Prerequisiti

- Calcolo Numerico (Eliminazione di Gauss,
  Matlab
- Analisi Matematica (Numeri complessi
  Sviluppo in serie di Taylor
  Derivate
  funzioni ad 1 e più variabili
  Min-Max funzioni ad 1 variabile
  Moltiplicatori di Lagrange → Min/Max funz. + var)
- Programmazione / Algoritmi / Architettura (complessità computazionale, aritmetica binaria, fonti di errore e loro propagazione negli algoritmi, operazioni in floating point

Aspetti fondamentali delle operazioni con i vettori che oggi costituiscono le basi dell'algebra lineare numerica e delle sue applicazioni in diversi problemi dell'informatica e delle scienze applicate in generale.
L'obiettivo delle nostre prime lezioni è quello di introdurre il linguaggio matriciale e le sue notazioni principali e quelle proprietà fondamentali che ci saranno utili per "comprendere" molti concetti che abbiamo "conosciuto" nel corso dei nostri studi di base.

→ DEF di scalare un numero complesso $\mathbb{R} \subset \mathbb{C}$
- $\mathbb{R}^n$ denota l'insieme delle n-uple di numeri reali
  ($\mathbb{C}^n$ " " " " " complessi)
- $\mathbb{R}^2$ → l'insieme di tutte le coppie ordinate di numeri reali (il piano cartesiano standard)

  $\mathbb{R}^3$ → spazio cartesiano 3 dimensionale

Analogamente $\mathbb{R}^{m \times n}$ ($\mathbb{C}^{m \times n}$) denota l'insieme delle matrici $m \times n$ che contengono numeri reali (complessi)

- DEF. Matrici uguali $A = [a_{ij}]$ $B = [b_{ij}]$ $a_{ij} = b_{ij}$ $\forall \begin{array}{l} i = 1, \dots, m \\ j = 1, \dots, n \end{array}$
  le definizioni si applicano ovviamente anche ai vettori riga ($1 \times n$) o colonna ($m \times 1$)

N.B. $u = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ $v = (1, 2, 3)$ sebbene descrivano lo stesso punto nello spazio 3D non sono matrici uguali perchè hanno diverse forme (vettore riga / colonna)

- DEF Somma di matrici $A+B$
  $+$ denota addizione tra scalari e matrici
  (quindi 2 diverse operazioni

- DEF $-A$ (inversa additiva) $\Rightarrow$ permette la sottrazione
- DEF Differenza tra matrici $A-B$
- Propr. della somma
  $A, B, C \in \mathbb{R}^{m\times n}$ $(\mathbb{C}^{m\times n})$
  chiusura $A+B \in \mathbb{K}^{m\times n}$ $(\mathbb{K}=\mathbb{R} \vee \mathbb{K}=\mathbb{C})$
  associatività $(A+B)+C = A+(B+C)$
  commutatività $A+B = B+A$
  identità con 0 $A+0 = A$
  elemento inverso x l'addit. (matrice opposta) $(-A)+A = 0$

- Moltiplicazione x uno scalare
  $\alpha \in \mathbb{K}$ $A \in \mathbb{K}^{m\times n}$ $[\alpha A]_{ij} = [\alpha a_{ij}]$
  propr. moltipl. scalare
  chiusura $\alpha A \in \mathbb{K}^{m\times n}$
  associativ. $\alpha[\beta A] = (\alpha\beta)A$
  distrib. $\alpha(A+B) = \alpha A + \alpha B$
  distrib. $(\alpha+\beta)A = \alpha A + \beta A$
  propr. di identità $1A = A$

- Trasposta $A^T \rightarrow$ scambia righe e colonne
  $(A^T)^T = A$

N.B. Quando $A \in \mathbb{C}^{m\times n}$ ~~l'~~op. complesso coniugato è accompagnato dall'op. di trasposizione
$$[A^H]_{ij} = [\bar{a}_{ji}] \qquad \bar{a} = Re(a) - i\, Im(a)$$

Propr. della trasposta

$$(A+B)^T = A^T + B^T$$
$$(A+B)^H = A^H + B^H$$
$$(\alpha A)^T = \alpha A^T$$
$$(\alpha A)^H = \bar{\alpha} A^H$$

dim. banali

Simmetrie

- Matrice simmetrica

$A \in \mathbb{K}^{n \times n}$ (quadrata)

A è simmetrica $A^T = A$

A hermitiana $A^H = A$

- antisimmetrica

$A^T = -A$ $(A = -A^T)$

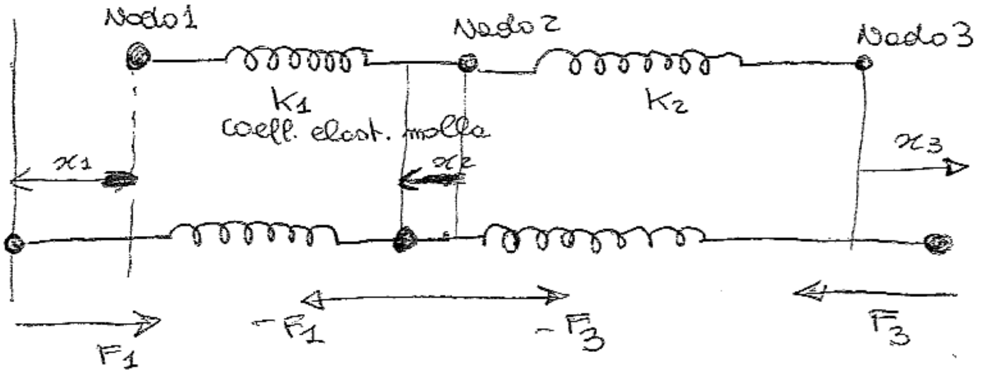
- antihermitiana $A = -A^H$

concetti di simmetria di matrici riflettono molto spesso simmetrie fisiche

→ Esempio

Esempio di simmetria

2 molle connesse come in figura



sulla 1 molla non è esercitata alcuna forza (non c'è tensione)

Supponiamo che le molle sono allungate o compresse in modo che i nodi si dispongano come in figura

L'allungamento/compressione crea una forza su ciascun nodo "Legge di Hooke" $F = Kx$ forza esercitata sulla molla è direttamente proporzionale allo spostamento subito (x) ~~se~~ il coeff. di proporzionalità K è detto costante di stiffness caratteristica del materiale di costituzione della molla

Supponendo k_1 k_2 siano i coeff. di stiffness ed F_i le forze sull'i-simo nodo (quando agisce la compres/allung)

- con la convenzione che lo spostamento a sinistra è positivo ed a destra è negativo

la forza diretta a destra positiva, quella a sinistra negativa

Se il nodo 1 si sposta di x_1 e il nodo 2 di x_2 allora la molla di sinistra avrà subito uno spostamento totale dovuto alla forza F_1 sul nodo 1

$$F_1 = k_1 (x_1 - x_2)$$

Analogamente se il nodo 2 si sposta di una quantità x_2 ed il nodo 3 di una quantità x_3 la molla a destra avrà subito uno spost. totale $x_2 - x_3$ quindi la forza sul nodo 3

$$F_3 = -k_2 (x_2 - x_3)$$

→ forza diretta a sinistra

Forza a sinistra del nodo 2 è opposta a quella alla ~~...~~ forza sul nodo 1, la forza a ~~d~~destra del nodo 2 è opposta a quella del nodo 3

$$F_2 = -F_1 - F_3 \qquad F_2 = -k_1 x_1 + k_1 x_2 + k_2 x_2 - k_2 x_3$$

Organizzando il tutto sotto forma di sist. lineare

$$\begin{aligned} k_1 x_1 - k_1 x_2 &= F_1 \\ -k_1 x_1 + (k_1+k_2) x_2 - k_2 x_3 &= F_2 \\ -k_2 x_2 + k_2 x_3 &= F_3 \end{aligned} \Rightarrow \begin{pmatrix} k_1 & -k_1 & 0 \\ -k_1 & (k_1+k_2) & -k_2 \\ 0 & -k_2 & k_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix}$$

↓

matrice simmetrica

Il punto di questo esempio è che la simmetria del problema fisico si trasferisce nella simmetria della matrice di stiffness K

(se le 2 molle sono uguali $k_1 = k_2 = k$)

(... ... ... in modo "naturale")

Funzioni lineari

$f: D \to \mathbb{R}$ → rango di f

↳ dominio

una funzione lineare è una particolare funzione caratt. da

D, R insiemi che posseggono addizione e moltipl. scalare

f lineare
$$f(x+y) = f(x) + f(y) \qquad \forall x, y \in D$$
$$f(\alpha x) = \alpha f(x)$$

le due condizioni si combinano in una unica propr.

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) \qquad \forall \alpha \in \mathbb{R} \quad \forall {x \atop y} \in D$$

Es. di funz. lineari → rette, piani } grafici di funz. lineari

in generale $f(x_1, \dots x_n) = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_i x_i + \dots + \alpha_n x_n$

rappresenta una sup. lineare o "piatta" (flat) passante x l'origine di un sist. di assi coordinati

<u>Traccia di una matrice</u> $n \times n$ (sommatoria degli elementi sulla diagonale)

$$\text{trace}(A) = \sum_i a_{ii}$$ è una funzione lineare

Consideriamo 2 funz. lineari $f, g: \mathbb{R}^2 \to \mathbb{R}^2$

$$f(x) = f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a x_1 + b x_2 \\ c x_1 + d x_2 \end{pmatrix} \qquad g\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} A x_1 + B x_2 \\ C x_1 + D x_2 \end{pmatrix}$$

componiamo f e g per creare una nuova funz lineare

$$h(x) = f(g(x)) = f\begin{pmatrix} A x_1 + B x_2 \\ C x_1 + D x_2 \end{pmatrix} = \begin{pmatrix} a(A x_1 + B x_2) + b(C x_1 + D x_2) \\ c(A x_1 + D x_2) + d(C x_1 + D x_2) \end{pmatrix}$$

$$= \begin{pmatrix} (aA + bC) x_1 + (aB + bD) x_2 \\ (cA + dC) x_1 + (cD + dD) x_2 \end{pmatrix}$$

Utilizzando la notazione matriciale per rappresentare queste funzioni lineari 6

$f \to F = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ $\quad g \to G = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ $\quad \cancel{h = H = AB}$

$h \to H = \begin{pmatrix} aA + bC & aB + bD \\ cA + dC & cB + dD \end{pmatrix}$

Che cosa è H ?

H composizione (prodotto) di F e G

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} aA+bC & aB+bD \\ cA+dC & cB+dD \end{pmatrix}$$

N.B Il prodotto di 2 matrici rappresenta la composizione delle due funzioni lineari associate

Moltiplicazione tra matrici

$R = (r_1 \; r_2 \; \dots \; r_n) \qquad C = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$

prodotto righe x colonne "prodotto interno standard"
prodotto scalare tra vettori

$$RC = \sum_{i=1}^{n} r_i c_i$$

La moltiplicazione tra matrici può essere riformulata in termini di prodotto scalare tra le righe e le colonne di due matrici di dimensioni conformi per la moltiplicazione

$A \in \mathbb{K}^{m \times p} \qquad B \in \mathbb{K}^{p \times n}$

$$(AB)_{ij} = A(i,:)B(:,j) = A_{i*}B_{*j} = \sum_{k=1}^{p} a_{ik} b_{kj}$$

NON COMMUTATIVA $\quad AB \neq BA$

NON VALE LA LEGGE DELL'ANNULLAMENTO DEL PRODOTTO

Righe e colonne della matrice prodotto

$A = [a_{ij}] \in \mathbb{K}^{m \times p} \qquad B = [b_{ij}] \in \mathbb{K}^{p \times n}$

$[AB]_{i*} \overset{(AB)(i,:)}{=} A_{i*}B \quad$ (l'i-riga di AB = (i riga A) * B)

$[AB]_{*j} = AB_{*j} \quad$ (j col. di AB) = A * (j colonna di B)

$[AB]_{i*} = a_{i1} B_{1*} + a_{i2} B_{2*} + \dots + a_{ip} B_{p*} = \sum_{k=1}^{p} a_{ik} B_{k*}$

$[AB]_{*j} = A_{*1} b_{1j} + A_{*2} b_{2j} + \dots + A_{*p} b_{pj} = \sum_{k=1}^{p} A_{*k} b_{kj}$

righe di AB sono combinazione delle righe di B

colonne di AB " " " colonne di A

Proprietà moltipl. tra matrici

$A(B+C) = AB + AC \qquad$ (distributività a sinistra)

$(D+E)F = DF + EF \qquad$ (" a destra)

$A(BC) = (AB)C \qquad$ (associatività)

Matrice identità $\quad I = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}_{n \times n}$

$AI = A \qquad$ (se $A \in \mathbb{K}^{m \times n} \quad AI_n = A \ \wedge \ I_m A = A$)

$(AB)^T = B^T A^T$

$(AB)^H = B^H A^H$

$\forall A \in \mathbb{K}^{m \times n} \implies A^T A \ \wedge \ AA^T$ sono matrici simmetr

$\boxed{\text{trace}(AB) = \text{trace}(BA)}$

Matrici a blocchi e loro moltiplicazione

Siano A, B due matrici di dimensione conforme partizionate a blocchi ovvero in sottomatrici

$$A = \begin{pmatrix} A_{11} & A_{12} \dots & A_{1r} \\ A_{21} & A_{22} \dots & A_{2r} \\ \vdots & & \\ A_{s1} & A_{s2} \dots & A_{sr} \end{pmatrix} \qquad B = \begin{pmatrix} B_{11} & B_{12} \dots & B_{1t} \\ B_{21} & B_{22} & B_{2t} \\ B_{r1} & B_{r2} & B_{rt} \end{pmatrix}$$

Se ciascuna coppia di blocchi (A_{ik}, B_{kj}) ha dimensione conforme A e B sono dette "partizionate in modo conforme"

$\Rightarrow AB$ può essere partizionata a blocchi e l' (i,j)-blocco di AB è dato da

$$A_{i1}B_{1j} + A_{i2}B_{2j} + \dots A_{ir}B_{rj}$$

Matrice inversa $\quad A \in \mathbb{K}^{n\times n}$ se $\exists B \in \mathbb{K}^{n \times n}$ t.c.

