

AUTOVALORI E AUTOVETTORI

1. Definizioni

Sia $A \in \mathbb{C}^{n \times n}$, il numero $\lambda \in \mathbb{C}$, reale o complesso, è detto **autovalore** di A se esiste un vettore $x \in \mathbb{C}^n$, $x \neq 0$, tale che valga la relazione

$$Ax = \lambda x \tag{1}$$

Allora il vettore x è detto **autovettore** di A corrispondente all'autovalore λ .

L'insieme degli autovalori di una matrice A costituisce lo *spettro* di A e il modulo massimo degli autovalori di A è detto *raggio spettrale* di A e si indica con $r(A)$.

Il sistema (1) può essere riscritto nella forma

$$(A - \lambda I)x = 0. \tag{2}$$

Per il teorema fondamentale dei sistemi lineari esso ammette soluzioni non nulle se e solo se

$$\det(A - \lambda I) = 0,$$

cioè se

$$\det(A - \lambda I) = P_n(\lambda) = a_0 \lambda^n + a_1 \lambda^{n-1} + \dots + a_n = 0, \tag{3}$$

in cui

$$a_0 = (-1)^n$$

$$a_1 = (-1)^{n-1} \text{tr}(A) = (-1)^{n-1} \sum_{i=1}^n a_{ii} \quad e \quad a_n = \det(A).$$

Il polinomio $P_n(\lambda)$ è detto *polinomio caratteristico* di A e l'equazione $P(\lambda)=0$ è detta *equazione caratteristica* di A.

Gli autovalori di A sono tutti e soli i valori che annullano $P_n(\lambda)$, cioè le radici di $P_n(\lambda)$. Poiché un polinomio di grado n ammette sempre n radici reali o complesse, distinte o coincidenti, una matrice $n \times n$ ha sempre n autovalori, non necessariamente distinti.

Dalle relazioni che legano i coefficienti e le radici di un'equazione algebrica risulta che

$$\sum_{i=1}^n I_i = tr(A) \quad e \quad \prod_{i=1}^n I_i = \det(A).$$

Gli autovettori corrispondenti agli autovalori di A sono le soluzioni non nulle del sistema lineare omogeneo (2). Quindi un autovettore corrispondente ad un autovalore λ risulta determinato a meno di una costante moltiplicativa $\alpha \neq 0$, cioè se x è un autovettore di A anche αx è un autovettore di A corrispondente allo stesso autovalore.

Esempio1:

Il polinomio caratteristico della matrice A

$$A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$$

si ricava dal determinante

$$\det(A - II) = \det \begin{bmatrix} 1-I & 3 \\ 3 & 1-I \end{bmatrix} = (1-I)^2 - 9 = I^2 - 2I - 8.$$

L'equazione caratteristica corrispondente

$$I^2 - 2I - 8 = 0$$

ha come radici $\lambda_1 = -2$ e $\lambda_2 = 4$ che sono gli autovalori della matrice A.

L'autovalore corrispondente all'autovalore $\lambda_1=-2$ si calcola risolvendo il sistema (2) che in questo caso diventa

$$\begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

Dalla prima equazione si ottiene $x_1+x_2=0$ da cui $x_1=-x_2$

Da cui segue che qualunque vettore

$$x = \alpha \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Con $\alpha \neq 0$ è un autovettore di A corrispondente a λ_1 .

2. Proprietà degli autovalori

- Gli autovalori di una matrice diagonale o triangolare sono uguali agli elementi diagonali.
- Se λ è un autovalore di una matrice A non singolare e x un autovettore corrispondente, allora risulta
 1. $\lambda \neq 0$
 2. $1/\lambda$ è autovalore di A^{-1} con x autovettore corrispondente. Infatti dalla (1) si ha

$$x = \lambda A^{-1} x$$

e quindi

$$I \neq 0 \quad e \quad A^{-1} x = \frac{1}{\lambda} x.$$

- Per il raggio spettrale di A vale $\rho(A) \leq \|A\|$

Infatti abbiamo $\|Ax\| = |\lambda| \cdot \|x\|$

perciò vale $\|Ax\| \leq \|A\| \cdot \|x\| \quad \Rightarrow \quad |\lambda| \cdot \|x\| \leq \|A\| \cdot \|x\| \quad \Rightarrow \quad |\lambda| \leq \|A\|$

- Se λ è un autovalore di una matrice A , allora esso è anche autovalore di A^T .
Infatti, poiché

$$\det A^T = \det A,$$

si ha

$$0 = \det(A - \lambda I) = \det(A - \lambda I)^T = \det(A^T - \lambda I).$$

- Se λ è un autovalore di una matrice A ortogonale, cioè tale che $A^T = A^{-1}$, allora risulta $|\lambda| = 1$. Infatti dalla relazione (1) si ha

$$(Ax)^T = (Ix)^T$$

e quindi

$$x^T A^T = Ix^T,$$

da cui si ha

$$x^T A^T Ax = I Ix^T x.$$

Poiché A è ortogonale, $A^T A = I$ e quindi si ha

$$x^T x = I^2 x^T x.$$

Essendo $x^T x \neq 0$, segue che

$$I^2 = 1, \quad e \quad quindi \quad |I| = 1.$$

- Se λ è un autovalore di una matrice A , allora λ^k è anche autovalore di A^k .

Infatti dalla relazione $Ax = \lambda x$ si ottiene

$$A^k x = \underbrace{A \cdot A \dots A}_{k \text{ volte}} x = \underbrace{A \cdot A \dots A}_{k-1 \text{ volte}} I x = \underbrace{A \cdot A \dots A}_{k-2 \text{ volte}} I^2 x = \dots = I^k x$$

Poiché $P(\lambda)$ è un polinomio di grado n , esso ammette n radici reali o complesse, non necessariamente tutte distinte, quindi una matrice A $n \times n$, se contiamo le radici multiple di $P(\lambda)$ secondo la loro molteplicità, ammette n autovalori non necessariamente distinti. In particolare vale il seguente teorema:

Def.

Si definisce **molteplicità algebrica** di un autovalore λ , e si indica con il simbolo $\sigma(\lambda)$, la molteplicità di λ in quanto radice dell'equazione caratteristica (3) .

Esempio 2: Se consideriamo la matrice

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Si ha $P(\lambda)=(1-\lambda)^3$.

L'autovalore $\lambda=1$ ha molteplicità algebrica pari a 3. In corrispondenza di questo autovalore si possono però determinare 3 autovettori linearmente indipendenti

$$x_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad x_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad x_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Esempio 3:Se consideriamo la matrice

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Si ha $P(\lambda)=(1-\lambda)^3$.

L'autovalore $\lambda=1$ ha molteplicità algebrica pari a 3 ma in corrispondenza di questo autovalore si può determinare un solo autovettore linearmente indipendente, cioè

$$x_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Tutti gli altri sono multipli di esso.

Per poter differenziare questi due casi, che corrispondono a differenti proprietà delle matrici corrispondenti, si definisce **molteplicità geometrica** di un autovalore λ , e si indica con il simbolo $\mu(\lambda)$, il numero di autovettori linearmente indipendenti corrispondenti a λ .

La molteplicità algebrica e molteplicità geometrica sono legate dalla seguente relazione

$$l \leq m(l) \leq s(l) \leq n.$$

3. Localizzazione degli autovalori

Si vuole ora dare una locazione preliminare dello spettro della matrice nel piano complesso.

Def.

Sia $A \in C^{n \times n}$. I cerchi del piano complesso

$$K_i = \left\{ z \in C : |z - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \right\}, i = 1, 2, \dots, n$$

di centro a_{ii} e raggio $r_i = \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|$ sono detti Cerchi di Gerschgorin.

