Capitolo 6

METODI PER IL CALCOLO DI AUTOVALORI E AUTOVETTORI

1. Teoremi di localizzazione

Poiché gli autovalori di una matrice A sono gli zeri del suo polinomio caratteristico $P(\lambda) = \det(A - \lambda I)$, il calcolo numerico degli autovalori può essere effettuato applicando un qualsiasi metodo di iterazione funzionale (metodo delle tangenti, delle secanti, ecc.) all'equazione $P(\lambda) = 0$. Questo modo di procedere può essere conveniente se sono disponibili dei metodi efficienti per il calcolo del valore che la funzione $P(\lambda)$ (ed eventualmente la sua derivata prima) assume in un punto, come nel caso che la matrice abbia alcune proprietà di struttura (si veda il paragrafo 6 per il caso in cui la matrice A sia tridiagonale). Un'altra possibilità potrebbe essere quella di calcolare i coefficienti del polinomio caratteristico e poi applicare un metodo numerico per la risoluzione dell'equazione $P(\lambda) = 0$. Anche se il calcolo dei coefficienti del polinomio caratteristico ha lo stesso costo asintotico della moltiplicazione di matrici, cioè $O(n^{2.38})$, questo modo di procedere non è conveniente essenzialmente per due motivi: da una parte il costo computazionale rimane comunque elevato anche per valori grandi di n, dall'altra perché gli errori di arrotondamento generati nel calcolo dei coefficenti di $P(\lambda)$ possono indurre elevate variazioni degli zeri del polinomio. Quest'ultimo inconveniente non si presenta nel caso di matrici hermitiane, considerando gli autovalori come funzioni degli elementi della matrice (si veda il paragrafo 2). È comunque evidente che in generale il calcolo numerico degli autovalori di una matrice può essere fatto solamente con un procedimento iterativo. In questo capitolo verranno descritti i principali metodi numerici per il calcolo degli autovalori di una matrice.

Inizialmente verranno esposti alcuni teoremi di *localizzazione* che permettono di determinare facilmente sottoinsiemi del piano complesso in cui si trovano gli autovalori. Di questi teoremi i più importanti sono i teoremi di Gershgorin 2.35, 2.37 e 2.38, che per la loro generalità e semplicità di applicazione sono stati anticipati nel secondo capitolo.

6.1 Teorema (di Hirsch). *Sia $A \in \mathbf{C}^{n\times n}$ e sia $\| \cdot \|$ una qualsiasi norma matriciale indotta. Allora il cerchio*

$$\{\ z \in \mathbf{C}\ :\ |z| \leq \|A\|\ \}$$

contiene tutti gli autovalori di A.

Dim. La tesi segue dal teorema 3.10. ∎

6.2 Teorema. *Siano $A \in \mathbf{C}^{n\times n}$ normale, $\mathbf{x} \in \mathbf{C}^n, \mathbf{x} \neq \mathbf{0}$ e f una funzione razionale definita su un sottoinsieme del piano complesso contenente gli autovalori di A. Allora esiste almeno un autovalore λ di A tale che*

$$|f(\lambda)| \leq \frac{\|f(A)\mathbf{x}\|_2}{\|\mathbf{x}\|_2}. \tag{1}$$

Dim. Poiché A è normale, per il teorema 2.28 esiste una matrice unitaria U tale che

$$A = UDU^H,$$

dove D è la matrice diagonale il cui i-esimo elemento principale è uguale a λ_i, autovalore di A, per $i = 1, \ldots, n$. Quindi

$$\|A\mathbf{x}\|_2 = \|UDU^H\mathbf{x}\|_2 = \|DU^H\mathbf{x}\|_2,$$

in quanto U è unitaria. Posto $\mathbf{y} = U^H\mathbf{x}$, è $\|\mathbf{y}\|_2 = \|\mathbf{x}\|_2$ e

$$\frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \frac{\|D\mathbf{y}\|_2}{\|\mathbf{y}\|_2} = \sqrt{\frac{\sum_{j=1}^{n} |d_{jj}y_j|^2}{\sum_{j=1}^{n} |y_j|^2}} \geq \sqrt{\frac{\min_{i=1,\ldots,n} |\lambda_i|^2 \sum_{j=1}^{n} |y_j|^2}{\sum_{j=1}^{n} |y_j|^2}} = \min_{i=1,\ldots,n} |\lambda_i|.$$

Per la (20) del capitolo 2 è

$$f(A) = Uf(D)U^H$$

e quindi

$$\frac{\|f(A)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \frac{\|Uf(D)U^H\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \frac{\|f(D)\mathbf{y}\|_2}{\|\mathbf{y}\|_2} \geq \min_{i=1,\ldots,n} |f(\lambda_i)|. \quad ∎$$

La (1) può essere utilizzata per determinare delle maggiorazioni a posteriori dell'errore che si commette approssimando gli autovalori di una matrice normale.

Sia ad esempio σ un'approssimazione di un autovalore di una matrice A normale e sia $\mathbf{x}$ un vettore che approssima l'autovettore corrispondente. Allora se

$$f(z) = z - \sigma,$$

dalla (1) si ha che esiste un autovalore λ di A tale che

$$|\lambda - \sigma| \leq \frac{\|(A - \sigma I)\mathbf{x}\|_2}{\|\mathbf{x}\|_2}; \tag{2}$$

se invece

$$f(z) = \frac{z - \sigma}{z},$$

e A è non singolare, dalla (1) si ha che esiste un autovalore λ di A tale che

$$\left|\frac{\lambda - \sigma}{\lambda}\right| \leq \frac{\|(A - \sigma I)A^{-1}\mathbf{x}\|_2}{\|\mathbf{x}\|_2};$$

e ponendo $A^{-1}\mathbf{x} = \mathbf{z}$ è

$$\left|\frac{\lambda - \sigma}{\lambda}\right| \leq \frac{\|(A - \sigma I)\mathbf{z}\|_2}{\|A\mathbf{z}\|_2}. \tag{3}$$

La (2) e la (3) danno una stima facilmente calcolabile dell'errore assoluto e relativo che si commette assumendo σ come approssimazione dell'autovalore λ, e possono essere anche usate come criterio di arresto per i metodi iterativi che approssimano gli autovalori.

Utilizzando il teorema 6.2 si dimostra un teorema di localizzazione in cui interviene il *quoziente di Rayleigh* di una matrice A relativo ad un vettore $\mathbf{x} \neq \mathbf{0}$:

$$r_A(\mathbf{x}) = \frac{\mathbf{x}^H A\mathbf{x}}{\mathbf{x}^H\mathbf{x}}. \tag{4}$$

6.3 Teorema (di Weinstein). *Siano $A \in \mathbf{C}^{n\times n}$ normale e $\mathbf{x} \in \mathbf{C}^n, \mathbf{x} \neq \mathbf{0}$. Allora esiste almeno un autovalore λ di A nel cerchio*

$$\left\{ z \in \mathbf{C} \ : \ |z - r_A(\mathbf{x})| \leq \sqrt{\frac{\mathbf{x}^H A^H A\mathbf{x}}{\mathbf{x}^H\mathbf{x}} - |r_A(\mathbf{x})|^2} \right\}.$$

Dim. Dal teorema 6.2, ponendo

$$f(z) = z - r_A(\mathbf{x}),$$

risulta che esiste un autovalore λ di A tale che

$$|\lambda - r_A(\mathbf{x})| \leq \frac{\|[A - r_A(\mathbf{x})I]\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \sqrt{\frac{\mathbf{x}^H[A^H - \overline{r_A(\mathbf{x})}\ I]\ [A - r_A(\mathbf{x})\ I]\ \mathbf{x}}{\mathbf{x}^H\mathbf{x}}}$$

$$= \sqrt{\frac{\mathbf{x}^H A^H A\mathbf{x} - r_A(\mathbf{x})\mathbf{x}^H A^H\mathbf{x} - \overline{r_A(\mathbf{x})}\mathbf{x}^H A\mathbf{x} + |r_A(\mathbf{x})|^2\mathbf{x}^H\mathbf{x}}{\mathbf{x}^H\mathbf{x}}},$$

da cui, poiché $\dfrac{\mathbf{x}^H A^H\mathbf{x}}{\mathbf{x}^H\mathbf{x}} = \overline{r_A(\mathbf{x})}$, segue che

$$|\lambda - r_A(\mathbf{x})| \leq \sqrt{\frac{\mathbf{x}^H A^H A\mathbf{x}}{\mathbf{x}^H\mathbf{x}} - |r_A(\mathbf{x})|^2}.$$

∎

2. Teoremi di perturbazione

In questo paragrafo, in modo analogo a quanto fatto nel primo paragrafo del capitolo 4 per il problema della risoluzione dei sistemi lineari, si studia 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. Tali risultati permettono di valutare l'errore inerente del problema del calcolo degli autovalori, generato dalla rappresentazione dei dati con un numero finito di cifre.

6.4 Teorema (di Bauer-Fike). *Sia $\|\ .\ \|$ una norma matriciale indotta che verifichi la seguente proprietà*

$$\|D\| = \max_{i=1,\ldots,n} |d_{ii}|$$

per ogni matrice diagonale $D \in \mathbf{C}^{n\times n}$ (una tale norma viene detta norma assoluta, le norme $\|\ .\ \|_1$, $\|\ .\ \|_2$ e $\|\ .\ \|_\infty$ sono assolute). Sia $A \in \mathbf{C}^{n\times n}$ una matrice diagonalizzabile, cioè tale che

$$A = TDT^{-1},$$

con D diagonale e T non singolare. Se $\delta A \in \mathbf{C}^{n\times n}$ e ξ è un autovalore di $A + \delta A$, allora esiste almeno un autovalore λ di A tale che

$$|\lambda - \xi| \leq \mu(T)\ \|\delta A\|,$$

dove $\mu(T) = \|T\|\ \|T^{-1}\|$.

Dim. Se ξ fosse autovalore di A, la tesi sarebbe verificata. Altrimenti la matrice $A - \xi I$ risulta non singolare e dalla relazione

$$(A + \delta A)\mathbf{y} = \xi\mathbf{y},$$

dove $\mathbf{y}$ è autovettore di $A + \delta A$, si ha

$$\delta A\mathbf{y} = -(A - \xi I)\mathbf{y},$$

da cui

$$(A - \xi I)^{-1}\delta A\mathbf{y} = -\mathbf{y}$$

e quindi

$$\|(A - \xi I)^{-1}\delta A\| \geq 1. \tag{5}$$

Poiché

$$(A - \xi I)^{-1} = T(D - \xi I)^{-1}T^{-1},$$

si ha dalla (5)

$$1 \leq \|T(D - \xi I)^{-1}T^{-1}\delta A\| \leq \|T\|\ \|T^{-1}\|\ \|(D - \xi I)^{-1}\|\ \|\delta A\|,$$

e poiché $\|\ .\ \|$ è una norma assoluta, ne segue

$$1 \leq \mu(T)\frac{1}{\min\limits_{i=1,\ldots,n}|\lambda_i - \xi|}\ \|\delta A\|, \tag{6}$$

in cui i λ_i, $i = 1, \ldots, n$, sono gli autovalori di A e quindi gli elementi principali di D. Dalla (6) segue che

$$\min_{i=1,\ldots,n}|\lambda_i - \xi| \leq \mu(T)\|\delta A\|,$$

da cui la tesi. ∎

Il teorema 6.4 esprime un risultato di perturbazione: perturbando gli elementi di una matrice A, gli autovalori cambiano al più proporzionalmente all'entità della perturbazione δA. Il condizionamento del problema del calcolo degli autovalori di A è legato al numero di condizionamento della matrice T le cui colonne sono gli autovettori di A: quindi il problema del calcolo degli autovalori è tanto meglio condizionato quanto più basso è il numero di condizionamento $\mu(T)$. Se A è una matrice normale, allora T è unitaria, per cui $\mu_2(T) = 1$ e dal teorema 6.4 si ha

$$|\lambda - \xi| \leq \|\delta A\|_2,$$

ossia il problema del calcolo degli autovalori per matrici normali è ben condizionato per tutti gli autovalori.