$AB = BA = I \Rightarrow B$ viene detta inversa di A e si denota $B \equiv A^{-1}$

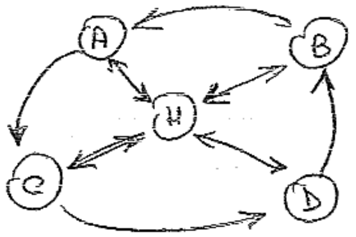
una matrice invertibile è detta nonsingolare

Propr. dell' inversa

- $(A^{-1})^{-1} = A$
- A, B sono $\boxed{\text{non singolari}} \Rightarrow AB$ è non sing. e $(AB)^{-1} = B^{-1}A^{-1}$
- $(A^{-1})^T = (A^T)^{-1}$ e $(A^H)^{-1} = (A^{-1})^H$
- $A_1, A_2 \dots A_k$ non singolari $\quad (A_1 \dots A_k)^{-1} = A_k^{-1} A_{k-1}^{-1} \dots A_2^{-1} A_1^{-1}$

un problema di connessioni aeree

- una linea aerea serve 5 città A, B, C, D, e H in cui H è la città hub. Le diverse connessioni tra le città sono illustrate in fig.



Esercizio assegnato → 6 città

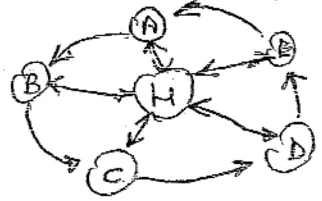


FIG 1

Supponiamo di voler viaggiare da A a B effettuando almeno 2 voli di connessione

Il volo $(A \to H)$ e $(H \to B)$ fornisce il volo con il # minimo di connessioni. Qualora per motivi di disponibilità di posti non sia possibile utilizzare uno di questi voli il transito da A a B deve essere effettuato con + voli.

Determinare un modello decisionale che vi permetta di indicare quanti modi di connessione tra A e B esistono con esattamente 3, 4, 5 ecc. voli

(grafico piccolo → la risposta può essere data manualmente osservando il graf)

Suggerimento: Creare la matrice di connessione

$C = [a_{ij}]$ (matrice di adiacenza)

$$a_{ij} = \begin{cases} 1 & \text{se la città } i \text{ è connessa con la } j \\ 0 & \text{altrimenti} \end{cases}$$

operare sulla matrice C per generare tale modello

Quale operazione matriciale serve?

- Risoluzione di quello assegnato (FIG 1)

La matrice C è

$$C = \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \\ H \end{array} \begin{array}{c} \begin{array}{cccccc} A & B & C & D & E & H \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix} \end{array}$$

C e le sue potenze $C^2, C^3, C^4, \ldots$ forniscono tutte le informazioni necessarie per analizzare il network.

Infatti c_{ik} rappresenta il # di percorsi (voli) diretti dalla città i alla k, c_{kj} " " " " dalla città k alla j, allora $c_{ik}c_{kj}$ rappresenta il # di percorsi con 2 voli dalla città i alla città j attraverso la connessione k

Quindi l'elemento (i,j) nel prodotto $C^2 = CC$

$$[C^2]_{ij} = \sum_{k=1}^{6} c_{ik}c_{kj} = \text{\# totale di connessioni con 2 voli tra la città } i \text{ e la città } j$$

Analogamente

$$[C^3]_{ij} = C^2C = \sum_{k=1}^{6}(C^2)_{ik}c_{kj} = \sum_{k=1}^{6}\Big(\sum_{\ell=1}^{6} c_{i\ell}c_{\ell k}\Big)c_{kj}$$
$$= \sum_{k,\ell} c_{i\ell}c_{\ell k}c_{kj}$$
$$= \text{\# totale di connessioni con 3 voli tra le città } i \text{ e } j$$

In generale

$$[C^n]_{ij} = \sum_{k_1,k_2,\dots k_{n-1}=1}^{6} c_{ik_1}c_{k_1k_2}c_{k_2k_3}\cdots c_{k_{n-2}k_{n-1}}c_{k_{n-1}j}$$

$=$ # totale di voli con n connessioni (cambi) tra le città i e j

Il # totale di voli tra la città i e j che non richiedono più di n ~~voli~~ collegamenti è dato da

$$[C]_{ij} + [C^2]_{ij} + [C^3]_{ij} + \dots + [C^n]_{ij} = [C + C^2 + C^3 + \dots C^n]_{ij}$$

Nel caso particolare del network assegnato

$$C^2 = \begin{bmatrix} 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 1 & 1 & 1 & 2 & 1 \\ 1 & 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 5 \end{bmatrix} \qquad C^3 = \begin{bmatrix} 2 & 2 & 2 & 3 & 2 & 6 \\ 2 & 2 & 2 & 2 & 3 & 6 \\ 3 & 2 & 2 & 2 & 2 & 6 \\ 2 & 3 & 2 & 2 & 2 & 6 \\ 2 & 2 & 3 & 2 & 2 & 6 \\ 6 & 6 & 6 & 6 & 6 & 5 \end{bmatrix}$$

$$C^4 = \begin{bmatrix} 8 & 8 & 8 & 8 & 9 & 11 \\ 9 & 8 & 8 & 8 & 8 & 11 \\ 8 & 9 & 8 & 8 & 8 & 11 \\ 8 & 8 & 9 & 8 & 8 & 11 \\ 8 & 8 & 8 & 9 & 8 & 11 \\ 11 & 11 & 11 & 11 & 11 & 30 \end{bmatrix}$$

Una permutazione $p = (p_1, p_2, \dots p_n)$ dei numeri $(1, 2, \dots n)$ (1-
è semplicemente un loro riordinamento.
Una sequenza di $(1, \dots n)$ ha $n!$ diverse permutazioni
Definiamo segno della permutazione p

$$\sigma(p) = \begin{cases} +1 & \text{se } p \text{ può essere riportata all'ordinamento naturale con un numero pari di scambi} \\ -1 & \text{" " " " " " " dispari " "} \end{cases}$$

il segno dell'ordinamento naturale è $+1$.

<u>DEF</u>. $A \in \mathbb{R}^{n \times n}$ $A = [a_{ij}]$ si definisce $\det(A)$

$$\det(A) = \sum_p \sigma(p)\, a_{1p_1} a_{2p_2} \cdots a_{np_n}$$

la somma su tutte $(n!)$ le possibili permutazioni $p = (p_1 p_2 \dots p_n)$ di $(1, 2, \dots n)$. OSS Il termine $a_{1p_1} a_{2p_2} \dots a_{np_n}$ contiene esattamente un solo elemento per ciascuna riga e colonna di A
N.B. Il det di una matrice rettangolare non è definito

Propr. del determinante

- $\det(A) = \det(A^T)$
- T matrice triangolare $\Rightarrow \det(T) = \prod_{i=1}^{n} t_{ii}$

Sia B la matrice ottenuta
- Scambiando la riga i con j $\Rightarrow \det(B) = -\det(A)$
- moltiplicando una riga i per α $\Rightarrow \det(B) = \alpha \det(A)$
- Sommando α volte la riga i alla riga j $\Rightarrow \det(B) = \det(A)$

Se A è nonsingolare $\Leftrightarrow \det(A) \neq 0$
- $\det(I) = 1$

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

$$\det\begin{pmatrix} A & B \\ 0 & D \end{pmatrix} = \det(A)\det(D)$$
se A, D quadrate

<u>OSS</u> Determinanti vicini a zero non equivale a dire che la matrice è singolare (o almeno non sempre) ovvero non necessariamente il $\det(A)$ è una misura di quanto A è vicina ad una matrice singolare

Infatti matrici vicine a matrici singolari non hanno necessariamente il det piccolo (12)

$$A = \begin{pmatrix} n & 0 \\ 0 & \frac{1}{n} \end{pmatrix} \implies \det(A) = 1$$

matrice vicina ad una matrice singolare

DEF Minore

Sia $A \in \mathbb{R}^{m \times n}$ si dice minore di A il determinante di ogni sottomatrice quadrata $k \times k$ di A

DEF RANK

$rank(A)$ = dimensione del più grande minore di A diverso da $\emptyset$

Sebbene la definizione del $\det(A)$ è puramente algebrica, ne possiamo fornire una interpretazione geometrica.

Il $\det(A)$ è il volume del parallelepipedo n-dimensionale definito dalle colonne di A (se $A \in \mathbb{R}^{n \times n}$)

Se $A \in \mathbb{R}^{m \times n}$ $\quad \det(A^T A)^{1/2}$ = Volume " " " " " "

$$\det(A^H) = \overline{\det(A)}$$

Determinante di una matrice a blocchi

A, D matrici quadrate

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{cases} \det(A) \det(D - CA^{-1}B) & \text{se } \exists A^{-1} \\ \det(D) \det(A - BD^{-1}C) & \text{se } \exists D^{-1} \end{cases}$$

le matrici $D - CA^{-1}B$ e $A - BD^{-1}C$ sono detti complementi di Shur di A e D

Spazi vettoriali (Spazi e sottospazi)

Molte entità matematiche diverse dalle matrici ne condividono alcune proprietà. Per esempio i punti in $\mathbb{R}^2$ o $\mathbb{R}^3$, i polinomi, le funzioni continue, quelle differenziabili ecc. soddisfano le stesse proprietà additive e di moltiplicazione per uno scalare che abbiamo ricapitolato nella lez. precedente.
Pertanto risulterà utile introdurre in modo assiomatico la definizione di spazio vettoriale.

