Algebra lineare e geometria*

Questi appunti sono stati presi durante le lezioni dell'insegnamento "Algebra lineare e geometria" tenuto dal prof. Lella durante l'A. A. 2021/22. Non sono stati revisionati da alcun docente (potrebbero contenere errori, in forma e in sostanza, di qualsiasi tipo) e non sono in alcun modo sostitutivi della frequentazione delle lezioni. L'ordine seguito nella presentazione degli argomenti è quello del libro *Algebra lineare e geometria* di E. Schlesinger.

*mariachiara.menicucci@mail.polimi.it per segnalare errori, richiedere il codice LaTeX ecc.

Indice

1 Algebra delle matrici

Si chiama $\mathbf{K}$ un generico campo ($\mathbb{Q}$, $\mathbb{R}$ o $\mathbb{C}$).

Def Una matrice di tipo (m, n) su $\mathbf{K}$ è una tabella di elementi di $\mathbf{K}$ disposti su m righe e n colonne.

$M_{\mathbf{K}}(m, n)$ indica l'insieme delle matrici su $\mathbf{K}$ di tipo (m, n); M sta per matrice. Poiché $\mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$, $M_{\mathbb{Q}}(m, n) \subseteq M_{\mathbb{R}}(m, n) \subseteq M_{\mathbb{C}}(m, n)$.

In generale, data la matrice $A \in M_{\mathbf{K}}(m, n)$, il generico elemento di A si indica con il nome della matrice in minuscolo e un pedice il cui primo numero è l'indice di riga e il cui secondo numero è l'indica di colonna: a_{ij} è l'elemento generico di A e si trova nella riga i e nella colonna j.

$$A = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{m1} & a_{m2} & ... & a_{mn} \end{bmatrix} = [a_{ij}]$$

dove $[a_{ij}]$ indica la matrice degli elementi a_{ij}.

Gli elementi a_{1j} sono tutti gli elementi della prima riga, gli elementi a_{i2} tutti gli elementi della seconda colonna ecc.

Ci sono alcuni casi particolari di m e n:

- $m = n = 1$: la matrice $A = [a_{11}]$ contiene un solo elemento. $M_{\mathbf{K}}(1, 1) = \mathbf{K}$.
- $m = n > 1$: la matrice A è quadrata e $M_{\mathbf{K}}(n, n)$ è l'insieme delle matrici quadrate di ordine n.
- $m = 1$: $M_{\mathbf{K}}(1, n)$ è l'insieme dei vettori riga con n componenti, e $A = [a_{11}|a_{12}|...|a_{1n}]$. $M_{\mathbf{K}}(1, n) = \mathbf{K}^n$ è l'insieme delle n-uple di elementi di $\mathbf{K}$.
- $n = 1$: $M_{\mathbf{K}}(m, 1)$ è l'insieme dei vettori colonna con n componenti, e $A = \begin{bmatrix} a_{11} \\ a_{21} \\ ... \\ a_{m1} \end{bmatrix}$. $M_{\mathbf{K}}(m, 1) = \mathbf{K}^m$ è l'insieme delle m-uple di elementi di $\mathbf{K}$.

1.1 Operazioni tra matrici

Uguaglianza tra matrici Date $A \in M_{\mathbf{K}}(m, n)$, $B \in M_{\mathbf{K}}(p, q)$, se $m \neq p$ o $n \neq q$, A e B sono diverse. Se $m = n$ e $p = q$, cioè $A, B \in M_{\mathbf{K}}(m, n)$, $A = B \Longleftrightarrow a_{ij} = b_{ij}$ $\forall$ $i = 1, ..., m$ e $\forall$ $j = 1, ..., n$.

1.1.1 Somma tra matrici

L'operazione $+ : M_{\mathbf{K}}(m,n) \times M_{\mathbf{K}}(m,n) \to M_{\mathbf{K}}(m,n)$ è tale che, date $A = [a_{ij}]$ e $B = [b_{ij}]$, $A + B = [a_{ij} + b_{ij}]$. Dunque $(A, B) \mapsto (A + B)$.

1. $A = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 1 \\ 3 & 4 \end{bmatrix}$, $A + B = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$.

Proprietà

- Associativa: $\forall\ A, B, C \in M_{\mathbf{K}}(m,n)$, $A + (B + C) = (A + B) + C$.

- Commutativa: $\forall\ A, B \in M_{\mathbf{K}}(m,n)$, $A + B = B + A$.

- Esistenza dell'elemento neutro: $\forall\ A \in M_{\mathbf{K}}(m,n)\ \exists\ 0_M \in M_{\mathbf{K}}(m,n)$: vale $A + 0_M = A$. $0_M \in M_{\mathbf{K}}(m,n)$ è la matrice nulla $0_M = \begin{bmatrix} 0 & 0 & ... & 0 \\ 0 & 0 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 0 \end{bmatrix}$.

- Esistenza dell'elemento opposto: $\forall\ A \in M_{\mathbf{K}}(m,n)\ \exists\ B \in M_{\mathbf{K}}(m,n)$: vale $A + B = 0_M$. Data $A = [a_{ij}]$, $B = [-a_{ij}]$.

$(M_{\mathbf{K}}(m,n), +)$ è un gruppo abeliano.

1.1.2 Prodotto scalare per matrice

L'operazione $\cdot : \mathbf{K} \times M_{\mathbf{K}}(m,n) \to M_{\mathbf{K}}(m,n)$ è tale che, date $A = [a_{ij}]$ e $t \in \mathbf{K}$, $tA = [ta_{ij}]$. Dunque $(t, A) \mapsto tA$.

1. $A = \begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix}$, $t = 2$, $tA = \begin{bmatrix} 2 & -2 \\ 4 & 6 \end{bmatrix}$.

Proprietà

- Distributiva a sinistra: $\forall\ t, s \in \mathbf{K}$, $\forall\ A \in M_{\mathbf{K}}(m,n)$ $(t + s) A = tA + sA$.

  Al lato sinistro il $+$ indica una somma in $\mathbf{K}$, al lato destro una somma in $M_{\mathbf{K}}(m,n) \times M_{\mathbf{K}}(m,n)$.

- Distributiva a destra: $\forall\ t \in \mathbf{K}$, $\forall\ A, B \in M_{\mathbf{K}}(m,n)$ $t(A + B) = tA + tB$.

  Al lato destro e sinistro il $+$ indica una somma in $M_{\mathbf{K}}(m,n) \times M_{\mathbf{K}}(m,n)$.

- Associativa: $\forall\ t, s \in \mathbf{K}$, $\forall\ A \in M_{\mathbf{K}}(m,n)$ $(ts)\,A = t\,(sA)$.

Al lato sinistro c'è un prodotto in $\mathbf{K}$, al lato destro un prodotto in $\mathbf{K} \times M_{\mathbf{K}}(m,n)$.

- Esistenza dell'elemento neutro: $\forall\ A \in M_{\mathbf{K}}(m,n)\ \exists\ 1 \in \mathbf{K}$: vale $1_{\mathbf{K}}A = A$.

Queste operazioni e proprietà si possono usare per scrivere più sinteticamente un sistema lineare.

1. $\begin{cases} x - 3y = 2 \\ 4x + 5y = 1 \end{cases} \iff \begin{bmatrix} x - 3y \\ 4x + 5y \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \iff \begin{bmatrix} x \\ 4x \end{bmatrix} + \begin{bmatrix} -3y \\ 5y \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \iff x \begin{bmatrix} 1 \\ 4 \end{bmatrix} + y \begin{bmatrix} -3 \\ 5 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

1.1.3 Prodotto matriciale righe per colonne

L'operazione $\cdot\ : M_{\mathbf{K}}(m,n) \times M_{\mathbf{K}}(n,p) \to M_{\mathbf{K}}(m,p)$ assegna a $A \in M_{\mathbf{K}}(m,n)$ e $B \in M_{\mathbf{K}}(n,p)$ una matrice $C = AB \in M_{\mathbf{K}}(m,p)$, tale che $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$, cioè il generico elemento c_{ij} è uguale al prodotto scalare tra il vettore riga i di A e il vettore colonna j di B.

1. $A = [a_{11}|a_{12}...|a_{1n}]$ è $1 \times n$, $B = \begin{bmatrix} b_{11} \\ b_{21} \\ ... \\ b_{n1} \end{bmatrix}$ è $n \times 1$, $C = AB \in M_{\mathbf{K}}(1,1)$: $c_{11} = \sum_{k=1}^{n} a_{1k}b_{k1} = a_{11}b_{11} + a_{12}b_{21} + ... + a_{1n}b_{n1}$.

2. $A = [a_{11}a_{12}...a_{1n}]$ è $1 \times n$, $B = \begin{bmatrix} b_{11} & ... & b_{1p} \\ ... & ... & ... \\ b_{n1} & ... & b_{np} \end{bmatrix}$ è $n \times p$, $C = AB \in M_{\mathbf{K}}(1,p)$: $c_{1p} = \sum_{k=1}^{n} a_{1k}b_{kp} = a_{11}b_{1p} + a_{12}b_{2p} + ... + a_{1n}b_{np}$.

3. $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \in M_{\mathbf{K}}(2,2)$, $B = \begin{bmatrix} 0 & 1 & -1 \\ 0 & -1 & 2 \end{bmatrix} \in M_{\mathbf{K}}(2,3)$, $AB = \begin{bmatrix} 0 & -1 & 3 \\ 0 & -1 & 5 \end{bmatrix} \in M_{\mathbf{K}}(2,3)$.

Proprietà Si suppone che le matrici coinvolte abbiano dimensioni conformabili. Il prodotto matriciale ha le seguenti proprietà:

1. Proprietà distributiva a destra: $\forall\ A \in M_{\mathbf{K}}(m,n)$, $\forall\ B, C \in M_{\mathbf{K}}(n,p)$, $A\,(B + C) = AB + AC$.

   A sinistra il prodotto è in $M_{\mathbf{K}}(n,p)$, a destra in $M_{\mathbf{K}}(m,p)$.

2. Proprietà distributiva a sinistra: $\forall\ A, B \in M_{\mathbf{K}}(m,n)$, $\forall\ C \in M_{\mathbf{K}}(n,p)$, $(A+B)\,C = AC + BC$.

   A sinistra il prodotto è in $M_{\mathbf{K}}(m,n)$, a destra in $M_{\mathbf{K}}(m,p)$.

3. Proprietà associativa: $\forall\ A \in M_{\mathbf{K}}(m,n)$, $\forall\ B \in M_{\mathbf{K}}(n,p)$, $\forall\ C \in M_{\mathbf{K}}(p,q)$, $(AB)\,C = A\,(BC)$.

   A sinistra il primo prodotto è in $M_{\mathbf{K}}(m,p)$ e il secondo in $M_{\mathbf{K}}(m,q)$, a destra il secondo prodotto è in $M_{\mathbf{K}}(n,q)$ e il primo in $M_{\mathbf{K}}(m,q)$.

4. Omogeneità: $\forall\ t \in \mathbf{K}$, $\forall\ A \in M_{\mathbf{K}}(n,m)$, $\forall\ B \in M_{\mathbf{K}}(m,p)$, $t\,(AB) = (tA)\,B = A\,(tB)$.

5. Esistenza dell'elemento neutro: se voglio che $A? = A$, la matrice incognita deve avere dimensione $n \times n$, se voglio $?A = A$ deve avere dimensione $m \times m$. La matrice incognita è la matrice identità, quadrata, con a pedice la dimensione: $Id_n = \begin{bmatrix} 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 1 \end{bmatrix} = \begin{cases} 1 \text{ se } i = j \\ 0 \text{ se } i \neq j \end{cases}$. Vale $AId_n = A = Id_mA$.

Per il prodotto matriciale non vale la legge di annullamento del prodotto, cioè non è vero che se AB è nulla allora A è nulla o B è nulla: l'elemento neutro rispetto alla somma si può ottenere senza che A o B siano nulle. Infatti:

1. $A = \begin{bmatrix} 1 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, $AB = [0]$

2. $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, $AB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Inoltre il prodotto tra matrici in generale non ha la proprietà commutativa. Infatti, se $A \in M_{\mathbf{K}}(m,n)$ e $B \in M_{\mathbf{K}}(n,p)$, si può calcolare AB, ma se $p \neq m$ non si può calcolare BA. Suppongo allora $m = p$: $AB \in M_{\mathbf{K}}(m,m)$, mentre $BA \in M_{\mathbf{K}}(n,n)$; quindi, se $m \neq n$, $AB \neq BA$ e non vale la proprietà commutativa. Suppongo allora $m = n = p$: ma la proprietà commutativa non vale neanche per matrici quadrate. Infatti:

1. $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}$, $AB = \begin{bmatrix} 5 & 6 \\ 11 & 12 \end{bmatrix} \neq BA = \begin{bmatrix} 1 & 2 \\ 11 & 16 \end{bmatrix}$.

In alcuni casi il prodotto può essere commutativo: data $A \in M_{\mathbf{K}}(n,n)$, $0_M \in M_{\mathbf{K}}(n,n)$, $A0_{n,n} = 0_{n,n}A = 0_M \in M_{\mathbf{K}}(n,n)$; data $A \in M_{\mathbf{K}}(n,n)$, $AId_n = Id_nA = A \in M_{\mathbf{K}}(n,n)$; $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, $AB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = BA = 0_M \in M_{\mathbf{K}}(2,2)$.

1. $A = B$ non è equivalente a $CA = CB$. $A = B \Longrightarrow CA = CB$, ma non vale in generale l'implicazione opposta: e. g. $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$, ma $\begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \neq \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$. $A = B \Longleftrightarrow CA = CB$ quando C rappresenta un'applicazione lineare iniettiva (?).

1.2 Matrice inversa

Def Una matrice $A \in M_{\mathbf{K}}(m,n)$ si dice invertibile a destra se esiste $B \in M_{\mathbf{K}}(n,m) : AB = Id_m$. Una matrice $A \in M_{\mathbf{K}}(m,n)$ si dice invertibile a sinistra se esiste $C \in M_{\mathbf{K}}(n,m) : CA = Id_n$. Una matrice $A \in M_{\mathbf{K}}(m,n)$ si dice invertibile se è invertibile sia a destra che a sinistra.

Non tutte le matrici sono invertibili, quindi l'insieme delle matrici non è un campo con l'operazione di prodotto matriciale. Una matrice non invertibile si dice anche singolare.

Proposizione

Hp: $A \in M_{\mathbf{K}}(m,n)$ è invertibile

Ts: (i) L'inversa destra è uguale all'inversa sinistra

(ii) L'inversa è unica

Dim (1) Essendo A invertibile, $\exists\ B \in M_{\mathbf{K}}(n,m) : AB = Id_m$ e $\exists\ C \in M_{\mathbf{K}}(n,m) : CA = Id_n$. Considero $B \in M_{\mathbf{K}}(n,m)$: poiché esiste l'elemento neutro, $B = Id_nB = (CA)B$ per ipotesi. Poiché vale la proprietà associativa e B è l'inversa destra di A, $(CA)B = C(AB) = CId_m = C$, dunque $B = C$ e le inverse sono uguali.

(2) Dimostro che, date due generiche inverse di A, queste sono uguali. Considero due generiche inverse $B_1, B_2 \in M_{\mathbf{K}}(n,m)$: vale $AB_1 = AB_2 = Id_m$ e, poiché l'inversa destra è uguale alla sinistra, $B_1A = B_2A = Id_n$. Allora $B_1 = Id_nB_1 = (B_2A)B_1 = B_2(AB_1) = B_2Id_m = B_2$. ∎

Si indica con A^{-1} l'unica matrice inversa di una matrice invertibile A.

1. Provo a invertire una matrice $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in M_{\mathbf{K}}(2,3)$. Esiste un'inversa destra, cioè una matrice $X \in M_{\mathbf{K}}(3,2) : AX = Id_2$? Voglio $X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} : AX = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Calcolo $AX = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} = \begin{bmatrix} x_{11} + 2x_{21} + 3x_{31} & x_{12} + 2x_{22} + 3x_{32} \\ 4x_{11} + 5x_{21} + 6x_{31} & 4x_{12} + 4x_{22} + 6x_{32} \end{bmatrix}$. Eguaglio questa matrice a Id_2, elemento per elemento: $\begin{cases} x_{11} + 2x_{21} + 3x_{31} = 1 \\ x_{12} + 2x_{22} + 3x_{32} = 0 \\ 4x_{11} + 5x_{21} + 6x_{31} = 0 \\ 4x_{12} + 4x_{22} + 6x_{32} = 1 \end{cases}$. E' un sistema a sei incognite e quattro equazioni: l'intuizione suggerisce che abbia infinite soluzioni, perché i gradi di libertà (cioè il numero di equazioni) sono più dei vincoli.

Esiste un'inversa sinistra, cioè una matrice $X \in M_{\mathbf{K}}(3,2) : XA = Id_3$? Voglio $X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} : XA = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Calcolo $XA = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} x_{11} + 4x_{12} & 2x_{11} + 5x_{12} & 3x_{11} + 6x_{12} \\ x_{21} + 4x_{22} & 2x_{21} + 5x_{22} & 3x_{21} + 6x_{22} \\ x_{31} + 4x_{32} & 2x_{31} + 5x_{32} & 3x_{31} + 6x_{32} \end{bmatrix}$. Uguaglia questa matrice a Id_3, elemento per elemento, e ottengo un sistema a sei incognite e nove equazioni: l'intuizione suggerisce che non abbia soluzioni, perché i gradi di libertà sono meno dei vincoli. Il sistema è l'unione di tre sistemi, ciascuno dei quali con due incognite: $\begin{cases} x_{11} + 4x_{12} = 1 \\ 2x_{11} + 5x_{12} = 0 \\ 3x_{11} + 6x_{12} = 0 \end{cases}$, $\begin{cases} x_{21} + 4x_{22} = 0 \\ 2x_{21} + 5x_{22} = 1 \\ 3x_{21} + 6x_{22} = 0 \end{cases}$, $\begin{cases} x_{31} + 4x_{32} = 0 \\ 2x_{31} + 5x_{32} = 0 \\ 3x_{31} + 6x_{32} = 1 \end{cases}$. Dal primo, sostituendo, si ha $\begin{cases} x_{11} + 4x_{12} = 1 \\ 2x_{11} + 5x_{12} = 0 \\ x_{11} = -2x_{12} \end{cases} \iff \begin{cases} -2x_{12} + 4x_{12} = 1 \\ 2x_{11} + 5x_{12} = x_{12} = 0 \\ x_{11} = -2x_{12} \end{cases}$. Poiché dalle prime due equazioni si ricava che $x_{12} = x_{11} = 0$ e sostituendo nella prima si ha $0 = 1$, si ricava che il sistema è impossibile. Dunque la matrice non è invertibile.

Quindi, se $m \neq n$, ricercando l'inversa destra si ottiene un sistema con più incognite che equazioni, ricercando l'inversa sinistra un sistema con più equazioni che incognite. Rimane la speranza di poter calcolare la matrice inversa se $m = n$.

2 Sistemi lineari

Dato il sistema $\begin{cases} x - 3y = 2 \\ 4x + 5y = 1 \end{cases}$, se ne può dare una rappresentazione matriciale raccogliendo le incognite: $\begin{bmatrix} x - 3y \\ 4x + 5y \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \iff \begin{bmatrix} x \\ 4x \end{bmatrix} + \begin{bmatrix} -3y \\ 5y \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \iff x \begin{bmatrix} 1 \\ 4 \end{bmatrix} + y \begin{bmatrix} -3 \\ 5 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ (primo tipo di rappresentazione matriciale del sistema). Utilizzando il prodotto tra matrici, vale anche $x \begin{bmatrix} 1 \\ 4 \end{bmatrix} + y \begin{bmatrix} -3 \\ 5 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \iff \begin{bmatrix} 1 & -3 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$: quest'ultimo è il secondo tipo di rappresentazione matriciale del sistema. Ma x e y sono variabili mute, quindi è inutile scrivere il vettore delle incognite. Si può quindi unire tutto in una matrice $\begin{bmatrix} 1 & -3 & 2 \\ 4 & 5 & 1 \end{bmatrix}$, giustapponendo a destra il vettore dei termini noti: questo è il terzo tipo di rappresentazione matriciale di un sistema.

In generale, un sistema lineare è un insieme di m equazioni con n incognite, del tipo

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

che è una matrice $m \times 1$ dove $x_1, x_2, ..., x_n$ sono le n incognite. Tale sistema si può rappresentare usando le matrici in tre modi:

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

2. $\begin{bmatrix} a_{11} & ... & a_{1n} \\ ... & ... & ... \\ a_{m1} & ... & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ ... \\ b_m \end{bmatrix}$ dove $\begin{bmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{bmatrix}$ è il vettore delle incognite. Il generico sistema lineare, in forma estremamente sintetica, si scrive $A\mathbf{x} = \mathbf{b}$, dove A è la matrice dei coefficienti, $\mathbf{x}$ il vettore delle incognite, $\mathbf{b}$ il vettore dei termini noti.

3. $\begin{bmatrix} a_{11} & ... & a_{1n} & b_1 \\ ... & ... & ... & ... \\ a_{m1} & ... & a_{mn} & b_m \end{bmatrix} \in M_{\mathbf{K}}(m, n+1)$ è la matrice completa o estesa del sistema, che non riporta le incognite.

2.1 Metodi di risoluzione e matrici a scala

Tra i sistemi $\begin{cases} x + y - 2z = 0 \\ x - y + 3z = 0 \\ 2x - 2y + z = -7 \end{cases}$ e $\begin{cases} x + y - 3z = 4 \\ y + 2z = 1 \\ z = -4 \end{cases}$, il secondo è più facile da risolvere perché la seconda equazione è in due incognite e la terza in una sola incognita. Questo è molto evidente scrivendo la rappresentazione matriciale del terzo tipo dei due sistemi: $\begin{bmatrix} 1 & 1 & -2 & 0 \\ 1 & -1 & 3 & 1 \\ 2 & -2 & 1 & -7 \end{bmatrix}$ e $\begin{bmatrix} 1 & 1 & -3 & 4 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & -4 \end{bmatrix}$. La matrice dei coefficienti del secondo sistema è tale che le ultime due righe costituiscono un sistema di due equazioni in due incognite e l'ultima riga è un'equazione in un'incognita: questo tipo di matrice prende il nome di matrice a scala.

Def Sia U una matrice in $M_{\mathbf{K}}(m, n)$, siano U_1, U_2,..., U_m le sue righe, con $U_i \in M_{\mathbf{K}}(1, n)$. Si chiama pivot della riga U_i il (eventuale?) primo elemento non nullo, a partire da sinistra, di U_i. La matrice U si dice a scala se

- date due righe consecutive U_i, U_{i+1}, entrambe non nulle, l'indice di colonna del pivot di U_i è minore dell'indice di colonna del pivot di U_{i+1} (il pivot di U_i compare a sinistra rispetto al pivot di U_{i+1})

- se U_i è nulla, allora sono nulle tutte le righe successive.

Tutte le matrici con una riga sono a scala.

1. $\begin{bmatrix} 0 & p_1 & * & * & * & * & * \\ 0 & 0 & p_2 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & p_3 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ è una matrice a scala i cui pivot sono p_1, p_2 e p_3; i $*$ indicano elementi irrilevanti; e. g. p_2 non può essere nella prima colonna.

Risolvo il primo sistema col metodo di riduzione di Gauss: l'obiettivo è rendere la matrice dei coefficienti una matrice a scala. $\begin{cases} x + y - 2z = 0 \\ x - y + 3z = 0 \\ 2x - 2y + z = -7 \end{cases} \iff (A|\mathbf{b})^{(1)} = \begin{bmatrix} 1 & 1 & -2 & 0 \\ 1 & -1 & 3 & 1 \\ 2 & -2 & 1 & -7 \end{bmatrix}$. Considero il pivot della prima riga: il pivot d ella seconda riga deve essere a destra, quindi il primo elemento della seconda riga dev'essere nullo. Analogamente per la terza riga: faccio $(3) = (3) - 2\,(2)$ e $(2) = (2) - (1)$, e ottengo $(A|\mathbf{b})^{(2)} = \begin{bmatrix} 1 & 1 & -2 & 0 \\ 0 & -2 & 5 & 1 \\ 0 & 0 & -5 & -9 \end{bmatrix}$.

La matrice dei coefficienti è una matrice a scala: si è ottenuta con il **metodo di riduzione di Gauss**. Adesso cerco di avere tutti i pivot uguali a 1: $(2) = \frac{-1}{2}\,(2)$, $(3) = \frac{-1}{5}\,(3)$ e ottengo $\begin{bmatrix} 1 & 1 & -2 & 0 \\ 0 & 1 & -\frac{5}{2} & -\frac{1}{2} \\ 0 & 0 & 1 & \frac{9}{5} \end{bmatrix}$. Adesso, seguendo il metodo di riduzione di Gauss-Jordan, cerco di rendere la matrice dei coefficienti una matrice identità: modifico le righe dal basso verso l'alto, perché gli elementi nulli prima del pivot permettono di conservare la parte di matrice che non voglio modificare. Faccio $(2) = (2) + \frac{5}{2}\,(3)$ e $(1) = (1) + 2\,(3)$ e ottengo $\begin{bmatrix} 1 & 1 & 0 & \frac{18}{5} \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & \frac{9}{5} \end{bmatrix}$. Manca un solo elemento da modificare: faccio $(1) = (1) - (2)$ e ottengo $\begin{bmatrix} 1 & 0 & 0 & -\frac{2}{5} \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & \frac{9}{5} \end{bmatrix}$. Avere i pivot uguali a 1 e la matrice dei coefficienti identità permette di ottenere il valore esatto delle incognite nel vettore dei termini noti: infatti quest'ultima matrice è la rappresentazione del sistema $\begin{cases} x = -\frac{2}{5} \\ y = 4 \\ z = \frac{9}{5} \end{cases}$

Quindi è possibile usare il metodo di Gauss e risolvere il sistema partendo dalla sua rappresentazione con matrice a scala e sostituendo a ritroso, oppure si può seguire il metodo di Gauss-Jordan e risolvere il sistema lavorando direttamente sulla matrice. Il primo richiede meno operazioni e permette di capire più in fretta se il sistema ha soluzione, mentre il secondo permette di ottenere direttamente la soluzione e porta inoltre a un'unica matrice (SE c'è il risultato), a differenza del metodo di Gauss.

In generale, dato un sistema lineare, il metodo di riduzione permette di:

1. moltiplicare un'equazione (una riga) per una costante non nulla.
2. scambiare la posizione di due righe, e. g. se si trova una riga nulla;
3. sostituire a una riga U_i la somma di U_i e un multiplo di un'altra riga;

Proposizione

Hp: $A\mathbf{x} = \mathbf{b}$ è un sistema lineare; $A'\mathbf{x} = \mathbf{b}'$ è un sistema lineare ottenuto dal sistema di partenza mediante una sequenza di operazioni sulle righe

Ts: $A\mathbf{x} = \mathbf{b}$ e $A'\mathbf{x} = \mathbf{b}'$ hanno le stesse soluzioni

Questo significa che le operazioni sulle righe che si fanno col metodo di Gauss non modificano le soluzioni del sistema (lo spazio delle righe si conserva). L'idea della dimostrazione è mostrare che le operazioni sulle righe sono reversibili e in un singolo passaggio le soluzioni non cambiano.

Teorema

Hp: $A \in M_{\mathbf{K}}(m, n)$

Ts: A può essere trasformata in una matrice a scala con una sequenza finita di operazioni sulle righe

Dim La dimostrazione è costruttiva e si basa su un procedimento ricorsivo. Se A è a scala, non devo fare niente. Suppongo che A non sia a scala: questo garantisce che ci sia almeno un elemento non nullo (una matrice nulla è a scala). Fisso l'indice j della prima colonna da sinistra contenente un elemento non nullo; fisso l'indice i di una qualsiasi riga tale che $a_{ij} \neq 0$ e scambio la prima riga con la riga i-esima (voglio "schiacciare" l'informazione in alto). Ora l'elemento non nullo nella prima riga è un pivot, che chiamo p_1; cancello tutti gli elementi non nulli che si trovano sotto al primo pivot (nella colonna j), facendo (per ogni riga i sotto la prima tale che l'elemento sulla j-esima colonna è non nullo) l'operazione $R_i - \frac{a_{ij}}{p_1}R_1 \to R_i$. Dopo queste operazioni, tutti gli elementi sotto p_1 sono nulli; ripeto il procedimento per la sottomatrice formata dalle righe $(2, ..., m)$ e dalle colonne $(j+1, ..., n)$. L'algoritmo termina perché a ogni iterazione del procedimento la nuova matrice da ridurre ha meno righe: si arriverà ad avere quindi una riga. $\blacksquare$

L'elemento $\frac{a_{ij}}{p_1}$ è detto moltiplicatore.

2.2 Soluzioni di un sistema

Def Un sistema lineare si dice omogeneo se tutti i termini noti sono nulli: $A\mathbf{x} = \mathbf{0}$. Dato un sistema lineare qualsiasi $A\mathbf{x} = \mathbf{b}$, si dice sistema omogeneo associato ad esso il sistema $A\mathbf{x} = \mathbf{0}$.

Un sistema omogeneo ammette sempre almeno una soluzione: c'è sempre la soluzione nulla $\mathbf{x} = \mathbf{0}$. In generale, ammette una soluzione o infinite soluzioni: se $\mathbf{v}$ è una soluzione, anche $\lambda\mathbf{v}$ lo è; se $\mathbf{v}, \mathbf{w}$ sono soluzioni, anche $\mathbf{v} + \mathbf{w}$ lo è.

Teorema (struttura generale della soluzione di un sistema lineare)

Hp:$A\mathbf{x} = \mathbf{b}$ è un sistema lineare, $A\mathbf{x} = \mathbf{0}$ è il sistema omogeneo

ad esso associato, $\mathbf{v}_0$ è una soluzione di $A\mathbf{x} = \mathbf{b}$

Ts: tutte e sole le soluzioni di $A\mathbf{x} = \mathbf{b}$ si ottengono sommando

a $\mathbf{v}_0$ una soluzione $\mathbf{v}_h$ di $A\mathbf{x} = \mathbf{0}$, cioè $\{\mathbf{x} \in \mathbf{K}^n : A\mathbf{x} = \mathbf{b}\} =$

$\{\mathbf{x} \in \mathbf{K}^n : \mathbf{x} = \mathbf{v}_0 + \mathbf{v}_h$, al variare di $\mathbf{v}_h$ nell'insieme delle soluzioni del sistema omogeneo$\}$

Dim Devo dimostrare che due insiemi coincidono, quindi ne mostro la mutua inclusione. (1) Mostro che se ho una soluzione particolare e considero una soluzione del sistema omogeneo associato, posso determinare una soluzione del sistema lineare $A\mathbf{x} = \mathbf{b}$. Sia $\mathbf{v}_h$ una soluzione di $A\mathbf{x} = \mathbf{0}$: allora $\mathbf{v}_0 + \mathbf{v}_h$ è una soluzione di $A\mathbf{x} = \mathbf{b}$. Infatti $A(\mathbf{v}_0 + \mathbf{v}_h) = A\mathbf{v}_0 + A\mathbf{v}_h$ per la proprietà distributiva a destra del prodotto matriciale, e $A\mathbf{v}_0 + A\mathbf{v}_h = \mathbf{b} + \mathbf{0} = \mathbf{b}$ per ipotesi e per l'esistenza dell'elemento neutro della somma.

(2) Mostro che se considero una seconda soluzione $\mathbf{v}_1$ di $A\mathbf{x} = \mathbf{b}$, questa può essere scritta come somma di $\mathbf{v}_0$ e una soluzione del sistema omogeneo, cioè $\exists\ \mathbf{v}_h : A\mathbf{v}_h = \mathbf{0}$ tale che $\mathbf{v}_1 = \mathbf{v}_0 + \mathbf{v}_h$. Questo è equivalente a mostrare che $\mathbf{v}_h = \mathbf{v}_1 - \mathbf{v}_0$ è soluzione del sistema omogeneo. Infatti, $A(\mathbf{v}_1 - \mathbf{v}_0) = A\mathbf{v}_1 - A\mathbf{v}_0 = \mathbf{b} - \mathbf{b} = \mathbf{0}$, per la proprietà distributiva a destra del prodotto matriciale, per ipotesi e per esistenza dell'elemento opposto. ■

Questo teorema è estremamente utile per risolvere sistemi lineari che condividono la matrice dei coefficienti.

1. Dati $a, b \in \mathbb{R}$, il sistema $\begin{cases} 3x - 2y = 1 \\ 6x + ay = b \end{cases}$ è risolubile? Se sì, che soluzioni ha?

   La matrice completa del sistema è $\begin{bmatrix} 3 & -2 & 1 \\ 6 & a & b \end{bmatrix}$: la riduco a scala facendo $(2) = (2) - 2(1)$ e ottengo $\begin{bmatrix} 3 & -2 & 1 \\ 0 & a+4 & b-2 \end{bmatrix}$. Ci sono tre casi di risolubilità del sistema:

(a) Se $a+4 \neq 0$, posso procedere con il MEGJ e determinare le soluzioni del sistema: $\begin{bmatrix} 3 & -2 & 1 \\ 0 & a+4 & b-2 \end{bmatrix} \iff$ $\begin{bmatrix} 3 & -2 & 1 \\ 0 & 1 & \frac{b-2}{a+4} \end{bmatrix} \iff \begin{bmatrix} 3 & 0 & 1+2\frac{b-2}{a+4} \\ 0 & 1 & \frac{b-2}{a+4} \end{bmatrix} \iff \begin{bmatrix} 1 & 0 & \frac{1}{3}+\frac{2}{3}\frac{b-2}{a+4} \\ 0 & 1 & \frac{b-2}{a+4} \end{bmatrix}$. La soluzione è quindi unica. (Caso in cui nel nucleo c'è solo la soluzione nulla, vettori linearmente indipendenti, $r\,(A|b) = n$)

(b) Se $a+4=0$ e $b-2 \neq 0$, il sistema $\begin{bmatrix} 3 & -2 & 1 \\ 0 & 0 & b-2 \end{bmatrix}$ è impossibile.

(c) Se $a+4=0$ e $b-2=0$, il sistema $\begin{bmatrix} 3 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ presenta una riga non informativa: la seconda, che si può ignorare. Si deve allora risolvere il sistema $\{3x-2y=1$, che è risolubile con infinite soluzioni, dipendenti da un parametro. Scrivo le soluzioni in dipendenza di un parametro $t \in \mathbb{R}$: pongo $y=t$ e risolvo $\begin{cases} 3x-2y=1 \\ y=t \end{cases}$: $\begin{bmatrix} 3 & -2 & 1 \\ 0 & 1 & t \end{bmatrix} \iff \begin{bmatrix} 3 & 0 & 1+2t \\ 0 & 1 & t \end{bmatrix} \iff \begin{bmatrix} 1 & 0 & \frac{1}{3}+\frac{2}{3}t \\ 0 & 1 & t \end{bmatrix}$. Dunque $x = \frac{1}{3}+\frac{2}{3}t$, $y=t$, al variare di $t \in \mathbb{R}$. Rappresento le soluzioni in forma matriciale: $\begin{bmatrix} x \\ y \end{bmatrix} =$ $\begin{bmatrix} \frac{1}{3}+\frac{2}{3}t \\ t \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ 0 \end{bmatrix} + t\begin{bmatrix} \frac{2}{3} \\ 1 \end{bmatrix}$. Questo modo di scrivere la soluzione ricorda la struttura delle soluzioni enunciata nel teorema appena visto. Questo fa pensare che il contributo del parametro t non dipenda dal vettore dei termini noti, ma costituisca l'inseme delle soluzioni del sistema omogeneo associato (infatti $\begin{bmatrix} \frac{2}{3} \\ 1 \end{bmatrix}$ è soluzione del sistema omogeneo: $\begin{bmatrix} 3 & -2 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \frac{2}{3} \\ 1 \end{bmatrix} = \begin{bmatrix} 2-2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, e ogni vettore $t\begin{bmatrix} \frac{2}{3} \\ 1 \end{bmatrix}$ è soluzione del sistema omogeneo). $\begin{bmatrix} \frac{1}{3} \\ 0 \end{bmatrix}$ sarà invece una soluzione del sistema $A\mathbf{x} = \mathbf{b}$.

Dal punto di vista geometrico $3x-2y=1$ è l'equazione cartesiana di una retta nel piano $\mathbb{R}^2$: tale soluzione è rappresentata come $\begin{bmatrix} \frac{1}{3} \\ 0 \end{bmatrix} + t\begin{bmatrix} \frac{2}{3} \\ 1 \end{bmatrix}$. Se scrivessi analogamente $3x-2y=-2$ (cambiando il termine noto), otterrei $\begin{bmatrix} -\frac{2}{3} \\ 0 \end{bmatrix} + t\begin{bmatrix} \frac{2}{3} \\ 1 \end{bmatrix}$: modificare il termine noto equivale ad avere come soluzione una retta traslata verticalmente rispetto alla prima trovata. Per tale sistema quindi la soluzione è sempre una retta di coefficiente angolare fisso (determinato dal sistema omogeneo associato) e intercetta variabile a seconda del vettore dei termini noti $\mathbf{v}_0$. Infatti $3x-2y=1 \iff y = -\frac{1}{2}+\frac{3}{2}x$: mettere un altro numero al posto dell'1 inficia solo l'intercetta.

Nel primo caso il numero di pivot di A è uguale al numero di pivot di $[A|\mathbf{b}]$, e c'è un'unica soluzione. Nel secondo il numero di pivot di A è diverso da quello di $[A|\mathbf{b}]$, e non c'è alcuna soluzione. Nel terzo il numero di pivot è uguale e c'è una riga nulla: ci sono infinite soluzioni. Il numero di pivot di A è in relazione con la risolubilità del sistema.

2.3 Rango e traccia

Def Data una matrice a scala U, si definisce rango di U il suo numero di pivot. Data una matrice qualsiasi A, si definisce rango di A il numero di pivot di una qualsiasi matrice a scala ottenuta a partire da A con il metodo di eliminazione di Gauss.

Il rango della matrice A si indica con $r(A)$, $rk(A)$, $rg(A)$, $\rho(A)$.

La definizione è ben posta perché, nonostante si possano ottenere diverse matrici a scala da A con il MEG, tutte hanno lo stesso numero di pivot (si vedrà più avanti che $\dim(rowA) = \dim(rowU)$).

Dalla definizione segue che, data $A \in M_{\mathbf{K}}(m,n)$, $r(A) \geq 0$, $r(A) \in \mathbb{N}$; dal momento che nella matrice a scala c'è al più un pivot per riga, $r(A) \leq m$; poiché c'è al più un pivot per colonna, $r(A) \leq n$. Dunque $0 \leq r(A) \leq \min\{m,n\}$.

Il rango di A è il numero di righe non nulle di A ridotta a scala: se A è la matrice incompleta di un sistema, $r(A)$ rappresenta il numero di equazioni significative.

Def Data una matrice $A \in M_{\mathbf{K}}(n,n)$, si definisce traccia $tr(A)$ la somma degli elementi sulla sua diagonale: $tr(A) = \sum_{i=1}^{n} a_{ii}$.

2.4 Teorema di Rouché-Capelli

$$\text{(1) Hp: } A\mathbf{x} = \mathbf{b} \text{ è un sistema lineare con } n \text{ incognite}$$

$$\text{Ts: } A\mathbf{x} = \mathbf{b} \text{ ammette almeno una soluzione } \iff r(A|\mathbf{b}) = r(A)$$

$$\text{(2) Hp: } A\mathbf{x} = \mathbf{b} \text{ è un sistema lineare con } n \text{ incognite; } r(A|\mathbf{b}) = r(A)$$

$$\text{Ts: esistono } n-r+1 \text{ vettori colonna } \mathbf{v}_0, \mathbf{v}_1, ..., \mathbf{v}_{n-r} \in \mathbf{K}^n \text{ tali che la generica soluzione}$$

$$\text{del sistema è descritta da } \mathbf{x} = \mathbf{v}_0 + t_1\mathbf{v}_1 + ... + t_{n-r}\mathbf{v}_{n-r},\ t_i \in \mathbf{K}$$

$\mathbf{v}_0$ è usato con lo stesso significato del teorema di struttura.

Dim Dimostro l'implicazione da sinistra a destra della prima tesi in forma negata. Riduco a scala la matrice completa del sistema $[A|\mathbf{b}]$ e ottengo, con il MEG, $[U|\mathbf{b}']$. Pongo $r(A) = r(U) = r$, che è il numero di pivot di

U. A questo punto, $[A|\mathbf{b}]$ ha due possibili configurazioni: o la colonna dei termini noti non contiene alcun pivot, oppure ne contiene uno (non può contenerne più d'uno perché due pivot non possono stare sulla stessa colonna).

$$\begin{bmatrix} 0 & p_1 & * & * & * & \\ 0 & 0 & p_2 & * & * & \\ ... & ... & ... & ... & ... & \\ 0 & 0 & 0 & 0 & p_r & \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

cioè tutti i pivot sono contenuti di U e nessuno in $\mathbf{b}'$: in questo caso $r(A) = r(A|\mathbf{b})$; oppure

$$\begin{bmatrix} 0 & p_1 & * & * & * & \\ 0 & 0 & p_2 & * & * & \\ ... & ... & ... & ... & ... & \\ 0 & 0 & 0 & 0 & p_r & \\ 0 & 0 & 0 & 0 & 0 & b'_{r+1} \end{bmatrix}$$

con $b'_{r+1} \neq 0$, cioè la colonna dei termini noti contiene un pivot e $r(A|\mathbf{b}) = r(A) + 1$. Per definizione di pivot, tutti gli elementi precedenti a b'_{r+1} devono essere nulli, quindi la riga $r+1$ descrive l'equazione $0 = b'_{r+1}$, che è impossibile. Dunque, se $r(A|\mathbf{b}) = r(A) + 1$, il sistema non ha soluzione. Così ho dimostrato l'implicazione da sinistra a destra della prima tesi.

Per dimostrare la seconda tesi e l'implicazione da destra a sinistra della prima tesi, suppongo ora $r(A) = r(A|\mathbf{b})$: cancello le eventuali righe nulle e riparto da un sistema con r equazioni in n incognite. Chiamo variabili dipendenti le variabili corrispondenti alle r colonne contenenti i pivot; chiamo le altre variabili indipendenti. Suppongo senza perdita di generalità che le variabili dipendenti siano le prime: $x_1, ..., x_r$ (se non lo sono, le riordino: sono variabili mute). A questo punto, nella parte sinistra della matrice (le prime r colonne) ho una scala perfetta. Le altre $n-r$ colonne sono quelle corrispondenti a variabili indipendenti.

$$\begin{bmatrix} p_1 & * & * & * & * & b'_1 \\ 0 & p_2 & * & * & * & b'_2 \\ 0 & 0 & p_3 & * & * & b'_3 \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & p_3 & * & b'_{r+1} \end{bmatrix}$$

Considero le seguenti equazioni, una per ogni variabile indipendente: $x_{r+1} = t_1, ..., x_n = t_{n-r}$, con $t_i \in \mathbf{K}$, e le aggiungo al sistema. Ottengo una matrice a scala con n righe e $n+1$ colonne.

$$\begin{bmatrix} p_1 & * & * & * & * & * & * & * & b'_1 \\ 0 & p_2 & * & * & * & * & * & * & b'_2 \\ 0 & 0 & ... & * & * & * & * & * & ... \\ 0 & 0 & 0 & p_r & * & * & * & * & b'_r \\ 0 & 0 & 0 & 0 & 1 & * & * & * & t_1 \\ 0 & 0 & 0 & 0 & 0 & 1 & * & * & t_2 \\ ... & ... & ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & t_{n-r} \end{bmatrix}$$

Le equazioni dalla $r+1$ alla n sono risolte in partenza. Quindi uso il metodo di Gauss-Jordan dal basso verso l'alto per trovare il valore (parametrico) di ogni variabile $x_1, ..., x_n$, e ottengo

$$\begin{cases} x_1 = v_{10} + v_{11}t_1 + ... + v_{1,n-r}t_{n-r} \\ ... \\ x_r = v_{r0} + v_{r1}t_1 + ... + v_{r,n-r}t_{n-r} \\ x_{r+1} = t_1 \\ ... \\ x_n = t_{n-r} \end{cases}$$

dove $v_{ij} \in \mathbf{K}$ per ogni $i = 1, ..., r,\ j = 0, ..., n-r$. Quindi ogni componente del vettore soluzione ha una parte costante: la componente i-esima ha per parte costante v_{i0}. Essendo $r(A|\mathbf{b}) = r(A)$, esiste almeno una soluzione, e ne ho scritta almeno una; inoltre ho mostrato che esistono gli $n-r+1$ vettori della seconda tesi tali che $\mathbf{x} = \mathbf{v}_0 + t_1\mathbf{v}_1 + ... + t_{n-r}\mathbf{v}_{n-r}$ è soluzione $\forall\ (t_1, ..., t_{n-r}) \in \mathbf{K}$.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ ... \\ x_r \\ x_{r+1} \\ ... \\ x_n \end{bmatrix} = \begin{bmatrix} v_{10} + \sum_{k=1}^{n-r} v_{1k}t_k \\ ... \\ v_{r0} + \sum_{k=1}^{n-r} v_{rk}t_k \\ t_1 \\ ... \\ t_{n-r} \end{bmatrix} = \begin{bmatrix} v_{10} \\ ... \\ v_{r0} \\ 0 \\ ... \\ 0 \end{bmatrix} + t_1 \begin{bmatrix} v_{11} \\ ... \\ v_{r1} \\ 1 \\ ... \\ 0 \end{bmatrix} + ... + t_{n-r} \begin{bmatrix} v_{1,n-r} \\ ... \\ 0 \\ ... \\ 1 \end{bmatrix} = \mathbf{v}_0 + t_1\mathbf{v}_1 + ... + t_{n-r}\mathbf{v}_{n-r}$$

■

Il numero di equazioni m non ha alcuna influenza sull'enunciato. Se $r(A|\mathbf{b}) = r(A) = n$, ci sono n pivot e, essendoci n colonne, non esistono variabili indipendenti: dunque la matrice incompleta a scala è quadrata ed esiste un'unica soluzione, priva di parametri. $\mathbf{v}_0$ (che viene detto soluzione speciale o particolare) è lo stesso vettore del teorema di struttura: questo significa che $t_1\mathbf{v}_1 + ... + t_{n-r}\mathbf{v}_{n-r}$ è l'insieme delle soluzioni del sistema omogeneo associato. Infatti, essendo $\mathbf{v}_0$ $(t_1 = ... = t_{n-r} = 0)$ e $\mathbf{x} = \mathbf{v}_0 + t_1\mathbf{v}_1 + ... + t_{n-r}\mathbf{v}_{n-r}$ soluzioni di $A\mathbf{x} = \mathbf{b}$, si dimostra che $\mathbf{x} - \mathbf{v}_0$ è soluzione del sistema omogeneo: $A(\mathbf{x} - \mathbf{v}_0) = A\mathbf{x} - A\mathbf{v}_0 = \mathbf{b} - \mathbf{b} = 0$, per la proprietà distributiva a destra del prodotto matriciale.

1. *Studiare le soluzioni del sistema* $\begin{bmatrix} 5 & 2 & 1 & 1+k \\ 2 & 2 & -2 & k \\ 1 & -2 & 5 & 1+2k \end{bmatrix}$ *al variare di* $k \in \mathbb{R}$.

   Dovendo le soluzioni dipendere da un parametro, mi aspetto che non ce ne sia una sola (con $r(A) = r(A|b)$ avrei $3 - 3 = 0$ parametri liberi), e quindi che $r(A) = r(A|b) < 3$. Riduco a scala la matrice completa: scambio (1) e (3), faccio $(2) = (2) - 2(1)$ e $(3) = (3) - 5(1)$, faccio $(3) = (3) - 2(2)$.

$$\begin{bmatrix} 5 & 2 & 1 & 1+k \\ 2 & 2 & -2 & k \\ 1 & -2 & 5 & 1+2k \end{bmatrix} \Longleftrightarrow \begin{bmatrix} 1 & -2 & 5 & 1+2k \\ 2 & 2 & -2 & k \\ 5 & 2 & 1 & 1+k \end{bmatrix} \Longleftrightarrow \begin{bmatrix} 1 & -2 & 5 & 1+2k \\ 0 & 6 & -12 & -3k-2 \\ 0 & 12 & -24 & -4-9k \end{bmatrix} \Longleftrightarrow \begin{bmatrix} 1 & -2 & 5 & 1+2k \\ 0 & 6 & -12 & -3k-2 \\ 0 & 0 & 0 & -3k \end{bmatrix}$$

   Se $-3k \neq 0$, $r(A) = 2 \neq r(A|b) = 3$, dunque il sistema non ha soluzione, per il teorema di Rouché-Capelli.

   Se $-3k = 0$, $r(A) = r(A|b) = 2$ ed esistono infinite soluzioni dipendenti da un parametro. Il sistema diventa

$$\begin{bmatrix} 1 & -2 & 5 & 1 \\ 0 & 6 & -12 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

   dove x e y sono le variabili dipendenti, z è la variabile indipendente. Aggiungo al sistema l'equazione $z = t$ e uso il MEGJ: faccio $(2) = (2) + 12(3)$ e $(1) = (1) - 5(3)$, $(2) = \frac{1}{6}(2)$, $(1) = (1) + 2(2)$

$$\begin{bmatrix} 1 & -2 & 5 & 1 \\ 0 & 6 & -12 & -2 \\ 0 & 0 & 1 & t \end{bmatrix} \Longleftrightarrow \begin{bmatrix} 1 & -2 & 0 & 1-5t \\ 0 & 6 & 0 & -2+12t \\ 0 & 0 & 1 & t \end{bmatrix} \Longleftrightarrow \begin{bmatrix} 1 & -2 & 0 & 1-5t \\ 0 & 1 & 0 & -\frac{1}{3}+2t \\ 0 & 0 & 1 & t \end{bmatrix} \Longleftrightarrow \begin{bmatrix} 1 & 0 & 0 & \frac{1}{3}-t \\ 0 & 1 & 0 & -\frac{1}{3}+2t \\ 0 & 0 & 1 & t \end{bmatrix}$$

   Quindi $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{1}{3}-t \\ -\frac{1}{3}+2t \\ t \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ -\frac{1}{3} \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} = \mathbf{v}_0 + t_1\mathbf{v}_1.$

Se avessi un sistema con due parametri, ad esempio $\begin{bmatrix} 0 & 1 & -1 & 3 \end{bmatrix}$, dovrei agire similmente aggiungendo due equazioni: $\begin{bmatrix} 0 & 1 & -1 & 3 \\ 1 & 0 & 0 & t_1 \\ 0 & 0 & 1 & t_2 \end{bmatrix}$ con x, z variabili indipendenti. Poi risolverei normalmente scambiando (1) e (2), ecc.:

$$\begin{bmatrix} 1 & 0 & 0 & t_1 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 1 & t_2 \end{bmatrix} \iff \begin{bmatrix} 1 & 0 & 0 & t_1 \\ 0 & 1 & 0 & 3+t_2 \\ 0 & 0 & 1 & t_2 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix} + t_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

2. Se $A \in M_{\mathbf{K}}(m, n)$, $\mathbf{b} \in \mathbf{K}^m$, $r(A) = r$:

   Il sistema lineare $A\mathbf{x} = \mathbf{b}$ ammette almeno una soluzione per ogni $\mathbf{b}$ se e solo se $r = m$ (perché questo accada è necessario $m \leq n$, essendo $r(A) \leq \min\{m, n\}$), cioè $col(A) = \mathbf{K}^m$ e $\dim(Im(_A)) = m$ ($_A : \mathbf{K}^n \to \mathbf{K}^m$ è suriettiva, cioè $m \leq n$, cioè A è invertibile a destra).

   Il sistema lineare $A\mathbf{x} = \mathbf{b}$ ammette al più una soluzione per ogni $\mathbf{b}$ se e solo se la soluzione non esiste o dipende da zero parametri, quindi $r = n$, cioè le colonne di A sono linearmente indipendenti e $\ker(A) = \{\mathbf{0}\}$ ($_A$ è iniettiva, cioè $n \leq m$, cioè A è invertibile a sinistra).

   Il sistema lineare $A\mathbf{x} = \mathbf{b}$ ammette esattamente una soluzione per ogni $\mathbf{b}$ se e solo se $r = m$ e $r = n$, cioè A è quadrata e ha rango massimo ($_A$ è biunivoca; se $\dim V \neq \dim W$ l'applicazione lineare non può essere invertibile).

   Dunque, se $m = n$ e $\exists\ \mathbf{b}_1 \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b}_1$ ha più di una soluzione, significa che $r < n$, dunque esiste $\mathbf{b}_2 : A\mathbf{x} = \mathbf{b}_2$ non ha alcuna soluzione; se $m = n$ e $\exists\ \mathbf{b}_1 \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b}_1$ non ha soluzione, significa che $r < n$, dunque esiste $\mathbf{b}_2 : A\mathbf{x} = \mathbf{b}_2$ ha infinite soluzioni (basta prendere $\mathbf{b}_2 = \mathbf{0}$). Se $\exists\ \mathbf{b}_1 \in \mathbb{R}^m : A\mathbf{x} = \mathbf{b}_1$ ha esattamente una soluzione, significa che $r = n$, ma può esistere $\mathbf{b}_2 : A\mathbf{x} = \mathbf{b}_2$ non ha alcuna soluzione (basta prendere una matrice non quadrata e un termine noto con l'ultima componente non nulla).

Corollario (teorema di Cramer)

Hp: $A\mathbf{x} = \mathbf{b}$ è un sistema lineare con n equazioni e n incognite; $r(A) = n$

Ts: $\forall\ \mathbf{b} \in \mathbf{K}^n$ esiste un'unica soluzione di $A\mathbf{x} = \mathbf{b}$

Dim Riducendo a scala $A|\mathbf{b}$, si ottiene $U|\mathbf{b}'$, con U che è necessariamente una "scala perfetta"

$$\begin{bmatrix} p_1 & * & ... & * & b_1' \\ 0 & p_2 & ... & * & b_2' \\ ... & ... & ... & ... & ... \\ 0 & 0 & ... & p_n & b_n' \end{bmatrix}$$

dato che ci sono n pivot, uno su ciascuna riga, uno su ciascuna colonna, perché $r(A) = n$. Non è possibile $r(A) \neq r(A|\mathbf{b})$ perché non ci possono essere pivot sulla stessa riga o sulla stessa colonna. Dunque, se $A \in M_{\mathbf{K}}(n,n)$ e $r(A) = n$, ne segue $r(A) = r(A|\mathbf{b})$: per il teorema di Rouché-Capelli esiste almeno una soluzione. Ma poiché $n - r = 0$, non c'è alcun parametro libero e dunque esiste un'unica soluzione. ■

Il sistema omogeneo associato ha dunque come unica soluzione $\mathbf{x} = \mathbf{0}$.

2.5 Fattorizzazione LU

E' un modo per "memorizzare" il processo di una riduzione a scala di una matrice A, così da non doverlo ripetere per ogni $\mathbf{b}$.

Def Una matrice $A \in M_{\mathbf{K}}(n,n)$ si dice triangolare superiore se $a_{ij} = 0\ \forall\ i > j$. Una matrice $A \in M_{\mathbf{K}}(n,n)$ si dice triangolare inferiore se $a_{ij} = 0\ \forall\ i < j$.

$$\text{MT superiore: } \begin{bmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & * \end{bmatrix};\ \text{MT inferiore: } \begin{bmatrix} * & 0 & 0 \\ * & * & 0 \\ * & * & * \end{bmatrix}$$

Una matrice quadrata a scala è una matrice triangolare superiore. Non vale l'implicazione inversa: una matrice può essere triangolare superiore e avere due pivot sulla stessa colonna.

1. $\begin{bmatrix} 0 & 1 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$ è triangolare superiore ma non a scala.

Teo

Hp: $A \in M_{\mathbf{K}}(m,n)$ si può ridurre a scala senza l'utilizzo dello scambio di due righe

Ts: $\exists\ L \in M_{\mathbf{K}}(m,m)$ triangolare inferiore e $U \in M_{\mathbf{K}}(m,n)$ a scala tali che $A = LU$

Se $m = n$, U è triangolare superiore: in tal caso L sta per "lower" e U per "upper".

L'idea è codificare in L le operazioni del metodo di eliminazione di Gauss.

1. **Scambio di due righe**

   Sia $A \in M_{\mathbf{K}}(m,n)$ una matrice di cui si vogliono scambiare le righe R_i e R_j; sia B la matrice A dopo lo scambio. Si cerca una matrice $S : SA = B$.

   $S \in M_{\mathbf{K}}(m,m)$. Si consideri l'esempio di $A = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$: per ottenere $B = \begin{bmatrix} d & e & f \\ a & b & c \end{bmatrix}$ calcolando SA dev'essere $S = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Se $A = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}$, per ottenere $B = \begin{bmatrix} e & f \\ c & d \\ a & b \end{bmatrix}$ dev'essere $S = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$.

   In generale, se S è la matrice incognita, $s_{lk} = \begin{cases} 1 \text{ se } (l = i, k = j) \text{ o } (l = j, k = i)\text{, se } l = k \neq i, j \\ 0 \text{ altrimenti} \end{cases}$: S è la matrice identità con le righe i e j scambiate.

   Lo scambio è reversibile: se $SA = B$, $SB = A$. Inoltre, dato che $S(SA) = A$, $S^2 A = A$, quindi $S^2 = Id_m$ e $S = S^{-1}$.