Per matrici non normali il problema del calcolo degli autovalori può essere ben condizionato o mal condizionato, a seconda delle proprietà dell'autovalore considerato. Per questa ragione è bene analizzare il comportamento del problema per un singolo autovalore, distinguendo il caso di un autovalore di molteplicità algebrica uno dal caso di un autovalore di molteplicità algebrica maggiore di uno.

6.5 Teorema. *Sia $A \in \mathbf{C}^{n\times n}$, λ un autovalore di A di molteplicità algebrica uno, $\mathbf{x}, \mathbf{y} \in \mathbf{C}^n$, $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = 1$, tali che*

$$A\mathbf{x} = \lambda\mathbf{x}$$

$$\mathbf{y}^H A = \lambda \mathbf{y}^H.$$

Allora è $\mathbf{y}^H\mathbf{x} \neq 0$ ed inoltre per ogni $F \in \mathbf{C}^{n\times n}$ esiste nel piano complesso un intorno V dello zero e una funzione $\lambda(\epsilon) : V \to \mathbf{C}$, analitica, tale che

a) $\lambda(\epsilon)$ *è autovalore con molteplicità algebrica uno di* $A + \epsilon F$,

b) $\lambda(0) = \lambda$,

c) $\lambda'(0) = \dfrac{\mathbf{y}^H F\mathbf{x}}{\mathbf{y}^H\mathbf{x}}$,

d) a meno dei termini di ordine superiore in ϵ *è*

$$\lambda(\epsilon) - \lambda = \epsilon\, \frac{\mathbf{y}^H F\mathbf{x}}{\mathbf{y}^H\mathbf{x}}\ .$$

Per la dimostrazione si veda [26] (si veda anche l'esercizio 6.9). ∎

Anche in questo caso risulta che la variazione nell'autovalore dovuta alla perturbazione ϵF di A è proporzionale ad ϵ. Inoltre il condizionamento del problema dipende dalla quantità

$$\left|\frac{\mathbf{y}^H F\mathbf{x}}{\mathbf{y}^H\mathbf{x}}\right|,$$

che, data F, è tanto più grande quanto più piccolo è $|\mathbf{y}^H\mathbf{x}|$. Nel caso delle matrici normali è $\mathbf{y}^H\mathbf{x} = 1$, in accordo con i risultati del teorema 6.4.

Se λ è un autovalore di molteplicità algebrica $\sigma(\lambda) > 1$ e di molteplicità geometrica $\tau(\lambda)$, a cui corrispondono i blocchi di Jordan

$$C^{(1)}, C^{(2)}, \ldots, C^{(\tau(\lambda))},$$

di ordine massimo η, allora si può dimostrare che esiste un intorno V di zero e una costante $\gamma > 0$, tale che per $\epsilon \in V$ la matrice $A + \epsilon F$ ha autovalori $\lambda_i(\epsilon),\ i = 1, \ldots, \sigma(\lambda)$, tali che

$$|\lambda_i(\epsilon) - \lambda| \leq \gamma\, |\epsilon|^{1/\eta}.$$

Se $\eta > 1$, il problema del calcolo dell'autovalore λ può essere fortemente mal condizionato.

6.6 Esempio. Siano

$$A = \begin{bmatrix} \mu & 1 & 0 & 0 \\ 0 & \mu & 1 & 0 \\ 0 & 0 & \mu & 1 \\ 0 & 0 & 0 & \mu \end{bmatrix}, \quad F = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

allora

$$A + \epsilon F = \begin{bmatrix} \mu & 1 & 0 & 0 \\ 0 & \mu & 1 & 0 \\ 0 & 0 & \mu & 1 \\ \epsilon & 0 & 0 & \mu \end{bmatrix}.$$

Si ha

$$\det(A + \epsilon F - \lambda I) = (\mu - \lambda)^4 - \epsilon,$$

da cui risulta che gli autovalori λ_j di $A + \epsilon F$ soddisfano alla relazione

$$|\lambda_j - \mu| = \sqrt[4]{|\epsilon|}, \quad j = 1, \ldots, 4.$$

Quindi ad una perturbazione ϵ dell'elemento a_{41} corrisponde una variazione di modulo $\sqrt[4]{|\epsilon|}$ negli autovalori. Se ad esempio fosse $\epsilon = 10^{-8}$ (inferiore alla precisione di macchina quando si opera con 6 cifre significative esadecimali), risulterebbe

$$|\lambda_j - \mu| = 10^{-2}, \quad j = 1, \ldots, 4. \qquad \blacksquare$$

3. Caso delle matrici hermitiane

Nel caso delle matrici hermitiane il quoziente di Rayleigh (4) assume una notevole importanza, in quanto è legato a proprietà interessanti, utili anche per il calcolo. È opportuno richiamare alcune proprietà riguardanti i sottospazi introdotti nei paragrafi 2 e 6 del capitolo 1.

1 - Se S è un sottospazio di $\mathbf{C}^n$, allora il sottospazio ortogonale $S^\perp$ ha dimensione

$$\dim S^\perp = n - \dim S. \tag{7}$$

2 - Se S e T sono due sottospazi di $\mathbf{C}^n$, allora per il sottospazio $S \cap T$ vale

$$\dim(S \cap T) \geq \max \{0, \dim S + \dim T - n\}. \tag{8}$$

3 - Se $A \in \mathbf{C}^{m \times n}, m \leq n$, e S è un sottospazio di $\mathbf{C}^n$, allora per le dimensioni dei sottospazi

$$T = \{ \mathbf{x} \in \mathbf{C}^m \text{ tali che } \mathbf{x} = A\mathbf{y}, \mathbf{y} \in S \}$$

e

$$N = \{ \mathbf{x} \in S \text{ tali che } A\mathbf{x} = \mathbf{0} \}$$

vale la relazione

$$\dim T + \dim N = \dim S. \tag{9}$$

6.7 Teorema (di Courant-Fischer o del minimax). *Sia $A \in \mathbf{C}^{n\times n}$ una matrice hermitiana con autovalori*

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n.$$

Allora risulta

$$\lambda_{n-k+1} = \min_{V_k} \max_{\substack{\mathbf{x}\neq\mathbf{0} \\ \mathbf{x}\in V_k}} r_A(\mathbf{x}), \tag{10}$$

$$\lambda_k = \max_{V_k} \min_{\substack{\mathbf{x}\neq\mathbf{0} \\ \mathbf{x}\in V_k}} r_A(\mathbf{x}), \tag{11}$$

dove V_k è un qualunque sottospazio di $\mathbf{C}^n$ di dimensione k, per $k = 1, \ldots, n$.

Dim. Siano $\mathbf{x}_i, i = 1, \ldots, n$, autovettori ortonormali di A corrispondenti agli autovalori λ_i e, fissato un indice k, sia S il sottospazio di dimensione $n-k+1$ generato dagli $n-k+1$ vettori $\mathbf{x}_k, \ldots, \mathbf{x}_n$. Per la (8) è

$$\dim(S \cap V_k) \geq 1$$

e quindi l'intersezione fra S e V_k non può ridursi al solo vettore nullo. Sia allora

$$\mathbf{x} = \sum_{i=k}^{n} \alpha_i \mathbf{x}_i \neq \mathbf{0}$$

elemento di $S \cap V_k$. Poiché i vettori $\mathbf{x}_i$ sono ortonormali e vale

$$A\mathbf{x}_i = \lambda_i \mathbf{x}_i, \quad i = 1, \ldots, n,$$

allora

$$r_A(\mathbf{x}) = \frac{\mathbf{x}^H A\mathbf{x}}{\mathbf{x}^H\mathbf{x}} = \frac{\sum_{j=k}^{n} |\alpha_j|^2 \lambda_j}{\sum_{j=k}^{n} |\alpha_j|^2} \leq \frac{\max_{i=k,\ldots,n} \lambda_i \sum_{j=k}^{n} |\alpha_j|^2}{\sum_{j=k}^{n} |\alpha_j|^2} = \max_{i=k,\ldots,n} \lambda_i = \lambda_k.$$

Quindi si ha

$$\min_{\substack{\mathbf{x}\neq\mathbf{0} \\ \mathbf{x}\in V_k}} r_A(\mathbf{x}) \leq \lambda_k. \tag{12}$$

D'altra parte, se V_k è proprio lo spazio generato da $\mathbf{x}_1, \ldots, \mathbf{x}_k$, il vettore $\mathbf{x}_k$ è elemento di V_k e vale

$$r_A(\mathbf{x}_k) = \lambda_k,$$

quindi nella (12) vale il segno di uguaglianza, da cui segue la (11). Per dimostrare la (10) è sufficiente applicare la (11) alla matrice $-A$. ∎

Si osservi che dal teorema del minimax si ottiene in particolare

$$\lambda_1 = \max_{\mathbf{x}\neq\mathbf{0}} \; r_A(\mathbf{x}),$$

$$\lambda_n = \min_{\mathbf{x}\neq\mathbf{0}} \; r_A(\mathbf{x}).$$

Inoltre è facile verificare che se A è una matrice reale simmetrica, allora la funzione $f \; : \; \mathbf{R}^n \to \mathbf{R}$,

$$f(\mathbf{x}) = r_A(\mathbf{x})$$

è stazionaria nel punto $\mathbf{v} \in \mathbf{R}^n$ se e solo se

$$A\mathbf{v} = \lambda\mathbf{v}, \; \lambda = f(\mathbf{v}).$$

Dal teorema del minimax seguono i seguenti teoremi.

6.8 Teorema. *Sia $A \in \mathbf{C}^{n\times n}$ hermitiana, e $U \in \mathbf{C}^{n\times(n-1)}$, tale che $U^H U = I_{n-1}$. Sia inoltre $B = U^H AU$. Allora per gli autovalori λ_i, $i = 1, \ldots, n$, di A con $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, e μ_i, $i = 1, \ldots, n-1$, di B con $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_{n-1}$, vale la relazione*

$$\lambda_1 \geq \mu_1 \geq \lambda_2 \geq \mu_2 \geq \ldots \geq \mu_{n-1} \geq \lambda_n.$$

Tale proprietà viene anche espressa dicendo che gli autovalori di B separano gli autovalori di A.

