

Capitolo 5

Autovalori ed autovettori (teoria)

5.1 Definizioni

Abbiamo visto che applicare una matrice ad un vettore può essere visto come una trasformazione, in certi casi conservando proprietà geometriche (ribaltamenti, rotazioni, omotetie, eccetera). Usando i vettori canonici si può per esempio studiare il comportamento della matrice (espansivo, contrattivo o conservativo) per determinate direzioni.

Definizione Sia $A \in \mathbb{R}^{n \times n}$, se $Ax = \lambda x$ allora λ si dice *autovalore* di A e x si dice *autovettore* di A ((λ, x) si dice *autocoppia* di A). Per le abbreviazioni si usano i termini tedeschi *eigenwert* (autovalore, ew) e *eigenvektor* (autovettore, ev).

Consideriamo $x = 0$, $Ax = 0 = \lambda x \forall \lambda$, non ha molto senso in questo caso, quindi alla definizione si aggiunge $x \neq 0$.

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

$$x = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \Rightarrow Ax = \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 3x : x \text{ è l'autovettore e } 3 \text{ l'autovalore}$$

$$x' = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \Rightarrow Ax' = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 1x' : x' \text{ è l'autovettore e } 1 \text{ l'autovalore}$$

$$A = \alpha I \Rightarrow x \rightarrow \alpha x \forall x \in \mathbb{R}^n: \text{ogni } x \text{ è autovettore e } \alpha \text{ l'autovalore.}$$

Autovettori diversi vengono associati allo stesso autovalore.

$$A = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} \\ 1 & 0 & 0 & 0 \end{pmatrix} \text{ Questa matrice è } \textit{stocastica per righe} \text{ (la somma sulle righe è 1 e tutti gli}$$

elementi appartengono all'intervallo $[0, 1]$

$$x = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \Rightarrow Ax = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = 1x \text{ (in generale, c'è sempre l'autovalore 1)}$$

Prendiamo la matrice della riflessione di Householder: tutti i vettori che giacciono sul piano perpendicolare a w vengono lasciati fermi e quindi sono autovettori della matrice P , l'autovalore vale

1. Invece $Pw = -w = (-1)w$.

$$A = uv^T \text{ con } u, v \in \mathbb{R}^n$$

$$Ax = \lambda x \quad \forall x \in \mathbb{R}^n$$

$$Ax = u(v^T x) = (v^T x)u \text{ questo deve equivalere a } \lambda x: (v^T x)u = \lambda x$$

Ci sono due possibili casi:

1. $\lambda = 0 = v^T x$: essendo nullo l'autovalore abbiamo anche indicazione che $x \perp v$ (prodotto scalare nullo), quindi tutti i vettori dell'iperpiano perpendicolare sono autovettori e appartengono all'autospazio relativo a $\lambda = 0$. L'autovalore ha molteplicità geometrica $n - 1$
2. $x = \alpha u \Rightarrow Ax = (v^T x)u = (v^T \alpha u)u = \alpha(v^T u)u$, $\lambda x = \lambda \alpha u \Rightarrow \lambda = v^T u$ autovalore di A con molteplicità geometrica 1

Caso particolare: $v^T u = 0 \Rightarrow \exists! ew = 0$ con molteplicità geometrica $n - 1$ (αu è già perpendicolare a v).

In generale, consideriamo una autocoppia di A : $cx = y$ è un altro autovettore di A , infatti $Ay = A(cx) = cAx = c\lambda x = \lambda cx = \lambda y$. Pertanto, gli autovettori sono (per lo meno) determinati a meno di una scalatura.

Sia (λ, x) una autocoppia di A , allora x è autovettore anche delle seguenti matrici A^2 , αA , A^k , A^{-1} (se $\det A \neq 0$), $A + cI$. Ad esempio:

$$Ax = \lambda x \Rightarrow A^2 x = A(Ax) = A\lambda x = \lambda Ax = \lambda \lambda x = \lambda^2 x$$

$$(\alpha A)x = \alpha Ax = \alpha \lambda x$$

Osservazione: A tale che il suo determinante sia nullo, allora $\exists v \in \mathfrak{N}(A) \Rightarrow Av = 0 \Rightarrow 0 \cdot v \Rightarrow v$ è un autovettore di A corrispondente all'autovalore 0

5.2 Calcolo

Ricavare l'autovalore dall'autovettore Si definisce il *quoziente di Rayleigh*: $\lambda = \frac{x^T Ax}{x^T x} =: r(x)$

$$x \text{ autovettore} \Rightarrow Ax = \lambda x \Rightarrow r(x) = \frac{x^T \lambda x}{x^T x} = \frac{\lambda x^T x}{x^T x}; \text{ se } x \neq 0 \Rightarrow \|x\|_2 \neq 0 \Rightarrow \|x\|_2^2 = x^T x$$

Ricavare l'autovettore dall'autovalore $A \in \mathbb{R}^{n \times n}$, λ autovalore di A , $Ax - \lambda x = 0$ la riscrivo come $Ax - \lambda Ix = 0$, ovvero $(A - \lambda I)x = 0$ sistema lineare omogeneo. $x \in \mathfrak{N}(A - \lambda I)$

Definisco $\mathfrak{N}(A - \lambda I)$ come l'insieme di tutti gli autovettori di A associati a λ più l'origine: è detto *autospazio* di A relativo a λ . $\dim \mathfrak{N}(A - \lambda I)$ prende il nome di *molteplicità geometrica* dell'autovalore λ .

Esempio $A = \alpha I \Rightarrow \exists! \alpha$ autovalore associato ad ogni $x \in \mathbb{R}^n \Rightarrow \mathfrak{N}(A - \alpha I) = \mathbb{R}^n$, α ha molteplicità geometrica n .

$P = I - 2ww^T$ l'autovalore $\lambda = 1$ ha molteplicità geometrica $n - 1$ poichè l'autospazio è l'iperpiano perpendicolare a w , l'autovalore $\lambda = -1$ corrisponde all'autovettore w che appartiene alla retta che coincide con l'autospazio.

Osservazione λ autovalore di $A \in \mathbb{R}^{n \times n} \Leftrightarrow \det(A - \lambda I) = 0$ (altrimenti il nucleo sarebbe banale)

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \rightarrow A - \lambda I = \begin{pmatrix} 2 - \lambda & 1 \\ 1 & 2 - \lambda \end{pmatrix}$$

$$\det(A - \lambda I) = (2 - \lambda)^2 - 1 = \lambda^2 - 4\lambda + 3 = 0 \rightarrow \lambda_1 = 3, \lambda_2 = 1,$$

autospazi: $\left\langle \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle, \left\langle \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\rangle.$

$$A = \begin{pmatrix} 4 & 1 \\ -1 & 2 \end{pmatrix} \rightarrow A - \lambda I = \begin{pmatrix} 4 - \lambda & 1 \\ -1 & 2 - \lambda \end{pmatrix}$$

$$\det(A - \lambda I) = (4 - \lambda)(2 - \lambda) + 1 = \lambda^2 - 6\lambda + 9 = 0 \rightarrow \lambda = 3 \text{ con molteplicità } 2 \text{ (nel senso che è radice doppia).}$$

Cerco $\mathfrak{N}(A - \lambda I) = \mathfrak{N}(A - 3I)$

x autovettore di $A \Leftrightarrow (A - 3I)x = 0$

$$\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ Deduciamo che tutti i vettori con i componenti opposti sono autovettori}$$

$$\left\{ \begin{pmatrix} -x_2 \\ x_2 \end{pmatrix} \right\} = \left\langle \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\rangle, \text{ molteplicità geometrica } 1.$$

Teorema $\det(A - \lambda I)$ è un polinomio in λ di grado n detto *polinomio caratteristico*

Corollario Le radici del polinomio caratteristico sono gli autovalori della matrice

Teorema fondamentale dell'algebra Ogni polinomio a coefficienti complessi di grado n ha n radici in \mathbb{C} contate con le loro molteplicità.

La *molteplicità algebrica* di un autovalore è quante volte la radice risolve un polinomio caratteristico.

Teorema molteplicità geometrica \leq molteplicità algebrica

Radici complesse Abbiamo visto che λ autovalore di $A \iff \det(A - \lambda I) = 0$, che l'autovettore relativo a λ si ricava ponendo $(A - \lambda I)x = 0$, e che la molteplicità geometrica è $\dim \mathfrak{N}(A - \lambda I)$.