2. **Moltiplicazione di una riga per un coefficiente non nullo**

   Sia $A \in M_{\mathbf{K}}(m,n)$ una matrice di cui si vuole sostituire la riga R_i con λR_i, con $\lambda \in \mathbf{K} - \{0\}$; sia B la matrice A dopo la moltiplicazione. Si cerca una matrice $S : SA = B$.

   $S \in M_{\mathbf{K}}(m,m)$. Si consideri l'esempio di $A = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$: per ottenere $B = \begin{bmatrix} a & b & c \\ \lambda d & \lambda e & \lambda f \end{bmatrix}$ dev'essere $S = \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix}$.

   In generale, se S è la matrice incognita, $s_{lk} = \begin{cases} \lambda \text{ se } l = k = i \\ 1 \text{ se } l = k \neq i \\ 0 \text{ altrimenti} \end{cases}$, cioè S è la matrice identità con λ invece che 1 nella posizione (i,i).

L'operazione è reversibile: se $SA = B$ e si è fatto λR_i, $S'B = A$, con $s'_{lk} = \begin{cases} \frac{1}{\lambda} \text{ se } l = k = i \\ 1 \text{ se } l = k \neq i \\ 0 \text{ altrimenti} \end{cases}$. S' è la matrice inversa di S.

3. **Sostituzione di una riga con la sua somma con un multiplo di un'altra**

   Sia $A \in M_{\mathbf{K}}(m,n)$ una matrice di cui si vuole sostituire la riga R_i con $R_i + \lambda R_j$, con $\lambda \in \mathbf{K} - \{0\}$; sia B la matrice A dopo l'operazione. Si cerca una matrice $E : EA = B$.

   $E \in M_{\mathbf{K}}(m,m)$. Si consideri l'esempio di $A = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}$, per ottenere $A = \begin{bmatrix} a & b \\ c+\lambda a & d+\lambda b \\ e & f \end{bmatrix}$ dev'essere $E = \begin{bmatrix} 1 & 0 & 0 \\ \lambda & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

   In generale, se E è la matrice incognita, $e_{lk} = \begin{cases} \lambda \text{ se } l = i \text{ e } k = j \\ 1 \text{ se } l = k \\ 0 \text{ altrimenti} \end{cases}$.

   L'operazione è reversibile: se $EA = B$ e si è fatto $R_i = R_i + \lambda R_j$, $E^{-1}B = A$, con $e^{-1}_{lk} = \begin{cases} -\lambda \text{ se } l = i \text{ e } k = j \\ 1 \text{ se } l = k \\ 0 \text{ altrimenti} \end{cases}$.

   E^{-1} è quindi la matrice inversa di E: $E^{-1}B = A \iff EE^{-1}B = EA \iff EE^{-1}B = B \iff EE^{-1} = Id_m$.

   Ogni matrice E^{-1} (o E) siffatta è triangolare: inferiore se $i > j$, superiore se $i < j$.

Dim Nell'ipotesi del teorema, per la riduzione si possono usare solo operazioni di tipo 3: cioè, eseguendo t operazioni, si fa $E_t(...(E_3(E_2(E_1 A)))) = U$, con U a scala. In ciascuna operazione, si somma a una riga R_i della matrice a destra λR_j, con $j < i$ (questo perché nel MEG si procede dall'alto verso il basso, azzerando gli elementi sotto ciascun pivot: quindi si avrà $l > k$). Per la proprietà associativa del prodotto matriciale vale $E_t...E_3E_2E_1A = U$. Voglio spostare a destra le E: poiché E_t ha per inversa E_t^{-1}, premoltiplico lato sinistro e destro per E_t^{-1}:

$$E_t^{-1}(E_t...E_3E_2E_1A) = E_t^{-1}U \iff \left(E_t^{-1}E_t\right)...E_3E_2E_1A = E_t^{-1}U \iff Id_nE_{t-1}...E_3E_2E_1A = E_t^{-1}U \iff E_{t-1}...E_3E_2E_1A = E_t^{-1}$$

Procedo così anche per $E_{t-1}, ..., E_1$ e ottengo

$$A = E_1^{-1} E_2^{-1} E_3^{-1} ... E_{t-1}^{-1} E_t^{-1} U$$

Poiché ciascuna E_i^{-1} è triangolare inferiore e il prodotto di matrici triangolari inferiori è una matrice triangolare inferiore*, chiamo L la matrice triangolare inferiore $E_1^{-1} E_2^{-1} E_3^{-1} ... E_{t-1}^{-1} E_t^{-1}$ e vale $A = LU$. ■

***Proposizione**

$$\begin{array}{rcl} \text{Hp} & : & L_1, L_2 \text{ sono matrici triangolari inferiori} \\ \text{Ts} & : & L_1 L_2 \text{ è triangolare inferiore, cioè } i < j \Longrightarrow (L_1 L_2)_{ij} = 0 \end{array}$$

Dim Cerco un'espressione per scrivere il generico elemento $(L_1 L_2)_{ij}$, con $i < j$. $(L_1 L_2)_{ij} = \sum_{k=1}^{m} (L_1)_{ik} (L_2)_{kj} = \sum_{k=1}^{i} (L_1)_{ik} (L_2)_{kj} + \sum_{k=i+1}^{m} (L_1)_{ik} (L_2)_{kj} = 0 + 0$: il secondo fattore del primo addendo si annulla perché $k \leq i < j$ e L_2 è triangolare inferiore, il primo fattore del secondo addendo si annulla perché $i < k$ e L_1 è triangolare inferiore. ■

Corollario

$$E_1^{-1} E_2^{-1} E_3^{-1} ... E_{t-1}^{-1} E_t^{-1} \text{ è una matrice triangolare inferiore}$$

Si nota che la generica matrice E_i, che incorpora l'operazione di sostituzione di una riga con la sua somma con un multiplo di un'altra, ha tutti gli elementi sulla diagonale uguali a 1. Questa proprietà si estende al prodotto di due matrici siffatte.

$$\begin{array}{rl} \text{Hp:} & L_1, L_2 \text{ sono matrici triangolari inferiori tali che } (L_1)_{ii} = (L_2)_{ii} = 1 \\ \text{Ts:} & (L_1 L_2)_{ii} = 1 \end{array}$$

Dim Cerco un'espressione per scrivere il generico elemento $(L_1 L_2)_{ii}$. $(L_1 L_2)_{ii} = \sum_{k=1}^{m} (L_1)_{ik} (L_2)_{ki} = \sum_{k=1}^{i-1} (L_1)_{ik} (L_2)_{ki} + (L_1)_{ii} (L_2)_{ii} + \sum_{k=i+1}^{m} (L_1)_{ik} (L_2)_{ki} = 0 + 1 \cdot 1 + 0$: il secondo fattore del primo addendo si annulla perché $k < i$ e L_2 è triangolare inferiore, il primo fattore del terzo addendo si annulla perché $i < k$ e L_1 è triangolare inferiore, il secondo addendo è 1. ■

Si può quindi riesprimere il teorema sulla fattorizzazione LU aggiungendo un dettaglio:

Teo

$$\text{Hp: } A \in M_{\mathbf{K}}(m, n) \text{ si può ridurre a scala senza l'utilizzo dello scambio di due righe}$$

$$\text{Ts: } \exists\ L \in M_{\mathbf{K}}(m, m), \text{ triangolare inferiore con 1 sulla diagonale}$$

$$\text{principale, e } U \in M_{\mathbf{K}}(m, n), \text{ a scala, tali che } A = LU$$

1. Voglio scrivere $A=\begin{bmatrix}1&-4&1\\2&-6&5\\1&-2&5\end{bmatrix}$ come LU, dove U è A ridotta a scala e L è il prodotto delle matrici triangolari necessarie per la riduzione. Voglio mettere a zero a_{21} e a_{31}: devo fare quindi $(2)=(2)-2\,(1)$ e $(3)=(3)-(1)$, dunque la prima matrice necessaria è $E_1=\begin{bmatrix}1&0&0\\-2&1&0\\0&0&1\end{bmatrix}$, la seconda è $E_2=\begin{bmatrix}1&0&0\\0&1&0\\-1&0&1\end{bmatrix}$, il loro prodotto è $E_1E_2=\begin{bmatrix}1&0&0\\-2&1&0\\-1&0&1\end{bmatrix}$. Il risultato è $E_2E_1A=\begin{bmatrix}1&-4&1\\0&2&3\\0&2&4\end{bmatrix}$. Ora devo fare $(3)=(3)-(2)$: la matrice necessaria è $E_3=\begin{bmatrix}1&0&0\\0&1&0\\0&-1&1\end{bmatrix}$ e ottengo $E_3E_2E_1A=\begin{bmatrix}1&-4&1\\0&2&3\\0&0&1\end{bmatrix}$, che è la matrice U ridotta a scala. Ora calcolo il prodotto $E_1^{-1}E_2^{-1}E_3^{-1}U$, sapendo quali sono le matrici inverse:

$$\begin{bmatrix}1&0&0\\-2&1&0\\0&0&1\end{bmatrix}^{-1}\begin{bmatrix}1&0&0\\0&1&0\\-1&0&1\end{bmatrix}^{-1}\begin{bmatrix}1&0&0\\0&1&0\\0&-1&1\end{bmatrix}^{-1}\begin{bmatrix}1&-4&1\\0&2&3\\0&0&1\end{bmatrix}=\begin{bmatrix}1&0&0\\2&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&0\\1&0&1\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&0\\0&1&1\end{bmatrix}\begin{bmatrix}1&-4\\0&2\\0&0\end{bmatrix}$$

e $L=\begin{bmatrix}1&0&0\\2&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&0\\1&0&1\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&0\\0&1&1\end{bmatrix}=\begin{bmatrix}1&0&0\\2&1&0\\1&0&1\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&0\\0&1&1\end{bmatrix}=\begin{bmatrix}1&0&0\\2&1&0\\1&1&1\end{bmatrix}$.

2.5.1 Applicazione ai sistemi lineari

Sia $A\in M_{\mathbf{K}}(n,n)$: in tal caso A ridotta a scala è U triangolare superiore (una matrice ridotta a scala non è triangolare superiore in generale perché la definizione di triangolare si applica a matrici quadrate). Voglio risolvere $A\mathbf{x}=\mathbf{b}$, equivalente a $(LU)\,\mathbf{x}=\mathbf{b}\Longleftrightarrow L\,(U\mathbf{x})=\mathbf{b}$, che è equivalente a risolvere prima $L\mathbf{y}=\mathbf{b}$, con $\mathbf{y}$ nuovo vettore incognita, e poi $U\mathbf{x}=\mathbf{y}$. Risolvere questi due sistemi è più vantaggioso perché L e U sono matrici triangolari, quindi è più facile risolvere i sistemi associati.

Lo stesso vale se A non è quadrata e U è a scala.

E' poi opportuno chiedersi come applicare la fattorizzazione LU nel caso in cui siano necessari scambi di riga per ridurre A a scala: questa situazione non entra nelle ipotesi del teorema. Si può però fare una prima riduzione a scala identificando gli scambi necessari, che si possono effettuare con le matrici $S_1,...,S_t$, del tipo 1. Si calcola $S_t....S_2S_1A$:

questa matrice può essere ridotta a scala senza scambi, e ad essa si applica il teorema: $S_t....S_2S_1A = LU$.

2.6 Calcolo della matrice inversa

Considero il caso di matrici quadrate: devo risolvere $A\mathbf{x} = \mathbf{b}$, con $A \in M_{\mathbf{K}}(n,n)$. Se A fosse invertibile a sinistra, potrei premoltiplicare a sinistra e a destra per A^{-1}: $A^{-1}A\mathbf{x} = A^{-1}\mathbf{b} \iff Id_n\mathbf{x} = A^{-1}\mathbf{b} \iff \mathbf{x} = A^{-1}\mathbf{b}$. Quindi conoscere l'inversa permette di risolvere molto in fretta il sistema, facendo il prodotto tra $A^{-1} \in M_{\mathbf{K}}(n,n)$ e $\mathbf{b} \in \mathbb{R}^n$. Questo metodo è vantaggioso solo in alcuni casi, come si vedrà ora.

Sia $A \in M_{\mathbf{K}}(2,2)$, $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, per quali $a,b,c,d \in \mathbf{K}$ esiste A^{-1}? Cerco l'inversa destra $X : AX = Id_2$, $X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$. Dev'essere $AX = \begin{bmatrix} ax_{11}+bx_{21} & ax_{12}+bx_{22} \\ cx_{11}+dx_{21} & cx_{12}+dx_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$: apparentemente cercare l'inversa destra si è tradotto in un sistema lineare a quattro incognite e quattro equazioni. In realtà però è un insieme di due sistemi, ciascuno dei quali a due incognite e due equazioni: $\begin{bmatrix} ax_{11}+bx_{21} \\ cx_{11}+dx_{21} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ e $\begin{bmatrix} ax_{12}+bx_{22} \\ cx_{12}+dx_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. La matrice dei coefficienti è la stessa.

Se ne ricava che calcolare l'inversa destra di una matrice $n \times n$ equivale a risolvere n sistemi lineari quadrati, cioè con n equazioni in n incognite.

$$A\mathbf{x} = Id_n \iff A\begin{bmatrix} x_{11} \\ x_{21} \\ ... \\ x_{n1} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ ... \\ 0 \end{bmatrix}, A\begin{bmatrix} x_{12} \\ x_{22} \\ ... \\ x_{n2} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ ... \\ 0 \end{bmatrix}, ..., A\begin{bmatrix} x_{1n} \\ x_{2n} \\ ... \\ x_{nn} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ ... \\ 1 \end{bmatrix}$$

Tutti condividono la matrice dei coefficienti e tutti, per il teorema di Cramer, sono risolvibili se (e solo se?) $r(A) = n$.

Poiché invertire A richiede di risolvere n sistemi lineari, è conveniente risolvere dei sistemi lineari con il metodo dell'inversa solo se occorre risolvere più di n sistemi.

Tornando al caso 2×2, voglio risolvere simultaneamente due sistemi: $\begin{bmatrix} a & b & 1 \\ c & d & 0 \end{bmatrix}$ e $\begin{bmatrix} a & b & 0 \\ c & d & 1 \end{bmatrix}$; dal momento che il processo di riduzione è determinato unicamente dalla matrice dei coefficienti, i due sistemi possono essere risolti in parallelo usando un'unica matrice $\begin{bmatrix} a & b & 1 & 0 \\ c & d & 0 & 1 \end{bmatrix} = [A|Id_2]$. Applico il MEG. Suppongo che la prima colonna sia non nulla e $a \neq 0$; faccio $(1) = \frac{1}{a}(1)$ e $(2) = (2) - c(1)$: ottengo $\begin{bmatrix} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & d - c\frac{b}{a} & -\frac{c}{a} & 1 \end{bmatrix}$.

Se $d - c\frac{b}{a} = 0$, la seconda equazione del secondo sistema diventa $0 = 1$, quindi il secondo sistema è impossibile e la matrice non è invertibile.

Se $d - c\frac{b}{a} \neq 0$ $(ad - bc \neq 0)$, faccio $(2) = \frac{1}{d - \frac{bc}{a}}(2)$ e ottengo $\begin{bmatrix} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & 1 & -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix}$: uso il MEGJ e faccio $(1) = (1) - \frac{b}{a}(2)$, quindi ho $\begin{bmatrix} 1 & 0 & \frac{d}{ad-bc} & -\frac{b}{ad-bc} \\ 0 & 1 & -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix} = \left[Id_2 | A^{-1}\right]$. La formula generica per l'inversa destra di una matrice 2×2 è quindi $A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. La matrice $A^* = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ è detta matrice aggiunta o matrice dei complementi algebrici. Si nota che $A^* = (trA)\, Id - A$.

Si nota che per l'invertibilità è opportuno supporre che la prima colonna sia non nulla, perché affinché la formula valga, dovendo essere $ad - bc \neq 0$, se $a = 0$ allora $b, c \neq 0$. Nel caso di $a = 0$ e $c \neq 0$ si ottiene infatti la stessa formula. Inoltre nel caso 2×2 l'invertibilità dipende da un unico numero calcolato con gli elementi della matrice: $ad - bc$.

2.7 Determinante

Def Il determinante è una funzione scalare $\det : M_{\mathbf{K}}(n,n) \to \mathbf{K}$, che associa a ogni matrice quadrata A un numero, definita ricorsivamente nel modo seguente:

- se $n = 1$ $\det A = \det([a_{11}]) = a_{11}$
- se $n > 1$ $\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det \hat{A}_{ij}$

dove i è un indice di riga fissato, a_{ij} è l'elemento di A alla i-esima riga e j-esima colonna, $\hat{A}_{ij}$ è una matrice $(n-1)\times(n-1)$ ottenuta a partire da A, cancellando l'i-esima riga e la j-esima colonna. Il cappello sopra una lettera indica un oggetto ricavato da un altro mediante cancellazioni. Equivalentemente $\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det \hat{A}_{ij}$, dove j è un indice di colonna fissato: il determinante non dipende dalla scelta della riga o della colonna.

1. Considero $A \in M_{\mathbf{K}}(2,2)$, $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, e calcolo $\det A$ usando la riga 2. $\det A_{i=2} = \sum_{j=1}^{2} (-1)^{2+j} a_{2j} \det \hat{A}_{2j} = (-1)^{2+1} a_{21} \det \hat{A}_{21} + (-1)^{2+2} a_{22} \det \hat{A}_{22} = -a_{21}a_{12} + a_{22}a_{11} = a_{11}a_{22} - a_{12}a_{21}$. Se chiamo $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $\det A = ad - bc$. Ora calcolo $\det A$ usando la colonna 1. $\det A_{j=1} = \sum_{i=1}^{2} (-1)^{i+1} a_{i1} \det \hat{A}_{i1} = (-1)^{1+1} a_{11} \det \hat{A}_{11} + (-1)^{2+1} a_{21} \det \hat{A}_{21} = a_{11}a_{22} - a_{21}a_{12}$.

2. Considero $A \in M_{\mathbf{K}}(3,3)$, $A = \begin{bmatrix} 9 & 2 & -2 \\ 2 & 2 & 0 \\ -2 & 0 & 2 \end{bmatrix}$, e calcolo $\det A$ usando la colonna 1. $\det A_{j=1} = \sum_{i=1}^{3} (-1)^{i+1} a_{i1} \det \hat{A}_{i1} =$ $(-1)^{1+1} 9 \det \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} + (-1)^{2+1} 2 \det \begin{bmatrix} 2 & -2 \\ 0 & 2 \end{bmatrix} + (-1)^{3+1} (-2) \det \begin{bmatrix} 2 & -2 \\ 2 & 0 \end{bmatrix}$, cioè, usando la formula appena trovata, $9 \cdot 4 - 2 \cdot 4 - 2(4) = 20$. Calcolo ora $\det A$ usando la riga 2: $\det A_{i=2} = \sum_{j=1}^{3} (-1)^{2+j} a_{2j} \det \hat{A}_{2j} =$ $(-1)^{2+1} 2 \det \begin{bmatrix} 2 & -2 \\ 0 & 2 \end{bmatrix} + (-1)^{2+2} 2 \det \begin{bmatrix} 9 & -2 \\ -2 & 2 \end{bmatrix} + (-1)^{2+3} 0 \det \begin{bmatrix} 9 & 2 \\ -2 & 0 \end{bmatrix} = -2 \cdot 4 + 2(18-4) = 20$.

Ha senso chiedersi come scegliere la linea con cui calcolare il determinante. Usando la riga 2 nell'esempio precedente, grazie alla presenza di uno zero, si è potuto evitare di calcolare il determinante del terzo addendo: in generale, è sensato usare la linea con più zeri.

2.7.1 Proprietà e MEG

1. Se A ha una linea (riga o colonna) nulla, $\det A = 0$. Si dimostra calcolando il determinante usando quella linea.

2. Se U è triangolare, $\det U = u_{11}u_{22}...u_{nn}$. Se U è triangolare superiore, si dimostra calcolando il determinante usando la prima colonna: $\det U_{j=1} = (-1)^{1+1} u_{11} \det \hat{U}_{11} = u_{11}\left((-1)^{2+2} u_{22} \det \hat{U}_{22}\right) = ... = u_{11}u_{22}...u_{nn}$. Se U è triangolare inferiore, si dimostra usando la prima riga.

   (a) Essendo Id_n triangolare, $\det Id_n = 1$: questo fatto si esprime dicendo che il determinante è normalizzato, o si chiama proprietà di normalizzazione del determinante; il determinante è una funzione che associa l'elemento neutro del prodotto in $M_{\mathbf{K}}(n,n)$ all'elemento neutro del prodotto in $\mathbf{K}$.

   (b) Se U è una matrice a scala (quadrata, altrimenti non ha senso parlare di determinante) con $r(U) = n$, $U = \begin{bmatrix} p_1 & & & \\ & p_2 & & \\ & & ... & \\ & & & p_n \end{bmatrix}$ e, essendo U triangolare superiore, $\det U = p_1 p_2 ... p_n \neq 0$ per definizione di pivot. Se $r(U) < n$, c'è almeno una colonna priva di pivot e l'elemento di quella colonna che giace sulla diagonale principale è nullo. Quindi $\det U = 0$.

3. Come si comporta il determinante rispetto al prodotto matriciale?

Teorema (Binet)

$$\begin{aligned} \text{Hp:} \quad & A, B \in M_{\mathbf{K}}(n,n) \\ \text{Ts:} \quad & \det(AB) = \det A \det B \end{aligned}$$

Corollario

$$\begin{aligned} \text{Hp} \quad : \quad & A, B \in M_{\mathbf{K}}(n,n) \\ \text{Ts} \quad : \quad & \det(AB) = \det(BA) \end{aligned}$$

Infatti $\det(AB) = \det A \det B = \det B \det A = \det(BA)$, perché il prodotto sul campo $\mathbf{K}$ è commutativo.

4. Come si comporta il determinante rispetto al prodotto per scalari?

   Data $A \in M_{\mathbf{K}}(n,n)$, $\lambda \in \mathbf{K}$, $\det(\lambda A) = \lambda^n \det A$. Infatti $\det(\lambda A) = \det(\lambda Id_n A) = \det(\lambda Id_n) \det A = \lambda^n \det A$, perché λId_n è una matrice con tutti λ sulla diagonale e zero altrove. Ne segue che il determinante non è lineare rispetto ad A.

5. Come si comporta il determinante rispetto alla somma?

   Data $A \in M_{\mathbf{K}}(n,n)$, scrivo la prima riga di A come somma dei vettori $\mathbf{v}$ e $\mathbf{w}$.

$$A = \begin{bmatrix} \mathbf{v}+\mathbf{w} \\ R_2(A) \\ ... \\ R_n(A) \end{bmatrix}$$

Definisco $A_{\mathbf{v}} = \begin{bmatrix} \mathbf{v} \\ R_2(A) \\ ... \\ R_n(A) \end{bmatrix}$, $A_{\mathbf{w}} = \begin{bmatrix} \mathbf{w} \\ R_2(A) \\ ... \\ R_n(A) \end{bmatrix}$. Allora $\det A = \det A_{\mathbf{v}} + \det A_{\mathbf{w}}$. Lo stesso vale per una qualsiasi riga o colonna di A che possa essere scritta come somma di due vettori: si dice che il determinante è multilineare, cioè, fissate arbitrariamente tutte le righe (colonne) tranne una, è una funzione lineare della riga (colonna) rimanente.

Ne segue che, sostituendo una linea con la sua somma con un'altra linea, il determinante non cambia.

- $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$. E' possibile scrivere $\begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} = \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} + \begin{bmatrix} c_{11} \\ c_{21} \\ c_{31} \end{bmatrix}$: $A = \begin{bmatrix} b_{11}+c_{11} & a_{12} & a_{13} \\ b_{21}+c_{21} & a_{22} & a_{23} \\ b_{31}+c_{31} & a_{32} & a_{33} \end{bmatrix}$.

  Calcolo $\det A$ usando la prima colonna: $\det A_{j=1} = (-1)^{1+1}(b_{11}+c_{11})\det \hat{A}_{11} + (-1)^{2+1}(b_{21}+c_{21})\det \hat{A}_{21} + (-1)^{3+1}(b_{31}+c_{31})\det \hat{A}_{31}$. Usando la proprietà distributiva, si ha $\det A = (-1)^{1+1} b_{11}\det \hat{A}_{11} + (-1)^{2+1} b_{21}\det \hat{A}_{21} + (-1)^{3+1} b_{31}\det \hat{A}_{31} + (-1)^{1+1} c_{11}\det \hat{A}_{11} + (-1)^{2+1} c_{21}\det \hat{A}_{21} + (-1)^{3+1} c_{31}\det \hat{A}_{31}$, che è $\det A_{\mathbf{b}} + \det A_{\mathbf{c}}$, ciascuno calcolato usando la prima colonna.

6. Come si comporta il determinante rispetto al MEG?

   (a) **Scambio di due righe**

   Data la matrice A, ottengo la matrice B da A scambiandone la riga R_i con la riga R_j. Si può scrivere $B = S_{ij}A$, dove S_{ij} è una matrice del tipo 1. Per il teorema di Binet, $\det B = \det(S_{ij}A) = \det S_{ij} \det A$. Calcolo $\det S_{ij}$ usando una colonna che contiene 1 sulla diagonale: $\det S_{ij} = (-1)^{1+1} 1 \cdot \det \hat{S}_{ij}$. Gli 1 fuori dalla diagonale sono gli unici elementi non nulli sulla loro linea. Continuo a ridurre il problema scegliendo righe con 1 sulla diagonale ($i + j \equiv 0 mod 2$). Poiché continuo a moltiplicare per 1, infine avrò $\det S_{ij} = (-1)^{1+1} 1 \cdot \det \hat{S}_{ij} = ... = \det \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, cioè il determinante è lo stesso della matrice che si usa per scambiare le due righe di una matrice 2×2. Poiché $\det \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = -1 = \det S_{ij}$, $\det B = -\det A$ (proprietà di alternanza del determinante). Lo stesso vale se si scambiano due colonne invece di due righe.

   Di conseguenza, se una matrice ha due linee uguali, scambiandole si ottiene $B = A$: $\det A = -\det A \iff \det A = 0_{\mathbf{K}}$. Infatti, se una matrice ha due righe uguali, sottraendole si ottiene una matrice con una riga nulla, che se usata per calcolare il determinante produce zero.

   (b) **Moltiplicazione di una riga per $\lambda \neq 0$**

   Data la matrice A, ottengo la matrice B da A moltiplicando la riga R_i per λ. Si può scrivere $B = M_\lambda A$, dove M_λ è una matrice del tipo 2. Per il teorema di Binet, $\det B = \det(M_\lambda A) = \det M_\lambda \det A = \lambda \det A$, perché M_λ è una matrice triangolare il cui determinante è il prodotto degli elementi sulla diagonale.

   (c) **Sostituzione di un riga con la sua somma con il multiplo di un'altra**

   Data la matrice A, ottengo la matrice B da A sostituendo alla riga R_i $R_i + \lambda R_j$. Si può scrivere $B = E_{i,\lambda,j}A$, dove $E_{i,\lambda,j}$ è una matrice del tipo 3. Per il teorema di Binet, $\det B = \det(E_{i,\lambda,j}A) =$

$\det E_{i,\lambda,j} \det A = \det A$, perché $E_{i,\lambda,j}$ è una matrice triangolare inferiore $(i > j)$ il cui determinante è il prodotto degli elementi sulla diagonale.

Def Data $A \in M_{\mathbf{K}}(n,n)$, si dice matrice dei complementi algebrici o matrice dei cofattori di A, e si indica con $\mathrm{cof}A$, la matrice in $M_{\mathbf{K}}(n,n)$ avente come elemento ij $(-1)^{i+j} \det \hat{A}_{ij}$, dove $(-1)^{i+j} \det \hat{A}_{ij}$ è detto cofattore di A relativo alla posizione i,j.

Per una matrice $A \in M_{\mathbf{K}}(n,n)$ vale $A^{-1} = \frac{1}{\det A}(\mathrm{cof}A)^T$.

2.8 Invertibilità di una matrice

Hp: $A \in M_{\mathbf{K}}(n,n)$

Ts: le seguenti affermazioni sono equivalenti:

(i) A è invertibile

(ii) $r(A) = n$

(iii) $\det A \neq 0$

Questo significa (i)$\Longleftrightarrow$(ii)$\Longleftrightarrow$(iii). Per mostrare ciò è sufficiente dimostrare (i)$\Longrightarrow$(iii)$\Longrightarrow$(ii)$\Longrightarrow$(i).

Dim (1) Considero A invertibile e mostro che $\det A \neq 0$.

Per ipotesi (scelgo l'inversa destra, ma sarebbe analogo con l'inversa sinistra) $\exists\ A^{-1} : AA^{-1} = Id_n$. Per il teorema di Binet $\det\left(AA^{-1}\right) = \det Id_n \Longleftrightarrow \det A \det A^{-1} = 1$. Questo significa che $\det A$ è un elemento invertibile nel campo $\mathbf{K}$: $\det A = \frac{1}{\det A^{-1}}$. Poiché su un campo gli elementi invertibili sono tutti e soli gli elementi non nulli, $\det A \neq 0$.

(2) Considero $\det A \neq 0$ e mostro che $r(A) = n$.

Si è già verificato nel caso di A a scala. Nel caso di A generica, riduco A a scala con il MEG, usando tutti e tre i tipi di operazioni: se effettuo s scambi e k moltiplicazioni per scalare, ho $\det U = (-1)^s \lambda_1 \cdot \ldots \cdot \lambda_k \det A \Longleftrightarrow \det A = (-1)^s \frac{1}{\lambda_1} \cdot \ldots \cdot \frac{1}{\lambda_k} \det U$. Poiché $\det A \neq 0 \Longleftrightarrow \det U \neq 0$ e $r(A) = r(U)$ per definizione di rango, dall'ipotesi $\det A \neq 0$ segue $\det U \neq 0$, il che implica $r(U) = n$ e anche $r(A) = n$.

(3) Considero $r(A) = n$ e mostro che A è invertibile.

Mostro che esiste l'inversa destra e che, se esiste l'inversa destra, allora esiste anche l'inversa sinistra. Inizio cercando l'inversa destra, cioè $X : AX = Id_n$: questa equazione è equivalente agli n sistemi, ciascuno dei quali con

n equazioni in n incognite, $A\begin{bmatrix} x_{11} \\ x_{21} \\ ... \\ x_{n1} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ ... \\ 0 \end{bmatrix}$, $A\begin{bmatrix} x_{12} \\ x_{22} \\ ... \\ x_{n2} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ ... \\ 0 \end{bmatrix}$,..., $A\begin{bmatrix} x_{1n} \\ x_{2n} \\ ... \\ x_{nn} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ ... \\ 1 \end{bmatrix}$. Per il teorema di Cramer applicato a ciascuno dei sistemi (essendo $r(A) = n$), per ciascuno di essi esiste un unico vettore soluzione: aggregando i vettori soluzione si ottiene l'unica matrice inversa destra B. Adesso dovrei cercare $Y : YA = Id_n$: cerco invece $Y : BY = Id_n$ e mi chiedo qual è $r(B)$, in modo da poter applicare il teorema di Cramer. Osservo che $r(B) = n \iff B\mathbf{x} = \mathbf{0}$ ha un'unica soluzione (si ricava dai teoremi di Cramer e Rouché-Capelli); mostro allora che $B\mathbf{x} = \mathbf{0}$ ha un'unica soluzione. $B\mathbf{x} = \mathbf{0} \Longrightarrow A(B\mathbf{x}) = A\mathbf{0} \iff (AB)\mathbf{x} = \mathbf{0} \iff Id_n\mathbf{x} = \mathbf{0} \iff \mathbf{x} = \mathbf{0}$: dunque $r(B) = n$ e per il teorema di Cramer esiste un'unica $Y : BY = Id_n$; chiamo $Y = C$. Quindi B risulta invertibile: ha come inversa sinistra A e come inversa destra C: per unicità dell'inversa si ha $A = C$ e, essendo $AB = Id_n$ e $BC = BA = Id_n$, B è anche inversa sinistra di A e A ha come unica inversa B. ■

Se invece si considerano $A, B \in M_{\mathbf{K}}(n, n)$, le ipotesi affinché AB sia invertibile si ricavano dal teorema: AB è invertibile $\iff \det(AB) = \det A \det B \neq 0$, che è equivalente a $\det A, \det B \neq 0$ (perché su $\mathbf{K}$ vale la legge di annullamento del prodotto): $\det A, \det B \neq 0$ è equivalente a A e B invertibili. Quindi AB è invertibile se e solo se A e B sono invertibili.

Se AB è invertibile, qual è l'inversa? Cerco $X : (AB)X = X(AB) = Id_n$. Noto che $X = B^{-1}A^{-1}$ è tale che $(AB)\left(B^{-1}A^{-1}\right) = A\left(BB^{-1}\right)A^{-1} = AId_nA^{-1} = AA^{-1} = Id_n$, e analogamente se si calcola $X(AB)$. Dunque l'inversa del prodotto è il prodotto, in ordine scambiato, delle inverse.

1. Calcolo l'inversa di $B = \begin{bmatrix} 9 & 2 & -2 \\ 2 & 2 & 0 \\ -2 & 0 & 2 \end{bmatrix}$. $\det B = 20$, quindi è invertibile. Calcolo l'inversa risolvendo i tre sistemi in parallelo: $\begin{bmatrix} 9 & 2 & -2 & 1 & 0 & 0 \\ 2 & 2 & 0 & 0 & 1 & 0 \\ -2 & 0 & 2 & 0 & 0 & 1 \end{bmatrix}$. Scambio le prime due righe e normalizzo la nuova prima riga: $\begin{bmatrix} 1 & 1 & 0 & 0 & \frac{1}{2} & 0 \\ 9 & 2 & -2 & 1 & 0 & 0 \\ -2 & 0 & 2 & 0 & 0 & 1 \end{bmatrix}$. Faccio $(2) = (2) - 9(1)$ e $(3) = (3) + 2(1)$: $\begin{bmatrix} 1 & 1 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & -7 & -2 & 1 & -\frac{9}{2} & 0 \\ 0 & 2 & 2 & 0 & 1 & 1 \end{bmatrix}$;

normalizzo la terza riga e scambio la seconda e terza riga: $\begin{bmatrix} 1 & 1 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & -7 & -2 & 1 & -\frac{9}{2} & 0 \end{bmatrix}$; faccio $(3) = (3) + 7\,(2)$ e ottengo $\begin{bmatrix} 1 & 1 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 5 & 1 & -1 & \frac{7}{2} \end{bmatrix}$; $(3) = \frac{1}{5}\,(3)$: $\begin{bmatrix} 1 & 1 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{5} & -\frac{1}{5} & \frac{7}{10} \end{bmatrix}$; $(2) = (2) - (3)$: $\begin{bmatrix} 1 & 1 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & -\frac{1}{5} & \frac{7}{10} & -\frac{1}{5} \\ 0 & 0 & 1 & \frac{1}{5} & -\frac{1}{5} & \frac{7}{10} \end{bmatrix}$; $(1) = (1) - (2)$: $\begin{bmatrix} 1 & 0 & 0 & \frac{1}{5} & -\frac{1}{5} & \frac{1}{5} \\ 0 & 1 & 0 & -\frac{1}{5} & \frac{7}{10} & -\frac{1}{5} \\ 0 & 0 & 1 & \frac{1}{5} & -\frac{1}{5} & \frac{7}{10} \end{bmatrix}$. La matrice inversa è la matrice 3×3 a destra.

3 Geometria delle rette e dei piani nel piano e nello spazio euclideo

Def Un vettore applicato è un segmento con un verso assegnato.

Un vettore applicato $\vec{AB}$ è caratterizzato da un punto di applicazione (A), una direzione, un verso e un modulo.

Direzione e verso individuano una semiretta che parte dal punto di applicazione; il modulo è la distanza del secondo estremo dal punto di applicazione.

Def Un vettore libero è una grandezza caratterizzata da direzione, verso e modulo.

[Una relazione di equivalenza è una relazione tra due elementi di un insieme che ha le proprietà:

1. riflessiva: ogni elemento $\vec{AB}$ è equivalente a se stesso
2. simmetrica: se $\vec{AB}$ è equivalente a $\vec{CD}$, allora $\vec{CD}$ è equivalente ad $\vec{AB}$
3. transitiva: se $\vec{AB}$ è equivalente a $\vec{CD}$, e $\vec{CD}$ è equivalente a $\vec{EF}$, allora $\vec{AB}$ è equivalente a $\vec{EF}$]

Si chiama V_A l'insieme dei vettori applicati. Dati due elementi di V_A $\vec{AB}$ e $\vec{CD}$, si dice che $\vec{AB}$ è equivalente a $\vec{CD} \iff \vec{AB}$ e $\vec{CD}$ hanno stessa direzione, verso e modulo. Una direzione nello spazio è una classe di equivalenza di rette parallele: due rette hanno la stessa direzione se sono parallele (l'insieme di tutte le rette nello spazio può essere ripartito in classi di equivalenza di rette parallele).

Dato un vettore $\vec{AB}$, si definisce $\left[\vec{AB}\right]$ l'insieme dei vettori equivalenti ad $\vec{AB}$, $\left[\vec{AB}\right]$ si dice classe di equivalenza di $\vec{AB}$ e $\vec{AB}$ si dice rappresentante della classe.

Se $\vec{CD} \in \left[\vec{AB}\right]$, allora $\vec{AB} \in \left[\vec{CD}\right]$, cioè $\left[\vec{AB}\right] = \left[\vec{CD}\right]$. V_A si può dunque dividere in infinite classi di equivalenza, che non hanno intersezione tra loro (altrimenti si violerebbe la proprietà transitiva). Quindi si può meglio dare la definizione di vettore libero:

Def Un vettore libero è una classe di equivalenza (di vettori applicati, cioè di V_A).

Se si fissa un sistema di riferimento con origine O, si possono identificare:

- i punti dello spazio, del tipo (x, y, z)
- i vettori applicati $\vec{OP}$ del tipo $x\mathbf{u} + y\mathbf{v} + z\mathbf{w}$
- i vettori liberi $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$

3.1 Operazioni tra vettori liberi

3.1.1 Somma tra vettori e proprietà

Se $\mathbf{v} = \vec{OP}$ e $\mathbf{w} = \vec{OQ}$, il vettore somma è la diagonale del parallelogramma i cui due lati non paralleli sono i vettori da sommare: è il vettore $\vec{OR}$, dove R è il quarto vertice del parallelogramma.

$\forall$ terna di vettori liberi $\mathbf{v}, \mathbf{u}, \mathbf{w}$ valgono le proprietà:

- Associativa: $(\mathbf{v} + \mathbf{w}) + \mathbf{u} = \mathbf{v} + (\mathbf{w} + \mathbf{u})$.
- Commutativa: $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$.
- Esistenza dell'elemento neutro: $\vec{OO} = \vec{O}$ (vettore che parte nell'origine e arriva nell'origine).
- Esistenza dell'elemento opposto: $\forall\ \mathbf{v}\ \exists\ -\mathbf{v} : \mathbf{v} - \mathbf{v} = \vec{O}$. $-\mathbf{v}$ ha uguale direzione e modulo e verso opposto rispetto a $\mathbf{v}$.

3.1.2 Prodotto per scalari e proprietà

Dato $t \in \mathbb{R}$, $\mathbf{v}$ vettore libero, si ha $t\mathbf{v} = 0$ se $t = 0$; altrimenti $t\mathbf{v}$ ha modulo pari al modulo di $\mathbf{v}$ per $|t|$, stessa direzione di $\mathbf{v}$, stesso verso di $\mathbf{v}$ se $t > 0$, verso opposto se $t < 0$. Ha le seguenti proprietà:

- Distributiva a sinistra: $(t + s)\,\mathbf{v} = t\mathbf{v} + s\mathbf{v}$.
- Distributiva a destra: $t\,(\mathbf{v} + \mathbf{w}) = t\mathbf{v} + t\mathbf{w}$.

- Associativa: $(ts)\,\mathbf{v} = t\,(s\mathbf{v})$.

- Esistenza dell'elemento neutro: $1_{\mathbb{R}}\mathbf{v} = \mathbf{v}$.

Sono operazioni perfettamente analoghe a quelle viste per matrici.

3.1.3 Sistemi di riferimento

Piano H

Un sistema di riferimento di H è dato dalla scelta di un punto del piano, che fungerà da origine del sistema, e di due vettori liberi non paralleli (non con la stessa direzione: devono essere linearmente indipendenti per poter costituire una base) $\mathbf{u}$ e $\mathbf{v}$, che fungeranno da assi del sistema. Una tale scelta divide il piano in quattro parti. Per ogni punto P del piano, è possibile trovare un parallelogramma con lati paralleli a $\mathbf{u}$ e $\mathbf{v}$ e un vertice coincidente con P: allora $O\vec{P} = \alpha\mathbf{u} + \beta\mathbf{v}$, $\alpha, \beta \in \mathbb{R}$. A seconda della regione di piano in cui si trova P, α o β o entrambi dovranno essere negativi.

Dunque i punti del piano H sono descritti da tutte le possibili scelte di $\alpha, \beta \in \mathbb{R}$, attraverso l'identificazione $O\vec{P} = \alpha\mathbf{u} + \beta\mathbf{v}$.

$\{\mathbf{u}, \mathbf{v}\}$ si dice base del sistema di riferimento.

Spazio S

Un sistema di riferimento di S è dato dalla scelta di un punto dello spazio, che fungerà da origine del sistema, e di tre vettori liberi $\mathbf{u}, \mathbf{v}, \mathbf{w}$, tali che $\mathbf{u}, \mathbf{v}$ non sono paralleli e $\mathbf{w}$ non appartiene al piano individuato da $\mathbf{u}, \mathbf{v}$ (devono essere linearmente indipendenti); $\mathbf{u}, \mathbf{v}, \mathbf{w}$ fungeranno da assi del sistema. Una tale scelta divide lo spazio in otto parti. Per ogni punto P dello spazio, è possibile trovare un parallelepipedo con lati paralleli a $\mathbf{u}, \mathbf{v}, \mathbf{w}$ e un vertice coincidente con P: allora $O\vec{P} = \alpha\mathbf{u} + \beta\mathbf{v} + \gamma\mathbf{w}$, $\alpha, \beta, \gamma \in \mathbb{R}$. A seconda della regione di piano in cui si trova P, alcuni dei coefficienti possono essere negativi.

Dunque i punti dello spazio S sono descritti da tutte le possibili scelte di $\alpha, \beta, \gamma \in \mathbb{R}$, attraverso l'identificazione $O\vec{P} = \alpha\mathbf{u} + \beta\mathbf{v} + \gamma\mathbf{w}$. $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ si dice base del sistema di riferimento.

3.1.4 Prodotto scalare e proprietà

Per ogni coppia di vettori liberi $\mathbf{v}, \mathbf{w}$, si definisce prodotto scalare (o prodotto interno) di $\mathbf{v}$ e $\mathbf{w}$ il numero reale

$$\mathbf{v} \cdot \mathbf{w} := ||\mathbf{v}||\,||\mathbf{w}|| \cos\theta$$

dove θ è l'angolo compreso tra i due vettori e $||\mathbf{v}||$ è il modulo di $\mathbf{v}$. Ha le seguenti proprietà:

- Bilinearità: $(\mathbf{u}+\mathbf{v})\cdot\mathbf{w}=\mathbf{u}\cdot\mathbf{w}+\mathbf{v}\cdot\mathbf{w}$; $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w})=\mathbf{u}\cdot\mathbf{v}+\mathbf{u}\cdot\mathbf{w}$; $(t\mathbf{u})\cdot\mathbf{v}=t\,(\mathbf{u}\cdot\mathbf{v})=\mathbf{u}\cdot(t\mathbf{v})$

- Commutatività: $\mathbf{u}\cdot\mathbf{v}=\mathbf{v}\cdot\mathbf{u}$

- Positività: $\mathbf{v}\cdot\mathbf{v}\geq 0$; $\mathbf{v}\cdot\mathbf{v}=0 \Longleftrightarrow \mathbf{v}=\mathbf{0}$. Infatti $\mathbf{v}\cdot\mathbf{v}=||\mathbf{v}||\,||\mathbf{v}||\cos 0=||\mathbf{v}||^2\geq 0$.

Il prodotto scalare dà un criterio di perpendicolarità: due vettori sono perpendicolari se e solo se il loro prodotto scalare è nullo (il vettore nullo è considerato perpendicolare a tutti i vettori).

Usando la perpendicolarità si possono introdurre le coordinate cartesiane come sistema di riferimento del piano e dello spazio.

Nel piano la terna che descrive il sistema di riferimento è $\{O,\mathbf{i},\mathbf{j}\}$, dove $\mathbf{i}$ è un vettore di modulo 1 e $\mathbf{j}$ è il vettore che si ottiene ruotando $\mathbf{i}$ di $\frac{\pi}{2}$ in senso antiorario.

Nello spazio il sistema di riferimento è $\{O,\mathbf{i},\mathbf{j},\mathbf{k}\}$, dove $\mathbf{i},\mathbf{j}$ hanno le stesse caratteristiche di prima e $\mathbf{k}$ ha modulo 1, è perpendicolare a $\mathbf{i},\mathbf{j}$ - quindi al piano che li contiene - e verso dato dalla regola della mano destra ($\mathbf{i},\mathbf{j},\mathbf{k}$ si dice infatti terna destrorsa). Questo sistema di riferimento semplifica i calcoli e permette di ottenere una formula analitica per il prodotto scalare, sfruttando l'ortogonalità dei tre assi: dati $\mathbf{v}=x_1\mathbf{i}+y_1\mathbf{j}+z_1\mathbf{k}$, $\mathbf{w}=x_2\mathbf{i}+y_2\mathbf{j}+z_2\mathbf{k}$:

$$\begin{aligned}
\mathbf{v}\cdot\mathbf{w} &= (x_1\mathbf{i}+y_1\mathbf{j}+z_1\mathbf{k})\cdot(x_2\mathbf{i}+y_2\mathbf{j}+z_2\mathbf{k})=x_1\mathbf{i}\cdot(x_2\mathbf{i}+y_2\mathbf{j}+z_2\mathbf{k})+\\
&\quad +y_1\mathbf{j}\cdot(x_2\mathbf{i}+y_2\mathbf{j}+z_2\mathbf{k})+z_1\mathbf{k}\cdot(x_2\mathbf{i}+y_2\mathbf{j}+z_2\mathbf{k})\\
&= x_1x_2\mathbf{i}\cdot\mathbf{i}+x_1y_2\mathbf{i}\cdot\mathbf{j}+x_1z_2\mathbf{i}\cdot\mathbf{k}+x_2y_1\mathbf{i}\cdot\mathbf{j}+y_1y_2\mathbf{j}\cdot\mathbf{j}+y_1z_2\mathbf{j}\cdot\mathbf{k}+z_1x_2\mathbf{k}\cdot\mathbf{i}+z_1y_2\mathbf{k}\cdot\mathbf{j}+z_1z_2\mathbf{k}\cdot\mathbf{k}=\\
&= x_1x_2+0+0+0+y_1y_2+0+0+0+z_1z_2=x_1x_2+y_1y_2+z_1z_2
\end{aligned}$$

Questa formula è conseguenza della definizione geometrica di prodotto scalare e della scelta del sistema di riferimento. Si può interpretare il prodotto scalare anche come prodotto ta matrici:

$$\mathbf{v}\cdot\mathbf{w}=\begin{bmatrix} x_1 & y_1 & z_1 \end{bmatrix}\begin{bmatrix} x_2 \\ y_2 \\ z_1 \end{bmatrix}$$

3.1.5 Prodotto vettoriale e proprietà

Per ogni coppia di vettori liberi $\mathbf{v},\mathbf{w}$, si dice prodotto vettoriale (o prodotto esterno) di $\mathbf{v}$ e $\mathbf{w}$ il vettore libero $\mathbf{v}\times\mathbf{w}$ (indicato equivalentemente con $\mathbf{v}\wedge\mathbf{w}$), definito come segue.

Se $\mathbf{v}=\mathbf{0}$ o $\mathbf{w}=\mathbf{0}$ o $\mathbf{v},\mathbf{w}$ sono paralleli, $\mathbf{v}\times\mathbf{w}=\mathbf{0}$. Se $\mathbf{v},\mathbf{w}\neq\mathbf{0}$ e $\mathbf{v},\mathbf{w}$ non sono paralleli, il modulo di $\mathbf{v}\times\mathbf{w}$ è l'area del parallelogramma di lati $\mathbf{v},\mathbf{w}$, la direzione è perpendicolare al piano contenente $\mathbf{v},\mathbf{w}$, il verso è tale che $\{\mathbf{v},\mathbf{w},\mathbf{v}\times\mathbf{w}\}$ sia una terna destrorsa. Ha le seguenti proprietà:

- Bilinearità: $(\mathbf{u}+\mathbf{v}) \times \mathbf{w} = \mathbf{u} \times \mathbf{w} + \mathbf{v} \times \mathbf{w}$; $\mathbf{u} \times (\mathbf{v}+\mathbf{w}) = \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w}$; $(t\mathbf{u}) \times \mathbf{v} = t\,(\mathbf{u} \times \mathbf{v})$

- Anticommutatività: $\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}$

Dalla definizione segue $\mathbf{i} \times \mathbf{i} = \mathbf{j} \times \mathbf{j} = \mathbf{k} \times \mathbf{k} = \mathbf{0}$, perché sono coppie di vettori paralleli; $\mathbf{i} \times \mathbf{j} = \mathbf{k}$ ($\mathbf{i}$ e $\mathbf{j}$ sono lati di un quadrato), $\mathbf{j} \times \mathbf{k} = \mathbf{i} = -\mathbf{k} \times \mathbf{j}$, $\mathbf{k} \times \mathbf{i} = \mathbf{j} = -\mathbf{i} \times \mathbf{k}$.

Dalle proprietà segue che $\mathbf{r} \times (\lambda \mathbf{r}) = \mathbf{0}$ e viceversa se due vettori hanno prodotto vettoriale nullo sono paralleli, quindi il prodotto vettoriale dà un criterio di parallelismo: due vettori sono paralleli se e solo se il loro prodotto vettoriale è nullo (il vettore nullo è considerato parallelo a tutti i vettori).

Di nuovo, le coordinate cartesiane permettono di ottenere una formula analitica per il prodotto vettoriale, sfruttando l'ortogonalità dei tre assi: dati $\mathbf{v} = x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k}$, $\mathbf{w} = x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k}$:

$$\mathbf{v} \times \mathbf{w} = (x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k}) \times (x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k}) = x_1\mathbf{i} \times (x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k}) +$$

$$+ y_1\mathbf{j} \times (x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k}) + z_1\mathbf{k} \times (x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k})$$

$$= x_1x_2\mathbf{i} \times \mathbf{i} + x_1y_2\mathbf{i} \times \mathbf{j} + x_1z_2\mathbf{i} \times \mathbf{k} + y_1x_2\mathbf{j} \times \mathbf{i} + y_1y_2\mathbf{j} \times \mathbf{j} + y_1z_2\mathbf{j} \times \mathbf{k} + z_1x_2\mathbf{k} \times \mathbf{i} + z_1y_2\mathbf{k} \times \mathbf{j} + z_1z_2\mathbf{k} \times \mathbf{k} =$$

$$= (x_1y_2 - x_2y_1)\,\mathbf{k} - (x_1z_2 - x_2z_1)\,\mathbf{j} + (y_1z_2 - z_1y_2)\,\mathbf{i} = \det \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix} \mathbf{k} - \det \begin{bmatrix} x_1 & x_2 \\ z_1 & z_2 \end{bmatrix} \mathbf{j} + \det \begin{bmatrix} y_1 & y_2 \\ z_1 & z_2 \end{bmatrix} \mathbf{i}$$

che è uguale, con abuso di notazione, a $\det \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{bmatrix}$ (è un'uguaglianza solo formale).

Senza abuso di notazione, se $\mathbf{v} = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}$ e $\mathbf{w} = \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}$, $\mathbf{v} \times \mathbf{w} = \begin{pmatrix} y_1z_2 - z_1y_2 \\ x_2z_1 - x_1z_2 \\ x_1y_2 - x_2y_1 \end{pmatrix} = \begin{pmatrix} \det[R_2, R_3] \\ \det[R_3, R_1] \\ \det[R_1, R_2] \end{pmatrix}$, dove $[R_i, R_j]$ è la matrice che si ottiene accostando la riga i e la riga j, in quest'ordine. Dati $\mathbf{v}, \mathbf{w} \neq \mathbf{0}$, $\langle \mathbf{v}, \mathbf{v} \times \mathbf{w} \rangle = 0$ e $\langle \mathbf{w}, \mathbf{v} \times \mathbf{w} \rangle = 0$, coerentemente con la richiesta che $\mathbf{v} \times \mathbf{w}$ sia ortogonale sia a $\mathbf{v}$ che a $\mathbf{w}$; quindi $\mathbf{v}, \mathbf{v} \times \mathbf{w}$ sono linearmente indipendenti e $\mathbf{w}, \mathbf{v} \times \mathbf{w}$ sono linearmente indipendenti.

$\mathbf{v} \times \mathbf{w} = \mathbf{0}$ se e solo se $\mathbf{v}, \mathbf{w}$ sono paralleli: infatti $\mathbf{v} \times \mathbf{w} = \mathbf{0}$ significa che ogni sottomatrice 2×2 di $\begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \\ z_1 & z_2 \end{bmatrix}$ ha determinante nullo, quindi per il teorema di Kronecker il rango della matrice non può essere 2.

3.1.6 Prodotto misto

Per ogni terna di vettori liberi $\mathbf{u}, \mathbf{v}, \mathbf{w}$, si definisce prodotto misto di $\mathbf{u}, \mathbf{v}$ e $\mathbf{w}$ il numero reale

$$\langle \mathbf{u}, \mathbf{v} \times \mathbf{w} \rangle$$

E' noto che $\mathbf{v} \times \mathbf{w} = \det \begin{bmatrix} y_1 & z_1 \\ y_2 & z_2 \end{bmatrix} \mathbf{i} - \det \begin{bmatrix} x_1 & z_1 \\ x_2 & z_2 \end{bmatrix} \mathbf{j} + \det \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} \mathbf{k}$. Dato $\mathbf{u} = x_0\mathbf{i} + y_0\mathbf{j} + z_0\mathbf{k}$, si ha

$$\langle \mathbf{u}, \mathbf{v} \times \mathbf{w} \rangle = x_0 \det \begin{bmatrix} y_1 & z_1 \\ y_2 & z_2 \end{bmatrix} - y_0 \det \begin{bmatrix} x_1 & z_1 \\ x_2 & z_2 \end{bmatrix} + z_0 \det \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} = \det \begin{bmatrix} x_0 & y_0 & z_0 \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{bmatrix}$$

Ne segue che $\langle \mathbf{u}, \mathbf{v} \times \mathbf{w} \rangle = \langle \mathbf{v} \times \mathbf{w}, \mathbf{u} \rangle = \langle \mathbf{w} \times \mathbf{u}, \mathbf{v} \rangle$, perché scambiando un numero pari di righe il determinante non cambia.

$\langle \mathbf{u}, \mathbf{v} \times \mathbf{w} \rangle = 0$ se e solo se $\mathbf{u}$ appartiene al piano che contiene $\mathbf{w}$ e $\mathbf{v}$, cioè se $\mathbf{u}, \mathbf{v}, \mathbf{w}$ sono complanari (linearmente dipendenti): in tal caso infatti $\mathbf{v} \times \mathbf{w}$ dà un vettore perpendicolare al piano che li contiene, e se il prodotto scalare è nullo significa che $\mathbf{u}$ è ortogonale a tale vettore, cioè $\mathbf{u}$ appartiene allo stesso piano di $\mathbf{w}$ e $\mathbf{v}$. Se $\langle \mathbf{u}, \mathbf{v} \times \mathbf{w} \rangle \neq 0$, $\mathbf{u}, \mathbf{v}, \mathbf{w}$ non sono complanari e $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ può essere la base di un sistema di riferimento: il prodotto misto è diverso da zero se e solo se i vettori sono linearmente indipendenti. Come si vedrà più avanti, tre vettori di $\mathbb{R}^3$ sono linearmente indipendenti se e solo se il loro prodotto misto è non nullo, cioè il determinante della matrice che essi formano è non nullo, cioè il suo rango è massimo.

Il prodotto misto è quindi un criterio per capire se una terna di vettori può essere scelta come base di un sistema di riferimento nello spazio.

3.2 Rette

Una retta è univocamente determinata da due suoi punti distinti, oppure da un suo punto e una direzione.

Nel **caso 1**, se ho due punti distinti A e B, posso ricondurmi al secondo caso: la direzione si ricava dalla direzione del vettore applicato $A\vec{B}$. Altrimenti posso scrivere direttamente le equazioni cartesiane della retta: cioè equazioni che, dato un generico punto P, permettano di determinare se P appartiene alla retta o no.

Per appartenere alla retta P dev'essere tale che $A\vec{P}//A\vec{B}$.

Se sono nel piano ($n = 2$), $A = (x_A, y_A)$, $B = (x_B, y_B)$, $P = (x, y)$. Il vettore $A\vec{B}$ è il vettore che va da A a B, che, se applicato all'origine, è il vettore differenza ed è $A\vec{B} = \begin{bmatrix} x_B - x_A \\ y_B - y_A \end{bmatrix}$. Analogamente $A\vec{P} = \begin{bmatrix} x - x_A \\ y - y_A \end{bmatrix}$.

