quanto è stato detto vale indipendentemente dal fatto che L sia o meno una trasformazione lineare. Se L è una trasformazione lineare, allora vale il seguente importante risultato

Se L è una trasformazione lineare biunivoca tra R^n ed R^m, l'inversa, L^{-1}, è una trasformazione lineare biunivoca tra R^m ed R^n.

La dimostrazione è la seguente.

Poniamo

$$L(\mathbf{x}_1) = \mathbf{y}_1, L(\mathbf{x}_2) = \mathbf{y}_2$$

conseguentemente si ha

$$L^{-1}(\mathbf{y}_1) = \mathbf{x}_1, L^{-1}(\mathbf{y}_2) = \mathbf{x}_2$$

Per la linearità di L risulta

$$L(\mathbf{x}_1 + \mathbf{x}_2) = L(\mathbf{x}_1) + L(\mathbf{x}_2) = \mathbf{y}_1 + \mathbf{y}_2$$

Quindi sarà

$$L^{-1}(\mathbf{y}_1 + \mathbf{y}_2) = \mathbf{x}_1 + \mathbf{x}_2 = L^{-1}(\mathbf{y}_1) + L^{-1}(\mathbf{y}_2)$$

che è la nostra tesi.

Ma allora l'inversa di L, L^{-1}, deve ammettere una matrice.

Indicata con A_n la matrice di L, la matrice di L^{-1} sarà indicata con A_n^{-1} e sarà detta matrice inversa della matrice A.

Ricordando che per definizione la matrice del prodotto di due trasformazioni lineari è la matrice prodotto delle matrici delle trasformazioni lineari, ricordando le prime proprietà dell'inversa di una corrispondenza (quelle che si sono dette essere di fatto valide per ogni corrispondenza biunivoca) si vede che per la matrice inversa valgono le seguenti proprietà

$$\begin{aligned} AA^{-1} &= A^{-1}A = \Delta \\ (A^{-1})^{-1} &= A \end{aligned}$$

A partire da tali proprietà è possibile il calcolo degli elementi della matrice inversa di A in funzione degli elementi di A. Omettiamo tali calcoli, e diamo direttamente ilrisultato. Precisamente

Indicando con $'a^i_j$ l'elemento che nella matrice inversa si trova sulla $i-ma$ riga e sull'$j-ma$ colonna, si ha che tale elemento è dato dalla formula

$$'a^i_j = \frac{\mathcal{A}^i_j}{\det A}$$

dove $\mathcal{A}^i_j$ è il complemento algebrico, si faccia attenzione, non dell'elemento a^i_j che nella matrice A si trova sulla $i-ma$ riga e sulla $j-ma$ colonna, ma dell'elemento a^i_j che nella matrice A si trova sulla $j-ma$ riga e sulla $i-ma$ colonna.

Nella formula scritta appare il determinante di A a denominatore. Può sorgere il dubbio se la formula sia sempre valida visto che nel caso di una matrice a determinante nullo la formula perde di senso. Ma la formula è sempre valida poiché non si dà mai il caso per una trasformazione lineare biunivoca che il determinante della sua matrice sia nullo. Infatti tornando alla caratterizzazione delle trasformazioni lineari biunivoche, vediamo che deve essere $rgA = n$ ed il rango di una matrice quadrata di ordine n è proprio uguale ad n se e solo se il determinante della matrice è non nullo.

Anche ora nei ragionamenti fatti è implicita la definizione di matrice inversa di una matrice quadrata con determinante diverso da zero, indipendentemente dal fatto che la matrice venga o meno considerata associata ad una trasformazione lineare. Precisamente

Data una matrice quadrata a determinante non nullo, A, si definisce inversa di A e la si indica con A^{-1}, la matrice il cui generico elemento $'a^i_j$, è dato n funzione della matrice A dalla formula sopra scritta.

Una matrice a determinante non nullo viene chiamata anche invertibile, o non degenere.

Naturalmente per una tale inversa continuano a valere le proprietà indicate.

Completiamo il presente paragrafo elencando altre proprietà del prodotto tra matrici, omettendone la dimostrazione. Si ricordi che con A^t si indica la trasposta della matrice A.

$$\begin{aligned}(AB)^t &= B^tA^t \\ (AB)^{-1} &= B^{-1}A^{-1} \\ (A^t)^{-1} &= (A^{-1})^t\end{aligned}$$

Le matrici nella prima uguaglianza sono matrici arbitrarie purché moltiplicabili tra di loro. Nelle altre due uguaglianze le matrici debbono essere quadrate dello stesso ordine ed entrambe a determinante diverso da zero (invertibili o non degeneri)

5. Simbolismo matriciale.

Le operazioni che abbiamo definito tra matrici permettono di semplificare notevolmente i calcoli, almeno quelli di un certo tipo. Dovremo però organizzare tali operazioni in quello che si chiama simbolismo matriciale e che andiamo ad illustrare.

Da ora in poi le $n-ple$ di numeri saranno indicate con matrici colonna. Precisamente l'$n-pla$ $\mathbf{x} = (x^1, x^2, ..., x^n)$ verrà indicata con la matrice colonna

$$X_{(n,1)} = \begin{pmatrix} x^1 \\ x^2 \\ . \\ . \\ x^n \end{pmatrix}$$

Riscriviamo un arbitrario sistema di equazioni lineari nella forma sin qui utilizzata

$$\begin{cases} a^1_1x^1 + a^1_2x^2 + ... + a^1_nx^n = b^1 \\ a^2_1x^1 + a^2_2x^2 + ... + a^2_nx^n = b^2 \\ . \\ . \\ a^m_1x^1 + a^m_2x^2 + ... + a^m_nx^n = b^m \end{cases}$$

Ricordiamo che la matrice incompleta del sistema è la matrice seguente

$$A_{(m,n)=}\begin{pmatrix} a^1_1 & a^1_2 & . & a^1_n \\ a^2_1 & a^2_2 & . & a^2_n \\ . & . & & . \\ . & . & & . \\ a^m_1 & a^m_2 & . & a^m_n \end{pmatrix}$$

Indichiamo poi con

$$\begin{pmatrix} b^1 \\ b^2 \\ . \\ . \\ b^m \end{pmatrix}$$

la colonna dei termini noti del sistema. Con tali simboli, facendo uso delle definizioni delle operazioni tra matrici, il sistema stesso può scriversi

$$A_{(m,n)}X_{(n,1)} = B_{(m,1)}$$

Nel caso di un sistema quadrato con matrice dei coefficienti a determinante diverso da zero, si può considerare la matrice inversa, A_n^{-1}, della matrice dei coefficienti. Moltiplicando a sinistra ambo i membri dell'uguaglianza matriciale che esprime il sistema in questo caso, cioè dell'uguaglianza

$$A_n X_{(n,1)} = B_{(n,1)}$$

per la matrice A_n^{-1}, otteniamo

$$X_{(n,1)} = A_n^{-1}B_{(n,1)}$$

Come si può controllare subito la formula scritta, una volta esplicitata, fornisce la formula nota della soluzione del teorema con la regola di Cramer. Notiamo che questa maniera semplice per il calcolo della soluzione del sistema, vale solo nel caso detto: un sistma quadrato con matrice dei coefficienti a determinante diverso da zero.

Il simbolismo matriciale può essere utilizzato anche nel caso delle trasformazioni lineari. Scriviamo una generica trasformazione lineare nel modo in cui siamo stati abituati

$$\begin{cases} y^1 = a^1_1x^1 + a^1_2x^2 + ... + a^1_nx^n \\ y^2 = a^2_1x^1 + a^2_2x^2 + ... + a^2_nx^n \\ . \\ . \\ y^m = a^m_1x^1 + a^m_2x^2 + ... + a^m_nx^n \end{cases}$$

Indichiamo con $A_{(m,n)}$ la matrice della trasformazione lineare

$$A_{(m,n)} = \begin{pmatrix} a^1_1 & a^1_2 & . & a^1_n \\ a^2_1 & a^2_2 & . & a^2_n \\ . & . & & . \\ . & . & & . \\ a^m_1 & a^m_2 & . & a^m_n \end{pmatrix}$$

Indichiamo poi $X_{(n,1)}$ la matrice dell'$n-pla$ $\mathbf{x}$ e con $Y_{(m,1)}$ la matrice dell'$m-$ *pla* $\mathbf{y}$

$$X_{(n,1)} = \begin{pmatrix} x^1 \\ x^2 \\ . \\ . \\ x^n \end{pmatrix}, \; Y_{(m,1)} = \begin{pmatrix} y^1 \\ y^2 \\ . \\ . \\ y^m \end{pmatrix}$$

Una verifica diretta mostra che la trasformazione lineare può scriversi.

$$A_{(m,n)}X_{(n,1)} = Y_{(m,1)}$$

Inoltre se ricordiamo che il nucleo della trasformazione lineare è l'insieme delle $n-pl$e che sono trasformate dalla trasformazione lineare nell'$m-pla$ nulla, possiamo dire che l'equazione martriciale

$$A_{(m,n)}X_{(n,1)} = 0_{(m,1)}$$

fornisce le *equazioni del nucleo della trasformazione lineare*. (A dire il vero per equazioni del nucleo della trasformazione lineare non si intendono proprio le equazioni scritte, ma quelle consistenti solo nelle equazioni indipendenti in numero massimo del sistema scritto).

Useremo il simbolismo matriciale anche nella teoria degli spazi vettoriali che costituirà l'argomento del prossimo capitolo. Avremo però bisogno di introdurre oltre alle matrici sin qui considerate, fatte di numeri, matrici numeriche, anche matrici fatte di vettori, matrici vettoriali. Per questo il simbolismo matriciale nel contesto degli spazi vettoriali, verrà rinviato al capitolo che segue.

SPAZI VETTORIALI

1. Preliminari

Si è visto in corsi precedenti che nell'insieme dei vettori dello spazio è definita tanto un'operazione di somma tra vettori, quanto un'operazione di prodotto di un vettore per uno scalare. Noi abbiamo visto nel precedente capitolo che una somma può essere definita anche tra $n-ple$ di numeri reali per le quali può essere definito anche un prodotto per uno scalare. Operazioni di somma e prodotti per scalari si definiscono anche per altri enti matematici. Limitiamoci a due esempi.

Primo esempio. Si fissino due interi positivi, m, n, e si considerino solo le matrici ad m righe ed n colonne. Indichiamo l'insieme di tali matrici con $M_{(m,n)}$. Fissate in $M_{(m,n)}$ due matrici, diciamole $A_{(m,n)} = ((a_{ij}))\, , B_{(m,n)} = ((b_{ij}))$, definiamo la loro somma come la matrice $C_{(m,n)} = ((c_{ij}))$, anche essa ad m righe ed n colonne che, ha sulla $i-ma$ riga e l'$j-ma$ colonna la somma degli elementi che si trovano sulla $i-ma$ riga e l'$j-ma$ colonna dele due matrici. In formule

$$c_{ij} = a_{ij} + b_{ij}$$

Definiamo poi prodotto di una matrice $A_{(m,n)} = ((a_{ij}))$ per uno scalare λ, la matrice, sempre ad m righe ed n colonne che indicheremo con $\lambda A_{(m,n)}$, avente come generico elemento il generico elemento di $A_{(m,n)}$, moltiplicato per λ:

$$\lambda A_{(m,n)} = ((\lambda a_{ij}))$$

Nell'insieme $M_{(m,n)}$ la matrice i cui elementi sono nulli è detta matrice nulla ed indicata con $O_{(m,n)}$. Con $-A_{(m,n)}$ si indica la matrice, detta l'opposta della matrice $A_{(m,n_)}$, i cui elementi sono gli opposti dei corrisponenti elementi della matrice $A_{(m,n_)}$.

Cogliamo l'occasione per fare l'osservazione seguente, di fatto non collegata alla struttura di spazio vettoriale dello spazio delle matrici ad m righe ed n colonne.

Le operazioni di somma tra matrici e di prodotto di una matrice per uno scalare sono legate alla precedente definizione di prodotto tra matrici dalle due seguenti uguaglianze

$$\begin{aligned} (A_{(p,m)} + B_{(p,m)})C_{(m,q)} &= A_{(p,m)}C_{(m,q)} + B_{(p,m)}C_{(m,q)} \text{ \textit{distributività}} \\ \lambda(A_{(p,m)}C_{(m,q)}) &= (\lambda A_{(p,m)})C_{(m,q)} = A_{(p,m)}(\lambda C_{(m,q)}) \end{aligned}$$

Inoltre la definizione di prodotto di una matrice per uno scalare è legata alla nozione di determinante, nel caso di matrici quadrate, dall'uguaglianza

$$\det(\lambda A_n) = \lambda^n \det A_n$$

Secondo esempio. Si consideri l'insieme delle funzioni reali di variabile reale, continue in fissato intervallo I. Indicato tale insieme con $c^0(I)$ e prese due funzioni f, g di $c^0(I)$, si definisca loro somma la funzione, indicata con $f + g$, definita in I dall'uguaglianza

$$(f + g)(x) = f(x) + g(x)$$

E' noto che anche $f + g$ è una funzione continua in I, ovvero appartiene a $c^0(I)$. Inoltre dato uno scalare λ ed assegnata una funzione f, si definisce prodotto della funzione per lo scalare la funzione, indicata con λf, anche essa continua in I

$$(\lambda f)(x) = \lambda f(x)$$

Tra le funzioni continue in I vi è la funzione costantemente nulla. Tale funzione viene detta la funzione nulla ed indicata con 0. Inoltre assieme ad ogni funzione continua f, vi è la funzione che assume su ogni punto dell'intervallo di definizione come valore l'opposto del valore assunto sullo stesso punto da f. Tale funzione viene detta la funzione opposta di f ed indicata con $-f$.

Quello che avviene in tutti i casi indicati, già visto nel caso dei vettori dello spazio geometrico e delle $n - ple$ di numeri reali, e che è comunque di verifica immediata anche per i due esempi appena riportati, è che per le operazioni di somma e di prodotto per scalari, valgono le seguenti proprietà

$(\lambda\mu)\mathbf{u} = \lambda(\mu\mathbf{u})$*associatività rispetto al prodotto*
$(\lambda + \mu)\mathbf{u} = \lambda\mathbf{u} + \mu\mathbf{u}$*distributività del prodotto rispetto alla somma*
$1\mathbf{u} = \mathbf{u}$*elemento neutro rispetto al prodotto*
$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$*associatività della somma*
$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$*commutatività della somma*
$\mathbf{u} + \mathbf{0} = \mathbf{u}$*elemento neutro rispetto alla somma*
$\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$*opposto nella somma*
$\lambda(\mathbf{u} + \mathbf{v}) = \lambda\mathbf{u} + \lambda\mathbf{v}$*distributività del prodotto rispetto alla somma*

dove in quello che precede con le lettere in grassetto si indicano gli elementi dei diversi insiemi considerati, o vettori geometrici, o $n - ple$, o matrici o funzioni.

In generale

Se in un insieme sono definite due operazioni, una che associa un elemento dell'insieme a due elementi dell'insieme, un'altra che associa un elemento dell'insieme ad un elemento dell'insieme e ad uno scalare, in modo che siano soddisfatte tutte le proprietà sopra scritte, tale insieme viene detto spazio vettoriale.

2. Combinazioni lineari. Dipendenza lineare. Basi. Coordinate.

Da ora in poi indicheremo gli elementi di un qualunque spazio vettoriale con delle lettere grassette quali $\mathbf{u}, \mathbf{v}, \mathbf{w}, ...$Gli scalari saranno indicaticon delle lettere greche quali, $\lambda, \mu, \nu, ...$ Le prime nozioni che daremo saranno una ripetizione di nozioni analoghe date in precedenza relativamente ai casi considerati. Pertanto non ci soffermeremo su tali nozioni in modo particolare. Con V indicheremo un arbitrario spazio vettoriale.

Dati p vettori di V, $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p$ *e p* (≥ 1) *scalari* , $\lambda_1, \lambda_2, ..., \lambda_p$, *il vettore di* V

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + ...+\lambda_p\mathbf{v}_p$$

è detto combinazione lineare di $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p$ *con coefficienti* $\lambda_1, \lambda_2, ..., \lambda_p$.