C'è la possibilità di trovare radici complesse nonostante il polinomio caratteristico abbia tutti coefficienti reali.

$$\lambda \in \mathbb{R} \Rightarrow \exists x \in \mathbb{R}^n \text{ autovettore}$$

$\lambda \in \mathbb{C}$ autovalore di $A \Rightarrow \bar{\lambda}$ autovalore di A (poichè è la radice complessa coniugata) e $x \in \mathbb{C}^n$ (non reale) è l'autovettore relativo a λ

Possiamo coniugare tutti gli elementi del sistema complesso:

$(\bar{A} - \bar{\lambda}I)\bar{x} = 0 \Rightarrow (A - \bar{\lambda}I)\bar{x} = 0 \Rightarrow \bar{x} \in \mathfrak{N}(A - \bar{\lambda}I)$ è l'autospazio relativo a $\bar{\lambda}$: gli autovettori relativi a $\bar{\lambda}$ sono i coniugati degli autovettori relativi a λ .

$$G = \begin{pmatrix} c & -s \\ s & c \end{pmatrix} \text{ dove } c = \cos \theta \text{ e } s = \sin \theta$$

$$G - \lambda I = \begin{pmatrix} c - \lambda & -s \\ s & c - \lambda \end{pmatrix} \rightarrow \det(G - \lambda I) = (c - \lambda)^2 + s^2$$

$$\lambda = c \pm \sqrt{c^2 - 1} = c \pm is \rightarrow \lambda_1 = e^{i\theta}, \lambda_2 = e^{-i\theta}$$

$$\lambda = \cos \theta + i \sin \theta \Rightarrow G - \lambda I = \begin{pmatrix} c - \lambda & -s \\ s & c - \lambda \end{pmatrix} = \begin{pmatrix} -is & -s \\ s & -is \end{pmatrix}$$

$$\text{autovettore relativo a } \lambda \ x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ix_2 \\ x_2 \end{pmatrix} \text{ autospazio} = \left\langle \begin{pmatrix} i \\ 1 \end{pmatrix} \right\rangle$$

$\bar{\lambda} = \cos \theta - i \sin \theta$ ha autospazio $\left\langle \begin{pmatrix} -i \\ 1 \end{pmatrix} \right\rangle$

Osservazione Non esistono formule risolutive per equazioni di grado ≥ 5

Esercizio Sia $p(t) = c_0 + c_1 t + \dots + c_{n-1} t^{n-1} + t^n$, definisco la *matrice di Frobenius* $F = \begin{pmatrix} 0 & 1 & & & 0 \\ & 0 & \ddots & & \\ \vdots & & \ddots & \ddots & \\ 0 & & & 0 & 1 \\ c_0 & c_1 & \cdots & & c_{n-1} \end{pmatrix} \in$

$\mathbb{R}^{n \times n}$, si può dimostrare che $\det(F - \lambda I) = (-1)^n p(\lambda)$.

Con la matrice di Frobenius troviamo che i suoi autovalori sono le radici del polinomio da cui siamo partiti. Il problema di calcolare le radici di un polinomio è della stessa difficoltà di calcolare gli autovalori.

5.3 Matrici diagonali e triangolari

Se una matrice è diagonale, si può scrivere per brevità $A = \text{diag}(\alpha_1, \dots, \alpha_n)$

Proposizione: Se A è diagonale $\Rightarrow \forall k = 1, \dots, n \ A e_k = \alpha_k e_k$ ($\Rightarrow \alpha_1, \dots, \alpha_n$ sono autovalori di A e i vettori canonici sono possibili autovettori associati).

Dimostrazione: $A e_k = k\text{-esima colonna di } A = \begin{pmatrix} 0 \\ \vdots \\ \alpha_k \\ \vdots \\ 0 \end{pmatrix} = \alpha_k \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \alpha_k e_k$

Proposizione: sia A triangolare superiore allora $a_{11}, a_{22}, \dots, a_{nn}$ sono autovalori di A . Se a_{kk} è autovalore di A allora x è autovettore con k posizioni occupate e il resto zeri.

Dimostrazione: $\det(A - \lambda I) = \det \begin{pmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ & a_{22} - \lambda & & \\ \vdots & & \ddots & \\ 0 & & & a_{nn} - \lambda \end{pmatrix} = (a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda) =$

0 Le radici del polinomio caratteristico sono per forza a_{11}, \dots, a_{nn} .

$(A - a_{kk} I)x = 0$

$$(A - a_{kk} I)x = \begin{pmatrix} a_{11} - a_{kk} & a_{12} & \cdots & & & a_{1n} \\ & a_{22} - a_{kk} & & & & \\ & & \ddots & & & \\ & & & a_{k-1,k-1} - a_{kk} & & \\ & & & & 0 & \\ & & & & & a_{k+1,k+1} - a_{kk} \\ & & & & & \ddots \\ 0 & & & & & & a_{nn} - a_{kk} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Il vettore x è costituito da k componenti e poi zeri.

Il trucco è spezzare verticalmente il vettore delle componenti e la matrice orizzontalmente (il risultato si ottiene dalla somma dei prodotti delle parti). La parte a destra della matrice non pesa sul risultato perchè si moltiplica per zeri.

$$\begin{pmatrix} a_{11} - a_{kk} & a_{12} & \cdots & & \\ & a_{22} - a_{kk} & & & \\ & & \ddots & & \\ & & & a_{k-1,k-1} - a_{kk} & \\ & & & & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \Leftrightarrow B \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} = 0 \Leftrightarrow \mathfrak{N}(B) \neq \{0\} \Leftrightarrow \det B = 0$$

Questo risultato è corretto poichè la matrice ha un elemento nullo sulla diagonale.

5.3.1 Matrici diagonali a blocchi

$$\begin{pmatrix} \underbrace{A_1}_{k \times k} & 0 \\ 0 & \underbrace{A_2}_{(n-k) \times (n-k)} \end{pmatrix}$$

Se λ è autovalore di A , allora i casi sono due:

- λ è autovalore di $A_1 \Leftrightarrow A_1 y = \lambda y$
- λ è autovalore di $A_2 \Leftrightarrow A_2 z = \lambda z$

$$x = \begin{pmatrix} y \\ 0 \end{pmatrix} \in \mathbb{R}^n \Rightarrow Ax = \begin{pmatrix} A_1 y \\ 0 \end{pmatrix} = \begin{pmatrix} \lambda y \\ 0 \end{pmatrix} = \lambda x \text{ Analogamente per } A_2.$$

Il procedimento si può ripetere se anche A_1 o A_2 è diagonale a blocchi.

Esempio Matrice a blocchi

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

$$\text{blocco } A_2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}: \det \begin{pmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{pmatrix} = (1-\lambda)^2 - 1 = \lambda^2 - 2\lambda = 0, \lambda_1 = 0, \lambda_2 = 2$$

$$\lambda = 0 \rightarrow \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} y = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \text{ Autovettori di } A: \left\langle \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right\rangle$$

$$\lambda = 2 \rightarrow \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad y = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{Autovettori di } A: \left\langle \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right\rangle$$

$$\text{blocco } A_4 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}: \det \begin{pmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{pmatrix} = (2-\lambda)^2 - 1 = \lambda^2 - 4\lambda + 3 = 0, \lambda_1 = 1, \lambda_2 = 3$$

$$\lambda = 1 \rightarrow \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad y = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \text{Autovettori di } A: \left\langle \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} \right\rangle$$

$$\lambda = 3 \rightarrow \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad y = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{Autovettori di } A: \left\langle \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \right\rangle$$

Gli altri blocchi sono scalari e quindi coincidono con gli autovalori.

$$A_1 = \lambda = 3 \quad \text{Autovettori di } A: \left\langle \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right\rangle$$

$$A_3 = \lambda = 1 \quad \text{Autovettori di } A: \left\langle \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right\rangle$$

$$A_5 = \lambda = 1 \quad \text{Autovettori di } A: \left\langle \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\rangle$$

Mettendo insieme le informazioni, abbiamo:

$$\begin{aligned}
\lambda = 0 : \text{molteplicità algebrica 1, autospazio } & \left\langle \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right\rangle \text{ (molteplicità geometrica 1)} \\
\lambda = 2 : \text{molteplicità algebrica 1, autospazio } & \left\langle \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right\rangle \text{ (molteplicità geometrica 1)} \\
\lambda = 1 : \text{molteplicità algebrica 3, autospazio } & \left\langle \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\rangle \text{ (molteplicità geometrica 3)} \\
\lambda = 3 : \text{molteplicità algebrica 2, autospazio } & \left\langle \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right\rangle \text{ (molteplicità geometrica 2).}
\end{aligned}$$

Esempio Sia $A = \begin{pmatrix} A_1 & A_3 \\ 0 & A_2 \end{pmatrix}$ dimostrare che:

1. se (λ_1, y) è una autocoppia di $A_1 \Rightarrow \left(\lambda_1, \begin{pmatrix} y \\ 0 \end{pmatrix} \right)$ è un'autocoppia di A
2. se (λ_2, z) è una autocoppia di $A_2 \Rightarrow \lambda_2$ è un'autovalore di A ed inoltre x è autovettore relativo a λ_2 della forma $\begin{pmatrix} y \\ 0 \end{pmatrix}$ oppure $\begin{pmatrix} y \\ z \end{pmatrix}$.

5.4 Similitudine

L'eliminazione gaussiana non conserva autovalori ed autovettori, c'è una trasformazione che le conservi?

$$A : v \mapsto Av =: w$$

Una matrice diagonale ha le direzioni privilegiate corrispondenti ai vettori canonici. Per mantenere invariata l'azione della matrice, dobbiamo variare il sistema di riferimento.