I due vettori sono paralleli se e solo se il loro prodotto vettoriale è nullo. Con abuso di notazione,

$$\vec{AB} \times \vec{AP} = \det \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_B - x_A & y_B - y_A & 0 \\ x - x_A & y - y_A & 0 \end{bmatrix} = \mathbf{k} \left[(x_B - x_A)(y - y_A) - (x - x_A)(y_B - y_A) \right]$$

E' nullo se e solo se $(x_B - x_A)(y - y_A) - (x - x_A)(y_B - y_A) = 0 \iff \frac{y - y_A}{y_B - y_A} = \frac{x - x_A}{x_B - x_A} \iff y = \frac{y_B - y_A}{x_B - x_A}(x - x_A) + y_A$, se $x_A \neq x_B, y_A \neq y_B$.

Se sono nello spazio ($n = 3$), $A = (x_A, y_A, z_A)$, $B = (x_B, y_B, z_B)$, $P = (x, y, z)$. $\vec{AB} = \begin{bmatrix} x_B - x_A \\ y_B - y_A \\ z_B - z_A \end{bmatrix}$, $\vec{AP} = \begin{bmatrix} x - x_A \\ y - y_A \\ z - z_A \end{bmatrix}$.

I due vettori sono paralleli se e solo se il loro prodotto vettoriale è nullo. Con abuso di notazione,

$$\vec{AB} \times \vec{AP} = \det \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_B - x_A & y_B - y_A & z_B - z_A \\ x - x_A & y - y_A & z - z_A \end{bmatrix}$$

E' nullo se e solo se $\det \begin{bmatrix} x_B - x_A & y_B - y_A \\ x - x_A & y - y_A \end{bmatrix} = \det \begin{bmatrix} x_B - x_A & z_B - z_A \\ x - x_A & z - z_A \end{bmatrix} = \det \begin{bmatrix} y_B - y_A & z_B - z_A \\ y - y_A & z - z_A \end{bmatrix} = 0$. Se $x_A \neq x_B, y_A \neq y_B, z_A \neq z_B$, si ottiene $\frac{y - y_A}{y_B - y_A} = \frac{x - x_A}{x_B - x_A}$ e $\frac{x - x_A}{x_B - x_A} = \frac{z - z_A}{z_B - z_A}$ (la terza è ridondante).

In generale, una retta è definita da $n - 1$ equazioni lineari, dove n è la dimensione dello spazio ambiente e 1 è il numero di parametri che servono per descrivere la retta. Quindi nello spazio una retta è descritta da un sistema lineare la cui matrice incompleta ha due equazioni, tre incognite (x, y, z) e rango uguale a 2 (la soluzione del sistema deve dipendere da un parametro).

Nel **caso 2**, ho un punto A e una direzione, che è data da un vettore libero $\mathbf{v}$. Allora un generico punto P appartiene alla retta se e solo se $\exists\, t \in \mathbb{R} : \vec{AP} = t\mathbf{v}$.

Se sono nel piano ($n = 2$), $A = (x_A, y_A)$, $P = (x, y)$, $\vec{AP} = \begin{bmatrix} x - x_A \\ y - y_A \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} x_A \\ y_A \end{bmatrix}$. $\vec{AP} = t\mathbf{v} \iff \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} x_A \\ y_A \end{bmatrix} = t \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \iff \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x_A \\ y_A \end{bmatrix} + t \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$. Coerentemente con la solita equazione della retta, la direzione determina il coefficiente angolare e individua una direzione (una classe di rette parallele), il punto A individua la specifica retta (opportunamente traslata).

Se sono nello spazio ($n = 3$), $A = (x_A, y_A, z_A)$, $P = (x, y, z)$, $\vec{AP} = \begin{bmatrix} x - x_A \\ y - y_A \\ z - z_A \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} x_A \\ y_A \\ z_A \end{bmatrix}$.

$\vec{AP} = t\mathbf{v} \iff \begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} x_A \\ y_A \\ z_A \end{bmatrix} = t \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \iff \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x_A \\ y_A \\ z_A \end{bmatrix} + t \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$. La retta è descritta con un parametro libero. Si può passare dalla rappresentazione cartesiana a quella parametrica (con un parametro) risolvendo il sistema di $n-1$ equazioni e usando il teorema di Rouché-Capelli. Essendo le $n-1$ equazioni indipendenti per ipotesi ed avendo 3 incognite, il vettore soluzione si scrive con $n - (n-1) + 1 = 2$ vettori, uno dei quali è il coefficiente di un parametro libero: A è il $\mathbf{v}_0$, la direzione è il vettore $\mathbf{v}_1$. Se $A = (0, 0, 0)$, il sistema che si risolve è un sistema omogeneo e $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = t \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$ è l'unica retta di direzione assegnata passante per l'origine. Per passare dalla rappresentazione parametrica a quella cartesiana, si esplicita il parametro in un'equazione aggiuntiva e si sostituisce nelle altre due.

3.3 Piani

In equazioni cartesiane ci sono $n - 2$ equazioni, in equazioni parametriche ci sono due parametri liberi: il piano è il luogo dei punti (x, y, z) generato da due direzioni $\mathbf{v}, \mathbf{w}$ e un punto A al variare di $t, s \in \mathbb{R}$:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x_A \\ y_A \\ z_A \end{bmatrix} + t \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} + s \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$$

Un piano è univocamente determinato da tre suoi punti A, B, C non allineati (un punto P appartiene al piano se e solo se $\vec{AB}$, $\vec{AC}$, $\vec{AP}$ sono complanari, cioè se $\vec{AB} \cdot \left(\vec{AC} \times \vec{AP}\right) = 0$) o un suo punto e due direzioni non parallele ($\mathbf{v}, \mathbf{w}$ nella scrittura precedente), o un suo punto A e una direzione $\mathbf{v}$ ortogonale al piano (un punto P appartiene al piano se e solo se $\vec{AP} \perp \mathbf{v}$, cioè $\vec{AP} \cdot \mathbf{v} = 0$). La prima e la terza condizione si usano per scrivere le equazioni cartesiane, la seconda per scrivere le equazioni parametriche.

3.4 Posizione reciproca delle rette in $\mathbb{R}^3$

Date due rette r, s, esse si dicono:

1. se non sono complanari, sghembe

2. se sono complanari:

    (a) se non si incontrano, parallele e distinte

    (b) se si incontrano in un solo punto, incidenti

    (c) se si incontrano in infiniti punti, parallele e coincidenti

Considero r, s in equazioni cartesiane: a ogni retta - attraverso le sue equazioni cartesiane - è associato un sistema lineare, a r $[A_1|\mathbf{b}_1]$ e a s $[A_2|\mathbf{b}_2]$. $A_1, A_2 \in M_{\mathbb{R}}(2,3)$ e $r(A_1) = r(A_2) = 2$ (il sistema lineare deve avere una soluzione dipendente da un parametro). Cerco eventuali punti d'intersezione, i. e. punti (x, y, z) che soddisfano entrambi i sistemi e quindi appartengono a entrambe le rette. Accosto quindi i due sistemi: devo risolvere $\begin{bmatrix} A_1 & \mathbf{b}_1 \\ A_2 & \mathbf{b}_2 \end{bmatrix} \in M_{\mathbb{R}}(4,4)$. Ogni eventuale soluzione del sistema soddisfa entrambi i sistemi originari, quindi appartiene a entrambe le rette.

Per il teorema di Rouché-Capelli, affinché la soluzione esista è necessario $r\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = r\begin{pmatrix} A_1 & \mathbf{b}_1 \\ A_2 & \mathbf{b}_2 \end{pmatrix}$.

2b Affinché la soluzione sia unica è inoltre necessario che il rango sia pari al numero di incognite, cioè $r\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = 3$. In tal caso le rette sono complanari e incidenti.

2c Affinché le soluzioni siano infinite (cioè la soluzione sia l'equazione di una retta) ci dev'essere un parametro libero, cioè $r\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = r\begin{pmatrix} A_1 & \mathbf{b}_1 \\ A_2 & \mathbf{b}_2 \end{pmatrix} = 2$. In tal caso le rette sono parallele e coincidenti.

Se $r\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \neq r\begin{pmatrix} A_1 & \mathbf{b}_1 \\ A_2 & \mathbf{b}_2 \end{pmatrix}$, non esiste soluzione. Poiché la differenza tra i due ranghi può solo essere 1 (può esserci al massimo un pivot nella colonna dei termini noti) e $r(A_1) = r(A_2) = 2$, $2 \leq r\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \leq 3$ e $2 \leq r\begin{pmatrix} A_1 & \mathbf{b}_1 \\ A_2 & \mathbf{b}_2 \end{pmatrix} \leq 4$, dunque ci sono solo due possibilità: $r\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = 2 \neq r\begin{pmatrix} A_1 & \mathbf{b}_1 \\ A_2 & \mathbf{b}_2 \end{pmatrix} = 3$ e $r\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = 3 \neq r\begin{pmatrix} A_1 & \mathbf{b}_1 \\ A_2 & \mathbf{b}_2 \end{pmatrix} = 4$. Allora considero i sistemi omogenei associati: significa che ora ciascuno dei sistemi $[A_1|\mathbf{0}]$

e $[A_2|\mathbf{0}]$ descrive la retta di prima traslata in modo da passare per l'origine: infatti $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ è sempre soluzione di un sistema omogeneo.

2a Se $r\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = 2 = r\begin{pmatrix} A_1 & \mathbf{0} \\ A_2 & \mathbf{0} \end{pmatrix}$, ci sono infinite soluzioni dipendenti da un parametro. Questo significa che le rette di partenza erano parallele distinte: traslate nell'origine coincidono, e la soluzione del sistema omogeneo è appunto una retta.

1 Se $r\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = 3 = r\begin{pmatrix} A_1 & \mathbf{0} \\ A_2 & \mathbf{0} \end{pmatrix}$, c'è un'unica soluzione. Questo significa che le rette di partenza erano sghembe: traslate nell'origine si intersecano in un solo punto, e la soluzione del sistema omogeneo è appunto unica (l'origine).

Non può essere $r\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \neq r\begin{pmatrix} A_1 & \mathbf{0} \\ A_2 & \mathbf{0} \end{pmatrix}$, perché nella colonna dei termini noti, essendoci solo zeri, non possono esserci pivot.

4 Spazi vettoriali

Finora si sono incontrati ripetutamente insiemi in cui sono definite operazioni con le medesime proprietà: matrici, vettori liberi, soluzioni di un sistema omogeneo... Tutti questi insiemi sono casi particolari di un oggetto generale.

Def Un insieme V si dice spazio vettoriale su $\mathbf{K}$ se V è dotato (cioè in V sono definite) due operazioni: un'operazione interna $+ : V \times V \to V$ e un'operazione esterna $\cdot : \mathbf{K} \times V \to V$ che soddisfano le seguenti proprietà:

V1 Proprietà associativa della somma: $\forall\ \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in V\ (\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 = \mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3)$

V2 Proprietà commutativa della somma: $\forall\ \mathbf{v}_1, \mathbf{v}_2 \in V\ \mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1$

V3 Esistenza dell'elemento neutro della somma: $\forall\ \mathbf{v} \in V\ \exists\ \mathbf{0}_V : \mathbf{v} + \mathbf{0}_V = \mathbf{0}_V + \mathbf{v} = \mathbf{v}$

V4 Esistenza dell'elemento opposto della somma: $\forall\ \mathbf{v} \in V\ \exists\ \mathbf{v}' \in V : \mathbf{v} + \mathbf{v}' = \mathbf{v}' + \mathbf{v} = \mathbf{0}_V$

V5 Proprietà distributiva a destra: $\forall\ \lambda \in \mathbf{K}, \forall\ \mathbf{v}_1, \mathbf{v}_2 \in V,\ \lambda(\mathbf{v}_1 + \mathbf{v}_2) = \lambda\mathbf{v}_1 + \lambda\mathbf{v}_2$

V6 Proprietà distributiva a sinistra: $\forall\ \lambda, \mu \in \mathbf{K}, \forall\ \mathbf{v} \in V,\ (\lambda + \mu)\mathbf{v} = \lambda\mathbf{v} + \mu\mathbf{v}$

V7 Proprietà associativa del prodotto: $\forall\ \lambda, \mu \in \mathbf{K}, \forall\ \mathbf{v} \in V,\ (\lambda\mu)\mathbf{v} = \lambda(\mu\mathbf{v}) = (\lambda\mathbf{v})\mu$

V8 Esistenza dell'elemento neutro del prodotto: $\forall\ \mathbf{v} \in V\ \exists\ \mathbf{1_K} : \mathbf{1_K v} = \mathbf{v}$.

1. $\mathbf{K} = \mathbb{R}$, $V = \{f : \mathbb{R} \to \mathbb{R}\} = (\mathbb{R})$ è uno spazio vettoriale su $\mathbb{R}$. Infatti, date $f, g \in V$, si definisce $(f+g)(x)$ come la funzione $f(x) + g(x)$ (è una somma puntuale su $\mathbb{R}$: si valuta $\forall\ x \in \mathbb{R}$); data $f \in V$ e $\lambda \in \mathbb{R}$, si definisce $(\lambda f)(x)$ come la funzione $\lambda f(x)\ \forall\ x \in \mathbb{R}$ (prodotto puntuale in $\mathbb{R}$). Per tali operazioni valgono le proprietà viste, perché le operazioni sono definite tramite le operazioni di $\mathbb{R}$, che è uno spazio vettoriale.

2. $C^0(\mathbb{R})$, $C^1(\mathbb{R})$,..., $C^k(\mathbb{R})$ sono spazi vettoriali.

3. Dato un generico campo $\mathbf{K}$, si definisce l'insieme dei polinomi nella variabile t a coefficienti in $\mathbf{K}$ $\mathbf{K}[t] = \{a_0 + a_1 t + ... + a_d t^d$, al variare di $a_0, a_1, ..., a_d \in \mathbf{K}\}$. Si definisce la somma $\sum_{i=1}^{d} a_i t^i + \sum_{i=1}^{e} b_i t^i := \sum_{i=1}^{\max\{d,e\}} (a_i + b_i) t^i$; il prodotto per scalari è $\lambda \left(\sum_{i=1}^{d} a_i t^i\right) := \sum_{i=1}^{d} (\lambda a_i) t^i$. $\mathbf{K}[t]$ è uno spazio vettoriale su $\mathbf{K}$.

4. L'insieme delle soluzioni di un sistema omogeneo $S = \{\mathbf{x} \in \mathbf{K}^n : A\mathbf{x} = \mathbf{0}\}$ è uno spazio vettoriale. Infatti contiene $\mathbf{0}$ e inoltre, dati $\mathbf{x}_1, \mathbf{x}_2 \in S$, $A(\mathbf{x}_1 + \mathbf{x}_2) = A\mathbf{x}_1 + A\mathbf{x}_2 = \mathbf{0} + \mathbf{0}$, quindi anche $\mathbf{x}_1 + \mathbf{x}_2 \in S$. Analogamente, dato $\mathbf{x} \in S$, $t \in \mathbf{K}$, $A(t\mathbf{x}) = t(A\mathbf{x}) = t \cdot \mathbf{0} = \mathbf{0}$, dunque $t\mathbf{x} \in S$.

Le proprietà delle operazioni definite in uno spazio vettoriale hanno varie conseguenze.

In primo luogo, si possono risolverere equazioni lineari: dati $\mathbf{x}, \mathbf{v} \in V$, $\lambda, \mu \in \mathbf{K}$, tali che $\lambda\mathbf{x} + \mu\mathbf{v} = \mathbf{0}$, vale

$$\lambda\mathbf{x} + \mu\mathbf{v} = \mathbf{0} \Longleftrightarrow \lambda\mathbf{x} + \mu\mathbf{v} + (-\mu\mathbf{v}) = \mathbf{0} + (-\mu\mathbf{v})\ \text{(V4)}$$

$$\lambda\mathbf{x} + (\mu\mathbf{v} - \mu\mathbf{v}) = \mathbf{0} - \mu\mathbf{v}\ \text{(V1)}$$

$$\lambda\mathbf{x} + \mathbf{0} = \mathbf{0} - \mu\mathbf{v}\ \text{(V4)}$$

$$\lambda\mathbf{x} = -\mu\mathbf{v}\ \text{(V3)}$$

$$\text{se } \lambda \neq 0,\ \left(\frac{1}{\lambda}\right)\lambda\mathbf{x} = \left(\frac{1}{\lambda}\right)(-\mu\mathbf{v})\ \text{ perché } \mathbf{K} \text{ è un campo}$$

$$\left(\frac{1}{\lambda}\lambda\right)\mathbf{x} = \left(\frac{1}{\lambda}(-\mu)\right)\mathbf{v}\ \text{(V7)}$$

$$\mathbf{1x} = \frac{-\mu}{\lambda}\mathbf{v}$$

$$\mathbf{x} = -\frac{\mu}{\lambda}\mathbf{v}\ \text{(V8)}$$

Inoltre l'elemento opposto di $\mathbf{v}$ è $(-1)\mathbf{v}$ ed è indicato con $-\mathbf{v}$.

Vale poi la legge di annullamento del prodotto:

$$\lambda\mathbf{v} = \mathbf{0} \Longrightarrow \lambda = 0 \vee \mathbf{v} = \mathbf{0}$$

Alcuni spazi vettoriali visti negli esempi precedenti sono sottinsiemi di altri spazi vettoriali: e. g. $C^k(\mathbb{R}) \subseteq C^{k-1}(\mathbb{R}) \subseteq ... \subseteq C^1(\mathbb{R}) \subseteq C^0(\mathbb{R}) \subseteq (\mathbb{R})$, e l'insieme dei polinomi a coefficienti in $\mathbf{K}$ di grado al più d è $\mathbf{K}[t]_{\leq d} \subseteq \mathbf{K}[t]$; l'insieme delle soluzioni di un sistema omogeneo è un sottinsieme di $\mathbf{K}^n$.

Def Sia $(V, +, \cdot)$ uno spazio vettoriale sul campo $\mathbf{K}$. Un sottinsieme non vuoto di V $H \subseteq V$ si dice sottospazio vettoriale se H è uno spazio vettoriale con le operazioni ereditate da V.

Tutti gli esempi fatti prima sono sottospazi vettoriali. La definizione precedente è equivalente a chiedere che H sia chiuso rispetto alle operazioni di V, cioè che, usando le operazioni di V in H, si ottenga ancora un elemento di H.

Come capire se un insieme è un sottospazio vettoriale o no?

Una condizione necessaria affinché H sia un sottospazio vettoriale è $\mathbf{0} \in H$: infatti, se $t = 0$, $t\mathbf{v} = \mathbf{0}$, e se $\mathbf{0} \notin H$ H non è chiuso rispetto al prodotto per scalare. Quindi $\mathbf{0} \notin H$ implica che H non è un sottospazio vettoriale.

1. $S = \{\mathbf{x} \in \mathbf{K}^n : A\mathbf{x} = \mathbf{b}, \text{ con } \mathbf{b} \neq \mathbf{0}\}$ non è uno spazio vettoriale perché non contiene $\mathbf{0}_V$.

2. Se $V = \mathbb{R}^2$, $H_1 = \{x\mathbf{i} + y\mathbf{j} : x \geq 0, y \geq 0\}$ non è un sottospazio vettoriale di V perché è chiuso rispetto alla somma ma non rispetto al prodotto per scalare: $(-1)\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \notin H$; $H_1 = \left\{x\vec{i} + y\vec{j} : xy \geq 0\right\}$, cioè primo e terzo quadrante, non è un sottospazio vettoriale di V perché è chiuso rispetto al prodotto per scalare ma non rispetto alla somma: $\begin{pmatrix} -3 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 3 \end{pmatrix} = \begin{pmatrix} -3 \\ 3 \end{pmatrix} \notin H$.

   Tutti e soli i sottospazi di $\mathbb{R}^2$ sono le rette passanti per l'origine e l'origine stessa, che in equazioni cartesiane sono infatti le soluzioni di un sistema omogeneo.

3. $V = \{A \in M_{\mathbb{R}}(n,n) : \det(A) = 0\} \subset M_{\mathbb{R}}(n,n)$ non è un sottospazio vettoriale. Infatti $A = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix}$ e $B = \begin{bmatrix} 0 & 3 \\ 0 & 4 \end{bmatrix}$ hanno determinante nullo ma la loro somma no.

4. Se $V = \mathbb{R}^2$, $H = \left\{x, y \in \mathbb{R}^2 : xy = 0\right\}$ non è un sottospazio vettoriale di V, perché non è chiuso rispetto alla somma: infatti $H = \left\{\mathbf{x} \in \mathbb{R}^2 : x = 0\right\} \cup \left\{\mathbf{y} \in \mathbb{R}^2 : y = 0\right\}$, e, come si vedrà più avanti, l'unione di due sottospazi in generale non è un sottospazio perché non è chiusa rispetto alla somma.

5. $\mathbb{Q}$ è uno spazio vettoriale sul campo $\mathbb{Q}$, ma non è un sottospazio vettoriale di $\mathbb{R}$, perché non è chiuso rispetto al prodotto per scalare.

In verità i sottospazi sono tutti e soli gli insiemi che si possono scrivere come nucleo di una matrice.

[**Def** Un insieme G si dice gruppo rispetto all'operazione $*$ se G è dotato di (cioè in G è definita) un'operazione interna $* : G \times G \to G$ che soddisfa le seguenti proprietà:

G1 Proprietà associativa: $\forall\ \mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3 \in G\ (\mathbf{g}_1 * \mathbf{g}_2) * \mathbf{g}_3 = \mathbf{g}_1 * (\mathbf{g}_2 * \mathbf{g}_3)$

G2 Esistenza dell'elemento neutro: $\forall\ \mathbf{g} \in G\ \exists\ e : \mathbf{g} * e = e * \mathbf{g} = \mathbf{g}$

G3 Esistenza dell'elemento inverso: $\forall\ \mathbf{g} \in G\ \exists\ \mathbf{g}' \in G : \mathbf{g} * \mathbf{g}' = e$

Un gruppo che soddisfa anche la proprietà commutativa ($\forall\ \mathbf{g}_1, \mathbf{g}_2 \in G\ \mathbf{g}_1 * \mathbf{g}_2 = \mathbf{g}_2 * \mathbf{g}_1$) si dice gruppo abeliano o commutativo. Un gruppo è quindi una coppia $(G, *)$.

1. L'insieme delle funzioni biunivoche f nella variabile x è un gruppo rispetto al prodotto di composizione, con elemento neutro la funzione identità x ed elemento inverso f^{-1}.

2. $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ sono gruppi abeliani rispetto alla somma.

3. $\mathbb{Q} - \{0\}, \mathbb{R} - \{0\}, \mathbb{C} - \{0\}$ sono gruppi abeliani rispetto al prodotto.

4. Ogni spazio vettoriale è un gruppo abeliano rispetto alla somma tra vettori.]

4.1 Combinazioni lineari, generatori, dipendenza lineare

Def Sia $(V, +, \cdot)$ uno spazio vettoriale sul campo $\mathbf{K}$, siano $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ vettori di V. Si dice combinazione lineare dei vettori $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ un vettore del tipo $\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + ... + \lambda_n\mathbf{v}_n$, con $\lambda_1, \lambda_2, ..., \lambda_n \in \mathbf{K}$. Si dice spazio vettoriale generato (*linear span*) da $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ l'insieme di tutte le combinazioni lineari di $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$. Tale insieme è $\{\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + ... + \lambda_n\mathbf{v}_n$, al variare di $\lambda_1, \lambda_2, ..., \lambda_n \in \mathbf{K}\}$ e si denota equivalentemente con $Span\,(\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n)$ o $\langle \mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n \rangle$ o $\langle \mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n \rangle$.

Se $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ sono vettori di V, $Span\,(\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n)$ è un sottospazio vettoriale di V. Infatti contiene $\mathbf{0}$, per le proprietà V1, V2, V6 la somma di due combinazioni lineari di $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ è una combinazione lineare di $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ e il prodotto di una combinazione lineare di $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ per uno scalare è una combinazione lineare di $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$, per le proprietà V5 e V7.

$$\begin{aligned}
(\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + ... + \lambda_n\mathbf{v}_n) + (\mu_1\mathbf{v}_1 + \mu_2\mathbf{v}_2 + ... + \mu_n\mathbf{v}_n) &= (\lambda_1 + \mu_1)\,\mathbf{v}_1 + (\lambda_2 + \mu_2)\,\mathbf{v}_2 + ... + (\lambda_n + \mu_n)\,\mathbf{v}_n \\
\mu\,(\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + ... + \lambda_n\mathbf{v}_n) &= \mu\lambda_1\mathbf{v}_1 + \mu\lambda_2\mathbf{v}_2 + ... + \mu\lambda_n\mathbf{v}_n
\end{aligned}$$

1. Un generico polinomio di $\mathbf{K}[t]_{\leq d}$ si scrive $a_0 + a_1 t + a_2 t^2 + ... + a_d t^d$. Questa scrittura è analoga a $\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + ... + \lambda_n \mathbf{v}_n$. Infatti $Span\left(1, t, t^2, ..., t^d\right) = \mathbf{K}[t]_{\leq d}$: $1, t, t^2, ..., t^d$ si dicono vettori generatori di $\mathbf{K}[t]_{\leq d}$.

2. Se A è una matrice $m \times n$, $C_1(A), ..., C_n(A)$ sono le sue colonne e $\mathbf{v} \in \mathbf{K}^n$, il prodotto $A\mathbf{v}$ è una combinazione lineare delle colonne di A. Se A è una matrice $m \times n$, $R_1(A), ..., R_n(A)$ sono le sue righe e $\mathbf{v} \in \mathbf{K}^m$, il prodotto $\mathbf{v}A$ è una combinazione lineare delle righe di A, del tipo $v_1(a_{11}|a_{12}|...|a_{1n}) + v_2(a_{21}|a_{22}|...|a_{2n}) + ... + v_m(a_{m1}|a_{m2}|...|a_{mn})$.

Def Un insieme di vettori $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ di uno spazio vettoriale V si dice insieme di generatori di V se $Span(\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n) = V$.

1. $Span\left(1, t, t^2, ..., t^d\right) = \mathbf{K}[t]_{\leq d}$: $1, t, t^2, ..., t^d$ sono vettori generatori di $\mathbf{K}[t]_{\leq d}$.

2. $V = \mathbb{R}^3$ è lo spazio vettoriale dei vettori liberi nello spazio. Considero $\{\mathbf{i}, \mathbf{j}\}$: $Span(\mathbf{i}, \mathbf{j}) = \left\{x\mathbf{i} + y\mathbf{j} \in \mathbb{R}^3 : x, y \in \mathbb{R}\right\}$. $Span(\mathbf{i}, \mathbf{j}) \subsetneq V$. Infatti $\mathbf{k} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \notin Span(\mathbf{i}, \mathbf{j})$. Quindi $\{\mathbf{i}, \mathbf{j}\}$ non è un insieme di generatori di $\mathbb{R}^3$; lo è $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$. Osservo che i tre vettori sono "indispensabili": ogni vettore libero $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ si descrive in modo unico come una combinazione lineare di $\mathbf{i}, \mathbf{j}, \mathbf{k}$: si prendono $\lambda_1 = x$, $\lambda_2 = y$, $\lambda_3 = z$.

3. $V = \mathbb{R}^3 = M_{\mathbf{K}}(3, 1)$. Considero $S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} \right\}$. S è un insieme di generatori di V? Questo equivale a chiedersi se $\forall\ \mathbf{v} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \in \mathbb{R}^3$ esistono $\lambda_1, \lambda_2, \lambda_3, \lambda_4 : \lambda_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + \lambda_3 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + \lambda_4 \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ (questa è la rappresentazione del primo tipo di un sistema lineare), cioè se

il seguente sistema lineare ha soluzione: $\begin{bmatrix} 1 & -2 & 0 & 1 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \iff \begin{bmatrix} 1 & -2 & 0 & 1 & a \\ 0 & 1 & 1 & -1 & b \\ 0 & 0 & 1 & 2 & c \end{bmatrix}$.

A è già ridotta a scala; per il teorema di Rouché-Capelli esistono infinite soluzioni dipendenti da $n-r = 4-3$ parametri. Quindi ogni vettore di $\mathbb{R}^3$ può essere scritto in infiniti modi come combinazione lineare dei vettori dati, che costituiscono dunque un insieme di generatori. Tutti i vettori sono indispensabili?

Considero $\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$. Risolvo il sistema lineare aggiungendo il parametro t per la variabile indipendente e uso il MEGJ. Risolvere il sistema omogeneo esplicita le relazioni lineari tra le colonne. $\begin{bmatrix} 1 & -2 & 0 & 1 & 0 \\ 0 & 1 & 1 & -1 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 & t \end{bmatrix} \to$

$\begin{bmatrix} 1 & -2 & 0 & 0 & -t \\ 0 & 1 & 1 & 0 & t \\ 0 & 0 & 1 & 0 & -2t \\ 0 & 0 & 0 & 1 & t \end{bmatrix} \to \begin{bmatrix} 1 & -2 & 0 & 0 & -t \\ 0 & 1 & 0 & 0 & 3t \\ 0 & 0 & 1 & 0 & -2t \\ 0 & 0 & 0 & 1 & t \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 0 & 5t \\ 0 & 1 & 0 & 0 & 3t \\ 0 & 0 & 1 & 0 & -2t \\ 0 & 0 & 0 & 1 & t \end{bmatrix}$. $\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} = \begin{pmatrix} 5 \\ 3 \\ -2 \\ 1 \end{pmatrix} t$. Scelgo $t = 1$: $5 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 3 \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} - 2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + 1 \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$. Esplicito la combinazione lineare rispetto alla variabile indipendente (ho scelto $t = 1$ in modo da poterlo fare comodamente): $\begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} = -5 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - 3 \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$. Quindi il quarto generatore, quello cui corrisponde la colonna priva di pivot nella matrice ridotta a scala, è $\begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} \in Span \left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right)$, dunque è ridondante:

non aggiunge nuove combinazioni lineari e gli altri tre vettori sono sufficienti per generare $\mathbb{R}^3$.

Def Un insieme di vettori $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ di uno spazio vettoriale $(V, +, \cdot)$ su $\mathbf{K}$ si dice insieme di vettori linearmente indipendenti se $\lambda_1 \mathbf{v}_1 + ... + \lambda_n \mathbf{v}_n = \mathbf{0}_V$ implica $\lambda_1 = ... = \lambda_n = 0$. Se al contrario esiste una soluzione non nulla dell'equazione $\lambda_1 \mathbf{v}_1 + ... + \lambda_n \mathbf{v}_n = \mathbf{0}_V$, i vettori si dicono linearmente dipendenti.

Nel secondo caso c'è almeno un vettore ridondante, cioè che non aggiunge combinazioni lineari.

1. I vettori $\begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ sono linearmente dipendenti: infatti una soluzione non nulla dell'equazione $\lambda_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + \lambda_3 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + \lambda_4 \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ è quella appena vista $\begin{pmatrix} 5 \\ 3 \\ -2 \\ 1 \end{pmatrix}$.

2. Dato un insieme di n vettori, per verificare se sono linearmente indipendenti è sufficiente risolvere il sistema omogeneo associato: se l'unica soluzione è il vettore nullo allora sono indipendenti. Questo accade quando il rango della matrice ottenuta accostando gli n vettori colonna è $r(A) = n$, per il teorema di Rouché-Capelli.

3. $\mathbb{R}^3 = Span(\mathbf{i}, \mathbf{j}, \mathbf{k})$. $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ sono linearmente indipendenti. Infatti $\lambda_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \lambda_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} =$ $\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ ha come unica soluzione $\lambda_1 = \lambda_2 = \lambda_3 = 0$.

L'indipendenza lineare di un insieme di generatori garantisce che ogni vettore dello spazio generato possa essere rappresentato in modo unico come combinazione lineare di tali vettori, come si spiega nel teorema successivo.

4.2 Basi e dimensione

Def Sia $(V, +, \cdot)$ uno spazio vettoriale su $\mathbf{K}$, sia $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ un insieme di vettori generatori di V. Se $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ sono linearmente indipendenti, l'insieme $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ si dice base di V.

Teorema (unicità della rappresentazione rispetto a una base)

$$\text{Hp: } B = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\} \text{ è una base di } V$$

$$\text{Ts: ogni vettore } \mathbf{v} \in V \text{ si scrive in modo unico}$$

$$\text{come combinazione lineare di } \mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$$

La tesi può anche essere espressa dicendo che il sistema lineare $[\mathbf{v}_1|\mathbf{v}_2|...|\mathbf{v}_n] \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ ... \\ \lambda_n \end{bmatrix} = \mathbf{v}$ ha un'unica soluzione per ogni $\mathbf{v} \in V$. Come si vedrà più avanti, questo significa che l'applicazione lineare $_A : \mathbf{K}^n \to V$ che usa $A = [\mathbf{v}_1|\mathbf{v}_2|...|\mathbf{v}_n]$ è biunivoca, e garantisce che la mappa delle coordinate sia una funzione.

Dim Per mostrare che esiste un unico modo di scrivere $\mathbf{v}$ come combinazione lineare dei vettori della base, considero due modi e mostro che coincidono. Poiché per ipotesi $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ sono generatori di V, $\forall\ \mathbf{v} \in V$ $\exists\ \lambda_1, \lambda_2, ..., \lambda_n : \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + ... + \lambda_n\mathbf{v}_n = \mathbf{v}$. Suppongo di considerare una seconda combinazione lineare $\mu_1\mathbf{v}_1 + \mu_2\mathbf{v}_2 + ... + \mu_n\mathbf{v}_n = \mathbf{v}$ e mostro che i coefficienti devono coincidere uno a uno. Infatti, sottraendo la seconda dalla prima, si ha $(\lambda_1 - \mu_1)\,\mathbf{v}_1 + (\lambda_2 - \mu_2)\,\mathbf{v}_2 + ... + (\lambda_n - \mu_n)\,\mathbf{v}_n = \mathbf{0}$. Ma poiché i vettori sono linearmente indipendenti, per definizione l'unica soluzione a tale equazione è il vettore nullo e quindi tutti i coefficienti sono nulli, cioè $\lambda_1 - \mu_1 = 0, ..., \lambda_n - \mu_n = 0$. Dunque i coefficienti coincidono uno a uno e le due rappresentazioni sono uguali. ■

1. Per verificare se un insieme di vettori $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ è una base di $\mathbf{K}^n$ bisogna verificare che la matrice $A = [\mathbf{v}_1|...|\mathbf{v}_n]$ abbia rango uguale al numero di righe (affinché il sistema lineare $A\mathbf{x} = \mathbf{v}$ abbia soluzione per ogni $\mathbf{v}$, cioè $\mathbf{v}_1, ..., \mathbf{v}_n$ siano generatori) e al numero di colonne (affinché la soluzione sia unica, per il teorema di Cramer).

Tutti gli spazi vettoriali hanno una base? In generale, quante basi hanno gli spazi vettoriali?

Lemma fondamentale

$$\text{Hp: } (V, +, \cdot) \text{ è uno spazio vettoriale; } T = \{\mathbf{v}_1, ..., \mathbf{v}_n\} \text{ è un insieme di generatori di } V;$$

$$S = \{\mathbf{w}_1, ..., \mathbf{w}_m\} \text{ è un insieme di } m \text{ vettori di } V \text{, con } m > n$$

$$\text{Ts: } S \text{ è un insieme di vettori linearmente dipendenti}$$

Dim Per ipotesi $V = Span\left(\mathbf{v}_1, ..., \mathbf{v}_n\right)$. In particolare $\mathbf{w}_1 = a_{11}\mathbf{v}_1 + ... + a_{1n}\mathbf{v}_n$,..., $\mathbf{w}_m = a_{m1}\mathbf{v}_1 + ... + a_{mn}\mathbf{v}_n$, con $a_{ij} \in \mathbf{K}$. Cerco di capire se i vettori di S sono linearmente dipendenti o no: considero $\lambda_1\mathbf{w}_1 + ... + \lambda_m\mathbf{w}_m = \mathbf{0}$ e mi chiedo se esistano soluzioni $\begin{bmatrix} \lambda_1 \\ ... \\ \lambda_m \end{bmatrix}$ non nulle. L'equazione è equivalente a $\lambda_1\left(a_{11}\mathbf{v}_1 + ... + a_{1n}\mathbf{v}_n\right) + ... + \lambda_m\left(a_{m1}\mathbf{v}_1 + ... + a_{mn}\mathbf{v}_n\right) = \mathbf{0} \Longleftrightarrow \left(\lambda_1 a_{11} + ... + \lambda_m a_{m1}\right)\mathbf{v}_1 + ... + \left(\lambda_1 a_{1n} + ... + \lambda_m a_{mn}\right)\mathbf{v}_n = \mathbf{0}$. Questa equazione è sicuramente soddisfatta se si annullano simultaneamente tutti i coefficienti: questa condizione permette quindi di trovare un sottinsieme delle soluzioni dell'equazione originaria.

$$\begin{cases} \lambda_1 a_{11} + ... + \lambda_m a_{m1} = 0 \\ ... \\ \lambda_1 a_{1n} + ... + \lambda_m a_{mn} = 0 \end{cases} \Longleftrightarrow \begin{bmatrix} a_{11} & ... & a_{1m} \\ ... & ... & ... \\ a_{1n} & ... & a_{mn} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ ... \\ \lambda_m \end{bmatrix} = \begin{bmatrix} 0 \\ ... \\ 0 \end{bmatrix}$$

Questo è un sistema lineare con n righe e m colonne. Poiché per ipotesi $m > n$, $r(A) \leq n < m$; per il teorema di Rouché-Capelli, le soluzioni dipendono quindi da $m - r(A) > 0$ parametri liberi (almeno uno). Dunque ci sono infinite soluzioni non nulle dell'equazione $\lambda_1\mathbf{w}_1 + ... + \lambda_m\mathbf{w}_m = \mathbf{0}$ e S è un insieme di vettori linearmente dipendenti. ■

Questo lemma è fondamentale per dimostrare il prossimo teorema.

Teorema della dimensione

Hp: $(V, +, \cdot)$ è uno spazio vettoriale su $\mathbf{K}$; $T = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$, $S = \{\mathbf{w}_1, ..., \mathbf{w}_m\}$ sono basi di V

Ts: $m = n$

cioè tutte le basi di uno spazio vettoriale hanno la stessa cardinalità.

Dim Dimostrare $m = n$ è equivalente a dimostrare $m \leq n$ e $n \leq m$.

Dimostro $m \leq n$. Vedo T come un insieme di generatori, S come un insieme di vettori linearmente indipendenti. Se fosse $m > n$ i vettori di S sarebbero linearmente dipendenti per il lemma fondamentale, ma questo è assurdo, quindi dev'essere $m \leq n$.

Dimostro $n \leq m$. Vedo S come un insieme di generatori, T come un insieme di vettori linearmente indipendenti. Se fosse $n > m$ i vettori di T sarebbero linearmente dipendenti per il lemma fondamentale, ma questo è assurdo, quindi dev'essere $n \leq m$. Dunque $n = m$. ■

Def Sia $(V, +, \cdot)$ uno spazio vettoriale su $\mathbf{K}$. Si dice dimensione di V, e si indica con $\dim V$, il numero di elementi di una base di V.

La definizione è ben data perché tutte le basi hanno la stessa cardinalità. Se V contiene solo il vettore nullo, la dimensione è 0 per convenzione.

1. $V = \mathbb{R}^2$; $\{\mathbf{i}, \mathbf{j}\}$ è una base di V. Dunque $\dim V = 2$.

2. $V = \mathbf{K}[t]_{\leq d}$. $\{1, t, t^2, ..., t^d\}$ è una base di V. Dunque $\dim V = d + 1$. La base $\{1, t, t^2, ..., t^d\}$ viene detta base canonica.

3. $V = M_{\mathbf{K}}(2,3)$. La generica matrice di $M_{\mathbf{K}}(2,3)$ è $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} = a_{11}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + a_{12}\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} +$ $... + a_{32}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Quindi $\left\{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right\}$ è una base di V. In generale, $\dim(M_{\mathbf{K}}(m,n)) = mn$.

Ci sono ulteriori conseguenze del teorema della dimensione e del lemma fondamentale.

Se $\dim V = n$, n è il massimo numero di vettori linearmente indipendenti in V (per il lemma fondamentale in forma negata) e n è il numero minimo di vettori generatori di V: infatti, se si avessero meno vettori generatori, se essi fossero linearmente indipendenti sarebbero una base e $\dim V$ sarebbe minore di n, se fossero linearmente dipendenti sarebbe possibile toglierne alcuni in modo da avere una base e $\dim V$ sarebbe sempre minore di n.

Inoltre, considero $T = \{\mathbf{v}_1, ..., \mathbf{v}_k\}$, con $k < n$, linearmente indipendenti. T è un insieme di generatori? Per le conseguenze del lemma fondamentale, esiste $\mathbf{w} : S = T \cup \{\mathbf{w}\}$ è un insieme di vettori linearmente indipendenti. Quindi $w \notin Span(\mathbf{v}_1, ..., \mathbf{v}_k)$: porta un'informazione nuova, perciò $Span(\mathbf{v}_1, ..., \mathbf{v}_k) \subsetneqq V$ e T non è un insieme di generatori.

In generale, col teorema della dimensione e il lemma fondamentale si dimostra che se $\dim V = n$ e un insieme di vettori di V di cardinalità n è un insieme di vettori linearmente indipendenti, allora tale insieme è una base; se $\dim V = n$ e un insieme di vettori di V di cardinalità n è un insieme di generatori, allora tale insieme è una base (perché ogni insieme di generatori contiene una base).

Ci si chiedeva prima se tutti gli spazi vettoriali avessero una base. Ci sono spazi vettoriali con insiemi di vettori linearmente indipendenti di cardinalità arbitraria; tali spazi non hanno una base con un numero finito di elementi, perché non c'è un massimo numero di vettori linearmente indipendenti.

1. $V = \mathbf{K}[t]$ è lo spazio dei polinomi, senza limitazioni sul grado. $T_d = \{1, t, t^2, ..., t^d\}$ è una base di $\mathbf{K}[t]_{\leq d}$: in questo caso però la cardinalità dell'insieme dev'essere infinita; se ci si fermasse a d non si potrebbero generare i polinomi di grado superiore. Si dice quindi che $\mathbf{K}[t]$ è uno spazio vettoriale di dimensione infinita.

Uno spazio vettoriale con base si dice finitamente generato; altrimenti si dice non finitamente generato.

1. Considero $V = \mathbf{K}[t]_{\leq 2}$; mi chiedo se $S = \left\{1+t, 1+t-t^2, 2-t^2\right\}$ è una base di $\mathbf{K}[t]$. Per stabilirlo devo verificare due proprietà: se i vettori sono linearmente indipendenti (cioè $\lambda_1(1+t) + \lambda_2\left(1+t-t^2\right) + \lambda_3\left(2-t^2\right) = 0_V$ implica $\lambda_1 = \lambda_2 = \lambda_3 = 0$) e se sono generatori di V (cioè $\forall\ a_0 + a_1 t + a_2 t^2\ \exists\ \lambda_1, \lambda_2, \lambda_3 : \lambda_1(1+t) + \lambda_2\left(1+t-t^2\right) + \lambda_3\left(2-t^2\right) = a_0 + a_1 t + a_2 t^2$). Per quanto detto sulle conseguenze del teorema fondamentale, dato che la dimensione di V è nota, una di queste due condizioni è sufficiente.

   (a) $\lambda_1(1+t) + \lambda_2\left(1+t-t^2\right) + \lambda_3\left(2-t^2\right) = 0_V \iff \lambda_1 + \lambda_2 + 2\lambda_3 + (\lambda_1 + \lambda_2)\, t + (-\lambda_2 - \lambda_3)\, t^2 = 0_V$. Un polinomio è nullo se i coefficienti di tutti i suoi termini sono nulli, quindi risolvo il sistema omogeneo $\begin{cases} \lambda_1 + \lambda_2 + 2\lambda_3 = 0 \\ \lambda_1 + \lambda_2 = 0 \\ -\lambda_2 - \lambda_3 = 0 \end{cases}$ e, usando il MEG e il teorema di Rouché-Capelli, se capisco che c'è un'unica soluzione, allora è $\lambda_1 = \lambda_2 = \lambda_3 = 0$ e i vettori sono linearmente indipendenti.

   (b) $\forall\ a_0 + a_1 t + a_2 t^2\ \exists\ \lambda_1, \lambda_2, \lambda_3 : \lambda_1(1+t) + \lambda_2\left(1+t-t^2\right) + \lambda_3\left(2-t^2\right) = a_0 + a_1 t + a_2 t^2$, cioè $\lambda_1 + \lambda_2 + 2\lambda_3 + (\lambda_1 + \lambda_2)\, t + (-\lambda_2 - \lambda_3)\, t^2 = a_0 + a_1 t + a_2 t^2$? Equivale a chiedersi se il sistema lineare $\begin{cases} \lambda_1 + \lambda_2 + 2\lambda_3 = a_0 \\ \lambda_1 + \lambda_2 = a_1 \\ -\lambda_2 - \lambda_3 = a_2 \end{cases}$ ammette almeno una soluzione per ogni a_0, a_1, a_2. La matrice dei coefficienti è $A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 0 \\ 0 & -1 & -1 \end{bmatrix}$. E' evidente che la prima condizione è equivalente a questa, perché se i vettori sono linearmente indipendenti il sistema omogeneo considerato ha un'unica soluzione e $r(A) = n$ (in questo caso 3), per il teorema di Rouché-Capelli. Allora, per il teorema di Cramer, c'è un'unica soluzione per ogni a_0, a_1, a_2.

   In questo caso, riducendo a scala A si ottiene $\begin{bmatrix} 1 & 1 & 2 \\ 0 & -1 & -1 \\ 0 & 0 & -2 \end{bmatrix}$: essendo effettivamente $r(A) = 3$, entrambe le condizioni sono soddisfatte e la terna di polinomi data è una base di V.

   (c) n vettori di $\mathbf{K}^n$ sono una sua base se e solo se il rango della matrice quadrata che essi formano è n, che è equivalente a chiedere che il suo determinante sia non nullo.

(d) *Sia* $V=\{p(x)=a+bx+cx^2 : a,b,c\in\mathbb{R}\}$ *lo spazio vettoriale dei polinomi reali di grado* ≤ 2, *e sia* $H=\{p(x)\in V : p(2)=0\}$ *un suo sottospazio. Si determini una base di* H *il cui primo elemento sia* $p_1(x)=x-2$.

Il generico polinomio di H ha coefficienti $a,b,c : a+2b+4c=0$, quindi è della forma $-2b-4c+bx+cx^2$. Risolvendo il sistema lineare aggiungendo i due parametri liberi si trova $\begin{pmatrix} a \\ b \\ c \end{pmatrix} = t\begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix} + s\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}$, e sostituendo in $a+bx+cx^2$ si ottiene $t\left(x^2-4\right)+s\left(x-2\right)$. $\left\{x^2-4, x-2\right\}$ è una base di H: i vettori sono linearmente indipendenti e generatori di H (per ottenere il generico polinomio basta prendere $t=c$, $s=b$), e $\dim H=2$. In realtà, avendo intuito che $\dim H=2$ (la condizione $p\left(2\right)=0$ toglie un grado di libertà in V), per trovare una base di H contenente $x-2$ è sufficiente prendere un qualsiasi polinomio di H linearmente indipendente (cioè non multiplo) da $x-2$, e. g. x^2+x-6.

Questo esempio mostra come un problema presentato nello spazio dei polinomi si traduca naturalmente in un problema matriciale.

4.3 Spazi generati da una matrice

Def Data $A\in M_{\mathbf{K}}\left(m,n\right)$, si definisce spazio delle righe di A, e si indica con $row\left(A\right)$, il sottospazio vettoriale di $\mathbf{K}^n$ (cioè $M_{\mathbf{K}}\left(1,n\right)$) generato dalle righe di A; si definisce spazio delle colonne di A, e si indica con $col\left(A\right)$, il sottospazio vettoriale di $\mathbf{K}^m$ (cioè $M_{\mathbf{K}}\left(m,1\right)$) generato dalle colonne di A; si definisce nucleo di A, e si indica con $\ker\left(A\right)$, il sottospazio vettoriale di $\mathbf{K}^n$ (cioè $M_{\mathbf{K}}\left(n,1\right)$) costituito dalle soluzioni del sistema omogeneo avente A come matrice dei coefficienti.

1. $A=\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$. $row\left(A\right)=Span\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}\right)\subseteq\mathbb{R}^3$. $col\left(A\right)=Span\left(\begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \end{bmatrix}\right)\subseteq\mathbb{R}^2$. $\ker\left(A\right)=\left\{\begin{bmatrix} x \\ y \\ z \end{bmatrix}\in\mathbb{R}^3 : A\begin{bmatrix} x \\ y \\ z \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\right\}$. Risolvo il sistema omogeneo per scrivere esplicitamente il

nucleo: riducendo a scala A ottengo $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \end{bmatrix}$, e aggiungendo il parametro per la variabile indipendente e risolvendo col MEGJ si ottiene $\begin{bmatrix} 1 & 0 & 0 & t \\ 0 & 1 & 0 & -2t \\ 0 & 0 & 1 & t \end{bmatrix} : \begin{bmatrix} x \\ y \\ z \end{bmatrix} = t \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$, quindi $\ker(A) = Span\left(\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}\right)$.

2. $V = \mathbf{K}^3$. $S = \left\{ \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 7 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} \right\}$. S è un insieme di generatori di V? Se sì, come eliminare la ridondanza (dato che $\dim V = 3$), cioè come estrarne una base? Ci sono due possibilità:

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 3 & 4 & 7 \\ 0 & -1 & 2 \\ 1 & 0 & -2 \end{bmatrix} \text{ e } row(A) = Span(S)$$

$$A = \begin{bmatrix} 1 & 3 & 0 & 1 \\ -1 & 4 & -1 & 0 \\ 2 & 7 & 2 & -2 \end{bmatrix} \text{ e } col(A) = Span(S)$$

Lemma (invarianza dello spazio delle righe rispetto al MEG)

Hp: $A \in M_{\mathbf{K}}(m,n)$; $B \in M_{\mathbf{K}}(m,n)$ è ottenuta da A con una sequenza di operazioni del MEG

Ts: $row(A) = row(B)$

Ha senso che lo spazio delle righe che si conservi, perché con le operazioni del MEG si annullano le righe che sono combinazioni lineari delle altre, cioè che non aggiungono nessuna combinazione lineare.

Dim $row(A) = Span(R_1(A), R_2(A), ..., R_m(A))$; $row(B) = Span(R_1(B), R_2(B), ..., R_m(B))$. Mostrare che $row(A) = row(B)$ equivale a mostrare $row(A) \subseteq row(B)$ e $row(B) \subseteq row(A)$.

Mostro $row(A) \subseteq row(B)$, i. e. che ogni combinazione lineare delle righe di A è anche una combinazione lineare delle righe di B: è sufficiente mostrare che tutti i generatori di $row(A)$ appartengono a $row(B)$, cioè $R_i(A) \in row(B)\ \forall\ i = 1, ..., m$, perché allora anche ogni loro combinazione lineare apparterrà a $row(B)$, essendo $row(B)$ un sottospazio. Lo stesso vale per mostrare $row(B) \subseteq row(A)$. E' sufficiente inoltre verificarlo per una singola operazione del MEG (mi permette di scrivere una catena di uguaglianze con $row(A)$ all'inizio e $row(B)$ alla fine), cioè come se B fosse stata ottenuta da A con una singola operazione del MEG.

E' ovvio nel caso dello scambio di righe: sta cambiando solo l'ordine dei generatori, che è irrilevante. Nel caso della moltiplicazione per $\lambda \neq 0$, si ha che $R_i(B) = \lambda R_i(A)$: in particolare vale $R_i(B) = 0R_1(A) + ... + \lambda R_i(A) + ... + 0R_m(A)$, l'unico generatore che differisce nei due sottospazi è $R_i(B)$, e $R_i(B) \in Span(R_1(A), ..., R_m(A)) = row(A)$. Inoltre $R_i(A) = \frac{1}{\lambda} R_i(B)$, quindi analogamente $R_i(A) \in row(B)$. Nel caso della sostituzione di una riga con la sua somma con un multiplo di un'altra, si ha che $R_i(B) = R_i(A) + \lambda R_j(A) \in row(A)$ e $R_i(A) = R_i(B) - \lambda R_j(B) \in row(B)$ (perché R_j è uguale sia in A che in B). $\blacksquare$

Quindi, se riduco A a scala e ottengo U, le righe non nulle di U sono una base di $row(U) = row(A)$. Infatti le righe di A sono un insieme di generatori, ma possibilmente non linearmente indipendenti: le operazioni del MEG fanno sì che quelle che sono combinazione lineari delle altre diventino nulle, cioè eliminano quelle ridondanti, per cui le rimanenti costituiscono una base. Se U non ha righe nulle, cioè $r(U) = \min\{m, n\}$, significa che le righe di A erano linearmente indipendenti.

Teorema del rango

$$\text{Hp: } A \in M_{\mathbf{K}}(m, n),\ r = r(A)$$

$$\text{Ts: } (row(A) = row(U))$$

$$\text{(i) } \dim(row(U)) = \dim(row(A)) = r$$

$$\text{(ii) } \dim(\ker(A)) = n - r$$

$$\text{(iii) } \dim(col(A)) = r$$

Quindi il rango di una matrice è la dimensione dello spazio delle righe (o delle colonne).

Dim Dall'ultima lemma visto è noto $row(A) = row(U)$.

(i) Voglio mostrare che le righe contenenti i pivot (che sono r) formano una base di $row(A)$. Mostro che sono generatori di $row(U)$. Un generico vettore $\mathbf{v} \in row(U)$ si scrive come $\lambda_1 R_1(U) + ... + \lambda_r R_r(U) + \lambda_{r+1} 0 + ... + \lambda_m 0$: quindi $Span(R_1(U), ..., R_m(U)) = Span(R_1(U), ..., R_r(U))$. Mostro che sono linearmente indipendenti. Prendo la matrice ridotta a scala U e tolgo le righe nulle: voglio mostrare che $\lambda_1 R_1(U) + ... + \lambda_r R_r(U) = \mathbf{0} \Longrightarrow \lambda_1 = ... = \lambda_r = 0$. Mostro che $\lambda_1 = 0$ considerando la posizione della colonna contenente il primo pivot. Dal momento che nella colonna del primo pivot p_1 è l'unico elemento non nullo, in tal colonna si deve avere $\lambda_1 p_1 + 0 = 0$, dunque $\lambda_1 = 0$. Per dimostrare $\lambda_2 = 0$ considero la colonna del secondo pivot; so che $\lambda_1 = 0$. Allora la somma sarà $\lambda_1 u_{1k} + \lambda_2 p_2 + 0 = 0$: $\lambda_2 = 0$ e così via. Dunque $\lambda_1 = ... = \lambda_r = 0$ e $R_1(U), ..., R_r(U)$ sono linearmente indipendenti: perciò $\dim(row(U)) = r$.

(ii) Per definizione $\ker(A) = \{\mathbf{x} \in \mathbf{K}^n : A\mathbf{x} = \mathbf{0}\}$. Per il teorema di Rouché-Capelli $\exists\ \mathbf{v}_1, ..., \mathbf{v}_{n-r} \in \mathbf{K}^n : \mathbf{x} = \lambda_1\mathbf{v}_1 + ... + \lambda_{n-r}\mathbf{v}_{n-r} \in \ker(A)$. Quindi $Span\left(\mathbf{v}_1, ..., \mathbf{v}_{n-r}\right) = \ker(A)$: sono generatori. Mostro che $\mathbf{v}_1, ..., \mathbf{v}_{n-r}$ sono linearmente indipendenti. Supponiamo che le variabili indipendenti del sistema $A\mathbf{x} = \mathbf{0}$ siano le ultime $n-r$: $r(A) = r$. Scrivo il generico elemento del kernel $\lambda_1\mathbf{v}_1 + ... + \lambda_{n-r}\mathbf{v}_{n-r} = \begin{bmatrix} * \\ \lambda_1 \\ \lambda_2 \\ ... \\ \lambda_{n-r} \end{bmatrix} = \lambda_1 \begin{bmatrix} * \\ 1 \\ 0 \\ ... \\ 0 \end{bmatrix} + ... + \lambda_{n-r} \begin{bmatrix} * \\ 0 \\ ... \\ 0 \\ 1 \end{bmatrix}$ dove $*$ sono le componenti dei vettori nelle righe $1, ..., r$. $\lambda_1\mathbf{v}_1 + ... + \lambda_{n-r}\mathbf{v}_{n-r} = \mathbf{0}$ implica (nella riga $r+1$) $\lambda_1 = 0$, (nella riga $r+2$) $\lambda_2 = 0, ..., \lambda_{n-r} = 0$.

(iii) [In generale $col(A) \neq col(U)$: ad esempio se $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$ e $U = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, $col(A) = Span\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right)$ è una retta passante per l'origine con pendenza 2, $col(U) = Span\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)$.]