Se tra q (≥ 2) *vettori ve ne è uno combinazione lineare degli altri, i* q *vettori sono detti dipendenti. Un vettore è detto dipendente se e solo se è il vettore nullo.*

La definizione data di vettori dipendenti equivale alla seguente

q (≥ 1) *vettori sono dipendenti se e solo il vettore nullo risulta essere loro combinazione lineare con coefficienti non tutti nulli.*

La dimostrazione di tale fatto è identica a quella data nel caso delle $n-ple$ e non la ripetiamo.

Ora due nuove definizioni

Un insieme di vettori è detto essere un sistema di generatori dello spazio vettoriale se ogni vettore dello spazio è esprimibile come combinazione lineare dei vettori dell'insieme.

Si osservi che nel precedente enunciato non si esclude che l'insieme sia fatto da infiniti vettori.

Un sistema di generatori è detto essere una base se, comunque si prenda un numero finito di generatori, questi sono sempre indipendenti.

Si può dimostrare che ogni spazio vettoriale ammette una base, ma tale base non è necessariamente costituita da un numero finito di vettori.

In tutto quello che segue considereremo solo spazi vettoriali con basi costituite da un numero finito di vettori. Tali spazi vettoriali vengono detti di dimensione finita.

Nel seguito considereremo solo spazi vettoriali di dimensione finita.

Sia V uno spazio vettoriale di dimensione finita, quindi con una base costituita da un numero finito di vettori. Si dimostra che

Se uno spazio vettoriale ammette anche una sola base costituita da un numero finito di vettori, tutte le basi (che sono sempre infinite) non solo hanno sempre un numero finito di vettori, ma in più tale numero è sempre lo stesso per ogni base.

Il comune numero di vettori che costituiscono le basi di uno spazio vettoriale di dimensione finita viene detto dimensione dello spazio vettoriale.

Se si indica con n la dimensione di uno spazio vettoriale V di dimensione finita lo spazio vettoriale stesso viene indicato con V^n.

La nozione di base è molto importante. Infatti permetterà di ricondurre i conti relativi ad un qualunque spazio vettoriale, al caso dello spazio vettoriale numerico delle $n-ple$ di numeri. Vogliamo chiarire tale fatto con quello che segue.

Sia $\mathbf{E} = (e_1, e_2, ..., e_n)$ una base di uno spazio vettoriale di dimensione finita $n : V^n$. Ogni vettore $\mathbf{v} \in V^n$ si può esprimere come combinazione lineare dei vettori di $\mathbf{E}$

$$\mathbf{v} = v^1\mathbf{e}_1 + v^2\mathbf{e}_2 + ... + v^n\mathbf{e}_n$$

questo perché in particolare una base è un sistema di generatori e quindi ogni vettore dello spazio è esprimibile come loro combinazione lineare. Inoltre non esistono due combinazioni lineari dei vettori della base che esprimono lo stesso vettore con coefficienti diversi, ovvero

I coefficienti della combinazione lineare dei vettori di una base che esprime un dato vettore, sono individuati dal vettore stesso una volta data la base.

Infatti se risultasse tanto

$$\mathbf{v} = v^1\mathbf{e}_1 + v^2\mathbf{e}_2 + ... + v^n\mathbf{e}_n$$

quanto

$$\mathbf{v} = w^1\mathbf{e}_1 + w^2\mathbf{e}_2 + ... + w^n\mathbf{e}_n$$

facendo la differenza a membro a membro delle due uguaglianze, si otterrebbe

$$\mathbf{0} = (v^1 - w^1)\mathbf{e}_1 + (v^2 - w^2)\mathbf{e}_2 + ... + (v^n - w^n)\mathbf{e}_n$$

e siccome i vettori della base sono indipendenti, l'unica loro combinazione lineare che dà il vettore nullo è quella con tutti i coefficienti nulli, quindi si deve avere

$$v^1 = w^1, v^2 = w^2, ..., v^n = w^n$$

cioè la tesi.

Riassumendo possiamo dire che, se si fissa una base in uno spazio vettoriale, per ogni vettore rimangono individuati i coefficienti dell'unica combinazione lineare che esprime il vettore come combinazione lineare dei vettori della base. Tali coefficienti sono tanti quanti gli elementi della base, cioè sono pari alla dimensione dello spazio vettoriale, e vengono detti *coordinate del vettore nella base fissata.*

Il ragionamento fatto permette di stabilire, per ogni spazio vettoriale di dimensione n, qualora in tale spazio sia stata fissata una base, una corrispondenza biunivoca tra i vettori dello spazio e le $n-ple$ di numeri reali ovvero i vettori dello spazio vettoriale numerico delle $n-ple$ di numeri. Segnaliamo subito che, dato lo spazio vettoriale, la corrispondenza che ad un vettore associa le sue coordinate, dipende in maniera essenziale dalla base fissata. Vedremo in seguito come cambia la corrispondenza al variare della base, cioè come cambiano le coordinate di un vettore al variare della base.

Ora vogliamo segnalare la proprietà fondamentale della corrispondenza vettore-coordinate.

Assegnato uno spazio vettoriale V^n *e scelta in tale spazio una base* $\mathbf{E} = (e_1, e_2, ..., e_n)$, *le coordinate del vettore* $\mathbf{u} + \mathbf{v}$ *somma dei due vettori* $\mathbf{u}, \mathbf{v}$, *sono la somma delle coordinate dei due vettori* $\mathbf{u}, \mathbf{v}$. *Analogamente il vettore*

$\lambda \mathbf{v}$, *prodotto del vettore* $\mathbf{v}$ *per lo scalare* λ, *ha per coordinate le coordinate di* $\mathbf{v}$ *moltiplicate per lo scalare* λ.

Se indichiamo con $\varphi_{\mathbf{E}}$ la corrispondenza che ad ogni vettore $\mathbf{v}$ di V^n associa l'$n-pla$ delle coordinate $\mathbf{v}$ nella base $\mathbf{E}$, il risultato esposto può esprimersi nelle due formule seguenti

$$\varphi_{\mathbf{E}}(\mathbf{u}+\mathbf{v}) = \varphi_{\mathbf{E}}(\mathbf{u}) + \varphi_{\mathbf{E}}(\mathbf{v})$$
$$\varphi_{\mathbf{E}}(\lambda\mathbf{v}) = \lambda\varphi_{\mathbf{E}}(\mathbf{v})$$

Vederemo in seguito che corrispondenze tra spazi vettoriali diversi, come è quella tra i vettori di V^n ed i vettori di R^n, lo spazio vettoriale delle $n-ple$, verificanti condizioni come quella ora scritta, vengono detto omomorfismi tra spazi vettoriali od anche operatori lineari.

L'importanza del risultato consiste nel fatto che ogni problema di dipendenza lineare tra vettori di uno spazio vettoriale qualunque, può essere studiata studiando la dipendenza lineare delle loro coordinate, dipendenza che, come oramai sappiamo, si studia facendo uso della nozione di rango. Infatti dalla proprietà appena indicata segue senza difficoltà che

Se un vettore è combinazione lineare di altri vettori con certi coefficienti, le sue coordinate sono combinazione lineare delle coordinate dei vettori con gli stessi coefficienti

e quindi discende subito anche che

p vettori di un qualunque spazio vettoriale (di dimensione finita) sono dipendenti se e solo se tali sono le loro coordinate relative ad una fissata base.

In particolare se diamo p vettori $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p$, e se indichiamo con

$$(v_1^1, v_1^2, ..., v_1^n), (v_2^1, v_2^2, ..., v_2^p), ..., (v_p^1, v_p^2, ..., v_p^p)$$

le loro coordinate, abbiamo che

I p vettori sono dipendenti se e solo se

$$rg \begin{pmatrix} v_1^1 & v_2^1 & . & . & v_p^1 \\ v_1^2 & v_2^2 & . & . & v_p^2 \\ . & . & & & . \\ v_p^n & v_p^n & . & . & v_p^n \end{pmatrix} < p$$

Da questo fatto discende subito che

In un qualunque spazio vettoriale di dimensione n, più di n vettori sono sempre dipendenti.

Nel seguito chiameremo

Rango di un certo insieme di vettori di un assegnato spazio vettoriale, il massimo numero di vettori indipendenti dell'insieme.

E' chiaro che per quanto appena detto

Il rango di un insieme di vettori uguaglia il rango dell'insieme delle $n-$ple coordinate dei vettori relativamente ad una fissata base.

Ora che abbiamo illustrato l'importanza della nozione di base, torniamo a tale nozione, per dare due caratterizzazioni delle basi che torneranno utili e che, in generale, contribuiscono a farsi un'idea di quali siano le basi. Le caratterizzazioni sono le seguenti:

Prima caratterizzazione delle basi: Le basi di uno spazio vettoriale di dimenzione finita n, sono tutte e sole le n-ple di vettori indipendenti.

Seconda caratterizzazione delle basi: Se $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n)$ è una base di uno spazio vettoriale V^n, le altre basi di V^n sono tutte e sole le $n-ple$ di vettori che si ottengono dalla base fissata tramite la formula

$$\begin{cases} \mathbf{f}_1 = c_1^1\mathbf{e}_1 + c_1^2\mathbf{e}_2 + ... + c_1^n\mathbf{e}_n \\ \mathbf{f}_2 = c_2^1\mathbf{e}_1 + c_2^2\mathbf{e}_2 + ... + c_2^n\mathbf{e}_n \\ . \\ . \\ \mathbf{f}_n = c_n^1\mathbf{e}_1 + c_n^2\mathbf{e}_2 + ... + c_n^n\mathbf{e}_n \end{cases}$$

con

$$C_n = \begin{pmatrix} c_1^1 & c_2^1 & . & c_n^1 \\ c_1^2 & c_2^2 & . & c_n^2 \\ . & & & . \\ . & & & . \\ c_1^n & c_2^n & . & c_n^n \end{pmatrix}$$

matrice a determinante diverso da zero. In particolare due basi sono sempre legate da tale formula. La formula scritta viene detta formula di cambiamento di base: dalla base $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n)$ alla base $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_n)$. La matrice C_n è detta matrice del cambiamento di base.

La dimostrazione della prima proposizione è la seguente. Innanzi tutto siccome per definizione le basi sono fatte da vettori indipendenti ed è stato detto (senza dimostrazione) che tutte le basi hanno lo stesso numero n di vettori, che per definizione è la dimensione dello spazio vettoriale, è chiaro che ogni base è un'$n-pla$ di vettori indipendenti. Il punto è dimostrare il viceversa e far vedere che, presi n vettori indipendenti, questi formano una base, ovvero oltre ad essere indipendenti formano un sistema di generatori dello spazio. Indichiamo dunque con $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ n vettori indipendenti dello spazio V^n e facciamo vedere che ogni vettore di V^n si può esprimere come loro combinazione lineare. Se $\mathbf{v}$ è un qualunque vettore di V^n, $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n, \mathbf{v}$, saranno $n+1$ vettori. Sappiamo che in uno spazio vettoriale di dimensione n, $n+1$ vettori sono sempre dipendenti. Quindi uno di loro sarà combinazione lineare dei rimanenti. Vogliamo far vedere che proprio $\mathbf{v}$ si può esprimere come combinazione lineare degl altri. Una condizione equivalente alla dipendenza lineare è l'esistenza di una combinazione lineare dei vettori dati, con coefficienti non tutti nulli uguale al vettore nullo. Sia

$$\alpha^1\mathbf{x}_1 + \alpha^2\mathbf{x}_2 + ...+\alpha^n\mathbf{x}_n + \lambda\mathbf{v} = \mathbf{0}$$

una tale combinazione lineare. Se fosse $\lambda = 0$ si avrebbe

$$\alpha^1\mathbf{x}_1 + \alpha^2\mathbf{x}_2 + ...+\alpha^n\mathbf{x}_n= \mathbf{0}$$

con almeno un altro dei coefficienti rimanenti diverso da zero. Ma ciò non può essere perché in tal caso avremmo una combinazione lineare d i $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$, a coefficienti non tutti, nulli uguale al vettore nullo ed i vettori $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$, sarebbero dipendenti contro l'ipotesi. Dunque λ non è nullo. Pertanto dividendo ambo i membri di

$$\alpha^1\mathbf{x}_1 + \alpha^2\mathbf{x}_2 + ...+\alpha^n\mathbf{x}_n + \lambda\mathbf{v} = \mathbf{0}$$

per λ e lasciando $\mathbf{v}$ a primo membro, si viene ad esprimere $\mathbf{v}$ come combinazione lineare degli n vettori assegnati. La tesi.

La dimostrazione della seconda caratterizzazione è la seguente.

Se $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n)$ è una base allora la matrice C_n è la matrice le cui colonne sono le ccordinate dei vettori $\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_n$ rispetto alla base $\mathbf{E}$. Essendo per ipotesi il determinante della matrice C_n diverso da zero, il rango di C_n sarà n. Ma allora gli n vettori $\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_n$ saranno indipendenti. Per la proposizione appena dimostrata, essendo n vettori indipendenti, saranno una base. Viceversa dati n vettori $\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_n$, intanto ognuno di essi si potrà esprimere come combinazione dei vettori $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n$ giacché questi per ipotesi formano una base. Quindi intanto potremo scrivere

$$\begin{cases} \mathbf{f}_1 = c_1^1\mathbf{e}_1+c_1^2\mathbf{e}_2 + ...+c_1^n\mathbf{e}_n \\ \mathbf{f}_2 = c_2^1\mathbf{e}_1+c_2^2\mathbf{e}_2 + ...+c_2^n\mathbf{e}_n \\ . \\ . \\ \mathbf{f}_n = c_n^1\mathbf{e}_1+c_n^2\mathbf{e}_2 + ...+c_n^n\mathbf{e}_n \end{cases}$$

e le colonne della matrice C_n saranno le coordinate dei vettori $\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_n$ nella base $\mathbf{E}$. Se poi si fa l'ipotesi che i vettori $\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_n$, siano una base, in particolare dovranno essere indipendenti e quindi la matrice delle loro coordinate, C_n, avrà determinante diverso da zero.

3. Cambiamenti di base e cambiamenti di cordinate. Simbolismo matriciale.

Si è appena visto che, date due basi di un fissato spazio vettoriale, $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n)$ ed $\mathbf{F}$ =$(\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_n)$, il legame che passa tra le due basi è dato dalle seguenti formule

$$\begin{cases} \mathbf{f}_1 = c_1^1\mathbf{e}_1 + c_1^2\mathbf{e}_2 + ... + c_1^n\mathbf{e}_n \\ \mathbf{f}_2 = c_2^1\mathbf{e}_1 + c_2^2\mathbf{e}_2 + ... + c_2^n\mathbf{e}_n \\ . \\ . \\ \mathbf{f}_n = c_n^1\mathbf{e}_1 + c_n^2\mathbf{e}_2 + ... + c_n^n\mathbf{e}_n \end{cases}$$

dove

$$C_n = \begin{pmatrix} c_1^1 & c_2^1 & . & c_n^1 \\ c_1^2 & c_2^2 & . & c_n^2 \\ . & . & . & . \\ . & . & . & . \\ c_1^n & c_2^n & . & c_n^n \end{pmatrix}$$

è quella che si chiama matrice del cambiamento di base. Sottolineiamo il fatto che le colonne di C_n sono le coordinate dei vettori della nuova base calcolate rispetto alla vecchia base. In particolare tutte le colonne di C_n son indipendenti in quanto coordinate di vettori indipendenti, e pertantoil determinante di C_n è diverso da zero.