Dim. Si dimostra dapprima che per $k = 2, \ldots, n$ è

$$\lambda_k \leq \mu_{k-1}. \tag{13}$$

Dalla (11) segue che esiste un sottospazio Z_k di $\mathbf{C}^n$ di dimensione k tale che

$$\lambda_k = \min_{\substack{\mathbf{x}\neq\mathbf{0} \\ \mathbf{x}\in Z_k}} \; r_A(\mathbf{x}). \tag{14}$$

Sia S il sottospazio di dimensione $n-1$ generato dalle colonne di U. Per la (8) è

$$\dim(Z_k \cap S) \geq k - 1,$$

per cui, poiché $k \geq 2$, esistono vettori $\mathbf{x} \neq \mathbf{0}$ appartenenti a $Z_k \cap S$ e quindi dalla (14)

$$\lambda_k \leq \min_{\substack{\mathbf{x}\neq\mathbf{0} \\ \mathbf{x}\in Z_k\cap S}} \; r_A(\mathbf{x}).$$

Per ogni vettore $\mathbf{x} \in S$, con $\mathbf{x} \neq \mathbf{0}$, esiste un solo vettore $\mathbf{y} \in \mathbf{C}^{n-1},\ \mathbf{y} \neq \mathbf{0}$, tale che

$$\mathbf{x} = U\mathbf{y},$$

per cui

$$\mathbf{y} = U^H\mathbf{x}.$$

Si considera allora il sottospazio

$$W = \{\ \mathbf{y} \in \mathbf{C}^{n-1} \text{ tali che } \mathbf{y} = U^H\mathbf{x},\ \mathbf{x} \in Z_k \cap S\ \}.$$

Poiché $U^HU = I$, il nucleo di U^H

$$N = \{\ \mathbf{x} \in S \text{ tali che } U^H\mathbf{x} = \mathbf{0}\ \}$$

ha dimensione nulla; per la (9) risulta

$$\dim W = \dim(Z_k \cap S) \geq k - 1$$

e quindi

$$\lambda_k \leq \min_{\substack{\mathbf{x}\neq\mathbf{0} \\ \mathbf{x}\in Z_k\cap S}} r_A(\mathbf{x}) = \min_{\substack{\mathbf{x}\neq\mathbf{0} \\ \mathbf{x}\in Z_k\cap S}} \frac{\mathbf{x}^HA\mathbf{x}}{\mathbf{x}^H\mathbf{x}} = \min_{\substack{\mathbf{y}\neq\mathbf{0} \\ \mathbf{y}\in W}} \frac{\mathbf{y}^HU^HAU\mathbf{y}}{\mathbf{y}^HU^HU\mathbf{y}}$$

$$= \min_{\substack{\mathbf{y}\neq\mathbf{0} \\ \mathbf{y}\in W}} \frac{\mathbf{y}^HB\mathbf{y}}{\mathbf{y}^H\mathbf{y}} \leq \max_{V_{k-1}} \min_{\substack{\mathbf{y}\neq\mathbf{0} \\ \mathbf{y}\in V_{k-1}}} r_B(\mathbf{y}) = \mu_{k-1},$$

dove V_{k-1} è un qualunque sottospazio di $\mathbf{C}^{n-1}$ di dimensione $k-1$, da cui segue la (13).

Applicando la (13) alle matrici $-A$ e $-B$, i cui autovalori sono

$$-\lambda_n \geq -\lambda_{n-1} \geq \ldots \geq -\lambda_1$$

e

$$-\mu_{n-1} \geq -\mu_{n-2} \geq \ldots \geq -\mu_1,$$

si ha

$$-\lambda_{n+1-k} \leq -\mu_{(n-1)+1-(k-1)},\ \text{per } k = 2, \ldots, n,$$

e quindi

$$\lambda_{n+1-k} \geq \mu_{n+1-k},\ \text{per } k = 2, \ldots, n,$$

cioè

$$\lambda_i \geq \mu_i,\ \text{per } i = 1, \ldots, n-1.$$

∎

6.9 Teorema. *Sia $A \in \mathbf{C}^{n\times n}$ hermitiana e per $m \leq n$ sia $U_m \in \mathbf{C}^{n\times m}$ tale che $U_m^H U_m = I_m$. Allora indicati con λ_1 e λ_n il massimo e il minimo autovalore di A e con μ_1 e μ_m il massimo e il minimo autovalore di $U_m^H A U_m$, valgono le relazioni*

$$\lambda_1 \geq \mu_1, \qquad \mu_m \geq \lambda_n.$$

Dim. Sia $B = U_m^H A U_m$. Si ha

$$\mu_1 = \max_{\substack{\mathbf{y}\neq\mathbf{0}\\ \mathbf{y}\in\mathbf{C}^m}} r_B(\mathbf{y}) = \max_{\substack{\mathbf{y}\neq\mathbf{0}\\ \mathbf{y}\in\mathbf{C}^m}} \frac{\mathbf{y}^H U_m^H A U_m \mathbf{y}}{\mathbf{y}^H\mathbf{y}} = \max_{\substack{\mathbf{x}\neq\mathbf{0}\\ \mathbf{x}=U_m\mathbf{y}\\ \mathbf{y}\in\mathbf{C}^m}} \frac{\mathbf{x}^H A\mathbf{x}}{\mathbf{x}^H\mathbf{x}}$$

$$\leq \max_{\substack{\mathbf{x}\neq\mathbf{0}\\ \mathbf{x}\in\mathbf{C}^n}} r_A(\mathbf{x}) = \lambda_1.$$

Analogamente si procede per μ_m. ∎

6.10 Teorema. *Sia $A \in \mathbf{C}^{n\times n}$ hermitiana, e sia A_k la sottomatrice principale di testa di ordine k di A. Allora gli autovalori di A_k separano gli autovalori di A_{k+1}, per $k = 1, \ldots, n-1$.*

Dim. Si osservi che se

$$U = \begin{bmatrix} I_k \\ \mathbf{0}^H \end{bmatrix} \begin{matrix} \} & k \text{ righe} \\ \} & 1 \text{ riga} \end{matrix}$$

allora $U^H U = I_k$ e $A_k = U^H A_{k+1} U$. La tesi segue dal teorema 6.8. ∎

6.11 Esempio. Come illustrazione del teorema 6.10 si consideri la matrice hermitiana

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 & 3 \\ 1 & 2 & 3 & 4 & 4 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}.$$

Gli autovalori delle sottomatrici A_k principali di testa di ordine k di A sono

k	λ_1	λ_2	λ_3	λ_4	λ_5
1	1				
2	2.618033	0.3819660			
3	5.048913	0.6431029	0.3079774		
4	8.290849	1.	0.4260219	0.2831172	
5	12.34352	1.448682	0.5829639	0.3532520	0.2715528

Gli autovalori sono stati calcolati con il metodo di Jacobi (si veda il paragrafo 9). ∎

Facendo ancora uso del teorema del minimax si dimostrano i seguenti risultati.

6.12 Teorema. *Sia* $\mathbf{u} \in \mathbf{C}^n, \sigma \in \mathbf{R}, \sigma \geq 0$, *e siano* $A, B \in \mathbf{C}^{n\times n}$ *hermitiane, tali che*

$$B = A + \sigma \mathbf{u}\mathbf{u}^H.$$

Per gli autovalori

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$$

di A, *e*

$$\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$$

di B, *vale la relazione*

$$\lambda_1 + \sigma \mathbf{u}^H\mathbf{u} \geq \mu_1 \geq \lambda_1 \geq \mu_2 \geq \lambda_2 \geq \ldots \geq \lambda_{n-1} \geq \mu_n \geq \lambda_n.$$

Dim. Se $\mathbf{u} = \mathbf{0}$, la tesi è banale; si suppone allora che $\mathbf{u} \neq \mathbf{0}$ e si dimostra che $\mu_i \geq \lambda_i$, per $i = 1, \ldots, n$. Dalla (10) si ha che per $k = 1, \ldots, n$, esiste un sottospazio Z_k di dimensione k, tale che

$$\mu_{n-k+1} = \max_{\substack{\mathbf{x}\neq\mathbf{0}\\ \mathbf{x}\in Z_k}} r_B(\mathbf{x}) = \max_{\substack{\mathbf{x}\neq\mathbf{0}\\ \mathbf{x}\in Z_k}} \frac{\mathbf{x}^H B\mathbf{x}}{\mathbf{x}^H\mathbf{x}} = \max_{\substack{\mathbf{x}\neq\mathbf{0}\\ \mathbf{x}\in Z_k}} \left[\frac{\mathbf{x}^H A\mathbf{x}}{\mathbf{x}^H\mathbf{x}} + \sigma\, \frac{|\mathbf{x}^H\mathbf{u}|^2}{\mathbf{x}^H\mathbf{x}}\right]$$

$$\geq \max_{\substack{\mathbf{x}\neq\mathbf{0}\\ \mathbf{x}\in Z_k}} \frac{\mathbf{x}^H A\mathbf{x}}{\mathbf{x}^H\mathbf{x}} \geq \min_{V_k} \max_{\substack{\mathbf{x}\neq\mathbf{0}\\ \mathbf{x}\in V_k}} r_A(\mathbf{x}) = \lambda_{n-k+1}.$$

Inoltre dalla (10) si ha che esiste un sottospazio W_k di dimensione k tale che

$$\lambda_{n-k+1} = \max_{\substack{\mathbf{x}\neq\mathbf{0}\\ \mathbf{x}\in W_k}} r_A(\mathbf{x}).$$

Sia S il sottospazio di $\mathbf{C}^n$ generato dal vettore $\mathbf{u}$ e sia $\mathbf{x}$ un vettore del sottospazio $T = W_k \cap S^\perp$, cioè tale che

$$\mathbf{x}^H\mathbf{u} = 0.$$

Allora è

$$\mathbf{x}^H B\mathbf{x} = \mathbf{x}^H A\mathbf{x} + \sigma|\mathbf{x}^H\mathbf{u}|^2 = \mathbf{x}^H A\mathbf{x}$$

e quindi

$$\lambda_{n-k+1} \geq \max_{\substack{\mathbf{x}\neq\mathbf{0}\\ \mathbf{x}\in T}} r_A(\mathbf{x}) = \max_{\substack{\mathbf{x}\neq\mathbf{0}\\ \mathbf{x}\in T}} r_B(\mathbf{x}). \tag{15}$$

Poiché $\dim S = 1$, per la (7) è $\dim S^{\perp} = n - 1$, e quindi per la (8) è $\dim T \geq k - 1$. Dalla (15) segue che

$$\lambda_{n-k+1} \geq \min_{V_{k-1}} \max_{\substack{\mathbf{x}\neq\mathbf{0} \\ \mathbf{x}\in V_{k-1}}} r_B(\mathbf{x}) = \mu_{n-k+2}.$$

Inoltre dal teorema del minimax si ottiene

$$\mu_1 = \max_{\mathbf{x}\neq\mathbf{0}} r_B(\mathbf{x}) = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^H B\mathbf{x}}{\mathbf{x}^H\mathbf{x}} = \max_{\mathbf{x}\neq\mathbf{0}} \left[\frac{\mathbf{x}^H A\mathbf{x}}{\mathbf{x}^H\mathbf{x}} + \sigma \frac{|\mathbf{x}^H\mathbf{u}|^2}{\mathbf{x}^H\mathbf{x}}\right],$$

e poiché per la disuguaglianza di Cauchy-Schwarz (1) cap. 1, è

$$\max_{\mathbf{x}\neq\mathbf{0}} \frac{|\mathbf{x}^H\mathbf{u}|^2}{\mathbf{x}^H\mathbf{x}} = \mathbf{u}^H\mathbf{u},$$

ne segue che

$$\mu_1 \leq \max_{\mathbf{x}\neq\mathbf{0}} r_A(\mathbf{x}) + \sigma\mathbf{u}^H\mathbf{u} = \lambda_1 + \sigma\mathbf{u}^H\mathbf{u}. \qquad \blacksquare$$

6.13 Esempio. La matrice

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 & 3 \\ 1 & 2 & 3 & 4 & 4 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}$$

ha gli autovalori $\lambda_1 = 12.34352$, $\lambda_2 = 1.448682$, $\lambda_3 = 0.5829639$, $\lambda_4 = 0.3532520$, $\lambda_5 = 0.2715528$ (si veda l'esempio 6.11).

Se $\sigma = 1$ e $\mathbf{u} = [1,\ \text{-}1,\ 0,\ \text{-}1,\ 0]^T$, la matrice $B = A + \sigma\mathbf{u}\mathbf{u}^T$ ha gli autovalori $\mu_1 = 12.96301$, $\mu_2 = 2.736083$, $\mu_3 = 1.448030$, $\mu_4 = 0.5076391$, $\mu_5 = 0.3451974$, che separano gli autovalori di A e che sono tali che $\mu_i > \lambda_i$, per $i = 1, \ldots, 5$, e $\mu_1 < \lambda_1 + \sigma\mathbf{u}^H\mathbf{u} = \lambda_1 + 3$.