Vale il seguente teorema.

Teorema: (Primo teorema di Gerschgorin)

Gli autovalori della matrice A di ordine n sono tutti contenuti in

$$\bigcup_{i=1, \dots, n} K_i.$$

Osservazione:

Poiché il teorema precedente può essere applicato anche alla matrice A^T che ha gli stessi autovalori di A , si osserva che gli autovalori di A appartengono anche all'unione dei cerchi

$$H_j = \left\{ z \in \mathbb{C} : |z - a_{jj}| \leq \sum_{\substack{i=1 \\ i \neq j}}^n |a_{ji}| \right\}, j = 1, 2, \dots, n$$

e quindi gli autovalori di A appartengono all'insieme

$$\left(\bigcup_{i=1, \dots, n} K_i \right) \cap \left(\bigcup_{j=1, \dots, n} H_j \right)$$

Teorema: (Secondo teorema di Gerschgorin)

Se l'unione M_1 di k cerchi di Gerschgorin è disgiunta dall'unione M_2 dei rimanenti $n-k$, allora k auto valori appartengono a M_1 e $n-k$ auto valori appartengono a M_2 .

Esempio: Si consideri la matrice

$$A = \begin{bmatrix} 15 & -2 & 2 \\ 1 & 10 & -3 \\ -2 & 1 & 0 \end{bmatrix}$$

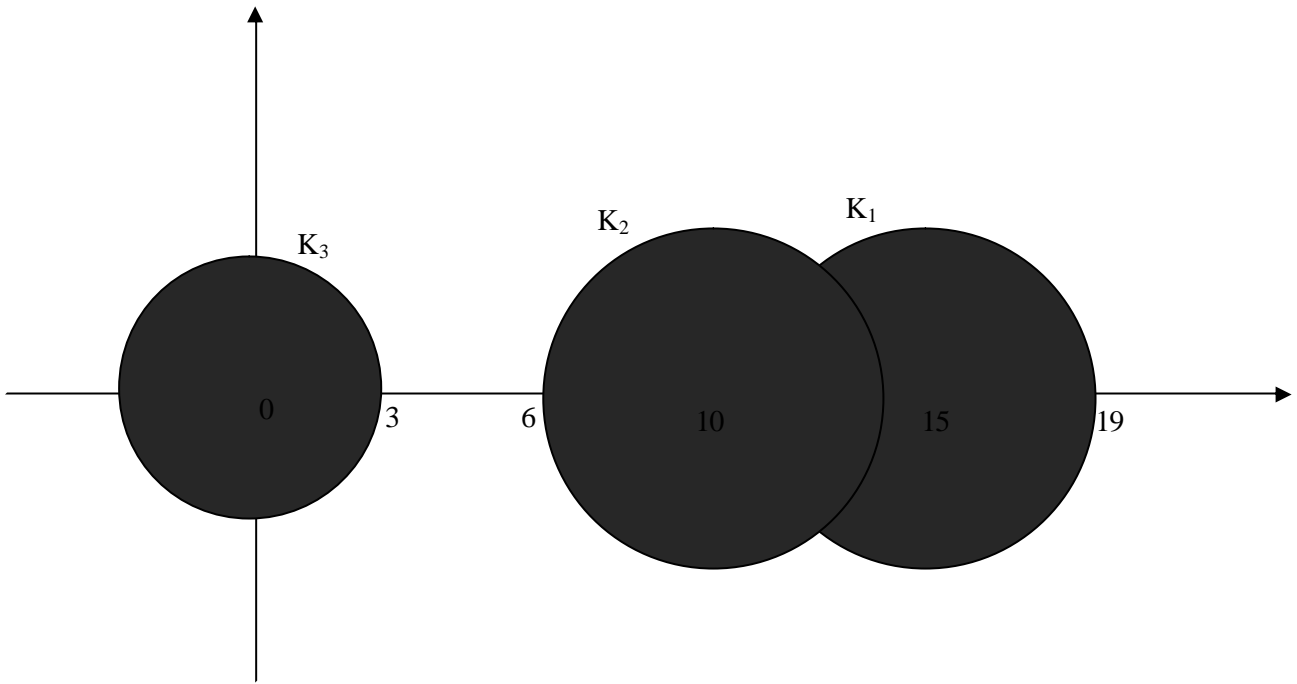
alla quale sono associati i seguenti cerchi di Gerschgorin

$$K_1 = \{ z \in \mathbb{C} : |z - 15| \leq 4 \}$$

$$K_2 = \{ z \in \mathbb{C} : |z - 10| \leq 4 \}$$

$$K_3 = \{ z \in \mathbb{C} : |z| \leq 3 \}$$

rappresentati nella figura seguente



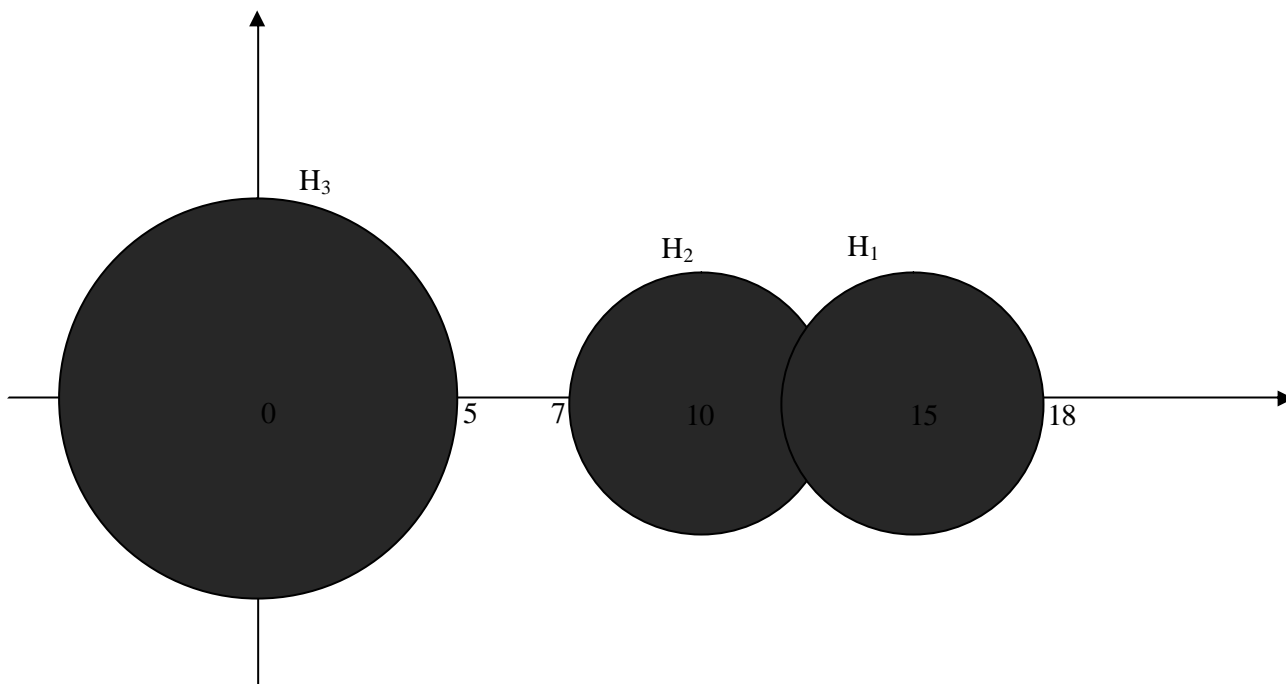
Per il primo teorema di Gerschgorin possiamo affermare che gli autovalori di A stanno nelle aree grigie. Si considerino poi i cerchi associati alla matrice A^T