Il problema al quale ci accingiamo a dare risposta è come cambino le coordinate di un vettore fissato genericamente, quando la base nella quale tali vengono calcolate, cambi. Precisamente ci domandiamo che relazione passi tra le coordinate, diciamole $(v^1, v^2, ..., v^n)$ di un dato vettore $\mathbf{v}$ nella base $\mathbf{E}$ e le coordinate, diciamole $(w^1, w^2, ..., w^n)$ dello stesso vettore, nella base $\mathbf{F}$. Proviamo a ricavare detta relazione facendo uso del simbolismo matriciale per i vettori, che adesso introduciamo. Indichiamo l'$n - pla$ delle coordinate del vettore $\mathbf{v}$ nella base $\mathbf{E}$ con la matrice colonna

$$V_{(n,1)} = \begin{pmatrix} v^1 \\ v^2 \\ . \\ . \\ v^n \end{pmatrix}$$

Indichiamo le coordinate, sempre di $\mathbf{v}$, ma nella base $\mathbf{F}$ con la matrice colonna

$$W_{(n,1)} = \begin{pmatrix} w^1 \\ w^2 \\ . \\ . \\ w^n \end{pmatrix}$$

Indichiamo la base $\mathbf{E}$ con la matrice vettoriale riga

$$\mathbf{E}_{(1,n)} = (\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n)$$

e la base $\mathbf{F}$ con la matrice vettoriale riga

$$\mathbf{F}_{(1,n)} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_n)$$

Infine identifichiamo il vettore $\mathbf{v}$ con la matrice vettoriale ad una riga ed na colonna $\mathbf{v}_{(1)} = (\mathbf{v})$. Con tali simbolismi il vettore $\mathbf{v}$ potrà tanto scriversi

$$\mathbf{v} = \mathbf{v}_{(1)} = \mathbf{E}_{(1,n)} V_{(n,1)}$$

quanto scrversi

$$\mathbf{v} = \mathbf{v}_{(1)} = \mathbf{F}_{(1,n)} W_{(n,1)}$$

mentre la formula di cambiamento di base si scriverà

$$\mathbf{F}_{(1,n)} = \mathbf{E}_{(1,n)} C_n$$

Avremo le uguaglianze

$$\mathbf{v} = \mathbf{F}_{(1,n)} W_{(n,1)} = \mathbf{E}_{(1,n)} C_n W_{(n,1)}$$

Quindi

$$\mathbf{v} = \mathbf{E}_{(1,n)} V_{(n,1)} = \mathbf{E}_{(1,n)} C_n W_{(n,1)}$$

Siccome le coordinate di un vettore in una fissata base sono individuate, dovranno essere uguali le matrici delle coordinate del vettore $\mathbf{v}$ che appaiono nel primo e nel secondo membro, in quanto sono coordinate dello stesso vettore nella stessa base. Si dovrà pertanto avere

$$V_{(n,1)} = C_n W_{(n,1)}$$

Da queste, moltiplicando ambo i membri a sinistra per l'inversa, C_n^{-1}, della matrice C_n, certo esistente stando l'invertibilità di C_n, otteniamo la relazione cercata

$$W_{(n,1)} = C_n^{-1} V_{(n,1)}$$

Per maggior chiarezza esplicitiamo le uguaglianze tra numeri implicite nell'uguaglianza matriciale scritta. Troviamo

$$\begin{cases} w_1 =' c_1^1 v_1 +' c_1^2 v_2 + ... +' c_1^n v_n \\ w_2 =' c_2^1 v_1 +' c_2^2 v_2 + ... +' c_2^n v_n \\ . \\ . \\ w_n =' c_n^1 v_1 +' c_n^2 v_2 + ... +' c_n^n v_n \end{cases}$$

dove con ${}'c^i_j$, indichiamo il generco elementi dell'*inversa* della matrice C_n.

Possiamo enunciare

Le formule che fanno passare dalle coordinate di un dato vettore nella base $\mathbf{E}_{(1,n)}$ alle coordinate dello stesso vettore nella base $\mathbf{F}_{(1,n)} = \mathbf{E}_{(1,n)}C_n$, sono date dalla trasformazione lineare (omogenea ed invertibile) $W_{(n,1)} = C_n^{-1}V_{(n,1)}$, che ha come matrice la matrice C_n^{-1}, inversa della matrice C_n, del cambiamento di base.

Notiamo come dall'uguaglianza

$$\mathbf{F}_{(1,n)} = \mathbf{E}_{(1,n)}C_n$$

l'inversa di moltiplicando a destra per l'inversa di C_n, si ottiene

$$\mathbf{E}_{(1,n)} = \mathbf{F}_{(1,n)}C_n^{-1}$$

e si deduce che

Se C_n è la matrice del cambiamento di base dalla base $\mathbf{E}_{(1,n)}$ alla base $\mathbf{F}_{(1,n)}$, allora la matrice del cambiamento di base inverso, dalla base $\mathbf{F}_{(1,n)}$ alla base $\mathbf{E}_{(1,n)}$, ha come matrice di cambiamento di base l'inversa, C_n^{-1}, della matrice, C_n, del cambiamento di base da $\mathbf{E}_{(1,n)}$ ad $\mathbf{F}_{(1,n)}$.

4. Sottospazi vettoriali. Somme dirette.

Cominciamo con il rivolgere l'attenzione a dei sottoinsiemi particolari di spazi vettoriali a noi noti. Consideriamo lo spazio vettoriale V^3, lo spazio vettoriale dei vettori dello spazio geometrico tridimensionale. E' noto che l'insieme dei vettori paralleli ad una direzione assegnata, d, sia costituito da vettori tutti tra loro proporzionali. Questo comporta che, tanto se si moltiplica un vettore parallelo alla direzione assegnata per un fattore di proporzionalità, quanto se si sommano vettori paralleli alla direzione fissata, si ottengano ancora un vettori paralleli alla direzione fissata e che pertanto lo stesso avvenga facendo combinazioni lineari di vettori tutti paralleli sempre alla direzione d. La stessa situazione si ripresenta nel caso dei vettori paralleli ad un piano π: la somma ed il prodotto per scalari di vettori paralleli tutti a π è sempre un vettore parallelo a π e quindi tale risulta anche essere una qualunque combinazione lineare di vettori paralleli al piano. Insomma tanto i vettori paralleli ad un piano fissato, quanto i vettori paralleli ad una retta fissata, costituiscono sempre *sottoinsiemi chiusi* rispetto alle combinazioni lineari intendendo con ciò che, facendo combinazioni lineari di vettori di tali insiemi, non si esce da tali insiemi. Il fatto illustrato equivale al fatto seguente. Se si fissano i vettori paralleli ad un piano, o quelli paralleli ad una retta, allorché si restringono le operazioni di somma e di prodotto per scalari ai sottoinsiemi indicati, ognuno di tali sottonsiemi viene a costituire uno spazio vettoriale proprio rispetto alla somma ed al prodotto per scalari definiti per tutti i vettori dello spazio. Per capire meglio la situazione si consideri un sottoinsieme diverso dai precedenti, quindi *non chiuso* rispetto alla somma ed al prodotto per scalari, ovvero non chiuso rispetto alle combinazioni lineari, il che è lo stesso. In tal caso presi comunque due vettori del sottoinsieme si potrà ugualmente fare la loro somma od anche una loro combinazione lineare, ma per il fatto che il sottoinsieme non è chiuso rispetto alle combinazioni lineari, il risultato della somma o della combinazione lineare non sarà sempre un vettore del sottoinsieme. Pertanto il sottoinsieme, con le operazioni di somma tra vettori e di prodotto di un vettore per uno scalare, non sarà uno spazio vettoriale.

Passiamo ad un altro caso: quello delle soluzioni di un sistema omogeneo. Sappiamo che, assegnato un sistema omogeneo in n incognite, le soluzioni del sistema costituiscono un sottoinsieme, dello spazio vettoriale R^n, chiuso rispetto alla somma tra $n - ple$ e rispetto al prodotto di un'$n - pla$ per uno scalare. Ovvero anche, chiuso rispetto alle combinazioni lineari. Di nuovo questo comporta il fatto equivalente che, se nell'insieme delle soluzioni di un sistema omogeneo, si considerano le operazioni di somma tra $n - ple$ e di prodotto di un'$n - pla$ per uno scalare, tale insieme viene ad essere uno spazio vettoriale. Tanto nel caso precedente, quello dei vettori paralleli ad una retta o paralleli ad un piano, quanto nel caso delle soluzioni di un sistema omogeneo, si parla di *sottospazio vettoriale.* Nel primo caso si tratta di sottospazi vettoriali dello spazio vettoriale dei vettori geometrici V^3, nel secondo caso di sottospazi vettoriali dello spazio vettoriale numerico R^n.

Veniamo alla definizione generale di sottospazio vettoriale di uno spazio vettoriale.

Dato uno spazio vettoriale V, un sottoinsieme U di V è detto sottospazio vettoriale di V, se ogni combinazione lineare di vettori di U è ancora un vettore di U (si usa dire: se U è chiuso rispetto alla somma ed al prodotto)

E' immediato constatare che

Un sottoinsieme U di uno spazio vettoriale V è un sottospazio vetoriale di V se e solo se U è uno spazio vettoriale rispetto alle operazioni di somma tra vettori e di prodotto di un vettore per uno scalare, definite in V.

Ci limitiamo ad accennare, senza dimostrazioni ai fatti seguenti.

In V^3 non vi sono altri sottospazi vettoriali oltre a quelli indicati, ovvero diversi da quelli costituiti o dai vettori paralleli ad una retta, o dai vettori paralleli ad un piano.

Questo se si escludono due sottospazi vettoriali, che esistono sempre qualunque sia lo spazio vettoriale considerato, e che vengono detti *sottospazi vettoriali banali*, ossia lo spazio vettoriale tutto intero, pensato come sottoinsieme di se stesso, e lo spazio vettoriale costituito dal solo vettore nullo, che ovviamente verifica la definizione di sottospazio vettoriale. Inoltre

In R^n non vi sono sottospazi vettoriali diversi da quelli costituiti dalle soluzioni di un sistema omogeneo.

Quest'ultimo risultato è, nel senso che ora chiariremo, la situazione generale per un arbitrario sottospazio vettoriale, U, di un arbitrario spazio vettoriale, V, purché lo spazio vettoriale V abbia dimensione finita: $V = V^n$. Precisamente si ha il risultato seguente (non dimostreremo neanche questo)

Sia U un arbitrario sottospazio vettoriale di uno spazio vettoriale di dimensione finita n: V^n. Sia $\mathbf{E}$ una base di V^n fissata. I vettori di U sono tutti e soli i vettori di V^n le cui coordinate, rispetto alla base $\mathbf{E}$, sono soluzioni di un opportuno sistema omogeneo in n incognite. Viceversa, ogni sottoinsieme di vettori di V^n costituito dai vettori le cui coordinate nella base $\mathbf{E}$ sono soluzioni di un fissato sistema omogeneo in n incognite, costituiscono un sottospazio vettoriale di V^n.

Facciamo due importanti esempi di sottospazi vettoriali, associati ad una trasformazione lineare $L : R^n \longrightarrow R^m$.

Data una trasformazione lineare $L : R^n \longrightarrow R^m$, il nucleo di L, $\ker(L)$, è un sottospazio di R^n mentre l'immagine, $Im(L)$, è un sottospazio di R^m.

La dimostrazione è la seguente.

Per il nucleo si ha che se $\mathbf{u}, \mathbf{v} \in \ker(L)$, allora $L(\mathbf{u}) = L(\mathbf{v}) = \mathbf{0}$. Se poi λ, μ sono scalari, risulta per la linearità di L

$$L(\lambda\mathbf{u}+\mu\mathbf{v}) = \lambda L(\mathbf{u})+\mu L(\mathbf{v}) = \mathbf{0}$$

dunque anche $\lambda\mathbf{u}+\mu\mathbf{v} \in \ker(L)$. Il nucleo è chiuso rispetto alle combinazioni lineari. Il nucleo è un sottospazio vettoriale di R^n.

Per l'immagine si ha che se $\mathbf{w}, \mathbf{z} \in Im(L)$, debbono esistere $\mathbf{u}, \mathbf{v} \in R^n$ tali che $L(\mathbf{u}) = \mathbf{w}$ e $L(\mathbf{v}) = \mathbf{z}$. Ma allora si avrà anche, sempre per la linearità di L, qualunque siano gli scalari λ, μ

$$\lambda\mathbf{w}+\mu\mathbf{z} = \lambda L(\mathbf{u}) + \mu L(\mathbf{v}) = L(\lambda\mathbf{u}+\mu\mathbf{v})$$

ossia anche ogni combinazione lineare dei due vettori $\mathbf{w}, \mathbf{z}$ presi in $Im(L)$, è immagine di un vettore, ovvero è ancora nell'immagine. L'immagine è chiusa rispetto alle combinazioni lineari, è un sottospazio vettoriale di R^m.

Una maniera generale per costruire sottospazi vettoriali consiste nel "generarli" per combinazioni lineari. Illustriamo tale metodo.

Si fissino a piacere r vettori di uno spazio vettoriale V: $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r$. Si consideri l'insieme dei vettori di V che sono combinazioni lineari di tali vettori ovvero l'insieme dei vettori del tipo

$$\lambda^1\mathbf{u}_1 + \lambda^1\mathbf{u}_2 + ...+\lambda^r\mathbf{u}_r$$

Indicheremo tale insieme con

$$\langle \mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r \rangle$$

Presi due vettori di tale insieme: $\lambda^1\mathbf{u}_1 + \lambda^2\mathbf{u}_2 + ...+\lambda^r\mathbf{u}_r$ e $\mu^1\mathbf{u}_1 + \mu^2\mathbf{u}_2 + ...+\mu^r\mathbf{u}_r$, si consideri una loro combinazione lineare:$\alpha(\lambda^1\mathbf{u}_1+\lambda^2\mathbf{u}_2+...+\lambda^r\mathbf{u}_r)+ \beta(\mu^1\mathbf{u}_1 + \mu^2\mathbf{u}_2 + ...+\mu^r\mathbf{u}_r)$. Si ha

$$\begin{aligned}&\alpha(\lambda^1\mathbf{u}_1 + \lambda^2\mathbf{u}_2 + ...+\lambda^r\mathbf{u}_r) + \beta(\mu^1\mathbf{u}_1 + \mu^2\mathbf{u}_2 + ...+\mu^r\mathbf{u}_r) = \\ &= (\alpha\lambda^1 + \beta\mu^1)\mathbf{u}_1 + (\alpha\lambda^2 + \beta\mu^2)\mathbf{u}_2 + ... + (\alpha\lambda^r + \beta\mu^r)\mathbf{u}_r\end{aligned}$$

Dunque

Una combinazione lineare di elementi di $\langle \mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r \rangle$ *è ancora un elemento di*

$\langle \mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r \rangle$. *Ossia* $\langle \mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r \rangle$ è *un sottospazio vettoriale di* V. *Tale sottospazio vettoriale viene chiamato sottospazio generato da* $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r$.

Si noti come $\langle \mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r \rangle$ è generato da $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r$ e pertanto $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r$ sono un sistema di generatori di $\langle \mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r \rangle$, ma in generale non sono una base di $\langle \mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r \rangle$. Chiaramente sono anche una base se e solo se sono indipendenti. Se non sono indipendenti una base di $\langle \mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r \rangle$ può ottenersi prendendo tra i vettori $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r$ dei vettori indipendenti in numero massimo, e la dimensione di $\langle \mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r \rangle$ sarà chiaramente il rango dell'insieme di vettori $(\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r)$:

$$\dim \langle \mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r \rangle = rg(\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r)$$

Vedremo in seguito altri importanti esempi di sottospazi lineari collegati agli operatori lineari. Si tratterà dei cosidetti autospazi di un operatore lineare. Per ora passiamo ad alcune importanti proprietà dei sottospazi.

Sia U un sottospazio vettoriale di uno spazio vettoriale V^n di dimensione finita n. p vettori di U sono linearmente dipendenti in U se e solo se tali sono in V^n. In particolare in U non possono esistere più vettori indipendenti di quanti non ne esistano in V^n, ovvero non si possono trovare in U più di n vettori indipendenti. Ne consegue che la dimensione di un sottospazio vettoriale di uno spazio vettoriale di dimensione fionita n, V^n, è sempre minore o uguale della dimensione di V^n. Di fatto la dimensione di un sottospazio vettoriale U di V^n è uguale ad n se e solo se $U = V^n$.

La proposizione enunciata è tutta ovvia e non ha bisogno di essere dimostrata. Salvo la parte finale:

La dimensione di U è uguale alla dimensione di V^n se e solo se $U = V^n$.

Naturalmente se U e V^n coincidono non possono che avere la stessa dimensione. Viceversa se le dimensioni di U e di V^n coincidono, allora in U

posso trovare una base costituita da n vettori indipendenti. Ma tali n vettori indipendenti in U sono anche n vettori indipendenti in V^n. Senonché sappiamo che le basi di V^n sono le $n-ple$ di vettori indipendenti e pertanto i vettori di una base di U saranno anche i vettori di una base di V^n. Questo comporta che facendo le combinazion lineari dei vettori di una base di U, che è anche una base di V^n, si ottengono tanto tutti i vettori di U quanto tutti i vettori di V^n e quindi V^n ed U coincidono.