Se $\sigma = -2$, la matrice $C = A + \sigma\mathbf{u}\mathbf{u}^T$ ha gli autovalori $\eta_1 = 11.65525$, $\eta_2 = 1.448380$, $\eta_3 = 0.5175053$, $\eta_4 = 0.3456874$, $\eta_5 = -4.966838$, che separano ancora gli autovalori di A e sono tali che $\eta_i < \lambda_i$, per $i = 1, \ldots, 5$, e $\eta_5 > \lambda_5 + \sigma\mathbf{u}^H\mathbf{u} = \lambda_5 - 6$. $\blacksquare$

6.14 Teorema. *Siano $A, B, C \in \mathbf{C}^{n\times n}$ hermitiane, tali che $C = A + B$. Per gli autovalori*

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$$

di A,

$$\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$$

di B,

$$\nu_1 \geq \nu_2 \geq \ldots \geq \nu_n$$

di C, vale la relazione

$$\lambda_k + \mu_n \leq \nu_k \leq \lambda_k + \mu_1, \quad k = 1, \ldots, n.$$

Dim. Dalla (11) si ha per $k = 1, \ldots, n$

$$\nu_k = \max_{V_k} \min_{\substack{\mathbf{x}\neq\mathbf{0} \\ \mathbf{x}\in V_k}} r_C(\mathbf{x}) = \max_{V_k} \min_{\substack{\mathbf{x}\neq\mathbf{0} \\ \mathbf{x}\in V_k}} [r_A(\mathbf{x}) + r_B(\mathbf{x})]$$

$$\geq \max_{V_k} \min_{\substack{\mathbf{x}\neq\mathbf{0} \\ \mathbf{x}\in V_k}} [r_A(\mathbf{x}) + \mu_n] = \max_{V_k} \min_{\substack{\mathbf{x}\neq\mathbf{0} \\ \mathbf{x}\in V_k}} r_A(\mathbf{x}) + \mu_n$$

$$= \lambda_k + \mu_n.$$

L'altra relazione si ricava applicando lo stesso procedimento alla matrice $A = C - B$, per cui si ottiene

$$\lambda_k \geq \nu_k + (-\mu_1).$$ ∎

6.15 Esempio. Si considerino le due matrici A e $B \in \mathbf{R}^{5\times 5}$ simmetriche a banda

$$A = \begin{bmatrix} 6 & -4 & 0 & 0 & 0 \\ -4 & 6 & -4 & 0 & 0 \\ 0 & -4 & 6 & -4 & 0 \\ 0 & 0 & -4 & 6 & -4 \\ 0 & 0 & 0 & -4 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

Gli autovalori di A sono dati da

$$\lambda_k = 6 + 8\cos\frac{k\pi}{6}, \ k = 1, \ldots, 5$$

(si veda l'esercizio 2.40), cioè $\lambda_1 = 6 + 4\sqrt{3} = 12.92820,$

$$\lambda_2 = 10, \ \lambda_3 = 6, \ \lambda_4 = 2, \ \lambda_5 = 6 - 4\sqrt{3} = -0.9282032.$$

La matrice B ha il polinomio caratteristico

$$p(\lambda) = -\lambda\,(\lambda^2 - 1)\,(\lambda^2 - 2)$$

e quindi ha gli autovalori

$$\mu_1 = \sqrt{2} = 1.414214, \ \mu_2 = 1, \ \mu_3 = 0, \ \mu_4 = -1, \ \mu_5 = -\sqrt{2} = -1.414214.$$

La matrice $C = A + B$ è una matrice pentadiagonale i cui autovalori sono

$$\nu_1 = 14.10892, \nu_2 = 9.531118, \nu_3 = 4.678975, \nu_4 = 1.468864, \nu_5 = 0.2120767$$

e soddisfano le disuguaglianze

$$\lambda_k - \sqrt{2} < \nu_k < \lambda_k + \sqrt{2}, \quad k = 1, \ldots, 5. \qquad \blacksquare$$

Il teorema 6.14 fornisce anche un risultato di perturbazione. Se A e $B \in \mathbf{C}^{n\times n}$ sono matrici hermitiane e $C = A + \epsilon B,\ \epsilon > 0$, allora per gli autovalori $\nu_1, \ldots, \nu_n$ di C vale la relazione

$$\lambda_k + \epsilon\mu_n \leq \nu_k \leq \lambda_k + \epsilon\mu_1,$$

dove $\lambda_1, \ldots, \lambda_n$ sono gli autovalori di A e $\mu_1, \ldots, \mu_n$ sono gli autovalori di B, ordinati in ordine non crescente. Cioè la variazione sugli autovalori della matrice perturbata C è proporzionale all'entità della perturbazione ϵB.

6.16 Esempio. Sia

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 & 3 \\ 1 & 2 & 3 & 4 & 4 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}$$

la matrice dell'esempio 6.11 e sia $B \in \mathbf{R}^{5\times 5}$ tale che $b_{ij} = 1$ per $i,\ j = 1, \ldots, 5$. Scegliendo per ϵ i valori $10^{-r},\ r = 1, \ldots, 4$, si ottengono per la matrice perturbata $C = A + \epsilon B$ gli autovalori riportati nella seguente tabella.

ϵ	ν_1	ν_2	ν_3	ν_4	ν_5
10^{-1}	12.78558	1.491278	0.5942757	0.3566023	0.2722294
10^{-2}	12.38752	1.453030	0.5841669	0.3536236	0.2716302
10^{-3}	12.34792	1.449116	0.5830847	0.3532901	0.2715611
10^{-4}	12.34396	1.448725	0.5829763	0.3532561	0.2715544

$\blacksquare$

4. Introduzione ai metodi

Nei prossimi paragrafi vengono presentati metodi numerici per calcolare gli autovalori e gli autovettori di una matrice. Fra i diversi metodi considerati alcuni hanno carattere generale e sono convenientemente applicabili a matrici dense e senza stuttura, altri utilizzano in modo specifico eventuali proprietà di struttura o sparsità della matrice, permettendo di trattare problemi anche con dimensioni molto elevate. Alcuni dei metodi esposti possono essere utilizzati per calcolare tutti gli autovalori e autovettori di una matrice, altri invece servono per calcolare solo alcuni autovalori, per esempio quelli che si trovano all'estremità dello spettro, ed i corrispondenti autovettori, come è richiesto in molte applicazioni.

I metodi per il calcolo degli autovalori possono essere divisi in due classi.

1) Metodi in cui il calcolo viene effettuato in due fasi: riduzione con metodi diretti della matrice A in una matrice simile B, di cui sia più agevole calcolare gli autovalori, e calcolo degli autovalori di B con un metodo iterativo. Questi metodi si applicano in generale a problemi di piccole dimensioni, per i quali tutti i dati su cui si opera possono essere contenuti nella memoria centrale del calcolatore.

2) Metodi completamente iterativi, che richiedono ad ogni passo la moltiplicazione di una matrice per un vettore, o la risoluzione di un sistema lineare. Questi metodi si applicano in generale a problemi di grandi dimensioni, anche nel caso in cui non sia possibile contenere tutti i dati nella memoria centrale del calcolatore.

Nei metodi della prima classe per la riduzione della matrice A nella matrice B si utilizzano metodi diretti analoghi a quelli descritti per la fattorizzazione delle matrici. Nel caso più generale la matrice B che si ottiene è tale che

$$b_{ij} = 0, \quad \text{per} \quad i > j + 1,\ i,\ j = 1, \ldots, n.$$

Una matrice B con questa proprietà è detta essere *in forma di Hessenberg superiore*. Se la matrice A è hermitiana, e la trasformazione viene eseguita con matrici unitarie, la matrice B risulta hermitiana e tridiagonale.

Se $B = T^{-1}AT$ e A è diagonalizzabile, cioè $A = SDS^{-1}$, dove D è la matrice diagonale i cui elementi principali sono gli autovalori di A, allora anche B è diagonalizzabile e risulta

$$B = (T^{-1}S)D(T^{-1}S)^{-1}.$$

La matrice $T^{-1}S$ ha per colonne gli autovettori di B. Poiché per il teorema 6.4 il condizionamento del problema del calcolo degli autovalori di una matrice diagonalizzabile è legato al numero di condizionamento della matrice

degli autovettori, è opportuno determinare la matrice T in modo tale che il numero di condizionamento di $T^{-1}S$ sia minore o uguale al numero di condizionamento di S. Ciò è senz'altro vero se $\mu(T) = \|T\| \, \|T^{-1}\| = 1$, in tal caso infatti

$$\begin{aligned}\mu(T^{-1}S) &= \|T^{-1}S\| \, \|(T^{-1}S)^{-1}\| \\ &\leq \|T\| \, \|T^{-1}\| \, \|S\| \, \|S^{-1}\| = \mu(T) \, \mu(S) = \mu(S).\end{aligned}$$

In generale conviene utilizzare trasformazioni per similitudine in cui $\mu(T)$ sia piccolo.

La trasformazione per similitudine della matrice A nella matrice B è fatta per passi successivi

$$A^{(k+1)} = T_k^{-1} A^{(k)} T_k, \;\; k = 1, 2, \ldots, m-1, \tag{16}$$

dove

$$A^{(1)} = A \quad \text{e} \quad A^{(m)} = B,$$

per cui, posto $T = T_1 T_2 \ldots T_{m-1}$, risulta $B = T^{-1}AT$, e se $\mathbf{x}$ è autovettore di B, $T\mathbf{x}$ è autovettore di A.

Le matrici T_k sono di solito matrici elementari di Gauss o di Householder oppure matrici di Givens. Se T_k è una matrice di Householder o di Givens, risulta

$$\|T_k\|_2 \, \|T_k^{-1}\|_2 = 1,$$

se T_k è una matrice di Gauss con elementi non principali di modulo minore o uguale ad 1 (massimo pivot per colonne), risulta

$$\|T_k\|_\infty \, \|T_k^{-1}\|_\infty \leq 4.$$

I metodi iterativi per il calcolo degli autovalori di B potrebbero essere anche applicati direttamente alla matrice A. Trasformando però prima la matrice A nella matrice B, si abbassa notevolmente il numero delle operazioni richieste da ogni iterazione (ad esempio per il metodo QR, descritto nel paragrafo 8, si passa da un numero di operazioni dell'ordine n^3 ad uno dell'ordine di n^2).

Per il calcolo degli autovalori della matrice B, due sono le tecniche più usate:

a) se sono richiesti solo pochi autovalori rispetto alla dimensione della matrice (non più del 25%), conviene usare un metodo iterativo che calcoli un singolo autovalore per volta, come ad esempio un metodo di iterazione funzionale applicato all'equazione caratteristica o il metodo delle potenze inverse (paragrafo 11). È questo il modo migliore di

procedere per matrici hermitiane tridiagonali o per matrici in forma di Hessenberg superiore e sparse;

b) se sono richiesti tutti o molti degli autovalori, il metodo migliore è in generale il QR (paragrafo 8).

I metodi della seconda classe sono basati sul calcolo di successioni di vettori del tipo $\mathbf{x}_{k+1} = B\mathbf{x}_k$, dove la matrice B può essere, a seconda del metodo considerato, la A, la A^{-1} oppure una matrice $(A-\alpha I)^{-1}$. In questo modo ad ogni passo viene effettuata sempre la stessa trasformazione sul vettore corrente $\mathbf{x}_k$. Se la matrice B ha particolari proprietà di struttura o di sparsità questa trasformazione può essere fatta senza dover memorizzare tutti gli elementi della matrice nella memoria principale. Questi metodi sono particolarmente adatti a problemi di grosse dimensioni con matrici sparse, quando si richiede il calcolo di un numero limitato di autovalori e autovettori. Se la matrice A ha proprietà di struttura, ad esempio è una matrice a banda, è possibile ridurre il numero delle operazioni richieste ad ogni passo sfruttando queste proprietà. Un metodo classico, molto semplice, appartenente a questa classe è il metodo delle potenze (paragrafo 10) che approssima l'autovalore di modulo massimo e il corrispondente autovettore. Opportune varianti del metodo delle potenze consentono di calcolare anche altri autovalori e autovettori della matrice. In particolare la variante delle potenze inverse di Wielandt (paragrafo 11) è la più usata per calcolare l'autovettore corrispondente ad un autovalore di cui è nota un'approssimazione. Fra i metodi di questa seconda classe il metodo di Lanczos (paragrafo 13) è il più importante per calcolare gli autovalori di matrici reali, simmetriche e sparse, di grosse dimensioni.