$$H_1 = \{z \in C : |z - 15| \leq 3\}$$

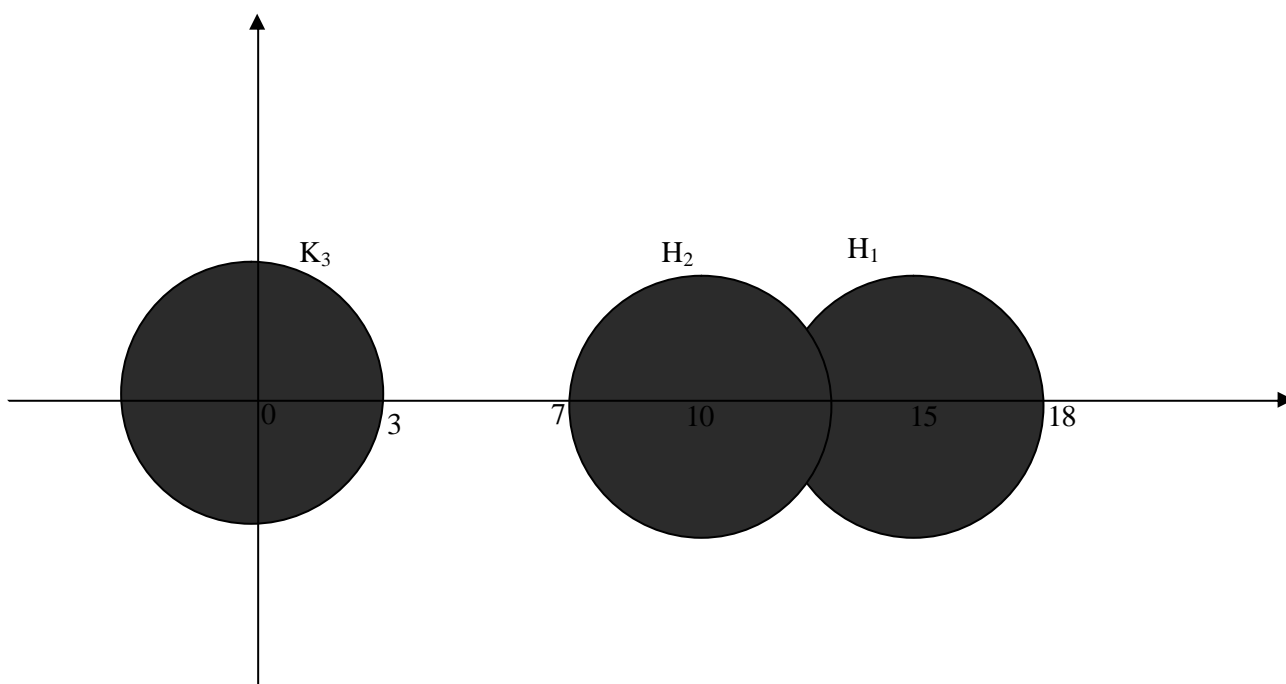
$$H_2 = \{z \in C : |z - 10| \leq 3\}$$

$$H_3 = \{z \in C : |z| \leq 5\}$$

e rappresentati nella figura sotto



Quindi gli autovalori di A stanno nell'intersezione dei due insiemi di cerchi, ovvero



Per il secondo teorema di Gerschgorin si può affermare che un auto valore di A è contenuto in K_3 mentre gli altri due appartengono ad $H_1 \cup H_2$.

4. Trasformazioni per similitudine

Poiché abbiamo visto che tutti gli autovalori di una matrice sono le radici del polinomio caratteristico considerate con la loro molteplicità, si potrebbe pensare che il calcolo degli autovalori di una matrice sia riconducibile al calcolo delle radici del suo polinomio caratteristico. Questa strada però non è la migliore dal punto di vista dell'Analisi Numerica; infatti il calcolo delle radici di un polinomio è tipicamente un problema mal condizionato, in cui a piccole variazioni sui coefficienti del polinomio corrispondono forti variazioni delle radici.

Un esempio tipico è dato dal polinomio $p(x)=(x-1)(x-2)\dots(x-20)=x^{20}-210x^{19}+\dots$

Le radici di questo polinomio sono $1,2,\dots,20$, quindi ben distinte tra loro. Consideriamo di calcolarle con un calcolatore che lavora in base 2 e con $t=30$ cifre per la mantissa. Per vedere come le radici dipendono fortemente dai coefficienti modifichiamo la trentesima cifra binaria relativa al valore -210 ottenendo $-210+2^{-23}$.

Eseguendo i conti con $t=90$ si ottiene che i primi quattro valori sono esatti, poi cominciano a peggiorare fino ad arrivare a 8.91725 anziché 9 e per i valori da 10 a 19 si ottengono addirittura dei numeri complessi.

Nel calcolo degli autovalori come radici del polinomio caratteristico, i coefficienti di quest'ultimo sono il risultato dei calcoli necessari allo sviluppo del determinante $\det(A - \lambda I)$, quindi sono necessariamente dei valori affetti da errori dovuti al fatto che operiamo con numeri finiti. Data la sensibilità delle radici a piccole variazioni dei coefficienti del polinomio, il calcolo degli autovalori in questo modo **non risulta affidabile**.

Si preferisce operare sulla matrice effettuando delle **trasformazioni che ne lasciano invariato lo spettro**, cioè gli autovalori, ma che la trasformino in una matrice per la quale il calcolo degli autovalori sia immediato, ad esempio matrici diagonali o matrici triangolari.

Vediamo quindi come si caratterizzano queste trasformazioni che lasciano invariato lo spettro.

Def: Matrici Simili

Siano A e B due matrici quadrate dello stesso ordine, si dice che B è simile ad A, o che B è ottenuta da A mediante una **trasformazione di similitudine** se esiste una matrice quadrata non singolare T tale che

$$B = T^{-1}AT$$

Si osserva che la similitudine tra matrici è una relazione di equivalenza, cioè

- A è simile ad A
- Se A è simile a B \Rightarrow B è simile ad A
- Se A è simile a B e B è simile a C \Rightarrow A è simile a C

Proprietà di matrici simili:

- Due matrici simili hanno lo stesso spettro, cioè gli stessi autovalori. Inoltre hanno gli autovettori legati tra loro dalla matrice di similitudine T. Infatti è facile verificare che se (λ, x) è una coppia autovalore-autovettore di A allora $(\lambda, T^{-1}x)$ lo è di $B=T^{-1}AT$.

Se λ è autovalore di A ed x è il relativo autovettore vale

$$Ax = \lambda x$$

Sia ora $y = T^{-1}x$ allora si ha

$$By = T^{-1}ATT^{-1}x = T^{-1}Ax = T^{-1}\lambda x = \lambda y$$

da cui

$$By = \lambda y$$

cioè A e B hanno gli stessi autovalori e autovettori legati dalla matrice di similitudine.

- Due matrici simili A e B hanno lo stesso polinomio caratteristico, cioè

$$P_A(\lambda) = P_B(\lambda)$$

Infatti si ha

$$\begin{aligned} P_B(I) &= \det(B - II) = \det[(T^{-1}AT - IT^{-1}T)] = \det[(T^{-1}(A - II)T)] \\ &= \det(T^{-1}) \det(A - II) \det(T) = \det(A - II) \det(T^{-1}T) \\ &= \det(A - II) = P_A(I) \end{aligned}$$

- Se A e B sono due matrici simili, allora la molteplicità algebrica $\sigma(\lambda)$ e la molteplicità geometrica degli autovalori $\mu(\lambda)$ rimane invariata.