Prendiamo un sottospazio vettoriale di V^n, diverso da V^n, quindi propriamente contenuto in V^n. Supponiamo di indicare con p la sua dimensione e con U^p il sottospazio vettoriale stesso. Oramai sappiamo che dovrà risultare $p < n$. In relazione a tale situazione si ha il seguente

Teorema di completamento di una base.

Comunque si fissi una base in U^p, $\mathbf{H} = (\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_p)$, si possono sempre trovare $n-p$ vettori in $V^n - U^p$, $(\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_{n-p})$ tali che

$\mathbf{E} = (\mathbf{h_1}, \mathbf{h_2}, ..., \mathbf{h_p}, \mathbf{g_1}, \mathbf{g_2}, ..., \mathbf{g_{n-p}})$ *sia una base di V^n.*

Se U^p è contenuto propriamente in V^n, vi saranno vettori di V^n che non si troveranno in U^p. Tali vettori non potranno ottenersi come combinazioni lineari dei vettori $(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_p)$ in quanto questi costituiscono una base di U^p e le loro combinazioni lineari danno solo vettori di U^p. Quindi ognuno dei vettori di $V^n - U^p$ è indipendente dai vettori $(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_p)$. Aggiungnendo uno qualunque dei vettori che si trovano in $V^n - U^p$ ai vettori $(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_p)$, diciamolo $\mathbf{g}_1$, otterremo $p + 1$ vettori indipendenti $(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_p, \mathbf{g}_1)$. Se $n = p + 1$ questi vettori costituiscono una base per V^n ed il teorema è dimostrato. Altrimenti, se $p + 1 < n$, i vettori $(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_p, \mathbf{g}_1)$, non possono costituire una base perché le basi di V^n sono sempre costituite da n vettori. Quindi, siccome i vettori $(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_p, \mathbf{g}_1)$ sono indipendenti, vuol dire che non sono un sistema di generatori. Ma questo è quanto dire che in V^n esistono vettori indipendenti dai vettori $(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_p, \mathbf{g}_1)$. Prendendo a piacere uno qualunque di tali vettori, diciamolo $\mathbf{g}_2$, ed aggiungendolo ai vettori $(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_p, \mathbf{g}_1)$, si ottengono i $p + 2$ vettori $(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_p, \mathbf{g}_1, \mathbf{g}_2)$. A questo punto si ripete il ragionamento fatto in precedenza, fino ad arrivare ad aggiungere $n - p$ vettori e si conclude con la tesi.

Vogliamo concludere il paragrafo con un argomento anche questo utile nel successivo capitolo sugli operatori lineari: *La somma diretta di sottospazi vettoriali.*

Nella precedente proposizione si è visto come, fissato un sottospazio vettoriale U^p di V^n con $p < n$, per ogni scelta di una base $\mathbf{H} = (\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_p)$ in U^p, si possano trovare $n-p$ vettori indipendenti di $V^n - U^p$, $(\mathbf{g_1}, \mathbf{g_2}, ..., \mathbf{g_{n-p}})$ tali che

$\mathbf{E} = (\mathbf{h_1}, \mathbf{h_2}, ..., \mathbf{h_p}, \mathbf{g_1}, \mathbf{g_2}, ..., \mathbf{g_{n-p}})$ costituisca una base di V^n. Se si considera il sottospazio vettoriale, generato dai vettori $(\mathbf{g_1}, \mathbf{g_2}, ..., \mathbf{g_{n-p}})$ ossia il sottospazio che indichiamo con $\langle \mathbf{g_1}, \mathbf{g_2}, ..., \mathbf{g_{n-p}} \rangle$, si ha intanto che i vettori $(\mathbf{g_1}, \mathbf{g_2}, ..., \mathbf{g_{n-p}})$ sono una base di $\langle \mathbf{g_1}, \mathbf{g_2}, ..., \mathbf{g_{n-p}} \rangle$ (sono indipendenti ed un sistema di generatori per definizione), inoltre potremo dire che una base di V^n, la base $\mathbf{E}$, è costituita dall'unione di una base $\mathbf{H}$ di U^p e di una base $\mathbf{G} = (\mathbf{g_1}, \mathbf{g_2}, ..., \mathbf{g_{n-p}})$ di $\langle \mathbf{g_1}, \mathbf{g_2}, ..., \mathbf{g_{n-p}} \rangle$. Tale situazione si esprime dicendo che

Il sottospazio $\langle \mathbf{g_1}, \mathbf{g_2}, ..., \mathbf{g_{n-p}} \rangle$ *è un addendo diretto del sottospazio* U^p *in* V^n*, ovvero anche che lo spazio* V^n *è somma diretta di* U^p *e di* $\langle \mathbf{g_1}, \mathbf{g_2}, ..., \mathbf{g_{n-p}} \rangle$.

Notiamo che, dato il sottospazio U^p, questo non individua quello che abbiamo chiamato il suo addendo diretto. Infatti l'addendo diretto dipende chiaramente dalla determinazione di $n-p$ vettori indipendenti in $V^n - U^p$. Quindi dato un sottospazio vettoriale U^p, questo in generale ammette più addendi diretti in V^n, ovvero uno spazio vettoriale V^n è somma diretta in più modi di due suoi sottospazi vetgtoriali. Sulla base della terminologia ora usata, possiamo dare le definizioni seguenti

Dato uno spazio vettoriale V^n*, si dice che* V^n *è somma diretta di due suoi sottospazi vettoriali* U^p *e* W^q*, ovvero anche che due suoi sottospazi* U^p *e* W^q *sono due addendi diretti in* V^n*, se si possono scegliere in* U^p *ed in* W^q *due basi, rispettivamente*

$\mathbf{H} = (\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_p)$ *e* $\mathbf{G} = (\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_q)$*, tali che l'unione delle due basi,* $(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_p, \mathbf{g}_2, ..., \mathbf{g}_q)$*, sia una base* $\mathbf{E}$ *di* V^n.

Una semplice dimostrazione, che omettiamo, mostra che

Se U^p *e* W^q *sono due addendi diretti di uno spazio vettoriale* V^n*, allora comunque si scelgano in* U^p *ed in* W^q *due basi, l'unione di tale basi è sempre una base di* V^n.

La definizione data viene generalizzata nella seguente

Dati r *sottospazi vettoriali di* V^n*,* $U_1, U_2, ..., U_r$ *si dice che questi sono addendi diretti in* V^n*, ovvero che* V^n *è somma diretta dei suoi sottospazi vettoriali* $U_1, U_2, ..., U_r$*, se comunque si fissino delle basi in* $U_1, U_2, ..., U_r$*, l'unione disgiunta (se due basi diverse hanno due elementi uguali nell'unione disgiunta l'elemento comune viene ripetuto due volte) di tali basi è una base di* V^n.

Se V^n *è somma diretta dei suoi sottospazi vettoriali* $U_1, U_2, ..., U_r$*, si pone*

$$V^n = U_1 \oplus U_2 \oplus ... \oplus U_r$$

Notiamo che se $V^n = U_1 \oplus U_2 \oplus ... \oplus U_r$, siccome una base di V^n si può ottenere facendo l'unione di basi di $U_1, U_2, ..., U_r$, il numero degli elementi che compongono una base di V^n, ovvero la dimensione di V^n, dovrà essere uguale alla somma del numero degli elementi che compongono le basi di $U_1, U_2, ..., U_r$, ossia dovrà essere uguale alla somma delle dimensioni degli addendi diretti della decomposizione di V^n in somma diretta:

$$n = \dim(V^n) = \dim(U_1) + \dim(U_2) + ... + \dim(U_r)$$

In base alle definizioni date risulta che

Se si fissa una decomposizione di V^n in somma diretta di suoi r sottospazi, $V^n = U_1 \oplus U_2 \oplus ... \oplus U_r$, allora comunque si scelga in V^n un vettore $\mathbf{v}$ *tale vettore può univocamente essere espresso, in un'unica maniera, come somma di r vettori,* $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r$, *rispettivamente contenuti in* $U_1, U_2, ..., U_r$

$$\mathbf{v} = \mathbf{u}_1 + \mathbf{u}_2 + ... + \mathbf{u}_r$$

Per semplificare la dimostrazione facciamola nel caso $r = 2$. Il caso generale è del tutto identico. Considerata in V^n una base, $\mathbf{E} = (\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_p, \mathbf{g}_2, ..., \mathbf{g}_q)$, fatta dall'unione di due basi una, $\mathbf{H} = (\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_p)$ in U_1, l'altra $\mathbf{G} = (\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_q)$ in U_2, ogni vettore $\mathbf{v}$, verrà ad essere espresso come combinazione lineare dei vettori della base $\mathbf{E}$

$$\mathbf{v} = \lambda^1\mathbf{h}_1 + \lambda^2\mathbf{h}_2 + ... + \lambda^p\mathbf{h}_p + \mu^1\mathbf{g}_1 + \mu^2\mathbf{g}_2 + ... + \mu^q\mathbf{g}_q$$

Ponendo $\mathbf{u}_1 = \lambda^1\mathbf{h}_1 + \lambda^2\mathbf{h}_2 + ... + \lambda^p\mathbf{h}_p$ e $\mathbf{u}_2 = \mu^1\mathbf{g}_1 + \mu^2\mathbf{g}_2 + ... + \mu^q\mathbf{g}_q$, intanto si ha che ogni vettore $\mathbf{v}$ è esprimibile come somma di un vettore $\mathbf{u}_1$ di U_1 e di un vettore $\mathbf{u}_2$ di U_2. Con qualche ragionamento in più, ragionamento che omettiamo, si vede che di fatto i due vettori $\mathbf{u}_1$ e $\mathbf{u}_2$ non dipendono, come potrebbe sembrare, dalla particolare scelta delle basi in U_1 e U_2.

Vediamo di crearci delle immagini geometriche delle nozioni appena illustrate. Consideriamo lo spazio vettoriale V^3. In V^3 consideriamo i vettori paralleli ad un fissato piano π, indichiamo tale insieme con $V^2(\pi)$, ed i vettori paralleli ad una data retta r. Indichamo quest'altro insieme con $V^1(r)$. Se π ed r sono paralleli, allora prendendo una base in $V^2(\pi)$ ed una base in $V^1(r)$, si otterranno tre vettori complanari, quindi dipendenti, e lo spazio vettoriale V^3 non sarà somma diretta dei due suoi sottospazi $V^2(\pi)$ e $V^1(r)$: mettendo assieme una base di uno ed una base dell'altro, non si ottiene mai una base di tutto V^3 in quanto si ottengono sempre vettori dipendenti. Supponiamo oea che la retta ed il piano non siano paralleli. In tal caso prendendo una base di $V^2(\pi)$, due vettori non paralleli di π ed una base di $V^1(r)$, un vettore non nullo di r, si ottengono tre vettori non complanari quindi formanti una base per V^3. Questo qualunque sia la scelta fatta delle basi in $V^2(\pi)$ ed in $V^1(r)$. Dunque, in base alla definizione,

V^3 è somma diretta di $V^2(\pi)$ e di $V^1(r)$, se π ed r non sono paralleli.

Questo risultato ci fa capire come l'addendo diretto di un sottospazio non sia individuato: fissato un piano, ed il sottospazio vettoriale dei vettori paralleli al piano, comunque si fissi una retta non parallela al piano, il corrispondente sottospazio vettoriale è sempre un addendo diretto del sottospazio vettoriale dei vettori del piano fissato.

Altra immagine geometrica di quanto asserito in precedenza è la seguente: dato un vettore dello spazio, questo può sempre decomporsi in maniera unica come somma di un vettore del piano π e di un vettore della retta r. Precisamente, dato $\mathbf{v}$ in V^3, rimangono individuati due vettori $\mathbf{v}_\pi$ e $\mathbf{v}_r$, il primo parallelo a π il secondo parallelo ad r, tali che

$$\mathbf{v} = \mathbf{v}_\pi + \mathbf{v}_r$$

I due vettori $\mathbf{v}_\pi$ e $\mathbf{v}_r$, vengono rispettivamente detti, la proiezione del vettore $\mathbf{v}$ sul piano π parallelamente alla retta r, e la proiezione del vettore $\mathbf{v}$ sulla retta r parallelamente al piano π. Nel caso particolare in cui π ed r, siano stati presi perpendicolari, si parla di proiezione ortogonale del vettore $\mathbf{v}$ sul piano, il vettore $\mathbf{v}_\pi$**,** o di proiezione ortogonale di $\mathbf{v}$ sulla retta, il vettore $\mathbf{v}_r$.

Notiamo come, sempre facendo riferimento allo spazio vettoriale V^3, vi sia la possibilità di fare un esempio di somma diretta con tre addendi. Si fissino tre rette, r, s, t, non parallele ad uno stesso piano, e si considerino i sottospazi costituiti dai vettori paralleli ad r. $V^1(r)$, dei vettori paralleli ad s, $V^1(s)$, dei vettori paralleli a t, $V^1(t)$. Prendendo in ognuno di tali sottospazi un vettore non nullo, cioè una base di tale sottospazio vettoriale, l'ipotesi che le tre rette non siano parallele ad uno stesso piano, assicura che i tre vettori non nulli presi parallelamente alle tre rette, non sono complanari, e vengono pertanto a costituire una base per i vettori dello spazio. Di nuovo ci troviamo di fronte al caso di tre sottospazi vettoriali per i quali avviene che l'unione delle loro basi dà una base dello spazio. Di nuovo abbiamo che lo spazio vettoriale V^3 è somma diretta dei sottospazi $V^1(r), V^1(s), V^1(t)$

$$V^3 = V^1(r) \oplus V^1(s) \oplus V^1(t)$$

In questo caso la decomposizione di un vettore in vettori paralleli ai sottospazi, dà luogo alla decomposizione di ogni vettore dello spazio nella somma di tre vettori: dato $\mathbf{v} \in \mathbf{V}^3$, rimangono individuati $\mathbf{v}_r \in V^1(r)$, $\mathbf{v}_s \in V^1(s)$, $\mathbf{v}_t \in V^1(t)$, tali che

$$\mathbf{v} = \mathbf{v}_r + \mathbf{v}_s + \mathbf{v}_t$$

I vettori $\mathbf{v}_r$**,** $\mathbf{v}_s$**,** $\mathbf{v}_t$, vengono rispettivamente detti la proiezione di $\mathbf{v}$ su $V^1(r)$, su $V^1(s)$, su $V^1(t)$.

In generale quando si fissa una decomposizione di uno spazio vettoriale V nella somma diretta dei suoi sottospazi, cioè si ha

$$V^n = U_1 \oplus U_2 \oplus ... \oplus U_r$$

per modo che per un qualunque vettore $\mathbf{v}$ rimangono individuati i vettori $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r$, tali che

$$\mathbf{v} = \mathbf{u}_1 + \mathbf{u}_2 + ... + \mathbf{u}_r$$

la corrispondenza che ad ogni vettore $\mathbf{v}$ associa il vettore $\mathbf{u}_h$, $h = 1, 2, ..., r$, viene detta proiezione del vettore $\mathbf{v}$ sull'$h-mo$ addendo diretto della decomposizione. Indicheremo tale proiezione con Π_h e porremo

$$\Pi_h(\mathbf{v}) = \mathbf{u}_h$$

Quello che ci preme far rilevare è la seguente proprietà della proiezione. Dati due vettori $\mathbf{v}, \mathbf{w}$, ed uno scalare λ, risulta

$$\begin{aligned} \Pi_h(\mathbf{v} + \mathbf{w}) &= \Pi_h(\mathbf{v}) + \Pi_h(\mathbf{w}) \\ \Pi_h(\lambda \mathbf{v}) &= \lambda \Pi_h(\mathbf{v}) \end{aligned}$$

Facciamo la dimostrazione nel caso $n = 2$. Al solito nel caso generale tutto è identico.