$\ker(A) = \ker(U)$ perché l'insieme delle soluzioni del sistema omogeneo $A\mathbf{x} = \mathbf{0}$ (e in generale di un sistema lineare) è lasciato immutato dalle operazioni elementari sulle righe. $\ker(A) = \{$relazioni di dipendenza lineare tra le colonne di $A\}$: infatti $\ker(A)$ è l'insieme degli $\begin{pmatrix} x_1 \\ ... \\ x_n \end{pmatrix}$ che soddisfano $\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n = 0 \\ ... \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n = 0 \end{cases}$: $\begin{bmatrix} a_{11} \\ a_{21} \\ ... \\ a_{m1} \end{bmatrix} x_1 + \begin{bmatrix} a_{12} \\ a_{22} \\ ... \\ a_{m2} \end{bmatrix} x_2 + ... + \begin{bmatrix} a_{1n} \\ a_{2n} \\ ... \\ a_{nn} \end{bmatrix} x_n = \mathbf{0}$.

$\mathbf{x} \in \ker(A)$ se e solo se soddisfa tale equazione. Dire che $\ker(A) = \ker(U)$ è equivalente a dire che se $\{C_{i_1}(A), ..., C_{i_k}(A)\}$ (dove $C_{i_1}(A)$ identifica la prima colonna di A scelta arbitrariamente) è un insieme di vettori linearmente indipendenti, allora anche $\{C_{i_1}(U), ..., C_{i_k}(U)\}$ è un insieme di vettori linearmente indipendenti. [Per esempio le prime tre colonne di A sono indipendenti se e solo se le prime tre colonne di U sono indipendenti, perché questo significa che nel nucleo non si trova nessun vettore con le prime tre componenti non tutte nulle e le altre uguali a zero.] Si può interpretare $\dim(col(A))$ come il massimo numero di vettori colonna linearmente indipendenti di A. Allora, scegliendo come $\{C_{i_1}(A), ..., C_{i_k}(A)\}$ l'insieme dei vettori colonna linearmente indipendenti di cardinalità massima, anche $\{C_{i_1}(U), ..., C_{i_k}(U)\}$ è un insieme di vettori linearmente indipendenti di U di cardinalità massima, quindi $\dim(col(A)) = \dim(col(U))$. Ora mostro che $\dim(col(U)) = r$. Voglio mostrare che le colonne non contenenti un pivot si esprimono come combinazione lineare delle colonne contenenti i pivot. Prendo la soluzione di $U\mathbf{x} = \mathbf{0}$ con $t_1 = 1$, $t_2 = ... = t_{n-r} = 0$: $\mathbf{v}_1 = \begin{pmatrix} v_{11} \\ ... \\ v_{1r} \\ 1 \\ 0 \\ ... \\ 0 \end{pmatrix}$. Allora $v_{11}C_1(U) + ... + v_{1r}C_r(U) + 1 \cdot C_{r+1}(U) + ... + 0C_n(U) = \mathbf{0}$ perché $\mathbf{v}_1$ è una soluzione (sto supponendo che i pivot siano tutti a sinistra). Quindi $C_{r+1}(U) = -v_{11}C_1(U) - ... - v_{1r}C_r(U)$. ... Prendo la soluzione di $U\mathbf{x} = \mathbf{0}$ con $t_1 = ... = t_{n-r-1} = 0$, $t_{n-r} = 1$: $\mathbf{v}_{n-r} = \begin{pmatrix} v_{n-r,1} \\ ... \\ v_{n-r,r} \\ 0 \\ 0 \\ ... \\ 1 \end{pmatrix}$. Allora $v_{n-r,1}C_1(U) + ... + v_{n-r,r}C_r(U) + 0C_{r+1}(U) + ... + 0C_{n-1}(U) + 1C_n(U) = \mathbf{0}$ perché $\mathbf{v}_{n-r}$ è una soluzione. Quindi $C_n(U) = -v_{n-r,1}C_1(U) - ... - v_{n-r,r}C_r(U)$. Ho mostrato che $C_1(U), ..., C_r(U)$ sono generatori di $col(U)$: le ultime $n-r$ colonne non aggiungono combinazioni lineari, perché sono combinazioni lineari delle altre. Sono anche linearmente indipendenti: mostro che $\lambda_1 C_1(U) + ... + \lambda_r C_r(U) = \mathbf{0}$ implica $\lambda_1 = ... = \lambda_r = 0$.

$$\lambda_1 \begin{bmatrix} p_1 \\ 0 \\ ... \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} * \\ p_2 \\ ... \\ 0 \end{bmatrix} + ... + \lambda_r \begin{bmatrix} * \\ ... \\ p_r \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ ... \\ 0 \end{bmatrix}$$

Considero la riga r: corrisponde all'equazione $\lambda_r p_r = 0$, $p_r \neq 0$, quindi dev'essere $\lambda_r = 0$. Considero la riga $r-1$: $\lambda_{r-1}p_{r-1} + \lambda_r u_{r-1,r} = 0$, ma $\lambda_r = 0$ e $p_{r-1} \neq 0$, quindi $\lambda_{r-1} = 0$, e così via fino alla riga 1. Quindi i vettori sono linearmente indipendenti, quindi $\dim(col(U)) = r = \dim(col(A))$. ■

Ne segue che:

- Possiamo usare il MEG per l'estrazione di una base trasformando i vettori riga in vettori colonna, oppure no. Se $S = \{\mathbf{v}_1, ..., \mathbf{v}_k\}$, $\mathbf{v}_i \in \mathbf{K}^n$, e tali vettori sono generatori dello spazio vettoriale di cui si cerca la base (quindi $k > n$), si considera $A = [\mathbf{v}_1|\mathbf{v}_2|...|\mathbf{v}_k]$. Col MEG si ottiene U: la base estratta è formata dai vettori $\mathbf{v}_i$ corrispondenti alle colonne contenenti i pivot, perché lo spazio delle colonne non si preserva ma l'indipendenza lineare sì. Se invece si considera $A = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ ... \\ \mathbf{v}_k \end{bmatrix}$ usando i vettori come righe, le righe non nulle (= le righe che contengono i pivot) di U sono i vettori che formano la base (posso anche prendere le righe originarie nelle posizione delle righe coi pivot, ma sono meno semplici da usare perché hanno meno componenti nulle).

- Vogliamo completare un insieme di vettori linearmente indipendenti in modo che siano una base. Se $S = \{\mathbf{v}_1, ..., \mathbf{v}_k\}$, $\mathbf{v}_i \in \mathbf{K}^n$, e tali vettori sono linearmente indipendenti (quindi $k < n$), si considera $A = \begin{bmatrix} \mathbf{v}_1 \\ ... \\ \mathbf{v}_k \\ \mathbf{e}_1 \\ ... \\ \mathbf{e}_n \end{bmatrix}$, a cui ho aggiunto gli n vettori della base canonica di $\mathbf{K}^n$; A ha $k+n$ righe. Riduco A a scala col MEG. U ha k pivot nella prima metà perché i vettori $\mathbf{v}_1, ..., \mathbf{v}_k$ erano linearmente indipendenti. La parte con la base canonica ha $n-k$ pivot e k righe nulle: gli elementi della base canonica superflui sono k. Gli $n-k$ sono quelli necessari per ottenere una base insieme ai k vettori originari: insieme hanno cardinalità n e sono linearmente indipendenti. Le righe da $k+1$ a n forniscono il completamento, che consiste di $n-k$ vettori riga. Di fatto il problema del completamento a una base è ridotto al problema di estrazione, perché ho aggiunto un insieme di generatori. Poiché lo spazio dello righe si preserva, posso scegliere come base, oltre al completamento, le righe da 1 a k dopo la riduzione a scala, oppure i k vettori riga originari.

Posso fare lo stesso con le colonne, ma alla fine devo scegliere i vettori originari nelle posizioni corrispondenti alle colonne con i pivot.

- Qual è il significato di $col\,(A)$? Prendo un sistema $A\mathbf{x} = \mathbf{b}$ e suppongo $r\,(A) = r\,(A|\mathbf{b})$, cioè che il sistema sia risolubile. Allora esiste una soluzione $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ ... \\ v_n \end{pmatrix} : v_1C_1(A) + v_2C_2(A) + ... + v_nC_n(A) = \mathbf{b}$. Quindi $col(A)$, cioè l'insieme dei termini generici $v_1C_1(A) + v_2C_2(A) + ... + v_nC_n(A)$ al variare di $v_1, ..., v_n$, è l'insieme (lo spazio vettoriale) dei vettori $\mathbf{b}$ per cui il sistema ammette soluzione.

4.4 Operazioni tra sottospazi

Sia $(V, +, \cdot)$ uno spazio vettoriale su $\mathbf{K}$. Siano H_1, H_2 sottospazi vettoriali di V.

1. $V = \mathbf{K}^n$. $H_1 = \{\mathbf{x} \in \mathbf{K}^n : A\mathbf{x} = \mathbf{0}\}$, $H_2 = \{\mathbf{x} \in \mathbf{K}^n : B\mathbf{x} = \mathbf{0}\}$.

E' naturale chiedersi quali sono le caratteristiche geometriche degli spazi $H_1 \cap H_2$ e $H_1 \cup H_2$.

4.4.1 Intersezione

Considero $H_1 \cap H_2$: è un sottospazio vettoriale di V? $\mathbf{0} \in H_1, H_2$. $H_1 \cap H_2$ è chiuso rispetto alle operazioni di V?

Dati $\mathbf{v}, \mathbf{w} \in H_1 \cap H_2$, $\mathbf{v} + \mathbf{w} \in H_1 \cap H_2$? $\mathbf{v}, \mathbf{w} \in H_1$, $\mathbf{v}, \mathbf{w} \in H_2$, quindi, poiché H_1 e H_2 sono sottospazi, $\mathbf{v} + \mathbf{w} \in H_1$ e $\mathbf{v} + \mathbf{w} \in H_2$, quindi $\mathbf{v} + \mathbf{w} \in H_1 \cap H_2$.

Dati $\lambda \in \mathbf{K}$ e $\mathbf{v} \in H_1 \cap H_2$, $\lambda\mathbf{v} \in H_1 \cap H_2$? $\lambda\mathbf{v} \in H_1$, $\lambda\mathbf{v} \in H_2$, quindi $\lambda\mathbf{v} \in H_1 \cap H_2$. Perciò $H_1 \cap H_2$ è un sottospazio vettoriale.

1. Nell'esempio sopra, $H_1 \cap H_2 = \{\mathbf{x} \in \mathbf{K}^n : C\mathbf{x} = \mathbf{0}\}$, dove $C = \begin{bmatrix} A \\ B \end{bmatrix}$.

4.4.2 Somma

Considero $H_1 \cup H_2$: è un sottospazio vettoriale di V? $\mathbf{0} \in H_1, H_2$. $H_1 \cup H_2$ è chiuso rispetto alle operazioni di V?

Dati $\lambda \in \mathbf{K}$, $\mathbf{v} \in H_1 \cup H_2$, $\lambda\mathbf{v} \in H_1 \cup H_2$? $\lambda\mathbf{v} \in H_1$ o $\lambda\mathbf{v} \in H_2$, quindi $\lambda\mathbf{v} \in H_1 \cup H_2$.

Dati $\mathbf{v}, \mathbf{w} \in H_1 \cup H_2$, $\mathbf{v} + \mathbf{w} \in H_1 \cup H_2$? Ci sono quattro casi: $\mathbf{v}, \mathbf{w} \in H_1$, o $\mathbf{v}, \mathbf{w} \in H_2$, o $\mathbf{v} \in H_1$ e $\mathbf{w} \in H_2$, o $\mathbf{v} \in H_2$ e $\mathbf{w} \in H_1$. Nei primi due $\mathbf{v} + \mathbf{w} \in H_1 \cup H_2$; negli ultimi due non lo sappiamo.

Se $V = \mathbb{R}^2$ e considero come sottospazi due rette (la dimensione è 1), sommando due elementi delle rette è possibile ottenere un elemento che non appartiene a nessuna delle due. Quindi in generale $H_1 \cup H_2$ non è un sottospazio vettoriale: è un insieme troppo piccolo per essere chiuso rispetto alle operazioni. Ma allora, qual è il più piccolo sottospazio che contiene $H_1 \cup H_2$?

Se $H_1 = Span\,(\mathbf{v}_1, ..., \mathbf{v}_k)$, $H_2 = Span\,(\mathbf{w}_1, ..., \mathbf{w}_h)$, l'insieme dei vettori che sono combinazione lineare di $\mathbf{v}_1, ..., \mathbf{v}_k$ o di $\mathbf{w}_1, ..., \mathbf{w}_h$ non è un sottospazio: non è chiuso rispetto alla somma. Devo quindi aggiungere tutte le possibile somme: il più piccolo sottospazio che contiene $H_1 \cup H_2$ è $Span\,(\mathbf{v}_1, ..., \mathbf{v}_k, \mathbf{w}_1, ..., \mathbf{w}_h)$.

Infatti, dato $\mathbf{v} \in H_1$, $\mathbf{v} = \lambda_1\mathbf{v}_1 + ... + \lambda_k\mathbf{v}_k$, dato $\mathbf{w} \in H_1$, $\mathbf{w} = \mu_1\mathbf{w}_1 + ... + \mu_h\mathbf{w}_h$, $\mathbf{v} + \mathbf{w} = \lambda_1\mathbf{v}_1 + ... + \lambda_k\mathbf{v}_k + \mu_1\mathbf{w}_1 + ... + \mu_h\mathbf{w}_h$: affinché $\mathbf{w} + \mathbf{v}$ appartenga al sottospazio l'unione va estesa allo span di tutti gli elementi.

Def Si definisce spazio vettoriale somma $H_1 + H_2$ lo $Span\,(\mathbf{v}_1, ..., \mathbf{v}_k, \mathbf{w}_1, ..., \mathbf{w}_h)$.

Quindi $H_1 + H_2$ è l'insieme dei vettori $\mathbf{u}$ che si possono scrivere come $\mathbf{v} + \mathbf{w}$, dove $\mathbf{v} \in H_1$ e $\mathbf{w} \in H_2$.

1. $H_1 \cup H_2$ è un sottospazio se e solo se $H_1 \subseteq H_2$ vel $H_2 \subseteq H_1$.

2. L'intersezione $\bigcap_i S_i$ di un numero finito o infinito di sottospazi di V è un sottospazio di V.

3. Se $S = \{U_0, U_1, ...\}$ è un insieme di sottospazi di V e per ogni coppia di sottospazi $U_i, U_j \in S$ esiste un sottospazio $U_k \in S$ che li contiene entrambi, S si dice filtrante. In tal caso l'unione dei sottospazi U_i è un sottospazio. Se $U_0 \subseteq U_1 \subseteq ...$, S si dice catena e anche in tal caso l'unione dei sottospazi U_i è un sottospazio.

   Se $V = \mathbb{R}[x]$ e S è l'insieme dei sottospazi di V di dimensione finita, S è filtrante.

Teorema (formula di Grassman)

Hp:	$(V, +, \cdot)$ è uno spazio vettoriale su $\mathbf{K}$, H_1, H_2 sono sottospazi di V
Ts:	$\dim(H_1 + H_2) = \dim H_1 + \dim H_2 - \dim(H_1 \cap H_2)$

Dim Sia $B = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ una base di $H_1 \cap H_2$. $\mathbf{v}_1, ..., \mathbf{v}_n$ è un insieme di vettori linearmente indipendenti di $H_1 \cap H_2$, quindi anche di H_1 e H_2 singolarmente. Allora posso completare B a una base di H_1 e posso completare B a una base di H_2. $B_1 = B \cup \{\mathbf{w}_1, ..., \mathbf{w}_p\} = \{\mathbf{v}_1, ..., \mathbf{v}_n\} \cup \{\mathbf{w}_1, ..., \mathbf{w}_p\}$ è una base di H_1, $B_2 = \{\mathbf{v}_1, ..., \mathbf{v}_n\} \cup \{\mathbf{u}_1, ..., \mathbf{u}_q\}$ è una base di H_2. Si sta dicendo che $\dim(H_1 \cap H_2) = n$, $\dim H_1 = n + p$, $\dim H_2 = n + q$. Voglio mostrare che $\dim(H_1 + H_2) = n + p + q$: in particolare, che $\bar{B} = \{\mathbf{v}_1, ..., \mathbf{v}_n\} \cup \{\mathbf{w}_1, ..., \mathbf{w}_p\} \cup \{\mathbf{u}_1, ..., \mathbf{u}_q\}$ è una base di $H_1 + H_2$.

Mostro che tali vettori sono generatori di $H_1 + H_2$. In generale $H_1 + H_2$ è l'insieme dei vettori $\mathbf{v}$ che si possono scrivere come $\mathbf{w} + \mathbf{u}$, dove $\mathbf{w} \in H_1$ e $\mathbf{u} \in H_2$: $\mathbf{w} = \lambda_1\mathbf{v}_1 + ... + \lambda_n\mathbf{v}_n + \lambda_{n+1}\mathbf{w}_1 + ... + \lambda_{n+p}\mathbf{w}_p$, mentre $\mathbf{u} = \mu_1\mathbf{v}_1 + ... + \mu_n\mathbf{v}_n + \mu_{n+1}\mathbf{u}_1 + ... + \mu_{n+q}\mathbf{u}_q$. Dunque $\mathbf{w} + \mathbf{u} = \lambda_1\mathbf{v}_1 + ... + \lambda_n\mathbf{v}_n + \lambda_{n+1}\mathbf{w}_1 + ... + \lambda_{n+p}\mathbf{w}_p + \mu_1\mathbf{v}_1 + ... + \mu_n\mathbf{v}_n + \mu_{n+1}\mathbf{u}_1 + ... + \mu_{n+q}\mathbf{u}_q =$

$(\lambda_1 + \mu_1)\mathbf{v}_1 + ... + (\lambda_n + \mu_n)\mathbf{v}_n + \lambda_{n+1}\mathbf{w}_1 + ... + \lambda_{n+p}\mathbf{w}_p + \mu_{n+1}\mathbf{u}_1 + ... + \mu_{n+q}\mathbf{u}_q$. Quindi $H_1 + H_2$ è l'insieme delle combinazioni lineari dei vettori $\{\mathbf{v}_1, ..., \mathbf{v}_n\} \cup \{\mathbf{w}_1, ..., \mathbf{w}_p\} \cup \{\mathbf{u}_1, ..., \mathbf{u}_q\}$, che dunque generano $H_1 + H_2$.

Mostro che sono linearmente indipendenti. Voglio mostrare che $x_1\mathbf{v}_1 + ... + x_n\mathbf{v}_n + y_1\mathbf{w}_1 + ... + y_p\mathbf{w}_p + z_1\mathbf{u}_1 + ... + z_q\mathbf{u}_q = \mathbf{0}$ implica $x_1 = ... = x_n = y_1 = ... = y_p = z_1 = ... = z_q = 0$. Chiamo $\mathbf{z}$ la somma dei primi $n+p$ elementi (potrei anche prendere il primo e il terzo gruppo; l'idea è sfruttare l'indipendenza lineare dei vettori di una delle basi note): $\mathbf{z} = x_1\mathbf{v}_1 + ... + x_n\mathbf{v}_n + y_1\mathbf{w}_1 + ... + y_p\mathbf{w}_p = -(z_1\mathbf{u}_1 + ... + z_q\mathbf{u}_q)$. Dalla prima equazione si ha che $\mathbf{z} \in H_1$, dalla seconda che $\mathbf{z} \in Span\left(\mathbf{u}_1, ..., \mathbf{u}_q\right) \subseteq H_2$, dunque $\mathbf{z} \in H_1 \cap H_2$. Dato che $B = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ è una base di $H_1 \cap H_2$, posso scrivere $\mathbf{z} = \lambda_1\mathbf{v}_1 + ... + \lambda_n\mathbf{v}_n$. Allora si ottiene $\mathbf{z} = \lambda_1\mathbf{v}_1 + ... + \lambda_n\mathbf{v}_n = x_1\mathbf{v}_1 + ... + x_n\mathbf{v}_n + y_1\mathbf{w}_1 + ... + y_p\mathbf{w}_p$, che è equivalente a $(x_1 - \lambda_1)\mathbf{v}_1 + ... + (x_n - \lambda_n)\mathbf{v}_n + y_1\mathbf{w}_1 + ... + y_p\mathbf{w}_p = \mathbf{0}$. Ma siccome $\mathbf{v}_1, ..., \mathbf{v}_n, \mathbf{w}_1, ..., \mathbf{w}_p$ sono linearmente indipendenti perché sono una base di H_1, si ha $x_1 = \lambda_1, ..., x_n = \lambda_n, y_1 = 0, ..., y_p = 0$. Quindi nell'equazione $x_1\mathbf{v}_1 + ... + x_n\mathbf{v}_n + y_1\mathbf{w}_1 + ... + y_p\mathbf{w}_p + z_1\mathbf{u}_1 + ... + z_q\mathbf{u}_q = \mathbf{0}$ il secondo gruppo si annulla e si ottiene $x_1\mathbf{v}_1 + ... + x_n\mathbf{v}_n + z_1\mathbf{u}_1 + ... + z_q\mathbf{u}_q = \mathbf{0}$. Ma $\mathbf{v}_1, ..., \mathbf{v}_n, \mathbf{u}_1, ..., \mathbf{u}_q$ sono una base di H_2, quindi sono linearmente indipendenti e si ha $x_1 = ... = x_n = z_1 = ... = z_q = 0$. Dunque $\{\mathbf{v}_1, ..., \mathbf{v}_n\} \cup \left\{\mathbf{w}_1, ..., \mathbf{w}_p\right\} \cup \left\{\mathbf{u}_1, ..., \mathbf{u}_q\right\}$ è un insieme di vettori linearmente indipendenti. ■

Nel caso $H_1 \cap H_2 = \{\mathbf{0}\}$, $\dim(H_1 \cap H_2) = 0$ (non c'è alcuna "informazione" in comune). La formula di Grassman diventa allora $\dim(H_1 + H_2) = \dim H_1 + \dim H_2$.

Def Se $H_1 \cap H_2 = \{\mathbf{0}\}$, la somma $H_1 + H_2$ si dice somma diretta tra H_1 e H_2 e si indica con $H_1 \oplus H_2$.

Teo

$$\begin{aligned} \text{Hp} &: \quad H_1, H_2 \text{ sono sottospazi di } V,\ B_1, B_2 \text{ sono basi rispettivamente di } H_1 \text{ e } H_2 \\ \text{Ts} &: \quad H_1 \oplus H_2 = V \iff B = B_1 \cup B_2 \text{ è una base di } V \end{aligned}$$

Nel caso in cui $H_1 \oplus H_2 = V$, se B_1 è una base di H_1 e B_2 è una base di H_2, $B = B_1 \cup B_2$ è una base di V. Infatti ogni vettore di V si scrive in modo unico come somma di un vettore di H_1 e un vettore di H_2: i vettori di B_1 e i vettori di B_2 sono linearmente indipendenti, perché $H_1 \cap H_2 = \{\mathbf{0}\}$.

In generale, se $V = A \oplus B = A \oplus C$, non si può concludere che $B = C$: infatti, B_B e B_C forniscono entrambi un completamento di B_A a una base di V, ma non necessariamente generano lo stesso sottospazio. E. g., $\mathbb{R}^3$ è somma diretta di un piano e di una qualsiasi retta non appartenente al piano, ma ci sono infinite rette distinte non appartenenti al piano.

4.5 Applicazioni lineari

Come sono fatte le funzioni che si "comportano bene" rispetto alle proprietà caratterizzanti uno spazio vettoriale?

Def Siano V e W spazi vettoriali su un campo $\mathbf{K}$. Una funzione $L : V \to W$ si dice applicazione lineare se

(i) ha la proprietà di additività, i. e. $\forall\ \mathbf{v}_1, \mathbf{v}_2 \in V,\ L(\mathbf{v}_1 + \mathbf{v}_2) = L(\mathbf{v}_1) + L(\mathbf{v}_2)$;

(ii) ha la proprietà di omogeneità, i. e. $\forall\ \mathbf{v} \in V,\ \forall\ \lambda \in \mathbf{K},\ L(\lambda \mathbf{v}) = \lambda L(\mathbf{v})$.

L quindi "trasporta" la somma da V in W e il prodotto per scalare da V in W.

1. $V = \mathbf{K}^n$, $W = \mathbf{K}^m$. Data $A \in M_{\mathbf{K}}(m, n)$, si definisce ${}_A : \mathbf{K}^n \to \mathbf{K}^m$ la funzione che a ogni vettore colonna $\mathbf{v} \in \mathbf{K}^n$ associa il vettore colonna di $\mathbf{K}^m$ che si ottiene calcolando $A\mathbf{v}$. ${}_A$ è un'applicazione lineare, infatti ${}_A(\mathbf{v}_1 + \mathbf{v}_2) = A(\mathbf{v}_1 + \mathbf{v}_2) = A\mathbf{v}_1 + A\mathbf{v}_2 = {}_A(\mathbf{v}_1) + {}_A(\mathbf{v}_2)$, per la proprietà distributiva a destra del prodotto matriciale, e ${}_A(\lambda \mathbf{v}) = A(\lambda \mathbf{v}) = \lambda (A\mathbf{v}) = \lambda {}_A(\mathbf{v})$.

   Date le coordinate di $\mathbf{x}$, una funzione è del tipo ${}_A$ se le coordinate del vettore trasformato sono polinomi omogenei di primo grado nelle coordinate di $\mathbf{x}$.

2. $V = M_{\mathbf{K}}(m, n)$, $W = M_{\mathbf{K}}(n, m)$. Si definisce la funzione *Trasposizione* $: V \to W$ che a ogni matrice $A = \begin{bmatrix} R_1(A) \\ \dots \\ R_m(A) \end{bmatrix}$ associa la matrice trasposta $A^T = \begin{bmatrix} R_1(A) | & \dots & | R_m(A) \end{bmatrix}$, dove la prima colonna è la prima riga di A e così via. E. g. se $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$, $A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$. La trasposizione è un'applicazione lineare.

L è un'applicazione lineare se e solo se $\forall\ \mathbf{v}_1, \mathbf{v}_2 \in V,\ \forall\ \lambda_1, \lambda_2 \in \mathbf{K}\ L(\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2) = \lambda_1 L(\mathbf{v}_1) + \lambda_2 L(\mathbf{v}_2)$: si incorporano le due proprietà in una sola, che può essere ricondotta ad esse considerando rispettivamente $\lambda_1 = \lambda_2 = 1$ e $\lambda_2 = 0$.

Le applicazioni lineari si "comportano bene" rispetto a composizione e inversione.

Proposizione

Hp	:	V, W, U sono spazi vettoriali su $\mathbf{K}$; $L : V \to W, M : W \to U$ sono due applicazioni lineari
Ts	:	$M \circ L : V \to U$ è un'applicazione lineare

Dim Uso il criterio visto. $\forall\ \mathbf{v}_1, \mathbf{v}_2 \in V,\ \forall\ \lambda_1, \lambda_2 \in \mathbf{K},\ M \circ L(\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2) = M(L(\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2)) = M(\lambda_1 L(\mathbf{v}_1) + \lambda_2 L(\mathbf{v}_2))$ perché L è un'applicazione lineare. Inoltre $M(\lambda_1 L(\mathbf{v}_1) + \lambda_2 L(\mathbf{v}_2)) = \lambda_1 M(L(\mathbf{v}_1)) + \lambda_2 M(L(\mathbf{v}_2))$ perché M è un'applicazione lineare e si ottiene quindi $\lambda_1 M \circ L(\mathbf{v}_1) + \lambda_2 M \circ L(\mathbf{v}_2)$. ■

1. $V = \mathbf{K}^n$, $W = \mathbf{K}^m$, $U = \mathbf{K}^p$. Considero $A \in M_{\mathbf{K}}(m,n)$, $B \in M_{\mathbf{K}}(p,m)$. $\left({}_B \circ {}_A\right)(\mathbf{v}) = {}_B\left({}_A(\mathbf{v})\right) = {}_B(A\mathbf{v}) = B(A\mathbf{v}) = (BA)\mathbf{v}$. In generale la composizione di applicazioni lineari del tipo $\left({}_B \circ {}_A\right)(\mathbf{v})$ cattura il prodotto matriciale.

Proposizione

Hp : V, W sono spazi vettoriali su $\mathbf{K}$; $L : V \to W$ è un'applicazione lineare invertibile

Ts : $L^{-1} : W \to V$ è un'applicazione lineare

Nelle ipotesi un'applicazione lineare è considerata invertibile se lo è in quanto funzione, quindi se è iniettiva e suriettiva.

Dim Devo mostrare che $\forall\ \mathbf{w}_1, \mathbf{w}_2 \in W$, $\forall\ \lambda_1, \lambda_2 \in \mathbf{K}$, $L^{-1}(\lambda_1 \mathbf{w}_1 + \lambda_2 \mathbf{w}_2) = \lambda_1 L^{-1}(\mathbf{w}_1) + \lambda_2 L^{-1}(\mathbf{w}_2)$. Per ipotesi di invertibilità esistono e sono unici $\mathbf{v}_1, \mathbf{v}_2 \in V : L(\mathbf{v}_1) = \mathbf{w}_1$ e $L(\mathbf{v}_2) = \mathbf{w}_2$: $\mathbf{v}_1 = L^{-1}(\mathbf{w}_1)$, $\mathbf{v}_2 = L^{-1}(\mathbf{w}_2)$. Allora, usando il criterio nel verso meno consueto, $L^{-1}(\lambda_1 L(\mathbf{v}_1) + \lambda_2 L(\mathbf{v}_2)) = L^{-1}(L(\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2)) = \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 = \lambda_1 L^{-1}(\mathbf{w}_1) + \lambda_2 L^{-1}(\mathbf{w}_2)$. ■

1. $V = \mathbf{K}^n$, $W = \mathbf{K}^m$. Considero $A \in M_{\mathbf{K}}(m,n)$, ${}_A : \mathbf{K}^n \to \mathbf{K}^m$. Quando esiste ${}_A^{-1}$? Voglio ${}_A \circ {}_A^{-1} = id_{\mathbf{K}^m}$ e ${}_A^{-1} \circ {}_A = id_{\mathbf{K}^n}$. Se $m = n$ e A è invertibile, allora ${}_A^{-1} = {}_{A^{-1}}$. Infatti ${}_{A^{-1}}\left({}_A(\mathbf{v})\right) = {}_{A^{-1}}(A\mathbf{v}) = A^{-1}(A\mathbf{v}) = Id_n \mathbf{v} = \mathbf{v}$ e analogamente ${}_A^{-1}\left({}_{A^{-1}}(\mathbf{v})\right) = \mathbf{v}$. Quindi ${}_A^{-1} = {}_{A^{-1}}$ e ${}_A^{-1}$ esiste solo quando $m = n$ e A è invertibile.

Si fanno altri esempi generali di applicazioni lineari.

1. $V = C^1(\mathbb{R})$, $W = C^0(\mathbb{R})$, $L : V \to W$ che associa ad ogni funzione di V la sua derivata.

2. $V = C^0(\mathbb{R})$, $W = C^1(\mathbb{R})$, $\Sigma : V \to W$ che associa ad ogni funzione $g(x)$ di V la sua funzione integrale $\int_0^x g(t)\,dt$.

3. $V = C^k(\mathbb{R})$, $W = C^0(\mathbb{R})$, $L : V \to W$ che associa ad ogni funzione di V una combinazione lineare delle sue derivate: $L(y) = y^{(k)} + a_1 y^{(k-1)} + ... + a_{k-1} y' + a_k y$, con $a_i \in \mathbf{K}$.

Ci sono delle conseguenze notevoli delle proprietà della applicazioni lineari.

- $L(\mathbf{v}_1 - \mathbf{v}_2) = L(\mathbf{v}_1 + (-1)\mathbf{v}_2) = L(\mathbf{v}_1) + L((-1)\mathbf{v}_2) = L(\mathbf{v}_1) - L(\mathbf{v}_2)$ perché L è un'applicazione lineare.

- $L(\mathbf{0}_V) = L(\mathbf{v} - \mathbf{v}) = L(\mathbf{v}) - L(\mathbf{v}) = \mathbf{0}_W$. Che $L(\mathbf{0}_V) = \mathbf{0}_W$ è necessario affinché L sia un'applicazione lineare: se $L(\mathbf{0}_V) \neq \mathbf{0}_W$, L non è un'applicazione lineare.

Come si comporta un'applicazione lineare rispetto ai sottospazi di dominio e codominio?

Proposizione

$$\begin{aligned}
\text{(i) Hp} &: \quad L : V \to W \text{ è un'applicazione lineare; } H \subset V \text{ è un sottospazio vettoriale} \\
\text{Ts} &: \quad L(H) = \{\mathbf{w} \in W : \exists\ \mathbf{v} \in H : L(\mathbf{v}) = \mathbf{w}\} \text{ è un sottospazio vettoriale di } W \\
\text{(ii) Hp} &: \quad L : V \to W \text{ è un'applicazione lineare; } K \subset W \text{ è un sottospazio vettoriale} \\
\text{Ts} &: \quad L^{-1}(K) = \{\mathbf{v} \in V : L(\mathbf{v}) \in \mathbf{K}\} \text{ è un sottospazio vettoriale di } V
\end{aligned}$$

Quindi un'applicazione lineare preserva le proprietà dello spazio vettoriale. $L^{-1}(K)$ denota un insieme di vettori, non ha niente a che fare con l'applicazione lineare inversa.

Dim (i) Devo verificare che $L(H)$ è chiuso rispetto alle operazioni di somma e prodotto per scalare. Dati $\mathbf{w}_1, \mathbf{w}_2 \in L(H)$, esistono $\mathbf{v}_1, \mathbf{v}_2 \in H : L(\mathbf{v}_1) = \mathbf{w}_1$ e $L(\mathbf{v}_2) = \mathbf{w}_2$. Ma $\mathbf{w}_1 + \mathbf{w}_2 = L(\mathbf{v}_1) + L(\mathbf{v}_2) = L(\mathbf{v}_1 + \mathbf{v}_2)$. Poiché H è un sottospazio, $\mathbf{v}_1 + \mathbf{v}_2 \in H$, dunque $L(\mathbf{v}_1 + \mathbf{v}_2) \in L(H)$. Dato $\mathbf{w} \in L(H)$, $\exists\ \mathbf{v} \in H : L(\mathbf{v}) = \mathbf{w}$. Allora $\lambda\mathbf{w} = \lambda L(\mathbf{v}) = L(\lambda\mathbf{v})$: poiché H è un sottospazio, $\lambda\mathbf{v} \in H$ e $\lambda\mathbf{w} \in L(H)$.

(ii) Devo verificare che $L^{-1}(K)$ è chiuso rispetto alle operazioni di somma e prodotto per scalare. Dati $\mathbf{v}_1, \mathbf{v}_2 \in L^{-1}(K)$, $L(\mathbf{v}_1) \in K$ e $L(\mathbf{v}_2) \in K$. Ma $L(\mathbf{v}_1) + L(\mathbf{v}_2) = L(\mathbf{v}_1 + \mathbf{v}_2)$, poiché K è un sottospazio, appartiene a K, dunque $\mathbf{v}_1 + \mathbf{v}_2 \in L^{-1}(K)$. Dato $\mathbf{v} \in L^{-1}(K)$, $L(\mathbf{v}) \in K$. Allora $\lambda L(\mathbf{v}) = L(\lambda\mathbf{v})$, poiché K è un sottospazio, appartiene a K, dunque $\lambda\mathbf{v} \in L^{-1}(K)$. ■

1. Considero $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$, $_A : \mathbb{R}^3 \to \mathbb{R}^3$. Considero $H = Span\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right) \subset \mathbb{R}^3$. Cos'è $_A(H)$? Un generico elemento di H è $\mathbf{v} = \alpha \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$. Un generico elemento di $_A(H)$ è $_A(\mathbf{v}) =_A$

$$\left(\alpha \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right) = \alpha_A \left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) + \beta_A \left(\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right) = \alpha A \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \beta A \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 4 \end{bmatrix} + \beta \begin{bmatrix} 5 \\ 11 \end{bmatrix}.$$

Dunque $_A(H) = Span\left(\begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 5 \\ 11 \end{bmatrix}\right) = \mathbb{R}^2$, perché tali vettori sono due e sono linearmente indipendenti.

Considero $K = Span\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \subset \mathbb{R}^2$. ${}_A^{-1}(K) = \left\{\mathbf{x} \in \mathbb{R}^3 : A\mathbf{x} = \lambda \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\} = \left\{\mathbf{x} \in \mathbb{R}^3 : A\mathbf{x} = \begin{bmatrix} 0 \\ \lambda \end{bmatrix}\right\}$ e risolvo il sistema lineare $\left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 4 & 5 & 6 & \lambda \end{array}\right]$: la soluzione è un'espressione generica dell'elemento del sottospazio.

2. Data $L : V \to W$ e Z sottospazio di W, qual è $\dim\left(L^{-1}(Z)\right)$? Non si può dire molto in generale, perché, anche conoscendo Z, $\dim\left(L^{-1}(Z)\right)$ dipende dall'intersezione tra $Im(L)$ e Z. Se infatti l'unica intersezione tra $Im(L)$ e Z è il vettore nullo (l'intersezione non può essere vuota perché sono entrambi sottospazi), allora $L^{-1}(Z) = \ker(L)$ e $\dim\left(L^{-1}(Z)\right) = \dim(V) - \dim(Im(L))$. $\dim(Z)$ può essere qualsiasi.

3. Se $L = A\mathbf{x}$ e cerco $L^{-1}(V)$, posso scrivere le equazioni cartesiane di V, rappresentate da una matrice B: allora $\mathbf{y} \in V \iff B\mathbf{y} = \mathbf{0}$. Si ha quindi che $\mathbf{x} \in L^{-1}(V) \iff L(\mathbf{x}) \in V \iff A\mathbf{x} \in V \iff B(A\mathbf{x}) = 0$. Trovare $L^{-1}(V)$ significa quindi trovare $\ker(BA)$.

Ci sono poi alcuni casi notevoli generali.

Se $H = V$ invece che $H \subset V$, $L(H) = L(V) = ImL$.

Def Si definisce $ImL = L(V)$ l'immagine dell'applicazione lineare L.

Se L è ${}_A$, $ImL = col(A)$. Quindi, essendo $\dim(col(A)) = r(A)$, si ha $\dim(col(A)) \leq \min\{m, n\}$, in particolare $\dim(col(A)) \leq n$, che è la dimensione del dominio.

Considerato l'insieme $\{\mathbf{0}_W\} \subseteq W$, $L^{-1}(\{\mathbf{0}_W\}) = L^{-1}(\mathbf{0}_W) = \ker L$.

Def Si definisce nucleo dell'applicazione lineare L l'insieme $\ker L = L^{-1}(\{\mathbf{0}_W\})$.

Data ${}_A : \mathbf{K}^n \to \mathbf{K}^m$, $A \in M_{\mathbf{K}}(m, n)$, $\ker_A = {}_A^{-1}(\{\mathbf{0}_W\}) = \{\mathbf{x} \in \mathbf{K}^n : {}_A(\mathbf{x}) = \mathbf{0}_W\} = \{\mathbf{x} \in \mathbf{K}^n : A\mathbf{x} = \mathbf{0}\} = \ker A$.

In generale $\ker_A = \ker A$.

Se V, W sono spazi vettoriali su $\mathbf{K}$ e $L : V \to W$ è un'applicazione lineare, dato $\mathbf{w} \in W$, $L^{-1}(\mathbf{w}) = \{\mathbf{v} \in V : L(\mathbf{v}) = \mathbf{w}\}$. Nel caso di ${}_A : \mathbf{K}^n \to \mathbf{K}^m$, ${}_A(\mathbf{x}) = A\mathbf{x}$ con $A \in M_{\mathbf{K}}(m, n)$, $\mathbf{b} \in \mathbf{K}^m$, ${}_A^{-1}(\mathbf{b}) = \{\mathbf{v} \in \mathbf{K}^n : L(\mathbf{v}) = \mathbf{b}\} = \{\mathbf{v} \in \mathbf{K}^n : A\mathbf{v} = \mathbf{b}\}$: trovare la controimmagine di $\mathbf{b}$ attraverso ${}_A$ significa risolvere il sistema lineare $A\mathbf{v} = \mathbf{b}$.

Teorema di fibra

Hp: $L : V \to W$ è un'applicazione lineare, $\mathbf{w} \in W$, $\mathbf{v}_0 \in L^{-1}(\mathbf{w})$

Ts: $L^{-1}(\mathbf{w}) = \{\mathbf{v}_0 + \mathbf{v}_h$ al variare di $\mathbf{v}_h \in \ker(L)\}$

cioè tutte e sole le controimmagini di $\mathbf{w}$ si ottengono sommando a $\mathbf{v}_0$ un elemento del nucleo.

Dim E' analoga a quella del teorema di struttura. Mostro che se $\mathbf{v}$ è tale che $L(\mathbf{v}) = \mathbf{w}$, allora $\mathbf{v} - \mathbf{v}_0 \in \ker(L)$: infatti $L(\mathbf{v} - \mathbf{v}_0) = L(\mathbf{v}) - L(\mathbf{v}_0) = \mathbf{w} - \mathbf{w} = \mathbf{0}$. Mostro che $\mathbf{v}_0+\mathbf{v}_h$ è tale che $L(\mathbf{v}_0+\mathbf{v}_h) = \mathbf{w}$: infatti $L(\mathbf{v}_0+\mathbf{v}_h) = L(\mathbf{v}_0) + L(\mathbf{v}_h) = \mathbf{w} + \mathbf{0} = \mathbf{w}$. ■

$L^{-1}(\mathbf{w})$ si dice fibra di L sopra $\mathbf{w}$.

1. Dato F^1 insieme delle funzioni derivabili su $\mathbb{R}$, $D :^1\to$ è $D(f) = f'$. Allora, se $D(f) = g$, $D^{-1}(g) = G + c$, dove G è una soluzione particolare (cioè è una funzione tale che $D(G) = g$) e c è il generico elemento di $\ker(D)$, cioè l'insieme delle funzioni che hanno derivata nulla (quindi le funzioni costanti).

4.5.1 Iniettività, suriettività e invertibilità

Una funzione (quindi anche un'applicazione lineare) è invertibile se e solo se è iniettiva e suriettiva.

(i) *Suriettività* Un'applicazione lineare $L : V \to W$ è suriettiva se $\forall\ \mathbf{w} \in W\ \exists\ \mathbf{v} \in V : L(\mathbf{v}) = \mathbf{w}$, cioè se $\forall\ \mathbf{w} \in W\ L^{-1}(\mathbf{w}) \neq \varnothing$, cioè posso risolvere il sistema lineare $A\mathbf{x} = \mathbf{b}\ \forall\ \mathbf{b} \in \mathbf{K}^m$. Questo significa che $Im(L) = W$.

Proposizione

Hp: $L : V \to W$ è un'applicazione lineare; $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ è un insieme di generatori di V

Ts: $\{L(\mathbf{v}_1), ..., L(\mathbf{v}_n)\}$ è un insieme di generatori di ImL, i. e. $Span\{L(\mathbf{v}_1), ..., L(\mathbf{v}_n)\} = ImL$

Dim $\forall\ \mathbf{w} \in Im(L)$, $\exists\ \mathbf{v} \in V : L(\mathbf{v}) = \mathbf{w}$. Per ipotesi $V = Span(\mathbf{v}_1, ..., \mathbf{v}_n)$ e in particolare $\mathbf{v} = \lambda_1\mathbf{v}_1 + ... + \lambda_n\mathbf{v}_n$. Applico L a entrambi i lati: ottengo $L(\lambda_1\mathbf{v}_1 + ... + \lambda_n\mathbf{v}_n) = \lambda_1 L(\mathbf{v}_1) + ... + \lambda_n L(\mathbf{v}_n) = \mathbf{w}$, perché L è un'applicazione lineare. Dunque $\mathbf{w} \in Span\{L(\mathbf{v}_1), ..., L(\mathbf{v}_n)\}$, cioè ogni elemento dell'immagine può essere scritto come combinazione lineare di $L(\mathbf{v}_1), ..., L(\mathbf{v}_n)$, e $ImL \subseteq Span\{L(\mathbf{v}_1), ..., L(\mathbf{v}_n)\}$. Va dimostrata anche l'inclusione opposta, che è ovvia perché lo span di un insieme di vettori è un sottospazio. ■

Ci sono alcune conseguenze:

- $L : V \to W$ è un'applicazione lineare suriettiva se e solo $Im(L) = W$, cioè, applicando la proposizione, se l'immagine di un insieme di generatori di V è un insieme di generatori di W;

- la suriettività conserva la proprietà di essere generatori di tutto lo spazio

1. Considero $_A : \mathbf{K}^n\to \mathbf{K}^m$, con $A \in M_{\mathbf{K}}(m, n)$. $_A$ è suriettiva se e solo se $A\mathbf{x} = \mathbf{b}$ è risolubile per ogni $\mathbf{b} \in \mathbf{K}^n$. Per il teorema di Rouché-Capelli questo è possibile se e solo se $r(A) = m$ (infatti, per il teorema di nullità più rango, $\dim(Im(_A)) = r$ dev'essere m affinché sia suriettiva): in tal caso si ha per ogni $\mathbf{b}$ $r(A) = r(A|\mathbf{b})$.

Poiché $r(A) \leq \min\{m,n\}$, affinché $_A$ sia suriettiva occorre $m \leq n$. Se $n < m$ $_A$ non è suriettiva. Infatti, per nullità più rango, $\dim(Im(_A)) = n - \dim(\ker(_A)) \leq n$: se $m > n$ $\dim(Im(_A))$ non può essere m.

Prendendo la base canonica di $\mathbf{K}^n$ $B = \left\{\mathbf{e}^1, ..., \mathbf{e}^n\right\}$, si ha che B è un insieme di generatori di $\mathbf{K}^n$, quindi si applica la proposizione e $Im(_A) = Span\left\{_A\left(\mathbf{e}^1\right), ..., _A(\mathbf{e}^n)\right\} = Span\{C_1(A), ..., C_n(A)\} = col(A)$. $_A$ è suriettiva se e solo se $col(A) = \mathbf{K}^m$, cioè le colonne di A generano $\mathbf{K}^m$.

L, descritta dalla matrice A, è suriettiva se e solo se è invertibile a destra. Infatti, se è invertibile a destra, $\exists$ $B : AB = Id_m$: questo significa che $\begin{bmatrix} R_1(A) \\ ... \\ R_m(A) \end{bmatrix} B = \begin{bmatrix} R_1(A)B \\ ... \\ R_m(A)B \end{bmatrix}$. Dal momento che le m righe del risultato sono linearmente indipendenti, anche le righe di A sono indipendenti e $r(A) = m$.

(ii) *Iniettività* Un'applicazione lineare $L : V \to W$ è iniettiva se $L(\mathbf{v}_1) = L(\mathbf{v}_2)$ implica $\mathbf{v}_1 = \mathbf{v}_2$, o equivalentemente se per ogni $\mathbf{v}_1, \mathbf{v}_2 \in \mathbf{K}^n$, $\mathbf{v}_1 \neq \mathbf{v}_2$, si ha $L(\mathbf{v}_1) \neq L(\mathbf{v}_2)$.

Teorema: criterio di iniettività

Hp: $L : V \to W$ è un'applicazione lineare

Ts: L è iniettiva $\iff$ $\ker(L) = \{0_V\}$

Dim Mostro che L iniettiva implica $\ker(L) = \{0_V\}$. Poiché $\ker(L)$ è un sottospazio vettoriale, $0_V \in \ker(L)$. Ma L è iniettiva, quindi $L^{-1}(\{0_W\})$ contiene un solo elemento, che è pertanto 0_V, e $\ker(L) = \{0_V\}$.

Mostro che $\ker(L) = \{0_V\}$ implica L iniettiva: uso la prima definizione di iniettività. Considero $\mathbf{v}_1, \mathbf{v}_2 \in V : L(\mathbf{v}_1) = L(\mathbf{v}_2)$. Equivalentemente $L(\mathbf{v}_1) - L(\mathbf{v}_2) = \mathbf{0} \iff L(\mathbf{v}_1 - \mathbf{v}_2) = \mathbf{0}$ perché L è lineare, ma allora $\mathbf{v}_1 - \mathbf{v}_2 \in \ker(L)$, che però contiene solo $\{0_V\}$ per ipotesi, dunque $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{0} \iff \mathbf{v}_1 = \mathbf{v}_2$. ■

1. Considero $_A : \mathbf{K}^n \to \mathbf{K}^m$, con $A \in M_{\mathbf{K}}(m,n)$. $_A$ è iniettiva se e solo se $\ker(L) = \{0_V\}$. Per il teorema di Rouché-Capelli questo è possibile se e solo se $r(A) = n$ (il sistema omogeneo ha un'unica soluzione). Poiché $r(A) \leq \min\{m,n\}$, affinché $_A$ sia iniettiva occorre $n \leq m$. Se $m < n$ $_A$ non è iniettiva.

Riepilogando, affinché $_A$ sia invertibile occorre che sia suriettiva e iniettiva, per le quali proprietà è necessario rispettivamente $m \leq n$ e $n \leq m$, quindi $m = n$; se $m \neq n$ $_A$ non è invertibile.

1. Se $\dim V = \dim W = n$, L è iniettiva se e solo se è suriettiva: infatti $\dim(\ker(L)) = 0$ se e solo se $\dim(ImL) = n$, per il teorema di nullità più rango. Dunque è sufficiente verificare una delle due condizioni per stabilire se L è biunivoca.

Corollario

Hp : $L : V \to W$ è un'applicazione lineare iniettiva; $\mathbf{v}_1, ..., \mathbf{v}_n$ sono vettori linearmente indipendenti

Ts : $L(\mathbf{v}_1), ..., L(\mathbf{v}_n)$ sono vettori linearmente indipendenti

Dim Mostro che $\lambda_1 L(\mathbf{v}_1) + ... + \lambda_n L(\mathbf{v}_n) = \mathbf{0}_W$ implica $\lambda_1 = ... = \lambda_n = 0$. Per linearità tale equazione è equivalente a $L(\lambda_1 \mathbf{v}_1 + ... + \lambda_n \mathbf{v}_n) = 0_W$, quindi $\lambda_1 \mathbf{v}_1 + ... + \lambda_n \mathbf{v}_n \in \ker(L)$. Poiché L è iniettiva, $\ker(L) = \{0_V\}$ e $\lambda_1 \mathbf{v}_1 + ... + \lambda_n \mathbf{v}_n = 0_V$, e poiché tali vettori sono linearmente indipendenti $\lambda_1 = ... = \lambda_n = 0$. ■

Quindi l'iniettività preserva l'indipendenza lineare.

Non vale l'implicazione in assenza di iniettività: si consideri l'applicazione lineare $L(\mathbf{v}) = \mathbf{0}$.

1. Se $L(\mathbf{v}_1), ..., L(\mathbf{v}_n)$ sono vettori linearmente indipendenti, allora $\mathbf{v}_1, ..., \mathbf{v}_n$ sono vettori linearmente indipendenti, anche se L non è iniettiva. Infatti, se per assurdo $\mathbf{v}_1, ..., \mathbf{v}_n$ fossero linearmente dipendenti, si avrebbe $x_1\mathbf{v}_1 + ... + x_n\mathbf{v}_n = \mathbf{0}$ con i coefficienti non tutti nulli e $L(x_1\mathbf{v}_1 + ... + x_n\mathbf{v}_n) = L(\mathbf{0})$. Ma allora per linearità si otterrebbe $x_1 L(\mathbf{v}_1) + ... + x_n L(\mathbf{v}_n) = \mathbf{0}$ con coefficienti non tutti nulli, cioè $L(\mathbf{v}_1), ..., L(\mathbf{v}_n)$ sarebbero linearmente dipendenti,b b contro l'ipotesi.

2. Se $L : V \to W$ è un'applicazione lineare, considero una base B_1 di $\ker L$ e la completo a una base di V, usando B_2 (che genera il sottospazio $U \subseteq V$) come completamento. Allora $U \cap \ker L = \{\mathbf{0}\}$ perché $B_1 \cup B_2$ è un insieme di vettori linearmente indipendenti. Se per assurdo nell'intersezione ci fosse altro, esisterebbe $\mathbf{v} \neq \mathbf{0}$ che si può scrivere come combinazione lineare dei vettori di entrambe le basi, e portando dall'altro lato si avrebbe che si può ottenere il vettore nullo con coefficienti non tutti nulli, per cui i vettori non sarebbero linearmente indipendenti. Anche intuitivamente, se $U \cap \ker L$ non contenesse solo il vettore nullo significherebbe che B_2 non è un vero completamento, ma contiene dei vettori "inutili".

3. L, descritta dalla matrice A, è iniettiva se e solo se è invertibile a sinistra. Infatti, se è invertibile a sinistra, $\exists$ $B : BA = Id_n$: questo significa che $B[C_1(A)|...|C_n(A)] = [M(C_1(A))|...|M(C_n(A))]$. Dal momento che le n colonne del risultato sono linearmente indipendenti, anche le colonne di A sono indipendenti e $r(A) = n$.

Corollario

Hp: $L : V \to W$ è un'applicazione lineare iniettiva; $H \subset V$ è un sottospazio vettoriale

Ts : $\dim(L(H)) = \dim(H)$

Dim $\dim(H) = h$, cioè $B = \{\mathbf{v}_1, ..., \mathbf{v}_h\}$ è una base di H. Mostro che $L(B) = \{L(\mathbf{v}_1), ..., L(\mathbf{v}_h)\}$ è una base di $L(H)$. Per il precedente corollario $L(\mathbf{v}_1), ..., L(\mathbf{v}_h)$ sono vettori linearmente indipendenti. Per costruzione (per la proposizione??) sono generatori di $L(H)$: $\forall\ \mathbf{w} \in L(H)$, $\exists\ \mathbf{v} \in H : L(\mathbf{v}) = \mathbf{w}$. Per ipotesi $H = Span(\mathbf{v}_1, ..., \mathbf{v}_h)$ e in particolare $\mathbf{v} = \lambda_1\mathbf{v}_1 + ... + \lambda_h\mathbf{v}_h$. Applico L a entrambi i lati: ottengo $L(\lambda_1\mathbf{v}_1 + ... + \lambda_h\mathbf{v}_h) = \lambda_1 L(\mathbf{v}_1) + ... + \lambda_h L(\mathbf{v}_h) = \mathbf{w}$, perché L è un'applicazione lineare. Dunque $\mathbf{w} \in Span\{L(\mathbf{v}_1), ..., L(\mathbf{v}_h)\}$, cioè ogni elemento di $L(H)$ può essere scritto come combinazione lineare di $L(\mathbf{v}_1), ..., L(\mathbf{v}_h)$, che costituiscono dunque una base. ∎

Ci sono alcune conseguenze:

- $H \subseteq V$ implica $\dim(H) \leq \dim(V)$. Infatti, se considero una base di H $B = \{\mathbf{v}_1, ..., \mathbf{v}_h\}$ e l'applicazione lineare identità $id : H \to V$, che è iniettiva, per il primo corollario $\mathbf{v}_1 = id(\mathbf{v}_1)$,..., $\mathbf{v}_h = id(\mathbf{v}_h)$ sono linearmente indipendenti in V.(corollario 2?) Allora $h = \dim(H)$ è minore o uguale del massimo numero di vettori linearmente indipendenti in V, che è $\dim(V)$.

- Se $H \subseteq V$, $\dim(H) = \dim(V)$ implica $H = V$ (l'implicazione opposta è ovvia). Infatti si ha $\dim(H) \leq \dim(V)$, ma se per assurdo la relazione fosse stretta, si potrebbe completare una base di H a una base di V, e la dimensione cambierebbe. In alternativa si può considerare che, dato $\dim H = \dim V = n$, una base di H è formata da n elementi linearmente indipendenti sia in H che in V, essendo $H \subseteq V$, quindi è anche una base di V, ma se due spazi vettoriali hanno una base in comune coincidono.

4.5.2 Rappresentazione di un'applicazione lineare

Come definire un'applicazione lineare?

Proposizione

Hp: V, W sono spazi vettoriali; $B = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ è una base di V;

$S = \{\mathbf{w}_1, ..., \mathbf{w}_n\}$ è un insieme di vettori qualsiasi di W

Ts: esiste un'unica applicazione lineare $L : V \to W : L(\mathbf{v}_1) = \mathbf{w}_1, ..., L(\mathbf{v}_n) = \mathbf{w}_n$

Dim Se esiste $L : V \to W$ con le proprietà richieste, allora $\forall\ \mathbf{v} \in V$ $\mathbf{v} = \lambda_1\mathbf{v}_1 + ... + \lambda_n\mathbf{v}_n$ e $L(\mathbf{v}) = \lambda_1 L(\mathbf{v}_1) + ... + \lambda_n L(\mathbf{v}_n) = \lambda_1\mathbf{w}_1 + ... + \lambda_n\mathbf{w}_n \in W$. Rimane da verificare che una funzione definita in questo modo è additiva e omogenea (??). ∎

1. $V = \mathbf{K}^n$, $W = \mathbf{K}^m$. So già costruire applicazioni lineari $L : \mathbf{K}^n \to \mathbf{K}^m$ usando il prodotto matriciale ($L =_A$). Posso costruirne altre? Uso la proposizione, considerando base di $\mathbf{K}^n$ $B = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$. Scelgo le immagini

dei vettori di B $S = \{\mathbf{w}_1, ..., \mathbf{w}_n\}$: $L(\mathbf{e}_1) = \mathbf{w}_1, ..., L(\mathbf{e}_n) = \mathbf{w}_n$. Quindi $L(\mathbf{e}_1) = L \begin{pmatrix} 1 \\ 0 \\ ... \\ 0 \end{pmatrix} = \begin{pmatrix} w_{11} \\ ... \\ w_{m1} \end{pmatrix}$,

ecc. $\forall$ $\mathbf{v} \in V$ $\mathbf{v} = v_1\mathbf{e}_1 + ... + v_n\mathbf{e}_n = \begin{pmatrix} v_1 \\ ... \\ v_n \end{pmatrix}$. Calcolo $L(\mathbf{v}) = L(v_1\mathbf{e}_1 + ... + v_n\mathbf{e}_n) = v_1 L(\mathbf{e}_1) +$

$... + v_n L(\mathbf{e}_n) = v_1 \begin{pmatrix} w_{11} \\ ... \\ w_{m1} \end{pmatrix} + ... + v_n \begin{pmatrix} w_{1n} \\ ... \\ w_{mn} \end{pmatrix}$, che può essere riscritta usando il prodotto matriciale

$\begin{bmatrix} w_{11} & ... & w_{1n} \\ ... & ... & ... \\ w_{m1} & ... & w_{mn} \end{bmatrix} \begin{bmatrix} v_1 \\ ... \\ v_n \end{bmatrix}$. Quindi l'applicazione lineare costruita è in ogni caso codificata col prodotto matriciale.