Infatti l'invarianza della molteplicità algebrica segue dall'invarianza del polinomio caratteristico; mentre l'invarianza della molteplicità geometrica segue dal fatto che, essendo T matrice non singolare, i vettori x_1, x_2, \dots, x_μ sono linearmente indipendenti se e solo se lo sono i vettori $y_i = T^{-1}x_i$ $i=1, 2, \dots, \mu$.

Le trasformazioni di similitudine sono state introdotte perché nei più importanti metodi per il calcolo degli autovalori e autovettori di una matrice si eseguono una serie di trasformazioni di similitudine del tipo

$$A^{(i)} = T_i^{-1} A^{(i-1)} T_i \quad i = 1, 2, \dots \quad e \quad A^{(0)} = A$$

allo scopo di trasformare progressivamente la matrice A in una matrice simile ma il cui calcolo degli autovalori sia immediato o comunque facilmente eseguibile.

Alla base di questi algoritmi ci sono degli importanti risultati teorici che garantiscono che la matrice A può essere trasformata per similitudine in forme cosiddette "canoniche" in cui gli auto valori si calcolano facilmente.

1° Teorema di Schur

Sia A una matrice quadrata di ordine n, con elementi reali, allora esiste una trasformazione ortogonale U tale che trasforma A, per similitudine, in una B matrice triangolare superiore B, cioè

$$B = U^T A U = U^{-1} A U .$$

Poichè B è triangolare e simile ad A, gli autovalori di A sono gli elementi diagonali di B.

Il teorema di Schur si particolarizza quando anziché avere A una matrice qualunque si ha una matrice simmetrica, cioè quando $A = A^T$.

2° Teorema di Shur:

Sia A una matrice simmetrica di ordine n, cioè tale che $A = A^T$. Allora essa ha n autovalori reali $\lambda_1, \lambda_2, \dots, \lambda_n$, non necessariamente tutti distinti, ed n autovettori corrispondenti u_1, u_2, \dots, u_n , che formano una base ortogonale per R^n .

Indicando con U la matrice che ha u_1, u_2, \dots, u_n come colonne si ha

$$U^T A U = D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$$

con D matrice diagonale.

Esempio 4: Sia

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

Per calcolare gli autovalori di A consideriamo

$$\begin{aligned} P_A(\lambda) &= \det(A - \lambda I) = \det \begin{bmatrix} 2-\lambda & 1 & 0 \\ 1 & 3-\lambda & 1 \\ 0 & 1 & 2-\lambda \end{bmatrix} = \\ &= (2-\lambda) \begin{vmatrix} 3-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} - 1 \begin{vmatrix} 1 & 1 \\ 0 & 2-\lambda \end{vmatrix} + 0 \begin{vmatrix} 1 & 3-\lambda \\ 0 & 2-\lambda \end{vmatrix} = (2-\lambda)(\lambda^2 - 5\lambda + 4) = (2-\lambda)(4-\lambda)(1-\lambda) \end{aligned}$$

Quindi $P_A(\lambda) = 0 \Rightarrow \lambda_1 = 1 \quad \lambda_2 = 2 \quad \lambda_3 = 4$

Per calcolare un autovettore associato all'autovalore λ_1 si considera una soluzione del sistema $(A - \lambda_1 I) x = 0$, cioè

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Una soluzione è il vettore $x_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$.

In maniera del tutto analoga è possibile determinare gli auto vettori corrispondenti agli autovalori λ_2 e λ_3 , ottenendo

$$x_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, x_3 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}.$$

Se si valutano i prodotti scalari $\langle x_1, x_2 \rangle$, $\langle x_1, x_3 \rangle$, $\langle x_2, x_3 \rangle$ si può verificare che sono nulli, cioè sono auto vettori mutualmente ortogonali cioè linearmente indipendenti poiché relativi ad autovalori distinti (a conferma del teorema).

Prima di dare il teorema che caratterizza la forma canonica più importante, introduciamo la seguente matrice $v \times v$

$$J_n(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & \vdots \\ \vdots & \ddots & \ddots & 1 & 0 \\ \vdots & & \ddots & \lambda & 1 \\ 0 & \dots & \dots & 0 & \lambda \end{bmatrix} \quad \text{Blocco di Jordan} \quad (3)$$

Questa matrice ha un unico autovalore λ con molteplicità algebrica v e molteplicità geometrica 1.

Forma canonica di Jordan

Sia A una qualunque matrice quadrata di ordine n non singolare e $\lambda_1, \lambda_2, \dots, \lambda_k$, siano i suoi autovalori distinti con molteplicità $s(I_i)$ e $m(I_i)$ rispettivamente, $i=1, \dots, k$.

Per ogni autovalore λ_i esistono $m(I_i)$ numeri naturali distinti $n_j^{(i)}, j=1, 2, \dots, m(I_i)$, tali che

$$s(I_i) = n_1^{(i)} + n_2^{(i)} + \dots + n_{m(I_i)}^{(i)}$$

ed una matrice non singolare H che trasforma A in una matrice diagonale a blocchi J tale che

$$J = H^{-1}AH = \text{diag}(J_{s(I_1)}(I_1), \dots, J_{s(I_k)}(I_k))$$

nota come forma **canonica di Jordan**, dove I_j è autovalore di A e $J_{s(I_j)}(I_j)$ ha dimensione $s(I_j) \times s(I_j)$.

Il teorema afferma che, in sostanza, una matrice qualsiasi può essere trasformata per similitudine in una matrice diagonale a blocchi; il numero e la dimensione di questi blocchi sono legati alla molteplicità geometrica e algebrica degli autovalori della matrice. Infatti per ciascun autovalore abbiamo tanti blocchi di Jordan quanti sono gli autovalori linearmente indipendenti corrispondenti ad esso e la dimensione totale di questi blocchi è esattamente uguale alla molteplicità algebrica dell'autovalore stesso. Se la molteplicità algebrica degli autovalori è uguale alla molteplicità geometrica la forma canonica di Jordan è diagonale. In particolare si ha che ogni matrice A $n \times n$ con n autovalori distinti è diagonalizzabile mediante trasformazioni di similitudine, cioè

$$H^{-1}AH = D = \text{diag}(I_1, \dots, I_n)$$

e la matrice H ha come colonne gli auto vettori di A .

Poiché H è invertibile segue che

Autovettori corrispondenti ad autovalori distinti sono linearmente indipendenti.

5. Analisi di Stabilità e Condizionamento

Si vuole ora studiare il condizionamento del problema del calcolo degli autovalori di una matrice, cioè si analizza la variazione indotta sugli autovalori da una perturbazione degli elementi della matrice.

Sia $A \in \mathbb{C}^{n \times n}$ una matrice diagonalizzabile e sia H la matrice dei suoi autovettori.

Indichiamo con $\mathbf{d}A \in \mathbb{C}^{n \times n}$ una matrice perturbazione di A .

Vale il seguente teorema:

Teorema (Bauer-Fike)

Sia \mathbf{m} un autovalore della matrice $A + \mathbf{d}A$, allora esiste almeno un autovalore \mathbf{l} di A tale che

$$|\mathbf{l} - \mathbf{m}| \leq K_p(H) \|\mathbf{d}A\|_p$$

dove $\|\cdot\|_p$ è una qualunque norma p di matrice e $K_p(H) = \|H\|_p \|H^{-1}\|_p$ è detto indice di condizionamento spettrale di A .

Il teorema ci dice che perturbando gli elementi di una matrice A , gli autovalori cambiano proporzionalmente all'entità della perturbazione $\mathbf{d}A$. Il condizionamento del problema è legato all'indice di condizionamento della matrice H degli autovettori. Un'importante conseguenza di questo teorema è che se A è una matrice simmetrica i suoi auto vettori sono ortogonali e quindi $K(H)=1$, perciò il problema del calcolo degli autovalori per matrici simmetriche è ben condizionato per tutti gli autovalori.