Poniamo

$$\begin{aligned} V^n &= U_1 \oplus U_2 \\ \mathbf{v} &= \mathbf{u}_1 + \mathbf{u}_2 \\ \mathbf{w} &= \mathbf{t}_1 + \mathbf{t}_2 \end{aligned}$$

Con tali posizioni risulterà

$$\begin{aligned} \Pi_1(\mathbf{v}) &= \mathbf{u}_1 \text{ e } \Pi_1(\mathbf{w}) = \mathbf{t}_1 \\ \Pi_2(\mathbf{v}) &= \mathbf{u}_2 \text{ e } \Pi_2(\mathbf{w}) = \mathbf{t}_2 \end{aligned}$$

Sommando a membro a membro le due uguaglianze troviamo

$$\mathbf{v} + \mathbf{w} = (\mathbf{u}_1 + \mathbf{t}_1) + (\mathbf{u}_2 + \mathbf{t}_2) = (\Pi_1(\mathbf{v}) + \Pi_1(\mathbf{w})) + (\Pi_2(\mathbf{v}) + \Pi_2(\mathbf{w}))$$

Siccome U_1 e U_2 sono sottospazi vettoriali e $\mathbf{u}_1, \mathbf{t}_1 \in U_1$, $\mathbf{u}_2, \mathbf{t}_2 \in U_2$ si ha che $\mathbf{u}_1 + \mathbf{t}_1 = \Pi_1(\mathbf{v}) + \Pi_1(\mathbf{w}) \in U_1$e che $\mathbf{u}_2 + \mathbf{t}_2 = \Pi_2(\mathbf{v}) + \Pi_2(\mathbf{w} \in U_2$. Dunque l'uguaglianza

$$\mathbf{v} + \mathbf{w} = (\Pi_1(\mathbf{v}) + \Pi_1(\mathbf{w})) + (\Pi_2(\mathbf{v}) + \Pi_2(\mathbf{w})))$$

esprime il vettore somma $\mathbf{v} + \mathbf{w}$ come somma di due vettori: $\Pi_1(\mathbf{v}) + \Pi_1(\mathbf{w})$ in U_1 e $\Pi_2(\mathbf{v}) + \Pi_2(\mathbf{w})$ in U_2. D'altronde sappiamo che tale decomposizione è unica, per cui si deve avere

$$\begin{aligned} \Pi_1(\mathbf{v} + \mathbf{w}) &= \Pi_1(\mathbf{v}) + \Pi_1(\mathbf{w}) \\ \Pi_2(\mathbf{v} + \mathbf{w}) &= \Pi_2(\mathbf{v}) + \Pi_2(\mathbf{w}) \end{aligned}$$

Analogamente si trovano le uguaglianze

$$\begin{aligned}\Pi_1(\lambda\mathbf{v}) &= \lambda\Pi_1(\mathbf{v})\\ \Pi_2(\lambda\mathbf{v}) &= \lambda\Pi_2(\mathbf{v})\end{aligned}$$

OPERATORI LINEARI

1. Generalità e prime definizioni..

In quanto è stato sin qui svolto si sono incontrati alcuni esempi di operatori lineari. Facciamone un breve elenco.

i) Le trasformazioni lineari: $L : R^n \longrightarrow R^m$

ii) La corrispondenza che, fissata una base $\mathbf{E}$, *di uno spazio vettoriale di dimensione finita* V^n, *associa ad un vettore le sue coordinate relativamente alla base fissata*: $\varphi_{\mathbf{E}} : V^n \longrightarrow R^n$.

iii) Le proiezioni di uno spazio vettoriale sui suoi addendi diretti: $\Pi_h : V^n = U_1 \oplus U_2 \oplus ... \oplus U_r \longrightarrow U_h$

A questi esempi di operatori lineari, possiamo i seguenti

iv) Le rotazioni del piano di un angolo θ *attorno ad un punto fissato,* O, *e secondo un verso di rotazione stabilito*: $R_\theta : V^2(\pi) \longrightarrow V^2(\pi)$

v) Le omotetie di ragione $\rho \in R$, $I_\rho : V \longrightarrow V$ *di un qualunque spazio vettriale* V *in sé: le corrispondenze che ad un vettore di* V *associano il vettore stesso moltiplicato per uno scalare fissato* ρ *detto ragione dell'omotetia. Ovvero* $I_\rho(\mathbf{v}) = \rho\mathbf{v}$.

Per le corrispondenze *i), ii), iii)*, è stato segnalato di volta in volta, che valgono le due seguenti proprietà

$$\begin{aligned}T(\mathbf{u}+\mathbf{v}) &= T(\mathbf{u}) + T(\mathbf{v})\\ T(\lambda\mathbf{u}) &= \lambda T(\mathbf{u})\end{aligned}$$

dove con T si è indicata una qualunque delle corrispondenze sopracitate, e con $\mathbf{u}, \mathbf{v}$, vettori appartenenti agli spazi vettoriali in cui tali corrispondenze sono definite. Per le omotetie la verifica delle suddette peoprietà è immediata, mentre la stessa verifica per le rotazioni del piano richiede qualche ragionamento anche se semplice.

Sulla base degli esempi portati diamo la definizione seguente di *Operatore Lineare*:

Una applicazione $T : V \longrightarrow W$ tra due spazi vettoriali V, W, è detta Operatore Lineare, se verifica le due proprietà seguenti

$$\begin{aligned} T(\mathbf{u}+\mathbf{v}) &= T(\mathbf{u})+T(\mathbf{v}) \\ T(\lambda\mathbf{u}) &= \lambda T(\mathbf{u}) \end{aligned}$$

per $\mathbf{u}, \mathbf{v} \in V$ e $\lambda \in R$.

E' subito visto che dalle due proprietà segue che se T è un operatore lineare allora soddisfa la

$$T(\lambda\mathbf{u}+\mu\mathbf{v}) = \lambda T(\mathbf{u}) + \mu T(\mathbf{v})$$

e più in generale la

$$T(\lambda^1\mathbf{u}_1+\lambda^2\mathbf{u}_2+...+\lambda^r\mathbf{u}_r) = \lambda^1 T(\mathbf{u}_1)+\lambda^2\mathbf{T}(\mathbf{u}_2)+...+\lambda^r T(\mathbf{u}_r)$$

Ma vale ovviamente anche il viceversa per cui potremo dire che

$T : V \longrightarrow W$ è un operatore lineare se e solo se verifica l'uguaglianza

$$T(\lambda^1\mathbf{u}_1+\lambda^2\mathbf{u}_2+...+\lambda^r\mathbf{u}_r) = \lambda^1 T(\mathbf{u}_1)+\lambda^2\mathbf{T}(\mathbf{u}_2)+...+\lambda^r T(\mathbf{u}_r)$$

Per un operatore lineare $T : V \longrightarrow W$ si definiscono i due sottospazi lineari

$$\ker(T) = \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}\}$$

e

$$Im(T) = \{\mathbf{w} \in W : \exists \mathbf{v} \in V : T(\mathbf{v}) = \mathbf{w}\}$$

Il primo è un sottospazio di V, il secondo un sottospazio di W. La dimostrazione che si tratti di sottospazi vettoriali è identica a quella fatta per dimostrare che il nucleo e l'immagine di una trasformazione lineare sono sottospazi vettoriali.

Da ora in poi considereremo solo operatori lineari definiti ed a valori in uno stesso spazio vettoriale V di dimensione finita $n : V = V^n$.

In tal caso tanto $\ker(T)$, quanto $Im(T)$ sono sottospazi vettoriali dello stesso spazio vettoriale V^n, ed entrambi hanno dimensione finita minore od uguale ad n.

Si chiama rango di un operatore lineare T, e lo si indica con $rg(T)$, la dimensione del sottospazio immagine dell'operatore:

$$rg(T) = \dim(Im(T))$$

Vedremo tra breve che anche il rango di un operatore lineare qualunque può sempre essere pensato come il rango di una matrice, così come avveniva per il rango di una trasformazione lineare.

Per quanto riguarda il nucleo di un operatore lineare, si ha la seguente importante proprietà:

Un operatore lineare $T : V^n \longrightarrow V^n$ *è biunivoco se e solo se* $\ker(T) = \{\mathbf{0}\}$

Dimostriamo tale proposizione.

Supponiamo $T : V^n \longrightarrow V^n$ biunivoco. In tal caso deve essere $\ker(T) = \{\mathbf{0}\}$ infatti se ciò non fosse esisterebbe $\mathbf{v} \neq \mathbf{0}$ e $\mathbf{v} \in \ker(T)$. Ma allora dovrebbe risultare $T(\mathbf{v}) = \mathbf{0}$ e $T(\mathbf{0}) = \mathbf{0}$. Ossia i due vettori $\mathbf{v}, \mathbf{0}$ pur essendo diversi, avrebbero la stessa immagine, contro l'ipotesi di biunivocità dell'operatore.

Viceversa, se $\ker(T) = \{\mathbf{0}\}$ e risultasse $T(\mathbf{u}) = T(\mathbf{v})$, ne seguirebbe $T(\mathbf{u}) - T(\mathbf{v}) = \mathbf{0}$ e per la linearità di T si avrebbe $T(\mathbf{u} - \mathbf{v}) = \mathbf{0}$. Quindi $\mathbf{u} - \mathbf{v} \in \ker(T)$. Ma $\ker(T) = \{\mathbf{0}\}$ dunque $\mathbf{u} - \mathbf{v} = \mathbf{0}$ ossia $\mathbf{u} = \mathbf{v}$. Donde intanto l'iniettività di T. Per dimostrare anche la surgettività di T utilizziamo la seguente proposizione che ha interesse di per sé

Se il suo nucleo di un operatore lineare è nullo, l'operatore trasforma vettori indipendenti in vettori indipendenti.

Diamo subito la dimostrazione di tale fatto per tornare poi alla dimostrazione lasciata in sospeso.

Sia $\ker(T) = \{\mathbf{0}\}$. Siano $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r$, vettori indipendenti. Consideriamo una combinazione lineare delle loro immagini: $\lambda^1 T(\mathbf{u}_1) + \lambda^2 T(\mathbf{u}_2) + ... + \lambda^r T(\mathbf{u}_r)$. Supponiamo che tale combinazione lineare sia nulla: $\lambda^1 T(\mathbf{u}_1) + \lambda^2 T(\mathbf{u}_2) + ... + \lambda^r T(\mathbf{u}_r) = \mathbf{0}$. Per la linearità di T avremo anche

$$T(\lambda^1 \mathbf{u}_1 + \lambda^2 \mathbf{u}_2 + ... + \lambda^r \mathbf{u}_r) = \mathbf{0}$$

quindi $\lambda^1 \mathbf{u}_1 + \lambda^2 \mathbf{u}_2 + ... + \lambda^r \mathbf{u}_r \in \ker(T)$. Ma $\ker(T) = \{\mathbf{0}\}$. Dunque deve essere

$$\lambda^1 \mathbf{u}_1 + \lambda^2 \mathbf{u}_2 + ... + \lambda^r \mathbf{u}_r = \mathbf{0}$$

Ma i vettori $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r$ sono stati presi indipendenti quindi $\lambda^1 = \lambda^2 = ... = \lambda^r = 0$. Ossia l'unica combinazione lineare dei vettori $T(\mathbf{u}_1), T(\mathbf{u}_2), ..., T(\mathbf{u}_r)$ uguale al vettore nullo è quella a coefficenti tutti nulli: i vettori sono indipendenti.

Ora che abbiamo dimostrato l'ultima proposizione possiamo terminare facilmente la dimostrazione di quella precedente. Sappiamo che $\ker(T) = \{\mathbf{0}\}$ e quindi oramai che T trasforma vettori indipendenti in vettori indipendenti. Prendiamo in V^n una base

$\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n)$. Le immagini dei vettori della base, $T(\mathbf{e}_1), T(\mathbf{e}_2), ..., T(\mathbf{e}_n)$ saranno indipendenti. Ma n vettori indipendenti di V^n sono a loro volta una base di V^n. Quindi $T(\mathbf{e}_1), T(\mathbf{e}_2), ..., T(\mathbf{e}_n)$ sono una base di V^n. Siccome $Im(T)$ è un sottospazio vettoriale, tutte le combinazioni lineari dei vettori $T(\mathbf{e}_1), T(\mathbf{e}_2), ..., T(\mathbf{e}_n)$, che sono tutti i vettori di V^n poiché $T(\mathbf{e}_1), T(\mathbf{e}_2), ..., T(\mathbf{e}_n)$

formano una base, saranno ancora nell'immagine. Pertanto $Im(T) = V^n$ ossia T è surgettivo.

E' il caso di osservare che, per la proposizione di prima, se $\ker(T) = \{\mathbf{0}\}$ allora T trasforma vettori indipendenti in vettori indipendenti, ma se $\ker(T) \neq \{\mathbf{0}\}$, questo non è più vero: vettori indipendenti possono avere per immagine vettori dipendenti. Anzi riandando a vedere la dimostrazione della prposizione in questione, ci si convince subito che

Se i vettori $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n$ *formano una base, i vettori* $T(\mathbf{e}_1)$,$T(\mathbf{e}_2)$, ...,$T(\mathbf{e}_n)$ *formeranno una base se e solo se* $\ker(T) = \{\mathbf{0}\}$, *ossia se e solo se* T *è biunivoco.*

2. Matrice di un Operatore Lineare.

Immaginiamo di fissare nello spazio vettoriale V^n una base $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n)$.

Dato un operatore lineare $T : V^n \longrightarrow V^n$ si possono considerare le seguenti uguaglianze

$$\begin{cases} T(\mathbf{e}_1) = a_1^1\mathbf{e}_1 + a_1^2\mathbf{e}_2 + ... + a_1^n\mathbf{e}_n \\ T(\mathbf{e}_2) = a_2^1\mathbf{e}_1 + a_2^2\mathbf{e}_2 + ... + a_2^n\mathbf{e}_n \\ . \\ . \\ T(\mathbf{e}_n) = a_n^1\mathbf{e}_1 + a_n^2\mathbf{e}_2 + ... + a_n^n\mathbf{e}_n \end{cases}$$

Evidentemente le colonne

$$\begin{matrix} a_1^1 \\ a_1^2 \\ . \\ . \\ a_1^n \end{matrix} , \begin{matrix} a_2^1 \\ a_2^2 \\ . \\ . \\ a_2^n \end{matrix} , ..., \begin{matrix} a_n^1 \\ a_n^2 \\ . \\ . \\ a_n^n \end{matrix}$$

danno le colonne delle coordinate dei vettori $T(\mathbf{e}_1)$,$T(\mathbf{e}_2)$, ...,$T(\mathbf{e}_n)$, relativamente alla base $\mathbf{E}$. La matrice che ha per colonne le coordinate dei vettori $T(\mathbf{e}_1)$,$T(\mathbf{e}_2)$, ...,$T(\mathbf{e}_n)$ nella base $\mathbf{E}$, ovvero la matrice

$$A_{(n)} = \begin{pmatrix} a_1^1 & a_2^1 & . & . & a_n^1 \\ a_1^2 & a_2^2 & . & . & a_n^2 \\ . & . & . & . & . \\ . & . & . & . & . \\ a_1^n & a_2^n & . & . & a_n^n \end{pmatrix}$$

è detta *matrice dell'operatore* T *nella base* $\mathbf{E}$. Tale matrice svolge un ruolo fondamentale in tutta la teoria degli operatori lineari. Per tale motivo prima di andare avanti con ulteriori considerazioni sulla matrice di un operatore, vediamo di considerarne alcuni esempi.

Sia $L : R^2 \longrightarrow R^2$ un operatore lineare di R^2 in sé:

$$\begin{cases} y^1 = a_1^1 x^1 + a_2^1 x^2 \\ y^2 = a_1^2 x^1 + a_2^2 x^2 \end{cases}$$

Sappiamo che la matrice

$$\begin{pmatrix} a_1^1 & a_2^1 \\ a_1^2 & a_2^2 \end{pmatrix}$$

è la matrice di L come trasformazione lineare. Domandiamoci chi è la matrice di L come operatore lineare. La prima cosa da osservare è che per parlare di matrice di un operatore lineare bisogna preventivamente fissare una base. Fissiamo la base canonica $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2) = ((1,0),(0,1))$ di R^2 e calcoliamo, secondo la definizione, la matrice di L pensato come operatore lineare relativamente alla base $\mathbf{E}$. Dobbiamo calcolare i valori di L sui due vettori della base scelta, $\mathbf{e}_1, \mathbf{e}_2$, ed esprimere tali valori come combinazione lineare sempre dei vettori della base $\mathbf{E}$. I valori di L su $\mathbf{e}_1, \mathbf{e}_2$ sono le due coppie che si ottengono calcolando la trasformazione lineare prima sulla coppia $(1,0)$, poi sulla coppia $(0,1)$ e sono

$$\begin{aligned} L(\mathbf{e}_1) &= (a_1^1, a_1^2) \\ L(\mathbf{e}_2) &= (a_2^1, a_2^2) \end{aligned}$$

L'espressione della prima coppia, della coppia (a_1^1, a_1^2), come combinaione lineare delle due coppie della base $\mathbf{E}$, è data da

$$(a_1^1, a_1^2) = a_1^1 \mathbf{e}_1 + a_1^2 \mathbf{e}_2$$

quella della seconda coppia da

$$(a_2^1, a_2^2) = a_2^1 \mathbf{e}_1 + a_2^2 \mathbf{e}_2$$

Dunque risulta

$$\begin{aligned} L(\mathbf{e}_1) &= a_1^1 \mathbf{e}_1 + a_1^2 \mathbf{e}_2 \\ L(\mathbf{e}_2) &= a_2^1 \mathbf{e}_1 + a_2^2 \mathbf{e}_2 \end{aligned}$$

Pertanto la matrice di L come operatore lineare nella base canonica di R^2, è

$$\begin{pmatrix} a_1^1 & a_2^1 \\ a_1^2 & a_2^2 \end{pmatrix}$$

e quindi coincide con la matrice di L come trasformazione lineare.