Tutte e sole le applicazioni lineari $L : \mathbf{K}^n \to \mathbf{K}^m$ sono del tipo $L =_A$, con $A \in M_{\mathbf{K}}(m, n)$. $M_{\mathbf{K}}(m, n) = \hom(\mathbf{K}^n, \mathbf{K}^m)$ indica un omomorfismo, cioè l'insieme delle applicazioni lineari da $\mathbf{K}^n$ in $\mathbf{K}^m$. In generale $M_{\mathbf{K}}(m, n) = mor(\mathbf{K}^n, \mathbf{K}^m)$ indica un morfismo: nel caso dominio e codominio siano strutture algebriche si parla di omomorfismo.

Teorema (isomorfismo con $\mathbf{K}^n$)

Hp: V è uno spazio vettoriale su $\mathbf{K}$, $\dim(V) = n$

Ts: esiste un'applicazione lineare $L : V \to \mathbf{K}^n$ invertibile

Dim Uso la proposizione precedente: $B = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ è una base di V e definisco $L : V \to \mathbf{K}^n$ come l'unica applicazione lineare tale che $L(\mathbf{v}_i) = \mathbf{e}_i$ per ogni $i = 1, ..., n$. L è invertibile perché conserva sia la proprietà di essere generatori che l'indipendenza lineare, essendo $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ la base canonica di $\mathbf{K}^n$. ■

Un'applicazione lineare invertibile si dice isomorfismo; se esiste $L : V \to \mathbf{K}^n$ invertibile, V e $\mathbf{K}^n$ si dicono isomorfi, quindi il teorema può essere espresso più sinteticamente dicendo che ogni spazio vettoriale di dimensione n è isomorfo a $\mathbf{K}^n$.

L'applicazione lineare usata nella dimostrazione riveste un ruolo particolare.

Def Dato uno spazio vettoriale V su un campo $\mathbf{K}$ e una sua base $B = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$, si dice mappa delle coordinate l'applicazione lineare $X_B : V \to \mathbf{K}^n$ definita da $X_B(\mathbf{v}_i) = \mathbf{e}_i$ per $i = 1, ..., n$.

Questo significa che, se un generico vettore $\mathbf{v}$ si scrive rispetto ai vettori di B come $x_1\mathbf{v}_1 + ... + x_n\mathbf{v}_n$, $X_B(\mathbf{v}) = X_B(x_1\mathbf{v}_1 + ... + x_n\mathbf{v}_n) = x_1 X_B(\mathbf{v}_1) + ... + x_n X_B(\mathbf{v}_n) = \begin{pmatrix} x_1 \\ ... \\ x_n \end{pmatrix}$. Quindi la mappa delle coordinate rispetto a B di un vettore di V associa a ogni $\mathbf{v} \in V$ il vettore dei coefficienti della combinazione lineare dei vettori della base che permette di ottenere $\mathbf{v}$.

Tale applicazione lineare è un isomorfismo e l'applicazione lineare inversa si dice mappa di parametrizzazione $P_B : \mathbf{K}^n \to V$, definita da $P_B(\mathbf{e}_i) = \mathbf{v}_i$ per $i = 1, ..., n$.

Questo significa che, dato un vettore $\mathbf{w}$ delle coordinate rispetto ai vettori di B, $P_B(\mathbf{w}) = P_B(x_1\mathbf{e}_1 + ... + x_n\mathbf{e}_n) = x_1 P_B(\mathbf{e}_1) + ... + x_n P_B(\mathbf{e}_n) = x_1\mathbf{v}_1 + ... + x_n\mathbf{v}_n = \mathbf{v}$.

L'idea della mappa della coordinate è poter ricorrere a una "lingua universale", cioè $\mathbf{K}^n$, per esprimere qualsiasi vettore di uno spazio vettoriale e quindi comunicare tra spazi vettoriali diversi.

1. Considero $V = M_{\mathbb{R}}(2, 2)$ e la sua base canonica $B = \{E_{11}, E_{12}, E_{21}, E_{22}\}$. $X_B(E_{11}) = \mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, ..., X_B(E_{22}) = \mathbf{e}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$. Data $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = aE_{11} + bE_{12} + cE_{21} + dE_{22}$, $X_B(A) = aX_B(E_{11}) + bX_B(E_{12}) + cX_B(E_{21}) + dX_B(E_{22}) = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$: sono i coefficienti della combinazione lineare dei vettori della base che permette di ottenere A.

2. Considero $V = \mathbf{K}[t]_{\leq 2}$ e la sua base canonica $B = \{1, t, t^2\}$. La mappa di parametrizzazione $P_B : \mathbf{K}^3 \to$

$\mathbf{K}[t]_{\leq 2}$ associa a ogni vettore $\mathbf{w} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ di $\mathbf{K}^3$ il vettore di V che si ottiene combinando linearmente i vettori di B con coefficienti x, y, z: $P_B(\mathbf{w}) = P_B(x_1\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3) = x_1 \cdot 1 + y \cdot t + z \cdot t^2$. Al variare di $\mathbf{w} \in \mathbf{K}^3$ si descrive ogni vettore di $\mathbf{K}[t]_{\leq 2}$.

Teorema (isomorfismo tra spazi vettoriali con stessa dimensione)

Hp: V, W sono spazi vettoriali su $\mathbf{K}$

Ts: V, W sono isomorfi $\iff \dim V = \dim W$

Dim (i) Se V e W sono isomorfi, esiste un'applicazione lineare $L : V \to W$ invertibile. Considero una base di V $B = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$. Quindi, essendo L iniettiva, $L(\mathbf{v}_1), ..., L(\mathbf{v}_n)$ sono vettori di W linearmente indipendenti; essendo L suriettiva, $L(\mathbf{v}_1), ..., L(\mathbf{v}_n)$ sono generatori di W. Dunque $\{L(\mathbf{v}_1), ..., L(\mathbf{v}_n)\}$ è una base di W e $\dim W = \dim V = n$. In generale, un'applicazione lineare invertibile preserva le basi.

(ii) Vale $\dim V = \dim W = n$: mostro che esiste un'applicazione lineare invertibile da V in W. Chiamo B una base di V, B' una base di W. Considero $X_B : V \to \mathbf{K}^n$, $P_B : \mathbf{K}^n \to V$, $X_{B'} : W \to \mathbf{K}^n$, $P_{B'} : \mathbf{K}^n \to W$. Allora $L = P_{B'} \circ X_B$, $L : V \to W$ è un isomorfismo, perché composizione di isomorfismi, dunque V e W sono isomorfi. ■

Teorema di rappresentazione

Si vuole ora utilizzare la "lingua universale", la traduzione nella quale è permessa dalla mappa delle coordinate, per rappresentare un'applicazione lineare.

Sia V uno spazio vettoriale su $\mathbf{K}$ di dimensione n con base $B = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$, sia W uno spazio vettoriale su $\mathbf{K}$ di dimensione m con base $C = \{\mathbf{w}_1, ..., \mathbf{w}_m\}$. Considero un'applicazione lineare $L : V \to W$. Tutte le applicazione lineari possono essere rappresentate con una matrice: qual è la matrice $A \in M_{\mathbf{K}}(m, n) : X_C(L(\mathbf{v})) = AX_B(\mathbf{v}) =_A (X_B(\mathbf{v}))$, cioè qual è $A : P_C(AX_B(\mathbf{v})) = L(\mathbf{v})$?

Hp: V, W sono spazi vettoriali su $\mathbf{K}$, $L : V \to W$ è un'applicazione lineare,

$B = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ è una base di V, $\dim V = n$, C è una base di W, $\dim W = m$

Ts: la matrice $A \in M_{\mathbf{K}}(m, n)$ definita come $a_{ij} = i -$ esima coordinata rispetto alla base C dell'immagine mediante L del $j -$ esimo vettore di B $(L(\mathbf{v}_j))$ è l'unica matrice tale che, $\forall\ \mathbf{v} \in V$, $AX_B(\mathbf{v}) = X_C(L(\mathbf{v}))$

La tesi richiede che la matrice sia $A = [X_C(L(\mathbf{v}_1)) | .... | X_C(L(\mathbf{v}_n))]$. Significa che ogni applicazione lineare può essere rappresentata come un'applicazione lineare $_A : \mathbf{K}^n \to \mathbf{K}^m$.

Dim $B = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$, $C = \{\mathbf{w}_1, ..., \mathbf{w}_m\}$. Poiché $L(\mathbf{v}_i) \in W$, $L(\mathbf{v}_i)$ si può scrivere come combinazione lineare dei vettori di C: $L(\mathbf{v}_1) = a_{11}\mathbf{w}_1 + a_{21}\mathbf{w}_2 + ... + a_{m1}\mathbf{w}_m$ (in questo caso sto considerando il primo vettore della base B, $j = 1$, perciò ogni coefficiente dev'essere del tipo a_{i1}),..., $L(\mathbf{v}_n) = a_{1n}\mathbf{w}_1 + a_{2n}\mathbf{w}_2 + ... + a_{mn}\mathbf{w}_m$. Dunque

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

dove la colonna j è il vettore delle coordinate di $L(\mathbf{v}_j)$ rispetto a C, la riga i è la sequenza dei coefficienti di posto i in ciascun vettore delle coordinate. Ora che è noto com'è costruita la matrice, si mostra che soddisfa le proprietà richieste. Per ogni $\mathbf{v} \in V$ considero $\mathbf{v} = x_1\mathbf{v}_1 + ... + x_n\mathbf{v}_n$: $L(\mathbf{v}) = x_1 L(\mathbf{v}_1) + ... + x_n L(\mathbf{v}_n)$, perché L è lineare. Le coordinate di $L(\mathbf{v})$ rispetto a C sono $X_C(L(\mathbf{v})) = x_1 X_C(L(\mathbf{v}_1)) + ... + x_n X_C(L(\mathbf{v}_n))$, poiché X_C è lineare; inoltre $X_C(L(\mathbf{v}_i))$ è noto per ogni i perché colonna di A. Quindi $X_C(L(\mathbf{v})) = x_1 \begin{pmatrix} a_{11} \\ ... \\ a_{m1} \end{pmatrix} + ... + x_n \begin{pmatrix} a_{1n} \\ ... \\ a_{mn} \end{pmatrix}$, e questo è uguale a $\begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{m1} & a_{m2} & ... & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ ... \\ x_n \end{bmatrix} = AX_B(\mathbf{v})$. ■

Def La matrice A introdotta nel teorema di rappresentazione si dice matrice rappresentativa di L rispetto alle basi B e C e si indica con $A = M_B^C(L)$ (si pone a pedice la base del dominio, ad apice la base del codominio).

1. $V = \mathbf{K}[t]_{\leq 2}$, $\dim V = 3$, $B = \{1, t, t^2\}$, $W = Span(\mathbf{w}_1, \mathbf{w}_2)$, $C = \{\mathbf{w}_1, \mathbf{w}_2\}$. $L : V \to W$ è definita da $L(1) = \mathbf{w}_1 + \mathbf{w}_2$, $L(t) = \mathbf{w}_1$, $L(t^2) = -\mathbf{w}_1 + 2\mathbf{w}_2$. $X_C(L(1)) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $X_C(L(t)) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $X_C(L(t^2)) = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$, quindi $M_B^C(L) = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 0 & 2 \end{bmatrix}$.

2. $V = W = M_{\mathbb{R}}(2,2)$, $B = \{E_{11}, E_{12}, E_{21}, E_{22}\}$, $L : V \to W$, $L(X) = AX$ con $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. La matrice rappresentativa di L rispetto a B si ottiene calcolando $L(E_{11}) = AE_{11} = \begin{bmatrix} a & 0 \\ c & 0 \end{bmatrix} = aE_{11} + cE_{21}$,

$L\left(E_{12}\right)=AE_{12}=\begin{bmatrix} 0 & a \\ 0 & c \end{bmatrix}=aE_{12}+cE_{22}$, $L\left(E_{21}\right)=AE_{21}=\begin{bmatrix} b & 0 \\ d & 0 \end{bmatrix}=bE_{11}+dE_{21}$, $L\left(E_{22}\right)=AE_{22}=$ $\begin{bmatrix} 0 & b \\ 0 & d \end{bmatrix}=bE_{12}+dE_{22}$. Quindi $M_B^B\left(L\right)=\begin{bmatrix} a & 0 & b & 0 \\ 0 & a & 0 & b \\ c & 0 & d & 0 \\ 0 & c & 0 & d \end{bmatrix}$.

Il teorema di rappresentazione ha alcune conseguenze:

- Se $L:V\rightarrow V$, $L=id$, $B=\{\mathbf{v}_1,...,\mathbf{v}_n\}$ è una base di V, $M_B^B\left(L\right)=id$. Infatti $X_B\left(L\left(\mathbf{v}_1\right)\right)=X_B\left(\mathbf{v}_1\right)=\begin{pmatrix} 1 \\ 0 \\ ... \\ 0 \end{pmatrix}$,..., $X_B\left(L\left(\mathbf{v}_n\right)\right)=X_B\left(\mathbf{v}_n\right)=\begin{pmatrix} 0 \\ ... \\ 0 \\ 1 \end{pmatrix}$, quindi $M_B^B\left(id\right)=Id$ (id è l'applicazione lineare identità, Id è la matrice identità).

- V,W,U sono spazi vettoriali su $\mathbf{K}$, B,C,D sono basi di V,W,U rispettivamente, $L_1:V\rightarrow W$, $L_2:W\rightarrow U$. La matrice rappresentativa della composizione delle due applicazioni lineari è $M_B^D\left(L_2\circ L_1\right)=M_C^D\left(L_2\right)\cdot M_B^C\left(L_1\right)$, cioè il prodotto delle due matrici rappresentative.

- V,W sono spazi vettoriali isomorfi su $\mathbf{K}$, B,C sono basi di V,W rispettivamente (hanno la stessa dimensione per il teorema visto), $L:V\rightarrow W$, $L^{-1}:W\rightarrow V$. La matrice rappresentativa della composizione delle due applicazioni lineari è $M_B^B\left(L\circ L^{-1}\right)=M_C^B\left(L^{-1}\right)\cdot M_B^C\left(L\right)$: poiché $M_B^B\left(L\circ L^{-1}\right)=M_B^B\left(id\right)=Id$ per la prima conseguenza, $M_C^B\left(L^{-1}\right)\cdot M_B^C\left(L\right)=Id_n$, cioè $M_C^B\left(L^{-1}\right)=M_B^C\left(L\right)^{-1}$ (perché le due matrici sono quadrate, essendo V,W isomorfi, e se esiste l'inversa sinistra esiste anche l'inversa destra). La matrice rappresentativa dell'inversa di L è l'inversa della matrice rappresentativa di L.

- V,W sono spazi vettoriali su $\mathbf{K}$, B,C sono basi di V,W rispettivamente, $L:V\rightarrow W$, $X_B:V\rightarrow\mathbf{K}^n$, $X_C:W\rightarrow\mathbf{K}^m$, $A=M_B^C\left(L\right)$, ${}_A:\mathbf{K}^n\rightarrow\mathbf{K}^m$. E' noto che $\ker\left({}_A\right)=\ker\left(A\right)$ e $Im\left({}_A\right)=col\left(A\right)$, ma quali sono le relazioni con $\ker\left(L\right)$e $Im\left(L\right)$, rispettivamente? Vale $X_B\left(\ker\left(L\right)\right)=\ker\left(A\right)$ (si applica all'indietro la linearità di L) e $X_C\left(Im\left(L\right)\right)=col\left(A\right)$: coerentemente con il teorema di rappresentazione, gli elementi del nucleo di A sono i rappresentanti mediante coordinate degli elementi del nucleo di L, e analogamente per l'immagine. Si dimostra con il fatto che X_C,X_B sono isomorfismi.

4.5.3 Teorema di nullità più rango

Hp:	V, W sono spazi vettoriali su $\mathbf{K}$ di dimensione finita, $L : V \to W$ è un'applicazione lineare
Ts:	$\dim V = \dim(\ker(L)) + \dim(Im(L))$

che è una reinterpretazione del già visto teorema del rango fondata sul fatto che ogni applicazione lineare è rappresentata da una matrice. $\dim(\ker(L))$ è la dimensione della parte di informazione "buttata via".

Dim Chiamo B una base di V, $\dim V = n$, C una base di W, $\dim W = m$. Considero ${}_A : \mathbf{K}^n \to \mathbf{K}^m$, $A = M_B^C(L)$; per la conseguenza 4 $X_B(\ker(L)) = \ker({}_A)$: poiché X_B è iniettiva (in quanto invertibile), $\dim(X_B(\ker(L))) = \dim(\ker({}_A)) = \dim(\ker(A))$ ($\ker(L)$ e $\ker({}_A)$ sono isomorfi). Analogamente, $X_C(Im(L)) = col(A)$: poiché X_C è iniettiva, $\dim(X_C(Im(L))) = \dim(col(A))$ ($Im(L)$ e $col(A)$ sono isomorfi). Ma allora, per il teorema del rango, $\dim(\ker(A)) + \dim(col(A)) = n - r(A) + r(A) = n = \dim \mathbf{K}^n = \dim V$. ■

Def Data $L : V \to W$, si chiama rango di L e si indica con $r(L)$ la dimensione dell'immagine di L.

1. Se $A \in M_{\mathbf{K}}(m, n)$, $\dim(\ker A) = n \iff \dim(ImA) = r(A) = 0$, cioè il nucleo ha dimensione massima (i. e. coincide con $\mathbb{R}^n$) se e solo se A è la matrice nulla.

Corollario

Hp: V, W sono spazi vettoriali su $\mathbf{K}$, $\dim V = n$, $\dim W = m$,

$L : V \to W$ è un'applicazione lineare, $r = r(L)$

Ts: (i) L è iniettiva $\iff \dim V = \dim(Im(L)) \iff n = r$

(ii) L è suriettiva $\iff Im(L) = W \iff \dim(ImL) = \dim(W) \iff r = m$

(iii) L è invertibile $\iff r = m = n$

Se ne possono dedurre conseguenze necessarie per iniettività e suriettività: se $n > m$, $r \leq m$ e L non può essere iniettiva ($\dim(\ker(L)) \geq n - m$); se $n < m$, $r \leq n$ e L non può essere suriettiva ($\dim(Im(L)) \leq n$).

4.5.4 Cambiamenti di base

Ha senso chiedersi come cambia la matrice rappresentativa di un'applicazione lineare se si cambiano le basi con cui si costruisce.

1. $V = W = \mathbb{R}^2$, $L : \mathbb{R}^2 \to \mathbb{R}^2$ è la simmetria assiale rispetto a una retta data. Tale retta deve passare per l'origine, altrimenti la simmetria assiale non è un'applicazione lineare. Voglio descrivere L mediante una sua

matrice rappresentativa; fisso come base di $V=W$ la base canonica $B=\{\mathbf{e}_1,\mathbf{e}_2\}$. $M_B^B(L)$? Utilizzando la trigonometria si nota che $L(\mathbf{e}_1)=\begin{pmatrix}\cos 2\alpha\\ \sin 2\alpha\end{pmatrix}$ e $L(\mathbf{e}_2)=\begin{pmatrix}\cos\left(\frac{\pi}{2}-2\left(\frac{\pi}{2}-\alpha\right)\right)\\ \sin\left(\frac{\pi}{2}-2\left(\frac{\pi}{2}-\alpha\right)\right)\end{pmatrix}=\begin{pmatrix}\sin 2\alpha\\ -\cos 2\alpha\end{pmatrix}$, dove α è l'angolo tra la retta e l'asse. Quindi si ha $M_B^B(L)=\begin{bmatrix}\cos 2\alpha & \sin 2\alpha\\ \sin 2\alpha & -\cos 2\alpha\end{bmatrix}$. Tuttavia questa scelta ignora la natura geometrica dell'applicazione: considerandola ci si accorge che ci sono vettori la cui simmetria assiale è particolarmente semplice, e. g. qualsiasi vettore $\mathbf{v}$ giaccia completamente sulla retta (l'asse di riflessione) è tale che $L(\mathbf{v})=\mathbf{v}$, e un vettore $\mathbf{w}$ ortogonale all'asse di riflessione è tale che $L(\mathbf{w})=-\mathbf{w}$. $\mathbf{v},\mathbf{w}$ sono linearmente indipendenti, quindi $C=\{\mathbf{v},\mathbf{w}\}$ è una base di $\mathbb{R}^2$ (che dipende da L, cioè dalla retta). Adesso scrivere la matrice rappresentativa è molto più semplice: $M_C^C(L)=\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}$ è una matrice diagonale (cioè se $i\neq j$ $a_{ij}=0$), descritta da $n=2$ coefficienti, mentre $M_B^B(L)$ è una matrice "normale" descrita da $n^2=4$ coefficienti. Se si avesse genericamente $L:\mathbb{R}^n\to\mathbb{R}^n$, si starebbe confrontando una matrice con $n\cdot n$ coefficienti con una descritta dagli n coefficienti sulla diagonale.

Ha senso scegliere, per rappresentare un'applicazione lineare $L:\mathbb{R}^n\to\mathbb{R}^n$, una matrice diagonale.

Qual è la relazione tra due matrici che rappresentano la stessa applicazione lineare?

Siano V,W spazi vettoriali tali che $\dim V=n,\dim W=m$, B_1,B_2 sono basi di V, C_1,C_2 sono basi di W, $L:V\to W$ è un'applicazione lineare. Allora ci sono quattro possibili matrici rappresentative di L: $M_{B_1}^{C_1}(L)$ (rosso), $M_{B_1}^{C_2}(L)$ (verde), $M_{B_2}^{C_1}(L)$ (viola), $M_{B_2}^{C_2}(L)$ (blu). Questo significa che $X_{C_1}(L(\mathbf{v}))=M_{B_1}^{C_1}(L)X_{B_1}(\mathbf{v})$, $X_{C_1}(L(\mathbf{v}))=M_{B_2}^{C_1}(L)X_{B_2}(\mathbf{v})$ eccetera.dtbpF3.7137in2.134in0ptcambio di base.png

Tuttavia, se si conoscesse una matrice $S:X_{B_2}(\mathbf{v})=SX_{B_1}(\mathbf{v})$, si potrebbe scrivere anche $X_{C_1}(L(\mathbf{v}))=M_{B_2}^{C_1}(L)X_{B_2}(\mathbf{v})=M_{B_2}^{C_1}(L)SX_{B_1}(\mathbf{v})$, cioè, se fosse nota la matrice per passare dalla rappresentazione di $\mathbf{v}$ rispetto a una base alla rappresentazione di $\mathbf{v}$ rispetto a un'altra base, ci si potrebbe muovere liberamente nel diagramma. Questo è un problema indipendente dall'applicazione lineare con cui si ha a che fare. Qual è $S:X_{B_2}(\mathbf{v})=SX_{B_1}(\mathbf{v})$? Questo è un caso particolare della ricerca di $A:AX_B(\mathbf{v})=X_C(L(\mathbf{v}))$, in cui $L:V\to V$ è $L(\mathbf{v})=\mathbf{v}$, $B=B_1$ e $C=B_2$. La matrice cercata è quindi $M_{B_1}^{B_2}(id)$. Il ruolo di B_1 e B_2 è interscambiabile: $M_{B_1}^{B_2}(id)=\left(M_{B_2}^{B_1}(id)\right)^{-1}$.

Def Date due basi B_1,B_2 di uno spazio vettoriale V, la matrice $M_{B_2}^{B_1}(id)$ si dice matrice del cambio di base da B_2 a B_1.

Chiamo $P=M_{B_2}^{B_1}(id)$, $Q=M_{C_2}^{C_1}(id)$. Allora $M_{B_1}^{C_1}(L)=M_{C_2}^{C_1}(id)M_{B_1}^{C_2}(L)=QM_{B_1}^{C_2}(L)$ per la conseguenza due del teorema di rappresentazione; $M_{B_1}^{C_1}(L)=QM_{B_1}^{C_2}(L)=QM_{B_2}^{C_2}(L)P^{-1}$, perché $M_{B_1}^{C_2}(L)=M_{B_2}^{C_2}(L)M_{B_1}^{B_2}(id)$

sempre per la conseguenza due; $M_{B_1}^{C_1}(L) = M_{B_2}^{C_1}(L)\,P^{-1} = M_{B_2}^{C_1}(L)\,M_{B_1}^{B_2}(id)$ sempre per la conseguenza due.

1. Nell'esempio precedente, scegliendo $B = \{\mathbf{e}_1, \mathbf{e}_2\}$ si ha come matrice rappresentativa $A = M_B^B(L) = \begin{bmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{bmatrix}$, con $C = \{\mathbf{v}, \mathbf{w}\}$ si ha $B = M_C^C(L) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Scrivo una matrice del cambio di base: la più facile da scrivere è la $P = M_C^B(id) = \begin{bmatrix} \cos\alpha & \cos\left(\frac{\pi}{2}+\alpha\right) \\ \sin\alpha & \sin\left(\frac{\pi}{2}+\alpha\right) \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$, dove $\mathbf{v} = \cos\alpha\mathbf{e}_1 + \sin\alpha\mathbf{e}_2$ e $\mathbf{w} = -\sin\alpha\mathbf{e}_1 + \cos\alpha\mathbf{e}_2$. Quindi vale $M_B^B(L) = M_C^B(id)\,M_C^C(L)\,M_B^C(id) \iff A = PBP^{-1} \iff B = P^{-1}AP$.

Def Due matrici $A, B \in M_{\mathbf{K}}(n,n)$ si dicono simili se esiste una matrice invertibile $P \in M_{\mathbf{K}}(n,n) : B = P^{-1}AP$.

Se due matrici A, B rappresentano la stessa applicazione lineare, allora sono simili (implicazione inversa?): se $B = P^{-1}AP$, P è la matrice del cambio di base dalla base rispetto a cui B rappresenta l'applicazione lineare alla base canonica (rispetto cui A rappresenta l'endomorfismo), B rappresenta l'endomorfismo rispetto alla base formata dalle colonne di P.

Due matrici A, B simili hanno lo stesso determinante: infatti $\det B = \det P^{-1}AP = \det P^{-1}\det A \det P = \det A \det P^{-1} \det P = \det A \det Id = \det A$ usando due volte il teorema di Binet e la commutatività del prodotto su $\mathbf{K}$.

La relazione di similitudine tra matrici è una relazione di equivalenza. Infatti ha le proprietà:

1. Riflessiva: ogni matrice A è simile a se stessa. $A = Id_n^{-1}AId_n = Id_nAId_n$
2. Simmetrica: se A è simile a B allora B è simile ad A. Infatti se $B = P^{-1}AP$, si ha $A = PBP^{-1} = Q^{-1}BQ$, ponendo $Q = P^{-1}$.
3. Transitiva: se A è simile a B e B è simile a C, allora A è simile a C. Infatti se $B = P^{-1}AP$, $C = Q^{-1}BQ$, si ha $C = Q^{-1}BQ = Q^{-1}\left(P^{-1}AP\right)Q = \left(Q^{-1}P^{-1}\right)A(PQ) = (PQ)^{-1}A(PQ) = R^{-1}AR$, ponendo $R = PQ$.

Ogni relazione di equivalenza in un insieme induce una partizione dell'insieme in classi di equivalenza (come avviene anche e. g. per le frazioni): quindi si può dividere $M_{\mathbf{K}}(n,n)$ in classi di equivalenza di matrici simili, prive di intersezioni. Una classe di equivalenza di matrici simili, rappresentata dalla matrice A, si indica con $[A]$.

Come scegliere una opportuna matrice per rappresentare ogni classe di equivalenza (cioè ogni applicazione lineare con dominio e codominio isomorfi)?

4.6 Diagonalizzabilità, autovettori, autovalori, similitudine

Tutto quello che segue riguarda matrici quadrate.

Def Una matrice $A \in M_{\mathbf{K}}(n,n)$ si dice diagonalizzabile su $\mathbf{K}$ se esiste una matrice diagonale D a cui A è simile, i. e. se $\exists\ P \in M_{\mathbf{K}}(n,n)$ invertibile, $\exists\ D \in M_{\mathbf{K}}(n,n) : P^{-1}AP = D$.

Analogamente, dato uno spazio vettoriale V su un campo $\mathbf{K}$, sia $L : V \to V$ un'applicazione lineare. L si dice diagonalizzabile se esiste una base B di V tale che la matrice rappresentativa di L rispetto a B è diagonale.

Il campo $\mathbf{K}$ ha una ruolo rilevante nella definizione: data $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, se si considera $A \in M_{\mathbb{Q}}(2,2)$ o $M_{\mathbb{R}}(2,2)$ A non è diagonalizzabile, ma lo è su $\mathbb{C}$ (si trova una matrice diagonale a coefficienti complessi cui A è simile).

Se A è diagonalizzabile, il rappresentante della classe di equivalenza di matrici simili cui A appartiene è la matrice diagonale cui A è simile.

Def Sia V uno spazio vettoriale, $L : V \to V$ un'applicazione lineare. Un vettore $\mathbf{v} \in V$, $\mathbf{v} \neq \mathbf{0}$, si dice autovettore di L se $\exists\ \lambda \in \mathbf{K}$, detto autovalore di $\mathbf{v}$, tale che $L(\mathbf{v}) = \lambda \mathbf{v}$.

Un autovettore di $A \in M_{\mathbf{K}}(n,n)$ è un autovettore dell'applicazione lineare ${}_A$, quindi $\mathbf{v} \neq \mathbf{0} : \exists\ \lambda \in \mathbf{K} :_A (\mathbf{v}) = A\mathbf{v} = \lambda \mathbf{v}$.

1. Nell'esempio della simmetria assiale, $\mathbf{v}$ giacente sulla retta è un autovettore di L con autovalore 1, $\mathbf{w}$ ortogonale alla retta è un autovettore di L con autovalore -1 (gli autovalori sono gli elementi sulla diagonale). Si è inoltre notato che esiste una matrice rappresentativa di L diagonale, quindi la matrice $\begin{bmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{bmatrix}$ è diagonalizzabile su $\mathbb{R}$. La diagonalizzabilità è legata al numero di autovettori.

Primo criterio di diagonalizzabilità

Hp: $A \in M_{\mathbf{K}}(n,n)$

Ts: A è diagonalizzabile su $\mathbf{K} \iff$ esiste una base di $\mathbf{K}^n$ formata da autovettori di A

e in tal caso ogni matrice che rappresenti ${}_A$ rispetto a una base di autovettori di A è diagonale, con gli autovalori sulla diagonale: ad essa A è simile, e la matrice del cambio di base da una base di autovettori alla base canonica è formata proprio da tali autovettori.

Dim (i) Se A è diagonalizzabile, allora esiste una matrice diagonale $D = \begin{bmatrix} d_{11} & ... & 0 \\ ... & ... & ... \\ 0 & ... & d_{33} \end{bmatrix}$ cui A è simile: due matrici simili rappresentano la stessa applicazione lineare ${}_A : \mathbf{K}^n \rightarrow \mathbf{K}^n$, quindi esiste una base $B = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ di $\mathbf{K}^n$ tale che $M_B^B({}_A) = D$. Determino l'immagine attraverso ${}_A$ dei vettori di B: $X_B(\mathbf{v}_1) = \begin{pmatrix} 1 \\ 0 \\ ... \\ 0 \end{pmatrix}$, $X_B({}_A(\mathbf{v}_1)) = \begin{bmatrix} d_{11} & ... & 0 \\ ... & ... & ... \\ 0 & ... & d_{nn} \end{bmatrix} \begin{pmatrix} 1 \\ 0 \\ ... \\ 0 \end{pmatrix} = \begin{pmatrix} d_{11} \\ 0 \\ ... \\ 0 \end{pmatrix} = d_{11} X_B(\mathbf{v}_1)$, quindi $\mathbf{v}_1$ è un autovettore di A con autovalore d_{11}. Lo stesso vale per ogni altro vettore della base: per ogni $i = 1, ..., n$, $\mathbf{v}_i$ è un autovettore di A con autovalore d_{ii}. Quindi esiste una base di $\mathbf{K}^n$ formata da autovettori di A: è proprio la base B rispetto a cui la matrice diagonale rappresenta ${}_A$.

(ii) Esiste per ipotesi una base di $\mathbf{K}^n$ formata da autovettori di A: la chiamo $B = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$, e per ogni $i = 1, ..., n$ ${}_A(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$. Mi chiedo com'è fatta la matrice rappresentativa di ${}_A$ rispetto a tale base $M_B^B({}_A)$.

$$M_B^B({}_A) = \begin{bmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & 0 & \lambda_n \end{bmatrix}$$

perché ${}_A(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$, quindi $X_B({}_A(\mathbf{v}_1)) = \begin{pmatrix} \lambda_1 \\ 0 \\ ... \\ 0 \end{pmatrix}$. Poiché esiste una matrice rappresentativa diagonale di ${}_A$, esiste una matrice diagonale cui A è simile: $\exists\, P$ invertibile tale che $P^{-1}AP = D$, perciò A è diagonalizzabile. ■

Da questo teorema si evince che per studiare la diagonalizzabilità di una matrice si devono studiare i suoi autovettori.

1. Considero $L : \mathbf{K}[t]_{\leq 3} \rightarrow \mathbf{K}[t]_{\leq 3}$, $L(p(t)) = tp''(t) - p'(t)$ e la base canonica $B = \{1, t, t^2, t^3\}$.

Scrivo $M_B^B(L)$:

$$M_B^B(L) = A = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

A non è diagonale; è diagonalizzabile? Per il primo criterio, la diagonalizzabilità è equivalente all'esistenza di una base di $\mathbf{K}[t]_{\leq 3}$ (o $\mathbf{K}^4$) formata da autovettori di A.

Se ragiono su $\mathbf{K}[t]_{\leq 3}$ per trovare gli autovettori, significa che cerco $\lambda \in \mathbf{K}$ e $p(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3$: $L(p(t)) = \lambda p(t) \iff tp''(t) - p'(t) = \lambda p(t)$. Sto cercando simultaneamente autovalori e autovettori: ho parametri che ricoprono una funzione diversa.

Se invece ragiono su $\mathbf{K}^4$, sto cercando l'autovalore $\lambda \in \mathbf{K}$ e l'autovettore $\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} \in \mathbf{K}^4 : A \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \lambda \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}$. Voglio scrivere questa uguaglianza nella forma consueta di rappresentazione di un sistema lineare:

allora scrivo

$$A\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \lambda \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} \Longleftrightarrow A\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \lambda \left(Id_4 \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} \right)$$

$$A\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = (\lambda Id_4) \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} \Longleftrightarrow A\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} - (\lambda Id_4) \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = 0$$

$$(A - \lambda Id_4)\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = 0$$

Quindi ho un sistema lineare omogeneo con matrice dei coefficienti $A - \lambda Id_4$: trovare autovalori e autovettori di A si riduce a studiare le soluzioni di tale sistema. Ricordo che affinché $\mathbf{v} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}$ sia un autovettore dev'essere $\mathbf{v} \neq \mathbf{0}$: il sistema omogeneo non ammette solo la soluzione nulla (cioè esiste un autovettore) se e solo se $r(A - \lambda Id_4) < n \Longleftrightarrow \det(A - \lambda Id_4) = 0$.

In generale, λ è autovalore di $A \Longleftrightarrow r(A - \lambda Id) < n \Longleftrightarrow \det(A - \lambda Id) = 0 \Longleftrightarrow A - \lambda Id$ non è invertibile.

Def Data una matrice $A \in M_{\mathbf{K}}(n, n)$, il polinomio $p_A(x) = \det(A - xId_n)$ si dice polinomio caratteristico di A e l'equazione $p_A(x) = 0$ si dice equazione caratteristica di A.

Il polinomio caratteristico è di grado al più n.

Proposizione

$$\begin{array}{lcl} \text{Hp} & : & A \in M_{\mathbf{K}}(n, n) \\ \text{Ts} & : & \lambda \in \mathbf{K} \text{ è un autovalore di } A \Longleftrightarrow p_A(\lambda) = 0 \end{array}$$

cioè λ è un autovalore di A se e solo se è soluzione dell'equazione caratteristica.

Quindi la sequenza di passaggi da fare per trovare gli autovettori e gli autovalori di un'applicazione lineare è:

- scriverne una matrice rappresentativa

- determinare il polinomio caratteristico e le sue radici

- calcolare le soluzioni del sistema omogeneo sostituendo a λ i valori trovati al punto precedente

1. Riprendo l'esempio di prima: $p_A(x) = \det(A - \lambda Id_4) = \det \begin{bmatrix} -x & -1 & 0 & 0 \\ 0 & -x & 0 & 0 \\ 0 & 0 & -x & 3 \\ 0 & 0 & 0 & -x \end{bmatrix} = x^4$ perché $A - \lambda Id_4$ è triangolare superiore. C'è un'unica soluzione dell'equazione caratteristica: $\lambda = 0$. Determino gli autovettori risolvendo il sistema $A\mathbf{x} = \mathbf{0}$: se un autovalore è nullo, gli elementi del nucleo di A sono autovettori rispetto all'autovalore nullo: quindi esiste l'autovalore 0 se e solo se il nucleo non contine solo il vettore nullo, cioè se e solo se $r(A) < n$. Trovo che $\ker(A) = Span\left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}\right)$ è l'insieme degli autovettori di A. Esistono inoltre due autovettori di A linearmente indipendenti, che è il massimo numero di autovettori linearmente indipendenti di A, essendo $\dim(\ker(A)) = 2$. Questo significa che non esiste una base di $\mathbf{K}^4$ formata da autovettori di A, perché non si possono trovare quattro vettori di A linearmente indipendenti: quindi A non è diagonalizzabile.

2. Data $A \in M_{\mathbf{K}}(n, n)$, A^T ha lo stesso polinomio caratteristico di A: $\det\left(A^T - \lambda Id\right) = \det(A - \lambda Id)$. Infatti, posto $B = A - \lambda Id$, $\det B = \det(A - \lambda Id) = \det B^T$, e $\det B^T = \det(A - \lambda Id)^T = \det\left(A^T - \lambda Id^T\right) = \det\left(A^T - \lambda Id\right)$ per linearità della trasposizione.

Def Sia $A \in M_{\mathbf{K}}(n, n)$, sia λ un autovalore di A. Si definisce molteplicità algebrica di λ $\max\left\{k \in \mathbb{N} : (x - \lambda)^k \text{ divide } p_A(x)\right\}$ e si indica con a_λ o $a(\lambda)$. Si definisce autospazio associato a λ il sottospazio vettoriale degli autovettori con autovalore λ, cioè $\{\mathbf{v} \in \mathbf{K}^n : A\mathbf{v} = \lambda\mathbf{v}\} = \{\mathbf{v} \in \mathbf{K}^n : (A - \lambda Id_n)\mathbf{v} = 0\} = \ker(A - \lambda Id_n)$, e si indica con V_λ. Si definisce molteplicità geometrica di λ la dimensione dell'autospazio associato a λ $\dim V_\lambda$ e si indica con g_λ o $g(\lambda)$.

Vale $g(\lambda) = \dim V_\lambda = \dim(\ker(A - \lambda Id_n)) = n - r(A - \lambda Id_n)$, quindi, essendo $r(A - \lambda Id_n) < n$ (altrimenti l'autospazio V_λ conterrebbe solo il vettore nullo), $g(\lambda) \geq 1$. Inoltre $1 \leq g(\lambda) \leq a(\lambda)$; se $a(\lambda) = 1$, λ si dice autovalore semplice e $g(\lambda) = 1$. Se $a(\lambda) = g(\lambda)$, λ si dice autovalore regolare. Se un autovalore è semplice, è anche regolare.

L'autospazio associato a λ è l'insieme degli autovettori con autovalore λ, più il vettore nullo. L'insieme degli autovettori con autovalore λ è un sottospazio vettoriale perché, se $\mathbf{v}_1, \mathbf{v}_2$ sono autovettori con autovalore λ, anche $\mathbf{v}_1+\mathbf{v}_2$ è autovettore con autovalore λ: $A(\mathbf{v}_1+\mathbf{v}_2) = A\mathbf{v}_1 + A\mathbf{v}_2 = \lambda\mathbf{v}_1 + \lambda\mathbf{v}_2 = \lambda(\mathbf{v}_1+\mathbf{v}_2)$; se $\mathbf{v}_1$ è autovettore con autovalore λ, anche $t\mathbf{v}_1$ è autovettore con autovalore λ: $A(t\mathbf{v}_1) = tA\mathbf{v}_1 = t\lambda\mathbf{v}_1 = \lambda(t\mathbf{v}_1)$. In effetti è ovvio perché $V_\lambda = \ker(A - \lambda Id_n)$.

1. Nell'esempio di prima, la molteplicità algebrica dell'autovalore $\lambda = 0$ è $a_0 = 4$ e la molteplicità geometrica dell'autospazio associato a $\lambda = 0$ è $\dim(V_0) = g_0 = 2$.

2. Data una matrice $A \in M_{\mathbb{R}}(3,3)$ il cui polinomio caratteristico è $p_A(x) = -x^3 + x$, allora si può dire che la matrice $A + kId$ è invertibile per ogni $k \neq 0, 1, -1$. Infatti, $A + kId$ è invertibile $\Longleftrightarrow r(A + kId) = 3 \Longleftrightarrow \dim(\ker(A + kId)) = 0$, cioè l'autospazio relativo all'autovalore $-k$ contiene solo il vettore nullo (quindi $-k$ non è un autovalore di A). Questo accade quando $k \neq 0, 1, -1$, gli opposti degli autovalori di A.

In generale, λ è autovalore di $A \Longleftrightarrow r(A - \lambda Id) < n \Longleftrightarrow \det(A - \lambda Id) = 0 \Longleftrightarrow A - \lambda Id$ non è invertibile.

Se ${}_A : V \to V$ è un'applicazione lineare con $\dim V \geq 1$, A ha almeno un autovettore non nullo.

Una matrice triangolare A con tutti zeri sulla diagonale e almeno un elemento non nullo non è diagonalizzabile, perché ha come radice del polinomio caratteristico $0 : a_0 = n$, ma $r(A) \geq 1$.

Secondo criterio di diagonalizzabilità

$$A \in M_{\mathbf{K}}(n,n) \text{ è diagonalizzabile su } \mathbf{K} \iff$$

$$\text{(i) } p_A(x) = \det(A - \lambda Id_n) \text{ ha } n \text{ radici in } \mathbf{K} \text{ contate con molteplicità, i. e. } a(\lambda_1) + a(\lambda_2) + ... = n$$

$$\text{(ii) per ogni autovalore molteplicità algebrica e geometrica}$$

$$\text{coincidono, i. e. } a(\lambda_i) = g(\lambda_i) \text{ per ogni } i$$

La seconda tesi si può esprimere dicendo che ogni autovalore è regolare. Per il teorema fondamentale dell'algebra, che afferma che ogni polinomio di grado n a coefficienti in $\mathbb{C}$ ha esattamente n radici contate con molteplicità, la prima condizione è sempre soddisfatta se $\mathbf{K} = \mathbb{C}$.

Dim Si dimostra solo l'implicazione da destra a sinistra: la dimostrazione è divisa in quattro parti.

1) Per ogni autovalore λ, vale $1 \leq g(\lambda) \leq a(\lambda)$ (la seconda disuguaglianza è data per dimostrata). Se $g(\lambda_i) \leq a(\lambda_i)$ per ogni $i = 1, ..., s$ (per ogni autovalore λ_i), supponendo di avere s autovalori, vale $\sum_{i=1}^s g(\lambda_i) \leq \sum_{i=1}^s a(\lambda_i) = a(\lambda_1) + a(\lambda_2) + ... + a(\lambda_s) \leq n$ perché il polinomio caratteristico non può avere più di n radici; questo è vero in generale. Ma poiché per ipotesi $a(\lambda_i) = g(\lambda_i)$ per ogni i e $a(\lambda_1) + a(\lambda_2) + ... = n$, le due disuguaglianze

diventano due uguaglianze: $g(\lambda_1)+...+g(\lambda_s)=a(\lambda_1)+...+a(\lambda_s)=n$, cioè la somma delle dimensioni di ciascun autospazio è n.

2) Dimostro la proposizione

$$\begin{array}{ll} \text{Hp:} & \lambda_1, \lambda_2 \text{ sono due autovalori di } A \text{ tali che } \lambda_1 \neq \lambda_2 \\ \text{Ts:} & V_{\lambda_1} \cap V_{\lambda_2} = \{\mathbf{0}\} \end{array}$$

cioè se λ_1 e λ_2 sono due autovalori distinti, allora $V_{\lambda_1} \cap V_{\lambda_2} = \{\mathbf{0}\}$. $\mathbf{v} \in (V_{\lambda_1} \cap V_{\lambda_2}) \iff \mathbf{v} \in V_{\lambda_1}$, $\mathbf{v} \in V_{\lambda_2}$: significa che $\mathbf{v}$ è un autovettore rispetto a due autovalori diversi, cioè $A\mathbf{v} = \lambda_1\mathbf{v}$ e $A\mathbf{v} = \lambda_2\mathbf{v}$. Quindi vale $\lambda_1\mathbf{v} = \lambda_2\mathbf{v} \iff (\lambda_1 - \lambda_2)\mathbf{v} = \mathbf{0}$: per la legge di annullamento del prodotto, essendo $\lambda_1 - \lambda_2 \neq 0$, vale $\mathbf{v} = \mathbf{0}$. Poiché $V_{\lambda_1} \cap V_{\lambda_2} = \{\mathbf{0}\}$, $V_{\lambda_1} + V_{\lambda_2} = V_{\lambda_1} \oplus V_{\lambda_2}$, quindi $\dim(V_{\lambda_1} + V_{\lambda_2}) = \dim V_{\lambda_1} + \dim V_{\lambda_2} = g(\lambda_1) + g(\lambda_2)$.

3) Dimostro per induzione la proposizione

$$\begin{array}{ll} \text{Hp:} & \lambda_1, \lambda_2, ..., \lambda_s \text{ sono autovalori distinti di } A,\ \mathbf{v}_1 \in V_{\lambda_1}, ..., \mathbf{v}_s \in V_{\lambda_s} \\ \text{Ts:} & \{\mathbf{v}_1 \in V_{\lambda_1}, ..., \mathbf{v}_s \in V_{\lambda_s}\} \text{ è un insieme di vettori linearmente indipendenti} \end{array}$$

cioè se $\lambda_1, \lambda_2, ..., \lambda_s$ sono autovalori distinti e se considero un autovettore per ogni autospazio, allora tali autovettori sono linearmente indipendenti. Inizio con il caso base, mostrando che, se $\lambda_1 \neq \lambda_2$, $\{\mathbf{v}_1 \in V_{\lambda_1}, \mathbf{v}_2 \in V_{\lambda_2}\}$ è un insieme di vettori linearmente indipendenti: voglio mostrare che $t_1\mathbf{v}_1 + t_2\mathbf{v}_2 = \mathbf{0}$ implica $t_1 = t_2 = 0$. [In realtà l'indipendenza lineare di $\mathbf{v}_1 \in V_{\lambda_1}, \mathbf{v}_2 \in V_{\lambda_2}$ segue dal fatto che $V_{\lambda_1} \cap V_{\lambda_2} = \{\mathbf{0}\}$: se per assurdo fosse $\mathbf{v}_1 = t\mathbf{v}_2$, allora ciascuno dei due vettori sarebbe autovettore sia rispetto a λ_1 che rispetto a λ_2 e l'intersezione conterrebbe anche loro due: $L(\mathbf{v}_1) = \lambda_1\mathbf{v}_1$ e $L(\mathbf{v}_1) = L(t\mathbf{v}_2) = tL(\mathbf{v}_2) = t(\lambda_2\mathbf{v}_2) = \lambda_2(t\mathbf{v}_2) = \lambda_2\mathbf{v}_1$] Per sfruttare il fatto che $\mathbf{v}_1, \mathbf{v}_2$ sono autovettori, applico $_A$ a lato sinistro e destro: $A(t_1\mathbf{v}_1 + t_2\mathbf{v}_2) = A\mathbf{0} \iff t_1\lambda_1\mathbf{v}_1 + t_2\lambda_2\mathbf{v}_2 = \mathbf{0}$. Voglio avere λ_1 e λ_2 nello stesso coefficiente, quindi riscrivo l'equazione come $t_1\lambda_1\mathbf{v}_1 + t_2\lambda_2\mathbf{v}_2 - \lambda_2\mathbf{0} = \mathbf{0}$ e uso l'ipotesi $t_1\mathbf{v}_1 + t_2\mathbf{v}_2 = \mathbf{0}$: sostituendo ottengo $t_1\lambda_1\mathbf{v}_1 + t_2\lambda_2\mathbf{v}_2 - \lambda_2(t_1\mathbf{v}_1 + t_2\mathbf{v}_2) = \mathbf{0} \iff (\lambda_1 - \lambda_2)t_1\mathbf{v}_1 = \mathbf{0}$. $\lambda_1 \neq \lambda_2$ per ipotesi, $\mathbf{v}_1 \neq \mathbf{0}$ perché è un autovettore, quindi può essere solo $t_1 = 0$, per la legge di annullamento del prodotto. Allora, se $t_1\mathbf{v}_1 + t_2\mathbf{v}_2 = \mathbf{0}$ e $t_1 = 0$, di nuovo per la legge di annullamento del prodotto, essendo $\mathbf{v}_2 \neq \mathbf{0}$, $t_2 = 0$.

Allora, prendendo come ipotesi induttiva l'indipendenza lineare in $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$, si mostra l'indipendenza lineare dei vettori di $\{\mathbf{v}_1, ..., \mathbf{v}_k, \mathbf{v}_{k+1}\}$: voglio mostrare che $t_1\mathbf{v}_1 + ... + t_k\mathbf{v}_k + t_{k+1}\mathbf{v}_{k+1} = \mathbf{0}$ implica $t_1 = ... = t_k = t_{k+1} = 0$. Per sfruttare il fatto che $\mathbf{v}_1, ..., \mathbf{v}_{k+1}$ sono autovettori, applico $_A$ a lato sinistro e destro: $A(t_1\mathbf{v}_1 + ... + t_k\mathbf{v}_k + t_{k+1}\mathbf{v}_{k+1}) = A\mathbf{0} \iff t_1\lambda_1\mathbf{v}_1 + ... + t_k\lambda_k\mathbf{v}_k + t_{k+1}\lambda_{k+1}\mathbf{v}_{k+1} = \mathbf{0}$. Come prima, riscrivo l'equazione come $t_1\lambda_1\mathbf{v}_1 + ... + t_k\lambda_k\mathbf{v}_k + t_{k+1}\lambda_{k+1}\mathbf{v}_{k+1} - \lambda_{k+1}\mathbf{0} = \mathbf{0}$ e uso l'ipotesi $t_1\mathbf{v}_1 + ... + t_k\mathbf{v}_k + t_{k+1}\mathbf{v}_{k+1} = \mathbf{0}$: sostituendo ottengo $t_1\lambda_1\mathbf{v}_1 + ... + t_k\lambda_k\mathbf{v}_k + t_{k+1}\lambda_{k+1}\mathbf{v}_{k+1} - \lambda_{k+1}(t_1\mathbf{v}_1 + ... + t_k\mathbf{v}_k + t_{k+1}\mathbf{v}_{k+1}) = \mathbf{0} \iff (\lambda_1 - \lambda_{k+1})t_1\mathbf{v}_1 +$

$... + (\lambda_k - \lambda_{k+1})\, t_k \mathbf{v}_k = \mathbf{0}$. Per ipotesi induttiva $\mathbf{v}_1, ..., \mathbf{v}_k$ sono indipendenti, quindi $(\lambda_1 - \lambda_{k+1})\, t_1 = ... = (\lambda_k - \lambda_{k+1})\, t_k = 0$, ma gli autovalori sono tutti distinti, quindi può essere solo $t_1 = ... = t_k = 0$. Allora, se $t_1\mathbf{v}_1 + ... + t_k\mathbf{v}_k + t_{k+1}\mathbf{v}_{k+1} = \mathbf{0}$ e $t_1 = ... = t_k = 0$, per la legge di annullamento del prodotto, essendo $\mathbf{v}_{k+1} \neq \mathbf{0}$, $t_{k+1} = 0$.

4) Dimostro la proposizione

Hp:	B_1 è una base dell'autospazio V_1,..., B_s è una base di V_s
Ts:	$B_1 \cup ... \cup B_s$ è un insieme di vettori di A linearmente indipendenti

cioè che posso costruire un insieme di autovettori di A linearmente indipendenti con cardinalità $g(\lambda_1) + g(\lambda_2) + ... + g(\lambda_s)$. Questo si può fare sempre; in questo caso significherà che si può costruire un insieme di n autovettori linearmente indipendenti, che costituiranno la base di $\mathbf{K}^n$ cercata. Considero V_1 e la sua base di cardinalità $g(\lambda_1)$ $B_1 = \left\{\mathbf{w}_{11}, ..., \mathbf{w}_{1g(\lambda_1)}\right\}$,..., V_s e la sua base di cardinalità $g(\lambda_s)$ $B_s = \left\{\mathbf{w}_{s1}, ..., \mathbf{w}_{sg(\lambda_s)}\right\}$. Voglio mostrare che $B_1 \cup ... \cup B_s$ è un insieme di vettori linearmente indipendenti, cioè che $t_{11}\mathbf{w}_{11} + ... + t_{1g(\lambda_1)}\mathbf{w}_{1g(\lambda_1)} + ... + t_{s1}\mathbf{w}_{s1} + ... + t_{sg(\lambda_s)}\mathbf{w}_{sg(\lambda_s)} = \mathbf{0}$ implica $t_{11} = ... = t_{1g(\lambda_1)} = ... = t_{s1} = ... = t_{sg(\lambda_s)} = 0$. Pongo $\mathbf{v}_i = t_{i1}\mathbf{w}_{i1} + ... + t_{ig(\lambda_i)}\mathbf{w}_{ig(\lambda_i)}$, che è una generica combinazione lineare dei vettori della base di V_i, quindi un generico vettore dell'autospazio V_i. L'equazione precedente diventa $\mathbf{v}_1 + ... + \mathbf{v}_s = \mathbf{0}$. Ma $\mathbf{v}_1, ..., \mathbf{v}_s$ sono s vettori provenienti da s autospazi diversi, quindi sono linearmente indipendenti per quanto visto al punto 3. Dunque $\mathbf{v}_1 + ... + \mathbf{v}_s = \mathbf{0}$ può essere vera se e solo se $\mathbf{v}_1, ..., \mathbf{v}_s$ sono tutti nulli (altrimenti si potrebbe scrivere una relazione di dipendenza lineare tra di loro), perciò $\mathbf{v}_1 = ... = \mathbf{v}_s = \mathbf{0}$. Perciò per ogni i $\mathbf{v}_i = t_{i1}\mathbf{w}_{i1} + ... + t_{ig(\lambda_i)}\mathbf{w}_{ig(\lambda_i)} = \mathbf{0}$: ma $\left\{\mathbf{w}_{i1}, ..., \mathbf{w}_{ig(\lambda_i)}\right\}$ costituiscono una base di V_i, quindi sono linearmente indipendenti e $t_{i1} = ... = t_{ig(\lambda_i)} = 0$ per ogni $i = 1, ..., s$.

Si è quindi mostrato che se per ogni autospazio costruisco una base e metto insieme tutti i vettori di tutte le basi, questi vettori sono linearmente indipendenti.

In conclusione, poiché posso costruire un insieme di n autovettori di A linearmente indipendenti, questi costituiscono una base di $\mathbf{K}^n$ e A è diagonalizzabile per il primo criterio di diagonalizzabilità. ■

Nel punto (ii) si è visto che λ_1, λ_2 sono autovalori distinti allora $V_{\lambda_1} \cap V_{\lambda_2} = \{\mathbf{0}\}$ e $V_{\lambda_1} + V_{\lambda_2} = V_{\lambda_1} \oplus V_{\lambda_2}$. La somma è diretta qualunque numero di autospazi si consideri. Infatti, dati $\lambda_1, \lambda_2, \lambda_3$ tutti distinti, $V_{\lambda_1} + V_{\lambda_2} + V_{\lambda_3} = (V_{\lambda_1} + V_{\lambda_2}) + V_{\lambda_3} = (V_{\lambda_1} \oplus V_{\lambda_2}) + V_{\lambda_3}$. Per quanto visto in (iv) $\dim(V_{\lambda_1} + V_{\lambda_2} + V_{\lambda_3}) = g_{\lambda_1} + g_{\lambda_2} + g_{\lambda_3}$; $\dim(V_{\lambda_1} \oplus V_{\lambda_2}) = g_{\lambda_1} + g_{\lambda_2}$. Per la formula di Grassman, considerando $A = V_{\lambda_1} \oplus V_{\lambda_2}$ e $B = V_{\lambda_3}$, $\dim((V_{\lambda_1} \oplus V_{\lambda_2}) \cap V_{\lambda_3}) = g_{\lambda_1} + g_{\lambda_2} + g_{\lambda_3} - (g_{\lambda_1} + g_{\lambda_2}) - g_{\lambda_3} = 0$, quindi $V_{\lambda_1} + V_{\lambda_2} + V_{\lambda_3} = V_{\lambda_1} \oplus V_{\lambda_2} \oplus V_{\lambda_3}$.