Esempio5: Si considerino le matrici

$$A = \begin{bmatrix} 101 & -90 \\ 110 & -98 \end{bmatrix} \quad \text{e} \quad B = \begin{bmatrix} -1 & -1 \\ 0 & 0 \end{bmatrix}$$

$$\text{e sia } A(\mathbf{e}) = A + \mathbf{e}B = \begin{bmatrix} 101 - \mathbf{e} & -90 - \mathbf{e} \\ 110 & -98 \end{bmatrix}$$

gli autovalori di A sono $\lambda_1=1$ e $\lambda_2=2$; se prendiamo $\varepsilon=0.001$ si ottiene

$$A(\mathbf{e}) = \begin{bmatrix} 100.999 & -90.001 \\ 110 & -98 \end{bmatrix}$$

e i suoi auto valori sono $\lambda_1(\varepsilon)=1.298$ e $\lambda_2(\varepsilon)=1.701$.

Questo significa che ad una variazione dello 0.001% sui coefficienti di A corrisponde una variazione del 30% sugli autovalori di A.

In questo esempio la matrice H tale che $H^{-1}AH = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ è

$$H = \begin{bmatrix} 9 & -10 \\ 10 & -11 \end{bmatrix} \text{ e } H^{-1} = \begin{bmatrix} -11 & 10 \\ -10 & 9 \end{bmatrix}$$

con $K(H)=441$.

6. Metodi Numerici per il calcolo degli autovalori di una matrice

Ora vogliamo analizzare i metodi numerici per calcolare gli autovettori di una matrice. Fra questi alcuni sono convenientemente applicabili a matrici dense altri invece utilizzano proprietà o sparsità della matrice permettendo di trattare problemi di dimensioni elevate. Vi sono poi metodi che permettono di calcolare tutto lo spettro di A altri invece che permettono di calcolare solo alcuni autovalori.

Osservazione:

$$\text{Da } Ax = \lambda x \Rightarrow x^T Ax = \lambda x^T x$$

$$\Rightarrow \frac{x^T Ax}{x^T x} = \lambda = \frac{x^T Ax}{\|x\|_2^2}$$

Il metodo delle potenze

Il metodo delle potenze è un metodo iterativo per determinare l'autovalore di modulo massimo λ_1 e quello di modulo minimo λ_n di una matrice e i relativi autovettori.

La soluzione di tale problema risulta di grande interesse in numerosi problemi reali legati alla geosismica, alla statistica, allo studio delle reti elettriche,...

Calcolo di λ_1

Sia $A \in \mathbb{R}^{n \times n}$ diagonalizzabile e sia $X \in \mathbb{C}^{n \times n}$ la matrice le cui colonne sono gli autovettori x_i di A , per $i=1, \dots, n$ e siano gli autovalori di A ordinati in modo decrescente, cioè

$$|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_n|$$

con λ_1 di molteplicità algebrica 1, cioè λ_1 autovalore dominante di A .

Sia assegnato un vettore $q^{(0)}$ iniziale, arbitrario, di norma unitaria. Il metodo delle potenze considera le seguenti successioni:

Per $k=1, 2, \dots$

$$\begin{aligned} z^{(k)} &= Aq^{(k-1)} \\ q^{(k)} &= \frac{z^{(k)}}{\|z^{(k)}\|_2} \\ I^{(k)} &= (q^{(k)})^* Aq^{(k)} \end{aligned}$$

dove la $*$ indica il vettore trasporto coniugato.

Si vede facilmente che $q^{(k)}$ tende a disporsi nella direzione dell'autovettore x_1 .

Analizziamo le proprietà di convergenza del metodo.

Procediamo per induzione sul passo k . Si ha

$$q^{(k)} = \frac{Aq^{(k-1)}}{\|Aq^{(k-1)}\|_2} = \frac{AAq^{(k-2)}}{\|AAq^{(k-2)}\|_2} \cdot \frac{\|Aq^{(k-2)}\|_2}{\|Aq^{(k-2)}\|_2} = \frac{A^2q^{(k-2)}}{\|A^2q^{(k-2)}\|_2} = \dots = \frac{A^k q^{(0)}}{\|A^k q^{(0)}\|_2} \quad k \geq 1$$

Poichè A è diagonalizzabile i suoi autovettori x_i formano una base per \mathbb{C}^n .

Quindi $q^{(0)}$ può essere espresso come una combinazione lineare degli autovettori x_i , $i=1,\dots,n$, cioè

$$q^{(0)} = \sum_{i=1}^n \mathbf{a}_i x_i \quad \mathbf{a}_i \in C \quad i = 1, \dots, n.$$

Essendo $Ax_i = \lambda_i x_i$ si ha

$$A^k q^{(0)} = \sum_{i=1}^n A^k \mathbf{a}_i x_i = \sum_{i=1}^n \mathbf{a}_i A^k x_i = \sum_{i=1}^n \mathbf{a}_i \mathbf{I}_i^k x_i = \mathbf{a}_1 \mathbf{I}_1^k \left(x_1 + \sum_{i=2}^n \frac{\mathbf{a}_i}{\mathbf{a}_1} \left(\frac{\mathbf{I}_i}{\mathbf{I}_1} \right)^k x_i \right) \quad k = 1, \dots$$

Poiché λ_1 è autovalore dominante di A si ha che $\frac{\mathbf{I}_i}{\mathbf{I}_1} < 1$ e quindi

$$\left(\frac{\mathbf{I}_i}{\mathbf{I}_1} \right)^k \xrightarrow{k \rightarrow \infty} 0.$$

Ne segue che

$$q^{(k)} = \frac{A^k q^{(0)}}{\|A^k q^{(0)}\|_2} \xrightarrow{k \rightarrow \infty} \pm \frac{x_1}{\|x_1\|_2}.$$

Da questa relazione segue che la quantità $\mathbf{I}_1 = q^{(k)*} A q^{(k)}$ tende a λ_1 al tendere di k a ∞ .

e la convergenza è tanto più veloce quanto più è piccolo $\left| \frac{\mathbf{I}_2}{\mathbf{I}_1} \right|$.

Poiché lavoriamo con un metodo iterativo occorre valutare il criterio di arresto dell'algoritmo. Si definisce residuo al passo k :

$$r^{(k)} = A q^{(k)} - \lambda^{(k)} q^{(k)}$$

Fissata una tolleranza ε come condizione d'arresto del metodo si usa $|r^{(k)}| < \varepsilon$.

Calcolo di \mathbf{I}_n (Metodo delle Potenze Inverse)

Il metodo delle potenze fornisce un'approssimazione dell'autovalore di modulo di A minimo se si verifica che $|\mathbf{I}_n| < |\mathbf{I}_{n-1}|$ e se viene applicato ad A^{-1} . Infatti l'autovalore di modulo minimo di A risulta l'autovalore di modulo massimo di A^{-1} essendo gli autovalori di A^{-1} i reciproci degli auto valori di A .

Per $k=1,2,\dots$

$$z^{(k)} = A^{-1} q^{(k-1)}$$

$$q^{(k)} = \frac{z^{(k)}}{\|z^{(k)}\|_2}$$

$$I^{(k)} = (q^{(k)})^H A^{-1} q^{(k)}$$

Oss. Per non calcolare la matrice inversa A^{-1} si risolve, ad ogni passo, il sistema un sistema lineare $Az^{(k)} = q^{(k-1)}$, in cui la matrice A viene prima fattorizzata.