Importante è osservare che la coincidenza tra la matrice di L come trasformazione lineare e e la matrice di L come operatore lineare si ha se la base in cui si calcola la matrice di L come operatore è la base canonica di R^2 altrimenti ciò non è detto ed in generale non sarà vero.

Il risultato esposto si estende tale e quale al caso di una arbitraria trasformazione lineare L di R^n in sé;

La matrice di una trasformazione lineare $L : R^n \longrightarrow R^n$, pensata come operatore lineare, relativamente alla base canonica di R^n, coincide con quella che in precedenza era stata definita come matrice della trasformazione lineare.

Vediamo di calcolarci la matrice di una proiezione.

Facciamo il caso di una proiezione di uno spazio vettoriale decomposto nella somma diretta di due suoi sottopazi vettoriali, chiaramente ognuno unidimensionale (si ricordi come la dimensione di uno spazio decomposto in somma diretta deve uguagliare la somma delle dimensioni dei suoi addendi diretti)

$$V^2 = U^1 \oplus W^1$$

Prendiamo la proiezione sul primo addendo $\Pi_1 : V^2 \longrightarrow U^1$. Notiamo come, così concepita, la proiezione viene ad essere un operatore lineare tra due spazi vettoriali diversi: V^2 ed il suo addendo diretto U^1. Per applicare le definizioni che noi abbiamo dato, bisogna considerare operatori lineari di uno spazio vettoriale in sé. Dobbiamo pertanto interpretare l'operatore di proiezione, come un operatore di V^2 in V^2. Si può fare questo pensando ai vettori di U^1 come a particolari vettori di V^2 cosa che di fatto sono. Interpreteremo da ora la proiezione come un operatore di V^2 in sé. Come base rispetto alla quale calcolare la matrice della proiezione prendiamo una qualunque coppia di vettori non nulli di U^1, e di W^1: $\mathbf{u} \in U^1, \mathbf{w} \in W^1$. Un generico vettore $\mathbf{v}$ di V^2 si potrà decomporre in maniera unica come somma di un vettore di U^1 e di uno di W^1. Siccome i vettori di U^1 sono tutti proporzionali a $\mathbf{u}$ (U^1 ha dimensione 1) e quelli di W^1 sono tutti proporzionali a $\mathbf{w}$, ogni vettore $\mathbf{v}$ di V^2 si potrà scrivere

$$\mathbf{v} = \lambda\mathbf{u} + \mu\mathbf{w}$$

Calcoliamo la proiezione sulla base fissata $(\mathbf{u}, \mathbf{w})$ di V^2 tenendo conto del fatto che, per definizione, la proiezione Π_1 associa ad ogni vettore la sua componente sul primo addendo

$$\begin{aligned} \Pi_1(\mathbf{u}) &= \mathbf{u} \\ \Pi_1(\mathbf{w}) &= \mathbf{0} \end{aligned}$$

ma

$$\begin{aligned}\mathbf{u} &= 1\mathbf{u}+0\mathbf{w}\\ \mathbf{0} &= 0\mathbf{u}+0\mathbf{w}\end{aligned}$$

e dunque i valori della proiezione calcolata sui vettori di una base in funzione dei vettori della stessa base, sono

$$\begin{aligned}\Pi_1(\mathbf{u}) &= 1\mathbf{u}+0\mathbf{w}\\ \Pi_1(\mathbf{w}) &= 0\mathbf{u}+0\mathbf{w}\end{aligned}$$

Dal che si deduce che la matrice della proiezione sul primo addendo in una base fatta dall'unione di due basi prese nei singoli addendi, è

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

Il ragionamento fatto per calcolare la matrice di una proiezione su di un addendo diretto di una decomposizione di uno spazio vettoriale in somma diretta è del tutto generale. Con identici ragionamenti, complicati solo da fatti formali dovuti alle dimensini dei sottospazi della decomposizione in somma diretta, si arriva al seguente enunciato

Una matrice di una proiezione su di un addendo diretto di una decomposizione di uno spazio vettoriale in somma diretta, è sempre una matrice diagonale con sulla diagonale principale tutti zero fatta eccezzione dei posti della diagonale principale, corrispondenti ai vettori di una base dell'addendo sul quale si proietta, in cui la matrice ha degli uno.

Calcoliamoci la matrice di una rotazione del piano di un angolo θ attorno ad un punto del piano, O, in un fissato verso di rotazione. Indichiamo tale operatore con $R_\theta : \pi \longrightarrow \pi$.

Come base nel piano prendiamo una coppia ordinata di vettori entrambi di lunghezza uno e tra loro perpendicolari: quella che si chiama una base ortonormale del piano. Indichiamo detta base con $(\mathbf{i}, \mathbf{j})$. Facendo ruotare il versore (vuol dire vettore di lunghezza pari a uno) $\mathbf{i}$ di un angolo θ, otterremo un altro vettore di lunghezza unitaria (un altro versore), $R_\theta(\mathbf{i})$, che formerà con il versore $\mathbf{i}$, non ruotato, un angolo θ. Pertanto la componente di $R_\theta(\mathbf{i})$ rispetto ad $\mathbf{i}$, sarà $\cos\theta$, mentre la componente rispetto ad $\mathbf{j}$ sarà $\sin\theta$, per cui avremo

$$R_\theta(\mathbf{i}) = \cos\theta\mathbf{i} + \sin\theta\mathbf{j}$$

In maniera analoga si trova che $R_\theta(\mathbf{j})$, ruotato dello stesso angolo θ di $\mathbf{j}$, è dato da

$$R_\theta(\mathbf{j}) = -\sin\theta\mathbf{i} + \cos\theta\mathbf{j}$$

per cui i valori dell'operatore di rotazione sui vettori della base ortonormale $(\mathbf{i}, \mathbf{j})$, espressi come combinazione lineare dei vettori della base ortonormale $(\mathbf{i}, \mathbf{j})$, sono dati da

$$\begin{array}{l} R_\theta(\mathbf{i}) = \cos\theta\mathbf{i} + \sin\theta\mathbf{j} \\ R_\theta(\mathbf{j}) = -\sin\theta\mathbf{i} + \cos\theta \end{array}$$

e la matrice dell'operatore di rotazione relativamente alla base $(\mathbf{i}, \mathbf{j})$, è

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Richiamiamo l'attenzione sul fatto che le coordinate dei vettori immagine dei vettori della base sono disposte sulle colonne della matrice dell'operatore e non sulle righe come si sarebbe spinti a fare.

Calcoliamo la matrice di una omotetia di ragione ρ di uno spazio vettoriale V^n. $I_\rho : V^n \longrightarrow V^n$. Si fissi una base $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n)$ in V^n. Per la definizione stessa di omotetia si avrà $I_\rho(\mathbf{e}_i) = \rho\mathbf{e}_i$, $i = 1, ..., n$. Pertanto la matrice dell'omotetia sarà

$$A = \begin{pmatrix} \rho & 0 & . & . & 0 \\ 0 & \rho & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & \rho \end{pmatrix} = \rho\Delta_n$$

Si noti come nel caso delle omotetie, la matrice sia sempre la stessa qualunque sia la base. La matrice di un'omotetia non dipende dalla base. Si tratta di un caso unico. Precedentemente si era visto che la matrice di una rotazione era sempre la stessa prendendo una base ortonormale del piano. Ma se la base del piano non fosse stata ortonormale, la matrice della rotazione non sarebbe stata quella trovata.

Torniamo alle considerazioni generali sulla matrice dell'operatore.

Innanzitutto si deve avere ben chiaro che la matrice di un operatore dipende dalla base che si sceglie. Cambiando base la matrice cambierà. Almeno in generale. Tra poco vedremo come cambia la matrice di un operatore al cambiare della base. Una seconda osservazione è che, come visto in precedenza, i vettori $T(\mathbf{e}_1)$,$T(\mathbf{e}_2)$, ...,$T(\mathbf{e}_n)$, immagini secondo T dei vettori di una base, non sono in generale, indipendenti. Sono indipendenti, quindi sono a loro volta una base, se e solo se l'operatore è biunivoco. Ma l'indipendenza dei vettori $T(\mathbf{e}_1)$,$T(\mathbf{e}_2)$, ...,$T(\mathbf{e}_n)$ equivale evidentemente all'annullarsi o meno del determinante della matrice delle loro coordinate nella base $\mathbf{E}$, matrice che è la matrice dell'operatore. Quindi

Un operatore lineare è biunivoco se e solo se il determinante della sua matrice è diverso da zero

Si noti che, mentre la matrice di un operatore cambia al cambiare della base, la condizione esposta vale qualunque sia la base in cui la matrice dell'operatore viene calcolata: è indipendente dalla base.

Si è appena osservato che, la matrice di un operatore dipende dalla base in cui viene calcolata. Vediamo come tale matrice cambi al cambiare della base.

Fissiamo due basi: $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n)$ ed $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_n)$. Siano

$$\mathbf{F} = \mathbf{E}C$$

le formule del cambiamento di base, quindi C sia la matrice del cambiamento di base dalla base $\mathbf{E}$ alla base $\mathbf{F}$. Poniamo

$$T(\mathbf{E}) = \mathbf{E}A$$

in altre parole indichiamo con A la matrice dell'operatore T relativamente alla base $\mathbf{E}$, e poniamo

$$T(\mathbf{F}) = \mathbf{F}B$$

ossia indichiamo con B la matrice di T nella base $\mathbf{F}$. Vogliamo stabilire come si possa ottenere B da A. Abbiamo le uguaglianze

$$T(\mathbf{F}) = T(\mathbf{E}C) = T(\mathbf{E})C = \mathbf{E}AC$$

la prima uguaglianza e l'ultima essendo ottenute dalle precedenti uguaglianze, mentre la seconda uguaglianza essendo conseguenza dalla linearità dell'operatore. Inoltre risulta

$$T(\mathbf{F}) = \mathbf{F}B = \mathbf{E}CB$$

anche queste due uguaglianze derivando dalle uguaglianze di prima. Mettendo assieme i due risultati troviamo

$$\mathbf{E}AC = \mathbf{E}CB$$

Dunque

$$AC = CB$$

Moltiplicando ambo i membri a sinistra per l'inversa di C, inversa che sicuramente esiste poiché C è una matrice di un cambiamento di base ed in particolare è invertibile (si ricordi le colonne di C sono le coordinate dei vettori della nuova base rispetto alla vecchia base e quindi sono colonne indipendenti), otteniamo la relazione cercata tra A e B:

$$B = C^{-1}AC$$

A parole il risultato ottenuto si esprime dicendo che

La matrice, B, di un operatore lineare T in una base $\mathbf{F}=\mathbf{E}C$, si ottiene dalla matrice A, dello stesso operatore, nella base $\mathbf{E}$ moltiplicando la matrice A a sinistra per l'inversa, C^{-1}, della matrice, C, del cambiamento di base dalla base $\mathbf{E}$ alla base $\mathbf{F}$, ed a destra per la matrice, C, di detto cambiamento di base.

Segnaliamo subito che è una relazione di importanza fondamentale. Matrici legate da una relazione quale quella di sopra possono sempre essere interpretate come matrici di uno stesso operatore in basi diverse, ottenibili una dall'altra con matrice di cambiamento di base data dalla matrice C.

Passiamo ad un altro importante argomento: quello delle *equazioni di un operatore lineare.*

Anche questo argomento non può essere trattato senza l'assegnazione preventiva di una base. Fissiamo una base $\mathbf{E}=(\mathbf{e}_1,\mathbf{e}_2,...,\mathbf{e}_n)$ e continuiamo ad indicare con A la matrice dell'operatore T nella base $\mathbf{E}$, ovvero continuiamo a porre $T(\mathbf{E})=\mathbf{E}A$. Ogni vettore, $\mathbf{v}$ di V^n, ammetterà delle coordinate relativamente alla base $\mathbf{E}$. Poniamo

$$\mathbf{v}=v^1\mathbf{e}_1+v^2\mathbf{e}_2+...+v^n\mathbf{e}_n=\mathbf{E}V$$

dove ora con

$$V=V_{(n,1)}=\begin{pmatrix} v^1 \\ v^2 \\ . \\ . \\ v^n \end{pmatrix}$$

si è indicata la matrice colonna delle coordinate di $\mathbf{v}$ nella base $\mathbf{E}$. Poniamo poi $\mathbf{w}=T(\mathbf{v})$ e

$$\mathbf{w}=w^1\mathbf{e}_1+w^2\mathbf{e}_2+...+w^n\mathbf{e}_n=\mathbf{E}W$$

dove

$$W=W_{(n,1)}=\begin{pmatrix} w^1 \\ w^2 \\ . \\ . \\ w^n \end{pmatrix}$$

è la matrice colonna delle coordinate di $\mathbf{w}=T(\mathbf{v})$ sempre nella base $\mathbf{E}$. Avremo, in base alle posizioni fatte ed alla linearità di T

$$\mathbf{w}=T(\mathbf{v})=T(\mathbf{E}V)=T(\mathbf{E})V=\mathbf{E}AV$$

ed anche

$$\mathbf{w}=\mathbf{E}W$$

da cui, mettendo assieme le due uguaglianze appena trovate

$$\mathbf{E}W = \mathbf{E}AV$$

il che comporta

$$W = AV$$

L'uguaglianza matriciale scritta, esplicitata, porta alla scrittura della seguente trasformazione lineare

$$\begin{cases} w^1 = a^1_1 v^1 + a^1_2 v^2 + ... + a^1_n v^n \\ w^2 = a^2_1 v^1 + a^2_2 v^2 + ... + a^2_n v^n \\ . \\ . \\ w^n = a^n_1 v^1 + a^n_2 v^2 + ... + a^n_n v^n \end{cases}$$

Tale trasformazione lineare porta il nome di *equazioni dell'operatore lineare T relativamente alla base* $\mathbf{E}$.

Si noti come le equazioni di un operatore dipendano dalla base in cui vengono calcolate. Esse fanno passare dalle coordinate di un vettore a quelle della sua immagine, secondo l'operatore T, con la trasformazione lineare che ha per matrice la matrice dell'operatore relativamente alla base in cui si calcolano le coordinate dei vettori. Dunque:

Le coordinate, W, del vettore $\mathbf{w}$ $=T(\mathbf{v})$, *immagine di un vettore* $\mathbf{v}$ *secondo l'operatore T, si ottengono dalle coordinate, V, del vettore* $\mathbf{v}$, *con la trasformazione lineare che ha per matrice la matrice A dell'operatore T relativamente alla base* $\mathbf{E}$ *nella quale si calcolano tanto le coordinate V di* $\mathbf{v}$ *quanto le coordinate W di* $\mathbf{w}$.

Si osservi come, fissata una base, uno spazio vettoriale qualunque, V^n, si rappresenti con lo spazio vettoriale numerico R^n, di dimensione pari alla dimensione dello spazio vettoriale V^n, associando ad ogni vettore $\mathbf{v}$ l'$n-pla$ delle sue coordinate nella base $\mathbf{E}$. Inoltre un qualunque operatore lineare di V^n, si rappresenta con una trasformazione lineare di R^n in sé. Lo studio degli spazi vettoriali e quello degli operatori lineari di tali spazi vettoriali in sé, è in tal modo ricondotto allo studio di R^n e delle trasformazioni lineari di R^n in sé.