Dunque, se A è diagonalizzabile su $\mathbf{K}$, $\dim(V_{\lambda_1} + V_{\lambda_2} + ... + V_{\lambda_s}) = \dim(V_{\lambda_1} \oplus V_{\lambda_2} \oplus ... \oplus V_{\lambda_s}) = n$ e $\mathbf{K}^n$ si può decomporre in sottospazi invarianti, cioè si può decomporre in sottospazi

$$\mathbf{K}^n = V_{\lambda_1} \oplus V_{\lambda_2} \oplus ... \oplus V_{\lambda_s}$$

che sono invarianti per $_A$ perché per ogni sottospazio V_{λ_i}, per ogni $\mathbf{v} \in V_{\lambda_i}$, vale $A\mathbf{v} = \lambda_i \mathbf{v} \in V_{\lambda_i}$: V_{λ_i} è invariante rispetto all'azione dell'applicazione lineare $_A$.

Quindi per applicazioni lineari la ricerca di autovalori e autovettori coincide con la ricerca dei sottospazi invarianti.

1. Se $A = P^{-1}DP$ con D diagonale, si dice che P diagonalizza A. Esistono infinite matrici che diagonalizzano A, perché ogni matrice le cui colonne siano una base di autovettori di A la diagonalizzano (una volta trovata una base, posso prenderne qualsiasi multiplo).

2. Se A è una matrice tale che la somma degli elementi su ogni colonna i è uguale ad a, allora a è autovalore di A. Infatti in tal caso A^T è tale che la somma degli elementi su ciascuna riga è a: allora $A^T \begin{pmatrix} 1 \\ ... \\ 1 \end{pmatrix} = C_1(A) + ... + C_n(A) = \begin{pmatrix} a \\ ... \\ a \end{pmatrix}$. Poiché A e A^T hanno lo stesso polinomio caratteristico, a è autovalore di A.

Teorema (proprietà invarianti per similitudine)

Hp: A, B sono matrici simili

Ts: (i) A, B hanno lo stesso polinomio caratteristico: $p_A(x) = p_B(x)$

(ii) A, B hanno la stessa traccia e lo stesso determinante: $\det A = \det B$, $tr(A) = tr(B)$

(iii) A, B hanno gli stessi autovalori con le stesse molteplicità algebriche e geometriche

(iv) A, B hanno lo stesso rango

Dim Per ipotesi $\exists\ P$ invertibile tale che $B = P^{-1}AP$.

(i) $p_B(x) = \det(B - xId)$. Scrivo $B - xId = P^{-1}AP - x\left(P^{-1}P\right) = P^{-1}AP - P^{-1}(xId)P = P^{-1}(A - xId)P$. Allora $\det(B - xId) = \det\left(P^{-1}(A - xId)P\right) = \det P^{-1} \det(A - xId) \det P = \det P^{-1}P \det(A - xId) = \det(A - xId) = p_A(x)$, usando due volte il teorema di Binet, quindi i due polinomi caratteristici coincidono.

(ii) $\det B = \det\left(P^{-1}AP\right) = \det\left(P^{-1}P\right) \det A = \det A$, usando due volte il teorema di Binet. Si può dimostrare anche sfruttando l'uguaglianza dei polinomi caratteristici: $\det B = p_B(0) = p_A(0) = \det A$.

$tr\left(B\right)$ è il coefficiente del termine di grado $n-1$ di $p_B\left(x\right)$: $p_B\left(x\right)=\left(-1\right)^n x^n+\left(-1\right)^{n-1} tr\left(B\right)x^{n-1}+...+\det B$. Analogamente $p_A\left(x\right)=\left(-1\right)^n x^n+\left(-1\right)^{n-1} tr\left(A\right)x^{n-1}+...+c_1x+\det A$. Poiché $p_A\left(x\right)=p_B\left(x\right)$, i coefficienti devono coincidere uno a uno, e in particolare $tr\left(A\right)=tr\left(B\right)$.

(iii) Poiché $p_A\left(x\right)=p_B\left(x\right)$, i due polinomi hanno le stesse radici (e quindi autovalori) con la stessa molteplicità algebrica. Mostro che ogni autovalore ha la stessa molteplicità geometrica sia come autovalore di A che come autovalore di B. Se $\mathbf{v}$ è autovettore di A con autovalore λ, allora $P^{-1}\mathbf{v}$ è autovettore di B con autovalore λ (P^{-1} è la matrice del cambio di base da A a B: $P^{-1}=M_A^B\left(id\right)$): infatti $B\left(P^{-1}\mathbf{v}\right)=\left(BP^{-1}\right)\mathbf{v}=\left(P^{-1}A\right)\mathbf{v}=P^{-1}\left(A\mathbf{v}\right)=P^{-1}\lambda\mathbf{v}=\lambda\left(P^{-1}\mathbf{v}\right)$, sfruttando il fatto che $B=P^{-1}AP\iff BP^{-1}=P^{-1}A$. B non può avere autovettori che A non ha, perché se $\mathbf{v}$ è autovettore di B, $P\mathbf{v}$ è autovettore di A. Quindi, se $\left\{\mathbf{v}_1,...,\mathbf{v}_{g_\lambda}\right\}$ è una base dell'autospazio V_λ di A, $\left\{P^{-1}\mathbf{v}_1,...,P^{-1}\mathbf{v}_{g_\lambda}\right\}$ è una base dell'autospazio V_λ di B (hanno entrambi la stessa dimensione): $P^{-1}\mathbf{v}_1,...,P^{-1}\mathbf{v}_{g_\lambda}$ sono linearmente indipendenti perché $P^{-1}\mathbf{x}$ è un'applicazione lineare iniettiva (è la matrice del cambio di base da A a B), quindi associa vettori linearmente indipendenti a vettori linearmente indipendenti. Dunque la molteplicità geometrica dell'autovalore λ è la stessa sia come autovalore di A che di B, e questo vale per ogni autovalore.

(iv) Due matrici simili rappresentano la stessa applicazione lineare. $r\left(B\right)=\dim\left(Im\left(_B\right)\right)$ e $r\left(A\right)=\dim\left(Im\left(_A\right)\right)$, ma essendo l'applicazione lineare la stessa, la dimensione dell'immagine dev'essere uguale, cioè $\dim\left(Im\left(_B\right)\right)$ $=\dim\left(Im\left(_A\right)\right)$, quindi $r\left(A\right)=r\left(B\right)$. ■

Questo teorema permette di capire che si possono avere informazioni sugli autovalori di una matrice anche senza calcolarli direttamente. Data $A\in M_{\mathbf{K}}\left(n,n\right)$, A ha n autovalori complessi $\lambda_1,\lambda_2,...,\lambda_n$, e - consderando gli autovalori con molteplicità - vale $\det A=\lambda_1\lambda_2...\lambda_n$ e $tr\left(A\right)=\lambda_1+\lambda_2+...+\lambda_n$. Si dimostra facilmente se A è diagonalizzabile: considero la sua matrice diagonale D (A,D sono simili), $\det D=\lambda_1\lambda_2...\lambda_n$; ma $\det D=\det A$, quindi $\det A=\lambda_1\lambda_2...\lambda_n$; analogamente $tr\left(D\right)=\lambda_1+\lambda_2+...+\lambda_n$, ma $tr\left(A\right)=tr\left(D\right)$, quindi $tr\left(A\right)=\lambda_1+\lambda_2+...+\lambda_n$. Perciò posso ricavare informazioni direttamente da A sui suoi autovalori anche se A non è diagonalizzabile.

1. $A\in M_{\mathbf{K}}\left(2,2\right)$. $p_A\left(x\right)=x^2-tr\left(A\right)x+\det A$ ha due radici, eventualmente complesse. λ_1,λ_2 sono le soluzioni, e coerentemente con le relazioni note tra le soluzioni di un'equazione di secondo grado ($x_1+x_2=\frac{-b}{a}$, $x_1x_2=\frac{c}{a}$), $tr\left(A\right)=\lambda_1+\lambda_2$, $\det A=\lambda_1\lambda_2$. Infatti, se $A=\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $p_A\left(x\right)=\det\left(A-xId\right)=\det\begin{bmatrix} a-x & b \\ c & d-x \end{bmatrix}=\left(a-x\right)\left(d-x\right)-bc=x^2-\left(a+d\right)x+ad-bc=\left(-1\right)^2x^2+\left(-1\right)^1 tr\left(A\right)x+\det A$.

2. $A\in M_{\mathbb{R}}\left(n,n\right)$ ha ogni elemento uguale a 1. Per cercarne gli autovalori è in questo caso poco comodo usare il polinomio caratteristico, essendo n generico. Osservo che $r\left(A\right)=1$, quindi $\dim\left(\ker A\right)=n-1$. $\ker\left(A\right)$

è l'autospazio relativo all'autovalore $\lambda = 0$, quindi l'autovalore nullo ha molteplicità geometrica $g_0 = n-1$. Essendo $a_0 \geq g_0 = n-1$, sappiamo che 0 è una radice del polinomio caratteristico con molteplicità algebrica almeno pari a $n-1$. L'n-esima radice potrebbe essere 0 oppure no (comunque è reale perché le radici complesse sono sempre a coppie). Per cercare l'ultima radice, che è un autovalore, uso il fatto che $tr\left(A\right) = n = \lambda_1 + ... + \lambda_n = 0 + \lambda_n$, quindi $\lambda_n = n$; $a_{\lambda_n} = 1 = g_{\lambda_n}$, autovalore semplice e regolare. Anche 0 è un autovalore regolare. Dunque A è diagonalizzabile è la matrice diagonale è $D = \begin{bmatrix} n & 0 & ... & 0 \\ 0 & 0 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 0 \end{bmatrix}$. In questo caso l'autovalore n ha come autospazio l'immagine dell'applicazione lineare rappresentata da A: se $\mathbf{x}$ è il generico vettore dell'immagine, $\mathbf{x} = x_1 \begin{pmatrix} 1 \\ ... \\ 1 \end{pmatrix} + ... + x_n \begin{pmatrix} 1 \\ ... \\ 1 \end{pmatrix} = \begin{pmatrix} x_1 + ... + x_n \\ ... \\ x_1 + ... + x_n \end{pmatrix}$ e $A \begin{pmatrix} x_1 + ... + x_n \\ ... \\ x_1 + ... + x_n \end{pmatrix} = \begin{pmatrix} n\left(x_1 + ... + x_n\right) \\ ... \\ n\left(x_1 + ... + x_n\right) \end{pmatrix} = n\mathbf{x}$.

Se A è diagonalizzabile, per (iv) $r\left(A\right)$ è il numero di autovalori non nulli di A. Infatti A, D hanno lo stesso rango, e $r\left(D\right)$, dato che D ha gli autovalori sulla diagonale, è il numero di autovalori non nulli, ciascuno contato con la sua molteplicità algebrica. Oppure si può pensare che A diagonalizzabile implica $a\left(\lambda_1\right) + a\left(\lambda_2\right) + ... + a\left(\lambda_s\right) = n$ e $a\left(\lambda_i\right) = g\left(\lambda_i\right)$. $a\left(0\right) = g\left(0\right)$ è la dimensione di $\ker\left(A\right)$, che è $n - r\left(A\right)$, e la somma delle molteplicità algebriche degli autovalori non nulli dev'essere $r\left(A\right)$.

1. Considero la successione di Fibonacci $\{F_n\}$, definita per ricorrenza con $a_0 = 0$, $a_1 = 1$, $a_{n+2} = a_n + a_{n+1}$, e il vettore $\mathbf{v}_n = \begin{pmatrix} a_n \\ a_{n+1} \end{pmatrix}$. Voglio scrivere in forma chiusa, per ogni n, $\mathbf{v}_n$, mostrare che $a_n = \frac{\phi_1^n - \phi_2^n}{\sqrt{5}}$, con $\phi_1 = \frac{1+\sqrt{5}}{2}$ e $\phi_2 = 1 - \phi_1$, e che $\lim_{n\to+\infty} \frac{a_{n+1}}{a_n} = \phi_1$, la sezione aurea.

   $\mathbf{v}_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Voglio scrivere $\mathbf{v}_n$ usando una matrice: $\mathbf{v}_n = \begin{pmatrix} a_n \\ a_{n+1} \end{pmatrix} = \begin{pmatrix} a_n \\ a_{n-1} + a_n \end{pmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a_{n-1} \\ a_n \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \mathbf{v}_{n-1}$. La definizione per ricorrenza della successione è $\mathbf{v}_n = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \mathbf{v}_{n-1} = A\mathbf{v}_{n-1}$. Essendo $\mathbf{v}_{n-1} = A\mathbf{v}_{n-2}$ e così via, si ha $\mathbf{v}_n = A^n\mathbf{v}_0$: per calcolare $\mathbf{v}_n$ occorre calcolare A^n: in questo è utile la diagonalizzazione. Infatti, se A è diagonalizzabile e $D = P^{-1}AP$, $A = PDP^{-1}$ e $A^n = \left(PDP^{-1}\right)\left(PDP^{-1}\right)....\left(PDP^{-1}\right) =$

$PD\left(P^{-1}P\right)D\left(P^{-1}...\right)...(...P)\,DP^{-1} = PD^nP^{-1}$. Se $D = \begin{bmatrix} d_1 & ... & 0 \\ ... & ... & ... \\ 0 & ... & d_n \end{bmatrix}$, vale $D^n = \begin{bmatrix} d_1^n & ... & 0 \\ ... & ... & ... \\ 0 & ... & d_n^n \end{bmatrix}$, quindi è molto pù semplice da calcolare rispetto ad A^n. Quindi mi chiedo se A è diagonalizzabile e calcolo eventualmente D. $\det(A-\lambda Id) = \det \begin{bmatrix} -\lambda & 1 \\ 1 & 1-\lambda \end{bmatrix} = (-\lambda)(1-\lambda) - 1 = \lambda^2 - \lambda - 1$ (potevo anche notare $tr(A) = 1$, $\det A = -1$). $\lambda_1 = \phi_1 = \frac{1+\sqrt{5}}{2}$, $\lambda_2 = \phi_2 = \frac{1-\sqrt{5}}{2}$: sono due autovalori semplici e regolari, quindi A è diagonalizzabile con matrice diagonale $D = \begin{bmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{bmatrix}$ e $D^n = \begin{bmatrix} \phi_1^n & 0 \\ 0 & \phi_2^n \end{bmatrix}$. Cerco una base di ciascun autospazio.

$V_{\phi_1} = \ker(A-\phi_1 Id) = \ker\left(\begin{bmatrix} -\phi_1 & 1 \\ 1 & 1-\phi_1 \end{bmatrix}\right)$: so già che il rango della matrice non è massimo, altrimenti l'unico vettore in V_{ϕ_1} sarebbe il vettore nullo (impossibile per definizione di autovettore). Perciò la matrice ridotta a scala dev'essere $\begin{bmatrix} -\phi_1 & 1 \\ 0 & 0 \end{bmatrix}$: risolvendo trovo che il generico vettore in V_{ϕ_1} è $\mathbf{x} = t\begin{pmatrix} 1 \\ \phi_1 \end{pmatrix}$.

$V_{\phi_2} = \ker(A-\phi_2 Id) = \ker\left(\begin{bmatrix} -\phi_2 & 1 \\ 1 & 1-\phi_2 \end{bmatrix}\right)$: la matrice ridotta a scala dev'essere $\begin{bmatrix} -\phi_2 & 1 \\ 0 & 0 \end{bmatrix}$; risolvendo trovo che il generico vettore in V_{ϕ_2} è $\mathbf{x} = t\begin{pmatrix} 1 \\ \phi_2 \end{pmatrix}$.

Quindi una base di K^2 formata da autovettori è $P = \begin{bmatrix} 1 & 1 \\ \phi_1 & \phi_2 \end{bmatrix}$; $P^{-1} = \frac{1}{\phi_2-\phi_1}\begin{bmatrix} \phi_2 & -1 \\ -\phi_1 & 1 \end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix} -\phi_2 & 1 \\ \phi_1 & -1 \end{bmatrix}$.

Calcolo $\mathbf{v}_n = A^n\mathbf{v}_0 = PD^nP^{-1}\mathbf{v}_0 = \begin{bmatrix} 1 & 1 \\ \phi_1 & \phi_2 \end{bmatrix}\begin{bmatrix} \phi_1^n & 0 \\ 0 & \phi_2^n \end{bmatrix}\frac{1}{\sqrt{5}}\begin{bmatrix} -\phi_2 & 1 \\ \phi_1 & -1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix} \phi_1^n & \phi_2^n \\ \phi_1^{n+1} & \phi_2^{n+1} \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix} \phi_1^n - \phi_2^n \\ \phi_1^{n+1} - \phi_2^{n+1} \end{bmatrix} = \begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix}$. Quindi $a_n = \frac{\phi_1^n - \phi_2^n}{\sqrt{5}}$. $\lim_{n\to+\infty} \frac{\phi_1^{n+1} - \phi_2^{n+1}}{\phi_1^n - \phi_2^n} = \frac{\phi_1^{n+1}}{\phi_1^n} = \phi_1$ perché $|\phi_2| < 1$.

4.6.1 Polinomi di matrici

Dato $P(t) \in \mathbf{K}[t]$, $P(t) = a_0 + a_1 t + ... + a_d t^d$, e $A \in M_{\mathbf{K}}(n,n)$, $P(A) = a_0 A^0 + a_1 A + ... + a_d A^d = a_0 Id + a_1 A + ... + a_d A^d$ è una matrice che si ottiene combinando linearmente le potenze di A.

Teo

$$\text{Hp: } A \in M_{\mathbf{K}}(n,n)\,,\ \mathbf{v} \text{ è un autovettore di } A \text{ con autovalore } \lambda$$

$$\text{Ts: (i) } \mathbf{v} \text{ è un autovettore di } A^k \text{ con autovalore } \lambda^k$$

$$\text{(ii) se } \det A \neq 0,\ \mathbf{v} \text{ è un autovettore di } A^{-1} \text{ con autovalore } \frac{1}{\lambda}$$

$$\text{(iii) se } P(t) = a_0 + a_1 t + ... + a_d t^d,\ \mathbf{v} \text{ è un autovettore di } P(A) \text{ con autovalore } P(\lambda)$$

$$\text{(iv) se } A, B \text{ sono simili, } P(A)\,, P(B) \text{ sono simili}$$

Dim (i) Se $\mathbf{v}$ è tale che $A\mathbf{v} = \lambda\mathbf{v}$, allora $A^2\mathbf{v} = A(A\mathbf{v}) = A(\lambda\mathbf{v}) = \lambda(A\mathbf{v}) = \lambda^2\mathbf{v}$. Questo vale in generale: per induzione $A^k\mathbf{v} = A\left(A^{k-1}\mathbf{v}\right) = A\left(\lambda^{k-1}\mathbf{v}\right) = \lambda^{k-1}(A\mathbf{v}) = \lambda^k\mathbf{v}$.

(ii) Se $\det A \neq 0$, essendo $\det A = \prod_{i=1}^n \lambda_i = \lambda_1\lambda_2...\lambda_n$, ogni autovalore è non nullo (infatti se non c'è l'autovalore nullo $r(A) = n$, coerentemente con $\det A \neq 0$). Inoltre $\exists\ A^{-1}$. $A\mathbf{v} = \lambda\mathbf{v} \Longleftrightarrow A^{-1}(A\mathbf{v}) = A^{-1}(\lambda\mathbf{v}) \Longleftrightarrow \mathbf{v} = \lambda\left(A^{-1}\mathbf{v}\right) \Longleftrightarrow \frac{1}{\lambda}\mathbf{v} = A^{-1}\mathbf{v}$. λ è invertibile sul campo $\mathbf{K}$ perché non nullo. Quindi $\mathbf{v}$ è un autovettore di A^{-1} con autovalore $\frac{1}{\lambda}$.

(iii) $P(t) = a_0 + a_1 t + ... + a_d t^d$. $P(A) = a_0 Id + a_1 A + ... + a_d A^d$. $A\mathbf{v} = \lambda\mathbf{v}$. Voglio mostrare che $P(A)\mathbf{v} = P(\lambda)\mathbf{v}$. Quindi $P(A)\mathbf{v} = \left(a_0 Id + a_1 A + ... + a_d A^d\right)\mathbf{v} = a_0 Id\mathbf{v} + a_1(A\mathbf{v}) + ... + a_d\left(A^d\mathbf{v}\right)$ per proprietà distributiva del prodotto matriciale. Ottengo quindi $a_0\mathbf{v} + a_1(\lambda\mathbf{v}) + ... + a_d\left(\lambda^d\mathbf{v}\right)$ (utilizzando anche il punto (i)), che si riscrive, fattorizzando rispetto a $\mathbf{v}$, come $\left(a_0 + a_1\lambda + ... + a_d\lambda^k\right)\mathbf{v} = P(\lambda)\mathbf{v}$.

(iv) $B = S^{-1}AS$. Voglio mostrare che esiste una matrice invertibile tale che $P(B) = C^{-1}P(A)C$. $P(B) = a_0 Id + a_1 B + ... + a_d B^d = a_0\left(S^{-1}S\right) + a_1\left(S^{-1}AS\right) + ... + a_d\left(S^{-1}A^dS\right) =$ come prima, $a_0\left(S^{-1}IdS\right) + a_1\left(S^{-1}AS\right) + ... + a_d\left(S^{-1}A^dS\right) = S^{-1}\left(a_0 Id + a_1 A + ... + a_d A^d\right)S = S^{-1}P(A)S$. Quindi non solo esiste una matrice invertibile tale che $P(B) = C^{-1}P(A)C$, ma tale matrice è anche la stessa che stabilisce la similitudine tra A e B. ■

Quindi, se A è diagonalizzabile, per (iv) anche $P(A)$ è diagonalizzabile. Infatti, se si prende come B una matrice diagonale cui A è simile, si ha che $P(A)$ è simile a $P(D)$, cioè $P(A)$ è diagonalizzabile con gli autovalori $P(\lambda_1)\,, ..., P(\lambda_n)$.

1. Se $\lambda = 1$ è un autovalore di A con una certa molteplicità algebrica, non necessariamente $\lambda = 1$ è autovalore di A^k con la stessa molteplicità algebrica. Infatti, se A ha anche un autovalore -1 e k è pari, $V_1\left(A^k\right) = V_{-1}(A) \oplus V_1(A)$ e $a_1\left(A^k\right) = a_1(A) + a_{-1}(A)$.

In generale si ha $V_\lambda(A) \subseteq V_{\lambda^k}\left(A^k\right)$; l'inclusione può essere propria anche nel caso di matrici diagonalizzabili che abbiano tra i loro autovalori due autovalori opposti.

1. $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ è tale che $V_1(A) \subset V_1(A^2) = \mathbb{R}^2$.

Se A non è diagonalizzabile, $P(A)$ potrebbe invece essere diagonalizzabile; in generale, potrebbe avere autovettori che A non ha.

1. La matrice $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ non è diagonalizzabile e ha come base di V_0 $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$. $A^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ è invece diagonalizzabile e ha come base di V_0 $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$. In generale, se A non è diagonalizzabile ed è nilpotente, allora $\exists\ k : A^k = 0_M$, e A^k è diagonalizzabile. In tal caso $V_0(A) \subseteq V_0(A^k)$ diventa $V_0(A) \subset V_0(A^k)$.

2. A è nilpotente (cioè $\exists\ k : A^k = 0_M$) se e solo se tutti gli autovalori di A sono nulli. Infatti, se A è nilpotente, esiste $k : A^k = 0_M$, ma se λ è un qualsiasi autovalori di A vale $A\mathbf{v} = \lambda\mathbf{v}$, quindi $A^k v = A^{k-1}\lambda\mathbf{v} = 0_M$, da cui $\lambda = 0$.

Inoltre, se A è simmetrica, anche $P(A)$ è simmetrica, perché sia A che $P(A)$ sono ortogonalmente diagonalizzabili.

Se A ha autovalori non nulli, A^k ha autovalori non nulli.

Teorema (Cayley-Hamilton)

Hp: $A \in M_{\mathbf{K}}(n,n)\,,\ p_A(x) = \det(A - xId)$

Ts: A è radice del proprio polinomio caratteristico: $p_A(A) = 0_M$

Questo teorema mostra che se A ha tutti autovalori nulli (per cui $\det(A - \lambda Id) = k\lambda^n$), allora A è nilpotente: infatti $kA^n = 0_M$.

Dim Si dimostra nel caso particolare di A diagonalizzabile. Considero $B = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ base diagonalizzante (di autovettori): $A\mathbf{v}_i = \lambda_i\mathbf{v}_i\ \forall\ i = 1, ..., n$. $p_A(A)\mathbf{v}_i = p_A(\lambda_i)\mathbf{v}_i$ (si applica (iii) nel caso in cui il polinomio considerato è il polinomio caratteristico) $= \mathbf{0}$ perché $p_A(\lambda_i) = 0,\ \forall\ i = 1, ..., n$. Quindi $\mathbf{v}_i \in \ker(p_A(A))\ \forall\ i = 1, ..., n$ e $\ker(P_A(A)) = \mathbf{K}^n$, perché il nucleo dell'applicazione lineare $P_A(A)\mathbf{x}$ è un sottospazio di $\mathbf{K}^n$, ma se contiene n vettori linearmente indipendenti (è il massimo numero di vettori linearmente indipendenti: non posso averne di più perché $\dim(\ker(P_A(A))) \leq n$, quindi sono anche generatori, altrimenti potrei completarli a una base aggiungendo altri vettori linearmente indipendenti), anche $\mathbf{K}^n$ contiene quegli n vettori linearmente indipendenti, che costituiscono una base di entrambi gli spazi. Avendo la stessa dimensione ed essendo $\ker(P_A(A)) \subseteq \mathbf{K}^n$, i due spazi coincidono. Affinché sia $\dim\ker(P_A(A)) = n$, dev'essere $r(P_A(A)) = 0$: l'unica matrice che ha rango nullo è la matrice nulla, quindi $P_A(A) = \begin{bmatrix} 0 & ... & 0 \\ ... & ... & ... \\ 0 & ... & 0 \end{bmatrix}$. ■

Questo teorema si applica nel calcolo della matrice inversa. Considero A invertibile: allora $\det A \neq 0$. $p_A(x) = (-1)^n x^n + (-1)^{n-1} tr(A) x^{n-1} + ... + (\det A) x^0$. $p_A(A) = (-1)^n A^n + (-1)^{n-1} tr(A) A^{n-1} + ... + (\det A) Id = 0_M$ per il teorema visto. Postmoltiplico per A^{-1}: $(-1)^n A^{n-1} + (-1)^{n-1} tr(A) A^{n-2} + ... + a_1 AA^{-1} + (\det A) IdA^{-1} = 0_M$. Quindi $A^{-1} = \frac{(-1)^{n+1}A^{n-1} + (-1)^n tr(A) A^{n-2} + ... - a_1 Id}{\det A}$: si può calcolare la matrice inversa calcolando varie potenze di A. Può essere conveniente quando $p_A(x)$ ha molti termini nulli.

1. $A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$. $p_A(x) = \det \begin{bmatrix} 3-x & 1 \\ 0 & 2-x \end{bmatrix} = (3-x)(2-x) = x^2 - 5x + 6$: $\det A = 6$, $tr(A) = 5$.

   $p_A(A) = A^2 - 5A + 6 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ per il teorema di Cayley-Hamilton. Premoltiplicando per A^{-1}, si ottiene $A - 5Id + 6A^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, dunque $A^{-1} = \frac{5Id - A}{6} = \frac{1}{6}\begin{bmatrix} 2 & -1 \\ 0 & 3 \end{bmatrix}$, che è coerente con la formula vista.

4.7 Forma canonica di Jordan e pseudodiagonalizzazione

I teoremi visti sulla diagonalizzabilità permettono di stabilire che, se A è diagonalizzabile, il rappresentante canonico della classe di equivalenza di matrici simili ad A (quindi la matrice migliore, tra tutte quelle che rappresentano la stessa applicazione lineare) è la matrice diagonale con gli autovalori della matrice, ripetuti con la loro molteplicità algebrica, sulla diagonale.

(1) Cosa succede se A è quadrata e non diagonalizzabile? Come si trova un rappresentante della classe di equivalenza di matrici simili ad A, cioè come si sceglie la matrice migliore per rappresentare l'applicazione lineare?

Posso scegliere $\mathbf{K} = \mathbb{C}$ in modo che la prima ipotesi del secondo criterio di diagonalizzabilità sia sempre soddisfatta: data $A \in M_{\mathbb{C}}(n,n)$, $\lambda_1, ..., \lambda_s$ sono autovalori complessi tali che $a_{\lambda_1} + ... + a_{\lambda_s} = n$.

Se, data $A \in M_{\mathbb{C}}(n,n)$, $a_{\lambda_1} + ... + a_{\lambda_s} = n$, allora A è simile a una matrice complessa $B \in M_{\mathbb{C}}(n,n)$, detta matrice a blocchi di Jordan, del tipo $\begin{bmatrix} B_{\lambda_1} & 0 & ... & 0 \\ 0 & B_{\lambda_2} & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & B_{\lambda_s} \end{bmatrix}$, dove $B_{\lambda_1}, ..., B_{\lambda_s}$ sono matrici triangolari superiori $B_{\lambda_i} \in$

$M_{\mathbb{C}}(a_{\lambda_i}, a_{\lambda_i})$, con λ_i sulla diagonale principale, tali che $r(B_{\lambda_i} - \lambda_i Id) = a_{\lambda_i} - g_{\lambda_i}$ ($g_{\lambda_i} = \dim(\ker(B_{\lambda_i} - \lambda_i Id)) = a_{\lambda_i} - r(B_{\lambda_i} - \lambda_i Id)$: ordine della matrice meno rango della matrice). Se $a_{\lambda_i} = g_{\lambda_i}$, $r(B_{\lambda_i} - \lambda_i Id) = 0$, cioè $B_{\lambda_i} - \lambda_i Id = 0_M \iff B_{\lambda_i} = \lambda_i Id$, cioè B_{λ_i} è diagonale. Ogni matrice B_{λ_i} ha come unico autovalore λ_i, con molteplicità $a(\lambda_i)$.

Questa è una struttura "abbastanza simile" a quella di una matrice diagonale.

1. Si è visto, prima del secondo criterio di diagonalizzabilità, un esempio di matrice $A = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ non diagonalizzabile né su $\mathbb{R}$ né su $\mathbb{C}$ il cui polinomio caratteristico aveva molteplicità algebrica della radice $\lambda = 0$ $a_0 = 4$ e molteplicità geometrica $\dim(V_0) = g_0 = 2$. A non è diagonalizzabile, ma è (in generale, E' SIMILE A?) una matrice di Jordan con blocco triangolare superiore $B_0 \in M_{\mathbb{C}}(4,4)$, con gli autovalori 0 sulla diagonale e tale che $r(B_0) = 4 - 2 = 2$.

2. $A \in M_{\mathbb{C}}(3,3)$ ha due autovalori $\lambda_1, \lambda_2 : a_{\lambda_1} = 1 = g_{\lambda_1}$, ma $a_{\lambda_2} = 2 > 1 = g_{\lambda_2}$. A non è diagonalizzabile, ma come matrice complessa è simile a una matrice formata da due blocchi triangolari superiori $B_{\lambda_1} \in M_{\mathbb{C}}(1,1), B_{\lambda_2} \in M_{\mathbb{C}}(2,2)$. $B_{\lambda_1} = [\lambda_1]$, $B_{\lambda_2} = \begin{bmatrix} \lambda_2 & ? \\ 0 & \lambda_2 \end{bmatrix}$: $r(B_{\lambda_2} - \lambda_2 Id) = a_{\lambda_2} - g_{\lambda_2} = 1$. In ? non può esserci 0 (altrimenti A sarebbe diagonalizzabile, contro l'ipotesi); dev'essere $r(B_{\lambda_2} - \lambda_2 Id) = 1$, cioè $r\left(\begin{bmatrix} \lambda_2 - \lambda_2 & ? \\ 0 & \lambda_2 - \lambda_2 \end{bmatrix}\right) = 1$, quindi per avere rango uno è sufficiente che in ? ci sia un qualsiasi elemento non nullo, e. g. 1. Perciò $B = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 1 \\ 0 & 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} B_{\lambda_1} & 0 \\ 0 & B_{\lambda_2} \end{bmatrix}$.

(2) Cosa succede se A non è neanche quadrata? La diagonalizzazione di una matrice A serve a trovare, nella classe di equivalenza di matrici simili ad A (che rappresentano tutte le stessa applicazione lineare L), la matrice rappresentativa che semplifichi il più possibile i calcoli legati all'uso di L. Questa è la matrice diagonale, se A è diagonalizzabile. Ma se L non ha dominio e codominio isomorfi (cui quindi non si può applicare né la definizione di similitudine né di diagonalizzazione, essendo la matrice rappresentativa non quadrata), quale si sceglie come matrice rappresentativa di L che semplifichi i calcoli?

Si affronta quindi il caso di applicazione lineare $L : V \to W$, $\dim V = n$, $\dim W = m$.

Teorema di rappresentazione pseudodiagonale

Hp: $L : V \to W$ è un'applicazione lineare, $\dim V = n$, $\dim W = m$

Ts: $\exists$ una base B di V e una base C di W tali che $M_B^C(L) = \begin{bmatrix} 1 & 0 & 0 & ... & 0 \\ 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 0 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 0 \end{bmatrix}$

cioè la matrice rappresentativa non è diagonale (la definizione di matrice diagonale si applica a matrici quadrate), ma ha "molti zeri e un po' di uni sulla diagonale": è una matrice a blocchi $\begin{bmatrix} Id & 0 \\ 0 & 0 \end{bmatrix}$, con Id di dimensione $r(A)$ (come si vedrà nella dimostrazione).

Dim Considero una base dell'immagine $\{\mathbf{w}_1, ..., \mathbf{w}_r\}$, dove r è il rango di L (la dimensione dell'immagine): voglio completarla a una base di W: $C = \{\mathbf{w}_1, ..., \mathbf{w}_r\} \cup \{\mathbf{w}_{r+1}, ..., \mathbf{w}_m\}$. Considero $\{\mathbf{v}_1, ..., \mathbf{v}_r\}$ tali che $L(\mathbf{v}_1) = \mathbf{w}_1$,..., $L(\mathbf{v}_r) = \mathbf{w}_r$: essendo $\mathbf{w}_1, ..., \mathbf{w}_r$ linearmente indipendenti, mostro che $\mathbf{v}_1, ..., \mathbf{v}_r$ sono linearmente indipendenti (se fosse e. g. $\mathbf{v}_3 = \mathbf{v}_2 + \mathbf{v}_1$, si avrebbe $L(\mathbf{v}_3) = L(\mathbf{v}_1) + L(\mathbf{v}_2) = \mathbf{w}_1 + \mathbf{w}_2 = L(\mathbf{v}_3) = \mathbf{w}_3$, e $\mathbf{w}_1, ..., \mathbf{w}_r$ non sarebbero linearmente indipendenti). Infatti, considero $\lambda_1 \mathbf{v}_1 + ... + \lambda_r \mathbf{v}_r = \mathbf{0}$: $L(\lambda_1 \mathbf{v}_1 + ... + \lambda_r \mathbf{v}_r) = L(\mathbf{0}) \iff \lambda_1 \mathbf{w}_1 + ... + \lambda_r \mathbf{w}_r = \mathbf{0}$. Ma allora $\lambda_1 = ... = \lambda_r = 0$ perché $\mathbf{w}_1, ..., \mathbf{w}_r$ sono linearmente indipendenti. Considero una base di $\ker(L)$ $\{\mathbf{v}_{r+1}, ..., \mathbf{v}_n\}$: poiché $\mathbf{v}_1, ..., \mathbf{v}_r$ sono linearmente indipendenti, $Span\{\mathbf{v}_1, ..., \mathbf{v}_r\} \cap \ker(L) = \{\mathbf{0}\}$ $(L(\lambda_1 \mathbf{v}_1 + ... + \lambda_r \mathbf{v}_r) = \mathbf{0} \iff \lambda_1 \mathbf{w}_1 + ... + \lambda_r \mathbf{w}_r = \mathbf{0} \iff \lambda_1 = ... = \lambda_r = 0)$. Allora per la formula di Grassman $\dim(Span\{\mathbf{v}_1, ..., \mathbf{v}_r\} + \ker(L)) = \dim(Span\{\mathbf{v}_1, ..., \mathbf{v}_r\}) + \dim(\ker(L)) = r + n - r = n$ $(Span\{\mathbf{v}_1, ..., \mathbf{v}_r\} + \ker(L) = Span\{\mathbf{v}_1, ..., \mathbf{v}_r\} \oplus \ker(L))$, i. e. l'unione dei generatori di $Span\{\mathbf{v}_1, ..., \mathbf{v}_r\}$ e di $\ker(L)$ è un insieme di vettori linearmente indipendenti (altrimenti ci sarebbe un vettore non nullo nell'intersezione). Quindi $B =$

$\{\mathbf{v}_1, ..., \mathbf{v}_r\} \cup \{\mathbf{v}_{r+1}, ..., \mathbf{v}_n\}$ è una base di V. Posso costruire la matrice rappresentativa:

$$M_B^C(L) = \begin{bmatrix} L(\mathbf{v}_1) & L(\mathbf{v}_2) & ... & L(\mathbf{v}_r) & L(\mathbf{v}_{r+1}) & ... & L(\mathbf{v}_n) & \\ 1 & 0 & ... & 0 & 0 & ... & 0 & \mathbf{w}_1 \\ 0 & 1 & ... & 0 & 0 & ... & 0 & \mathbf{w}_2 \\ ... & ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & ... & 1 & 0 & ... & 0 & \mathbf{w}_r \\ 0 & 0 & ... & 0 & 0 & ... & 0 & \mathbf{w}_{r+1} \\ ... & ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & ... & 0 & 0 & ... & 0 & \mathbf{w}_n \end{bmatrix}$$

Non si possono avere più zeri: il rango di $M_B^C(L)$ dev'essere r. $\blacksquare$

La matrice rappresentativa pseudodiagonale è molto bella per fare i calcoli, ma c'è un lato negativo dovuto al fatto che non ci sono informazioni sulle proprietà di B, C rispetto a cui essa rappresenta L; se si usa la matrice rappresentativa di L rispetto alle basi canoniche, la matrice potrebbe non essere bella, ma le basi sono ben note. Quindi non è ovvio qual è la migliore scelta di rappresentazione di L.

5 Spazi euclidei

Il campo considerato d'ora in poi è $\mathbb{R}$.

Def Si dice spazio euclideo V uno spazio vettoriale sul campo $\mathbb{R}$ dotato di un prodotto scalare $\langle \_, \_ \rangle : V \times V \to \mathbb{R}$ che soddisfa le seguenti proprietà:

- commutatività: $\forall\ \mathbf{v}, \mathbf{w} \in V$, $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$
- bilinearità: $\forall\ \mathbf{v}, \mathbf{w}, \mathbf{u} \in V$, $\langle \mathbf{v} + \mathbf{w}, \mathbf{u} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{w}, \mathbf{u} \rangle$; $\forall\ \mathbf{v}, \mathbf{w}, \mathbf{u} \in V$, $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$; $\forall\ \mathbf{v}, \mathbf{w} \in V$, $\forall\ \lambda \in \mathbb{R}$, $\langle \lambda\mathbf{v}, \mathbf{w} \rangle = \lambda\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, \lambda\mathbf{w} \rangle$
- positività: $\forall\ \mathbf{v} \in V$, $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$; $\forall\ \mathbf{v} \in V$, $\langle \mathbf{v}, \mathbf{v} \rangle = 0 \Longleftrightarrow \mathbf{v} = \mathbf{0}$

1. Il prodotto scalare standard tra $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ e $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$, in $\mathbb{R}^2$, è $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v} \cdot \mathbf{w} = v_1 w_1 + v_2 w_2$. Il prodotto scalare standard tra $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$ e $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$, in $\mathbb{R}^3$, è $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v} \cdot \mathbf{w} = v_1 w_1 + v_2 w_2 + v_3 w_3$.

2. Il prodotto scalare standard tra $\mathbf{v} = \begin{pmatrix} v_1 \\ ... \\ v_n \end{pmatrix}$ e $\mathbf{w} = \begin{pmatrix} w_1 \\ ... \\ w_n \end{pmatrix}$, in $\mathbb{R}^n$, è $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + v_2 w_2 + ... + v_n w_n = \mathbf{v}^T \mathbf{w} = [v_1 | ... | v_n] \begin{pmatrix} w_1 \\ ... \\ w_n \end{pmatrix} = \sum_{i=1}^n v_i w_i$. La commutatività segue dalla commutatività in $\mathbb{R}$, la bilinearità si mostra con la proprietà distributiva del prodotto matriciale, ecc ($\langle \mathbf{v}, \mathbf{v} \rangle = v_1^2 + ... + v_n^2 \geq 0$).

3. Posto $I = [a, b] \subseteq \mathbb{R}$, $L^2(I) = \left\{ f : I \to \mathbb{R} : \int_a^b f^2(t)\, dt \text{ è finito} \right\}$ è lo spazio vettoriale delle funzioni quadrato-integrabili (come si mostra che $f + g$ è tale che $\int_a^b (f+g)^2\, dt = \int_a^b f^2(t)\, dt + \int_a^b g^2(t)\, dt + \int_a^b f(t)\, g(t)\, dt$: il terzo è finito?). Il prodotto scalare in $L^2(I)$ è definito come segue: $\forall\ f, g \in L^2(I)$, $\langle f, g \rangle = \int_a^b f(t)\, g(t)\, dt$ (si può vedere una funzione come un vettore con un'infinità non numerabile di componenti: dato t, $f(t)$ è una sua componente). $L^2(I)$ è quindi uno spazio euclideo dotato di tale prodotto scalare. La commutatività segue dalla commutatività del prodotto di funzioni, la bilinearità dalla linearità dell'integrazione definita, $\langle f, f \rangle = \int_a^b f^2(t)\, dt \geq 0$ per positività dell'integrale. Si mostra che $\langle f, f \rangle = 0 \iff f(t) = 0\ \forall\ t \in I$.

   $L^2(I)$ è un insieme fondamentale in analisi del segnale.

Def Dato uno spazio euclideo V dotato di un prodotto scalare $\langle \_, \_ \rangle$, si dice norma la funzione $||\_|| : V \to \mathbb{R}$ definita come $||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}\ \forall\ \mathbf{v} \in V$.

Ha le seguenti proprietà:

- positività e annullamento: $||\mathbf{v}|| \geq 0$; $||\mathbf{v}|| = 0 \iff \mathbf{v} = \mathbf{0}$
- omogeneità: $||t\mathbf{v}|| = |t|\, ||\mathbf{v}||$ (infatti $||t\mathbf{v}|| = \sqrt{\langle t\mathbf{v}, t\mathbf{v} \rangle} = \sqrt{t^2 \langle \mathbf{v}, \mathbf{v} \rangle} = |t| \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$).

Tale definizione ha senso perché la radice si può calcolare $\forall\ \mathbf{v}$, per positività del prodotto scalare. La norma è una generalizzazione del valore assoluto in $\mathbb{R}$: $\sqrt{x \cdot x} = |x|$, $|x| \geq 0$, ecc.

Teorema (disuguaglianza di Schwarz)

$$\text{Hp: } (V, \langle \_, \_ \rangle) \text{ è uno spazio euclideo}$$

$$\text{Ts: } \forall\ \mathbf{v}, \mathbf{w} \in V,\ |\langle \mathbf{v}, \mathbf{w} \rangle| \leq ||\mathbf{v}||\, ||\mathbf{w}||$$

E' una conseguenza naturale della struttura di prodotto scalare.

Dim Se $\mathbf{w} = \mathbf{0}$, $|\langle \mathbf{v}, \mathbf{0} \rangle| = |\langle \mathbf{v}, \mathbf{w} - \mathbf{w} \rangle| = |\langle \mathbf{v}, \mathbf{w} \rangle - \langle \mathbf{v}, \mathbf{w} \rangle| = 0$ (o per bilinearità) $\leq ||\mathbf{v}||\, ||\mathbf{0}|| = 0$. Se $\mathbf{w} \neq \mathbf{0}$, considero $\mathbf{v} + x\mathbf{w}$, $x \in \mathbb{R}$. $||\mathbf{v} + x\mathbf{w}||^2 = \langle \mathbf{v} + x\mathbf{w}, \mathbf{v} + x\mathbf{w} \rangle = \langle \mathbf{v} + x\mathbf{w}, \mathbf{v} \rangle + \langle \mathbf{v} + x\mathbf{w}, x\mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{v} \rangle + \langle x\mathbf{w}, \mathbf{v} \rangle + \langle \mathbf{v}, x\mathbf{w} \rangle + \langle x\mathbf{w}, x\mathbf{w} \rangle = ||\mathbf{v}||^2 +$

$2\langle \mathbf{v}, x\mathbf{w}\rangle + ||x\mathbf{w}||^2 = ||\mathbf{v}||^2 + 2\langle \mathbf{v}, x\mathbf{w}\rangle + x^2 ||\mathbf{w}||^2$ per commutatività e omogeneità del prodotto scalare. Poiché $||\mathbf{v}||^2 + 2\langle \mathbf{v}, x\mathbf{w}\rangle + x^2 ||\mathbf{w}||^2 = ||\mathbf{v} + x\mathbf{w}||^2 \geq 0$ per positività del prodotto scalare, $||\mathbf{v}||^2 + 2\langle \mathbf{v}, x\mathbf{w}\rangle + x^2 ||\mathbf{w}||^2 \geq 0$. Se si considera $||\mathbf{v}||^2 + 2\langle \mathbf{v}, x\mathbf{w}\rangle + x^2 ||\mathbf{w}||^2 = ||\mathbf{w}||^2 x^2 + 2\langle \mathbf{v}, \mathbf{w}\rangle x + ||\mathbf{v}||^2$ una funzione di x, $f(x) \geq 0 \; \forall \; x \in \mathbb{R} \Longleftrightarrow \Delta \leq 0$, i. e. $\Delta = 4\langle \mathbf{v}, \mathbf{w}\rangle^2 - 4\,||\mathbf{w}||^2\,||\mathbf{v}||^2 \leq 0 \Longleftrightarrow \langle \mathbf{v}, \mathbf{w}\rangle^2 \leq ||\mathbf{w}||^2\,||\mathbf{v}||^2 \Longleftrightarrow \sqrt{\langle \mathbf{v}, \mathbf{w}\rangle^2} \leq \sqrt{||\mathbf{w}||^2\,||\mathbf{v}||^2} \Longleftrightarrow |\langle \mathbf{v}, \mathbf{w}\rangle| \leq ||\mathbf{w}||\,||\mathbf{v}||$. ■

Dal teorema segue $|\langle \mathbf{v}, \mathbf{w}\rangle| \leq ||\mathbf{v}||\,||\mathbf{w}|| \Longleftrightarrow -\,||\mathbf{v}||\,||\mathbf{w}|| \leq \langle \mathbf{v}, \mathbf{w}\rangle \leq ||\mathbf{v}||\,||\mathbf{w}|| \Longleftrightarrow -1 \leq \frac{\langle \mathbf{v}, \mathbf{w}\rangle}{||\mathbf{v}||\,||\mathbf{w}||} \leq 1$. $\frac{\langle \mathbf{v}, \mathbf{w}\rangle}{||\mathbf{v}||\,||\mathbf{w}||}$ si può interpretare come coseno di un angolo.

Def Dati due vettori $\mathbf{v}, \mathbf{w} \neq \mathbf{0}$ in uno spazio euclideo, si definisce angolo tra $\mathbf{v}$ e $\mathbf{w}$ il numero reale, appartenente a $[0, \pi]$, $\hat{\mathbf{v}}\mathbf{w} = \arccos \frac{\langle \mathbf{v}, \mathbf{w}\rangle}{||\mathbf{v}||\,||\mathbf{w}||}$.

1. In $\mathbb{R}^2$ e $\mathbb{R}^3$ $\hat{\mathbf{v}}\mathbf{w} = \arccos \frac{\langle \mathbf{v}, \mathbf{w}\rangle}{||\mathbf{v}||\,||\mathbf{w}||} \Longleftrightarrow ||\mathbf{v}||\,||\mathbf{w}|| \cos \hat{\mathbf{v}}\mathbf{w} = \langle \mathbf{v}, \mathbf{w}\rangle$, e $\hat{\mathbf{v}}\mathbf{w}$ è l'angolo tra i due vettori secondo l'usuale interpretazione geometrica di angolo.

2. In $L^2(I)$ l'angolo tra le funzioni f e g è $\hat{f}g = \arccos \frac{\int_a^b f(t)g(t)dt}{\sqrt{\int_a^b f^2(t)dt}\sqrt{\int_a^b g^2(t)dt}}$.

Si fanno altri esempi di spazi euclidei.

1. Considero lo spazio euclideo $\mathbb{R}^n$ e una matrice simmetrica B (che è invariante per trasposizione). Dati due vettori $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, si definisce il prodotto tra $\mathbf{v}$ e $\mathbf{w}$ indotto dalla matrice B come $\langle \mathbf{v}, \mathbf{w}\rangle_B = \mathbf{v}^T B\mathbf{w}$. E' commutativo perché B è simmetrica, bilineare per la proprietà distributiva del prodotto matriciale. Affinché sia un prodotto scalare deve valere $\langle \mathbf{v}, \mathbf{v}\rangle_B = \mathbf{v}^T B\mathbf{v} \geq \mathbf{0}$ e $\langle \mathbf{v}, \mathbf{v}\rangle_B = \mathbf{0} \Longleftrightarrow \mathbf{v} = \mathbf{0}$: impareremo come capire se vale ciò, usando gli autovalori di B (B dev'essere definita positiva). Tale prodotto scalare, quando $B = Id$, si dice prodotto scalare standard di $\mathbb{R}^n$ ed è quello usato finora: se $\mathbf{v} = \begin{bmatrix} v_1 \\ ... \\ v_n \end{bmatrix}$ e $\mathbf{w} = \begin{pmatrix} w_1 \\ ... \\ w_n \end{pmatrix}$, $\langle \mathbf{v}, \mathbf{w}\rangle_{Id} = \mathbf{v}^T Id\mathbf{w} = \mathbf{v}^T\mathbf{w} = [v_1|...|v_n] \begin{bmatrix} w_1 \\ ... \\ w_n \end{bmatrix}$. Quindi $\mathbb{R}^n$ è uno spazio euclideo rispetto a entrambe tali definizioni di prodotto scalare: la seconda è una generalizzazione della prima.

2. Considero lo spazio euclideo $M_{\mathbb{R}}(m, n)$ e definisco $\langle A, B\rangle = tr\left(B^T A\right)$ (la trasposizione è necessaria per poter fare il prodotto). La bilinearità segue dalla linearità della trasposizione, dalla proprietà distributiva e associativa del prodotto matriciale e dalla linearità della traccia. La commutatività si può mostrare usando la descrizione analitica del prodotto matriciale, oppure notando che $\langle A, B\rangle = tr\left(B^T A\right) = tr\left(\left(B^T A\right)^T\right) =$

$tr\left(A^TB\right) = \langle B, A\rangle$ (la traccia è invariante per trasposizione). Per la positività serve che $tr\left(A^TA\right) \geq 0$ $\forall\ A \in M_{\mathbb{R}}(m,n)$ e che $tr\left(A^TA\right) = 0 \iff A = 0_M$. Se $A = [C_1(A)|...|C_n(A)]$ e $A^T = \begin{bmatrix} C_1(A) \\ ... \\ C_n(A) \end{bmatrix}$, $\left(A^TA\right)_{ii} = R_i\left(A^T\right)C_i(A) = \langle C_i(A), C_i(A)\rangle = ||C_i(A)||^2$ (secondo la definizione di norma nello spazio euclideo $\mathbb{R}^n$ con prodotto scalare standard). Quindi la traccia è $tr\left(A^TA\right) = \sum_{i=1}^n \left(A^TA\right)_{ii} = \sum_{i=1}^n ||C_i(A)||^2$: è maggiore o uguale di zero per ogni A perché somma di quadrati, ed è nulla se e solo se ogni quadrato è zero, cioè $||C_i(A)||^2 = 0\ \forall\ i$, cioè $||C_i(A)|| = 0\ \forall\ i$, cioè, per la positività della norma nello spazio euclideo $\mathbb{R}^n$ con prodotto scalare standard, $C_i(A) = \mathbf{0}\ \forall\ i$, cioè A è la matrice nulla.

3. *Dato lo spazio euclideo $M_{\mathbb{R}}(n,n)$, con $\langle A, B\rangle = tr\left(B^TA\right)$, determinare un'applicazione lineare $L : M_{\mathbb{R}}(n,n) \to \mathbb{R}^{n^2}$ tale che $\langle A, A\rangle = L(A) \cdot L(A)$, dove $\cdot$ indica il prodotto scalare euclideo standard di $\mathbb{R}^{n^2}$.*

   $\langle A, A\rangle = tr\left(A^TA\right) = \sum_{i=1}^n \left(A^TA\right)_{ii} = \sum_{i=1}^n \left(\sum_{j=1}^n a_{ij}^T a_{ji}\right) = \sum_{i=1}^n \left(\sum_{j=1}^n a_{ji}a_{ji}\right) = \sum_{i=1}^n \left(\sum_{j=1}^n a_{ji}^2\right)$. Questo significa sommare tutti i quadrati degli elementi della matrice: è equivalente a calcolare il prodotto scalare standard tra due vettori delle coordinate della matrice in base canonica. Quindi, posto $A = a_{11}E_{11} + ... + a_{nn}E_{nn}$, $L(E_{ij}) = \mathbf{e}_{ij}$ e $L(A) = \begin{pmatrix} a_{11} \\ a_{12} \\ ... \\ a_{nn} \end{pmatrix}$.

   Useremo la traccia per stabilire quando alcune matrici si assomigliano.

Teorema (disuguaglianza triangolare)

Hp: $(V, \langle\_,\_\rangle)$ è uno spazio euclideo; $\mathbf{v}, \mathbf{w} \in V$

Ts: $||\mathbf{v} + \mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}||$

Questo mostra che la norma non è un'applicazione lineare.

Dim Calcolo $||\mathbf{v} + \mathbf{w}||^2 = \langle \mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w}\rangle = \langle \mathbf{v} + \mathbf{w}, \mathbf{v}\rangle + \langle \mathbf{v} + \mathbf{w}, \mathbf{w}\rangle = \langle \mathbf{v}, \mathbf{v}\rangle + \langle \mathbf{w}, \mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{w}\rangle + \langle \mathbf{w}, \mathbf{w}\rangle = \langle v, v\rangle + 2\langle \mathbf{v}, \mathbf{w}\rangle + \langle \mathbf{w}, \mathbf{w}\rangle \leq \langle \mathbf{v}, \mathbf{v}\rangle + 2\left|\langle \mathbf{v}, \mathbf{w}\rangle\right| + \langle \mathbf{w}, \mathbf{w}\rangle$ (la maggiorazione è dovuta al fatto che $|\langle \mathbf{v}, \mathbf{w}\rangle| \geq \langle \mathbf{v}, \mathbf{w}\rangle$) $\leq ||\mathbf{v}||^2 + 2\,||\mathbf{v}||\,||\mathbf{w}|| + ||\mathbf{w}||^2$ (per la disuguaglianza di Schwarz) $= (||\mathbf{v}|| + ||\mathbf{w}||)^2$. Si è quindi scritto che $||\mathbf{v} + \mathbf{w}||^2 \leq (||\mathbf{v}|| + ||\mathbf{w}||)^2$: per la positività della norma, questo è equivalente a $||\mathbf{v} + \mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}||$. ■

Def Dato uno spazio euclideo $(V, \langle\_,\_\rangle)$, due vettori $\mathbf{v}, \mathbf{w} \in V$ si dicono ortogonali con tale prodotto scalare se $\langle \mathbf{v}, \mathbf{w}\rangle = 0$.

Quindi il vettore nullo è ortogonale a ogni vettore.

1. Prendo $V = L^2([0,2\pi])$, $f(x) = 1$, $g(x) = \cos x$. $\langle 1, \cos x\rangle = \int_0^{2\pi} \cos x dx = 0$, quindi f e g sono vettori ortogonali in $L^2([0,2\pi])$; lo stesso vale per 1 e $\sin x$, e per $\sin x$ e $\cos x$.

Teorema (Pitagora generalizzato)

Hp: $(V, \langle \_, \_\rangle)$ è uno spazio euclideo; $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ è un insieme di vettori non nulli di V tali che $\forall\ i, j = 1, ..., k$, $i \neq j$ implica $\langle \mathbf{v}_i, \mathbf{v}_j\rangle = 0$

Ts: (i) $||\alpha_1\mathbf{v}_1 + ... + \alpha_k\mathbf{v}_k||^2 = \alpha_1^2 ||\mathbf{v}_1||^2 + ... + \alpha_k^2 ||\mathbf{v}_k||^2$, con $\alpha_1, ..., \alpha_k \in \mathbb{R}$

(ii) $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ è un insieme di vettori linearmente indipendenti

L'ipotesi significa che ogni coppia di vettori distinti è fatta da vettori ortogonali. Si ha quindi un nuovo strumento per verificare l'indipendenza lineare.