Un metodo iterativo classico che non appartiene alle due classi sopra descritte è il metodo di Jacobi (paragrafo 9), che con successive trasformazioni unitarie costruisce una successione di matrici che converge a una matrice diagonale e consente quindi di calcolare tutti gli autovalori e gli autovettori contemporaneamente.

5. Riduzione di una matrice hermitiana in forma tridiagonale: i metodi di Householder, di Givens e di Lanczos

Una matrice hermitiana può essere trasformata in una matrice tridiagonale hermitiana mediante trasformazioni per similitudine unitarie utilizzando le matrici di Householder o quelle di Givens, oppure con il procedimento di Lanczos.

a) *Metodo di Householder.*

Sia $A \in \mathbf{C}^{n\times n}$ una matrice hermitiana; si considerino le trasformazioni (16), con $m = n-1$, in cui le matrici T_k siano matrici elementari di Householder (hermitiane e unitarie):

$$T_k = I - \beta_k \mathbf{u}_k \mathbf{u}_k^H,$$

costruite in modo che nella matrice $T_k A^{(k)}$ siano nulli tutti gli elementi della k-esima colonna, con l'indice di riga maggiore di $k+1$.

Al primo passo, posto

$$A^{(1)} = A = \begin{bmatrix} a_{11}^{(1)} & \mathbf{a}_1^H \\ \mathbf{a}_1 & B^{(1)} \end{bmatrix} \begin{matrix} \} & 1 \text{ riga} \\ \} & n-1 \text{ righe} \end{matrix}$$

si consideri la matrice elementare di Householder $P^{(1)} \in \mathbf{C}^{(n-1)\times(n-1)}$ tale che

$$P^{(1)}\mathbf{a}_1 = \alpha_1 \mathbf{e}_1,$$

dove $\mathbf{e}_1$ è il primo vettore della base canonica di $\mathbf{C}^{n-1}$. La matrice

$$T_1 = \begin{bmatrix} 1 & \mathbf{0}^H \\ \mathbf{0} & P^{(1)} \end{bmatrix},$$

è tale che nella matrice

$$A^{(2)} = T_1^{-1} A^{(1)} T_1 = T_1 A^{(1)} T_1$$

sono nulli tutti gli elementi della prima colonna con indice di riga maggiore di due e dei simmetrici elementi della prima riga.

Al k-esimo passo la sottomatrice principale di testa di ordine $k+1$ di $A^{(k)}$ risulta tridiagonale hermitiana e $A^{(k)}$ ha la forma

$$A^{(k)} = \begin{bmatrix} C^{(k)} & \mathbf{b}_k & O \\ \mathbf{b}_k^H & a_{kk}^{(k)} & \mathbf{a}_k^H \\ O & \mathbf{a}_k & B^{(k)} \end{bmatrix} \begin{matrix} \} & k-1 \text{ righe} \\ \} & 1 \text{ riga} \\ \} & n-k \text{ righe} \end{matrix}$$

dove $C^{(k)} \in \mathbf{C}^{(k-1)\times(k-1)}$ è tridiagonale hermitiana e $\mathbf{b}_k \in \mathbf{C}^{k-1}$ ha nulle le prime $k-2$ componenti. Sia $P^{(k)} \in \mathbf{C}^{(n-k)\times(n-k)}$ la matrice di Householder tale che

$$P^{(k)}\mathbf{a}_k = \alpha_k \mathbf{e}_1,$$

dove $\mathbf{e}_1$ è il primo vettore della base canonica di $\mathbf{C}^{n-k}$. Posto

$$T_k = \begin{bmatrix} I_k & \mathbf{0}^H \\ \mathbf{0} & P^{(k)} \end{bmatrix}$$

risulta

$$A^{(k+1)} = T_k^{-1} A^{(k)} T_k = T_k A^{(k)} T_k = \begin{bmatrix} C^{(k)} & \mathbf{b}_k & O \\ \mathbf{b}_k^H & a_{kk}^{(k)} & \mathbf{a}_k^H P^{(k)} \\ O & P^{(k)} \mathbf{a}_k & P^{(k)} B^{(k)} P^{(k)} \end{bmatrix}.$$

Poiché il vettore $P^{(k)}\mathbf{a}_k \in \mathbf{C}^{n-k}$ ha nulle le componenti di indice maggiore o uguale a due, la sottomatrice principale di testa di ordine $k+2$ della matrice $A^{(k+1)}$ è tridiagonale hermitiana. Applicando il procedimento $n-2$ volte si ottiene la matrice $B = A^{(n-1)}$ tridiagonale hermitiana.

Per calcolare $P^{(k)}B^{(k)}P^{(k)}$ non si utilizza esplicitamente la matrice $P^{(k)}$, ma si procede in modo analogo a quanto fatto nella risoluzione dei sistemi lineari, sfruttando inoltre il fatto che $B^{(k)}$ è una matrice hermitiana. Poiché

$$\begin{aligned} P^{(k)}B^{(k)}P^{(k)} &= (I - \beta_k \mathbf{u}_k \mathbf{u}_k^H) B^{(k)} (I - \beta_k \mathbf{u}_k \mathbf{u}_k^H) \\ &= B^{(k)} - \beta_k B^{(k)} \mathbf{u}_k \mathbf{u}_k^H - \beta_k \mathbf{u}_k \mathbf{u}_k^H B^{(k)} + \beta_k^2 \mathbf{u}_k (\mathbf{u}_k^H B^{(k)} \mathbf{u}_k) \mathbf{u}_k^H \\ &= B^{(k)} - [\beta_k B^{(k)} \mathbf{u}_k - \tfrac{1}{2}\beta_k (\mathbf{u}_k^H \beta_k B^{(k)} \mathbf{u}_k) \mathbf{u}_k] \mathbf{u}_k^H \\ &\quad - \mathbf{u}_k [\beta_k \mathbf{u}_k^H B^{(k)} - \tfrac{1}{2}\beta_k (\mathbf{u}_k^H \beta_k B^{(k)} \mathbf{u}_k) \mathbf{u}_k^H], \end{aligned}$$

ponendo

$$\mathbf{r}_k = \beta_k B^{(k)} \mathbf{u}_k$$

e

$$\mathbf{q}_k = \mathbf{r}_k - \tfrac{1}{2}\beta_k (\mathbf{r}_k^H \mathbf{u}_k) \mathbf{u}_k,$$

si ha:

$$P^{(k)}B^{(k)}P^{(k)} = B^{(k)} - \mathbf{q}_k \mathbf{u}_k^H - \mathbf{u}_k \mathbf{q}_k^H.$$

La trasformazione $A^{(k)} \to A^{(k+1)}$ richiede allora solo $2(n-k)^2$ operazioni moltiplicative. Il metodo di Householder per tridiagonalizzare una matrice hermitiana richiede dunque

$$\sum_{k=1}^{n-2} 2(n-k)^2 \simeq \frac{2}{3}\, n^3 \quad \text{operazioni moltiplicative.}$$

Il metodo di Householder per la riduzione di una matrice hermitiana in forma tridiagonale gode delle stesse proprietà di stabilità del corrispondente metodo per la risoluzione dei sistemi lineari. È infatti possibile dimostrare che la matrice B' effettivamente calcolata è simile ad una matrice "vicina" ad A, cioè che esiste una matrice ortogonale Q e una matrice E tali che

$$B' = Q^H(A + E)Q,$$

con $\|E\|_F \leq \gamma u \|A\|_F$, in cui $\gamma = \gamma(n)$ è una funzione polinomiale di n di grado basso e u è la precisione di macchina con cui sono stati effettuati i calcoli [28].

b) *Metodo di Givens.*

Sia $A \in \mathbf{C}^{n\times n}$ una matrice hermitiana; si considerino le trasformazioni (16), in cui le matrici T_k siano matrici di Givens:

$$T_k = G_{pq} = \begin{bmatrix} 1 & & & & & & & & & & \\ & \ddots & & & & & & & & & \\ & & 1 & & & & & & & & \\ & & & c & & & & -\overline{s} & & & \leftarrow p \\ & & & & 1 & & & & & & \\ & & & & & \ddots & & & & & \\ & & & & & & 1 & & & & \\ & & & s & & & & c & & & \leftarrow q \\ & & & & & & & & 1 & & \\ & & & & & & & & & \ddots & \\ & & & & & & & & & & 1 \end{bmatrix}$$
$$\begin{matrix} & & \uparrow & & \uparrow & & \\ & & p & & q & & \end{matrix}$$

in cui $c = \cos\phi$ e $s = \psi \sin\phi$, $\phi \in \mathbf{R}$ e $\psi \in \mathbf{C}$, con $|\psi| = 1$. È $G_{pq}^{-1} = G_{pq}^H$.

La matrice $A^{(k+1)} = G_{pq}^H A^{(k)} G_{pq}$ differisce da $A^{(k)}$ solo nella p-esima e q-esima riga e nella p-esima e q-esima colonna. Infatti, considerando per semplicità il caso in cui $A^{(k)}$ è reale e indicando con a_{rj} gli elementi di $A^{(k)}$ e con $\hat{a}_{rj}$ gli elementi di $A^{(k+1)}$, si ha

$$\begin{aligned}
&\hat{a}_{rp} = \hat{a}_{pr} = ca_{pr} + sa_{qr}, \qquad r \neq p,\, q; \quad r = 1, 2, \ldots, n,\\
&\hat{a}_{rq} = \hat{a}_{qr} = -sa_{pr} + ca_{qr}, \quad r \neq p,\, q; \quad r = 1, 2, \ldots, n,\\
&\hat{a}_{pp} = c^2 a_{pp} + s^2 a_{qq} + 2csa_{pq},\\
&\hat{a}_{qq} = c^2 a_{qq} + s^2 a_{pp} - 2csa_{pq},\\
&\hat{a}_{qp} = \hat{a}_{pq} = (c^2 - s^2)a_{pq} - cs(a_{pp} - a_{qq}),\\
&\hat{a}_{rj} = a_{rj}, \quad \text{altrimenti.}
\end{aligned} \tag{17}$$

La scelta di G_{pq} può essere effettuata in modo che per un assegnato valore di r risulti

$$\hat{a}_{qr} = \hat{a}_{rq} = 0.$$

Infatti se $a_{qr} = 0$, è sufficiente porre $G_{pq} = I$, se invece $a_{qr} \neq 0$, si possono ricavare c ed s dalla condizione

$$\hat{a}_{qr} = -sa_{pr} + ca_{qr} = 0,$$

ossia

$$c = \frac{a_{pr}}{\sqrt{a_{pr}^2 + a_{qr}^2}} \quad \text{ed} \quad s = \frac{a_{qr}}{\sqrt{a_{pr}^2 + a_{qr}^2}}\ ,$$

da cui, procedendo come nel paragrafo 16 del capitolo 4, si ottengono le formule più stabili:

$$\text{se } |a_{pr}| \geq |a_{qr}|, \text{ allora si pone } t = \frac{a_{qr}}{a_{pr}},\ c = \frac{1}{\sqrt{1+t^2}},\ s = tc,$$
$$\text{altrimenti si pone } t = \frac{a_{pr}}{a_{qr}},\ s = \frac{1}{\sqrt{1+t^2}},\ c = ts.$$

Il processo completo di riduzione in forma tridiagonale, richiede $m = \dfrac{(n-1)(n-2)}{2}$ trasformazioni, con la seguente scelta di indici (p, q) ed r:

$$\begin{array}{llll} (2,3) & (2,4) & \dots & (2,n) \quad \text{con } r = 1 \\ & (3,4) & \dots & (3,n) \quad \text{con } r = 2 \\ & & \ddots & \vdots \\ & & & (n-1,n) \quad \text{con } r = n-2. \end{array}$$

Ogni trasformazione $A^{(k)} \to A^{(k+1)}$ richiede $4(n-r)$ operazioni moltiplicative, e per ogni r il numero di trasformazioni richieste è $n - r$. In totale la riduzione di una matrice hermitiana in forma tridiagonale con il metodo di Givens richiede

$$\sum_{r=1}^{n-2} 4(n-r)^2 \simeq \frac{4}{3}\, n^3 \quad \text{operazioni moltiplicative.}$$

Dal punto di vista della stabilità numerica, il comportamento del metodo di Givens è analogo a quello del metodo di Householder. Il numero delle operazioni richieste dal metodo di Householder risulta inferiore a quello richiesto dal metodo di Givens. Però il metodo di Givens è più adatto a sfruttare l'eventuale presenza di elementi nulli nella matrice A e in quelle

generate durante il metodo, ed è quindi possibile in questi casi che il metodo di Givens richieda meno operazioni del metodo di Householder.