Uno spazio vettoriale richiede 4 cose:
- 2 insiemi: V ed $\mathbb{K}$
- 2 operazioni algebriche chiamate addizione e moltipl. x uno scalare

V = insieme non vuoto di oggetti che chiameremo vettori (per noi $V = \mathbb{R}^n$ oppure $V = \mathbb{R}^{m \times n}$

$\mathbb{K}$ = campo degli scalari $\mathbb{R}$ v $\mathbb{C}$

addizione tra vettori $(x+y)$ op. tra elementi di V

moltiplic. scalare (αx) op. tra $\mathbb{K}$ e V

La def. formale di spazio vettoriale fornisce ~~le~~ interrelazioni tra questi 4 oggetti

<u>DEF</u> V è detto spazio vettoriale su $\mathbb{K}$ quando $+, \cdot$ soddisfano le seguenti proprietà

$\forall x, y \in V$	$x+y \in V$ (chiusura di +)	A1
$\forall x, y, z \in V$	$(x+y)+z = x+(y+z)$	A2
$\forall x, y$	$x+y = y+x$	A3
$\exists\, 0 \in V$	t.c. $0+x = x+0 = x$	A4
$\forall x \in V$	$\exists\, (-x) \in V$ t.c. $x+(-x) = 0$	A5
$\alpha x \in V$	$\forall \alpha \in \mathbb{K}$ $\forall x \in V$ (chiusura della $\cdot$)	M1
$\forall \alpha, \beta \in \mathbb{K}$	$\forall x \in V$ $(\alpha\beta)x = \alpha(\beta x)$	M2
$\forall \alpha \in \mathbb{K}$	$\forall x, y \in V$ $\alpha(x+y) = \alpha x + \alpha y$	M3
$\forall \alpha, \beta \in \mathbb{K}$	$(\alpha+\beta)x = \alpha x + \beta x$	M4
$\forall x \in V$	$1x = x$	M5

Esempi $\mathbb{R}^{m\times n}$ è uno spazio vettoriale su $\mathbb{R}$

$\mathbb{R}^{n\times 1} \equiv \mathbb{R}^n$ " " " " " $\mathbb{R}$

L'insieme dei polinomi con coefficienti reali " " " "

$\mathcal{L} = \{(x,y) \mid y = \alpha x\} \subset \mathbb{R}^2$ spazio vettoriale delle rette passanti per l'origine

DEF di sottospazio

Sia $S \subset V$ $S \neq \{\emptyset\}$

S è detto sottospazio di V se munito $+, \cdot$ soddisfa le proprietà A1-A5 M1-M5

non è necessario verificare tutte le 10 condizioni per definire un sottospazio. E' suff. considerare (e verificare) le proprietà di chiusura A1 e M1

Ovvero

PROP. Un sottoinsieme non vuoto $S \subset V$ è un sottospazio di V $\Leftrightarrow$

A1 $\forall \, x, y \in S \Rightarrow x + y \in S$

M1 $\forall \alpha \in \mathbb{K} \; \forall x \in S \Rightarrow \alpha x \in S$

dim Se $S \subset V$ eredita automaticamente tutte le proprietà di V tranne A1, A4, A5, e M1

Ma osserviamo che

A1 $\wedge$ M1 $\Rightarrow$ A4 e A5

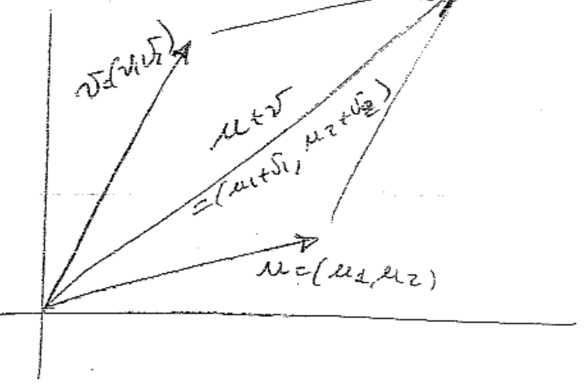
Infatti osserviamo che se M1 è vera $\Rightarrow$ $(-x) = (-1)x \in S$

Quindi A5 è soddisfatta.

Poiche x e $(-x) \in S$ allora se A1 è vera

$x + (-x) \in S$ ovvero $0 \in S$ quindi A4 è soddisfatta

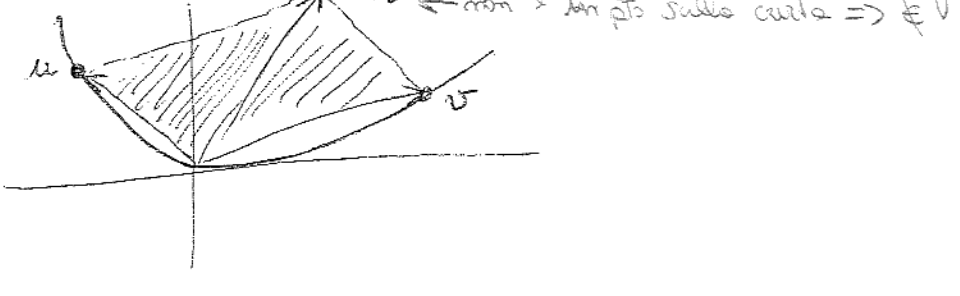
In $\mathbb{R}^2$ e $\mathbb{R}^3$ è facile visualizzare l'addizione vettoriale utilizzando la regola del parallelogramma (dati 2 vett. u, v la loro somma $u+v$ è il vettore def. dalla diag del parall.



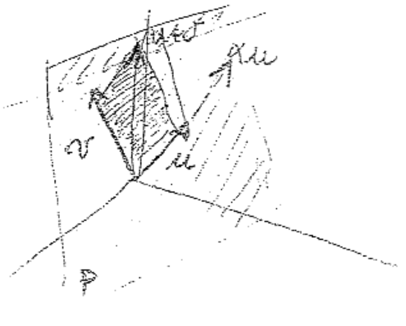
Abbiamo già osservato che le rette passanti per l'origine in $\mathbb{R}^2$ sono sottospazi.

- Che cosa possiamo dire delle rette che non passano per l'origine? Queste non sono sottospazi perché un sottospazio deve contenere il vettore nullo (ovvero l'origine del rif. cartesiano)

- Una linea curva passante per l'origine è un sottospazio di $\mathbb{R}^2$?
Anche in questo caso la risposta è NO.
Infatti possiamo vedere graficamente come la proprietà di chiusura A1 non può essere soddisfatta da curve perché esistono punti sulla curva per cui $u+v$ non è più un elemento della curva



- In $\mathbb{R}^3$ le rette passanti per l'origine ed i piani passanti per l'origine sono sottospazi di $\mathbb{R}^3$
Infatti se P è un piano per l'origine la regola del parallelogramma garantisce la propr. di chiusura rispetto l'addiz. (prop. A1)



La propr. di chiusura M1 è soddisfatta perchè moltiplicando ogni vettore del piano per uno scalare semplicemente lo si accorcia o lo si allunga o se ne cambia il verso, ma l'angolo da esso definito con il riferimento non cambia!

- Linee e superfici in $\mathbb{R}^3$ che hanno una curvatura non sono sottospazi di $\mathbb{R}^3$

Il concetto di sottospazio può essere interpretato visivamente in $\mathbb{R}^2$ e $\mathbb{R}^3$ $\rightarrow$ sottospazi sono superfici piatte (flat) passanti per l'origine

Sebbene non possiamo visualizzare la "flatness" (piattezza) in + dimensioni, la nostra mente la può concepire attraverso la nozione di sottospazio.
Pertanto ogni volta che parleremo di sottospazio immaginate una superficie piatta passante per l'origine

Sottospazi generati Span

L'intersezione di tutti i sottospazi contenenti un dato insieme S di vettori $\in \mathbb{K}^m$ è detto span ed è il più piccolo sottospazio di V contenente S

Consideriamo un insieme di vettori

$$S = \{v_1, v_2, \dots, v_r\} \subset V$$

l'insieme di tutte le possibili combinazioni lineari dei v_i si denota con

$$\mathrm{span}(S) = \{\alpha_1 v_1 + \dots + \alpha_r v_r \mid \alpha_i \in \mathbb{K}\}$$

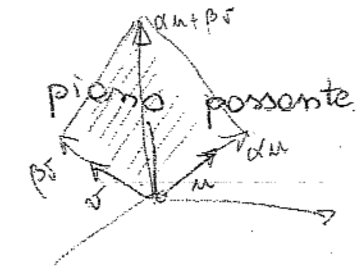
e si chiama sottospazio generato da S.
Infatti: span(S) è un sottospazio perchè soddisfa le due proprietà di chiusura A_1 e M_1

dim Siano $x = \sum \xi_i v_i$ $\quad y = \sum \eta_i v_i$ $\quad x+y = \sum(\xi_i + \eta_i)v_i \in \mathrm{span}(S)$

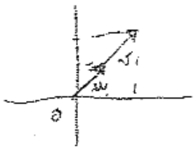
$$\beta x = \sum(\beta \xi_i)v_i \in \mathrm{span}(S)$$

Esempi di span

- $u \in \mathbb{R}^3$ $u \neq 0$ $span\{u\}$ è la retta passante per l'origine ed per ~~vettore u~~ il punto u
- $u, v \in \mathbb{R}^3 \neq 0$ $Span\{u, v\}$ piano passante per l'origine e per i punti u, v



- $S = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\}$ genera la retta $y = x$ di $\mathbb{R}^2$



- $S = \left\{ e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$ è $span(S) \equiv \mathbb{R}^3$
- $e_i \in \mathbb{R}^n$ $e_i = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \to$ i-sima $span\{e_i\} \equiv \mathbb{R}^n$
- L'insieme finito $\{1, x, x^2, \ldots x^n\}$ genera tutti i polinomi di grado $\leq n$

→ aggiungere (17 bis)

DEF Matrice diagonale, triangolare sup ed inf., matrice che commuta con A data, matrice idempotente $(A^2 = A)$, di traccia (A)

I sottoinsiemi precedentemente definiti costituiscono dei sottospazi di $\mathbb{R}^{m \times n}$

[$Ax = b$ si può risolvere se $b \in$ spazio colonne di A (se non appartiene al sottospazio non ci sono $x_1, \ldots, x_n$ che mi danno b per risultato) nel sottospazio (per via della chiusura)]

DEF Rango di $A \in \mathbb{R}^{m \times n}$

Sia $A \in \mathbb{R}^{m \times n}$, l'insieme $Range(A) = R(A) = \{Ax \mid x \in \mathbb{R}^n\}$ è detto range o immagine di A $\subseteq \mathbb{R}^m$

$R(A) = span\{\text{colonne di } A\}$ (spazio colonna di A) o spazio generato dalle colonne di A

$R(A^T) = span\{\text{righe di } A\}$ (spazio riga di A) spazio generato dalle righe di A

17 bis

Probleme Sia $S = \{a_1, a_2, \dots a_n\}$ un insieme di m vettori che in sottospazio $V \subset \mathbb{R}$

$a_i \in \mathbb{R}^m$ vettori m-dimensionale

Sia $A \in \mathbb{R}^{m \times n}$ la matrice contenente i vettori ai quali vettori colonna

~~$S = span\{S\}$~~

l'insieme di vettori a_i genera il sottospazio $V \iff$
$\forall b \in V \quad \exists x \in \mathbb{R}^n$ t.c. $Ax = b$ (il sistema $Ax = b$ è consistente $\forall b \in V$)

S genera $V \iff \forall b \in V \quad \exists \alpha_i \in \mathbb{R}$ t.c.

$$\alpha_1 a_1 + \alpha_2 a_2 + \dots + \alpha_m a_m = b \iff$$

$$(a_1 | a_2 | \dots | a_m) \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} = b \iff A \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} = b$$

$\iff$ posto $x = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix}$ $\quad Ax = b$

Questa semplice osservazione è utile a verificare se un insieme di vettori genera un determinato sottospazio. ES $S = \{(1,1,1), (1,-1,-1), (3,1,1)\}$ genera $\mathbb{R}^3$ si pongono tali vettori riga come colonne di 1 matrice A

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

e ci si domanda se il sistema $Ax = b$ è consistente.