3. Autovalori ed Autovettori.

Non ci è qui data la possibilità di accennare all'importanza che, in numerosi problemi collegati agli operatori lineari, rivestono le nozioni di autovalori ed autovettori. Ci limiteremo pertanto ad esporre per sommi capi tali nozioni.

Se per un vettore non nullo, $\mathbf{v}$*, esiste uno scalare* λ*(eventualmente anche nullo) tale che*

$$T(\mathbf{v}) = \lambda\mathbf{v}$$

allora si dice che $\mathbf{v}$ *è un autovettore e che* λ *è un autovalore. Inoltre* $\mathbf{v}$ *e* λ *si dicono associati.*

Poniamoci il seguente problema: dato uno scalare λ_0 quando è che avviene che questo sia un autovalore di un operatore T assegnato? ovvero quando esiste un vettore $\mathbf{v}$ non nullo in corrispondenza al quale risulti $T(\mathbf{v}) = \lambda_0\mathbf{v}$?

Prima di rispondere pienamente al problema segnaliamo che se λ_0 è un autovalore, ossia supposto che esista un vettore non nullo $\mathbf{v}$ per il quale si abbia $T(\mathbf{v}) = \lambda_0\mathbf{v}$, un tale vettore non è mai unico anzi risulta

Se λ_0 *è un autovalore allora l'insieme degli autovettori che sono associati a* λ_0*, con l'aggiunta del vettore nullo (che per definizione non è mai un autovettore) costituiscono un sottospazio vettoriale. Lo spazio vettoriale degli autovettori associati ad un dato autovalore* λ_0*, al quale insieme si aggiunga il vettore nullo, viene detto autospazio associato all'autovalore* λ_0*, ed indicato con* V_{λ_0}*.*

L'insieme di tutti gli autovalori di un fissato operatore T *(insieme che può essere anche vuoto), viene detto spettro dell'operatore* T *ed indicato con* Λ_T*.*

Dimostriamo la proposizione contenuta in quello che precede.

Supponiamo che tanto $\mathbf{u}$ quanto $\mathbf{v}$ siano o autovettori di uno stesso autovalore λ_0 o il vettore nullo. In ogni caso si avrà: $T(\mathbf{u}) = \lambda_0\mathbf{u}$, $T(\mathbf{v}) = \lambda_0\mathbf{v}$. Siano α, β due scalari. Per la linearità di T avremo le uguaglianze che seguono

$$T(\alpha\mathbf{u}+\beta\mathbf{v}) = \alpha T(\mathbf{u})+\beta T(\mathbf{v}) = \alpha\lambda_0\mathbf{u}+\beta\lambda_0\mathbf{v} = \lambda_0(\alpha\mathbf{u}+\beta\mathbf{v})$$

che mostra come, se due vettori o sono autovettori associati ad un dato autovalore, o sono il vettore nullo, anche una qualunque loro combinazione lineare, o è un autovettore associato allo stesso autovalore o è il vettore nullo. In altre parole l'insieme degli autovettori di un autovalore, con l'aggiunta del vettore nullo, è chiuso rispetto alle combinazioni lineari e costituisce pertanto un sottospazio vettoriale.

Si badi al fatto che mentre gli autovettori associati ad un dato autovalore sono infiniti,

Gli autovalori, λ*, associati ad un dato autovettore, diciamolo* $\mathbf{v}_0$*, cioè gli scalari* λ *per i quali avviene che* $T(\mathbf{v}_0) = \lambda\mathbf{v}_0$*, sono uno solo.*

Infatti se si avesse tanto $T(\mathbf{v}_0) = \lambda\mathbf{v}_0$, quanto $T(\mathbf{v}_0) = \mu\mathbf{v}_0$, il che implicherebbe

$$\lambda\mathbf{v}_0 = \mu\mathbf{v}_0$$

ovvero

$$(\lambda - \mu)\mathbf{v}_0 = \mathbf{0}$$

e, siccome stiamo supponendo $\mathbf{v}_0$ non nullo in quanto autovettore, deve per forza essere $\lambda = \mu$.

Torniamo al problema dal quale eravamo partiti: quando un dato scalare λ_0 è un autovalore per un operatore T assegnato?

Se ciò è vero deve esistere un vettore non nullo $\mathbf{v}$ in corrispondenza al quale risulta $T(\mathbf{v}) = \lambda_0\mathbf{v}$. Fissando una base, $\mathbf{E}$, in V^n, indicando con A la matrice di T nella base $\mathbf{E}$, avremo che le equazioni di T nella base $\mathbf{E}$, saranno

$$\begin{cases} w^1 = a_1^1 v^1 + a_2^1 v^2 + ... + a_n^1 v^n \\ w^2 = a_1^2 v^1 + a_2^2 v^2 + ... + a_n^2 v^n \\ . \\ . \\ w^n = a_1^n v^1 + a_2^n v^2 + ... + a_n^n v^n \end{cases}$$

Come sappiamo queste equazioni fanno passare dalle coordinate di un vettore alle coordinate della sua immagine. Dunque lo scalare λ_0 è un autovalore se e solo se esiste un vettore $\mathbf{v}$ non nullo le cui coordinate sono trasformate nelle sue stesse coordinate, moltiplicate per λ_0, ossia le cui coordinate soddisfino le

$$\begin{cases} \lambda_0 v^1 = a_1^1 v^1 + a_2^1 v^2 + ... + a_n^1 v^n \\ \lambda_0 v^2 = a_1^2 v^1 + a_2^2 v^2 + ... + a_n^2 v^n \\ . \\ . \\ \lambda_0 v^n = a_1^n v^1 + a_2^n v^2 + ... + a_n^n v^n \end{cases}$$

Portando tutto ad un membro, si ottiene

$$\begin{cases} 0 = (a_1^1 - \lambda_0) v^1 + a_2^1 v^2 + ... + a_n^1 v^n \\ 0 = a_1^2 v^1 + (a_2^2 - \lambda_0) v^2 + ... + a_n^2 v^n \\ . \\ . \\ 0 = a_1^n v^1 + a_2^n v^2 + ... + (a_n^n - \lambda_0) v^n \end{cases}$$

Possiamo allora dire che un certo scalare λ_0 è un autovalore per l'operatore T se e solo se, il sistema scritto, pensato nelle incognite $(v^1, v^2, ..., v^n)$, ammette una soluzione non nulla ovvero ammette autosoluzioni. Ma quello scritto è un sistema omogeneo di n equazioni in n incognite con matrice dei coefficienti la matrice $A - \lambda_0\Delta$, e tale sistema ammetterà autosoluzioni se e solo se

$$\det(A - \lambda_0\Delta) = 0$$

Riassumiamo quanto visto nell'enunciato che segue:

Uno scalare λ_0 *è un autovalore per un dato operatore* T, *se e solo se* λ_0 *è radice dell'equazione*

$$\det(A - \lambda_0 \Delta) = 0$$

dove A *è la matrice di* T *in una fissata base. L'equazione nell'incognita* λ : $\det(A - \lambda\Delta) = 0$ *viene detta equazione caratteristica dell'operatore.*

Soffermiamoci sull'equazione caratteristica di T.

Innanzitutto la condizione $\det(A - \lambda\Delta) = 0$ è espressa in funzione della matrice dell'operatore in una base assegnata ma, come è stato mostrato, non dipende da quale base si consideri. Senonché vale qualcosa di più:

L'equazione caratteristica di T *è sempre la stessa, qualunque sia la base a partire dalla quale viene calcolata.*

Dimostriamo la proposizione. Sappiamo che se dalla base $\mathbf{E}$ si passa alla base $\mathbf{F} = \mathbf{E}C$, la matrice dell'operatore passa dalla matrice A alla matrice $A^{-1}CA$. Calcolando l'equazione caratteristica di T relativamente alla base $\mathbf{F}$, avremo che questa sarà data dall'equazione

$$\det(A^{-1}CA - \lambda\Delta) = 0$$

Ma risulta

$$\begin{aligned} \det(A^{-1}CA - \lambda\Delta) &= \det(A^{-1}CA - \lambda A^{-1}A) = \det A^{-1}(CA - \lambda\Delta)A = \\ &= \det A^{-1} \det(CA - \lambda\Delta) \det A = \det(CA - \lambda\Delta) \end{aligned}$$

ossia la tesi: le due equazioni caratteristiche, quella scritta a partire dalla matrice dell'operatore nella base $\mathbf{E}$, ossia $\det(A - \lambda\Delta) = 0$ e quella scritta a partire dalla matrice dell'operatore nella base $\mathbf{F} = \mathbf{E}C$, ossia $\det(A^{-1}CA - \lambda\Delta) = 0$, coincidono.

Vediamo di che tipo è l'equazione caratteristica di un operatore. Cominciamo a vederlo nel caso di un operatore di uno spazio vettoriale di dimensione due $T : V^2 \longrightarrow V^2$. In tal caso una base sarà costituita da due vettori e la matrice di T in una data base sarà una matrice quadrata di ordine due. Poniamo che la matrice di T sia data dalla matrice

$$A = \begin{pmatrix} a_1^1 & a_2^1 \\ a_1^2 & a_2^2 \end{pmatrix}$$

La matrice $A - \lambda\Delta$ sarà allora la matrice

$$A - \lambda\Delta = \begin{pmatrix} a_1^1 - \lambda & a_2^1 \\ a_1^2 & a_2^2 - \lambda \end{pmatrix}$$

e conseguentemente l'equazione caratteristica sarà

$$\det A - \lambda\Delta = \det \begin{pmatrix} a_1^1 - \lambda & a_2^1 \\ a_1^2 & a_2^2 - \lambda \end{pmatrix} = \lambda^2 - (a_1^1 + a_2^2)\lambda + \det A = 0$$

ossia l'equazione caratteristica di T è data da un polinomio di secondo grado in λ uguagliato a zero.

Tale risultato si estende facilmente al caso di un operatore lineare di uno spazio vettoriale di dimensione n. Precisamente

L'equazione caratteristica di un operatore lineare $T : V^n \longrightarrow V^n$, è data da un polinomio di grado n, pari alla dimensione dello spazio sul quale l'operatore opera, uguagliato a zero. Il polinomio che uguagliato a zero dà l'equazione caratteristica dell'operatore T, si chiama polinomio caratteristico e lo si indica con $p_T(\lambda)$:

$$p_T(\lambda) = \det A - \lambda\Delta$$

Come osservato prima, tale polinomio non dipende dalla base che si utilizza per calcolarlo.

Vediamo alcune importanti conseguenze del risultato ottenuto.

Il polinomio caratteristico $p_T(\lambda)$ è un polinomio a coefficienti reali. Come sappiamo non sempre un polinomio a coefficenti reali ammette radici (nel campo reale). Pertanto non sempre un operatore ammetterà autovalori, dovendo questi, per quanto apena visto, essere radici reali del polinomio caratteristico dell'operatore. Veramente si potrebbe supporre che il polinomio caratteristico non sia arbitrario ma sempre di un tipo che ammette autovalori. Purtroppo così non è come si constata andando a considerare l'operatore di rotazione del piano che abbiamo indicato con R_θ. Abbiamo visto che la matrice di tale operatore rispetto ad una (qualunque) base ortonormale è data dalla matrice

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

L'equazione caratteristica dell'operatore R_θ, sarà

$$\det \begin{pmatrix} \cos\theta - \lambda & \sin\theta \\ -\sin\theta & \cos\theta - \lambda \end{pmatrix} = 0$$

dunque sarà data da

$$\lambda^2 - 2\cos\theta\lambda + 1 = 0$$

Il discriminante di tale equazione è

$$\cos^2\theta - 1$$

da cui si vede che tale discriminante, salvo il caso di $\theta = k\pi$, sarà sempre negativo e quindi l'equazione caratteristica del corrispondente operatore di rotazione, non ammetterà radici reali. Ma se $\theta = k\pi$ la rotazione corrispondente o è nulla, $\theta = 2h\pi$, oppure coincide con la simmetria dei vettori del piano rispetto all'origine $\theta = 2(h+1)\pi$. A parte questi due casi, ogni altra rotazione è del tutto priva di autovalori e di autovettori, cosa peraltro geometricamente più che evidente. Infatti se la rotazione ammettesse autovalori dovrebbe ammettere autovettori, ovvero vettori mutati dalla rotazione in vettori proporzionali. Ma se la rotazione non è né l'identità né la simmetria rispetto all'origine, non può evidentemente mutare un vettore in un altro proporzionale giacché quello proporzionale dovrebbe essere parallelo al vettore di partenza.

Nonostante quanto osservato, molti tra gli operatori che intervengono nei problemi che si presentano, ammettono autovalori ed autovettori.

Prima di chiudere il paragrafo enunciamo un importante teorema del quale non daremo la dimostrazione:

Se λ_1, λ_2, *sono due autovalori distinti, e se con* $V_{\lambda_1}, V_{\lambda_2}$, *si indicano i rispettivi autospazi, e con* n_1, n_2 *le rispettive dimensioni, presa una base in* V_{λ_1}: $\mathbf{E}_1 = (\mathbf{e}_{11}, \mathbf{e}_{12}, ..., \mathbf{e}_{1n_1})$ *ed una base in* V_{λ_2}: $\mathbf{E}_2 = (\mathbf{e}_{21}, \mathbf{e}_{22}, ..., \mathbf{e}_{2n_2})$ *, l'unione delle due basi, ovvero i vettori* $(\mathbf{e}_{11}, \mathbf{e}_{12}, ..., \mathbf{e}_{1n_1}, \mathbf{e}_{21}, \mathbf{e}_{22}, ..., \mathbf{e}_{2n_2})$, *sono vettori indipendenti.*

Il teorema si estende al caso di un numero qualunque di autovalori e si enuncia dicendo che

Mettendo assieme vettori costituenti basi di autospazi diversi, si ottengono sempre vettori indipendenti.

Dato un autovalore la dimensione dell'autospazio associato, è detta molteplicità geometrica dell'autovalore.

Nel paragrafo successivo studieremo delle conseguenze che si presentano allorché si considerano operatori con autovalori e quindi autovettori.

4. Operatori diagonalizzabili. Ricerca degli autovalori e degli autovettori.

Sia $T : R^n \longrightarrow R^n$ un operatore lineare. Si supponga che T ammetta un autovalore λ_1. Sia $\mathbf{u}_1$ un autovettore associato a λ_1. . Si consideri una base $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n)$ in cui il primo vettore coincida con l'autovettore $\mathbf{u}_1 : \mathbf{e}_1 = \mathbf{u}_1$. Questo è sempre possibile in base al teorema del completamento di una base. Nella bse in questione risulterà $T(\mathbf{u}_1) = \lambda_1 \mathbf{u}_1$. La matrice di T in tale base sarà pertanto del tipo

$$A = \begin{pmatrix} \lambda_1 & a_2^1 & . & . & a_n^1 \\ 0 & a_2^2 & . & . & a_n^2 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & a_2^n & . & . & a_n^n \end{pmatrix}$$

Supponiamo che T ammetta due autovalori distinti:λ_1, λ_2. Siano $\mathbf{u}_1$ ed $\mathbf{u}_2$ due autovettori associati rispettivamente al primo ed al secondo autovalore. Sappiamo che gli autovettori $\mathbf{u}_1$ e $\mathbf{u}_2$ sono indipendenti. Si consideri una base $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n)$ in cui i primi due vettori coincidano rispettivamente con $\mathbf{u}_1$ e $\mathbf{u}_2$: $\mathbf{e}_1 = \mathbf{u_1}, \mathbf{e}_2 = \mathbf{u}_2$. Al solito questo è possibile per il teorema di completamento della base. Risulterà $T(\mathbf{u}_1) = \lambda_1 \mathbf{u}_1$ e $T(\mathbf{u}_2) = \lambda_2 \mathbf{u}_2$ per cui in tale basela matrice di T sarà del tipo

$$A = \begin{pmatrix} \lambda_1 & 0 & . & . & a_n^1 \\ 0 & \lambda_2 & . & . & a_n^2 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & a_n^n \end{pmatrix}$$

Procedendo in tal modo si vede che se T ammette n autovalori distinti, $\lambda_1, \lambda_2, ..., \lambda_n$ con $\lambda_i \neq \lambda_j$ per $i \neq j$, si può prendere una base fatta tutta da autovettori, uno per ogni autovalore ed in tale base la matrice di T risulterà del tipoccia

$$A = \begin{pmatrix} \lambda_1 & 0 & . & . & 0 \\ 0 & \lambda_2 & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & \lambda_n \end{pmatrix}$$

ossia risulterà essere una matrice diagonale.