6.17 Esempio. Si consideri la matrice $A \in \mathbf{R}^{4\times 4}$

$$A = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 3 & 4 & 3 & 2 \\ 2 & 3 & 4 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix}.$$

Applicando il metodo di Householder, al primo passo si ottiene

$$\beta_1 = 0.03964327$$

$$\mathbf{u}_1 = [0,\ 6.741657,\ 2,\ 1]^T,$$

e quindi

$$A^{(2)} = \begin{bmatrix} 4 & \text{-}3.741657 & 0 & 0 \\ \text{-}3.741657 & 8.285713 & \text{-}1.301424 & \text{-}2.254283 \\ 0 & \text{-}1.301424 & 1.070671 & 0.9112844 \\ 0 & \text{-}2.254283 & 0.9112844 & 2.643615 \end{bmatrix};$$

al secondo passo si ottiene

$$\beta_2 = 0.09839517$$

$$\mathbf{u}_2 = [0,\ 0,\ \text{-}3.904409,\ \text{-}2.254283]^T,$$

e quindi

$$A^{(3)} = \begin{bmatrix} 4 & \text{-}3.741657 & 0 & 0 \\ \text{-}3.741657 & 8.285713 & 2.602978 & 0 \\ 0 & 2.602978 & 3.039586 & \text{-}0.2253981 \\ 0 & 0 & \text{-}0.2253981 & 0.6746988 \end{bmatrix}.$$

Applicando il metodo di Givens, al primo passo si pone $r = 1,\ p = 2,\ q = 3$ e si ottiene $c = 0.8320504,\ s = 0.5547003$ e quindi

$$A^{(2)} = \begin{bmatrix} 4 & 3.605551 & 0 & 1 \\ 3.605551 & 6.769231 & 1.153846 & 3.328201 \\ 0 & 1.153846 & 1.230768 & 1.386751 \\ 1 & 3.328201 & 1.386751 & 4 \end{bmatrix};$$

al secondo passo si pone $r = 1,\ p = 2,\ q = 4$ e si ottiene $c = 0.9636238,\ s = 0.2672612$ e quindi

$$A^{(3)} = \begin{bmatrix} 4 & 3.741654 & 0 & 0 \\ 3.741654 & 8.285707 & 1.482497 & 2.139555 \\ 0 & 1.482497 & 1.230768 & 1.027927 \\ 0 & 2.139555 & 1.027927 & 2.483513 \end{bmatrix};$$

al terzo passo si pone $r = 2,\ p = 3,\ q = 4$ e si ottiene $c = 0.5695390,\ s = 0.8219641$ e quindi

$$A^{(4)} = \begin{bmatrix} 4 & 3.741654 & 0 & 0 \\ 3.741654 & 8.285707 & 2.602977 & 0 \\ 0 & 2.602977 & 3.039581 & 0.2254009 \\ 0 & 0 & 0.2254009 & 0.6746972 \end{bmatrix}.$$

Si noti che la matrice $A^{(4)}$, non tenendo conto degli errori di arrotondamento, è uguale a quella ottenuta con il metodo di Householder, a meno di una matrice di fase reale, cioè, detta $H^{(3)}$ la matrice ottenuta con il metodo di Householder, è

$$H^{(3)} = D^{-1}A^{(4)}D,$$

dove D è una matrice diagonale con elementi principali uguali a 1 o a -1. ∎

c) *Metodo di Lanczos.*

Sia $A \in \mathbf{C}^{n\times n}$ una matrice hermitiana e $Q \in \mathbf{C}^{n\times n}$ una matrice unitaria le cui colonne sono $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$, tali che

$$Q^H AQ = T = \begin{bmatrix} \alpha_1 & \beta_1 & & \\ \beta_1 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \beta_{n-1} \\ & & \beta_{n-1} & \alpha_n \end{bmatrix}, \tag{18}$$

dove $\alpha_i \in \mathbf{R}$ per $i = 1, \ldots, n$, e $\beta_i \in \mathbf{R}$, $\beta_i \geq 0$ per $i = 1, \ldots, n-1$. Il metodo di Lanczos permette di generare, a partire dalla prima colonna $\mathbf{q}_1$ di Q, attraverso un processo di ortogonalizzazione, le rimanenti colonne di Q e gli elementi di T. Infatti scrivendo la (18) come

$$AQ = QT,$$

e confrontando le i-esime colonne, per $i = 1, 2, \ldots, n$, a primo e a secondo membro si ottengono le relazioni

$$\begin{aligned} A\mathbf{q}_1 &= \alpha_1\mathbf{q}_1 + \beta_1\mathbf{q}_2, \\ A\mathbf{q}_i &= \beta_{i-1}\mathbf{q}_{i-1} + \alpha_i\mathbf{q}_i + \beta_i\mathbf{q}_{i+1}, \quad i = 2, \ldots, n-1, \\ A\mathbf{q}_n &= \beta_{n-1}\mathbf{q}_{n-1} + \alpha_n\mathbf{q}_n, \end{aligned} \tag{19}$$

Sfruttando il fatto che i vettori $\mathbf{q}_i$ sono ortonormali, se $\beta_i \neq 0$, per $i = 1, \ldots, n-1$, si ottengono le relazioni

$$\begin{aligned}
&\alpha_1 = \mathbf{q}_1^H A \mathbf{q}_1, \quad \mathbf{q}_2 = \frac{(A - \alpha_1 I)\mathbf{q}_1}{\beta_1}, \quad \beta_1 = \|(A - \alpha_1 I)\mathbf{q}_1\|_2 \\
&\alpha_i = \mathbf{q}_i^H A \mathbf{q}_i, \quad \mathbf{q}_{i+1} = \frac{(A - \alpha_i I)\mathbf{q}_i - \beta_{i-1}\mathbf{q}_{i-1}}{\beta_i}, \\
&\qquad\qquad \beta_i = \|(A - \alpha_i I)\mathbf{q}_i - \beta_{i-1}\mathbf{q}_{i-1}\|_2, \quad i = 2, \ldots, n-1, \\
&\alpha_n = \mathbf{q}_n^H A \mathbf{q}_n,
\end{aligned} \tag{20}$$

che permettono di calcolare gli elementi di T e le colonne di Q se tutti i β_i sono non nulli. Se uno dei β_i fosse nullo, il procedimento può proseguire solo conoscendo il vettore $\mathbf{q}_{i+1}$.

Il procedimento di Lanczos comunque si può applicare a partire da un qualunque vettore $\mathbf{u} \in \mathbf{C}^n$, tale che $\|\mathbf{u}\|_2 = 1$, scegliendo, se uno dei β_i risultasse nullo, come $\mathbf{q}_{i+1}$ un qualsiasi vettore ortonormale ai vettori $\mathbf{q}_j$ già calcolati. Il procedimento può quindi essere portato a termine in ogni caso. I vettori $\mathbf{q}_i$ così calcolati sono ortonormali. Vale infatti il seguente teorema.

6.18 Teorema. *Sia $\mathbf{u} \in \mathbf{C}^n$, tale che $\|\mathbf{u}\|_2 = 1$. Scelto $\mathbf{q}_1 = \mathbf{u}$, i vettori $\mathbf{q}_1$ e $\mathbf{q}_i$, $i = 2, \ldots, n$, calcolati con la (20) (se $\beta_i = 0$ si sceglie come $\mathbf{q}_{i+1}$ un qualunque vettore ortogonale a $\mathbf{q}_1, \ldots, \mathbf{q}_i$, con $\|\mathbf{q}_{i+1}\|_2 = 1$), sono ortonormali. La matrice Q avente per colonne i vettori $\mathbf{q}_1, \ldots, \mathbf{q}_n$, e gli α_i, $i = 1, \ldots, n$, e i β_i, $i = 1, \ldots, n-1$, verificano la (18). Inoltre se i β_i sono tutti non nulli, la matrice Q così ottenuta è l'unica matrice per cui vale la (18) e tale che $Q\mathbf{e}_1 = \mathbf{u}$.*

Dim. I vettori $\mathbf{q}_1, \ldots, \mathbf{q}_n$ verificano la relazione $\|\mathbf{q}_i\|_2 = 1$ per costruzione. Per dimostrarne l'ortogonalità si procede per induzione. Si suppone che i vettori $\mathbf{q}_1, \ldots, \mathbf{q}_k$ siano ortogonali e si dimostra che $\mathbf{q}_{k+1}$ è ortogonale a $\mathbf{q}_1, \ldots, \mathbf{q}_k$. Per $k = 2$, i vettori $\mathbf{q}_1$ e $\mathbf{q}_2$ sono ortogonali per costruzione. Se $\beta_k = 0$ l'ortogonalità è verificata per costruzione. Altrimenti basta dimostrare che $\mathbf{q}_{k+1}$ è ortogonale a $\mathbf{q}_j$, $j = 1, \ldots, k-2$, perché $\mathbf{q}_{k+1}$ è ortogonale a $\mathbf{q}_k$ e $\mathbf{q}_{k-1}$ per le (20) e per l'ipotesi induttiva. Si ha dalla (19) per $j = 1, \ldots, k-2$,

$$\beta_k \mathbf{q}_{k+1}^H \mathbf{q}_j = \mathbf{q}_k^H A \mathbf{q}_j - \beta_{k-1}\mathbf{q}_{k-1}^H \mathbf{q}_j - \alpha_k \mathbf{q}_k^H \mathbf{q}_j,$$

e per l'ipotesi induttiva

$$\beta_k \mathbf{q}_{k+1}^H \mathbf{q}_j = \mathbf{q}_k^H A \mathbf{q}_j.$$

Poichè per la (19) è

$$A\mathbf{q}_j = \beta_{j-1}\mathbf{q}_{j-1} + \alpha_j\mathbf{q}_j + \beta_j\mathbf{q}_{j+1},$$

si ha

$$\mathbf{q}_k^H A\mathbf{q}_j = \beta_{j-1}\mathbf{q}_k^H\mathbf{q}_{j-1} + \alpha_j\mathbf{q}_k^H\mathbf{q}_j + \beta_j\mathbf{q}_k^H\mathbf{q}_{j+1},$$

e per l'ipotesi induttiva

$$\beta_k\mathbf{q}_{k+1}^H\mathbf{q}_j = 0,$$

da cui, poiché $\beta_k \neq 0$, segue che $\mathbf{q}_{k+1}^H\mathbf{q}_j = 0$. Per dimostrare che vale la (18) è sufficiente verificare che valgono le (19). Le prime $n-1$ relazioni in (19) sono verificate per costruzione dai vettori $\mathbf{q}_i$. L'ultima delle relazioni (19) è verificata perché il vettore

$$\mathbf{v} = A\mathbf{q}_n - \beta_{n-1}\mathbf{q}_{n-1} - \alpha_n\mathbf{q}_n$$

risulta nullo essendo ortogonale a $\mathbf{q}_1, \ldots, \mathbf{q}_n$. Infatti $\mathbf{q}_n^H\mathbf{v} = 0$ per la definizione di α_n e per $j = 1, \ldots, n-1$ è $\mathbf{q}_j^H\mathbf{v} = 0$, poiché

$$\begin{aligned}\mathbf{q}_j^H\mathbf{v} &= \mathbf{q}_j^H(A\mathbf{q}_n - \beta_{n-1}\mathbf{q}_{n-1} - \alpha_n\mathbf{q}_n) = \mathbf{q}_j^H A\mathbf{q}_n - \beta_{n-1}\mathbf{q}_j^H\mathbf{q}_{n-1}\\ &= (\beta_{j-1}\mathbf{q}_{j-1} + \alpha_j\mathbf{q}_j + \beta_j\mathbf{q}_{j+1})^H\mathbf{q}_n - \beta_{n-1}\mathbf{q}_j^H\mathbf{q}_{n-1}\\ &= \beta_j\mathbf{q}_{j+1}^H\mathbf{q}_n - \beta_{n-1}\mathbf{q}_j^H\mathbf{q}_{n-1}\end{aligned}$$

e quest'ultima relazione è nulla anche per $j = n-1$.