OSS $\boxed{Ax = b \text{ è consistente} \iff \text{rank}[A|b] = \text{rank}(A)}$

$\text{rank}(A) = 2$ (infatti A ha 2 righe uguali quindi $\det(A) = 0$)

$\text{rank}([A|b]) = \text{rank}\begin{bmatrix} 1 & 1 & 3 & b_1 \\ 1 & -1 & 1 & b_2 \\ 1 & 1 & 1 & b_3 \end{bmatrix} = 3$ per qualche b
(per esempio $b = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$)

Esempio Consideriamo la matrice $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{pmatrix}$ con colonne A_{*1}, A_{*2}, A_{*3}

$$R(A) = \text{span}\{A_{*1}, A_{*2}, A_{*3}\} =$$

$$= \text{span}\left\{\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}2\\4\end{pmatrix}, \begin{pmatrix}3\\6\end{pmatrix}\right\} = \left\{\alpha\begin{pmatrix}1\\2\end{pmatrix} + \beta\begin{pmatrix}2\\4\end{pmatrix} + \gamma\begin{pmatrix}3\\6\end{pmatrix} \,\middle|\, \alpha, \beta, \gamma \in \mathbb{R}\right\}$$

$$= \left\{\alpha\begin{pmatrix}1\\2\end{pmatrix} + 2\beta\begin{pmatrix}1\\2\end{pmatrix} + 3\gamma\begin{pmatrix}1\\2\end{pmatrix} \;\; \alpha, \beta, \gamma \in \mathbb{R}\right\} = \left\{\underbrace{(\alpha + 2\beta + 3\gamma)}_{\eta \in \mathbb{R} \text{ (scalare)}}\begin{pmatrix}1\\2\end{pmatrix}\right\}$$

= retta di $\mathbb{R}^2$ passante per l'origine ed il punto $\begin{pmatrix}1\\2\end{pmatrix}$

$$R(A^T) = \text{span}\{A_{1*}; A_{2*}\} = \text{span}\left\{\begin{pmatrix}1\\2\\3\end{pmatrix}, \begin{pmatrix}2\\4\\6\end{pmatrix}\right\}$$

$$= \left\{\alpha\begin{pmatrix}1\\2\\3\end{pmatrix} + \beta\begin{pmatrix}2\\4\\6\end{pmatrix} \,\middle|\, \alpha, \beta \in \mathbb{R}\right\} = \left\{\underbrace{(\alpha + 2\beta)}_{\eta \in \mathbb{R}}\begin{pmatrix}1\\2\\3\end{pmatrix}\right\}$$

= retta di $\mathbb{R}^3$ passante per l'origine ed il pto $\begin{pmatrix}1\\2\\3\end{pmatrix}$

η scalare

Spazio nullo di A

DEF. Sia $A \in \mathbb{R}^{m \times n}$ l'insieme

$$Null(A) = N(A) = \{x \in \mathbb{R}^{n \times 1} \mid Ax = 0\} \subseteq \mathbb{R}^n$$

si dice spazio nullo di A ed è un sottospazio di $\mathbb{R}^n$.

$N(A)$ è l'insieme delle soluzioni del sist. omogeneo $Ax = 0$

(lo spazio nullo corrisponde al Kernel della applicazione lineare associata ad A)

$N(A)$ è il sottospazio generato dalle soluzioni di $Ax = 0$

DEF Vettori linearmente indipendenti

Consideriamo un insieme di vettori $S = \{v_1, v_2, \dots v_n\} \in \mathbb{R}^n$

questi si dicono l.i. se e solo se $\sum_{i=1}^{n} \alpha_i v_i = 0 \iff \alpha_i = 0 \ \forall i$

PROP. Sia $A \in \mathbb{R}^{m\times n}$, le seguenti propr. sono equivalenti

- le colonne di A (A_{*i}) costituiscono un insieme di vett. l.i.
- $N(A) = \{0\}$
- $\boxed{rank(A) \overset{def}{=} \#\text{ delle colonne } \boxed{l.i} \text{ di } A = n}$

- le righe di A (A_{i*}) costituiscono un insieme di vett. l.i.
- $N(A^T) = \{0\}$
- $rank(A^T) = \#$ righe l.i. di $A = m$

Se $A \in \mathbb{R}^{n\times n}$

allora A è nonsingolare $\iff$ A_{i*} righe formano un insieme di velt. l.i

$\iff$ A_{*i} le colonne formano un ins. di velt. l.i.

N.B. Sia $A \in \mathbb{R}^{m\times n}$ e consideriamo le matrici prodotto (simmetriche)

A^TA e AA^T

Queste matrici assumono una particolare importanza perché compaiono in diverse applicazioni.

Per esse sono valide le seguenti proprietà

$rank(A^TA) = rank(A) = rank(AA^T)$

$R(A^TA) = R(A^T)$ spazio righe di A e $R(AA^T) = R(A)$ spazio colonne di A

$N(A^TA) = N(A)$ e $N(AA^T) = N(A^T)$

(mat. sinistra)

(mat. destra)

Un esempio ~~di~~ in ambito delle matrici prodotto A^TA e AA^T è la loro appl. alla risoluzione di un sistema lineare

$Ax = b \qquad A \in \mathbb{R}^{m\times n}$

$\rightarrow$ questo sist. può o no essere consistente.

Moltiplicando a sinistra per A^T si produce un sist. $n\times n$

$$A^T A x = A^T b$$

chiamato sist. associato delle eq. normali che possiede delle proprietà estremamente importanti.

- il sist. delle eq. normali è sempre consistente perché $A^T b \in R(A^T) = R(A^T A)$
- se $Ax = b$ è consistente allora $Ax = b$ e $A^T A x = A^T b$ hanno le stesse insieme di soluzioni
- se $Ax = b$ è consistente ed ha una unica sol. lo stesso accade per $A^T A x = A^T b$ e risulta $x = (A^T A)^{-1} A^T b$

(*) aggiungere (20 bis)

DEF BASE

Considerato uno spazio vettoriale V un insieme di vettori l.i. che generano tutto lo spazio è detto base di V

$\mathcal{B} = \{b_1, \dots b_n\}$ si l.i e $span\{b_1, \dots, b_n\} = V \overset{def}{\Longleftrightarrow}$

$\Longleftrightarrow \{b_i\}$ base di V

caratterizzazione di una base

Sia $V \subseteq \mathbb{R}^m$ sottospazio e $\mathcal{B} = \{b_1, b_2, \dots b_n\} \subseteq V$ le seguenti proprietà sono equivalenti

- $\mathcal{B}$ è una base di V $\quad (V = span(\mathcal{B}))$
- $\mathcal{B}$ è il $\overbrace{\text{più piccolo}}^{\text{minimale}}$ insieme che genera V
- $\mathcal{B}$ è il $\underbrace{\text{più grande}}_{\text{massimale}}$ insieme di vettori l.i di V (Tutti i vettori l.i. di V?)

N.B. La dimensione di uno spazio vettoriale V è def.

dim V = # di vettori di una base di V

= # " " " ogni sottospazio (minimale) generatore di V

= # " " " " sottoinsieme massimale di vett. l.i.

DEF vettori l.i. $S = \{v_1, v_2 \dots v_n\}$ è l.i. $\Leftrightarrow$ (20bis)

$\sum \alpha_i v_i = 0 \Leftrightarrow \alpha_i = 0 \ \forall i$

(l.i vettori che non contengono relazioni di dipendenza
l.d. se un vettore si può riscrivere come comb lin. degli altri vettori)

Matrici e lineare indipendenza

- $A \in \mathbb{R}^{m \times n}$ $\qquad A = [a_1 | a_2 | \dots a_n]$

le seguenti prop. sono equivalenti

- le colonne di A sono l.i.
- $N(A) = \{0\}$
- $\text{rank}(A) = n$

- le righe di A sono l.i
- $N(A^T) = \{0\}$
- $\text{rank}(A) = m$

se $A \in \mathbb{R}^{n \times n}$ le seguenti prop sono equiv.
- A è non singolare
- le colonne di A formano un insieme di vett. l.i
- le righe di A " " " " " "

Esempi

Matrice di Vandermonde

DEF
$$V_{m \times n} = \begin{pmatrix} 1 & x_1 & x_1^2 & x_1^3 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & x_2^3 & & x_2^{n-1} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & x_m & x_m^2 & x_m^3 & & x_m^{n-1} \end{pmatrix} \in \mathbb{R}^{m \times n}$$ è detta matrice di Vandermonde se

$\rightarrow x_i \neq x_j \quad \forall i \neq j$

Se $n \leq m$ le colonne di V costituiscono un insieme di vec. l.i

dim le colonne di V sono l.i $\Leftrightarrow N(V) = \{0\}$

Consideriamo $N(V) = \{\alpha \in \mathbb{R}^n \mid V\alpha = 0\}$ $\qquad \alpha := \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{n-1} \end{pmatrix}$

$$\begin{pmatrix} 1 & x_1 & x_1^2 \dots & x_1^n \\ 1 & x_2 & x_2^2 \dots & x_2^n \\ \vdots & & & \\ 1 & x_m & x_m^2 \dots & x_m^n \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_{n-1} \end{pmatrix} = \begin{pmatrix} 0 \\ | \\ 0 \end{pmatrix}$$

$\forall i = 1, \dots m \qquad \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \dots \alpha_{n-1} x_i^{n-1} = 0$

$\Rightarrow p(x) = \alpha_0 + \alpha_1 x + \dots + \alpha_{n-1} x^{n-1}$

il polinomio $p(x)$ di grado $\leq n-1$ e $p(x)$ ha m radici distinte x_i $\quad n-1 < n \leq m \Rightarrow$ dal teor. fondamentale dell'algebra $p(x)$ è necessariamente il polinomio nullo $\Rightarrow \alpha_i = 0$

Problema (20 tris)

Assegnato un insieme di m punti distinti di $\mathbb{R}^2$

$S = \{(x_1, y_1) \dots (x_m, y_m)\}$ $\quad x_i \neq x_j \quad \forall i \neq j$

$\exists!$ polinomio

$$\ell(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_{m-1} t^{m-1}$$ di grado $m-1$

passante per i punti (x_i, y_i) (interpolante tale insieme di punti)

Per det. il polinomio $\ell(t)$ dobbiamo calcolarne i suoi coefficienti α_i. Ossero dobbiamo risolvere il sistema

$$\ell(x_i) = y_i \qquad \forall i = 1 \dots m$$

$$\Leftrightarrow \overset{\ell(x_1)}{\alpha_0} + \alpha_1 x_1 + \alpha_2 x_1^2 + \dots \alpha_m x_1^{m-1} = y_1$$

$$\ell(x_2) \quad \alpha_0 + \alpha_1 x_2 + \alpha_2 x_2^2 + \dots + \alpha_{m-1} x_2^{m-1} = y_2$$

$$\vdots$$

$$\ell(x_m) = \alpha_0 + \alpha_1 x_m + \alpha_2 x_m^2 + \dots + \alpha_{m-1} x_m^{m-1} = y_m$$

$$\Leftrightarrow \underbrace{\begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{m-1} \\ 1 & x_2 & x_2^2 & & x_2^{m-1} \\ \vdots & & \vdots & & \\ 1 & x_m & x_m^2 & & x_m^{m-1} \end{pmatrix}}_{m \times m} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{m-1} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

matrice di Vandermonde che abbiamo visto essere non singolare

$\Leftrightarrow$ $\exists!$ sol. del precedente sist. lineare

Si prova

$$\ell(t) = \sum_{i=1}^{m} y_i \left(\frac{\prod_{j \neq i}^{m} (t - x_j)}{\prod_{j \neq i} (x_i - x_j)} \right)$$

polinomio interpolante di Lagrange di grado $m-1$

RELAZIONE TRA SOTTOSPAZI - DIMENSIONI - BASI

Sia $A \in \mathbb{R}^{m\times n}$ e' $\text{rank}(A) = r$.

Allora si prova che

- $\dim(R(A)) = r$
- $\dim(N(A)) = n - r$
- $\dim(R(A^T))\ r$
- $\dim(N(A^T)) = m - r$

TEOR $A \in \mathbb{R}^{m\times n}$

$$\dim(R(A)) + \dim(N(A)) = n$$

OSS

(da riprendere legame tra grafo e rank della matrice di incidenza associata ad un grafo connesso (~~attraverso~~ metodi di classific ??)

(vedi esempio di grafo (21 bis) e (21 Tris)

Ulteriori propr. di A^TA e AA^T

$A \in \mathbb{R}^{m\times n}$ le seguenti proprietà sono vere

- $\text{rank}(A^TA) = \text{rank}(A) = \text{rank}(AA^T)$
- $R(A^TA) = R(A^T) \quad \wedge \quad R(AA^T) = R(A)$
- $N(A^TA) = N(A) \quad \wedge \quad N(AA^T) = N(A^T)$

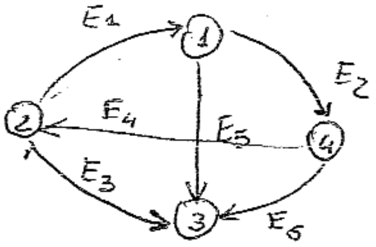
- Relazione tra rank e connettività

Si dice grafo
un insieme di punti (o nodi) $\{N_1, N_2, \dots N_m\}$ e un insieme di path (edge / percorsi) $\{E_1, E_2, \dots E_n\}$ tra i nodi

DEF Un <u>grafo connesso</u> è un grafo in cui esiste una sequenza di percorsi che connette ogni coppia di nodi

Un <u>grafo diretto</u> è un grafo in cui per ciascun percorso è assegnata una direzione di percorrenza.