In generale, il metodo delle potenze può essere applicato alla matrice $(A - mI)$ che ha come auto valori $\frac{1}{(I_1 - m)}$ che converge all'autovalore di A più vicino a m

Il metodo di Jacobi per matrici simmetriche

Il metodo di Jacobi nasce per determinare tutti gli autovalori di matrici Hermitiane o simmetriche. Riconducendoci solo al caso reale, se A è simmetrica $A=A^T$ allora oltre al metodo QR è possibile utilizzare un metodo più semplice, iterativo che si basa sulla costruzione di una successione di matrici ortogonali allo scopo di diagonalizzare A mediante trasformazioni di similitudine.

In particolare il metodo di Jacobi raggiunge in modo iterativo il risultato del teorema di Shur cioè di portare la matrice A simmetrica e definita positiva in forma diagonale mediante trasformazioni di similitudine ortogonali.

Il metodo di Jacobi consiste nell'operare successive trasformazioni di similitudine mediante matrici ortogonali di Givens $G_{p,q}$, che corrispondono a rotazioni piane, generando una successione di matrici $A^{(k)}$ in cui ad ogni passo viene eliminato un elemento fuori diagonale e il suo simmetrico. Questa successione di matrici si

$$a_{p,q}^{(k+1)} = a_{q,p}^{(k+1)} = 0.$$

Per scegliere l'angolo \mathbf{q} che permette di ottenere questo risultato si osserva che gli elementi di A che non appartengono alle righe p e q vengono lasciati inalterati, cioè

$$a_{i,j}^{(k+1)} = a_{i,j}^{(k)} \quad \forall i, j \neq p, q,$$

mentre vengono modificati gli elementi delle righe e colonne p -esima e q -esima. In particolare, per vedere come si modifica l'elemento che vogliamo azzerare consideriamo

$$\begin{aligned} \begin{bmatrix} a_{pp}^{(k+1)} & a_{pq}^{(k+1)} \\ a_{qp}^{(k+1)} & a_{qq}^{(k+1)} \end{bmatrix} &= \begin{bmatrix} \cos \mathbf{q} & -\sin \mathbf{q} \\ \sin \mathbf{q} & \cos \mathbf{q} \end{bmatrix} \begin{bmatrix} a_{pp}^{(k)} & a_{pq}^{(k)} \\ a_{qp}^{(k)} & a_{qq}^{(k)} \end{bmatrix} \begin{bmatrix} \cos \mathbf{q} & \sin \mathbf{q} \\ -\sin \mathbf{q} & \cos \mathbf{q} \end{bmatrix} = \\ &= \begin{bmatrix} a_{pp}^{(k)} \cos \mathbf{q} - a_{qp}^{(k)} \sin \mathbf{q} & a_{pq}^{(k)} \cos \mathbf{q} - a_{qq}^{(k)} \sin \mathbf{q} \\ a_{pp}^{(k)} \sin \mathbf{q} + a_{qp}^{(k)} \cos \mathbf{q} & a_{pq}^{(k)} \sin \mathbf{q} + a_{qq}^{(k)} \cos \mathbf{q} \end{bmatrix} \begin{bmatrix} \cos \mathbf{q} & \sin \mathbf{q} \\ -\sin \mathbf{q} & \cos \mathbf{q} \end{bmatrix} = \\ &= \begin{bmatrix} a_{pp}^{(k)} \cos^2 \mathbf{q} - a_{qp}^{(k)} \sin \mathbf{q} \cos \mathbf{q} - a_{pq}^{(k)} \sin \mathbf{q} \cos \mathbf{q} + a_{qq}^{(k)} \sin^2 \mathbf{q} & a_{pp}^{(k)} \sin \mathbf{q} \cos \mathbf{q} - a_{qp}^{(k)} \sin^2 \mathbf{q} + a_{pq}^{(k)} \cos^2 \mathbf{q} - a_{qq}^{(k)} \sin \mathbf{q} \cos \mathbf{q} \\ a_{pp}^{(k)} \sin \mathbf{q} \cos \mathbf{q} + a_{qp}^{(k)} \cos^2 \mathbf{q} - a_{pq}^{(k)} \sin^2 \mathbf{q} - a_{qq}^{(k)} \sin \mathbf{q} \cos \mathbf{q} & a_{pp}^{(k)} \sin^2 \mathbf{q} + a_{qp}^{(k)} \sin \mathbf{q} \cos \mathbf{q} + a_{pq}^{(k)} \sin \mathbf{q} \cos \mathbf{q} + a_{qq}^{(k)} \cos^2 \mathbf{q} \end{bmatrix} \end{aligned}$$

Affinché $a_{p,q}^{(k+1)} = a_{q,p}^{(k+1)} = 0$, essendo $a_{q,p}^{(k)} = a_{p,q}^{(k)}$, si deve imporre che

$$-a_{p,q}^{(k)} \sin^2 \mathbf{q} + a_{p,q}^{(k)} \cos^2 \mathbf{q} + (a_{p,p}^{(k)} - a_{q,q}^{(k)}) \sin \mathbf{q} \cos \mathbf{q} = 0 \quad (4)$$

da cui si ricavano i valori di $\cos \mathbf{q}$ e $\sin \mathbf{q}$ per costruire la matrice $G_{pq}^{(k)}$.

Dalla (4) si ottiene

$$\begin{aligned} a_{p,q}^{(k)} \left(\frac{\cos^2 \mathbf{q} - \sin^2 \mathbf{q}}{\cos^2 \mathbf{q}} \right) + (a_{p,p}^{(k)} - a_{q,q}^{(k)}) \frac{\sin \mathbf{q} \cos \mathbf{q}}{\cos^2 \mathbf{q}} &= 0 \\ a_{p,q}^{(k)} (1 - \tan^2 \mathbf{q}) + (a_{p,p}^{(k)} - a_{q,q}^{(k)}) \tan \mathbf{q} &= 0 \end{aligned}$$

Se poniamo

$$t = \tan \mathbf{q} \quad e \quad m = \frac{a_{pp}^{(k)} - a_{qq}^{(k)}}{2a_{pq}^{(k)}} \quad \text{si ha}$$

$$1 - t^2 + 2mt = 0$$

e quindi

$$t = \begin{cases} m + \sqrt{1+m^2} & \text{se } m > 0 \\ m - \sqrt{1+m^2} & \text{se } m < 0 \end{cases}$$

da cui

$$\cos \mathbf{q} = \frac{1}{\sqrt{1+m^2}} \quad \sin \mathbf{q} = t \cos \mathbf{q}.$$

In questo modo si annulla ad ogni passo un elemento non diagonale (e il suo simmetrico). Questo non esclude che ai passi successivi un elemento non diagonale nullo possa essere modificato ed assumere valore non nullo. Tuttavia si può dimostrare che la somma dei moduli degli elementi extradiagonali tende a 0 cioè che la successione $A(k)$ generata dal metodo di Jacobi converge alla matrice diagonale

$$D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$$

$\lambda_1, \lambda_2, \dots, \lambda_n$ sono gli auto valori di A .

Questo suggerisce come criterio di arresto:

$$\sum |a_{i,j}^{(k+1)}| \leq \epsilon \quad \text{o} \quad \max_{i>j} |a_{i,j}^{(k+1)}| \leq \epsilon \quad \text{con } \epsilon > 0 \text{ tolleranza prefissata.}$$

Le varie formulazioni del metodo di Jacobi differiscono per la scelta di quale elemento $a_{p,q}$ annullare ad ogni passo.

1. Metodo di Jacobi Classico:

Ad ogni passo viene annullato l'elemento di modulo massimo extradiagonale

$$a_{p,q}^{(k)} = \max_{i \neq j} |a_{i,j}^{(k)}|$$

Questa scelta non è conveniente dal punto di vista computazionale poiché si aggiunge il costo dovuto alla determinazione del max.