Considerati due sistemi di riferimento ed un vettore, si può scrivere il vettore come combinazione lineare delle sue proiezioni sugli assi scelti, ad esempio $v = \sum_{i=1}^n v_i e_i$ e $v = \sum_{i=1}^n \alpha_i s_i$, consideriamo ad esempio $n = 2$

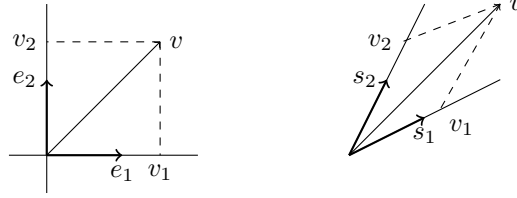


Figura 5.1: Sistemi di riferimento

Dati gli assi $s_1 \dots s_n \Rightarrow v = \sum_{i=1}^n \alpha_i s_i = S\alpha$ con $\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$ e $S = (s_1 \ s_2 \ \dots \ s_n)$.

Passare da una base canonica ad un'altra base significa risolvere il sistema lineare $S\alpha = v$ dove v è il termine noto (coordinate rispetto alla base canonica) e α le incognite (coordinate rispetto alla base trasformata).

Nella base canonica abbiamo $v \mapsto w = Av$, nella base S corrispondono rispettivamente a $S\alpha$ e $S\beta$. Ora notiamo che c'è un'uguaglianza e la sfruttiamo, essendo S invertibile (le sue colonne costituiscono una base): $\beta = S^{-1}Av = S^{-1}AS\alpha$.

Se nella base canonica A attua la trasformazione $v \rightarrow w$, nella base S la stessa trasformazione diventa $S^{-1}AS$, non variando l'azione geometrica e nemmeno autovalori ed autovettori.

5.4.1 Definizioni

A e B si dicono simili ($A \sim B$) se esiste S non singolare tale che $B = S^{-1}AS$.

Teorema Se $A \sim B \Rightarrow \det(A - \lambda I) = \det(B - \lambda I)$ e se x autovettore di $A \Rightarrow S^{-1}x$ autovettore di B

Con la similitudine gli autovalori si conservano e gli autovettori cambiano base.

Dimostrazione:

$$\begin{aligned} \det(B - \lambda I) &= \det(S^{-1}AS - \lambda I) = \det(S^{-1}AS - \lambda S^{-1}S) = \det[S^{-1}(A - \lambda I)S] \\ &= \det S^{-1} \cdot \det(A - \lambda I) \cdot \det S. \end{aligned}$$

I due determinanti ai lati si semplificano (per la regola di Binet) e si ottiene il risultato.

Essendo $Ax = \lambda x$ ($x \neq 0$): $B(S^{-1}x) = S^{-1}ASS^{-1}x = S^{-1}Ax = S^{-1}\lambda x = \lambda S^{-1}x$. Ci rimane da verificare se l'autovettore trasformato si può azzerare: $S^{-1}x = 0 \Rightarrow SS^{-1}x = 0 \Rightarrow x = 0$ assurdo quindi $S^{-1}x \neq 0$.

E' equivalente dire $B = S^{-1}AS \Leftrightarrow SB = AS \Leftrightarrow SBS^{-1} = A$

Esempio $\begin{pmatrix} a & b \\ b & a \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} a+b & 0 \\ 0 & a-b \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1} \Rightarrow \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} a+b & 0 \\ 0 & a-b \end{pmatrix} = \begin{pmatrix} a & b \\ b & a \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \rightarrow$

$\begin{pmatrix} a+b & a-b \\ a+b & b-a \end{pmatrix}$
 x autovettore di $A \Rightarrow S^{-1}x$ autovettore di $\begin{pmatrix} a+b & 0 \\ 0 & a-b \end{pmatrix} \Rightarrow S^{-1}x = e_k \quad k = 1..2 \Rightarrow x = Se_k \Rightarrow$
 $x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ e $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$

Esercizio Dimostrare nel caso delle matrici a blocchi che $\begin{pmatrix} A & B \\ B & A \end{pmatrix} = \begin{pmatrix} I & I \\ I & -I \end{pmatrix} \begin{pmatrix} A+B & 0 \\ 0 & A-B \end{pmatrix} \begin{pmatrix} I & I \\ I & -I \end{pmatrix}^{-1}$.

5.5 Diagonalizzabilità

Devo trovare S tale che $A = SDS^{-1}$ con D diagonale. E' sempre possibile farlo? La risposta è: no, e quando ci si riesce l'operazione è detta *diagonalizzazione*.

Teorema A diagonalizzabile se e solo se $A = X\Lambda X^{-1}$ con

1. $\Lambda = \text{diag}(\lambda_1 \dots \lambda_n)$ con λ_i autovalori di A
2. $X = (x_1 | \dots | x_n)$ con $Ax_i = \lambda_i x_i$ e $x_i \neq 0$
3. x_1, \dots, x_n linearmente indipendenti

Dimostrazione (\Rightarrow)

1. $A \sim \Lambda \Rightarrow$ autovalore di $A \equiv$ autovalore di Λ , Λ è diagonale e i suoi elementi sono autovalori (di A)
2. Riscrivo l'uguaglianza come $AX = X\Lambda$ e vado a leggere per colonne AXe_i (il primo membro) e il secondo membro $X\lambda_i e_i = \lambda_i Xe_i$. L'uguaglianza diventa $Ax_i = \lambda_i x_i$. Ipotizzando per assurdo $x_i = 0$ troviamo che $\det X = 0$ (non è invertibile) ma esiste X^{-1} che contraddice l'ipotesi
3. $\exists X^{-1} \iff \text{rk}(X) = n \Rightarrow X$ ha n colonne indipendenti $\Rightarrow x_1 \dots x_n$ indipendenti

(\Leftarrow)

$$\Lambda := \text{diag}(\lambda_1 \dots \lambda_n)$$

$$X := (x_1 | \dots | x_n)$$

$$\text{Hp: } x_1 \dots x_n \text{ indipendenti} \Rightarrow \text{rk}(X) = n \Rightarrow \det X \neq 0$$

$$\text{Th: } AX = X\Lambda \text{ (dimostrazione per colonne)}$$

Il primo membro è AXe_i , il secondo $X\lambda_i e_i = \lambda_i x_i$: le due matrici coincidono in ogni colonna.

Questo teorema non ha importanza pratica, infatti non dà criteri operativi per verificare la diagonalizzabilità anche se fornisce condizioni necessarie e sufficienti.

Teorema 2 Siano $\lambda_1 \dots \lambda_n$ autovalori di A , se $\lambda_i \neq \lambda_j \quad \forall i \neq j \implies A$ è diagonalizzabile.

Teorema 3 (teorema spettrale) Se $A = A^T \Rightarrow A$ è diagonalizzabile

Esempio: Blocco di Jordan $J = \begin{pmatrix} \alpha & 1 & \cdots & 0 \\ & \alpha & \ddots & \\ & & \ddots & 1 \\ 0 & & & \alpha \end{pmatrix}_{n \times n}$ $\det(J - \lambda I) = (\alpha - \lambda)^n$ autovalore

$\lambda = \alpha$ con molteplicità algebrica n

$$(J - \alpha I)x = 0$$

$$\begin{pmatrix} 0 & 1 & \cdots & 0 \\ & 0 & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix} x = 0 \Leftrightarrow x = \begin{pmatrix} c \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

La matrice non è diagonalizzabile (l'unico autospazio ha dimensione 1).

Per assurdo suppongo $J = X\Lambda X^{-1}$ (con $\Lambda = \alpha I$ necessariamente)

$= X\alpha IX^{-1} = \alpha XX^{-1} = \alpha I$ è falso poichè gli uno sono sulla sovradiagonale.

Teorema 4 (teorema di Schur) $\forall A \in \mathbb{R}^{n \times n} \exists S$ ($\det S \neq 0$) tale che $A = STS^{-1}$ dove T è triangolare superiore.

In generale, queste forme semplificate in cui viene trasformata A si dicono *forme canoniche*.

Ho una caratterizzazione per la diagonalizzabilità, ho condizioni sufficienti ed una forma triangolare, ci sono algoritmi semplici per trovare le matrici di passaggio?

5.5.1 Matrici simmetriche

Teniamo a mente il teorema spettrale (Se $A = A^T \Rightarrow A$ è diagonalizzabile).

Bisogna innanzitutto ampliare la definizione di quoziente di Rayleigh ($x \in \mathbb{R}^n \Rightarrow r(x) = \frac{x^T Ax}{x^T x}$ tale che x ev $\Rightarrow r(x) = \lambda$ ew) nel caso $x \in \mathbb{C}$. Se gli elementi x_i sono complessi, anche i quadrati sono complessi, non possiamo assicurare che $x^T x$ sia diverso da zero. Quindi ridefiniamo come

$$x \in \mathbb{C}^n \Rightarrow r(x) = \frac{x^* Ax}{x^* x}$$

dove $x^* = \bar{x}^T$ è un vettore trasposto coniugato.