Esempio di grafo connesso e diretto



La connettività di un grafo è indipendente dalla direzione assegnata ad ogni percorso.
Il concetto di connettività di un grafo e di rank di una matrice sono intimamente legati:

DEF Matrice di incidenza associata ad un grafo diretto contenente m nodi ed n percorsi è la matrice $E \in \mathbb{R}^{m \times n}$ i cui elementi sono

$$e_{kj} = \begin{cases} 1 & \text{se il percorso } E_j \text{ è diretto (entra) verso il nodo } N_k \\ -1 & \text{" " " esce dal nodo } N_k \\ 0 & \text{se " " non inizia né finisce nel nodo } N_k \end{cases}$$

La matrice di incidenza associata al grafo in figura è

$$E = \begin{array}{c} \\ N_1 \\ N_2 \\ N_3 \\ N_4 \end{array} \begin{array}{c} \begin{array}{cccccc} E_1 & E_2 & E_3 & E_4 & E_5 & E_6 \end{array} \\ \begin{pmatrix} 1 & -1 & 0 & 0 & -1 & 0 \\ -1 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & -1 & 0 & -1 \end{pmatrix} \end{array}$$

Ciascuna ~~…~~ linea di connessione in un grafo ~~…~~ diretto connette 2 nodi: la testa e la coda della connessione pertanto la matrice E sarà t.c. ciascuna sua colonna avrà 2 elementi diversi da 0 $(-1 \text{ e } +1)$

$\Rightarrow$ somma delle colonne $= 0$

$$e^T = (1, 1, \dots 1) \in \mathbb{R}^m \Rightarrow \overset{\text{somma colonne}}{e^T E} = 0 \Rightarrow e \in N(E^T)$$

e $\quad \text{rank}(E) = \text{rank}(E^T) = m - \dim(E^T) \leq m-1$

Il segno di disuguaglianza vale $\forall$ tipo di connettività associata al grafo

Il segno di uguaglianza si ha se il grafo è <u>connesso</u>

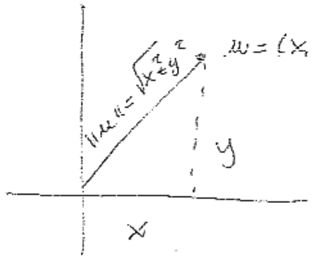
Si prova che

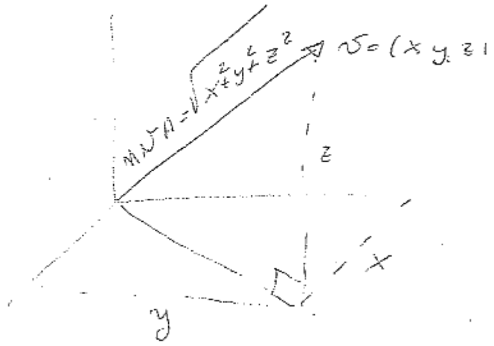
Theor. Sia G un grafo contenente m nodi.
Assegnate in modo arbitrario le direzioni alle sue linee di connessione in modo da rendere il grafo G connesso e detta E la matrice di incidenza associata a G si ha che

$$G \text{ è connesso} \Leftrightarrow \text{rank}(E) = m-1$$

NORME - PRODOTTI SCALARI - ORTOGONALITA'

L'algebra lineare (sua formalizzazione) cerca di generalizzare i concetti della geometria di base degli spazi $\mathbb{R}^2$ ed $\mathbb{R}^3$ agli spazi non visuali di grandi dimensione.
Per esempio la lungh. di un vettore $u \in \mathbb{R}^2$ o $v \in \mathbb{R}^3$ si ottiene facilmente applicando il teor. di Pitagora e calcolando la lungh. dell'ipotenusa di un triangolo come mostrato nella seguente figura





le misure
$\|u\| = \sqrt{x^2+y^2}$ e $\|v\| = \sqrt{x^2+y^2+z^2}$
sono chiamate norme euclidee dei vettori $u \in \mathbb{R}^2$ e $v \in \mathbb{R}^3$ e ammettono una ovvia estensione in $\mathbb{R}^n$

NORMA EUCLIDEA
Considerato lo spazio vettoriale $\mathbb{R}^n$ $\forall x \in \mathbb{R}^n$

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x^T x}$$

Nel caso complesso

$$\|x\|_2 = \sqrt{x^H x}$$

(Convenzione per noi i vettori saranno vettori colonna qualora non differentemente specificato)

- OSSERVAZIONI IMPORTANTI

La versione complessa di $\| \|_2$ include quella reale come caso special
N.B. $\|x\|_2$ è sempre reale anche se $x \in \mathbb{C}^n$

- La def. della $\| \|_2$ garantisce che
$\|x\| \geq 0$, $\|x\| = 0 \iff x = 0$ $\|\alpha x\| = |\alpha|\,\|x\|$ $\forall \alpha \in \mathbb{C}$

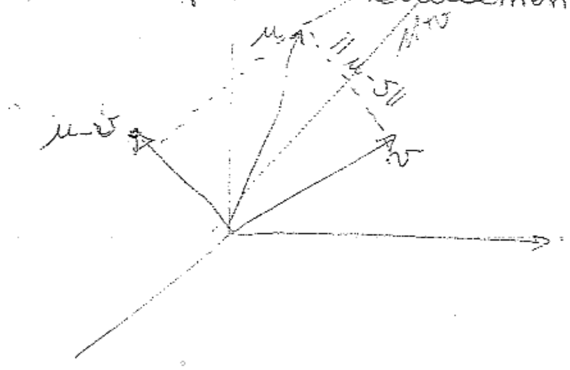
- Dato un vettore $x \neq 0$ può essere conveniente ottenere da x un altro vettore u puntato nella stessa direzione di x (cioè è un multiplo positivo di x) ma che abbia lunghezza unitaria. Per costruire questo vettore si normalizza il vettore x come segue

$$u = \frac{x}{\|x\|} \Rightarrow \|u\| = \left\|\frac{x}{\|x\|}\right\| = \frac{1}{\|x\|}\|x\| = 1$$

si normalizza tramite la norma

- La distanza tra vettori in $\mathbb{R}^3$ può essere visualizzata utilizzando la regola del parallelogramma, quindi in $\mathbb{R}^3$ ($\mathbb{C}^3$) la distanza tra $u, v \in \mathbb{R}^3$ si definisce naturalmente come $\|u - v\|$

23



Prodotto scalare $\forall \, x, y \in \mathbb{R}^n$ ($\mathbb{C}^n$)

lo scalare $x^T y = \sum_{i=1}^{n} x_i y_i \in \mathbb{R}$ ($x^H y = \sum \bar{x}_i y_i$)

è detto prodotto scalare interno su $\mathbb{R}^n$ ($\mathbb{C}^n$)

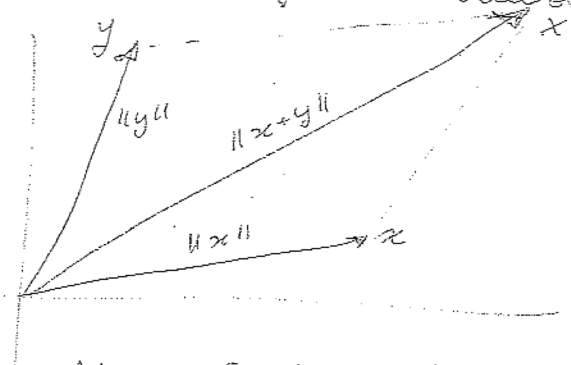
DISUGUAGLIANZA DI CAUCHY-SCHWARZ

$|x^H y| \leq \|x\| \|y\| \quad \forall \, x, y \in \mathbb{C}^n$

l'uguaglianza vale se e solo se $y = \alpha x$ $\quad \alpha = \frac{x^H y}{x^H x}$.

L'importanza della disuguaglianza di CS è quella di aiutare a definire una geometria negli spazi multidimensionali che sia consistente con gli spazi visuali $\mathbb{R}^2$ e $\mathbb{R}^3$

Consideriamo la seguente situazione



Supponiamo di muoverci dall'origine al punto x e poi di spostarci al punto $x+y$. Naturalmente abbiamo percorso una distanza che è almeno tanto grande quanto la distanza dall'origine a $x+y$ lungo la diagonale del parallelogramma.

È graficamente evidente $\|x+y\| \leq \|x\| + \|y\|$

Questa osservazione è nota come disuguaglianza triangolare

Negli spazi multidim. non possiamo visualizzare la geometria ma la disug. di CS ci permette di dimostrarla:

DISUG. TRIANG. $\quad \forall \, x, y \in \mathbb{R}^n \quad \|x+y\| \leq \|x\| + \|y\|$

Oltre la lungh. euclidea ci sono altre nozioni di misura su vettori (es. navigatore che segue una griglia e non la distanza in volo → norma della griglia è ~~la~~ ~~[illegible]~~ conosciuta come norma 1)

Norme p

Sia $p \geq 1$ su $\mathbb{C}^n$ p-norm è definita come

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

e come la norma euclidea verifica le seguenti propr.

$\|x\|_p \geq 0$

$\|x\|_p = 0 \iff x = 0$

$\|\alpha x\|_p = |\alpha| \|x\|_p \quad \forall \alpha \in \mathbb{R}$

$\|x+y\|_p \leq \|x\| + \|y\|_p$

disuguaglianza di Holder $\quad |x^H y| \leq \|x\|_p \|y\|_p$

generalizzazione della CS

In pratica solo 3 p-norme sono le più usate

$\|x\|_1 = \sum_{i=1}^{n} |x_i|$ (grid norm)

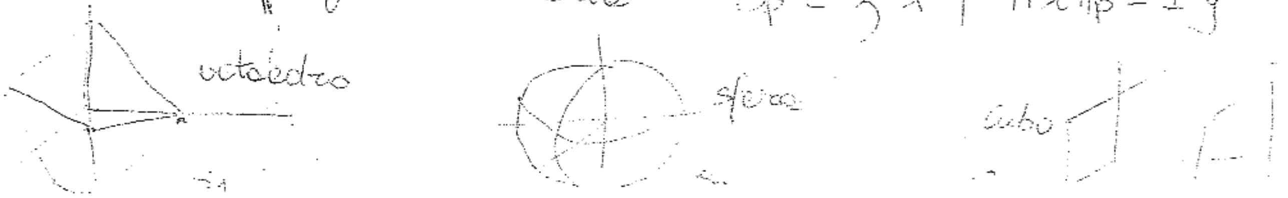
$\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}$ (norma euclidea)

$\|x\|_\infty = \max_i |x_i|$ (max norm o norma infinito $\lim_{p \to \infty} \|x\|_p$

Per avere una idea del significato geometrico delle 3 norme (tutte le p-norme sono definite in termini di coordinate quindi il loro uso è limitato agli spazi coordinato con un riferimento

Comunque il concetto di norma è generalizzato a uno spazio vettoriale V cioè è un concetto coordinate-free)

~~[illegible]~~ possiamo analizzare la forma e le dimensioni relative della ~~p~~ sfera unitaria $S_p = \{x \mid \|x\|_p = 1\}$



$S_1 \subset S_2 \subset S_\infty \Rightarrow \|x_1\| \geq \|x\|_2 \geq \|x\|_\infty \quad \forall x \in \mathbb{R}^3$

NORMA SU MATRICI

Poiche $\mathbb{C}^{m\times n}$ è uno spazio vettoriale di dim. mn la grandezza di una matrice può essere misurata utilizzando ogni norma vettoriale su $\mathbb{C}^{mn}$ (per esempio trasformando la matrice in un vettore colonna) ed applicando la norma euclidea

→ otteniamo la norma di Frobenius

$$A = \begin{pmatrix} 2 & 1 \\ -4 & -2 \end{pmatrix} \rightarrow \quad x_A = \begin{pmatrix} 2 \\ -4 \\ 1 \\ -2 \end{pmatrix}$$

$$\|x_A\|_2 = \sqrt{2^2+(-4)^2+1^2+(-2)^2} = 5$$

Questa è la nozione più semplice di norma matriciale ed è chiamata norma di Frobenius.