L'unicità della decomposizione (18) nel caso in cui $\beta_i \neq 0$ per $i = 1, \ldots, n-1$, segue dal fatto che la (18) e le (20) sono equivalenti se $\beta_i > 0$. ∎

In pratica se si genera solo la matrice T e non la matrice Q, il procedimento di Lanczos può essere implementato utilizzando solamente due vettori. Se il numero delle operazioni moltiplicative richieste dal prodotto della matrice A per un vettore è dato da hn (se A è una matrice piena è $h = n$, mentre se A è sparsa è $h << n$), il costo computazionale ad ogni passo è dato da $(5+h)n$ operazioni moltiplicative. Se A non è sparsa, il costo computazionale totale di questo metodo è dell'ordine di n^3 operazioni moltiplicative (quindi superiore a quello dei metodi di Householder e di Givens), se A è una matrice a banda, con $2p+1$ diagonali, allora il costo totale è di $(2p+6)n^2$ operazioni moltiplicative. Quindi tale metodo sembra essere molto indicato per matrici sparse e di dimensioni molto grandi.

Il metodo di tridiagonalizzazione di Lanczos presenta grossi problemi di stabilità numerica: infatti se uno dei β_i è piccolo, nel calcolo di $\mathbf{q}_{i+1}$ si possono presentare elevati errori di cancellazione, con una conseguente perdita

di ortogonalità dei vettori calcolati successivamente. Anche per questo motivo il metodo di Lanczos non è competitivo con i metodi di Givens e di Householder e non viene abitualmente usato per tridiagonalizzare matrici di dimensioni tali da poter essere contenute nella memoria del calcolatore. Il metodo di Lanczos risulta però particolarmente utile per il calcolo degli autovalori estremi dello spettro di matrici (si veda il paragrafo 13).

6.19 Esempio. Si applica il metodo di Lanczos alla matrice $A \in \mathbf{R}^{4\times 4}$

$$A = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 3 & 4 & 3 & 2 \\ 2 & 3 & 4 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix},$$

già vista nell'esempio 6.17, assumendo come vettore iniziale il vettore $\mathbf{q}_1 = \mathbf{e}_1$. Si ha

$$\alpha_1 = 4, \quad \mathbf{q}_2 = \begin{bmatrix} 0 \\ 0.8017837 \\ 0.5345225 \\ 0.2672612 \end{bmatrix}, \quad \beta_1 = 3.741658,$$

$$\alpha_2 = 8.285710, \quad \mathbf{q}_3 = \begin{bmatrix} 0 \\ \text{-}0.4987068 \\ 0.3520293 \\ 0.7920648 \end{bmatrix}, \quad \beta_2 = 2.602981,$$

$$\alpha_3 = 3.039589, \quad \mathbf{q}_4 = \begin{bmatrix} 0.1041032\ 10^{-4} \\ 0.3293015 \\ \text{-}0.7683433 \\ 0.5488251 \end{bmatrix}, \quad \beta_3 = 0.2253997,$$

$$\alpha_4 = 0.6746987.$$

La matrice tridiagonale così ottenuta risulta quindi

$$T = \begin{bmatrix} 4 & 3.741658 & 0 & 0 \\ 3.741658 & 8.285710 & 2.602981 & 0 \\ 0 & 2.602981 & 3.039589 & 0.2253997 \\ 0 & 0 & 0.2253997 & 0.6746987 \end{bmatrix}.$$

Se si sceglie $\mathbf{q}_1 = \frac{1}{2}\,[1\ ,\ 1\ ,\ 1\ ,\ 1]^T$, si ottiene

$$\alpha_1 = 11, \quad \mathbf{q}_2 = \frac{1}{2}\,[-1\ ,\ 1\ ,\ 1\ ,\ -1]^T, \quad \beta_1 = 1, \quad \alpha_2 = 1, \quad \beta_2 = 0.$$

A questo punto, per poter proseguire occorre scegliere un vettore $\mathbf{q}_3$ ortonormale a $\mathbf{q}_1$ e $\mathbf{q}_2$. Scegliendo $\mathbf{q}_3 = \frac{1}{2}\,[1\ ,\ 1\ ,\ -1\ ,\ -1]^T$ si ha

$$\alpha_3 = 3, \quad \mathbf{q}_4 = \frac{1}{2}\,[1\ ,\ -1\ ,\ 1\ ,\ -1]^T, \quad \beta_3 = 1, \quad \alpha_4 = 1.$$

Si è quindi ottenuta la seguente tridiagonalizzazione

$$A = QTQ^H,$$

dove

$$Q = \frac{1}{2}\begin{bmatrix} 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & -1 & -1 \end{bmatrix}, \quad T = \begin{bmatrix} 11 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \qquad \blacksquare$$

6. Calcolo degli autovalori delle matrici tridiagonali hermitiane con la successione di Sturm

Per calcolare gli autovalori di una matrice tridiagonale hermitiana conviene utilizzare metodi iterativi che facciano ricorso al polinomio caratteristico solo se il numero degli autovalori che si vogliono determinare è piccolo rispetto alle dimensioni della matrice. Sia $B_n \in \mathbf{C}^{n\times n}$ la matrice tridiagonale hermitiana definita da

$$B_n = \begin{bmatrix} \alpha_1 & \overline{\beta}_2 & & & \\ \beta_2 & \alpha_2 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \overline{\beta}_n \\ & & & \beta_n & \alpha_n \end{bmatrix}$$

e sia $P_n(\lambda) = \det(B_n - \lambda I)$ il suo polinomio caratteristico. Se la matrice B_n è riducibile, cioè se esiste almeno un indice $j, 2 \le j \le n$, tale che $\beta_j = 0$, allora il problema del calcolo degli autovalori di B_n è ricondotto al calcolo degli autovalori di due matrici di ordine inferiore. Infatti si ha

$$B_n = \begin{bmatrix} C_{j-1} & O \\ O & D_{n-j+1} \end{bmatrix},$$

in cui $C_{j-1} \in \mathbf{C}^{(j-1)\times(j-1)}, D_{n-j+1} \in \mathbf{C}^{(n-j+1)\times(n-j+1)}$ e quindi

$$\det(B_n - \lambda I) = \det(C_{j-1} - \lambda I_{j-1})\ \det(D_{n-j+1} - \lambda I_{n-j+1}).$$

Se le matrici C_{j-1} e D_{n-j+1} sono a loro volta riducibili, si procede in modo analogo.

Si consideri perciò il caso che B_n sia irriducibile cioè che $\beta_j \neq 0$ per $j = 2, 3, \ldots, n$. Calcolando $\det(B_n - \lambda I)$ con la regola di Laplace rispetto all'ultima riga, si ottengono le relazioni

$$\begin{aligned} &P_0(\lambda) = 1, \quad P_1(\lambda) = \alpha_1 - \lambda, \\ &P_i(\lambda) = (\alpha_i - \lambda)P_{i-1}(\lambda) - |\beta_i|^2 P_{i-2}(\lambda),\ i = 2, 3, \ldots, n, \end{aligned} \tag{21}$$

con cui è possibile calcolare il valore che il polinomio $P_n(\lambda)$ assume in un punto con $2(n-1)$ moltiplicazioni (supponendo di aver già calcolato $|\beta_i|^2,\ i = 2, 3, \ldots, n$).

6.20 Esempio. Si consideri la matrice $B_6 \in \mathbf{R}^{6\times 6}$ i cui elementi sono dati da:

$$b_{ij} = \begin{cases} 2 & \text{se } i = j, \\ 1 & \text{se } |i-j| = 1, \\ 0 & \text{altrimenti.} \end{cases}$$

Dalla (21) si ha

$$\begin{aligned} &P_0(\lambda) = 1, \quad P_1(\lambda) = 2 - \lambda, \\ &P_i(\lambda) = (2 - \lambda)P_{i-1}(\lambda) - P_{i-2}(\lambda),\ i = 2, 3, \ldots, 6, \end{aligned} \tag{22}$$

da cui

$$\begin{aligned} P_2(\lambda) &= \lambda^2 - 4\lambda + 3 \\ P_3(\lambda) &= -\lambda^3 + 6\lambda^2 - 10\lambda + 4 \\ P_4(\lambda) &= \lambda^4 - 8\lambda^3 + 21\lambda^2 - 20\lambda + 5 \\ P_5(\lambda) &= -\lambda^5 + 10\lambda^4 - 36\lambda^3 + 56\lambda^2 - 35\lambda + 6 \\ P_6(\lambda) &= \lambda^6 - 12\lambda^5 + 55\lambda^4 - 120\lambda^3 + 126\lambda^2 - 56\lambda + 7. \end{aligned}$$

Per calcolare il valore di $P_6(\lambda)$ in un punto sono richieste 5 moltiplicazioni e 6 addizioni sia con la relazione ricorrente (22) che con la regola di Ruffini-Horner che richiede però un lavoro preliminare per il calcolo dei coefficienti di $P_6(\lambda)$. Un aspetto particolarmente importante è che con la relazione ricorrente (22) si ottengono anche i valori $P_i(\lambda)$, $i = 1, \ldots, 5$, in un punto, che consentono di utilizzare un metodo semplice per calcolare gli zeri di $P_6(\lambda)$ (si veda il teorema 6.22).

La relazione ricorrente (22) e la regola di Ruffini-Horner possono generare errori algoritmici diversi per valori di λ vicini agli zeri di $P_6(\lambda)$. Ad esempio, per $\lambda = 3.8$ (uno zero di $P_6(\lambda)$ è 3.801973) si ottengono per $P_6(3.8)$ i valori

-0.3596973 con le (22) (sono esatte 4 cifre significative)

-0.3388882 con la regola di Ruffini (è esatta una sola cifra significativa). ▪

Gli autovalori di B_n vengono quindi calcolati risolvendo l'equazione caratteristica

$$P_n(\lambda) = 0, \tag{23}$$

con un metodo iterativo. Se si utilizza il metodo di Newton, il calcolo di $P'_n(\lambda)$ può essere fatto con le seguenti relazioni ricorrenti, ottenute derivando rispetto a λ entrambi i membri delle (21):

$$\begin{aligned} &P'_0(\lambda) = 0, \quad P'_1(\lambda) = -1, \\ &P'_i(\lambda) = (\alpha_i - \lambda)P'_{i-1}(\lambda) - P_{i-1}(\lambda) - |\beta_i|^2 P'_{i-2}(\lambda),\ i = 2, 3, \ldots, n. \end{aligned}$$

Quindi il rapporto $P_n(\lambda)/P'_n(\lambda)$, che interviene ad ogni passo del metodo di Newton, può essere calcolato con $4(n-1)$ moltiplicazioni e una divisione. Nel caso in cui si debba calcolare più di un autovalore, possono essere anche utilizzate delle tecniche di deflazione, quale la variante di *Maehly* della *deflazione implicita* (si veda [26]).