Il denominatore si annulla? No, infatti

$$x \in \mathbb{C}^n \Rightarrow x^* x = (\bar{x}_1 \quad \bar{x}_2 \quad \cdots \quad \bar{x}_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \bar{x}_1 x_1 + \bar{x}_2 x_2 + \dots + \bar{x}_n x_n = \sum_{i=1}^n |x_i|^2$$

Se $x \neq 0 \rightarrow$

$$x^* x > 0$$

Lemma $\forall A, B \quad (AB)^* = B^* A^*$

Teorema $A = A^T \Rightarrow$ ogni $\lambda \in \mathbb{R}$

Dimostrazione: Sia $x \neq 0$ autovettore di A e considero il suo quoziente di Rayleigh

$r(x) = \frac{x^*Ax}{x^*x} = \frac{x^*\lambda x}{x^*x} = \frac{\lambda x^*x}{x^*x} = \lambda$ Anche per il caso complesso restituisce un autovalore. Bisogna dimostrare che è reale (coincide con il suo complesso coniugato)

$$\lambda = \frac{x^*Ax}{x^*x} \Rightarrow \lambda^* = \frac{(x^*Ax)^*}{x^*x} = \frac{x^*A^*(x^*)^*}{x^*x} = \frac{x^*Ax}{x^*x} = \lambda \Rightarrow \lambda = \lambda^* \Rightarrow \lambda \in \mathbb{R}$$

Ne discende che anche x è nei reali, quindi non c'è bisogno di fare conti con numeri complessi.

Teorema $A = A^T \in \mathbb{R}^{n \times n} \Rightarrow \exists X$ ortogonale ($X^T = X^{-1}$) tale che $A = X\Lambda X^T \Leftrightarrow$ esiste una base di autovettori ortonormali

Dimostrazione: Esiste una base di n autovettori indipendenti (poichè essendo simmetrica è diagonalizzabile per il teorema spettrale) $x_1, \dots, x_n \in \mathbb{R}^n$

Considero x_i e x_j relativi a $\lambda_i = \lambda_j$ (autovalori con molteplicità almeno 2), $\mathfrak{N}(A - \lambda I)$ è un sottospazio di dimensione pari alla molteplicità geometrica, quindi esiste una sua base ortonormale (si può fare la fattorizzazione QR).

Considero $\lambda_i \neq \lambda_j$, ho due equazioni $Ax_i = \lambda_i x_i$ e $Ax_j = \lambda_j x_j$, posso scegliere $\|x_i\|_2 = \|x_j\|_2 = 1$.

$$Ax_i = \lambda_i x_i \xRightarrow{\text{moltiplicando per } x_j^T} x_j^T (Ax_i) = x_j^T (\lambda_i x_i) \Rightarrow x_j^T Ax_i = \lambda_i x_j^T x_i \Rightarrow (x_j^T Ax_i)^T = \lambda_i (x_j^T x_i) \text{ da cui}$$

$x_i^T A^T x_j = \lambda_i x_i^T x_j$ ed essendo A simmetrica e x_j autovettore $x_i^T A x_j = x_i^T \lambda_j x_j = \lambda_j x_i^T x_j$ ovvero $(\lambda_i - \lambda_j) x_i^T x_j = 0$, essendo gli autovalori distinti $x_i^T x_j = 0$ quindi gli autovettori sono ortogonali ed ortonormali (avendo lunghezza 1).

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

$$\lambda_1 = 3 \text{ (autospazio } \left\langle \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle) \lambda_2 = 2 \text{ (autospazio } \left\langle \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\rangle)$$

I vettori sono già ortogonali ma non sono scalati ad 1, quindi vanno normalizzati (divisi per la norma)

$$x_1 \Rightarrow \hat{x}_1 = \frac{x_1}{\|x_1\|_2} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

$$x_2 \Rightarrow \hat{x}_2 = \frac{x_2}{\|x_2\|_2} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$$

$$A = X\Lambda X^{-1} \text{ dove } \Lambda = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \text{ e } X = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$\hat{X} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \Rightarrow \hat{X}^T = \hat{X}^{-1}$$

$$A = \hat{X}\Lambda\hat{X}^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

Esempio $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ è diagonalizzabile? (sì, è simmetrica) Com'è la diagonalizzazione?

Calcolo gli autovalori

$$\det(A - \lambda I) = \det \begin{pmatrix} 1-\lambda & 1 & 1 \\ 1 & 1-\lambda & 1 \\ 1 & 1 & 1-\lambda \end{pmatrix} = (1-\lambda) \det \begin{pmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{pmatrix} - 1 \cdot \det \begin{pmatrix} 1 & 1 \\ 1 & 1-\lambda \end{pmatrix} + 1 \cdot \det \begin{pmatrix} 1 & 1-\lambda \\ 1 & 1 \end{pmatrix} = \lambda^2(3-\lambda)$$

Autovalori di A : $\lambda_1 = 0$ (con molteplicità 2) e $\lambda_2 = 3$ (con molteplicità 1)

$$\Lambda = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

Autovettori di A : $\lambda_1 \Rightarrow (A - \lambda_1 I)x = 0 \Rightarrow Ax = 0 \Rightarrow x_1 + x_2 + x_3 = 0$ ha molteplicità geometrica

2: $\mathfrak{N}(A) = \left\langle \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\rangle$ è una possibile base dell'autospazio.

$$\lambda_2 \Rightarrow (A - 3I)x = 0 \Rightarrow \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0 \text{ si riduce a } \begin{cases} x_2 = x_3 \\ x_1 = x_2 \end{cases} \text{ quindi } \mathfrak{N}(A - 3I) =$$

$$\left\{ \begin{pmatrix} x_1 \\ x_1 \\ x_1 \end{pmatrix} \right\} = \left\langle \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\rangle$$

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

$$A = X\Lambda X^{-1} = \hat{X}\Lambda\hat{X}^T$$

Le prime due colonne di X non nascono ortogonali, bisogna quindi trovare una base ortonormale sostitutiva tramite fattorizzazione QR. Siccome questo metodo è più lungo conviene calcolare l'inversa di X invece di passare a \hat{X} .

Caso complesso Cosa si conserva dalla teoria precedente? Nel caso $A = A^*$ (matrice hermitiana) ogni autovalore è reale, perciò $A = X\Lambda X^*$ con $X^* = X^{-1}$ (matrice unitaria).

5.5.2 Applicazioni

5.5.2.1 Coniche

$$y = \alpha x^2 + \beta x + \gamma \text{ parabola}$$

$$\frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} = 1 \text{ ellisse}$$

$$\frac{x^2}{\alpha^2} - \frac{y^2}{\beta^2} = 1 \text{ iperbole}$$

Nel caso generale

$$\alpha x^2 + \beta xy + \gamma y^2 + \delta x + \eta y + \varphi = 0$$

Posso dire qualcosa sulla conica partendo dai coefficienti dell'equazione? Considero solo la parte quadratica e scrivo

$$A = \begin{pmatrix} \alpha & \frac{\beta}{2} \\ \frac{\beta}{2} & \gamma \end{pmatrix} = Q\Lambda Q^T$$

$$\alpha x^2 + \beta xy + \gamma y^2 = \begin{pmatrix} x & y \end{pmatrix} A \begin{pmatrix} x \\ y \end{pmatrix} = \underline{x}^T Q\Lambda Q^T \underline{x}$$

Effettuo un cambio di coordinate definendo $\begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} := Q^T \begin{pmatrix} x \\ y \end{pmatrix}$, sostituendo si ha $\begin{pmatrix} \tilde{x} & \tilde{y} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = \lambda_1 \tilde{x}^2 + \lambda_2 \tilde{y}^2 \left(+\tilde{\delta}\tilde{x} + \tilde{\eta}\tilde{y} + \tilde{\varphi} = 0 \right)$

Guardando gli autovalori della matrice simmetrica costituita dalla parte quadratica possiamo classificare la conica riconoscendo un'equazione nota:

$\lambda_1, \lambda_2 > 0 \Rightarrow$ ellisse

$\lambda_1 > 0, \lambda_2 < 0 \Rightarrow$ iperbole

$\lambda_2 = 0 \Rightarrow$ parabola

5.5.2.2 Google

Consideriamo cinque pagine internet e i loro link.

I motori di ricerca generano una classifica delle pagine basata sulla loro importanza, definita in modo ricorsivo: essa si trasmette alle pagine linkate equamente, e per ogni pagina essa è la somma delle importanze trasmesse dalle pagine che puntano ad essa. Nella figura sotto ad esempio la pagina 2 ha importanza $x_2 = \frac{x_1}{2} + x_3 + \frac{x_4}{3}$.

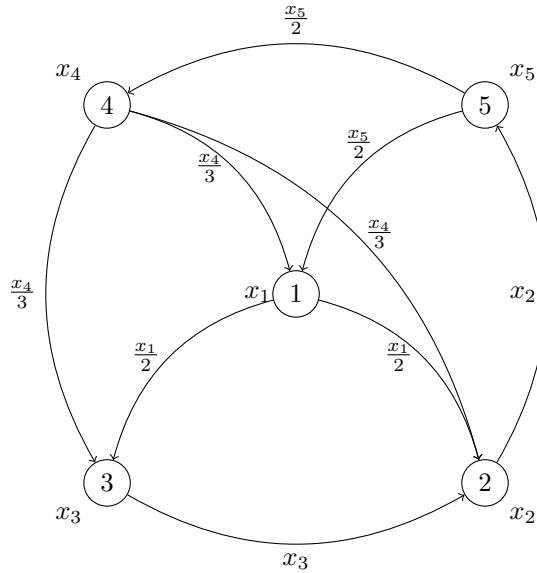


Figura 5.2: Grafo delle pagine

Si può costruire una *matrice di incidenza* quadrata con dimensione pari al numero dei nodi, costruita in modo tale che nel caso ci sia un link fra due pagine, nella posizione corrispondente ci sia 1.

$$\begin{cases} a_{ij} = 1 & j \rightarrow i \\ a_{ij} = 0 & j \nrightarrow i \end{cases}$$

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

Questa matrice viene scalata nelle colonne secondo il *grado del nodo* (il numero di archi uscenti) corrispondente:

$$G = \begin{pmatrix} 0 & 0 & 0 & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & 0 & 1 & \frac{1}{3} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Ogni riga è l'equazione della definizione implicita dell'importanza. Definito il vettore delle importanze $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}$, impostiamo il sistema $Gx = x$ riducendo il problema alla ricerca dell'autovettore, ma siamo sicuri che ci sia l'autovalore 1?