Dim (i) Usando la bilinearità si ha $||\alpha_1\mathbf{v}_1 + ... + \alpha_k\mathbf{v}_k||^2 = \langle \alpha_1\mathbf{v}_1 + ... + \alpha_k\mathbf{v}_k, \alpha_1\mathbf{v}_1 + ... + \alpha_k\mathbf{v}_k\rangle = \langle \alpha_1\mathbf{v}_1, \alpha_1\mathbf{v}_1\rangle + ...\langle \alpha_1\mathbf{v}_1, \alpha_k\mathbf{v}_k\rangle$ $\alpha_1^2\langle \mathbf{v}_1, \mathbf{v}_1\rangle + ... + \alpha_k^2\langle \mathbf{v}_k, \mathbf{v}_k\rangle + 2\langle \alpha_1\mathbf{v}_1, \alpha_2\mathbf{v}_2\rangle + ... + 2\langle \alpha_{k-1}\mathbf{v}_{k-1}, \alpha_k\mathbf{v}_k\rangle = \alpha_1^2\langle \mathbf{v}_1, \mathbf{v}_1\rangle + ... + \alpha_k^2\langle \mathbf{v}_k, \mathbf{v}_k\rangle + 2\alpha_1\alpha_2\langle \mathbf{v}_1, \mathbf{v}_2\rangle + ... + 2\alpha_{k-1}\alpha_k\langle \mathbf{v}$ $... + \alpha_k^2\langle \mathbf{v}_k, \mathbf{v}_k\rangle = \alpha_1^2 ||\mathbf{v}_1||^2 + ... + \alpha_k^2 ||\mathbf{v}_k||^2$ perché tutti i prodotti scalari misti sono nulli, essendo $\langle \mathbf{v}_i, \mathbf{v}_j\rangle = 0$ se $i \neq j$ per ipotesi.

(ii) Considero $\alpha_1\mathbf{v}_1 + ... + \alpha_k\mathbf{v}_k = \mathbf{0}$: è equivalente a $||\alpha_1\mathbf{v}_1 + ... + \alpha_k\mathbf{v}_k|| = 0$, che implica $||\alpha_1\mathbf{v}_1 + ... + \alpha_k\mathbf{v}_k||^2 = 0$, i. e., per il primo punto del teorema, $\alpha_1^2 ||\mathbf{v}_1||^2 + ... + \alpha_k^2 ||\mathbf{v}_k||^2 = 0$: una somma di quadrati è nulla se e solo se ogni quadrato è nullo, cioè $\alpha_i^2 ||\mathbf{v}_i||^2 = 0$ per ogni $i = 1, ..., k$, ma poiché $\mathbf{v}_i \neq \mathbf{0}$ per ipotesi, $\alpha_i = 0$ per ogni $i = 1, ..., k$ e i vettori sono linearmente indipendenti per definizione. ■

5.1 Basi ortogonali e proiezioni ortogonali

Def Dato uno spazio euclideo $(V, \langle \_, \_\rangle)$, una base $B = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ di V si dice base ortogonale di V se $\forall\ i, j = 1, ..., n$, $i \neq j$ implica $\langle \mathbf{v}_i, \mathbf{v}_j\rangle = 0$; si dice base ortonormale di V se $\forall\ i, j = 1, ..., n$, $i \neq j$ implica $\langle \mathbf{v}_i, \mathbf{v}_j\rangle = 0$ e $\langle \mathbf{v}_i, \mathbf{v}_i\rangle = 1$ per ogni $i = 1, ..., n$.

Quindi una base ortonormale è fatta di vettori a due a due ortogonali e con norma unitaria.

1. Considero $V = \mathbb{R}^n$, dotato di prodotto scalare standard. La base canonica di V $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ è ortonormale.

2. Considero lo spazio euclideo $L^2([0,2\pi])$, dotato di prodotto scalare $\langle f, g\rangle = \int_a^b f(t) g(t) dt$. $V = Span(\mathbf{1}, \sin x, \cos x)$ sottospazio di L^2 è euclideo rispetto al prodotto scalare di L^2. $\{1, \sin x, \cos x\}$ è una base ortogonale di V (si è calcolato sopra), ma non normale: $||\mathbf{1}||^2 = \langle \mathbf{1}, \mathbf{1}\rangle = \int_0^{2\pi} 1 dx = 2\pi \neq 1$.

Teorema (decomposizione rispetto a una base ortogonale)

Hp: V è uno spazio euclideo, $H \subset V$ è un sottospazio vettoriale di V;

$B = \{\mathbf{b}_1, ..., \mathbf{b}_h\}$ è una base ortogonale di H

Ts: (i) $\forall\ \mathbf{v} \in H$, $\mathbf{v} = \dfrac{\langle \mathbf{v}, \mathbf{b}_1 \rangle}{||\mathbf{b}_1||^2}\mathbf{b}_1 + ... + \dfrac{\langle \mathbf{v}, \mathbf{b}_h \rangle}{||\mathbf{b}_h||^2}\mathbf{b}_h$

(ii) $\forall\ \mathbf{v}, \mathbf{w} \in H$, se $\mathbf{v} = x_1\mathbf{b}_1 + ... + x_h\mathbf{b}_h$, $\mathbf{w} = y_1\mathbf{b}_1 + ... + y_h\mathbf{b}_h$,

$$\langle \mathbf{v}, \mathbf{w} \rangle = x_1 y_1 ||\mathbf{b}_1||^2 + ... + x_h y_h ||\mathbf{b}_h||^2 = (x_1|...|x_h) \begin{bmatrix} ||\mathbf{b}_1||^2 & ... & \mathbf{0} \\ ... & ... & ... \\ \mathbf{0} & ... & ||\mathbf{b}_h||^2 \end{bmatrix} \begin{pmatrix} y_1 \\ ... \\ y_h \end{pmatrix}$$

Quindi, le proprietà aggiuntive conferite a H dalla definizione di un prodotto scalare forniscono un nuovo strumento per determinare le coordinate di un vettore di H rispetto a una base ortogonale. Inoltre, (ii) dà una formula analitica per calcolare il prodotto scalare in V (rispetto a cui si ha una base ortogonale): si può calcolare il prodotto scalare tra due vettori senza ricorrere alla definizione, ma usando una formula analitica che descrive il prodotto scalare "localmente", cioè relativamente alla base scelta di H, sfruttando l'isomorfismo tra $Span\,(\mathbf{b}_1, ..., \mathbf{b}_h) = V$ e $\mathbb{R}^h$. Questo significa che ogni prodotto scalare può essere calcolato come prodotto scalare non standard in $\mathbb{R}^h$ indotto dalla matrice diagonale delle norme dei vettori della base.

$\frac{\langle \mathbf{v}, \mathbf{b}_i \rangle}{||\mathbf{b}_i||^2}$ si dice coefficiente di Fourier di $\mathbf{v}$ rispetto a $\mathbf{b}_i$.

Dim (i) La rappresentazione rispetto a una base è unica: $\mathbf{v} = x_1\mathbf{b}_1 + ... + x_n\mathbf{b}_h$. Calcolo $\langle \mathbf{v}, \mathbf{b}_i \rangle = \langle x_1\mathbf{b}_1 + ... + x_n\mathbf{b}_h, \mathbf{b}_i \rangle = x_1\langle \mathbf{b}_1, \mathbf{b}_i \rangle + ... + x_i\langle \mathbf{b}_i, \mathbf{b}_i \rangle + ... + x_n\langle \mathbf{b}_h, \mathbf{b}_i \rangle = x_i\langle \mathbf{b}_i, \mathbf{b}_i \rangle$ perché la base è ortogonale. Quindi $\langle \mathbf{v}, \mathbf{b}_i \rangle = x_i ||\mathbf{b}_i||^2$: $\mathbf{b}_i \neq \mathbf{0}$ perché è un vettore della base, quindi $||\mathbf{b}_i||^2 \neq 0$ e $x_i = \frac{\langle \mathbf{v}, \mathbf{b}_i \rangle}{||\mathbf{b}_i||^2}$.

(ii) $\mathbf{v} = x_1\mathbf{b}_1 + ... + x_h\mathbf{b}_h$, $\mathbf{w} = y_1\mathbf{b}_1 + ... + y_h\mathbf{b}_h$; $\langle \mathbf{v}, \mathbf{w} \rangle = \langle x_1\mathbf{b}_1 + ... + x_h\mathbf{b}_h, y_1\mathbf{b_1} + ... + y_h\mathbf{b_h} \rangle = \langle x_1\mathbf{b}_1, y_1\mathbf{b_1} \rangle + ... + \langle x_h\mathbf{b}_h, y_h\mathbf{b_h} \rangle + x_1$ (per ortogonalità della base) $\langle x_1\mathbf{b}_1, y_1\mathbf{b_1} \rangle + ... + \langle x_h\mathbf{b}_h, y_h\mathbf{b_h} \rangle = x_1 y_1 ||\mathbf{b}_1||^2 + ... + x_h y_h ||\mathbf{b}_h||^2$. ∎

Corollario

Hp: V è uno spazio euclideo, $H \subset V$ è un sottospazio vettoriale di V;

$B = \{\mathbf{b}_1, ..., \mathbf{b}_h\}$ è una base ortonormale di H

Ts: (i) $\forall\ \mathbf{v} \in H$, $\mathbf{v} = \langle \mathbf{v}, \mathbf{b}_1 \rangle \mathbf{b}_1 + ... + \langle \mathbf{v}, \mathbf{b}_h \rangle \mathbf{b}_h$

(ii) $\forall\ \mathbf{v}, \mathbf{w} \in H$, se $\mathbf{v} = x_1\mathbf{b}_1 + ... + x_h\mathbf{b}_h$, $\mathbf{w} = y_1\mathbf{b}_1 + ... + y_h\mathbf{b}_h$,

$$\langle \mathbf{v}, \mathbf{w} \rangle = x_1y_1 + ... + x_hy_h = (x_1|...|x_h)\begin{bmatrix} 1 & ... & \mathbf{0} \\ ... & ... & ... \\ \mathbf{0} & ... & 1 \end{bmatrix}\begin{pmatrix} y_1 \\ ... \\ y_h \end{pmatrix}$$

$$\text{e } ||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{x_1^2 + ... + x_h^2}$$

poiché $||\mathbf{b}_i||^2 = 1$ per ogni $i = 1, ..., h$.

Quindi fondamentalmente il prodotto scalare in un qualsiasi spazio euclideo può essere espresso come prodotto scalare standard in $\mathbb{R}^h$ tra i vettori delle coordinate rispetto a una base ortonormale: cioè l'isomorfismo tra V e $\mathbb{R}^h$ mediante una mappa delle coordinate rispetto a una base ortonormale preserva il prodotto scalare.

1. Considero $V = \mathbb{R}^2$ con la struttura euclidea standard. $B = \left\{ \begin{pmatrix} 2 \\ 7 \end{pmatrix}, \begin{pmatrix} 7 \\ -2 \end{pmatrix} \right\}$ è una base ortogonale di V, ma non normale. Cerco la decomposizione di $\begin{pmatrix} 13 \\ 5 \end{pmatrix}$ rispetto a B: senza il teorema, avrei risolto il sistema lineare $x_1 \begin{pmatrix} 2 \\ 7 \end{pmatrix} + x_2 \begin{pmatrix} 7 \\ -2 \end{pmatrix} = \begin{pmatrix} 13 \\ 5 \end{pmatrix}$. Ora so che $x_1 = \frac{\langle \begin{pmatrix} 13 \\ 5 \end{pmatrix}, \begin{pmatrix} 2 \\ 7 \end{pmatrix} \rangle}{\left\| \begin{pmatrix} 2 \\ 7 \end{pmatrix} \right\|^2} = \frac{61}{53}$ e $x_2 = \frac{\langle \begin{pmatrix} 13 \\ 5 \end{pmatrix}, \begin{pmatrix} 7 \\ -2 \end{pmatrix} \rangle}{\left\| \begin{pmatrix} 7 \\ -2 \end{pmatrix} \right\|^2} = \frac{81}{53}$.
2. Considero $L^2([0, 2\pi])$ e $V = Span(1, \cos x, \sin x)$ ($\{1, \cos, \sin\}$ è una base ortogonale, non normale), $f(x) = \lambda_1 + \lambda_2 \sin x + \lambda_3 \cos x$, $g(x) = \mu_1 + \mu_2 \sin x + \mu_3 \cos x$. $\langle f(x), g(x) \rangle = \int_0^{2\pi} (\lambda_1 + \lambda_2 \sin x + \lambda_3 \cos x)(\mu_1 + \mu_2 \sin x + \mu_3 \cos x)\, dx$ per definizione. Ma usando il teorema, si ha $\langle f(x), g(x) \rangle = \lambda_1\mu_1 ||1||^2 + \lambda_2\mu_2 ||\sin x||^2 + \lambda_3\mu_3 ||\cos x||^2 =$

$[\lambda_1|\lambda_2|\lambda_3]\begin{bmatrix} ||1||^2 & 0 & 0 \\ 0 & ||\sin x||^2 & 0 \\ 0 & 0 & ||\cos x||^2 \end{bmatrix}\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}$. Si osserva che, dato V isomorfo a $\mathbb{R}^3$, $(V, \langle\_,\_\rangle)$ è isomorfo a $\mathbb{R}^3$ con prodotto scalare non standard, indotto dalla matrice $\begin{bmatrix} ||1||^2 & 0 & 0 \\ 0 & ||\sin x||^2 & 0 \\ 0 & 0 & ||\cos x||^2 \end{bmatrix}$. Quindi anche la struttura di spazio euclideo può essere espressa in termini matriciali.

3. Se W è uno spazio euclideo e $f : V \to W$ è un'applicazione lineare iniettiva, f conferisce a V la struttura di spazio euclideo, con prodotto scalare tra $\mathbf{v}_1$ e $\mathbf{v}_2$ definito da $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle_V = \langle f(\mathbf{v}_1), f(\mathbf{v}_2) \rangle_W$ per ogni $\mathbf{v}_1, \mathbf{v}_2 \in V$. Quindi ogni spazio vettoriale V a dimensione finita, con base B, è dotato di un prodotto scalare indotto dalla mappa delle coordinate $X_B : V \to \mathbb{R}^n$, che è iniettiva: il prodotto scalare tra $\mathbf{v}_1, \mathbf{v}_2 \in V$ è $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle_V = \langle X_B(\mathbf{v}_1), X_B(\mathbf{v}_2) \rangle_{\mathbb{R}^n}$ (se $X_B(\mathbf{v}_1) = \mathbf{u}_1, X_B(\mathbf{v}_2) = \mathbf{u}_2$, $\langle \mathbf{u}_1, \mathbf{u}_2 \rangle_{\mathbb{R}^n} = \langle P(\mathbf{u}_1), P(\mathbf{u}_2) \rangle_V$ con P mappa delle coordinate). Perciò ogni spazio vettoriale V a dimensione finita può essere dotato di infinite strutture di spazio euclideo (una per ogni base).

Per il teorema visto, conoscere una base ortogonale di H semplifica molte cose, ad esempio permette di calcolare facilmente il prodotto scalare tra vettori di H. E' quindi ragionevole chiedersi se è possibile costruire una base ortogonale o ortonormale per qualsiasi sottospazio $H \subset V$: se sì, come si costruisce?

1. Considero H sottospazio di V spazio euclideo, $H = \mathit{Span}(\mathbf{v}, \mathbf{w})$, $\dim H = 2$ (H isomorfo a $\mathbb{R}^2$). $B = \{\mathbf{v}, \mathbf{w}\}$ è una base generica di H; posso trovare un insieme di vettori $\{\mathbf{b}_1, \mathbf{b}_2\}$ tale che $\langle \mathbf{b}_1, \mathbf{b}_2 \rangle = 0$ e che $\mathit{Span}(\mathbf{b}_1, \mathbf{b}_2) = \mathit{Span}(\mathbf{v}, \mathbf{w})$? Sto cercando una base ortogonale. Prendo $\mathbf{b}_1 = \mathbf{v}$ e cerco $\mathbf{b}_2 = x\mathbf{v} + y\mathbf{w}$ (lo scrivo in questa forma perché dev'essere $\mathit{Span}(\mathbf{b}_1, \mathbf{b}_2) = \mathit{Span}(\mathbf{v}, \mathbf{w})$), con $x \neq 0$: quindi posso considerare $\mathbf{b}_2 = \mathbf{v} + y\mathbf{w}$. Ora cerco $y : \langle \mathbf{b}_1, \mathbf{b}_2 \rangle = 0$. $\langle \mathbf{b}_1, \mathbf{b}_2 \rangle = \langle \mathbf{v}, \mathbf{v} + y\mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{v} \rangle + y\langle \mathbf{v}, \mathbf{w} \rangle = ||\mathbf{v}||^2 + y\langle \mathbf{v}, \mathbf{w} \rangle$. $\langle \mathbf{b}_1, \mathbf{b}_2 \rangle = 0 \iff ||\mathbf{v}||^2 + y\langle \mathbf{v}, \mathbf{w} \rangle = 0 \iff y = -\frac{||\mathbf{v}||^2}{\langle \mathbf{v}, \mathbf{w} \rangle}$: y è ben definito perché $\langle \mathbf{v}, \mathbf{w} \rangle \neq 0$, dato che si è partiti da una base non ortogonale. Quindi la base ortogonale trovata di H è $\left\{\mathbf{v}, \mathbf{v} - \frac{||\mathbf{v}||^2}{\langle \mathbf{v}, \mathbf{w} \rangle}\mathbf{w}\right\}$, o, considerando un multiplo del secondo vettore, $\left\{\mathbf{v}, \mathbf{w} - \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{||\mathbf{v}||^2}\mathbf{v}\right\}$: $\frac{\langle \mathbf{v}, \mathbf{w} \rangle}{||\mathbf{v}||^2}$ è il coefficiente di Fourier di $\mathbf{w}$ rispetto a $\mathbf{v}$. $\frac{\langle \mathbf{v}, \mathbf{w} \rangle}{||\mathbf{v}||^2}\mathbf{v}$ è il vettore proiezione ortogonale di $\mathbf{w}$ su $\mathbf{v}$. Infatti in $\mathbb{R}^2$ $\frac{\langle \mathbf{v}, \mathbf{w} \rangle}{||\mathbf{v}||^2}\mathbf{v} = \frac{||\mathbf{v}||||\mathbf{w}|| \cos\theta}{||\mathbf{v}||} \cdot \frac{\mathbf{v}}{||\mathbf{v}||} = ||\mathbf{w}|| \cos\theta \cdot \mathbf{1}$: $\frac{\mathbf{v}}{||\mathbf{v}||}$ è un vettore di norma unitaria.

Def Sia $(V, \langle\_,\_\rangle)$ uno spazio euclideo. Se S è un qualsiasi sottinsieme di V, il sottinsieme di V $S^\perp = \{\mathbf{v} \in V : \langle \mathbf{v}, \mathbf{w} \rangle = 0 \ \forall\ \mathbf{w} \in S\}$ è l'insieme dei vettori di V che sono ortogonali a ogni vettore di S.

Proposizione

$$\begin{aligned} \text{Hp} &: \quad V, \langle \_, \_ \rangle \text{ è uno spazio euclideo, } S \subset V \\ \text{Ts} &: \quad S^\perp \text{ è un sottospazio vettoriale di } V \end{aligned}$$

Dim Mostro che $\forall\, \mathbf{v}_1, \mathbf{v}_2 \in S^\perp$, per ogni $\lambda_1, \lambda_2 \in \mathbb{R}$, $\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 \in S^\perp$. Infatti per ogni $\mathbf{w} \in S$ $\langle \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2, \mathbf{w}\rangle = \lambda_1\langle \mathbf{v}_1, \mathbf{w}\rangle + \lambda_2\langle$ 0 perché $\mathbf{v}_1, \mathbf{v}_2 \in S^\perp$, quindi anche $\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 \in S^\perp$. ■

Ci interessa in particolare il caso di $S = H$ sottospazio vettoriale.

Proposizione

Hp: $V, \langle \_, \_ \rangle$ è uno spazio euclideo, H è un sottospazio vettoriale di V, $B = \{\mathbf{b}_1, ..., \mathbf{b}_h\}$ è una base di H

Ts: $H^\perp = B^\perp$

cioè per trovare il complemento ortogonale di H è sufficiente trovare quello di B, che ha un insieme finito di elementi.

Dim Mostro che $H^\perp \subseteq B^\perp$, cioè se $\mathbf{v}$ è tale che $\langle \mathbf{v}, \mathbf{w}\rangle = 0\ \forall\ \mathbf{w} \in H$, allora $\mathbf{v}$ è tale che $\langle \mathbf{v}, \mathbf{b}_i\rangle = 0\ \forall\ \mathbf{b}_i \in B$. Infatti, se $\mathbf{v}$ è ortogonale a ogni vettore di H, allora è ortogonale in particolare a ogni vettore di B, perché i vettori di B appartengono ad H. Mostro che $B^\perp \subseteq H^\perp$, cioè se $\mathbf{v}$ è tale che $\langle \mathbf{v}, \mathbf{b}_i\rangle = 0\ \forall\ \mathbf{b}_i \in B$, allora $\mathbf{v}$ è tale che $\langle \mathbf{v}, \mathbf{w}\rangle = 0\ \forall\, \mathbf{w} \in H$. Infatti, se $\mathbf{w} = x_1\mathbf{b}_1 + ... + x_h\mathbf{b}_h$, $\langle \mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{v}, x_1\mathbf{b_1} + ... + x_h\mathbf{b_h}\rangle = x_1\langle \mathbf{v}, \mathbf{b}_1\rangle + ... + x_h\langle \mathbf{v}, \mathbf{b}_h\rangle = 0$ perché $\mathbf{v}$ è ortogonale a ogni vettore di B. Quindi $H^\perp = B^\perp$. ■

1. Dati i sottospazi U, V, se $U \subseteq V$, allora $V^\perp \subseteq U^\perp$.

Teorema (proiezione ortogonale per spazi euclidei a dimensione finita)

Hp: $(V, \langle \_, \_ \rangle)$ è uno spazio euclideo, $\dim V < +\infty$, H è un sottospazio vettoriale di V

Ts: (i) $H^\perp \cap H = \{\mathbf{0}\}$

(ii) $\forall\ \mathbf{v} \in V\ \exists\ !\mathbf{v}_H \in H : \mathbf{v} - \mathbf{v}_H \in H^\perp$

(iii) $||\mathbf{v} - \mathbf{v}_H|| \leq ||\mathbf{v} - \mathbf{w}||$ per ogni $\mathbf{w} \in H$

cioè l'intersezione tra i due sottospazi è la più piccola possibile. $\mathbf{v}$ in (ii) è quello che nell'esempio precedente era $\mathbf{w}$ e $\mathbf{v}_H$ era la proiezione di $\mathbf{w}$ su $\mathbf{v}$. (iii) significa che $\mathbf{v}_H$, tra tutti quelli di H, è il vettore più vicino a $\mathbf{v}$, cioè è la migliore approssimazione di $\mathbf{v}$ che si può ricavare da H. (ii) significa che ogni vettore di V può essere scritto come somma di un vettore di H e un vettore di $H^\perp$: quindi $V = H + H^\perp$. Poiché questo è possibile in un unico modo (aggiungendo (i)), si ha $V = H \oplus H^\perp$.

Dim (i) Se $\mathbf{v} \in H^\perp \cap H$, $\mathbf{v}$ è un vettore di H che è ortogonale a tutti i vettori di H, incluso se stesso, quindi $\langle \mathbf{v}, \mathbf{v} \rangle = 0$: questo implica $\mathbf{v} = \mathbf{0}$ per positività del prodotto scalare.

(ii) Considero $B = \{\mathbf{b}_1, ..., \mathbf{b}_h\}$ base ortogonale di H (in realtà non abbiamo ancora mostrato che esiste). Dato $\mathbf{v} \in V$, cerco $\mathbf{v}_H = x_1\mathbf{b}_1 + ... + x_h\mathbf{b}_h : \mathbf{v} - \mathbf{v}_H = \mathbf{v} - x_1\mathbf{b}_1 - ... - x_h\mathbf{b}_h \in H^\perp$. $H^\perp = B^\perp$, quindi $\mathbf{v} - \mathbf{v}_H \in H^\perp \iff \mathbf{v} - \mathbf{v}_H \in B^\perp \iff \langle \mathbf{v} - \mathbf{v}_H, \mathbf{b}_i \rangle = 0 \ \forall \ \mathbf{b}_i \in B$. $\langle \mathbf{v} - \mathbf{v}_H, \mathbf{b}_i \rangle = \langle \mathbf{v} - x_1\mathbf{b}_1 - ...x_i\mathbf{b}_i - ... - x_h\mathbf{b}_h, \mathbf{b}_i \rangle = \langle \mathbf{v}, \mathbf{b}_i \rangle - x_1\langle \mathbf{b}_1, \mathbf{b}_i \rangle - ... - x_i\langle \mathbf{b}_i, \mathbf{b}_i \rangle - ... - x_h\langle \mathbf{b}_h, \mathbf{b}_i \rangle =$ (per ortogonalità della base) $\langle \mathbf{v}, \mathbf{b}_i \rangle - x_i\langle \mathbf{b}_i, \mathbf{b}_i \rangle$. Quindi $\langle \mathbf{v} - \mathbf{v}_H, \mathbf{b}_i \rangle = 0$ se e solo se $\langle \mathbf{v}, \mathbf{b}_i \rangle - x_i\langle \mathbf{b}_i, \mathbf{b}_i \rangle = \langle \mathbf{v}, \mathbf{b}_i \rangle - x_i \,||\mathbf{b}_i||^2 = 0$, cioè $x_i = \frac{\langle \mathbf{v}, \mathbf{b}_i \rangle}{||\mathbf{b}_i||^2}$: è il coefficiente di Fourier di $\mathbf{v}$ rispetto a $\mathbf{b}_i$. Dunque $\exists \ \mathbf{v}_H$ con le proprietà richieste: $\mathbf{v}_H = \frac{\langle \mathbf{v}, \mathbf{b}_1 \rangle}{||\mathbf{b}_1||^2}\mathbf{b}_1 + ... + \frac{\langle \mathbf{v}, \mathbf{b}_h \rangle}{||\mathbf{b}_h||^2}\mathbf{b}_h$. Mostro che è unico. Suppongo esista un secondo $\mathbf{v}'_H$ con le proprietà richieste: allora $\mathbf{v} = \mathbf{v}_H + (\mathbf{v} - \mathbf{v}_H)$ e $\mathbf{v} = \mathbf{v}'_H + (\mathbf{v} - \mathbf{v}'_H)$. Questo è equivalente a $\mathbf{v} - \mathbf{v} = \mathbf{0} = \mathbf{v}_H + (\mathbf{v} - \mathbf{v}_H) - \mathbf{v}'_H - (\mathbf{v} - \mathbf{v}'_H)$, cioè $\mathbf{v}_H - \mathbf{v}'_H = \mathbf{v} - \mathbf{v}'_H - (\mathbf{v} - \mathbf{v}_H)$. $\mathbf{v}_H - \mathbf{v}'_H \in H$, $\mathbf{v} - \mathbf{v}'_H - (\mathbf{v} - \mathbf{v}_H) \in H^\perp$ (sono entrambi sottospazi): l'unico elemento nell'intersezione è il vettore nullo, quindi può essere solo $\mathbf{v}_H = \mathbf{v}'_H$.

(iii) Mostro che $||\mathbf{v} - \mathbf{v}_H|| \leq ||\mathbf{v} - \mathbf{w}||$ per ogni $\mathbf{w} \in H$. Considero $\mathbf{v} - \mathbf{w} = \mathbf{v} - \mathbf{v}_H + \mathbf{v}_H - \mathbf{w} = (\mathbf{v} - \mathbf{v}_H) + (\mathbf{v}_H - \mathbf{w})$. $\mathbf{v}_H - \mathbf{w} \in H$ perché combinazione lineare di vettori di H, $\mathbf{v} - \mathbf{v}_H \in H^\perp$ per costruzione, quindi $\langle \mathbf{v}_H - \mathbf{w}, \mathbf{v} - \mathbf{v}_H \rangle = 0$. Posso applicare il teorema di Pitagora a $\{\mathbf{v}_H - \mathbf{w}, \mathbf{v} - \mathbf{v}_H\}$: $||\mathbf{v}_H - \mathbf{w}||^2 + ||\mathbf{v} - \mathbf{v}_H||^2 = ||\mathbf{v} - \mathbf{w}||^2$. $||\mathbf{v}_H - \mathbf{w}||^2 + ||\mathbf{v} - \mathbf{v}_H||^2 \geq ||\mathbf{v} - \mathbf{v}_H||^2$, quindi $||\mathbf{v} - \mathbf{w}||^2 \geq ||\mathbf{v} - \mathbf{v}_H||^2 \iff ||\mathbf{v} - \mathbf{w}|| \geq ||\mathbf{v} - \mathbf{v}_H||$. ■

Corollario

$$\text{Hp: } (V, \langle \_, \_ \rangle) \text{ è uno spazio euclideo, } \dim V < +\infty, \ H \text{ è un sottospazio vettoriale di } V$$

$$\text{Ts: (i) } H^\perp \oplus H = V$$

$$\text{(ii) } \dim H^\perp = \dim V - \dim H$$

$$\text{(iii) } \left(H^\perp\right)^\perp = H$$

(iii) non vale in spazi a dimensione infinita.

Dim (iii) Se $\mathbf{v} \in H$, $\mathbf{v}$ è ortogonale a tutti gli elementi di $H^\perp$: $\mathbf{v} \perp \mathbf{w}$ per ogni $\mathbf{w} \in H^\perp$, cioè $\mathbf{v} \in \left(H^\perp\right)^\perp$: quindi $H \subseteq \left(H^\perp\right)^\perp$. Per (i) applicata a $H^\perp$, vale $\left(H^\perp\right)^\perp \oplus H^\perp = V$, quindi $\dim \left(H^\perp\right)^\perp = \dim V - \dim H^\perp$. Poiché vale anche $H^\perp \oplus H = V$, $\dim H = \dim V - \dim H^\perp$: quindi $\dim \left(H^\perp\right)^\perp = \dim H$, ed essendo $H \subseteq \left(H^\perp\right)^\perp$, $H = \left(H^\perp\right)^\perp$. ■

$H^\perp$ si dice complemento ortogonale di H.

Dunque, ogni vettore $\mathbf{v} \in V$ può essere scritto nella forma $\mathbf{v} = \mathbf{v}_H + \mathbf{v}_{H^\perp}$, dove $\mathbf{v}_{H^\perp} \in H^\perp$ e $\mathbf{v}_H \in H$: data $B = \{\mathbf{b}_1, ..., \mathbf{b}_h\}$ base ortogonale di H, $\mathbf{v}_H = \frac{\langle \mathbf{v}, \mathbf{b}_1 \rangle}{||\mathbf{b}_1||^2}\mathbf{b}_1 + ... + \frac{\langle \mathbf{v}, \mathbf{b}_h \rangle}{||\mathbf{b}_h||^2}\mathbf{b}_h$ si dice proiezione ortogonale di $\mathbf{v}$ sul sottospazio H.

Teorema (algoritmo di Gram-Schmidt)

Hp: $(V, \langle \_, \_ \rangle)$ è uno spazio euclideo, $\{\mathbf{v}_1, ..., \mathbf{v}_d\}$ è un insieme di vettori linearmente indipendenti

Ts: esistono d vettori $\{\mathbf{b}_1, .., \mathbf{b}_d\} : \forall\ k = 1, ..., d\ \ \{\mathbf{b}_1, .., \mathbf{b}_k\}$ è una base ortogonale del sottospaziogenerato da $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$

Quindi considero $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$, con k che varia da 1 a d, e per ogni k posso trovare una base ortogonale di $Span\,(\mathbf{v}_1, ..., \mathbf{v}_k)$: se $k = 1$ $\{\mathbf{b}_1\}$, se $k = 2$ $\{\mathbf{b}_1, \mathbf{b}_2\}$, se $k = 3$ $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$,..., se $k = d$ $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, ..., \mathbf{b}_d\}$. Una base con un solo vettore è sempre ortogonale. Dato $k = 2$ e $\{\mathbf{v}_1, \mathbf{v}_2\}$, la base è formata da $\mathbf{b}_1 = \mathbf{v}_1$ e $\mathbf{b}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{b}_1 \rangle}{||\mathbf{b}_1||^2}\mathbf{b}_1$: $\frac{\langle \mathbf{v}_2, \mathbf{b}_1 \rangle}{||\mathbf{b}_1||^2}\mathbf{b}_1$ è la proiezione ortogonale di $\mathbf{v}_2$ su $Span\,(\mathbf{b}_1)$ (infatti voglio che $\mathbf{b}_2 \in H^\perp$ con $H = Span\,(\mathbf{b}_1)$, quindi, essendo $\mathbf{v}_2 = \mathbf{b}_H + \mathbf{b}_{H^\perp}$, prendo solo la componente ortogonale a H). Dato $k = 3$ e $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, la base è formata da $\mathbf{b}_1 = \mathbf{v}_1$, $\mathbf{b}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{b}_1 \rangle}{||\mathbf{b}_1||^2}\mathbf{b}_1$ e $\mathbf{b}_3 = \mathbf{v}_3 - \left(\frac{\langle \mathbf{v}_3, \mathbf{b}_1 \rangle}{||\mathbf{b}_1||^2}\mathbf{b}_1 + \frac{\langle \mathbf{v}_3, \mathbf{b}_2 \rangle}{||\mathbf{b}_2||^2}\mathbf{b}_2 \right)$: $\frac{\langle \mathbf{v}_3, \mathbf{b}_1 \rangle}{||\mathbf{b}_1||^2}\mathbf{b}_1 + \frac{\langle \mathbf{v}_3, \mathbf{b}_2 \rangle}{||\mathbf{b}_2||^2}\mathbf{b}_2$ è la proiezione ortogonale di $\mathbf{v}_3$ su $Span\,(\mathbf{b}_1, \mathbf{b}_2)$. Analogamente, $\mathbf{b}_d = \mathbf{v}_d - \sum_{i=1}^{d-1} \frac{\langle \mathbf{v}_d, \mathbf{b}_i \rangle}{||\mathbf{b}_i||^2}\mathbf{b}_i$.

Infatti, se scrivo un vettore $\mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{b}_1 \rangle}{||\mathbf{b}_1||^2}\mathbf{b}_1 + ... + \frac{\langle \mathbf{v}, \mathbf{b}_h \rangle}{||\mathbf{b}_h||^2}\mathbf{b}_h$, sottraendo $\sum_{i=1}^{d-1} \frac{\langle \mathbf{v}_d, \mathbf{b}_i \rangle}{||\mathbf{b}_i||^2}\mathbf{b}_i$ rimane la componente ortogonale a ogni $\mathbf{b}_i$ per $i = 1, ..., d-1$.

Se voglio avere una base ortonormale, normalizzo ogni vettore della base ortogonale moltiplicandolo per il reciproco della sua norma: la base ortonormale è $\left\{ \tilde{\mathbf{b}}_1, .., \tilde{\mathbf{b}}_d \right\} = \left\{ \frac{\mathbf{b}_1}{||\mathbf{b}_1||}, .., \frac{\mathbf{b}_d}{||\mathbf{b}_d||} \right\}$.

1. $V = L^2\,([0, 2\pi])$. Tutte le funzioni continue definite in $[0, 2\pi]$ appartengono a V, in particolare le funzioni trigonometriche $\{\sin(kx), \cos(kx), ...\}$. Interpreto $f(x) \in V$ come segnale di un suono: voglio comprimere il segnale, cioè scrivere f come combinazione lineare dei $2k + 1$ vettori $\{1, ..., \sin(kx), \cos(kx)\}$. Qual è l'elemento di $Span\,(1, ..., \sin(kx), \cos(kx))$ che meglio approssima f, cioè la g tale che $||f - g||$ è minima? Per il teorema visto g è la proiezione ortogonale di f su $Span\,(1, ..., \sin(kx), \cos(kx))$.

2. $V = \mathbb{R}[x]_{\leq 2}$, *con* $\langle p(x), q(x) \rangle = \int_0^2 p(x)\, q(x)\, dx$, $H = \{p(x) \in V : p(2) = 0\}$. *Trovare una base ortogonale di* H *il cui primo elemento sia* $p(x) = x - 2$.

   Si può procedere in due modi. Una volta appurato che $\dim H = 2$, si ha che una base generica di V, il cui primo elemento sia $p(x)$, è $\{x - 2, (x-2)(x-a)\}$ al variare di $a \in \mathbb{R}$, infatti $x - 2, (x-2)(x-a)$ sono due vettori linearmente indipendenti di H. Si cerca quindi $a : \int_0^2 (x-2)(x-2)(x-a)\, dx = 0$. Risolvendo l'equazione rispetto ad a, si trova $a = \frac{1}{2}$, quindi una base ortogonale è $\left\{x - 2, (x-2)\left(x - \frac{1}{2}\right)\right\}$.

Si può anche usare l'algoritmo di Gram-Schmidt su una base qualsiasi come $\left\{x-2, x^2-4\right\} = \{\mathbf{b}_1, \mathbf{b}_2\}$. Dato $\mathbf{v}_1 = x-2$, $\mathbf{v}_2 = \mathbf{b}_2 - P_{Span(\mathbf{v}_1)}(\mathbf{b}_2) = \mathbf{b}_2 - \frac{\langle \mathbf{b}_2, \mathbf{v}_1 \rangle}{||\mathbf{v}_1||}\mathbf{v}_1 = x-2 - \frac{\int_0^2 \left(x^2-4\right)(x-2)dx}{\int_0^2 (x-2)^2 dx}(x-2) = (x-2)\left(x-\frac{1}{2}\right)$.

3. *Dato lo spazio euclideo* $M_{\mathbb{R}}(2,2)$ *e definito* $\langle A, B\rangle = tr\left(B^T A\right)$, *trovare tutte le matrici ortogonali ad* $A = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$.

   E' richiesto di trovare $H^\perp$, dove $H = Span(A)$. Poiché $H^\perp = B^\perp$, dove $B = \{A\}$, cerco l'insieme delle matrici ortogonali ad A. Una generica matrice B è $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$: se è ortogonale ad A, vale $tr\left(B^T A\right) = tr\left(\begin{bmatrix} a & c \\ b & d \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}\right) = tr\begin{pmatrix} a & 2a \\ b & 2b \end{pmatrix} = a + 2b = 0$. Risolvendo il sistema si ottiene $\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = t\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + s\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} + u\begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix}$, quindi una base di $H^\perp$ è $\left\{\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -2 & 1 \\ 0 & 0 \end{bmatrix}\right\}$.

Quindi, se $(V, \langle\_,\_\rangle)$ è uno spazio euclideo a dimensione finita e $H \subset V$ è un sottospazio vettoriale, la funzione $P_H : V \to V$ che associa a ogni vettore $\mathbf{v}$ la sua proiezione ortogonale su H $\mathbf{v}_H$ (usato con lo stesso significato del teorema sulla proiezione ortogonale) è un'applicazione lineare, perché è sia omogenea che additiva. Se $B = \{\mathbf{b}_1, ..., \mathbf{b}_h\}$ è una base ortogonale di H, il vettore proiezione ortogonale di $\mathbf{v}$ su H è $\mathbf{v}_H = \frac{\langle \mathbf{v}, \mathbf{b}_1\rangle}{||\mathbf{b}_1||^2}\mathbf{b}_1 + ... + \frac{\langle \mathbf{v}, \mathbf{b}_h\rangle}{||\mathbf{b}_h||^2}\mathbf{b}_h$.

Ha quindi senso chiedersi quale nucleo, immagine, autovalori, autovettori ha P_H. Con osservazioni geometriche possiamo dire che $Im(P_H) = H$ (tutti gli elementi cui si applica la proiezione appartengono ad H) e $\ker(P_H)$ è l'insieme dei vettori ortogonali ad H: $\ker(P_H) = H^\perp$ (come si evince anche dalla formula analitica). Se si prende come H un piano passante per l'origine, $\ker(P_H)$ è la retta ortogonale al piano passante per l'origine. Il teorema di nullità più rango afferma che $\dim V = \dim(ImP_H) + \dim(\ker P_H) = \dim H + \dim H^\perp$, che è la stessa cosa affermata dalla formula di Grassman per la somma diretta $V = H \oplus H^\perp$.

Osservo che $P_H \circ P_H = P_H$, perché P_H associa a $\mathbf{v}$ la sua proiezione ortogonale su H $\mathbf{v}_H$, e P_H applicata a $\mathbf{v}_H$ lo lascia inalterato (si dimostra mostrando che P_H applicata a qualsiasi vettore di una base ortogonale di H lo lascia inalterato): $P_H^2 = P_H$, e P_H si dice pertanto applicazione lineare idempotente ($P_H^k = P_H$ per ogni $k \in \mathbb{N}$); la sua matrice rappresentativa è una matrice idempotente, cioè $P_H(\mathbf{v}) = \mathbf{v}\ \forall\ \mathbf{v} \in H$.

Per ricavare H avendo la matrice P_H è sufficiente trovare l'immagine dell'applicazione lineare descritta da P_H (oppure l'autospazio relativo all'autovalore 1).

Sia λ un autovalore di P_H e $\mathbf{v} \neq \mathbf{0}$ un autovettore relativo a λ. Allora $P_H(\mathbf{v}) = \lambda\mathbf{v}$ e $P_H(P_H(\mathbf{v})) = P_H(\lambda\mathbf{v}) = \lambda^2\mathbf{v}$: poiché P_H è idempotente, dev'essere $\lambda\mathbf{v} = \lambda^2\mathbf{v} \Longleftrightarrow \mathbf{v}\left(\lambda^2 - \lambda\right) = \mathbf{0}$. Poiché $\mathbf{v} \neq \mathbf{0}$, $\lambda = 1$ o $\lambda = 0$: significa che gli unici autovalori possibili per P_H (o per una matrice che rappresenta una proiezione ortogonale) sono 0 e 1 (ogni matrice di proiezione è semidefinita positiva). Gli autospazi relativi sono $V_0 = \ker P_H = H^\perp$ e $V_1 = \{\mathbf{v} \in V : P_H(\mathbf{v}) = \mathbf{v}\} = H$, che sono i vettori lasciati immutati dalla trasformazione. Quindi $V_0 \oplus V_1 = H + H^\perp = V$. Poiché gli autospazi di P_H sono in somma diretta, P_H è diagonalizzabile: esiste una base di V formata da autovettori di P_H che permette di ottenere una matrice rappresentativa diagonale; tale base è formata dall'unione delle basi degli autospazi di P_H. Grazie all'algoritmo di Gram-Schmidt si può costruire una base ortonormale B_H di H, una base ortonormale $B_{H^\perp}$ di $H^\perp$, e $B = B_{H^\perp} \cup B_H$ è una base ortonormale di V (per le caratteristiche di H e $H^\perp$, l'unione delle due basi ortonormali è ortonormale). Quindi si riesce a rappresentare P_H con una matrice diagonale rispetto a una base di autovettori, che è ortogonale e ha quindi tutte le buone proprietà della base canonica.

Def Una matrice $A \in M_{\mathbb{R}}(n,n)$ si dice ortogonalmente diagonalizzabile se è diagonalizzabile mediante una base ortogonale (o ortonormale: moltiplicare per scalari non cambia la proprietà di essere autovettori) di autovettori di A.

Quindi A è ortogonalmente diagonalizzabile se esiste una base ortonormale di $\mathbb{R}^n$ formata da autovettori di A. Se A è ortogonalmente diagonalizzabile, non tutte le matrici che rappresentano la stessa applicazione lineare descritta da A sono ortogonalmente diagonalizzabili (matrici simili non necessariamente hanno gli stessi autovettori).

Come riconoscere le matrici con queste proprietà?

Teorema fondamentale dell'ortogonalità

$$\text{Hp: } A \in M_{\mathbb{R}}(n,n)$$

$$\text{Ts: (i) } \ker A = (rowA)^\perp$$

$$\text{(ii) } rowA = (\ker A)^\perp$$

$$\text{(iii) } \ker A^T = (colA)^\perp$$

$$\text{(iv) } colA = \left(\ker A^T\right)^\perp$$

Dim (i) Se $A = \begin{bmatrix} \mathbf{v}_1 \\ ... \\ \mathbf{v}_m \end{bmatrix}$, $rowA = Span(\mathbf{v}_1, ..., \mathbf{v}_m) \in \mathbb{R}^n$. $\mathbf{x} \in \ker A \Longleftrightarrow A\mathbf{x} = \mathbf{0} \Longleftrightarrow \begin{bmatrix} \mathbf{v}_1 \\ ... \\ \mathbf{v}_m \end{bmatrix} \mathbf{x} = \mathbf{0}$

$\iff \begin{bmatrix} \langle \mathbf{v}_1, \mathbf{x} \rangle \\ ... \\ \langle \mathbf{v}_m, \mathbf{x} \rangle \end{bmatrix} = \mathbf{0} \iff \langle \mathbf{v}_i, \mathbf{x} \rangle = 0$ per ogni $i = 1, ..., m$, e questo è equivalente a dire che $\mathbf{x}$ è ortogonale a ogni riga di A (cioè a ogni vettore di una base di $rowA$), che è equivalente a dire che è ortogonale a $rowA$.

(ii) $\ker A = (rowA)^{\perp} \iff (\ker A)^{\perp} = rowA$: se due insiemi sono uguali, anche i loro complementi ortogonali coincidono.

(iii) $\ker A^T = \left(rowA^T\right)^{\perp}$ per (i), ma $rowA^T = colA$ perché la trasposizione scambia righe e colonne, quindi $\ker A^T = (colA)^{\perp}$.

(iv) $\ker A^T = (colA)^{\perp} \iff \left(\ker A^T\right)^{\perp} = colA$, come in (ii). ■

1. Dato $V = \mathbb{R}^n$, $H = \left(\begin{bmatrix} x_1 \\ ... \\ x_n \end{bmatrix} \in \mathbb{R}^n : a_1 x_1 + ... + a_n x_n = 0 \right)$ è un iperpiano di $\mathbb{R}^n$. Se $A = [a_1 | ... | a_n]$, $\ker A = H$ $\left((rowA)^{\perp}\right)$ e $H^{\perp} = (\ker A)^{\perp} = rowA = Span \begin{pmatrix} a_1 \\ ... \\ a_n \end{pmatrix}$. Quindi questo teorema permette di trovare molto facilmente $H^{\perp}$, senza bisogno di trovare una base di H né risolvere un sistema lineare.

5.2 Sistemi lineari sovradeterminati

Considero un sistema lineare $A\mathbf{x} = \mathbf{b}$ non risolubile: per il teorema di Rouché-Capelli, la non risolubilità è equivalente a dire che $r(A|\mathbf{b}) = r(A) + 1$, che è equivalente a dire che $\mathbf{b} \notin col(A)$. Dunque non c'è una soluzione: ha senso chiedersi qual è la migliore soluzione approssimata, i. e. $\tilde{\mathbf{x}} \in \mathbb{R}^n : A\tilde{\mathbf{x}}$ dà un vettore che è il più vicino possibile a $\mathbf{b}$, cioè $\tilde{\mathbf{x}} \in \mathbb{R}^n : ||A\tilde{\mathbf{x}} - \mathbf{b}|| \leq ||A\mathbf{x} - \mathbf{b}||$ per ogni $\mathbf{x} \in \mathbb{R}^n$.

Def Dato un sistema lineare $A\mathbf{x} = \mathbf{b}$, $A \in M_{\mathbb{R}}(m, n)$, si dice che $\tilde{\mathbf{x}}$ è una soluzione ai minimi quadrati del sistema se $A\tilde{\mathbf{x}}$ è la proiezione ortogonale di $\mathbf{b}$ su $colA$ ($A\tilde{\mathbf{x}} = P_{col(A)}(\mathbf{b})$).

E' un'applicazione del teorema della proiezione ortogonale con $H = colA$, $\mathbf{w} = A\mathbf{x}$, $\mathbf{v} = \mathbf{b}$. Quindi $A\tilde{\mathbf{x}}$ è, tra tutti i vettori di $colA$ (per cui quindi il sistema è risolubile), il più vicino a $\mathbf{b}$.

Quindi ora si ha un metodo più generale di risoluzione dei sistemi lineari: se $\mathbf{b} \in colA$ (cioè $P_{col(A)}(\mathbf{b}) = \mathbf{b}$), si risolve normalmente $A\mathbf{x} = \mathbf{b}$ (e ogni soluzione esatta è anche soluzione ai minimi quadrati), se $\mathbf{b} \notin colA$ si risolve $A\mathbf{x} = \mathbf{b}'$ con $P_{col(A)}(\mathbf{b}) = \mathbf{b}'$, che ha sicuramente soluzione perché $\mathbf{b}' \in colA$.

Teorema (soluzione ai minimi quadrati)

Hp: $A\mathbf{x}=\mathbf{b}$, $A\in M_{\mathbb{R}}(m,n)$

Ts: $\tilde{\mathbf{x}}\in\mathbb{R}^n$ è una soluzione ai minimi quadrati di $A\mathbf{x}=\mathbf{b}\iff\tilde{\mathbf{x}}$ è soluzione di $A^TA\mathbf{x}=A^T\mathbf{b}$

$A^TA\in M_{\mathbb{R}}(n,n)$. Risolvendo $A^TA\mathbf{x}=A^T\mathbf{b}$, si può poi calcolare $P_{col(A)}(\mathbf{b})=A\tilde{\mathbf{x}}$, con $\tilde{\mathbf{x}}$ soluzione ai minimi quadrati.

Dim (i) Per ipotesi $\tilde{\mathbf{x}}\in\mathbb{R}^n$ è una soluzione ai minimi quadrati di $A\mathbf{x}=\mathbf{b}$, quindi $A\tilde{\mathbf{x}}=P_{col(A)}(\mathbf{b})=\mathbf{b}'$. $\mathbf{b}-\mathbf{b}'=\mathbf{b}-P_{col(A)}(\mathbf{b})\in(colA)^{\perp}$ (per il teorema sulla proiezione ortogonale) e $(colA)^{\perp}=\ker\left(A^T\right)$. Quindi $\mathbf{b}-A\tilde{\mathbf{x}}\in\ker\left(A^T\right)$, cioè $A^T(\mathbf{b}-A\tilde{\mathbf{x}})=\mathbf{0}\iff A^TA\tilde{\mathbf{x}}=A^T\mathbf{b}$.

(ii) $\tilde{\mathbf{x}}$ è soluzione di $A^TA\mathbf{x}=A^T\mathbf{b}$, cioè $A^T(A\tilde{\mathbf{x}}-\mathbf{b})=\mathbf{0}$, quindi $\mathbf{b}-\mathbf{A}\tilde{\mathbf{x}}\in\ker\left(A^T\right)=(colA)^{\perp}$. Poiché $A\tilde{\mathbf{x}}\in colA$, dev'essere $A\tilde{\mathbf{x}}=P_{colA}(\mathbf{b})$. ■

In vari casi concreti, se e. g. si dispone di più osservazioni per il valore di alcuni parametri α,β,γ, non esistono α,β,γ che siano perfettamente in accordo con tutte le osservazioni: il sistema che ha come vettore dei termini noti le osservazioni per tali parametri (e. g. $\alpha^E_{i,i+1},\beta^E_{i,i+1},\gamma^E_{i,i+1},\alpha^E_{i+1,i+2},\beta^E_{i+1,i+2},\gamma^E_{i+1,i+2}$) e come incognite α,β,γ probabilmente non ha soluzione. Allora si risolve il sistema ai minimi quadrati associato e si trova $\begin{pmatrix}\alpha\\ \beta\\ \gamma\end{pmatrix}$ che si avvicina il più possibile ai valori osservati.

Def Dato il sistema $A\mathbf{x}=\mathbf{b}$, il sistema $A^TA\mathbf{x}=A^T\mathbf{b}$ si dice sistema delle equazioni normali associato a $A\mathbf{x}=\mathbf{b}$.

Se $\mathbf{b}\in colA$, ogni soluzione esatta è anche soluzione ai minimi quadrati, cioè $A\mathbf{x}=\mathbf{b}$ e $A^TA\mathbf{x}=A^T\mathbf{b}$ hanno le stesse soluzioni.

Quante soluzioni ha $A^TA\mathbf{x}=A^T\mathbf{b}$?

Proposizione

Hp: $A\in M_{\mathbb{R}}(m,n)$

Ts: (i) $\ker A=\ker\left(A^TA\right)$

(ii) $r(A)=r\left(A^TA\right)$

Dim (i) Mostro $\ker A\subseteq\ker\left(A^TA\right)$: $\mathbf{v}\in\ker A\iff A\mathbf{v}=\mathbf{0}\implies A^T(A\mathbf{v})=A^T\mathbf{0}\iff A^T(A\mathbf{v})=\mathbf{0}\implies\mathbf{v}\in\ker\left(A^TA\right)$. Mostro che $\ker\left(A^TA\right)\subseteq\ker A$: $\mathbf{v}\in\ker\left(A^TA\right)\iff A^TA\mathbf{v}=\mathbf{0}$. Premoltiplico per $\mathbf{v}^T$: $\mathbf{v}^TA^TA\mathbf{v}=\langle\mathbf{v},\mathbf{0}\rangle=\mathbf{0}$. D'altra parte $\mathbf{v}^TA^TA\mathbf{v}=\left(\mathbf{v}^TA^T\right)A\mathbf{v}=(A\mathbf{v})^TA\mathbf{v}=\langle A\mathbf{v},A\mathbf{v}\rangle$. Quindi $\langle A\mathbf{v},A\mathbf{v}\rangle=\mathbf{0}\iff||A\mathbf{v}||^2=\mathbf{0}\iff A\mathbf{v}=\mathbf{0}\iff\mathbf{v}\in\ker A$. Dunque $\ker A=\ker\left(A^TA\right)$.

(ii) Poiché $\ker A = \ker\left(A^T A\right)$, $\dim(\ker A) = \dim\left(\ker\left(A^T A\right)\right)$. Per il teorema di nullità più rango $r(A) = n - \dim(\ker A)$ e $r\left(A^T A\right) = n - \dim\left(\ker A^T A\right)$, quindi $r(A) = r\left(A^T A\right)$. ■

Corollario

$$\text{Hp: } A\mathbf{x} = \mathbf{b},\ A \in M_{\mathbb{R}}(m,n)$$

$$\text{Ts: la soluzione ai minimi quadrati di } A\mathbf{x} = \mathbf{b} \text{ è unica } \iff r(A) = n$$

Infatti la soluzione ai minimi quadrati di $A\mathbf{x} = \mathbf{b}$ è unica se e solo se $A^T A\mathbf{x} = A^T\mathbf{b}$ ha un'unica soluzione, cioè se e solo se $r\left(A^T A\right) = n = r(A)$, per il teorema di Cramer.

5.3 Matrice rappresentativa di una proiezione ortogonale

Data $A \in M_{\mathbb{R}}(m,n)$ e il sistema $A\mathbf{x} = \mathbf{b}$, il sistema delle equazioni normali a esso associato è $A^T A\mathbf{x} = A^T\mathbf{b}$; se $\tilde{\mathbf{x}}$ è una soluzione ai minimi quadrati di $A\mathbf{x} = \mathbf{b}$, $A\tilde{\mathbf{x}} = P_{col(A)}(\mathbf{b})$. Affinché la soluzione di $A^T A\mathbf{x} = A^T\mathbf{b}$ sia unica occorre $r(A) = n$. Con questa ipotesi si cerca di trovare la matrice rappresentativa di $P_{col(A)}(\mathbf{x})$.

Poiché $r\left(A^T A\right) = n$, $A^T A$ è invertibile, quindi $A^T A\mathbf{x} = A^T\mathbf{b} \iff \left(A^T A\right)^{-1}\left(A^T A\right)\mathbf{x} = \left(A^T A\right)^{-1} A^T\mathbf{b} \iff \mathbf{x} = \left(A^T A\right)^{-1} A^T\mathbf{b}$. Chiamo questa soluzione $\tilde{\mathbf{x}}$: dato che è una soluzione ai minimi quadrati, vale $A\tilde{\mathbf{x}} = P_{col(A)}(\mathbf{b})$. Quindi $P_{col(A)}(\mathbf{b}) = A\left(A^T A\right)^{-1} A^T\mathbf{b}$, e la matrice rappresentativa della proiezione ortogonale su $col(A)$, rispetto alla base canonica, è $A\left(A^T A\right)^{-1} A^T$. Ha una forma poco bella.

Tale formula si può usare per la matrice rappresentativa della proiezione ortogonale su qualsiasi sottospazio H, a patto di conoscere una sua base B: in tal caso la matrice A avrà come colonne i vettori di B, così che $col(A) = H$ e $r(A) = n$.

1. $B = \begin{bmatrix} 3 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & -2 & -1 \end{bmatrix}$: qual è la matrice rappresentativa della proiezione ortogonale su $colB$? Prima di applicare la formula devo assicurarmi che $r(B) = 3$. Riducendo a scala trovo che solo le prime due colonne sono linearmente indipendenti, quindi posso applicare la formula a $A = \begin{bmatrix} 3 & 0 \\ 1 & 1 \\ 1 & -2 \end{bmatrix}$ (il sottospazio su cui si calcola la proiezione ortogonale è lo stesso) e calcolo $A\left(A^T A\right)^{-1} A^T$.