NORMA DI FROBENIUS

$$\|A\|_F^2 = \sum_{ij} |a_{ij}|^2 = \sum_i \|A_{i*}\|_2^2 = \sum_j \|A_{*j}\|_2^2 = trace(A^H A)$$

DEF generale di norma matriciale su $\mathbb{C}^{m\times n}$

Norma 1 $\quad \|A\|_1 = \max_{\|x\|_1=1} \|Ax\|_1 = \max_j \sum_i |a_{ij}|$

= il + grande valore delle somme dei valori assoluti sulle colonne

$\|A\|_\infty = \max_i \sum_j |a_{ij}|$ → max somma dei val ass sulle righe

Riprendere la def di prodotto scalare

V spazio vettoriale reale

$\langle \cdot , \cdot \rangle : V \times V \to \mathbb{R} \ (\mathbb{C})$

$\forall x, y \quad \langle x, x \rangle \geq 0 \qquad \langle x, x \rangle = 0 \iff x = 0$

$\langle x, \alpha y \rangle = \alpha \langle x, y \rangle \quad \forall \alpha$

$\langle x+y, z \rangle = \langle x, y \rangle + \langle x, z \rangle$

$\langle x, y \rangle = \langle y, x \rangle \quad \left(\text{in } \mathbb{C} \ \langle x, y \rangle = \overline{\langle y, x \rangle} \right)$

~~DEF...~~

Abbiamo ricordato il prodotto scalare su $\mathbb{R}^n$

$\langle x, y \rangle = x^T y = \sum_{i=1}^{n} x_i y_i$

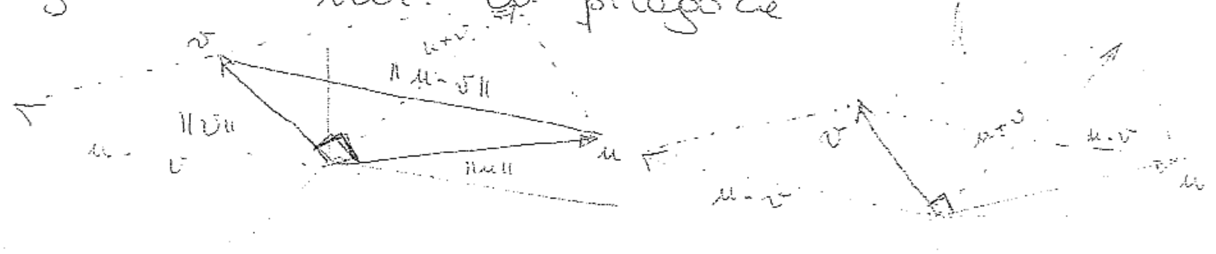
possiamo osservare come questo prodotto scalare definisca su $\mathbb{R}^n$ ~~la~~ una norma, detta norma indotta

$\|x\| = \sqrt{\langle x, x \rangle} = \langle x, x \rangle^{1/2} = (x^T x)^{\frac{1}{2}} \equiv \|x\|_2$

(la norma euclidea è indotta dal prodotto scalare su $\mathbb{R}^n$)

Su $\mathbb{R}^3$ due vettori sono ortogonali se l'angolo da essi formato è di 90°. La visualizzazione del concetto di angolo non è possibile in più dimensioni

comunque l'essenza del concetto di perpendicolarità è legata al teor. di Pitagora

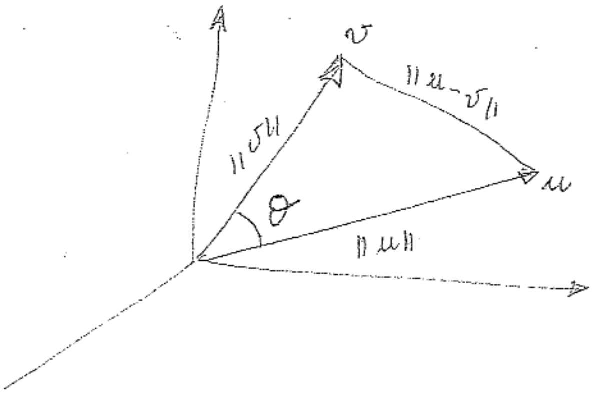


$\boxed{u \text{ e } v \text{ sono } \perp \text{ se } \|u\|^2 + \|v\|^2 = \|u \mp v\|^2}$

$$0 = u^Tu + v^Tv - (u-v)^T(u-v) =$$
$$= u^Tu + v^Tv - (u^Tu - u^Tv - v^Tu + v^Tv)$$
$$= \cancel{u^Tu} + \cancel{v^Tv} - \cancel{u^Tu} + u^Tv + v^Tu - \cancel{v^Tv}$$
$$= 2u^Tv \qquad \left(\begin{smallmatrix} u \\ v \end{smallmatrix} \in \mathbb{R}^3\right)$$

$\Rightarrow$ $\boxed{\text{due vettori sono ortogonali se e solo se il loro prodotto scalare è nullo } u^Tv = 0}$

Una volta che ha senso il concetto di angolo retto in spazi di $\mathbb{R}^n$ generalizziamo il concetto di angolo



la legge del coseno in $\mathbb{R}^2$ o $\mathbb{R}^3$ dice

$$\|u - v\|^2 = \|u\|^2 + \|v\|^2 - 2\|u\|\|v\| \cos\theta$$

$$\Rightarrow \cos\theta = \frac{-\|u-v\|^2 + \|u\|^2 + \|v\|^2}{2\|u\|\|v\|}$$

$$\cos\theta = \frac{\|u\|^2 + \|v\|^2 - \|u-v\|^2}{2\|u\|\|v\|}$$

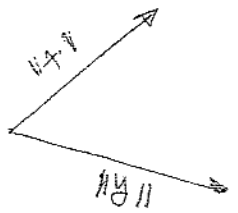
$$= \frac{u^Tu + v^Tv - (u-v)^T(u-v)^T}{2\|u\|\|v\|}$$

$$= \frac{2u^Tv}{2\|u\|\|v\|} = \frac{u^Tv}{\|u\|\|v\|}$$

EF Negli spazi vettoriali su $\mathbb{R}$ la misura in radianti dell'angolo compreso tra 2 vettori non nulli x, y definito come il numero $\theta \in [0, \pi]$ s' $\cos\theta = x^T y$ ...

in spazi reali $x, y \in \mathbb{R}^n$

$$\cos\theta = \frac{\langle x, y\rangle}{\|x\|\,\|y\|} = \frac{x^T y}{\|x\|\,\|y\|}$$



(similarità del coseno)

Linear correlation

Supponiamo di aver condotto un esperimento e che le osservazioni siano state salvate in due vettori

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \qquad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

sia $e = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

Problema: determiniamo con quale grado le componenti y_i sono linearmente correlate con x_i
cioè misuriamo quanto y è vicino ad una combinazione lineare $\beta_0 e + \beta_1 x$ elemento base canonica

Soluzione: Il coseno formato dai vettori x e y ci fornisce il grado di correlazione lineare tra questi vettori
Per capirne le motivazioni indichiamo con

$$\mu_x = \frac{\sum_i x_i}{n} = \frac{e^T x}{n}$$ media misura la tendenza principale dei dati

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu_x)^2}{n}} = \frac{\|x - \mu_x e\|_2}{\sqrt{n}}$$ deviazione standard misura il grado in cui i dati sono (distanti) sparpagliati

Frequentemente
Poiché i dati possono derivare da sorgenti diverse e quindi ~~[illegible]~~ ~~[illegible]~~ presentare diverse unità di misura. Generalmente si procede alla standardizzazione in quantità unitarie $\left(\frac{x - \mu}{\sigma}\right.$

DEF La standardizzazione di un vettore x t.c. $\sigma_x \neq 0$ è definita

$$z_x = \frac{x - \mu_x e}{\sigma_x}$$ vettore con componenti normalizzate

le componenti del vettore z_x si chiamano z-scores o punteggi standard
Tutti i vettori standardizzati sono t.c. $\|z\| = \sqrt{n}$ $\mu_z = 0$ $\sigma_z =$
Si può facilmente verificare che

Se x, y t.c. $\sigma_x \neq 0$ $\sigma_y \neq 0$ (deviazione standard non nulla)

$z_x = z_y \iff \exists \beta_0, \beta_1$ t.c. $y = \beta_0 e + \beta_1 x$ $\beta_1 > 0$

$z_x = -z_y \iff \exists \beta_0, \beta_1$ t.c. $y = \beta_0 e + \beta_1 x$ $\beta_1 < 0$

In altre parole $y = \beta_0 e + \beta_1 x$ per qualche $\beta_0, \beta_1 \iff$ $z_x = \pm z_y$ in tal caso diremo che y è perfettamente correlato in modo lineare con x

Poiché z_x dipende da x, z_x è tanto più vicino a $\pm z_y$ quanto più x è linearmente correlato a y.

Poiché $\|z_x\| = \|\pm z_y\| = \sqrt{n} \Rightarrow$ z_x e $\pm z_y$ differiscono solo nell'orientazione quindi $\cos\theta$ è una misura naturale di quanto z_x sia vicino a $\pm z_y$

$$\rho_{xy} = \cos\theta = \frac{z_x^T z_y}{\|z_x\|\|z_y\|} = \frac{z_x^T z_y}{n} = \frac{(x-\mu_x e)^T (y-\mu_y e)}{\|x-\mu_x e\| \|y-\mu_y e\|}$$

è detto coefficiente di correlazione lineare

$\rho_{xy} = 0$ $x \perp y \Rightarrow x, y$ sono completamente scorrelati

$|\rho_{xy}| = 1 \iff y$ è perfettamente correlato ad x

y positiv. correlato con x se $\beta_1 > 0$

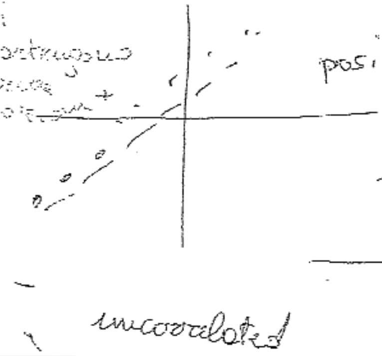
y negativ. correlato con x se $\beta_1 < 0$

$|\rho_{xy}|$ misura il grado in cui y è linearmente legato a x

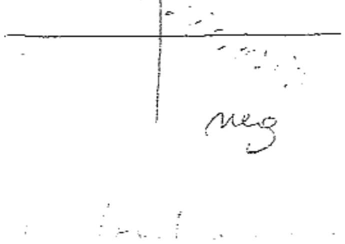
$\rho_{xy} \simeq 1$ posit. corr.

$\rho_{xy} \simeq -1$ neg. corr.

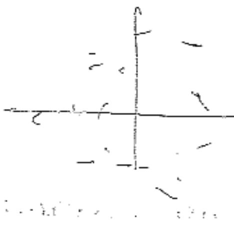
osservo i dati appartengono quasi a una retta con inclinazione +



posit



neg



uncorrelated

Rivedere il prodotto scalare

Rivedere le norme vettoriali

30

Procedura di Gram - Schmidt

[Assegnati n vettori $x_1, x_2, \dots x_n$ di una base $B = \{x_1, \dots x_n\}$ costruire una base
ortonormale $u_1, u_2, \dots u_n$

$K = 1$ $\qquad u_1 = \dfrac{x_1}{\|x_1\|}$

K

$$t_k = x_k - \sum_{i=1}^{k-1} \left(u_i^H x_k\right) u_i =$$

$$u_k = \frac{t_k}{\|t_k\|}$$

$\Rightarrow$ Fattorizzazione QR

Ogni matrice $A \in \mathbb{R}^{m \times n}$ (con colonne linearmente indip.) può essere univocamente fattorizzata nel prod.

$A = QR$

Q ortonormale
R triangolare superiore con elementi diagonali positivi

Quando consideriamo una matrice $A \in \mathbb{R}^{m \times n}$ $\{A_{*1}, A_{*2}, \dots A_{*n}\}$
matrice con colonne l.i. allora il
Il processo di Gram - Schmidt $\qquad a_1 \quad a_2 \quad a_n$
applicato alle colonne di A
costruisce una base $\{q_1, q_2, \dots q_n\}$ per $R(A)$

ove

$$q_1 = \frac{a_1}{\nu_1} \qquad \nu_1 = \|a_1\|$$

$$q_k = \frac{a_k - \sum_{i=1}^{k-1} (q_i^H a_k)\, q_i}{\nu_k}$$

$$\nu_k = \left\| a_k - \sum_{i=1}^{k-1} (q_i^T a_k)\, q_i \right\|$$

$$= \| a_k - \dots q_i \|$$

La relazione precedente può essere ~~scritta~~ ~~come~~
espressa formalmente come

$$(a_1|a_2|\dots a_n) = (q_1|q_2|\dots|q_n)\begin{pmatrix} v_1 & \langle q_1|a_2\rangle & \langle q_1|a_3\rangle & \dots & \langle q_1| \\ 0 & v_2 & \langle q_2|a_3\rangle & & \langle q_2, a \\ \vdots & & v_3 & & \langle q_3, a \\ & & & \ddots & \vdots \\ 0 & & & v_n & \langle q_n \\ & & & & v_n \end{pmatrix}$$

31

Questo ci dice che è possibile fattorizzare una matrice $\underset{m\times n}{A} = \underset{m\times n}{Q}\underset{n\times n}{R}$ ove le colonne di Q formano una base ortonormale di $R(A)$ ed R è una matrice triangolare sup con elementi diagonali positivi.