Un operatore che ammette una base in cui la sua matrice è una matrice diagonale, è detto operatore diagonalizzabile.

Si faccia attenzione al fatto che, come è stato appena visto

Se un operatore ammette n autovalori distinti, esiste una base in cui la matrice dell'operatore è una matrice dell'operatore è diagonale e pertanto un operatore con n autovalori distinti è diagonalizzabile,

Ma non è detto che per essere diagonalizzabile un operatore, per ammettere una base in cui la sua matrice sia una matrice diagonale, necessariamente debba ammettere n autovalori distinti. Riandiamo all'ultimo teorema enunciato (ma non dimostrato) del paragrafo precedente. Supponiamo di avere un cero numero, p, di autovalori dell'operatore T: $\lambda_1, \lambda_2, ..., \lambda_p$ e supponiamo inoltre che le dimensioni degli autospazi $V_{\lambda_1}, V_{\lambda_2}, ..., V_{\lambda_p}$, chiamiamole $n_1, n_2, ..., n_p$, sommate, siano uguali ad n, dimensione di V^n: $n = n_1 + n_2 + ... + n_p$. Per l'ultimo teorema del paragrafo precedente mettendo assieme delle basi dei vari autospazi si ottengono n vettori indipendenti di V^n e quindi una base di V^n. Evidentemente la matrice di T in una base siffatta sarà

$$A = \begin{pmatrix} \lambda_1 \Delta_{n_1} & 0 & . & . & 0 \\ 0 & \lambda_2 \Delta_{n_2} & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & \lambda_p \Delta_{n_p} \end{pmatrix}$$

dove con Δ_{n_i} si è indicata la matrice unità di ordine $n_i = \dim V_{\lambda_i}$. Dunque in una base siffatta l'operatore ammette una matrice diagonale e concludiamo, ricordando che la dimensione di V_{λ_i} è stata chiamata molteplicità geometrica dell'autovalore λ_i (vedi ultima definizione alla fine del paragrafo precedente),

Se la somma delle molteplicità geometriche di un operatore è uguale alla dimensione di V^n, allora l'operatore è diagonalizzabile.

Ma non sarebbe complicato mostrare che la proposizione si inverte e si conclude con

Un operatore è diagonalizzabile se e solo se la somma delle molteplicità geometriche dei suoi autovalori è uguale ad n.

Facendo uso della nozione di somma diretta, la situazione esposta può esprimersi dicendo che

Un operatore è diagonalizzabile se e solo se lo spazio vettoriale in cui è definito è somma diretta dei suoi autospazi.

Si noti come, sulla base di quanto appena esposto,

Se un operatore è diagonalizzabile, l'operatore viene ad avere come matrice associata una matrice diagonale se e solo se la base in cui la matrice è calcolata è tutta costituita da autovettori dell'operatore.

Inoltre

Se un operatore in una certa base ha come matrice associata una matrice diagonale (quindi la base è tutta fatta da autovettori dell'operatore), sulla diagonale principale appaiono gli autovalori dell'operatore ognuno ripetuto un numero di volte pari al numero degli autovettori che costituiscono una base per l'autospazio associato all'autovalore in considerazione, quindi un numero di volte pari alla dimensione del corrispondente autospazio, quindi ancora, un numero di volte pari alla molteplicità geometrica dell'autovalore considerato.

Vediamo come si calcolano gli autovalori e gli autovettori di un operatore.

Cosa si debba fare per calcolare gli autovalori è evidente: gli autovalori sono le radici reali del polinomio caratteristico dell'operatore. Quindi

I) Si fissa una base **E** in cui sia possibile calcolare la matrice dell'operatore.

II) Si calcola la matrice A dell'operatore relativamente alla base fissata.

III) Si calcola il polinomio caratteristico $p_T(\lambda) = \det(A - \lambda\Delta)$ di T.

IV) Si cercano le radici reali dell'equazione caratteristica $p_T(\lambda) = 0$.

Tutti i punti, eccetto l'ultimo, non presentano in generale difficoltà Per l'ultimo punto la cosa è diversa. Si è visto che il polinomio caratteristico di un operatore, è di grado pari alla dimensione dello spazio vettoriale in cui l'operatore opera. E' ben noto che esistono formule per il calcolo delle radici di un polinomio fino al quarto grado. Oltre non sempre è possibile calcolare le radici di un polinomio. Naturalmente a meno che non ci si trovi di fronte a polinomi particolari. Ma in generale la ricerca delle radici di un polinomio di grado maggiore di quattro non è possibile. Pertanto ogni qualvolta ci si accinge al calcolo degli autovalori di un operatore di uno spazio vettoriale di dimensione maggiore di quattro può succedere di trovarsi di fronte ad una difficoltà insormontabile.

Ma supponiamo di essere riusciti a calcolare almeno un autovalore per un dato operatore. Indichiamo con λ_0 un tale autovalore e proponiamoci di risolvere il secondo problema: quello del calcolo degli autovettori associati all'autovalore λ_0. Gli autovettori, $\mathbf{u}$, associati all'autovalore λ_0, sono caratterizzati dal soddisfare l'uguaglianza $T(\mathbf{u}) = \lambda_0\mathbf{u}$ e dall'essere non nulli. Imponendo alle coordinate di $\lambda_0\mathbf{u}$ di uguagliare le coordinate di $T(\mathbf{u})$, coordinate che si calcolano con le equazioni dell'operatore, si trova il sistema

$$\begin{cases} \lambda_0 v^1 = a_1^1 v^1 + a_2^1 v^2 + ... + a_n^1 v^n \\ \lambda_0 v^2 = a_1^2 v^1 + a_2^2 v^2 + ... + a_n^2 v^n \\ . \\ . \\ \lambda_0 v^n = a_1^n v^1 + a_2^n v^2 + ... + a_n^n v^n \end{cases}$$

che, riscritto portando tutto ad un membro diventa

$$\begin{cases} (a_1^1 - \lambda_0) v^1 + a_2^1 v^2 + ... + a_n^1 v^n = 0 \\ a_1^2 v^1 + (a_2^2 - \lambda_0) v^2 + ... + a_n^2 v^n = 0 \\ . \\ . \\ a_1^n v^1 + a_2^n v^2 + ... + (a_n^n - \lambda_0) v^n = 0 \end{cases}$$

ed appare come un sistema di n equazioni nelle n incognite date dalle coordinate di vettore. La matrice dei coefficienti di tale sistema è la matrice $A - \lambda\Delta$ il cui determinante è nullo, in quanto λ_0 è stato scelto proprio come radice dell'equazione $\det(A - \lambda\Delta) = 0$. Pertanto il sistema ammette autosoluzioni. Ogni autosoluzione del sistema fornisce le coordinate, nella base **E**, in cui si è calcolata la matrice A dell'operatore, di un autovettore associato all'autovalore λ_0. In realtà il sistema scritto dà le equazioni dell'autospazio V_{λ_0} associato a λ_0. Quindi, in particolare, calcolando un sistema fondamentale di autosluzioni del sistema, si ottengono le coordinate dei vettori di una base di V_{λ_0}. Tale considerazione torna particolarmente utile quando, avendo un operatore diagonalizzabile ci si propone il problema della ricerca di una base di autovettori per l'operatore, ovvero la ricerca di una base in cui la matrice dell'operatore è una matrice diagonale. Infatti, supponendo questa volta non di conoscere uno solo degli autovalori di T, ma di conoscerli tutti, per ognuno degli autovalori si determina una base per il corrispondente autospazio, così come appena illustrato. Dopodiché, mettendo assieme tutte le basi così trovate, si ottiene (siamo nell'ipotesi che l'operatore sia diagoalizzabile) una base per lo spazio vettoriale, tutta fatta da autovettori (anche se chiaramente non associati allo stesso autovalore). In tale base la matrice dell'operatore sarà una matrice diagonale con sulla diagonale principale gli autovalori dell'operatore stesso, ognuno ripetuto un numero di volte pari alla dimensione del corrispondente autospazio, ovvero un numero di volte pari alla molteplicità geometrica dell'autovalore.

5. Matrici Diagonalizzabili.

Concludiamo l'argomento degli operatori lineari con una applicazione dei risultati esposti. Il problema è il seguente

Data una matrice quadrata A di ordine n quando è possibile trovare una matrice quadrata C dello stesso ordine, ed a determinante diverso da zero, in corrispondenza alla quale la matrice $C^{-1}AC$ sia una matrice diagonale?

Diamo le definizioni seguenti

Se per una matrice A si può trovare una matrice invertibile C (a determinante non nullo) tale che la matrice $C^{-1}AC$ sia diagonale allora la matrice A è detta diagonalizzabile e una matrice C per la quale $C^{-1}AC$ è diagonale, si dice che è una matrice che diagonalizza la matrice A.

Anticipiamo alcune risposte che saranno illustrate subito.

Non tutte le matrici sono diagonalizzabili. Di fatto vedremo tra breve (ma sarà un risultato pressocché evidente)

Una matrice è diagonalizzabile se e solo se è una matrice associata ad un operatore diagonalizzabile.

Se A è una matrice diagonalizzabile, quindi se esiste una matrice C invertibile per la quale la matrice $C^{-1}AC$ sia diagonale, la matrice C non è mai unica. Di fatto vedremo tra breve che

Se A è diagonalizzabile, quindi è associata ad un operatore diagonalizzabile T, le matrici C che diagonalizzano A, sono le matrici le cui colonne sono le coordinate, relative alla base in cui A è associata a T, dei vettori di una base fatta da autovettori per T.

La chiave delle dimostrazione e delle spiegazioni che seguiranno, consiste nel fatto che, data una matrice A_n quadrata di ordine n, rimane individuata la trasformazione lineare, L, di R^n in sé, che ha la matrice A_n come matrice associata, ovvero la trasformazione lineare

$$Y_{(n,1)} = A_n X_{(n,1)}$$

Sappiamo bene che tale trasformazione lineare, L, è un operatore lineare. Se la trasformazione lineare L, come operatore lineare è diagonalizzabile, allora esiste una base (che poi sarà una base di autovettori per L) nella quale la matrice di L sarà una matrice diagonale con sulla diagonale principale gli autovalori di L. Diciamo D_n tale matrice e poniamo

$$D_n = \begin{pmatrix} \lambda_1 & 0 & . & . & 0 \\ 0 & \lambda_2 & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & . & . & . & \lambda_n \end{pmatrix}$$

Una base di R^n sono n $n-ple$ indipendenti. La base, diciamola $\mathbf{F}$, fatta da autovettori di L, sarà costituita da n $n-ple$ e la matrice, chiamiamola C_n, che ha le $n-ple$ per colonne, è anche la matrice le cui colonne sono le coordinate delle $n-ple$ nella base canonica, giacché gli elementi di un' $n-pla$ sono le coordinate dell' $n-pla$ nella base canonica. Ma allora la matrice C_n è la matrice del cambiamento di base dalla base canonica $\mathbf{E}$ alla base $\mathbf{F}$, ovvero

$$\mathbf{F} = \mathbf{E}C_n$$

Ricordando come sono collegate le matrici di uno stesso operatore in due diverse basi, potremo dire che, essendo, per definizione, la matrice A_n la matrice di L nella base canonica$\mathbf{E}$ di R^n, e D_n, la matrice dello stesso operatore, la trasformazione lineare L, nella base $\mathbf{F} = \mathbf{E}C_n$, le due matrici saranno collegate dall'uguaglianza

$$D_n = C_n^{-1} A_n C_n$$

Dunque, per la stessa definizione, la matrice A_n è diagonalizzabile e una matrice che la diagonalizza è proprio una matrice le cui colonne sono le coordinate dei vettori di una base di autovettori di L, calcolate nella base nella quale A_n è la matrice di L. Siccome di basi di autovettori per un operatore diagonalizzabile,ve ne sono infinite, si capisce come di matrici che diagonalizzano una matrice supposta diagonalizzabile, ve ne siano infinite.

Tutto il ragionamento fatto si inverte senza difficoltà. Prima di tutto si noti che l'ipotesi che la matrice C_n sia invertibile, assicura che $\mathbf{F} = \mathbf{E}C_n$ è una base. Ora se A_n è una matrice diagonalizzabile, allora deve esistere una matrice C_n invertibile, per la quale risulta

$$D_n = C_n^{-1} A_n C_n$$

con D_n diagonale. La trasformazione lineare L, di R^n in sé, che ha come matrice nella base $\mathbf{E}$ la matrice A_n, avrà come matrice nella base $\mathbf{F} = \mathbf{E}C_n$ come matrice la matrice

$$D_n = C_n^{-1} A_n C_n$$

ossia una matrice diagonale e, per definizione stessa, viene ad essere un operatore diagonalizzabile. Inoltre i vettori della base $\mathbf{F} = \mathbf{E}C_n$ sono tutti autovettori di L e la matrice C_n viene ad essere la matrice delle coordinate di una base di autovettori per l'operatore diagonalizzabile L.

Il problema di vedere se e quando una matrice è diagonalizzabile, consiste evidentemente nel problema di trovare gli autovalori e gli autovettori di un operatore che abbia per matrice la matrice assegnata. Andando a vedere il metodo illustrato per la ricerca degli autovalori e degli autovettori di un operatore, si vede che tale metodo fa uso solo della matrice, A, dell'operatore. Infatti, per prima cosa si calcola il polinomio caratteristico dell'operatore, $\det(A-\lambda\Delta)$, che dipende solo da A. Dopodiché si calcolano le radici dell'equazione caratteristica $\det(A-\lambda\Delta) = 0$, anche queste dipendenti solo da A, ed infine per ogni autovalore, λ_0, si vanno a cercare i corrispondenti autovettori cercando le autosoluzioni di un sistema omogeneo la cui matrice dei coefficienti è determinata sempre solo da A e dall'autovalore λ_0. Quindi volendo sapere se una matrice A è diagonalizzabile tutti i conti possono essere fatti a partire solo dalla matrice A. L'unico passaggio tecnicamente complicato, e che può fermare tutti i calcoli, è il calcolo delle radici del polinomio caratteristico. Per il resto ci si trova sempre di fronte a conti, eventualmente lunghi, ma sempre possibili. Senonché vi sono delle matrici per le quali il calcolo degli autovalori è più semplice e delle matrici per le quali, anche se il calcolo esplicito degli autovalori non è sempre possibile, si è sempre sicuri che si tratta di matrici diagonalizzabili.

Un esempio del primo caso è quello delle matrici triangolari. Un conto immediato mostra che se una matrice è una matrice triangolare, i suoi autovalori sono tutti e soli gli elementi della sua diagonale principale. Quindi in tal caso gli autovalori si calcolano subito. Per vedere poi se la matrice in questione è o meno diagonalizzbile bisognerà andare a vedere, con il metodo illustrato, se la somma delle molteplicità geometriche dei diversi autovalori è o meno uguale all'ordine della matrice.

Un esempio del secondo caso è quello delle matrici simmetriche: si dimostra, noi non lo dimostreremo, che ogni matrice simmetrica è diagonalizzabile.

Mettiamo in guardia su di un errore che si può fare in relazione a quanto appena detto. Si è osservato che, se si ha una matrice triangolare, gli autovalori della matrice sono gli elementi che si trovano sulla sua diagonale principale. Ricordando le prime cose fatte nel presente corso, si ricorderà come si sia usato un metodo semplice per triangolarizzare una matrice, ovvero per passare da una matrice arbitraria ad una matrice triangolare. Si usava un metodo che era una estensione del metodo di Gauss per triangolarizzare un sistema. Si potrebbe allora pensare che, per calcolare gli autovalori di una arbitraria matrice, si possa prima passare dalla matrice assegnata ad una matrice triangolare, con il metodo delle eliminazioni successive di Gauss, e poi calcolare gli autovalori della matrice triangolare ottenuta. Questo procedimento però non va. Quando si passa, con il metodo di Gauss, da una matrice ad una triangolare, si passa da una matrice ad un'altra con lo stesso determinante, ma non con gli stessi autovalori: gli autovalori di una matrice triangolare ottenuta da una matrice data con il metodo di Gauss, sono diversi dagli autovalori della matrice data, pur avendo le due matrici lo stesso determinante.