Consideriamo la trasposta della matrice di Google (che è stocastica per colonne, trasposta diventa stocastica per righe) $A = G^T$: della matrice A sappiamo che ha un autovettore di tutti 1 e l'autovalore 1 (per una proprietà della matrice stocastica per righe).

C'è una relazione fra le autocoppie di una matrice e della sua trasposta?

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

G e G^T hanno lo stesso polinomio caratteristico, quindi avranno gli stessi autovalori, ma diversi autovettori.

Google quindi cerca di calcolare l'autovettore relativo all'autovalore 1: la somma delle importanze dev'essere 1, e ciascuna di esse è positiva e compresa in $[0, 1]$: $\sum_{i=1}^n x_i = 1$

Per calcolare il pagerank costruisce la matrice delle pagine e ne calcola l'autovettore. Non usa l'eliminazione gaussiana ma un altro approccio di calcolo approssimato poichè la matrice è *sparsa* (ha tanti zeri).

5.5.2.3 Nota su Matlab

Il comando *eig* restituisce due matrici contenenti autovettori ed autovalori tramite diagonalizzazione: $[V, D] = \text{eig}(A)$ tale che $A = VDV^{-1}$.

5.5.3 Analisi dell'errore

Consideriamo una perturbazione di A (autovalori λ), la matrice B con autovalori μ .

Abbiamo visto la norma indotta per le matrici definita come $\|A\| := \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$.

In input l'errore assoluto sarà $\|A - B\|$ e quello relativo $\frac{\|A - B\|}{\|A\|}$, in output l'errore assoluto sarà $|\lambda - \mu|$ e quello relativo $\frac{|\lambda - \mu|}{|\lambda|}$.

Teorema di Bauer-Fike Sia A diagonalizzabile ($\Leftrightarrow A = X\Lambda X^{-1}$) allora

$$\forall \lambda \exists \mu \text{ tale che } |\lambda - \mu| < \mu(X) \|A - B\|$$

Consideriamo X ortogonale $\Rightarrow X^T X = I \Leftrightarrow X$ isometria $\rightarrow \|X\|_2 = \max_{x \neq 0} \frac{\|Xx\|}{\|x\|} = 1$ (se una matrice è ortogonale, la sua norma è sempre uguale ad 1).

Il teorema di Bauer-Fike applicato alla norma 2 dice che $|\lambda - \mu| \leq \|A - B\|_2$.

La condizione X ortogonale implica che la matrice A è simmetrica. Le matrici simmetriche hanno un problema agli autovalori ben condizionato.

Norma 2 per le matrici In generale

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)}$$

Si dimostra che gli autovalori sono tutti non negativi; si moltiplica A per la trasposta e si cerca l'autovalore massimo di cui si calcola la radice quadrata.

Capitolo 6

Decomposizione ai valori singolari

Nel caso di matrici simmetriche si ottiene la diagonalizzazione con matrici ortogonali come visto in precedenza. Nel caso in cui la matrice non sia simmetrica o sia rettangolare si procede ad usare la SVD (*Singular Values Decomposition*)

$$A = U\Sigma V^T$$

Ha un evidente apparentamento con la diagonalizzazione (U e V sono ortogonali e Σ è diagonale); viene introdotta da un teorema.

6.0.3.1 Teorema

$\forall A \in \mathbb{R}^{m \times n} \exists U \in \mathbb{R}^{m \times m}, V \in \mathbb{R}^{n \times n}$ ortogonali e $\exists \sigma_1, \dots, \sigma_p$ ($p = \min(m, n)$) tali che $A = U\Sigma V^T$

con $\Sigma = \begin{pmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_n & \\ & & & 0 \end{pmatrix}$ se $m \geq n$ oppure $\Sigma = \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_m & \\ 0 & & & 0 \end{pmatrix}$ se $m < n$, e inoltre $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$ sono detti valori singolari.

$U = (u_1 | \dots | u_m)$ le colonne sono dette vettori singolari sinistri

$V = (v_1 | \dots | v_n)$ le colonne sono dette vettori singolari destri

6.1 Proprietà algebriche della SVD

1

$A = U\Sigma V^T \Rightarrow AV = U\Sigma$ leggendo per colonne $\forall j = 1 \dots n$ $Av_j = U\Sigma_j$ bisogna distinguere due casi:

caso $m \geq n$ $U\Sigma e_j = U\sigma_j e_j = \sigma_j u_j$ quindi $\forall j = 1 \dots n$ $Av_j = \sigma_j u_j$

caso $m \leq n$ $\forall j = 1 \dots m$ $Av_j = \sigma_j u_j$ e $\forall j = m+1 \dots n$ $Av_j = 0$

2

$A = U\Sigma V^T \Rightarrow A^T = (U\Sigma V^T)^T = V\Sigma^T U^T \Rightarrow A^T U = V\Sigma^T$ letta per colonne $\forall j = 1 \dots m$ $A^T u_j = V\Sigma^T e_j$

$m \geq n$ $j = 1 \dots n$ $A^T u_j = \sigma_j v_j$

$j = n + 1 \dots m$ $A^T u_j = 0$
 $m \leq n$ $j = 1 \dots m$ $A^T u_j = \sigma_j v_j \Rightarrow AA^T u_j = \sigma_j A v_j = \sigma_j^2 u_j$ quindi σ_j^2 autovalore di AA^T e u_j autovettore

$m \geq n$ $A^T A v_j = A^T \sigma_j u_j = \sigma_j A^T u_j = \sigma_j^2 v_j$ quindi v_j autovettore di $A^T A$ e σ_j^2 autovalore.

Altro procedimento per ottenere lo stesso risultato:

$$A \in \mathbb{R}^{m \times n} \rightarrow A = U \Sigma V^T, A^T = V \Sigma^T U^T$$

$A^T A = V \Sigma^T U^T U \Sigma V^T = V \Sigma^T \Sigma V^T$ questa è di nuovo una diagonalizzazione $\Rightarrow v_1 \dots v_n$ sono autovettori di $A^T A$, $\sigma_1^2 \dots \sigma_m^2$ autovalori di $A^T A$ per i primi m autovettori; gli $n - m$ autovettori nel caso $m \leq n$ sono associati all'autovalore nullo.

$$\text{Infatti, nel caso } m \geq n \quad \Sigma^T \Sigma = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix}_{n \times n},$$

$$\text{nel caso } m \leq n \quad \Sigma^T \Sigma = \begin{pmatrix} \sigma_1^2 & & & & \\ & \ddots & & & \\ & & \sigma_n^2 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix}_{n \times n}$$

$AA^T = U \Sigma (V^T V) \Sigma^T U^T = U (\Sigma \Sigma^T) U^T$ $u_1 \dots u_m$ sono autovettori di AA^T . Le matrici sono uguali al caso precedente ma nei casi invertiti.

Algoritmo 6.1 Calcolo dei valori singolari e dei vettori singolari

calcolo λ_j di $A^T A$

$$\sigma_j := \sqrt{\lambda_j}$$

calcolo ev di $A^T A$ (vettori singolari destri)

calcolo λ_j di AA^T

$$\sigma_j := \sqrt{\lambda_j}$$

calcolo ev di AA^T (vettori singolari sinistri)

Nel caso $m = n$ $A = A^T \rightarrow$ autovalori di $A^T A = A^2$:

$$\text{Se } \lambda_j \text{ autovalore di } A \Rightarrow \lambda_j^2 \text{ autovalore di } A^2 \Rightarrow \sigma_j = \sqrt{\lambda_j^2} = |\lambda_j| = \pm \lambda_j$$

autovettore di $A^T A = AA^T = A^2 \equiv$ autovettore di A (vettori singolari destri e sinistri) $u_j, v_j = \pm x_j$ secondo la seguente regola:

$$\lambda_j \geq 0 \Rightarrow \sigma_j = \lambda_j, u_j = v_j = x_j;$$

$$\lambda_j < 0 \Rightarrow \sigma_j = -\lambda_j, v_j = x_j, u_j = -x_j \text{ (o viceversa).}$$

Sempre in questo caso, la SVD dà informazioni sulla norma 2: $A^T A$ ha autovalori $\sigma_1^2 \geq \dots \geq \sigma_n^2$ quindi $\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)} = \sqrt{\sigma_1^2} = \sigma_1$ e $\|A\|_F = \sqrt{\sigma_1^2 + \dots + \sigma_n^2}$.