→ passare agli autovalori ed autovettori

Autovalori ed Autovettori (aggiunta)

Vera definizione $Ax = \lambda x \quad \lambda \in \mathbb{C},\ x \neq 0 \quad A \in \mathbb{C}^{m \times n}$

$Ax = \lambda x \Leftrightarrow \underbrace{(A - \lambda I)} x = 0 \quad \Leftrightarrow x \neq 0 \quad x \in N(A - \lambda I)$

questa matrice deve essere singolare altrimenti l'unica sol. è data da $x = 0$

$N(A - \lambda I)$ contiene vettori diversi da $\emptyset$ solo se

$A - \lambda I$ è singolare $\Leftrightarrow$ λ sono i valori che rendono singolare $A - \lambda I$ $\Leftrightarrow$ $\det(A - \lambda I) = 0$

$A \in \mathbb{C}^{m \times n} \quad \lambda \in \mathbb{C} \quad x \in \mathbb{C}^n \quad \ni' \quad Ax = \lambda x$ (Ax vettore parallelo a x)

λ detto autovalore di A

x " autovettore " "

se A singolare, $\lambda = 0$ è un autovettore

Nota: $\sum_{i=1}^{m} \lambda_i = \text{traccia}(A)$

(λ, x) eigenpair di A

$\sigma(A)$ insieme degli autovalori distinti di A $\Rightarrow$ spettro di A

- $\lambda \in \sigma(A) \Leftrightarrow A - \lambda I$ è singolare $\Leftrightarrow \det(A - \lambda I) = 0$
- $\{x \neq 0 \mid x \in N(A - \lambda I)\}$ insieme degli autovettori associati a λ
  $N(A - \lambda I)$ detto eigenspace di A (autospazio di A)
- $y^H \ni' y^H (A - \lambda I) = 0$ autovalore sinistro di A

→ Geometricamente $Ax = \lambda x$ attraverso la trasformazione lineare associata di A gli autovettori sono quei particolari vettori che vengono modificati in grandezza e segno ma l'orientazione di Ax è la stessa di x

λ fornisce semplicemente la quantità con cui x viene allungato o accorciato dalla trasformazione A.

$p(\lambda) = \underbrace{\det(A - \lambda I)}_{\text{eq. caratteristica}}$ polinomio caratteristico di A

$p(\lambda) = 0$ eq. caratteristica di A

il grado di $p(\lambda)$ è n ed il coefficiente del termine di grado max è $(-1)^n \lambda^n$

- Gli autovalori di A sono le soluzioni del polinomio caratt. o equiv. le radici di $p(\lambda) = 0$

- A ha n autovalori, alcuni dei quali possono essere complessi (anche se la matrice $A \in \mathbb{R}^{n \times n}$) ed alcuni autovalori si possono ripetere

$$\boxed{\begin{array}{l} A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \qquad \lambda_{1/2} = 1 \pm i \\ N(A - \lambda I) \text{ per } \lambda = 1+i \equiv \text{span}\left\{ \begin{pmatrix} i \\ 1 \end{pmatrix} \right\} \\ N(A - \lambda I) \text{ per } \lambda = 1-i \equiv \text{span}\left\{ \begin{pmatrix} -i \\ 1 \end{pmatrix} \right\} \end{array}}$$

- Se $A \in \mathbb{R}^{n \times n}$ e $\lambda \in \sigma(A)$ allora $\bar{\lambda} \in \sigma(A)$

Gli autovalori di una matrice triangolare sono gli elementi diagonali

$$\boxed{\begin{array}{l} \text{trace}(A) = \sum_{\lambda_i \in \sigma(A)} \lambda_i \\ \det(A) = \prod_{\lambda \in \sigma(A)} \lambda_i \end{array}}$$

$$\rho(A) = \max_{\lambda \in \sigma(A)} |\lambda| \qquad \text{raggio spettrale}$$

$\rho(A) \leq \|A\|$ (maggiorazione di $\rho(A)$ economica da calcolare specialmente in $\|\ \|_1$ o $\|A\|_\infty$

Cerchi di Gershgorin

$$\sigma(A) \subset \bigcup G_r \qquad G_r = \left\{ z \in \mathbb{C} \;\middle|\; |z - a_{ii}| \leq r_i \quad r_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \right\} \quad i = 1 \dots n$$

in altre parole gli autovalori di una matrice sono contenuti in un insieme di cerchi ~~[illegible]~~ centrati in a_{ii} e raggio dato dalla somma dei valori assoluti di A in senza a_{ii}

SIMILARITA'

$A, B \in \mathbb{R}^{n \times n}$ sono simili $\iff$ $\exists P$ non singolare t.c. $P^{-1}AP = B$
$P^{-1}AP$ è detta trasformazione di similarità di A

Un problema fondamentale nell'alg. lin. è quello di ridurre una matrice ~~A~~ A assegnata ad una forma più semplice per mezzo di trasformazioni di similarità.

Le matrici diagonali hanno la forma più semplice pertanto è interessante chiedersi se e quando una matrice è simile ad una matrice diagonale

DEF A si dice <u>diagonalizzabile</u> $\iff$ A è simile ad una matrice diagonale

<u>PROP</u>
A è diagonalizzabile $\iff$ A possiede n autovettori l.i.
Inoltre t.c. $P^{-1}AP = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ $\iff$ le colonne di P costituiscono un insieme di autovettori l.i. associati ai λ_i (cioè ciascuna coppia (λ_i, P_{*i}) è una autocoppia (eigenpair) di A

PROP Matrici simili hanno gli stessi autovalori
DEF trasformazione unitaria → U è unitaria se [illegible]

<u>Teor. di Schur</u>

Ogni matrice quadrata è simile mediante una trasform. unitaria ad una matrice triangolare superiore $\forall A \in \mathbb{C}^{n \times n}$ $\exists U \in \mathcal{U}_n$ $(U^H U = I = U U^H)$ $\exists T$ triang. sup. t.c. $U^H A U = T$
Inoltre t_{ii} sono gli autovalori di A

MOLTEPLICITA'

sia
$\lambda \in \sigma(A) = \{\lambda_1, \lambda_2, \dots \lambda_s\}$

molteplicità algebrica $\text{alg mult}_A(\lambda)$ è il # di volte in cui λ è radice di $p(\lambda)$

molteplicità geometrica $\text{geo mult}_A(\lambda)$ è $\dim \mathcal{N}(A - \lambda I)$
ovvero il numero max di autovett. l.i. associati a λ.

λ si dice semisemplice se $\text{alg mult}_A(\lambda) = \text{geo mult}_A(\lambda)$

In generale

$$\text{geo mult}_A(\lambda) \le \text{alg mult}_A(\lambda)$$

Un autovalore λ si dice semplice $\text{alg mult}_A(\lambda) = 1$

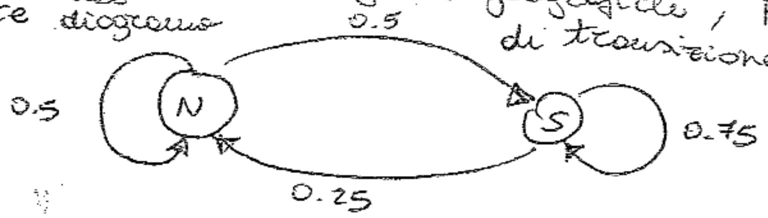
(N.B. λ è semplice $\Rightarrow$ λ è semisemplice non vale il viceversa)

PROP A è diagonalizzabile $\Leftrightarrow$ ogni autovalore di A è semisemplice

$$\text{alg mult}_A(\lambda) = \text{geo mult}_A(\lambda) \quad \forall \lambda \in \sigma(A)$$

Modello delle Migrazioni di una popolazione

Supponiamo che il grafico delle migrazioni di una popolazione ~~...~~ tra due regioni geografiche, p.e. N e S sia il rappresentato dal seguente diagramma di transizione



Ciascun anno il 50% della pop. del nord migra a sud
il 25 della " del sud " a Nord

Problema: Se il diagramma di transizione rimane invariato la popol. del Nord si sposterà completamente al Sud oppure la distribuzione della popolazione si stabilizzerà prima che il Nord rimanga deserto?

Soluz. Indichiamo con n_K ed s_K la porzione della pop. totale che vivono al Nord ed al Sud ~~...~~ al termine del K anno di osservazione e assumiamo che $n_K + s_K = 1$

Dal diagramma di transizione risulta che al termine dell'anno $K+1$ si ha

$$\begin{cases} n_{K+1} = 0.5\, n_K + 0.25\, s_K \\ s_{K+1} = 0.5\, n_K + 0.75\, s_K \end{cases} \quad \Leftrightarrow \quad p_K = \begin{pmatrix} n_K \\ s_K \end{pmatrix}$$

$$p_{K+1} = \begin{array}{c c} & \begin{array}{cc} N & S \end{array} \\ \begin{array}{c} N \\ S \end{array} & \begin{pmatrix} 0.5 & 0.25 \\ 0.5 & 0.75 \end{pmatrix} \end{array} p_K$$

$$p_{K+1} = T\, p_K \qquad T \text{ matrice di transizione}$$

Introducendo un vettore di popolazione iniziale p_0 (36)

$$p_1 = T p_0 \qquad p_2 = T p_1 = T T p_0 = T^2 p_0, \ \dots \qquad p_{k+1} = T p_k = T^k p_0$$

la popolazione dipende dalle potenze della matrice T

Questa sequenza $T^1, T^2 \dots T^k$ se calcoliamo le matrici prodotto sembra convergere alla matrice

$$T^\infty = \lim_{k \to \infty} T^k = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & 2/3 \end{pmatrix}$$

e quindi la distribuzione della pop. al limite è

$$p_\infty = \lim_{k \to \infty} p_k = \lim_{k \to \infty} T^k p_0 = \left(\lim_{k \to \infty} T^k \right) p_0 = \begin{pmatrix} 1/3 & 1/3 \\ 2/3 & 2/3 \end{pmatrix} \begin{pmatrix} n_0 \\ s_0 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{n_0 + s_0}{3} \\ \frac{2(n_0 + s_0)}{3} \end{pmatrix}$$

Questo significa che se il grafico di transizione rimane inalterato la distribuzione della popolazione si stabilizzerà con $\frac{1}{3}$ della pop. nel Nord e $\frac{2}{3}$ nel Sud indipendentemente dalla pop. iniziale.

Se calcolate le potenze di T vi accorgerete come il processo si stabilizza dopo circa 5 anni $\Rightarrow$ gli individui continuano a muoversi, ma la porzione della popol. rimane invariata (costante) dopo circa 5 anni.