Poiché si è visto che l'applicazione lineare proiezione ortogonale è idempotente, dev'esserlo anche la sua matrice rappresentativa: infatti $\left(A\left(A^T A\right)^{-1} A^T\right)\left(A\left(A^T A\right)^{-1} A^T\right) = A\left(A^T A\right)^{-1}\left(A^T A\right)\left(A^T A\right)^{-1} A^T = A\left(A^T A\right)^{-1} A^T$ poiché $\left(A^T A\right)\left(A^T A\right)^{-1} = Id_n$.

Proposizione

Hp: $(V, \langle \_, \_ \rangle)$ è uno spazio euclideo di dimensione finita, $H \subseteq V$ è un

sottospazio con $\dim H = h \leq n$, $B = \{\mathbf{b}_1, ..., \mathbf{b}_h\}$ è una base ortonormale di H

Ts: la matrice rappresentativa della proiezione ortogonale su H è AA^T, con $A = [\mathbf{b}_1|...|\mathbf{b}_h]$

La matrice di proiezione è sempre quadrata e di ordine uguale alla dimensione di V. In base alla forma vista prima $A\left(A^TA\right)^{-1}A^T$, se B è ortonormale, $\left(A^TA\right)^{-1} = Id_n$.

Dim Per costruzione, essendo B una base di H, $r(A) = h$, quindi la soluzione di $A^TA\mathbf{x} = A^T\mathbf{b}$ (soluzione ai minimi quadrati) è unica. La matrice rappresentativa della proiezione ortogonale su H è $A\left(A^TA\right)^{-1}A^T$, per quanto visto. Mostro che $\left(A^TA\right)^{-1} = Id_n$: calcolo $A^TA = \begin{bmatrix} \mathbf{b}_1 \\ ... \\ \mathbf{b}_h \end{bmatrix} [\mathbf{b}_1|...|\mathbf{b}_h] = \begin{bmatrix} \mathbf{b}_1^T\mathbf{b}_1 & \mathbf{b}_1^T\mathbf{b}_2 & ... \\ \mathbf{b}_2^T\mathbf{b}_1 & \mathbf{b}_2^T\mathbf{b}_2 & ... \\ ... & ... & ... \end{bmatrix} =$

$\begin{bmatrix} \langle \mathbf{b}_1, \mathbf{b}_1 \rangle & \langle \mathbf{b}_1, \mathbf{b}_2 \rangle & ... \\ \langle \mathbf{b}_2, \mathbf{b}_1 \rangle & \langle \mathbf{b}_2, \mathbf{b}_2 \rangle & ... \\ ... & ... & ... \end{bmatrix} = \begin{bmatrix} ||\mathbf{b}_1||^2 & 0 & ... \\ 0 & ||\mathbf{b}_2||^2 & ... \\ ... & ... & ... \end{bmatrix} = \begin{bmatrix} 1 & 0 & ... \\ 0 & 1 & ... \\ ... & ... & ... \end{bmatrix}$ perché la base è ortogonale e normale.

Quindi $A\left(A^TA\right)^{-1}A^T = AA^T$. ■

Quindi per calcolare la matrice rappresentativa di P_H ci sono più possibilità: se $\{\mathbf{v}_1, ..., \mathbf{v}_h\}$ è una base di H,

- pongo $A = [\mathbf{v}_1|...|\mathbf{v}_h]$ e calcolo $A\left(A^TA\right)^{-1}A^T$
- data la base $\{\mathbf{v}_1, ..., \mathbf{v}_h\}$, ne ricavo una base ortonormale $\{\mathbf{b}_1, ..., \mathbf{b}_h\}$ con l'algoritmo di Gram-Schmidt e calcolo BB^T, con $B = [\mathbf{b}_1|...|\mathbf{b}_h]$
- usando il teorema di rappresentazione, calcolo esplicitamente $P_H(\mathbf{e}_1), ...P_H(\mathbf{e}_h)$ utilizzando i coefficienti di Fourier, dopo aver ricavato una base ortonormale $\{\mathbf{b}_1, ..., \mathbf{b}_h\}$ con l'algoritmo di Gram-Schmidt.

1. Sia $V = \mathbb{R}^3$ dotato di prodotto scalare standard, $H = Span\left(\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}\right)$. Qual è la matrice rappresentativa di $P_H : \mathbb{R}^3 \to \mathbb{R}^3$? Uso la seconda tecnica: devo solo normalizzare l'unico generatore presente, quindi ottengo come base ortonormale di H $\left\{\frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}\right\}$ e $B = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$. Calcolo $AA^T = \frac{1}{5}\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}[1|2|0] =$

$\frac{1}{5}\begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. La matrice ottenuta è simmetrica, cioè è uguale alla sua trasposta.

2. Se P è la matrice rappresentativa della proiezione ortogonale su un sottospazio H, $Q = Id - P$ è la matrice rappresentativa della proiezione su $H^{\perp}$. Infatti, se $\mathbf{v}_H$ è la proiezione di $\mathbf{v}$ su H, $\mathbf{v}_{H^{\perp}} = \mathbf{v} - \mathbf{v}_H$ è la proiezione di $\mathbf{v}$ su $H^{\perp}$. Quindi la proiezione di $\mathbf{v}$ su $H^{\perp}$ è $Q\mathbf{v} = \mathbf{v}_{H^{\perp}} = \mathbf{v} - \mathbf{v}_H = Id\mathbf{v} - P\mathbf{v} = (Id - P)\,\mathbf{v}$.

In generale, per qualsiasi A, $P = AA^T$ è tale che $P^T = \left(AA^T\right)^T = AA^T = P$, quindi P è simmetrica. Dunque ogni matrice rappresentativa di una proiezione ortogonale è simmetrica (se ne ricava che una matrice simmetrica non necessariamente è invertibile), e, per quanto visto, diagonalizzabile: è simile a una matrice diagonale D ($P = SDS^{-1} \iff D = S^{-1}PS$) e tale che $P^T = P$. Quindi $\left(SDS^{-1}\right)^T = \left(S^{-1}\right)^T D^T S^T = \left(S^{-1}\right)^T DS^T$ dev'essere uguale a SDS^{-1}. Parrebbe quindi che $S^{-1} = S^T$.

1. Che A sia simmetrica è condizione necessaria, ma non sufficiente, affinché una matrice rappresenti una proiezione ortogonale. Infatti la matrice $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ è simmetrica, ma ha come autovalori $0, 1, -1$, quindi non rappresenta una proiezione ortogonale poiché la matrice rappresentativa di una proiezione ortogonale può avere solo $0, 1$ come autovalori.

5.3.1 Matrici ortogonali

Def Una matrice $Q \in M_{\mathbb{R}}(n, n)$ si dice ortogonale se $Q^TQ = Id_n$.

Significa che Q è invertibile a sinistra, con inversa sinistra uguale alla sua trasposta.

Proposizione

$$\text{Hp: } Q \in M_{\mathbb{R}}(n, n)$$

$$\text{Ts: } Q \text{ è ortogonale} \iff \text{le sue colonne descrivono una base ortonormale di } \mathbb{R}^n$$

Dim $Q = [\mathbf{q}_1|...|\mathbf{q}_n]$: $Q^TQ = \begin{bmatrix} \mathbf{q}_1 \\ ... \\ \mathbf{q}_n \end{bmatrix} [\mathbf{q}_1|...|\mathbf{q}_n] = \begin{bmatrix} \mathbf{q}_1^T\mathbf{q}_1 & \mathbf{q}_1^T\mathbf{q}_2 & ... \\ \mathbf{q}_2^T\mathbf{q}_1 & \mathbf{q}_2^T\mathbf{q}_2 & ... \\ ... & ... & ... \end{bmatrix} = \begin{bmatrix} ||\mathbf{q}_1||^2 & 0 & ... \\ 0 & ||\mathbf{q}_2||^2 & ... \\ ... & ... & ... \end{bmatrix} =$

$$\begin{bmatrix} 1 & 0 & ... \\ 0 & 1 & ... \\ ... & ... & ... \end{bmatrix} = Id$$ se e solo se $\{\mathbf{q}_1, ..., \mathbf{q}_n\}$ è una base ortonormale di $\mathbb{R}^n$ secondo il prodotto scalare standard.

■

1. La matrice che rappresenta la riflessione rispetto a una retta che forma un angolo θ con l'asse x (simmetria assiale) è $S_\theta = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}$: le due colonne costituiscono una base ortonormale di $\mathbb{R}^n$.

   La matrice che rappresenta la rotazione di un angolo θ in senso orario è $R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$: le due colonne costituiscono una base ortonormale di $\mathbb{R}^n$.

Se Q è ortogonale, anche Q^T è ortogonale, perché è anche inversa destra di Q, per cui $QQ^T = \left(Q^T\right)^T Q^T = Id$. Quindi, se le colonne di Q formano una base ortonormale di $\mathbb{R}^n$, anche le sue righe formano una base ortonormale di $\mathbb{R}^n$.

Proprietà

L'insieme delle matrici ortogonali è un gruppo rispetto al prodotto matriciale, con elemento neutro Id_n ed elemento inverso la matrice inversa. Infatti:

- Id_n è una matrice ortogonale: $Id_n^T Id_n = Id_n Id_n = Id_n$.
- Se Q_1, Q_2 sono ortogonali, Q_1Q_2 è ortogonale (cioè la composizione di isometrie è un'isometria): $(Q_1Q_2)^T (Q_1Q_2) = Q_2^T Q_1^T Q_1 Q_2 = Q_2^T Id_n Q_2$ (perché Q_1 è ortogonale) $= Q_2^T Q_2 = Id_n$ (perché Q_2 è ortogonale).
- $\det Q = \pm 1$ per ogni Q ortogonale. Infatti $\det\left(Q^T Q\right) = \det Q^T \det Q = \det^2 Q$ (una matrice e la sua trasposta hanno lo stesso determinante). Dev'essere $\det\left(Q^T Q\right) = \det Id = 1$, quindi $\det^2 Q = 1 \iff \det Q = 1 \vee \det Q = -1$.
- Se Q è ortogonale, Q è invertibile, infatti $\det Q \neq 0$, quindi $r(Q) = n$. Perciò, essendo Q invertibile e avendo come inversa sinistra Q^T, per il teorema dell'inversa si può concludere che la matrice inversa di Q è Q^T.

1. $\det S_\theta = -1$, $\det R_\theta = 1$. Il segno del determinante ha un significato geometrico. Infatti $\mathbf{e}_1 \times \mathbf{e}_2 = R(\mathbf{e}_1) \times R(\mathbf{e}_2)$, mentre $\mathbf{e}_1 \times \mathbf{e}_2 = -S(\mathbf{e}_1) \times S(\mathbf{e}_2)$: $\det R_\theta = 1$ significa che l'ordine dei due vettori è preservato dalla trasformazione (come nel caso di rotazioni), $\det S_\theta = -1$ significa che è invertito (come nel caso di simmetrie/riflessioni). Ne segue che la composizione di due riflessioni è una rotazione, mentre la composizione di riflessione e rotazione è una riflessione.

- Se Q è ortogonale, la trasformazione indotta da Q preserva il prodotto scalare in $\mathbb{R}^n$: dati $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, $\langle \mathbf{v}, \mathbf{w} \rangle = \langle Q\mathbf{v}, Q\mathbf{w} \rangle$. Infatti $\langle Q\mathbf{v}, Q\mathbf{w} \rangle = (Q\mathbf{v})^T Q\mathbf{w} = \mathbf{v}^T Q^T Q\mathbf{w} = \mathbf{v}^T Id_n \mathbf{w} = \mathbf{v}^T \mathbf{w} = \langle \mathbf{v}, \mathbf{w} \rangle$. Anche la norma è preservata: $||\mathbf{v}||^2 = \langle \mathbf{v}, \mathbf{v} \rangle = \langle Q\mathbf{v}, Q\mathbf{v} \rangle = ||Q\mathbf{v}||^2$. Quindi, secondo la definizione di angolo data, $\hat{\mathbf{v}}\mathbf{w} = Q(\mathbf{v})\hat{Q}(\mathbf{w})$, cioè Q preserva gli angoli: si dice che Q rappresenta un'isometria (lineare) di $\mathbb{R}^n$. Infatti le matrici ortogonali codificano tutte le trasformazioni rigide (senza dilatazioni o contrazioni), come rotazioni o simmetrie.

- Se Q è ortogonale e λ è un autovalore di Q con relativo autovettore $\mathbf{v}$, $Q\mathbf{v} = \lambda\mathbf{v}$ e $||Q\mathbf{v}|| = ||\lambda\mathbf{v}|| = |\lambda|\,||\mathbf{v}||$. Per quanto visto $||\mathbf{v}|| = ||Q\mathbf{v}||$, quindi dev'essere $|\lambda|\,||\mathbf{v}|| = ||\mathbf{v}|| \iff |\lambda| = 1$ ($\mathbf{v} \neq \mathbf{0}$). Quindi $\lambda = -1$ o $\lambda = 1$ sono i possibili autovalori di una matrice ortogonale.

Le matrici ortogonali che rappresentano simmetrie rispetto a un sottospazio sono simmetriche e quindi ortogonalmente diagonalizzabili.

In generale le matrici di proiezione su H non sono ortogonali, perché possono avere nucleo di dimensione non nulla, quindi rango non massimo e determinante nullo.

Le matrici ortogonali possono anche essere viste come matrici di cambiamento di base tra due basi ortonormali.

5.4 Teorema spettrale

Ora che si sono studiate le proprietà delle matrici ortogonali, riprendiamo la definizione di matrice ortogonalmente diagonalizzabile: A si dice ortogonalmente diagonalizzabile se è diagonalizzabile mediante una base ortonormale di autovettori di A.

Sia $B = \{\mathbf{b}_1, ..., \mathbf{b}_n\}$ una base ortonormale di $\mathbb{R}^n$ formata da autovettori di A: se A è diagonalizzabile mediante una base ortonormale di autovettori, data $P = [\mathbf{b}_1|...|\mathbf{b}_n]$, $D = P^{-1}AP$. Essendo B ortonormale, P è ortogonale, quindi $P^{-1} = P^T$. Perciò si può riformulare la definizione tenendo conto delle proprietà delle matrici ortogonali.

Def Una matrice $A \in M_{\mathbb{R}}(n, n)$ si dice ortogonalmente diagonalizzabile se esiste una matrice ortogonale $P \in M_{\mathbb{R}}(n, n)$ tale che $P^T AP$ è diagonale.

Lemma 1

Hp:	$A \in M_{\mathbb{R}}(n, n)$ è ortogonalmente diagonalizzabile
Ts:	A è simmetrica

Dim Per definizione $P^T AP = D$ è diagonale, con P matrice ortogonale. Mostro che $A = A^T$: $A = \left(P^T\right)^{-1} DP^{-1} = \left(P^{-1}\right)^{-1} DP^T = PDP^T$ perché P ha come inversa P^T. Allora $A^T = \left(PDP^T\right)^T = \left(P^T\right)^T D^T P^T = PDP^T = A$, dunque A è simmetrica. ■

Dunque la simmetria è condizione necessaria per l'ortogonale diagonalizzabilità. Si vedrà che è anche sufficiente.

Se è noto che A è ortogonalmente diagonalizzabile, allora lo è anche semplicemente, quindi si possono usare le tecniche standard per scrivere $\mathbb{R}^n$ come $V_{\lambda_1} \oplus ... \oplus V_{\lambda_s}$ e ottenere una base di $\mathbb{R}^n$ formata da autovettori di A: per ogni λ_i, $\left\{\mathbf{v}_{i1}, ..., \mathbf{v}_{ig(\lambda_i)}\right\}$ è una base di V_{λ_i}. Per costruire una base ortonormale, si usa l'algoritmo di Gram-Schmidt all'interno di ogni autospazio e si ricava una base ortonormale di ogni autospazio: $\left\{\mathbf{b}_{i1}, ..., \mathbf{b}_{ig(\lambda_i)}\right\}$ è una base ortonormale di V_{λ_i}. Non si può usare l'algoritmo sull'unione di tutte le basi degli autospazi, perché combinando linearmente autovettori di autospazi diversi non necessariamente si ottengono autovettori.

Lemma 2

$$\text{Hp: } A \in M_{\mathbb{R}}(n,n) \text{ è simmetrica, } \lambda, \mu \text{ sono autovalori distinti di } A,$$

$$\mathbf{v} \in V_\lambda, \mathbf{w} \in V_\mu \text{ (autospazi relativi agli autovalori } \lambda, \mu)$$

$$\text{Ts: } \langle \mathbf{v}, \mathbf{w} \rangle = 0 \ (V_\lambda \subseteq (V_\mu)^\perp)$$

Quindi autovettori di una matrice simmetrica relativi ad autospazi distinti sono ortogonali: dati $\mathbf{v} \in V_\lambda, \mathbf{w} \in V_\mu$, questi sono ortogonali; in questo senso si può dire che due autospazi distinti di una matrice simmetrica sono ortogonali. Perciò, avendo una base ortonormale di ogni autospazio, unendo tutte le basi ortonormali si ottiene una base formata da autovettori ancora ortonormale: $B = \left\{\mathbf{b}_{11}, ..., \mathbf{b}_{1g(\lambda_s)}\right\} \cup ... \cup \left\{\mathbf{b}_{s1}, ..., \mathbf{b}_{sg(\lambda_s)}\right\}$.

Dim Calcolo $\langle A\mathbf{v}, \mathbf{w} \rangle = (A\mathbf{v})^T \mathbf{w} = \left(\mathbf{v}^T A^T\right) \mathbf{w} = \mathbf{v}^T (A\mathbf{w}) = \langle \mathbf{v}, A\mathbf{w} \rangle$, usando il fatto che A è simmetrica. Calcolo $\langle A\mathbf{v}, \mathbf{w} \rangle = \langle \lambda\mathbf{v}, \mathbf{w} \rangle = \lambda \langle \mathbf{v}, \mathbf{w} \rangle$ e $\langle \mathbf{v}, A\mathbf{w} \rangle = \langle \mathbf{v}, \mu\mathbf{w} \rangle = \mu \langle \mathbf{v}, \mathbf{w} \rangle$. Poiché $\langle A\mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, A\mathbf{w} \rangle$, $\lambda \langle \mathbf{v}, \mathbf{w} \rangle = \mu \langle \mathbf{v}, \mathbf{w} \rangle \iff \lambda \langle \mathbf{v}, \mathbf{w} \rangle - \mu \langle \mathbf{v}, \mathbf{w} \rangle = 0 \iff (\lambda - \mu) \langle \mathbf{v}, \mathbf{w} \rangle = 0$: per la legge di annullamento del prodotto, essendo $\lambda \neq \mu$, dev'essere $\langle \mathbf{v}, \mathbf{w} \rangle = 0$. ■

1. Nel caso in cui A simmetrica abbia solo due autovalori e quindi solo due autospazi, si ha $V_\lambda = (V_\mu)^\perp$. Infatti $V_\lambda \subseteq (V_\mu)^\perp$, $\dim V_\lambda = n - \dim V_\mu$ (perché A è diagonalizzabile) e $\dim (V_\mu)^\perp = n - \dim V_\mu$, quindi $V_\lambda = (V_\mu)^\perp$. Si noti che quanto detto vale anche nel caso di A non simmetrica, tranne $V_\lambda \subseteq (V_\mu)^\perp$.

Def Dato uno spazio euclideo $V, \langle \_, \_ \rangle$ di dimensione finita e un'applicazione lineare $L : V \to V$, si dice operatore aggiunto di L un'applicazione lineare $L^* : V \to V$ tale che $\langle L(\mathbf{v}), \mathbf{w} \rangle = \langle \mathbf{v}, L^*(\mathbf{w}) \rangle \ \forall \ \mathbf{v}, \mathbf{w} \in V$.

La nozione di operatore aggiunto generalizza da matrici ad applicazioni lineari qualsiasi l'operazione di trasposizione: se $L(\mathbf{v}) = A\mathbf{v}$, $L^*(\mathbf{v}) = A^t\mathbf{v}$.

Def Dato uno spazio euclideo $V, \langle \_, \_ \rangle$ di dimensione finita, un'applicazione lineare $L : V \to V$ si dice autoaggiunta se $\langle L(\mathbf{v}), \mathbf{w} \rangle = \langle \mathbf{v}, L(\mathbf{w}) \rangle \ \forall \ \mathbf{v}, \mathbf{w} \in V$.

Quindi L è autoaggiunta se e solo se $L = L^*$.

Se $V = \mathbb{R}^n$ e $\langle \mathbf{v}, \mathbf{w} \rangle$ è il prodotto scalare canonico $\mathbf{v} \cdot \mathbf{w}$, L è autoaggiunta se e solo se $L =_A$ con A simmetrica. (se A è simmetrica tutte le matrici che rappresentano la stessa applicazione lineare sono simmetriche? No a meno che si usi come cambiamento di base una matrice ortogonale)

Lemma 3

$$\begin{array}{lll} \text{Hp} & : & A \in M_{\mathbb{R}}(n,n) \text{ è simmetrica} \\ \text{Ts} & : & A \text{ ha almeno un autovalore reale} \end{array}$$

Vedendo $A \in M_{\mathbb{R}}(n,n)$ come $A \in M_{\mathbb{C}}(n,n)$, per il teorema fondamentale dell'algebra $p_A(x) = \det(A - xId) \in \mathbb{C}[x]$ è tale che $p_A(x) = (x - \alpha_1) ... (x - \alpha_n)$, $\alpha_i \in \mathbb{C}$. Le radici di $p_A(x)$ potrebbero essere tutte complesse: il lemma aggiunge che se A è simmetrica, non sono tutte complesse.

Teorema spettrale per operatori autoaggiunti

$$\begin{array}{ll} \text{Hp:} & A \in M_{\mathbb{R}}(n,n) \\ \text{Ts:} & A \text{ è ortogonalmente diagonalizzabile} \Longleftrightarrow A \text{ è simmetrica} \end{array}$$

Quindi, essendo A diagonalizzabile su $\mathbb{R}$, ha n autovalori in $\mathbb{R}$ contati con molteplicità algebrica, i. e. i suoi autovalori sono tutti reali.

Dim L'implicazione da sinistra a destra è dimostrata nel lemma 1.

Si dimostra l'implicazione da destra a sinistra per induzione su n. Se $n = 1$, $A = [a_{11}]$ (necessariamente simmetrica) è ortogonalmente diagonalizzabile perché, essendo diagonale, è diagonalizzabile e ogni base di $\mathbb{R}^1 = \mathbb{R}$ è ortogonale, essendo fatta di un solo elemento.

Suppongo, come ipotesi induttiva, che il teorema sia vero per tutti i $k \leq n - 1$. Voglio trovare un'opportuna matrice $k \times k$ cui applicare l'ipotesi induttiva. Considero $A \in M_{\mathbb{R}}(n,n)$ simmetrica per ipotesi: per il lemma 3, ha almeno un autovalore $\lambda \in \mathbb{R}$; $V_\lambda = \{\mathbf{v} \in \mathbb{R}^n : A\mathbf{v} = \lambda\mathbf{v}\}$. Pongo $H = V_\lambda^\perp$ (è noto che $H \oplus V_\lambda = \mathbb{R}^n$); $g_\lambda \geq 1$, quindi $\dim H = n - \dim V_\lambda \leq n - 1$. Osservo che, $\forall \ \mathbf{w} \in H$, $A\mathbf{w} \in H$ $(A(H) \subseteq H)$, cioè $\langle A\mathbf{w}, \mathbf{v} \rangle = 0$ per ogni $\mathbf{v} \in V_\lambda$: infatti $\mathbf{w} \in H \Longleftrightarrow \langle \mathbf{v}, \mathbf{w} \rangle = 0$ per ogni $\mathbf{v} \in V_\lambda$; $\langle A\mathbf{w}, \mathbf{v} \rangle = \langle A\mathbf{v}, \mathbf{w} \rangle$ perché $_A$ è autoaggiunta; $\langle A\mathbf{v}, \mathbf{w} \rangle = \lambda \langle \mathbf{v}, \mathbf{w} \rangle = 0$ per ipotesi, quindi $\langle A\mathbf{w}, \mathbf{v} \rangle = \langle A\mathbf{v}, \mathbf{w} \rangle = 0$ e H è un sottospazio invariante rispetto all'azione di $_A$. Considero $_A|_H : H \to H$ (perché H è invariante per $_A$: mi serve per avere una matrice rappresentativa quadrata): questa nuova applicazione lineare $_{A'} : \mathbb{R}^k \to \mathbb{R}^k$ ha matrice rappresentativa $A' \in M_{\mathbb{R}}(k,k)$, con $\dim H = k$ ($A' = A|_H$: è la matrice rappresentativa dell'applicazione lineare descritta da A ristretta a H). ${}_{A'}$ è autoaggiunta: la definizione si applica considerando che il prodotto scalare tra $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^k$ sia $\langle P(\mathbf{u}_1), P(\mathbf{u}_2)\rangle$ con $P : \mathbb{R}^k \to H$ mappa di parametrizzazione. Infatti per ogni $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^k$ $\langle A'\mathbf{u}_1, \mathbf{u}_2\rangle = \langle P(A'\mathbf{u}_1), P(\mathbf{u}_2)\rangle = \langle AP(\mathbf{u}_1), P(\mathbf{u}_2)\rangle = \langle P(\mathbf{u}_1), AP(\mathbf{u}_2)\rangle = \langle \mathbf{u}_1, A'\mathbf{u}_2\rangle$. Poiché ${}_{A'}$ è autoaggiunta, A' è simmetrica e $A' = (A')^T$: dunque, per ipotesi induttiva, A' è ortogonalmente diagonalizzabile ed esiste una base ortonormale (rispetto al prodotto scalare $\langle P(\mathbf{u}_1), P(\mathbf{u}_2)\rangle$) di $\mathbb{R}^k$ formata da autovettori di A': $\{\mathbf{u}_1, ..., \mathbf{u}_k\}$. La base ortonormale corrispondente di $H \subseteq \mathbb{R}^n$ è $B_H = \{P(\mathbf{u}_1), ..., P(\mathbf{u}_k)\}$, formata da autovettori di A'. Data una base qualsiasi di V_λ, la ortonormalizzo usando l'algoritmo di Gram-Schmidt e ottengo una nuova base B_{V_λ}. Allora $B = B_H \cup B_{V_\lambda}$ è una base ortonormale di $\mathbb{R}^n$, perché B_H e B_{V_λ} sono ciascuna una base ortonormale e ogni vettore di H è ortogonale a ogni vettore di V_λ; è inoltre una base formata da autovettori di A: i vettori di B_{V_λ} lo sono per costruzione, e i vettori di $B_H = B_{V_\lambda^\perp}$ lo sono perché autovettori di A'. Quindi A è ortogonalmente diagonalizzabile. ■

Con spettro ci si riferisce all'insieme di autovalori di un'applicazione lineare.

1. Se $A \in M_{\mathbb{R}}(3,3)$ è simmetrica e $\mathbf{u}, \mathbf{v}, \mathbf{w}$ sono autovettori relativi agli autovalori - rispettivamente $0, 3, -2$, allora $\ker A = V_0 = Span(\mathbf{u})$: infatti, se A è simmetrica, è diagonalizzabile e quindi $g_0 + g_3 + g_{-2} = 3$, con $g_0 = 1$. Inoltre $colA = rowA = (\ker A)^\perp = V_0^\perp = V_3 \oplus V_{-2} = Span(\mathbf{v}, \mathbf{w})$.

Corollario (decomposizione spettrale)

Hp: $A \in M_{\mathbb{R}}(n,n)$ è simmetrica; $\lambda_1, ..., \lambda_s$ sono autovalori distinti di A;

P_{λ_i} è la matrice che descrive la proiezione ortogonale su V_{λ_i}, con $i = 1, ..., s$

Ts: (i) $Id_n = P_{\lambda_1} + ... + P_{\lambda_s}$

(ii) $A = \lambda_1 P_{\lambda_1} + ... + \lambda_s P_{\lambda_s}$

(iii) $P_{\lambda_i}^2 = P_{\lambda_i}$, $(P_{\lambda_i})^T = P_{\lambda_i}$; $P_{\lambda_i} P_{\lambda_j} = 0_M$ per ogni $i \neq j$

$\lambda_1, ..., \lambda_s$ sono tutti reali per conseguenza del teorema spettrale.

Dim (ii) Poiché $A = A^T$, A è ortogonalmente diagonalizzabile ed esiste una base ortonormale di $\mathbb{R}^n$ formata da autovettori di A, che è l'unione di basi ortonormali di ciascun autospazio. Sia $Q \in M_{\mathbb{R}}(n,n) = [\mathbf{q}_1|...|\mathbf{q}_n]$ la matrice ortogonale che ha come colonne i vettori di tale base: $D = Q^{-1}AQ \iff A = QDQ^T$. Penso a Q e Q^T come somma di n matrici: $Q = [\mathbf{q}_1|\mathbf{0}|...|\mathbf{0}] + ... + [\mathbf{0}|...|\mathbf{0}|\mathbf{q}_n]$, $Q^T = \begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{0} \\ ... \\ \mathbf{0} \end{bmatrix} + ... + \begin{bmatrix} \mathbf{0} \\ ... \\ \mathbf{0} \\ \mathbf{q}_n^T \end{bmatrix}$. Allora

$$A = QDQ^T = ([\mathbf{q}_1|\mathbf{0}|...|\mathbf{0}] + ... + [\mathbf{0}|...|\mathbf{0}|\mathbf{q}_n]) \begin{bmatrix} \lambda_1 & ... & 0 \\ ... & ... & ... \\ 0 & ... & \lambda_s \end{bmatrix} \left(\begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{0} \\ ... \\ \mathbf{0} \end{bmatrix} + ... + \begin{bmatrix} \mathbf{0} \\ ... \\ \mathbf{0} \\ \mathbf{q}_n^T \end{bmatrix} \right) =$$

$$([\lambda_1\mathbf{q}_1|\mathbf{0}|...|\mathbf{0}] + ... + [\mathbf{0}|...|\mathbf{0}|\lambda_s\mathbf{q}_n]) \left(\begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{0} \\ ... \\ \mathbf{0} \end{bmatrix} + ... + \begin{bmatrix} \mathbf{0} \\ ... \\ \mathbf{0} \\ \mathbf{q}_n^T \end{bmatrix} \right) = \lambda_1\mathbf{q}_1\mathbf{q}_1^T + ... + \lambda_s\mathbf{q}_n\mathbf{q}_n^T.$$ Ma $\mathbf{q}_1\mathbf{q}_1^T$ è la matrice che rappresenta la proiezione ortogonale su $Span\,(\mathbf{q}_1)$,..., $\mathbf{q}_n\mathbf{q}_n^T$ è la matrice che rappresenta la proiezione ortogonale su $Span\,(\mathbf{q}_n)$. $\lambda_1\mathbf{q}_1\mathbf{q}_1^T + ... + \lambda_s\mathbf{q}_n\mathbf{q}_n^T = \lambda_1 \left(\mathbf{q}_1\mathbf{q}_1^T + ... + \mathbf{q}_{g(\lambda_1)}\mathbf{q}_{g(\lambda_1)}^T\right) + ... + \lambda_s \left(... + \mathbf{q}_n\mathbf{q}_n^T\right)$, dove $\mathbf{q}_1, ..., \mathbf{q}_{g(\lambda_1)}$ è una base ortonormale di V_{λ_1}. Allora $\mathbf{q}_1\mathbf{q}_1^T + ... + \mathbf{q}_{g(\lambda_1)}\mathbf{q}_{g(\lambda_1)}^T$ è la matrice che rappresenta la proiezione ortogonale su V_{λ_1}, e $... + \mathbf{q}_n\mathbf{q}_n^T$ è la matrice che rappresenta la proiezione ortogonale su V_{λ_s}. Quindi $A = \lambda_1 P_{\lambda_1} + ... + \lambda_s P_{\lambda_s}$.

(i) Se $Q \in M_{\mathbb{R}}\,(n, n)$, $Q = [\mathbf{q}_1|...|\mathbf{q}_n]$ ha come colonne i vettori di una base ortonormale di $\mathbb{R}^n$, Q è ortogonale, quindi $QQ^T = Id_n$, perciò ripeto il calcolo precedente scrivendo sia Q sia Q^T come somma di n matrici.

$$Q = [\mathbf{q}_1|\mathbf{0}|...|\mathbf{0}] + ... + [\mathbf{0}|...|\mathbf{0}|\mathbf{q}_n],\ Q^T = \begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{0} \\ ... \\ \mathbf{0} \end{bmatrix} + ... + \begin{bmatrix} \mathbf{0} \\ ... \\ \mathbf{0} \\ \mathbf{q}_n^T \end{bmatrix}.$$ Allora $Id = QQ^T =$

$$([\mathbf{q}_1|\mathbf{0}|...|\mathbf{0}] + ... + [\mathbf{0}|...|\mathbf{0}|\mathbf{q}_n]) \left(\begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{0} \\ ... \\ \mathbf{0} \end{bmatrix} + ... + \begin{bmatrix} \mathbf{0} \\ ... \\ \mathbf{0} \\ \mathbf{q}_n^T \end{bmatrix} \right) = \mathbf{q}_1\mathbf{q}_1^T + ... + \mathbf{q}_n\mathbf{q}_n^T.$$ Ma $\mathbf{q}_1\mathbf{q}_1^T$ è la matrice che rappresenta la proiezione ortogonale su $Span\,(\mathbf{q}_1)$,..., $\mathbf{q}_n\mathbf{q}_n^T$ è la matrice che rappresenta la proiezione ortogonale su $Span\,(\mathbf{q}_n)$. $\mathbf{q}_1\mathbf{q}_1^T + ... + \mathbf{q}_n\mathbf{q}_n^T = \left(\mathbf{q}_1\mathbf{q}_1^T + ... + \mathbf{q}_{g(\lambda_1)}\mathbf{q}_{g(\lambda_1)}^T\right) + ... + \left(... + \mathbf{q}_n\mathbf{q}_n^T\right)$, dove $\mathbf{q}_1, ..., \mathbf{q}_{g(\lambda_1)}$ è una base ortonormale di V_{λ_1} (raggruppo gli autovettori che costituiscono una base dello stesso autospazio). Allora $\mathbf{q}_1\mathbf{q}_1^T + ... + \mathbf{q}_{g(\lambda_1)}\mathbf{q}_{g(\lambda_1)}^T$ è la matrice che rappresenta la proiezione ortogonale su V_{λ_1}, e $... + \mathbf{q}_n\mathbf{q}_n^T$ è la matrice che rappresenta la proiezione ortogonale su V_{λ_s}. Quindi $Id_n = P_{\lambda_1} + ... + P_{\lambda_s}$.

(iii) Si è già mostrato che ogni matrice di proiezione ortogonale è idempotente e simmetrica. Si dimostra che $P_{\lambda_i}P_{\lambda_j} = 0_M$ per ogni $i \neq j$. Per ogni $\mathbf{v} \in \mathbb{R}^n$ calcolo $P_{\lambda_j}\mathbf{v} \in V_{\lambda_j}$. $P_{\lambda_i}\left(P_{\lambda_j}\mathbf{v}\right) = \mathbf{0} \iff P_{\lambda_j}\mathbf{v} \in \ker P_{\lambda_i}$, ma $\ker P_{\lambda_i} = V_{\lambda_i}^{\perp}$, e $V_{\lambda_j} \subseteq V_{\lambda_i}^{\perp}$ per il lemma 2, quindi $P_{\lambda_j}\mathbf{v} \in V_{\lambda_j} \subseteq V_{\lambda_i}^{\perp}$ e $P_{\lambda_i}\left(P_{\lambda_j}\mathbf{v}\right) = \mathbf{0}$. Poiché questo vale per ogni $\mathbf{v} \in \mathbb{R}^n$, $P_{\lambda_i}P_{\lambda_j}$ dev'essere la matrice nulla. ■

Usare la decomposizione spettrale per interpretare una matrice simmetrica permette di calcolare facilmente potenze: $A^2 = (\lambda_1 P_{\lambda_1} + ... + \lambda_s P_{\lambda_s})(\lambda_1 P_{\lambda_1} + ... + \lambda_s P_{\lambda_s}) = \lambda_1^2 P_{\lambda_1}^2 + ... + \lambda_1 \lambda_s P_{\lambda_1} P_{\lambda_s} + ... + \lambda_s^2 P_{\lambda_s}^2 = \lambda_1^2 P_{\lambda_1}^2 + ... + \lambda_s^2 P_{\lambda_s}^2 = \lambda_1^2 P_{\lambda_1} + ... + \lambda_s^2 P_{\lambda_s}$ perché tutti i prodotti misti si annullano e ogni matrice di proiezione è idempotente. In generale,

$$A^k = \lambda_1^k P_{\lambda_1} + ... + \lambda_s^k P_{\lambda_s}$$

Questo si poteva già fare con A semplicemente diagonalizzabile, ma ora si può calcolare e. g. $\sqrt{A}$, cioè l'unica matrice B tale che $B^2 = A$: $\sqrt{A} = \sqrt{\lambda_1} P_{\lambda_1} + ... + \sqrt{\lambda_s} P_{\lambda_s}$, infatti $B^2 = \lambda_1 P_{\lambda_1} + ... + \lambda_s P_{\lambda_s}$. Quindi si possono calcolare potenze con esponente non intero.

1. Si può definire, data $A \in M_{\mathbb{R}}(n,n)$, e^A usando lo sviluppo di Taylor: $e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + ...$, e $e^A = Id + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + ...$ Se $A = A^T$, si usa la decomposizione spettrale e si ha

$$\begin{aligned} e^A &= (\lambda_1 P_{\lambda_1} + ... + \lambda_s P_{\lambda_s})^0 + \lambda_1 P_{\lambda_1} + ... + \lambda_s P_{\lambda_s} + \frac{1}{2}(\lambda_1 P_{\lambda_1} + ... + \lambda_s P_{\lambda_s})^2 + \frac{1}{3!}(\lambda_1 P_{\lambda_1} + ... + \lambda_s P_{\lambda_s})^3 ... = \\ & P_{\lambda_1} + ... + P_{\lambda_s} + \lambda_1 P_{\lambda_1} + ... + \lambda_s P_{\lambda_s} + \frac{1}{2}\left(\lambda_1^2 P_{\lambda_1} + ... + \lambda_s^2 P_{\lambda_s}\right) + \frac{1}{3!}\left(\lambda_1^3 P_{\lambda_1} + ... + \lambda_s^3 P_{\lambda_s}\right) ... \\ &= \left(1 + \lambda_1 + \frac{1}{2}\lambda_1^2 + \frac{1}{3!}\lambda_1^3 + ...\right) P_{\lambda_1} + ... + \left(1 + \lambda_s + \frac{1}{2}\lambda_s^2 + \frac{1}{3!}\lambda_s^3 + ...\right) P_{\lambda_s} = \\ & e^{\lambda_1} P_{\lambda_1} + ... + e^{\lambda_s} P_{\lambda_s} \end{aligned}$$

6 Forme quadratiche

Le forme quadratiche reali sono funzioni $q : \mathbb{R}^n \to \mathbb{R}$, polinomi omogenei in cui ogni monomio ha grado due, che sono combinazione lineare dei quadratii delle variabili e dei prodotti di due variabili diverse.

(Una forma lineare è $[a_1|...|a_n]\,\mathbf{x} = a_1 x_1 + ... + a_n x_n$)

1. $q(x_1, ..., x_n) = x_1^2 + ... + x_n^2$ è una forma quadratica che definisce la norma al quadrato del vettore $\mathbf{x} = \begin{pmatrix} x_1 \\ ... \\ x_n \end{pmatrix}$, secondo il prodotto scalare standard di $\mathbb{R}^n$.

2. $q(x_1, x_2) = x_1^2 + 2x_1x_2 + 3x_2^2$ è una forma quadratica

Questo tipo di funzioni può essere studiato usando le matrici.

1. $q(x_1, x_2) = x_1^2 + 2x_1x_2 + 3x_2^2 = x_1(x_1 + 2x_2) + x_2(3x_2) = [x_1|x_2] \begin{bmatrix} x_1 + 2x_2 \\ 3x_2 \end{bmatrix} = [x_1|x_2] \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

   Quindi q può essere rappresentata con una matrice 2×2. Questa matrice però non è l'unica possibile: infatti

$q(x_1,x_2) = x_1(x_1+2x_2)+x_2(3x_2) = [x_1+2x_2|3x_2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = [x_1|x_2]\begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ La matrice ottenuta così è la trasposta della precedente. Ci sono infinite matrici che sono associabili a questa forma quadratica. Vale anche $q(x_1,x_2) = x_1^2+x_1x_2+x_1x_2+3x_2^2 = x_1(x_1+x_2)+x_2(x_1+3x_2) = [x_1+x_2|x_1+3x_2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = [x_1|x_2]\begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. La matrice così ottenuta è simmetrica, ed è l'unica matrice simmetrica associabile a q.

Data una matrice simmetrica $A \in M_{\mathbb{R}}(n,n)$, si definisce la forma quadratica $q_A : \mathbb{R}^n \to \mathbb{R}$, $q_A(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, con $A = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{12} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{1n} & a_{2n} & ... & a_{nn} \end{bmatrix}$; l'espressione analitica della forma quadratica è $q_A(x_1,...,x_n) = \sum_{i=1}^n a_{ii}x_i^2 + 2\sum_{1\le i\le j\le n} a_{ij}x_ix_j$. Quindi sulla diagonale principale vanno i coefficienti dei quadrati, in a_{ij} va il coefficiente di x_ix_j diviso per due.

Proprietà

- $q_A(\mathbf{0}) = 0$
- $q_A(t\mathbf{x}) = (t\mathbf{x})^T A(t\mathbf{x}) = t^2\left(\mathbf{x}^T A\mathbf{x}\right) = t^2 q_A(\mathbf{x})$
- posto $\mathbf{x} \ne \mathbf{0}$, $q_A(t\mathbf{x})\,q_A(\mathbf{x}) \ge 0$, cioè $q_A(t\mathbf{x})$ e $q_A(\mathbf{x})$ sono concordi

Def Posto $q_A(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, $A \in M_{\mathbb{R}}(n,n)$ simmetrica, la forma quadratica q_A (o la matrice simmetrica A) si dice definita positiva se $q_A(\mathbf{x}) \ge 0\ \forall\ \mathbf{x} \in \mathbb{R}^n$ e $q_A(\mathbf{x}) = 0 \iff \mathbf{x} = \mathbf{0}$; definita negativa se $q_A(\mathbf{x}) \le 0\ \forall\ \mathbf{x} \in \mathbb{R}^n$ e $q_A(\mathbf{x}) = 0 \iff \mathbf{x} = \mathbf{0}$; semidefinita positiva se $q_A(\mathbf{x}) \ge 0\ \forall\ \mathbf{x} \in \mathbb{R}^n$ e $\exists\ \mathbf{v} \ne \mathbf{0} : q_A(\mathbf{v}) = 0$; semidefinita negativa se $q_A(\mathbf{x}) \le 0\ \forall\ \mathbf{x} \in \mathbb{R}^n$ e $\exists\ \mathbf{v} \ne \mathbf{0} : q_A(\mathbf{v}) = 0$; indefinita $\exists\ \mathbf{v}, \mathbf{w} \in \mathbb{R}^n : q_A(\mathbf{v}) > 0$ e $q_A(\mathbf{w}) < 0$.

Poiché, data $A \in M_{\mathbb{R}}(n,n)$, gli elementi sulla sua diagonale appartengono sempre all'immagine di $q_A(\mathbf{x})$ (per ottenere a_{ii} è sufficiente prendere $\mathbf{x} = \mathbf{e}_i$), se e. g. sulla diagonale ci sono elementi discordi si può già concludere che A è indefinita. Dunque affinché A sia definita positiva è necessario che tutti gli elementi sulla diagonale siano positivi.

Il segno delle forme quadratiche è interessante per vari motivi, ad esempio:

- studio di massimi e minimi di funzioni di più variabili reali: se $n = 1$ e x_0 è un punto stazionario di una funzione derivabile due volte, con $f''(x_0) \ne 0$, $f(x) = f(x_0) + \frac{1}{2}f''(x_0)(x-x_0)^2 + o(x-x_0)^2$. Per conoscere la natura del punto studio il segno di $f''(x_0)$: se $f''(x_0) > 0$ x_0 è un punto di minimo locale, se $f''(x_0) < 0$ x_0 è un punto di massimo locale. Lo stesso si può fare per $n \geq 2$: se $\mathbf{x}_0$ è un punto stazionario di una funzione, $f(\mathbf{x}) = f(\mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T Hf(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + ...$ $Hf(\mathbf{x}_0)$ è detta matrice hessiana ed è simmetrica sotto opportune ipotesi: in tal caso $(\mathbf{x} - \mathbf{x}_0)^T Hf(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$ è una forma quadratica e per capire la natura di $\mathbf{x}_0$ bisogna studiare il segno della forma quadratica $(\mathbf{x} - \mathbf{x}_0)^T Hf(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$.

- La funzione bilineare che definisce il prodotto scalare non standard in $\mathbb{R}^n$ è $\langle \_, \_ \rangle_B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, con B simmetrica: $\langle \mathbf{x}, \mathbf{y} \rangle_B = \mathbf{x}^T B \mathbf{y}$. Questo è un prodotto scalare se vale anche la proprietà di positività: $\langle \mathbf{x}, \mathbf{x} \rangle_B \geq 0 \ \forall \ \mathbf{x} \in \mathbb{R}^n$ e $\langle \mathbf{x}, \mathbf{x} \rangle_B = 0 \Longleftrightarrow \mathbf{x} = \mathbf{0}$. Richiedere che valga la positività è equivalente a chiedere che B sia definita positiva: quindi per definire tale prodotto scalare non standard di $\mathbb{R}^n$ occorre che B sia una matrice simmetrica definita positiva. Quindi i prodotti scalari in $\mathbb{R}^n$ sono in corrispondenza biunivoca con le matrici simmetriche definita positive.

Come cambia la matrice di una forma quadratica al variare delle coordinate? Sia $q_A(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$, con $\mathbf{x}$ vettore delle coordinate rispetto alla base canonica; fissata una nuova base di $\mathbb{R}^n$, sia P la matrice che ha come colonne i vettori della nuova base; allora $q_A(\mathbf{x}) = q_A(P\mathbf{y}) = (P\mathbf{y})^T A (P\mathbf{y}) = \mathbf{y}^T \left(P^T A P\right) \mathbf{y} = q_B(\mathbf{y})$ è la forma quadratica che assegna a $\mathbf{y} = X_P(\mathbf{x})$ lo stesso numero di $q_A(\mathbf{x})$, che opera rispetto alla base canonica. La forma $P^T A P$ ricorda la diagonalizzazione ortogonale ($D = P^T A P$): voglio reinterpretare q_A con le coordinate rispetto a una base ortonormale di $\mathbb{R}^n$ formata da autovettori di A, che so esistere per il teorema spettrale perché A è simmetrica. Si può infatti studiare il segno di una forma quadratica studiando la sua rappresentazione diagonale, dato che q_A e q_B producono lo stesso risultato e che la rappresentazione diagonale è particolarmente semplice. In tal caso infatti, se Q ha come colonne i vettori di una base ortonormale formata da autovettori di A, $q_D(\mathbf{y}) = \mathbf{y}^T D \mathbf{y} = [y_1|...|y_n] \begin{bmatrix} \lambda_1 & ... & 0 \\ ... & ... & ... \\ 0 & ... & \lambda_n \end{bmatrix} \begin{bmatrix} y_1 \\ ... \\ y_n \end{bmatrix} = [\lambda_1 y_1|...|\lambda_n y_n] \begin{bmatrix} y_1 \\ ... \\ y_n \end{bmatrix} = \lambda_1 y_1^2 + ... + \lambda_n y_n^2$, con $\mathbf{y}$ vettore delle coordinate rispetto alla base ortonormale.

Proposizione

$$\text{Hp: } A \in M_{\mathbb{R}}(n,n) \text{ simmetrica, } q_A(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$$

$$\text{Ts: (i) se } \lambda \text{ è un autovalore di } A \text{ e } \mathbf{v} \in V_\lambda,\ q_A(\mathbf{v}) = \lambda ||\mathbf{v}||^2$$

$$\text{(ii) } \forall\ \mathbf{x} \in \mathbb{R}^n\ \lambda_{\min} ||\mathbf{x}||^2 \leq q_A(\mathbf{x}) \leq \lambda_{\max} ||\mathbf{x}||^2$$

L'esistenza di $\lambda_{\min}, \lambda_{\max}$ è garantita dal fatto che in $\mathbb{R}$ è definita una relazione d'ordine totale.

Dim (i) $q_A(\mathbf{v}) = \mathbf{v}^T A\mathbf{v} = \mathbf{v}^T(\lambda\mathbf{v}) = \lambda\left(\mathbf{v}^T\mathbf{v}\right) = \lambda\,||\mathbf{v}||^2$.

(ii) Sia D la matrice diagonale associata ad A, con cambio di base (dalla base di autovettori alla base canonica) descritto dalla matrice ortogonale Q. Se $\mathbf{x}$ è il vettore delle coordinate rispetto alla base canonica e $\mathbf{y}$ è il vettore delle coordinate rispetto alla base ortonormale, $\mathbf{x} = Q\mathbf{y}$. $q_A(\mathbf{x}) = q_A(Q\mathbf{y}) = (Q\mathbf{y})^T A(Q\mathbf{y}) = \mathbf{y}^T\left(Q^T AQ\right)\mathbf{y} = \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + ... + \lambda_n y_n^2$ (i λ_i non sono necessariamente distinti). Tale termine può essere maggiorato prendendo come coefficiente di ogni termine $\lambda_{\max} = \max\{\lambda_1, ..., \lambda_n\}$, e analogamente minorato con $\lambda_{\min} = \min\{\lambda_1, ..., \lambda_n\}$:

$$\lambda_{\min}\left(y_1^2 + ... + y_n^2\right) = \lambda_{\min}y_1^2 + ... + \lambda_{\min}y_n^2 \leq \lambda_1 y_1^2 + ... + \lambda_n y_n^2 \leq \lambda_{\max}y_1^2 + ... + \lambda_{\max}y_n^2 = \lambda_{\max}\left(y_1^2 + ... + y_n^2\right)$$

$y_1^2 + ... + y_n^2$ è la norma al quadrato di $\mathbf{y}$, quindi si ha $\lambda_{\min}\,||\mathbf{y}||^2 \leq q_A(\mathbf{x}) \leq \lambda_{\max}\,||\mathbf{y}||^2$. Inoltre $||\mathbf{x}|| = ||\mathbf{y}||$ perché $\mathbf{x} = Q\mathbf{y}$ e le matrici ortogonali come Q sono isometrie: $||\mathbf{x}||^2 = ||Q\mathbf{y}||^2 = \langle Q\mathbf{y}, Q\mathbf{y}\rangle = (Q\mathbf{y})^T(Q\mathbf{y}) = \mathbf{y}^T\left(Q^T Q\right)\mathbf{y} = \mathbf{y}^T\mathbf{y} = ||\mathbf{y}||^2$. Dunque $\lambda_{\min}\,||\mathbf{x}||^2 \leq q_A(\mathbf{x}) \leq \lambda_{\max}\,||\mathbf{x}||^2$. $\blacksquare$

Per (ii), affinché $q_A(\mathbf{x})$ sia definita positiva è necessario e sufficiente che $\lambda_{\min} > 0$ (se $\mathbf{y} \neq \mathbf{0}$, $0 < \lambda_{\min}\,||\mathbf{y}||^2 \leq q_A(\mathbf{y})$; se $\lambda_{\min} = 0$, $\ker A$ non contiene solo il vettore nullo e $q_A(\mathbf{x}) = 0$ per ogni vettore in $\ker A$); affinché sia definita negativa che $\lambda_{\max} < 0$; affinché sia semidefinita positiva che $\lambda_{\min} = 0$; affinché sia semidefinita negativa che $\lambda_{\max} = 0$; affinché sia indefinita che $\lambda_{\min} < 0 < \lambda_{\max}$.

1. Le matrici che rappresentano la proiezione ortogonale su un sottospazio sono semidefinite positive.

Def Data $A \in M_{\mathbb{R}}(n, n)$ simmetrica, si dice segnatura di A la coppia di numeri interi non negativi (p, q), dove p è il numero di autovalori di A positivi e q il numero di autovalori di A negativi.

Se A è definita positiva, $(p, q) = (n, 0)$; se è definita negativa, $(p, q) = (0, n)$; se è semidefinita positiva, $(p, q) = (p, 0)$ con $0 < p < n$; se è semidefinita negativa, $(p, q) = (0, q)$ con $0 < q < n$; se è indefinita, (p, q) con $0 < p < n$ e $0 < q < n$.

1. Data $A \in M_{\mathbb{R}}(n, n)$ simmetrica, condizioni necessarie ma non sufficienti affinché i suoi autovalori siano tutti positivi (e quindi A sia definita positiva) è $tr(A) = \lambda_1 + ... + \lambda_n > 0$, $\det A = \lambda_1 ... \lambda_n > 0$. Tali condizioni diventano sufficienti se $n = 2$: infatti, se $\lambda_1\lambda_2 > 0$, λ_1, λ_2 potrebbero essere entrambi negativi, ma la positività di entrambi è garantita da $\lambda_1 + \lambda_2 > 0$.

6.0.1 Quoziente di Rayleigh

Posto $\mathbf{x} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$, $\lambda_{\min}\,||\mathbf{x}||^2 \leq q_A(\mathbf{x}) \leq \lambda_{\max}\,||\mathbf{x}||^2$ è equivalente a $\lambda_{\min} \leq \frac{q_A(\mathbf{x})}{||\mathbf{x}||^2} \leq \lambda_{\max}$. La funzione da $\mathbb{R}^n$ in $\mathbb{R}$

$$\frac{q_A(\mathbf{x})}{||\mathbf{x}||^2}$$

si dice quoziente di Rayleigh, e confronta la forma quadratica $q_A(\mathbf{x})$ con la forma quadratica $||\mathbf{x}||^2$. La diseguaglianza mostra che $\lambda_{\min}$ e $\lambda_{\max}$ sono rispettivamente minorante e maggiorante per il quoziente di Rayleigh; sono anche minimo e massimo, perché, se $\mathbf{v} \in V_{\lambda_{\min}}$, $\frac{q_A(\mathbf{v})}{||\mathbf{v}||^2} = \frac{\mathbf{v}^T A\mathbf{v}}{||\mathbf{v}||^2} = \frac{\lambda_{\min}\langle \mathbf{v},\mathbf{v}\rangle}{||\mathbf{v}||^2} = \lambda_{\min}$, e se $\mathbf{v} \in V_{\lambda_{\max}}$, $\frac{q_A(\mathbf{v})}{||\mathbf{v}||^2} = \frac{\mathbf{v}^T A\mathbf{v}}{||\mathbf{v}||^2} = \frac{\lambda_{\max}\langle \mathbf{v},\mathbf{v}\rangle}{||\mathbf{v}||^2} = \lambda_{\max}$. Quindi c'è un nuovo criterio per la ricerca degli autovalori di A; tale ricerca può essere notevolmente semplificata se si considerano solo $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\} : ||\mathbf{x}|| = 1$, perché $\frac{q_A(t\mathbf{x})}{||t\mathbf{x}||^2} = \frac{q_A(\mathbf{x})}{||\mathbf{x}||^2}$, infatti $q_A(t\mathbf{x}) = t^2 q_A(\mathbf{x})$ (quindi normalizzando un vettore il segno della forma quadratica, che è determinato dagli autovalori, non cambia e il quoziente di Rayleigh non cambia, dunque neanche il suo minimo e massimo). Dunque $\lambda_{\min} = \min\left\{\frac{\mathbf{x}^T A\mathbf{x}}{||\mathbf{x}||^2} : \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}\right\} = \min\left\{\mathbf{x}^T A\mathbf{x} : ||\mathbf{x}|| = 1\right\}$ e $\lambda_{\max} = \max\left\{\frac{\mathbf{x}^T A\mathbf{x}}{||\mathbf{x}||^2} : \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}\right\} = \max\left\{\mathbf{x}^T A\mathbf{x} : ||\mathbf{x}|| = 1\right\}$, e la ricerca degli autovalori è di molto semplificata. Gli $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\} : ||\mathbf{x}|| = 1$ sono, se $n = 2$, i punti di una circonferenza di raggio 1, se $n = 3$ sono i punti sulla superficie di una sfera di raggio 1; se $n > 3$ sono i punti sulla superficie di un'ipersfera di raggio 1. $\{\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\} : ||\mathbf{x}|| = 1\}$ è un sottinsieme compatto di $\mathbb{R}^n$, cioè chiuso e limitato (in $\mathbb{R}$, un insieme compatto è unione di intervalli del tipo $[a, b]$). Allora si può applicare il teorema di Weierstrass: data $f : \Omega \to \mathbb{R}$ continua, con $\Omega \subseteq \mathbb{R}^n$ compatto, f ammette massimo e minimo assoluti. Perciò, se si vuole mostrare l'esistenza di un autovalore reale di una matrice simmetrica A (lemma 3), si considera la forma quadratica associata e usando il teorema di Weierstrass si può dire che essa assume massimo e minimo assoluti, che sono $\lambda_{\max}$ e $\lambda_{\min}$, quindi esistono due autovalori reali.