6.1.0.2 Esempio

$$A = A^T = X \Lambda X^T \text{ con } X = \frac{1}{9} \begin{pmatrix} -4 & 7 & 4 \\ 8 & 4 & 1 \\ 1 & -4 & 8 \end{pmatrix} \text{ e } \Lambda = \text{diag}(-3, -1, 2).$$

I valori singolari devono essere ordinati, quindi $\Sigma = \text{diag}(3, 2, 1)$, perciò scambio l'ordine degli autovalori e degli autovettori.

$$\begin{aligned}\sigma_2 &= 2 = \lambda_3 \\ u_2 &= v_2 = \frac{1}{9} \begin{pmatrix} 4 \\ 1 \\ 8 \end{pmatrix} \\ \sigma_1 &= 3 = -\lambda_1 \\ v_1 &= \frac{1}{9} \begin{pmatrix} -4 \\ 8 \\ 1 \end{pmatrix} \quad u_1 = \frac{1}{9} \begin{pmatrix} 4 \\ -8 \\ -1 \end{pmatrix} \\ \sigma_3 &= 1 = -\lambda_2 \\ v_3 &= \frac{1}{9} \begin{pmatrix} 7 \\ 4 \\ -4 \end{pmatrix} \quad u_3 = \frac{1}{9} \begin{pmatrix} -7 \\ -4 \\ 4 \end{pmatrix} \\ V &= \frac{1}{9} \begin{pmatrix} -4 & 4 & 7 \\ 8 & 1 & 4 \\ 1 & 8 & -4 \end{pmatrix} \quad U = \frac{1}{9} \begin{pmatrix} 4 & 4 & -7 \\ -8 & 1 & -4 \\ -1 & 8 & 4 \end{pmatrix}\end{aligned}$$

6.1.0.3 Diadi

$$A = U\Sigma V^T = U(\sigma_1 e_1 v_1^T + \dots + \sigma_p e_p v_p^T) = \sigma_1 U e_1 v_1^T + \dots + \sigma_p U e_p v_p^T = \sigma_1 u_1 v_1^T + \dots + \sigma_p u_p v_p^T$$

$$A = \sigma_1 u_1 v_1^T + \dots + \sigma_p u_p v_p^T \text{ dove } u_p v_p^T \text{ è detto } \textit{diade}.$$

La matrice A $m \times n$ richiede $m \cdot n$ locazioni di memoria, scritta come somma di diadi richiede $p(1 + m + n)$, ponendo $p = \min(m, n)$ non conviene molto, perchè occupa più spazio; nei casi reali accade spesso che bastano $p \ll n$ diadi in quanto molti valori singolari sono trascurabili, il che porta ad un risparmio di memoria. Ad esempio $m = 640$, $n = 480$, scegliendo le prime 10 diadi sarà $10(1 + 640 + 480) = 11210$ invece di $640 \cdot 480 = 307200$.

6.2 Proprietà geometriche

La sfera unitaria $S = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$, trasformata secondo A , diventa un ellissoide e si dimostra che la misura dei semiassi di tale ellissoide sono i valori singolari.

6.2.0.4 Teorema

Sia $A = U\Sigma V^T$ con $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0 = \sigma_{r+1} = \dots = \sigma_p$ (supponiamo che ci siano r valori singolari positivi e il resto nulli), allora:

1. $\text{rk}(A) = r$
2. $\mathfrak{R}(A) = \langle u_1, \dots, u_r \rangle$
3. $\mathfrak{N}(A) = \langle v_{r+1}, \dots, v_n \rangle$

6.2.0.5 Esempio

$$m = 10, n = 6$$

$$\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6$$

$$100, 50, 10, 0.01, 0, 0$$

$$\text{Il rango è } \text{rk}(A) = 4$$

$$\mathfrak{R}(A) = \langle u_1, u_2, u_3, u_4 \rangle$$

$$\mathfrak{N}(A) = \langle v_5, v_6 \rangle$$

$$A^T = V \Sigma^T U^T$$

$$\mathfrak{R}(A^T) = \langle v_1, v_2, v_3, v_4 \rangle$$

$$\mathfrak{N}(A^T) = \langle u_5, u_6, u_7, u_8, u_9, u_{10} \rangle$$

6.2.0.6 Esempio

$$m = 4, n = 2$$

$$\sigma_j = 10, 1$$

$$\langle u_1, u_2 \rangle = \mathfrak{R}(A)$$

$$\mathfrak{N}(A) = \{0\}$$

$$\mathfrak{R}(A^T) = \langle v_1, v_2 \rangle$$

$$\mathfrak{N}(A^T) = \langle u_3, u_4 \rangle$$

$$A = \sigma_1 u_1 v_1^T + \dots + \sigma_p u_p v_p^T = \sigma_1 u_1 v_1^T + \dots + \sigma_r u_r v_r^T$$

Più una matrice ha rango basso, più è comprimibile una informazione: c'è una relazione fra numero di diadi e rango della matrice.

6.3 Affidabilità del rango

Considerando applicazioni reali, il rango non lo si conosce esattamente poichè c'è incertezza sulla matrice. Quanto è affidabile l'informazione sul rango?

6.3.0.7 Teorema

Sia $\text{rk}(A) = r$, $A = U \Sigma V^T$, consideriamo $k < r \Rightarrow A_k = \sigma_1 u_1 v_1^T + \dots + \sigma_k u_k v_k^T$, per costruzione A_k ha rango k , allora

$$\|A - A_k\|_2 = \min \{ \|A - B\|_2 : \text{rk}(B) = k \} = \sigma_{k+1}$$

Il rango della matrice A è k con un errore di σ_{k+1} ; il concetto di rango può essere quindi inteso come *rango numerico*.

6.4 Minimi quadrati

$$A \in \mathbb{R}^{m \times n}, m \geq n, \text{rk}(A) = n, b \in \mathbb{R}^n$$

Il problema ai minimi quadrati è

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2$$

Si dimostra che la soluzione è unica con le ipotesi già esposte.

Abbiamo visto due metodi di risoluzione:

1. equazioni normali: $(A^T A) x = A^T b$, $\det(A^T A) \neq 0$ con costo $\frac{1}{2}mn^2$ ed errore legato a $\mu(A^T A)$

2. fattorizzazione QR: $A = Q \begin{pmatrix} R \\ 0 \end{pmatrix}$, $Q^T b = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$, $Rx = h_1$, se $\det R \neq 0$ $\|Ax - b\|_2 = \|h_2\|$ con costo mn^2 ed errore legato a $\mu(R)$

Terzo metodo Questo metodo funziona sempre, ma deve essere introdotto da un teorema.

Teorema: sia $A = U\Sigma V^T$ con $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0 = \sigma_{r+1} = \dots = \sigma_n$, allora una soluzione del problema ai minimi quadrati è la *soluzione generalizzata* o *pseudosoluzione*

$$x^\dagger = \sum_{i=1}^r \left(\frac{u_i^T b}{\sigma_i} \right) v_i$$

Se $r < n$, allora x^\dagger è la soluzione di minima norma: $\|x^\dagger\| = \min\{\|x\|_2 : x \text{ risolve il problema ai minimi quadrati}\}$.

La pseudosoluzione ha componenti lungo le direzioni che non corrispondono a quelle del nucleo della matrice A .

Dimostrazione: $\|Ax - b\|_2^2 = \|U\Sigma V^T x - U U^T b\|_2^2 = \|U(\Sigma y - U^T b)\|_2^2 = \|\Sigma y - U^T b\|_2^2$

$$\Sigma y = \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & 0 & \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sigma_1 y_1 \\ \sigma_2 y_2 \\ \vdots \\ \sigma_r y_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$U^T b = \begin{pmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_m^T \end{pmatrix} b = \begin{pmatrix} u_1^T b \\ u_2^T b \\ \vdots \\ u_m^T b \end{pmatrix}$$

$$\Sigma y - U^T b = \begin{pmatrix} \sigma_1 y_1 - u_1^T b \\ \vdots \\ \sigma_r y_r - u_r^T b \\ -u_{r+1}^T b \\ \vdots \\ -u_m^T b \end{pmatrix} \Rightarrow \|\Sigma y - U^T b\|_2^2 = \sum_{i=1}^r (\sigma_i y_i - u_i^T b)^2 + \sum_{i=r+1}^m (u_i^T b)^2$$

Quest'ultima espressione va minimizzata rispetto a $y \in \mathbb{R}^n$, il primo termine si minimizza se $\forall i = 1 \dots r \quad \sigma_i y_i - u_i^T b = 0$

$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2 = \min_{y \in \mathbb{R}^n} \|\Sigma y - U^T b\|_2 \Leftrightarrow \forall i = 1 \dots r \quad y_i = \frac{u_i^T b}{\sigma_i}$ e $\forall i = r+1 \dots n \quad y_i$ arbitrario (non variano la norma del residuo).

$$\|x\|_2 = \|Vy\|_2 = \|y\|_2 \quad (V \text{ isometria}) \Rightarrow \min \|x\|_2 = \min \|y\|_2$$

$$\|y\|_2^2 = y_1^2 + \dots + y_r^2 + y_{r+1}^2 + \dots + y_n^2$$

dove la prima parte è fissata, devo cercare di porre a zero le ultime componenti per ottenere la minima soluzione: $y_i = 0 \quad \forall i = r+1, \dots, n \rightarrow \min \|y\|_2$

$$x = Vy = (v_1 | \dots | v_n) \begin{pmatrix} y_1 \\ \vdots \\ y_r \\ 0 \\ \vdots \end{pmatrix} = y_1 v_1 + \dots + y_r v_r = x^\dagger$$

Abbiamo dimostrato che la scelta delle y fa ottenere l'espressione a minima norma e abbiamo inoltre dimostrato che $\|Ax^\dagger - b\|_2^2 = \sum_{i=r+1}^m (u_i^T b)^2$

Quindi il terzo metodo è:

$A = U\Sigma V^T$, dopo di che calcolo $y_i = \frac{u_i^T b}{\sigma_i}$ $i = 1 \dots r$: $x^\dagger = y_1 v_1 + \dots + y_r v_r$ con costo rn

La SVD ha un costo proibitivo, ma ha il vantaggio di non funzionare sotto ipotesi: quando il rango non è massimo è l'unico metodo che funziona.