OSS

Se A è diagonalizzabile $\quad A = PDP^{-1} = P \operatorname{diag}(\lambda_1, \dots \lambda_n) P^{-1}$

$\Rightarrow \quad A^k = PD^kP^{-1} = P \operatorname{diag}(\lambda_1^k, \dots \lambda_n^k) P^{-1}$

* [illegible] è un autov. di A^k

Nel nostro esempio $T = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{3}{4} \end{pmatrix}$ possiamo calcolare $\lim_k T^k$ utilizzando la sua decomposizione spettrale

Calcoliamo gli autovalori di T

$\lambda_1 = 1 \qquad \lambda_2 = \frac{1}{4}$

gli autovettori associati $\quad x_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \qquad x_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

$P = [x_1 | x_2]$

$$\Rightarrow \quad T = P \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{4} \end{pmatrix} P^{-1} \qquad \Rightarrow \quad T^k = P \begin{pmatrix} 1^k & 0 \\ 0 & \left(\frac{1}{4}\right)^k \end{pmatrix} P^{-1}$$

OSS Il problema della migrazione della popolazione è un esempio di una classe più ampia di processi evolutivi noti come catene di Markov per le quali valgono le seguenti osservazioni

per l'esempio

1) la velocità con cui il processo evolutivo si stabilizza dipende da quanto $\left(\frac{1}{4}\right)^k \to 0$
in altri termini la grandezza dell'autovalore sottodominante determina la velocità di evoluzione
2) la distribuzione iniziale p_0 non influisce nel comportamento al limite, ma comunque caratterizza il comportamento transitorio del sistema (ovvero per k piccolo)

METODI PER CALCOLO DEGLI AUTOVALORI e/o AUTOVETTORI

Metodo delle potenze per il calcolo dell'autovalore e dell'autovec. dominante → metodo iterativo per approx l'autovalore di modulo max e il suo autovec. associato
Sulla base di questo metodo sono stati sviluppati altri algoritmi adatti ad approssimare autovalori di matrici sparse e di grandi dimensioni. Facile dimostrare la conv. nel caso di matrice diagonalizzabile che abbia un solo autovalore di modulo massimo

Sia $A \in \mathbb{C}^{n\times n}$
$\lambda_1, \lambda_2 \dots \lambda_n$ si con n. autovec. $x_1, x_2, \dots, x_n$ l.i. ed autovalori
$$|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_n|$$
ovvero alg molt$(\lambda_1) = 1$ e non esistono altri autovalori con lo stesso modulo (λ_1 reale)

Fissiamo un vettore iniziale $t_0 \in \mathbb{C}^n$ e consideriamo la successione $\{y_k\}$ $k = 1, 2, \dots$ definita

$$y_0 = t_0$$
$$y_k = A y_{k-1} \qquad k \geq 1$$

$x_1, x_2 \dots x_n$ sono l.i. $\Rightarrow$ $t_0 = \sum_{i=1}^{n} \alpha_i x_i$ con gli α_i si $\alpha_1 \neq 0$

$$\Rightarrow y_k = A y_{k-1} = \dots = A^k y_0 = A^k t_0 = A^k \sum^n \alpha_i x_i = \sum \alpha_i A^k x_i =$$

$$= \sum_{i=1}^{n} \alpha_i \lambda_i^k x_i = \alpha_1 \lambda_1^k x_1 + \alpha_2 \lambda_2^k x_2 + \dots + \alpha_n \lambda_n^k x_n$$

$$= \lambda_1^k \left[\alpha_1 x_1 + \alpha_2 \left(\frac{\lambda_2}{\lambda_1}\right)^k x_2 + \dots + \alpha_n \left(\frac{\lambda_n}{\lambda_1}\right)^k x_n \right]$$

$$= \lambda_1^k \left[\alpha_1 x_1 + \sum_{i=2}^{n} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^k x_i \right]$$

Indicando con $y_k(j)$ e $x_i(j)$ la j-sima componente dei vettori y_k e x_i si ha

$$y_k(j) = (A^k y_0)_{(j)} = \lambda_1^k \left[\alpha_1 x_1(j) + \sum_{i=2}^{n} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^k x_i(j) \right]$$

$$y_{k+1}(j) = \lambda_1^{k+1} \left[\alpha_1 x_1(j) + \sum_{i=2}^{m} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^{k+1} x_i(j) \right]$$

$\Rightarrow$ Se $x_1(j) \neq 0$

$$\frac{y_{k+1}(j)}{y_k(j)} = \lambda_1 \frac{\alpha_1 x_1(j) + \sum_{i=2}^{m} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^k x_i(j)}{\alpha_1 x_1(j) + \sum_{i=2}^{n} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^{k+1} x_i(j)}$$

$$\lim_{k \to \infty} \frac{y_{k+1}(j)}{y_k(j)} = \lambda_1$$

ossero da un certo indice k in poi l'autovalore dominante può essere approssimato mediante il rapporto tra le componenti dei vettori y_{k+1} ed y_k

Il metodo permette di approssimare anche l'autovec. dominante Infatti dalla relazione

$$y_k = \lambda_1^k \left[\alpha_1 x_1 + \sum_{i=2}^{n} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^k x_i \right]$$

otteniamo $\lim_{k} \frac{y_k}{\lambda_1^k} = \alpha_1 x_1$

quindi in termini di componenti $\lim \frac{y_k}{\;} = \frac{x_1}{\;}$ V

Poichè per K suff. elevato l'indice m della componente di modulo massimo di y_K rimane costante è la successione (33)

$$\frac{y_K}{y_K(m)} \rightsquigarrow \frac{x_1}{\|x_1\|_\infty}$$

OSS VANTAGGI Ciascuna iterata richiede solo una moltiplicazione matrice - vettore

SVANTAGGI Se si desiderano gli altri autovalori bisogna fare qualche altra cosa $\Rightarrow$ NON ADATTO x CALCOLARE TUTTI GLI AUTOVALORI

- La velocità di convergenza dipende da quanto velocemente il rapporto $\left(\frac{\lambda_2}{\lambda_1}\right)^k \to 0$ $\Rightarrow$ convergenza lenta se $|\lambda_1| \simeq |\lambda_2|$

Altre propr. di λ e x

$A \in \mathbb{C}^{n \times n}$

A nonsingolare $\Leftrightarrow$ $\det(A) \neq 0$ $\Leftrightarrow$ $\det(A) = \prod_{i=1}^{n} \lambda_i \neq 0$ $\Leftrightarrow$ $\lambda_i \neq 0$

1) (λ, x) è una autocoppia di A $\Leftrightarrow$ (λ^{-1}, x) è autocoppia di A^{-1}

2) $\forall \alpha \notin \sigma(A)$ x è autovettore di A $\Leftrightarrow$ x è autovett. di $(A-\alpha I)^{-1}$

dim

$Ax = \lambda x \Leftrightarrow Ax - \alpha x = \lambda x - \alpha x \Leftrightarrow (A-\alpha I)x = (\lambda-\alpha)x,$

$\Leftrightarrow (\lambda-\alpha)^{-1}x = (A-\alpha I)^{-1}x$

Varianti del metodo delle potenze : Metodo delle potenze inverse

- Assegnata una approx $\alpha \notin \sigma(A)$ di un $\lambda \in \sigma(A)$ l'algoritmo del metodo delle pot. inverse permette di calcolare una autocoppia (λ, x) di una matrice diagonalizzabile $A \in \mathbb{C}^{n \times n}$ applicando il metodo delle potenze alla matrice $B = (A-\alpha I)^{-1}$

Se $|\lambda-\alpha| < |\lambda_i-\alpha| \quad \forall \lambda_i \in \sigma(A)$ allora ④

$\Rightarrow (\lambda-\alpha)^{-1}$ è l'autovalore dominante di $B=(A-\alpha I)^{-1}$

infatti

$$|\lambda-\alpha|^{-1} > |\lambda_i-\alpha|^{-1}$$

Quindi applicando il metodo delle potenze a B si può calcolare la coppia (λ, x). In particolare considerato un vettore iniziale $t_0 \notin R(B-\lambda I)$?? il metodo delle potenze inverse è

$$(A-\alpha I)\, y_k = x_{k-1}$$

$$x_k = \frac{y_k}{\beta_k} \qquad \text{con } \beta_k \text{ scalare t.c. } \|x_k\|_\infty = 1$$

<u>OSS</u> Ad ogni passo il metodo richiede la risol. del sistema lineare $\Rightarrow$ poiché la matrice dei coeff. rimane invariata conviene applicare la fatt. LU

SVANTAGGI:

Solo una coppia può essere calcolata

Serve ogni volta una approx di λ

Convergenza (lenta) che dipende da quanto $\left[\frac{(\lambda-\alpha)}{(\lambda_i-\alpha)}\right]^n \longrightarrow 0$

... algoritmo QR per il calcolo degli autovalori di A

Idea: assegnata una generica matrice $A \in \mathbb{R}^{n\times n}$ l'idea è quella di alternare il calcolo dei fattori della fatt. QR e di scambiare l'ordine del loro prodotto
→ generare un meccanismo per trasformare la matrice A in una matrice simile con struttura triangolare

Sia $A_1 \equiv A \in \mathbb{R}^{n\times n}$

costruiamo la fatt. QR di A_1: $A_1 = Q_1 R_1$
poniamo : $A_2 = R_1 Q_1$
" " QR di A_2: $A_2 = Q_2 R_2$
poniamo : $A_3 = R_3 Q_3$

In generale
$A_{K+1} = R_K Q_K$ con R_K e Q_K fattori della fatt. QR di A_K

Osserviamo che posto
$P_K = Q_1 Q_2 \dots Q_K$
P_K è una matrice ortogonale che permette di esprimere ~~...~~ ogni matrice A_K in modo simile ad A:

Infatti risulta da
$A_{K+1} = R_K Q_K \Rightarrow R_K = A_{K+1} Q_K^T$
ma $A_K = Q_K R_K \Rightarrow A_K = Q_K A_{K+1} Q_K^T \quad \forall K$

⇒ le matrici della succ. $\{A_K\}$ sono tutte simili tra di loro.
Sotto opportune ipotesi la succ. converge ad una matrice triangol. sup che ha come elem. diag. gli autov. di A.

Teorema di convergenza per il metodo QR

<u>TEOR</u>. $A \in \mathbb{C}^{n\times n}$ e' λ_i abbiano tutti moduli distinti
$|\lambda_1| > |\lambda_2| > \dots > |\lambda_n| > 0$
Indicata con X la matrice degli autovettori di A e'
$A = XDX^{-1}$ con $D = \mathrm{diag}(\lambda_1, \dots \lambda_n)$
si supponga che $\exists$ una fatt. LU della matrice X^{-1}
Allora $\exists \{S_k\}_k$ successioni di matrici t.c. (S_k sono matrici di fase)

$$\lim_k S_k^H R_k S_{k-1} = \lim_{k\to\infty} S_{k-1}^H A_k S_{k-1} = T$$

$$\lim_k S_{k-1}^H Q_k S_k = I$$

con T matrice triang. sup. con elementi diagonali uguali λ_i
Quindi gli elementi diagonali di A_k tendono agli autovalori di A

Se A è hermitiana ($A^H = A$) allora T è diagonale.

- Costo computazionale

Il metodo QR applicato ad una matrice $n\times n$ richiede ad ogni passo $O(n^3)$ operazioni moltiplicative
Per ridurre il costo computazionale generalmente si trasforma la matrice A in una matrice in forma di Hessemberg superiore

<u>DEF</u> ~~Si dà~~ Una matrice $H \in \mathbb{C}^{n\times n}$ è detta di Hessemberg sup se $h_{ij} = \{ 0 \quad \forall j \le i-2$ ovvero

$$H = \begin{pmatrix} h_{11} & & & \\ h_{21} & h_{22} & & \\ 0 & h_{32} & h_{33} & \\ 0 & \ddots & & \\ 0 & 0 & & \end{pmatrix}$$

h_{11}
$h_{21}\ h_{22}$
$h_{31}\ h_{32}\ h_{33}$
$h_{i,i-1}\ h_{ii}$

h_{33}

- Velocità di convergenza

La velocità di convergenza del metodo QR (nelle ip. del teor. di convergenza) dipende dai rapporti $\left|\frac{\lambda_i}{\lambda_j}\right| \ \forall i>j$ e quindi se $|\lambda_1|>|\lambda_2|>\dots$ dal numero

$$\max_{1\leq i\leq m-1}\left|\frac{\lambda_{i+1}}{\lambda_i}\right| \qquad (*)$$

Se tale rapporto è vicino ad 1 la convergenza può essere lenta. Per accellerare si utilizza una tecnica di traslazione dello spettro di A, detta "tecnica dello shift"

Sia μ una approx di $\lambda\in\sigma(A)$, e consideriamo la matrice $A-\mu I$. Applichiamo il metodo QR a quest'ultima matrice (che ha autov. $\lambda-\mu$)

$$\begin{cases} A_k-\mu I = Q_kR_k & (1)\\ A_{k+1} = R_kQ_k+\mu I & (2)\end{cases}$$

da (1) $\Rightarrow$ $R_k = Q_k^H(A_k-\mu I)$ sostituiamo in (2)

$A_{k+1} = Q_k^H(A_k-\mu I)Q_k+\mu I = Q_k^HA_kQ_k - \cancel{\mu I}+\cancel{\mu I}$

A_{k+1} è simile ad A_k attraverso una matrice unitaria

155

Poichè gli autovalori di $A-\mu I$ sono del tipo $\lambda_i-\mu$ la velocità di convergenza è regolata dalla $(*)$ si può scegliere il parametro μ in modo da acc. la conv. del metodo QR con shift

Conviene scegliere $\mu\simeq\lambda_n$