2. Metodo di Jacobi Ciclico:

variante del metodo classico in cui si eliminano tutti gli elementi extradiagonali incontrati percorrendo per righe la matrice. L'operazione è ciclica perché può accadere che elementi annullati rinascano e quindi si ricominci.

3. Metodo di Jacobi con Soglia:

variante del metodo classico che consiste nel fissare un opportuno valore di soglia ed annullare per primi i valori maggiori della soglia.

Osservazione:

Il procedimento iterativo porta ad ottenere una matrice diagonale che ha per elementi gli autovalori di A

$$G^{(k)T} \cdot \dots \cdot G^{(2)T} \cdot G^{(1)T} \cdot A \cdot \underbrace{G^{(1)} G^{(2)} \cdot \dots \cdot G^{(k)}}_{H^{(k)}} = A^{(k+1)}$$

$$H^{(k)T} \cdot A \cdot H^{(k)} = A^{(k+1)}$$

$$A \cdot H^{(k)} = H^{(k)} A^{(k+1)}$$

$$\downarrow \qquad \downarrow$$

$$AH = HD$$

La matrice H invece è la matrice le cui colonne sono gli autovettori di A .

Il metodo basato sulla fattorizzazione QR

Il metodo QR consiste nel ridurre la matrice A , mediante opportune trasformazioni per similitudine, ad una forma per la quale il calcolo degli autovalori sia più agevole rispetto alla matrice di partenza.

Il metodo QR è un metodo iterativo per determinare la matrice unitaria U del teorema di Schur tale che $T=U^T A U$, con T triangolare superiore ed elementi diagonali pari agli autovalori di A .

Il metodo è abbastanza complicato anche se si basa su un semplice principio: riduzione preliminare di una matrice in forma triangolare o di Hessemberg superiore e si basa sulla fattorizzazione QR di una matrice.

Algoritmo base:

Nel metodo QR per il calcolo degli auto valori si genera una successione $\{A_k\}$ di matrici come segue:

- $A_1=A$
- Per $k=1,2,\dots$ si calcola la fattorizzazione QR della matrice A_k

$$A_k=Q_k R_k$$

dove Q_k matrice unitaria, R_k matrice triangolare superiore

- Si pone

$$A_{k+1}=R_k Q_k .$$

Per come definite le matrici della successione sono tutte simili tra loro; infatti si ha

$$A_{k+1}=Q_k^T A_k Q_k ,$$

e quindi

$$A_{k+1}=Q_k^T Q_{k-1}^T A_{k-1} Q_{k-1} Q_k =Q_k^T \dots Q_1^T A_1 Q_1 \dots Q_k = (Q_1 \dots Q_k)^T A_1 Q_1 \dots Q_k .$$

Ogni matrice A_k è ortogonalmente simile ad A . Si dimostra che la successione converge (*essenzialmente*) ad una matrice triangolare superiore con piccoli blocchi sulla diagonale.

Nel caso più semplice di autovalori semplici e separati in valore assoluto

$$|I_1| > |I_2| > \dots > |I_n|$$

Si dimostra che le matrici $Q_i \rightarrow I$ e le matrici A_i tendono ad una matrice triangolare superiore con λ_i sulla diagonale.

In particolare si dimostra che la convergenza a 0 degli elementi sotto diagonali $a_{i,j}$ è lineare ed è $O(\frac{I_i}{I_j})$ $i > j$, cioè tanto migliore quanto più gli auto valori di A sono separati in valore assoluto.

Costo Computazionale dell' algoritmo

Il metodo QR applicato ad una matrice di ordine n ha ad ogni passo un costo computazionale dell'ordine di n^3 operazioni moltiplicative. Per abbassare il costo computazionale globale conviene prima trasformare la matrice A in forma di Hessenberg superiore. Questa trasformazione iniziale viene applicata una sola volta perché poi la fattorizzazione QR mantiene la forma.

Def:

Una matrice A si dice **in forma di Hessenberg** superiore se

$$A(i,j) = 0 \text{ per ogni } i > j+1.$$

Il metodo QR applicato ad una matrice in forma di Hessenberg superiore ha ad ogni passo un costo computazionale di $2n^2$ operazioni moltiplicative.

Data una matrice A è possibile trasformarla per similitudine in forma di Hessenberg superiore con un costo computazionale di n^3 . L'algoritmo richiede complessivamente $n-2$ passi e la trasformazione per similitudine Q può essere calcolata come prodotto di matrici elementari di Householder.

Matrici Rettangolari

Nel caso di matrici rettangolari non si possono calcolare gli autovalori ma sono state introdotte altre grandezze note come valori singolari di una matrice.

Decomposizione in Valori Singolari :SVD

Tramite pre e post moltiplicazioni per matrici unitarie è possibile ridurre una matrice in forma diagonale.

La decomposizione in valori singolari di una matrice si basa sul seguente teorema:

Teorema:

Sia $A \in \mathbb{R}^{m \times n}$. Allora esistono due matrici unitarie $U \in \mathbb{R}^{m \times m}$ e $V \in \mathbb{R}^{n \times n}$ tali che

$$U^T A V = \Sigma = \text{diag}(\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_p) \quad (*)$$

dove $\Sigma \in \mathbb{R}^{m \times n}$ con $p = \min(m, n)$

$$\text{e } \mathbf{s}_1 \geq \mathbf{s}_2 \geq \dots \geq \mathbf{s}_p \geq 0.$$

La (*) è detta decomposizione in valori singolari (o Singular Value Decomposition) di A e i σ_i sono detti valori singolari di A, mentre le colonne di U e V sono dette vettori singolari sinistri e destri rispettivamente.

I valori singolari di una matrice hanno le seguenti proprietà:

- σ_i sono sempre reali e ≥ 0 ;
- il rapporto σ_1/σ_n ci fornisce l'indice di condizionamento della matrice rettangolare A;
- il numero di valori singolari non nulli rappresenta il rango della matrice A;
- inoltre

$$\mathbf{s}_i(A) = \sqrt{\lambda_i(A^T A)} \quad i = 1, \dots, n.$$

Infatti si ha che se $A = U\Sigma V^T$ allora $A^T = V\Sigma U^T$ e quindi essendo U e V unitarie si ha $A^T A = V\Sigma U^T U\Sigma V^T = V\Sigma^2 V^T$, cioè $A^T A$ è ortogonalmente simile alla matrice diagonale Σ^2 che quindi contiene i suoi autovalori.

Da ciò segue che

$$I_i(A^T A) = \mathbf{s}_i^2, i = 1, \dots, n.$$

cioè

$$\mathbf{s}_i = \sqrt{I_i(A^T A)}, i = 1, \dots, n.$$

Utilizzando i valori singolari e i corrispondenti vettori singolari si può ottenere la decomposizione spettrale di A nella forma