Esempio Supponiamo di avere incertezza sulla scelta del rango ($\sigma_r \sim 10^{-12}$) l' r -esimo termine è $\frac{u_r^T b}{\sigma_r} v_r \approx 10^{12}$: l'incertezza sull'inclusione dell' r -esimo termine introduce una grande differenza nella soluzione del problema ai minimi quadrati.

6.5 Pseudoinversa

$A \in \mathbb{R}^{n \times n}$, $\det A \neq 0 \Rightarrow \exists A^{-1} : AA^{-1} = A^{-1}A = I$, $Ax = b \Rightarrow x = A^{-1}b$

$\forall A \in \mathbb{R}^{m \times n} : \exists A^\dagger \in \mathbb{R}^{n \times m} : \min_{x \in \mathbb{R}^n} \|Ax - b\|_2 \Leftrightarrow x^\dagger = A^\dagger b$

$$A = U\Sigma V^T \Rightarrow A^\dagger = V\Sigma^\dagger U^T \text{ con } \Sigma^\dagger = \begin{pmatrix} \frac{1}{\sigma} & & & 0 \\ & \ddots & & \\ & & \frac{1}{\sigma_r} & \\ 0 & & & 0 \end{pmatrix}_{n \times m}, \quad \frac{1}{\sigma_1} \leq \frac{1}{\sigma_2} \leq \dots \leq \frac{1}{\sigma_r}$$

$$\Rightarrow A^\dagger = \frac{1}{\sigma_1} v_1 u_1^T + \dots + \frac{1}{\sigma_r} v_r u_r^T$$

Con l'inversa si poteva ottenere la scrittura alternativa della soluzione del sistema lineare come $Ax = b \Rightarrow x = A^{-1}b$

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2 \rightarrow x^\dagger = \sum_{i=1}^r \frac{u_i^T b}{\sigma_i} v_i$$

Adesso prendiamo la pseudoinversa e la moltiplichiamo per il termine noto dei minimi quadrati: $A^\dagger b = \sum_{i=1}^r \frac{1}{\sigma_i} v_i u_i^T b$ che è proprio la soluzione generalizzata.

$m = n$, $\det A \neq 0 \Rightarrow \exists A^{-1}$, $\text{rk}(A) = n$, i valori singolari sono tutti strettamente positivi

$$A^\dagger = V \begin{pmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma_n} \end{pmatrix} U^T = (V^T)^{-1} \Sigma^{-1} U^{-1} = (U\Sigma V^T)^{-1} = A^{-1} \text{ dove al secondo passaggio}$$

si sfrutta l'ortogonalità di V (la trasposta è uguale all'inversa), e al terzo il prodotto delle inverse è l'inversa del prodotto scambiando i fattori.

L'inversa generalizzata coincide con l'inversa classica (quando esiste).

Esaminiamo il caso in cui le righe sono maggiori delle colonne

$$m \geq n, n = \text{rk}(A) \Rightarrow A^T = V\Sigma^T U^T$$

proviamo a fare $A^T A = V (\Sigma^T \Sigma) V^T = V \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix} V^T$

$$(A^T A)^{-1} A^T = V \begin{pmatrix} \frac{1}{\sigma_1^2} & & \\ & \ddots & \\ & & \frac{1}{\sigma_n^2} \end{pmatrix} V^T V \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} U^T = V \begin{pmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma_n} \end{pmatrix} U^T \text{ che}$$

è la definizione di pseudoinversa.

$A^\dagger = (A^T A)^{-1} A^T$ se $m \geq n = \text{rk}(A)$ ricorda le equazioni normali del problema dei minimi quadrati $(A^T A) x^\dagger = A^T b \Rightarrow x^\dagger = (A^T A)^{-1} A^T b$

Capitolo 7

Metodi iterativi per il calcolo degli autovalori

Gli autovalori non possono essere calcolati esplicitamente, bisogna procedere per approssimazioni successive (usando un *metodo iterativo*).

Una successione $\{v_m\}_{m \in \mathbb{N}}$ è caratterizzata

- da un elemento di partenza v_0
- dalla legge $v_m \mapsto v_{m+1}$ che regola il passaggio da un elemento al successivo
- dal suo essere convergente o meno (ovvero $\exists x \lim_{m \rightarrow \infty} v_m = x$) e dalla velocità con cui tende al suo limite (ad esempio $\|v_m - x\| \leq \frac{1}{m}$ oppure $\leq \frac{1}{2^m}$)
- dal numero finito di passi in cui fermare il ciclo: ad esempio quando la differenza tra un passo e il successivo sia minore della tolleranza in input ($\|v_m - v_{m-1}\| < \varepsilon$) o una qualunque altra condizione che ne costituisca il *criterio di arresto*

7.0.1 Metodo delle potenze

Supponiamo di avere una matrice $A \in \mathbb{R}^{n \times n}$ con autovalori di cui solo uno di modulo massimo $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_n|$ e gli autovettori ad essi associati $x_1 \dots x_n$ linearmente indipendenti. Prendiamo adesso una combinazione lineare con coefficienti $\alpha_1, \dots, \alpha_n$ degli autovettori che genera un vettore v_0 e scriviamo la successione $v_{m+1} = Av_m$, ovvero $v_m = A^m v_0 = \alpha_1 A^m x_1 + \dots + \alpha_n A^m x_n = \alpha_1 \lambda_1^m x_1 + \dots + \alpha_n \lambda_n^m x_n$; raccogliendo il primo autovalore $v_m = \lambda_1^m \left(\alpha_1 x_1 + \dots + \alpha_n \left(\frac{\lambda_n}{\lambda_1} \right)^m x_n \right)$. Per ipotesi, il primo autovalore è il più grande, perciò i rapporti all'interno della parentesi sono sempre minori di 1, scalando e passando al limite otteniamo

$$\lim_{m \rightarrow \infty} \frac{v_m}{\lambda_1^m} = \alpha_1 x_1$$

ovvero $v_m \approx \lambda_1^m \alpha_1 x_1$

$$\lambda_1 \approx \frac{v_m^T A v_m}{v_m^T v_m} \text{ e converge con velocità } \left| \frac{\lambda_2}{\lambda_1} \right|^m$$

Siccome la potenza $|\lambda_1|^m$ può tendere a 0 o a ∞ a seconda che l'autovalore sia maggiore o minore di 1 in modulo, occorre normalizzare i termini della successione.

$$v_m \mapsto Av_m =: w_{m+1} \mapsto v_{m+1} := \frac{w_{m+1}}{\|w_{m+1}\|}$$

Algoritmo 7.1 Metodo delle potenze

$$\begin{aligned} y_0 &= (1, \dots, 1)^T \\ \forall m &= 0 \dots \max \\ w_{m+1} &= Ay_m \\ \lambda_1^{(m+1)} &= \frac{y_m^T w_{m+1}}{y_m^T y_m} \\ y_{m+1} &= \frac{w_{m+1}}{\|w_{m+1}\|} \end{aligned}$$

7.0.2 Metodo delle potenze inverse

Se $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_n|$ e ci interessa un λ_i diverso da quello dominante, dato $p \in \mathbb{R}$ tale che $\det(A - pI) \neq 0$ poniamo $B := (A - pI)^{-1}$ e cerchiamo i suoi autovalori.

Se (λ, x) è autocoppia di A ,

$$\begin{aligned} (A - pI)x &= Ax - pIx = \lambda x - px = (\lambda - p)x \\ (A - pI)^{-1}(A - pI)x &= (A - pI)^{-1}(\lambda - p)x \\ x &= (\lambda - p)(A - pI)^{-1}x \Rightarrow (A - pI)^{-1}x = \frac{1}{\lambda - p}x \end{aligned}$$

Essendo $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_n|$ autovalori di A e $|\mu_1| > |\mu_2| \geq |\mu_3| \geq \dots \geq |\mu_n|$ autovalori di B , $\forall \mu \exists \lambda_j$ tale che $\mu = \frac{1}{\lambda_j - p}$.

$\mu_1 \leftrightarrow \lambda_j$ tale che $|\lambda_j - p|$ è minimo

Applico il metodo classico delle potenze a $B = (A - pI)^{-1}$, fisso p (non è autovalore di A), allora il metodo converge a μ_1 per λ più vicino a p : $\lambda = p + \frac{1}{\mu_1}$.

$w_{m+1} := (A - pI)^{-1} y_m \rightarrow (A - pI) w_{m+1} = y_m$: risolvo il sistema lineare con matrice $(A - pI)$ e termine noto y_m (costo circa $\frac{n^3}{3}$)

$$\begin{aligned} \rho_m &:= \frac{y_m^T w_{m+1}}{y_m^T y_m} \approx \mu_1 \\ \tilde{\rho}_m &:= p + \frac{y_m^T y_m}{y_m^T w_{m+1}} \approx \lambda_j \end{aligned}$$

La velocità di convergenza del metodo, detto λ_k il secondo autovalore più vicino a p , è $\frac{|\mu_2|}{|\mu_1|} = \frac{|\lambda_j - p|}{|\lambda_k - p|}$

Nota. Affinché il metodo delle potenze inverse converga, non serve più la disuguaglianza $|\lambda_1| > |\lambda_2|$, ma diventa invece essenziale l'ipotesi $|\mu_1| > |\mu_2|$ che è equivalente a $|\lambda_j - p| < |\lambda_k - p|$ (ossia lo shift p è più vicino a λ_j di quanto lo sia a λ_k).

Algoritmo 7.2 Metodo delle potenze inverse

$$\begin{aligned} \forall m &= 0 \dots \max \\ \text{risolvo } (A - pI) w_{m+1} &= y_m \\ \rho_m &= \frac{y_m^T w_{m+1}}{y_m^T y_m} \\ y_{m+1} &= \frac{w_{m+1}}{\|w_{m+1}\|} \\ \tilde{\rho}_m &= p + \frac{1}{\rho_m} \end{aligned}$$

Esempio A ha autovalori 15, 10, 5, 1
 $p = 7$
 B ha autovalori $\frac{1}{8}, \frac{1}{3}, -\frac{1}{2}, -\frac{1}{6}$
 $\mu_1 \leftrightarrow \lambda_3, \mu_2 \leftrightarrow \lambda_2$

Teorema Se p viene variato ogni volta sostituendolo con l'approssimazione trovata ($\tilde{\rho}_m \rightarrow p$), allora $\frac{|\lambda_j - p|}{|\lambda_k - p|}$ decresce e vale
 $|\tilde{\rho}_m - \lambda_j| \approx |\tilde{\rho}_{m-1} - \lambda_j|^2$

7.0.3 Localizzazione degli autovalori

p viene scelto vicino all'autovalore che vogliamo individuare tramite *localizzazione degli autovalori*.

Cerchi di Gerschgorin $\mathcal{C} := \left\{ z \in \mathbb{C} : |z - a_{ii}| \leq \sum_{j \neq i} |a_{ij}| \right\}$

I teoremi di Gerschgorin servono a decidere il p iniziale dell'algoritmo delle potenze inverse.

Primo teorema di Gerschgorin Ogni autovalore di A è nell'unione dei cerchi corrispondenti alle righe di A

$$\{\lambda(A)\} \subset \bigcup_{i=1}^n C_i$$

Secondo teorema di Gerschgorin Se k cerchi sono disgiunti da $n - k$ cerchi, allora k autovalori sono nei primi k cerchi e $n - k$ sono nei rimanenti.

Dimostrazione del primo teorema Tesi: $\forall \lambda \exists i : \lambda \in C_i \Leftrightarrow |\lambda - a_{ii}| \leq \sum_{j \neq i} |a_{ij}|$

$Ax = \lambda x$ con $x \neq 0$

scriviamo l' i -esima riga

$$\sum_{j=1}^n a_{ij} x_j = \lambda x_i$$

per $i = 1 \dots n \Rightarrow a_{ii} x_i + \sum_{j \neq i} a_{ij} x_j = \lambda x_i$

$$(\lambda - a_{ii}) x_i = \sum_{j \neq i} a_{ij} x_j$$

$$|\lambda - a_{ii}| \cdot |x_i| = \left| \sum_{j \neq i} a_{ij} x_j \right| \leq \sum_{j \neq i} |a_{ij}| \cdot |x_j|$$

Scelgo i tale che $|x_i| \geq |x_j| \forall j \neq i$

$|\lambda - a_{ii}| |x_i| \leq \left(\sum_{j \neq i} |a_{ij}| \right) |x_i|$; $x \neq 0$ implica $|x_i| > 0$ da cui la tesi.

7.1 Metodo QR

Richiamando la similitudine $A \sim B \Leftrightarrow SAS^{-1} = B$ e la fattorizzazione QR si definisce il metodo QR come:

$$A \in \mathbb{R}^{n \times n} \rightarrow A = QR \Rightarrow Q^T A = R$$

$$A \sim Q^T A (Q^T)^{-1} = R (Q^{-1})^T = R (Q^T)^T = RQ \sim A$$

Procedimento:

Algoritmo 7.3 Metodo QR

 $A_1 := A \quad k = 1, 2, \dots$
 $A_k := Q_k R_k$
 $A_{k+1} := R_k Q_k = Q_k^T A_k Q_k$

$\{A_k\}_{k=1,2,\dots}$ sono simili ad A

La differenza col metodo delle potenze è che quello restituiva un solo autovalore, questo restituisce tutti gli autovalori.

Esempio $T = \text{diag}(12 : -3 : 0) + \text{diag}(\text{ones}(4, 1), 1) + \text{diag}(\text{ones}(4, 1), -1)$

 $[Q, R] = \text{qr}(T)$
 $Tk = R * Q$
 $[Q, R] = \text{qr}(Tk);$
 $Tk = R * Q; k = k + 1;$

Si iterano le due istruzioni fino ad ottenere una matrice vicina ad una diagonale a blocchi, con un blocco 4×4 ed uno 1×1 (ponendo un limite, ad esempio 10^{-8} , sotto il quale gli elementi fuori possono essere considerati nulli), a questo punto si salva l'elemento trovato nel vettore degli autovalori ($ew(5) = Tk(5, 5)$), si considera solo il blocco 4×4 (*deflazione*) e si ripete il procedimento fino ad arrivare ad una matrice 2×2

 $T43 = Tk(4, 3)$
 $\text{while } \text{abs}(T43) > 1e-8$
 $[Q, R] = \text{qr}(Tk); k = k + 1;$
 $Tk = R * Q;$
 $T43 = Tk(4, 3);$
 end
 $Tk = Tk(1 : 3, 1 : 3)$
 \dots
 $ew(1 : 2) = \text{diag}(Tk)$
 $\text{lambda} = \text{eig}(T)$
 $\text{abs}(\text{lambda} - \text{sort}(ew'))$ vediamo che gli errori sono $\sim 10^{-15}$

Teorema Se $|\lambda_1| > |\lambda_2| > \dots > |\lambda_n| \Rightarrow \lim_{k \rightarrow \infty} A_k = \begin{pmatrix} \lambda_1 & * & \dots & * \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \dots & 0 & \lambda_n \end{pmatrix}$

Se la matrice è simmetrica, genera la diagonale di autovalori, altrimenti genera una matrice triangolare superiore con sulla diagonale gli autovalori.

$(A_k)_{i,j} = O\left(\left|\frac{\lambda_i}{\lambda_j}\right|^k\right)$ Un elemento durante il metodo tende a zero come la successione all'interno della parentesi.

Quando il metodo QR non converge si può comunque notare qualcosa: infatti esso estrae gli autovalori che si possono estrarre e riconduce a matrici più piccole che contengono gli altri autovalori con stesso modulo (il loro rapporto è uguale ad 1).

7.1.1 Costo del metodo ed utilizzo di matrici di Hessenberg

Una fattorizzazione QR con riflessioni di Householder costa $\frac{2}{3}n^3$ nel caso di matrici arbitrarie, un passo del metodo costa n^3 .

Nel caso di matrici tridiagonali e simmetriche, si può vedere che ad ogni passo le matrici $\{T_k\}$ sono anch'esse tridiagonali e simmetriche: per la fattorizzazione delle matrici tridiagonali si usano le rotazioni di Givens, e ogni iterazione costa $10n$.

Nel caso di matrici di Hessenberg (matrici triangolari superiori con la prima sottodiagonale non nulla), una iterazione del metodo QR costa n^2 e si prova che tutte le matrici $\{A_k\}$ ad ogni passo sono matrici di Hessenberg.

Una matrice qualsiasi si può portare in un numero finito di passi in una matrice di Hessenberg tramite similitudine.

$$A = \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix} \quad \text{Una possibile similitudine (da non eseguire sulla prima colonna) è } G(2, 3, \theta) A G(2, 3, \theta)^{-1}$$

dove $G(2, 3, \theta)^{-1} = G(2, 3, -\theta)$ tocca le colonne invece che le righe, lasciando tale lo zero ricavato precedentemente.

In questo caso la matrice di Hessenberg è $H = G_{34}G_{24}G_{23}AG_{23}^TG_{24}^TG_{34}^T$. Questa similitudine costa n^3 ma viene eseguita solo una volta per preparare la matrice, garantendo che i successivi passi del metodo costino solo n^2 . Nel caso di matrice simmetrica $A = A^T \Rightarrow H = H^T \Rightarrow H$ è tridiagonale.