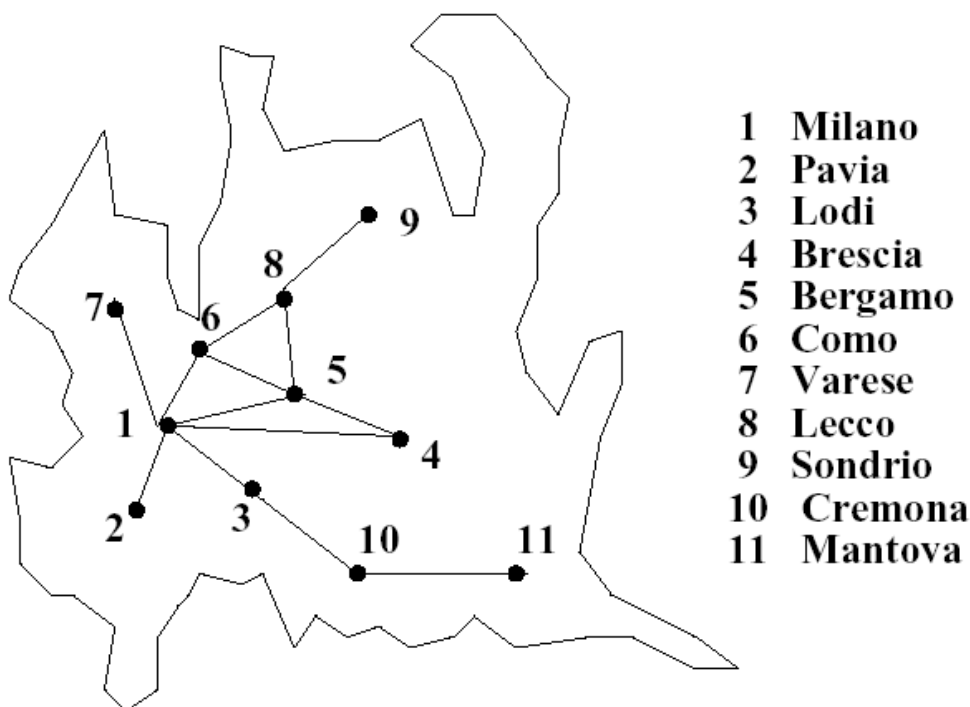
$$A = \sum_{j=1}^n \mathbf{s}_j u_j v_j^T .$$

Problema

Supponiamo che n città siano tra loro variamente collegate ed associamo ad esse una matrice quadrata A di dimensione n il cui generico elemento $a(i,j)$ è così definito

$$a_{i,j} = \begin{cases} 1 & \text{se la città } i \text{ è collegata alla città } j \\ 0 & \text{altrimenti} \end{cases}$$

Siamo interessati a determinare la facilità di accesso alle varie città. Si può dimostrare che i moduli delle componenti dell'autovettore (di norma unitaria) associato all'autovalore di modulo massimo sono una misura della facilità di accesso alle varie città, più significativa del contare semplicemente il numero di accessi possibile per ciascuna città. Consideriamo la rete ferroviaria collegante le 11 città capoluogo della Lombardia. Se associamo ad ogni città un numero possiamo facilmente generare la matrice A delle connessioni.



$$A = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

L' autovettore di norma unitaria associato all'autovalore di modulo massimo ha le seguenti componenti (in modulo)

0.5271	(Milano)	0.1590	(Pavia)
0.2165	(Lodi)	0.3579	(Brescia)
0.4690	(Bergamo)	0,3861	(Como)
0.1590	(Varese)	0.2837	(Lecco)
0.0856	(Sondrio)	0.1906	(Cremoma)
0.0575	(Mantova)		

Ordinandole in modo decrescente si deduce che subito dopo Milano, la città meglio collegata è Bergamo, seguita da Como e Brescia. Le peggio collegate sono Sondrio e Mantova. Si noti Pavia, Varese, Mantova e Sondrio presentano lo stesso numero di accessi, ma caratteristiche di accessibilità diverse (si noti che questa analisi non tiene conto della frequenza dei collegamenti, ma solo dell'esistenza o meno del collegamento).

Per calcolare tutti gli autovalori di una matrice ci serviamo della fattorizzazione QR di una matrice.

APPENDICE

Fattorizzazione QR di una matrice

La fattorizzazione QR di una matrice A consiste nel trasformare la matrice A in una matrice triangolare superiore mediante l'utilizzo di trasformazioni ortogonali che modificano la matrice lasciando inalterato lo spettro. In particolare data $A \in \mathbb{R}^{n \times n}$ si vuole determinare una matrice $Q \in \mathbb{R}^{n \times n}$ ortogonale tale che

$$Q^T A = R$$

Per determinare la matrice Q ortogonale della fattorizzazione si considera il metodo di Householder che costruisce Q come prodotto di una successione P^K di matrici elementari di Householder, ovvero

$$\underbrace{P^n \dots P^2 P^1}_{Q^T} A = R$$

dove ogni matrice elementare di Householder P^K è una matrice simmetrica e ortogonale tale che:

$$P = I - \frac{2ww^T}{\|w\|_2^2}.$$

Quindi per costruire la matrice P occorre definire il vettore w .

Le matrici elementari di Householder possono essere utilizzate per annullare un blocco di componenti di un dato vettore $x \in \mathbb{R}^n$. Se in particolare si volessero annullare tutte le componenti di x , tranne le m -esima, bisognerebbe scegliere

$$w = x \pm \|x\|_2 e_m$$

essendo e_m l' m -esimo versore di \mathbb{R}^n , in modo tale che

$$P \cdot x = [0, 0, \dots, \pm \|x\|_2, 0, \dots, 0]^T.$$

Se invece si vuole costruire la matrice P in modo da lasciare inalterate le prime k componenti di x e annullare tutte le componenti dalla $k+2$, bisognerebbe scegliere P tale che

$$P = \begin{bmatrix} I_k & \\ & \tilde{P}_{n-k} \end{bmatrix}, \quad \tilde{P}_{n-k} = I_{n-k} - \frac{2ww^T}{\|w\|_2^2}$$

dove I_k è la matrice identità di ordine k mentre \tilde{P}_{n-k} è la matrice elementare di Householder di ordine $n-k$ il cui vettore w è definito come

$$w = x^{n-k} \pm \|x^{n-k}\|_2 e_1^{n-k}$$

essendo $x^{n-k} \in \mathbb{R}^{n-k}$ il vettore coincidente con le ultime $n-k$ componenti di x e e_1^{n-k} il primo versore della base canonica di \mathbb{R}^{n-k} .

Le componenti di $y = Px$ risultano quindi:

$$\begin{cases} y_j = x_j & j = 1, \dots, k \\ y_j = 0 & j = k + 2, \dots, n \\ y_{k+1} = \pm \|x^{n-k}\|_2^2 \end{cases}$$

Pertanto data una matrice $A \in \mathbb{R}^{n \times n}$ del tipo

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & \ddots & & \vdots \\ \vdots & & & \\ a_{n1} & \dots & & a_{nn} \end{bmatrix} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$$

Si determina P^1 in modo tale da annullare gli elementi della prima colonna di a (\mathbf{a}_1) dal secondo elemento in poi, cioè

$$P^1 A = \begin{bmatrix} \tilde{a}_{11} & a'_{12} & \dots & a'_{1n} \\ 0 & \ddots & & \vdots \\ \vdots & & & \\ 0 & a'_{n2} & \dots & a'_{nn} \end{bmatrix} \quad \text{con } \tilde{a}_{11} = \pm \|\mathbf{a}_1\|_2$$

Al passo 2 si agisce sulla seconda colonna di A annullando gli elementi dal terzo in poi, cioè

$$P^2 P^1 A = \begin{bmatrix} \tilde{a}_{11} & a'_{12} & \dots & a'_{1n} \\ 0 & \tilde{a}'_{22} & a''_{23} & a''_{2n} \\ \vdots & 0 & \vdots & \vdots \\ & \vdots & & \\ 0 & 0 & a''_{n3} & a''_{nn} \end{bmatrix} \text{ dove } P^2 = \begin{bmatrix} I_1 & \\ & \tilde{P}_{n-1} \end{bmatrix}, \quad \tilde{P}_{n-1} = I_1 - \frac{2ww^T}{\|w\|_2^2}$$

Dopo n trasformazioni elementari di Householder si ottiene

$$\underbrace{P^n \dots P^2 P^1}_{Q^T} A = R$$

Da cui ricordando che il prodotto di matrici ortogonali è ancora una matrice ortogonale, $Q^T = Q^{-1}$ e quindi

$$A = QR$$