Per separare le radici di (23) conviene sfruttare le proprietà delle successioni di Sturm. Infatti nel seguente teorema si dimostra che i polinomi $P_i(\lambda)$ formano una successione di Sturm.

6.21 Teorema. *Se $\beta_i \neq 0$, per $i = 2, 3, \ldots, n$, la successione dei polinomi $P_i(\lambda)$, $i = 0, 1, \ldots, n$, verifica le seguenti proprietà:*

1) *$P_0(\lambda)$ non cambia segno;*
2) *se $P_i(\lambda) = 0$, allora $P_{i-1}(\lambda)P_{i+1}(\lambda) < 0$, per $i = 1, 2, \ldots, n-1$;*
3) *se $P_n(\lambda) = 0$, allora $P'_n(\lambda)P_{n-1}(\lambda) < 0$ (e quindi $P_n(\lambda)$ ha tutti zeri di molteplicità 1).*

Una successione di polinomi che verifica le proprietà 1), 2) e 3) è detta successione di Sturm.

Dim. La 1) è ovvia. Per la 2), si osservi che da (21) si ha $P_{i-1}(\lambda)P_{i+1}(\lambda) \leq 0$. Ma se fosse $P_{i-1}(\lambda)P_{i+1}(\lambda) = 0$ e $P_i(\lambda) = 0$, allora sarebbe $P_{i-1}(\lambda) = P_{i+1}(\lambda) = 0$, da cui, per ricorrenza, seguirebbe $P_0(\lambda) = 0$, che è assurdo. Da ciò segue anche che gli zeri $\lambda_j^{(n)}$, $j = 1, 2, \ldots, n$, di $P_n(\lambda)$ sono distinti dagli zeri $\lambda_j^{(n-1)}$, $j = 1, 2, \ldots, n-1$, di $P_{n-1}(\lambda)$ e quindi, per il teorema 6.10, gli zeri $\lambda_j^{(n-1)}$ separano strettamente gli zeri $\lambda_j^{(n)}$, cioè

$$\lambda_{j+1}^{(n)} < \lambda_j^{(n-1)} < \lambda_j^{(n)},\ j = 1, 2, \ldots, n-1.$$

Da questo fatto, tenendo presente che il coefficiente di λ^i in $P_i(\lambda)$ è $(-1)^i$, e quindi $\lim_{\lambda \to -\infty} P_i(\lambda) = +\infty$, segue la 3) (si veda la figura 6.1 per il caso $n = 4$). ∎

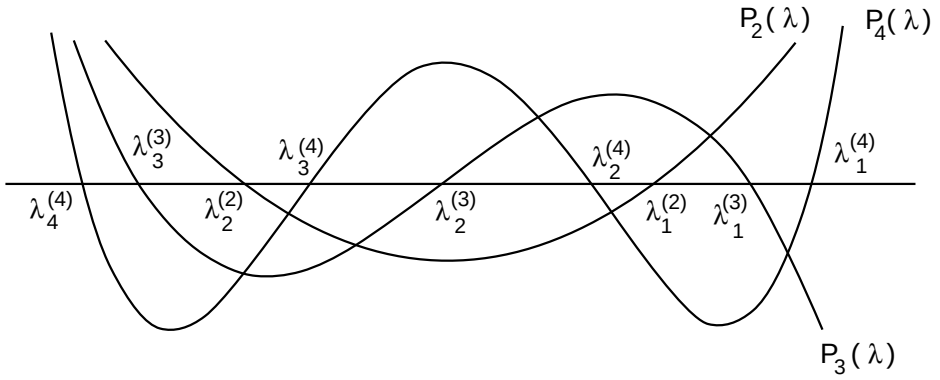

Fig. 6.1 - Grafico dei polinomi $P_2(\lambda)$, $P_3(\lambda)$ e $P_4(\lambda)$.

Si consideri, in un punto λ^*, la successione $P_0(\lambda^*), P_1(\lambda^*), \ldots, P_n(\lambda^*)$ (se fosse $P_i(\lambda^*) = 0$ per un indice $i \geq 1$, si attribuisca a tale valore il segno di $P_{i-1}(\lambda^*)$) e si indichi con $w(\lambda^*)$ il numero di cambiamenti di segno di tale successione. Vale il seguente teorema.

6.22 Teorema. *Se* $\{P_i(\lambda)\}, i = 0, 1, \ldots, n$, *è una successione di Sturm, il numero* $w(b) - w(a)$ *è uguale al numero di zeri di* $P_n(\lambda)$ *appartenenti all'intervallo* $[a, b)$.

Dim. Si faccia variare λ con continuità da a verso b. Si può avere una variazione nel numero $w(\lambda)$ solo quando λ incontra uno zero di uno dei polinomi $P_i(\lambda)$. Si consideri perciò un λ^* tale che $P_i(\lambda^*) = 0$ per un indice i. Per la proprietà 1) del teorema 6.21 deve essere $i \neq 0$. Si distinguono allora i due casi:

a) $i \neq n$. In questo caso, per la proprietà 2) del teorema 6.21 si ha

$$P_{i-1}(\lambda^*)P_{i+1}(\lambda^*) < 0.$$

Esiste perciò un numero h tale che nell'intervallo $[\lambda^* - h, \lambda^* + h]$ è ancora

$$P_{i-1}(\lambda)P_{i+1}(\lambda) < 0$$

e

$$P_i(\lambda) \neq 0,$$

eccetto che nel punto λ^*. Poiché per ogni $\lambda \in [\lambda^* - h, \lambda^* + h]$ i due polinomi $P_{i-1}(\lambda)$ e $P_{i+1}(\lambda)$ hanno segno discorde, $P_i(\lambda)$ deve avere in

questo intervallo segno concorde con uno dei due e discorde con l'altro. Quindi nella sequenza $P_{i-1}(\lambda), P_i(\lambda), P_{i+1}(\lambda)$ vi è una sola variazione di segno in tutto l'intervallo $[\lambda^* - h, \lambda^* + h]$, cioè il fatto che $P_i(\lambda)$ si annulli in λ^* non comporta variazioni del numero $w(\lambda)$.

b) $i = n$. In questo caso, poiché per la proprietà 3) del teorema 6.21 il polinomio $P_n(\lambda)$ ha radici semplici, la sua derivata $P'_n(\lambda)$ non si annulla in λ^* ed esiste un numero h tale che nell'intervallo $[\lambda^* - h, \lambda^* + h]$ $P'_n(\lambda)$ ha lo stesso segno che $P_n(\lambda)$ ha in $\lambda^* + h$ e segno opposto a quello che $P_n(\lambda)$ ha in $\lambda^* - h$. Se h è tale che nell'intervallo $[\lambda^* - h, \lambda^* + h]$ anche $P_{n-1}(\lambda)$ non si annulla, poiché per la proprietà 3) del teorema 6.21 $P_{n-1}(\lambda)$ ha segno opposto a quello di $P'_n(\lambda)$ per $\lambda \in [\lambda^* - h, \lambda^* + h]$, la sequenza $P_{n-1}(\lambda^* + h),\ P_n(\lambda^* + h)$ presenta una variazione di segno, mentre la sequenza $P_{n-1}(\lambda^* - h),\ P_n(\lambda^* - h)$ non presenta alcuna variazione di segno.

Se ne conclude che il numero di variazioni di segno in tutta la sequenza $P_0(\lambda), P_1(\lambda), \ldots, P_n(\lambda)$ può cambiare solo nei punti in cui si annulla $P_n(\lambda)$, ed esattamente aumenta di 1 ogni volta che si annulla $P_n(\lambda)$.

Nella tesi del teorema l'intervallo $[a, b)$ è aperto a destra perché se fosse $P_n(b) = 0$, poiché a $P_n(b)$ viene assegnato lo stesso segno assunto in b da $P_{n-1}(\lambda)$, che è diverso da zero in un intorno sinistro di b, $w(\lambda)$ non cambia in tale intorno. Perciò la radice b non altera il numero di variazioni di segno.

∎

Poiché

$$\lim_{\lambda \to -\infty} P_i(\lambda) = +\infty, \quad \text{per } i = 1, 2, \ldots, n,$$

esiste $\mu \in \mathbf{R}$ tale che per ogni $\lambda \leq \mu$ è $w(\lambda) = 0$. Quindi per ogni λ^* il numero di cambiamenti di segno $w(\lambda^*)$, per il teorema 6.22, fornisce il numero di autovalori di B_n minori di λ^*.

Sul teorema 6.22 è basato il seguente procedimento per calcolare il k-esimo autovalore λ_k di una matrice B_n tridiagonale, hermitiana e irriducibile, i cui autovalori sono $\lambda_1 > \lambda_2 > \ldots > \lambda_k > \ldots > \lambda_n$:

1) sia $[a_0, b_0)$ tale che $\lambda_k \in [a_0, b_0)$,

2) per $j = 0, 1, \ldots$, sia $\xi = \frac{1}{2}(a_j + b_j)$,
   se $w(\xi) \geq n - k + 1$, allora $a_{j+1} = a_j$ e $b_{j+1} = \xi$,
   se $w(\xi) < n - k + 1$, allora $a_{j+1} = \xi$ e $b_{j+1} = b_j$.

Questo procedimento, basato sul principio della bisezione, fornisce una successione di intervalli $[a_j, b_j)$ di ampiezza $2^{-j}(b_0 - a_0)$ che contengono λ_k ed è utile per separare gli autovalori di B_n, cioè per determinare intervalli che contengono un solo autovalore della matrice. Per l'effettiva approssimazione di un autovalore conviene in generale usare il metodo di Newton.

6.23 Esempio. Nel caso della successione di Sturm ottenuta nell'esempio 6.20 si ha

λ	$P_0(\lambda)$	$P_1(\lambda)$	$P_2(\lambda)$	$P_3(\lambda)$	$P_4(\lambda)$	$P_5(\lambda)$	$P_6(\lambda)$	$w(\lambda)$
0	1	2	3	4	5	6	7	0
1	1	1	0	-1	-1	0	1	2
2	1	0	-1	0	1	0	-1	3
3	1	-1	0	1	-1	0	1	4
4	1	-2	3	-4	5	-6	7	6

Dall'ultima colonna risulta che tutti gli autovalori sono positivi, che ve ne sono due nell'intervallo (0,1), uno nell'intervallo (1,2), uno nell'intervallo (2,3), due nell'intervallo (3,4). Poiché inoltre $w(0.5) = 1$ e $w(3.5) = 5$, risulta

$$0 < \lambda_6 < 0.5 < \lambda_5 < 1 < \lambda_4 < 2 < \lambda_3 < 3 < \lambda_2 < 3.5 < \lambda_1 < 4.$$

Se si vuole ridurre l'intervallo di separazione di λ_4, si può applicare l'algoritmo di bisezione all'intervallo [1,2], ottenendo per ξ e $w(\xi)$ successivamente i valori

ξ	$w(\xi)$
1.5	2
1.75	3
1.625	3
1.5625	3
1.53125	2
1.546875	2
1.554688	2
1.558594	3

da cui si ha che

$$1.554688 < \lambda_4 < 1.558594.$$

Per approssimare λ_1 si può applicare il metodo di Newton. Scegliendo come approssimazione iniziale il punto $x_0 = 4$, si ottiene la successione

i	x_i
1	3.875000
2	3.816162
3	3.802620
4	3.801939
5	3.801973
6	3.801973

■

7. Riduzione di una matrice in forma di Hessenberg superiore

Se applicati ad una matrice A non hermitiana, i metodi di Householder e di Givens, forniscono una matrice $B = T^{-1}AT$ in forma di Hessenberg superiore. Il costo computazionale è in questo caso di $5n^3/3$ operazioni moltiplicative per il metodo di Householder e di $10n^3/3$ operazioni moltiplicative per il metodo di Givens.

6.24 Esempio. Si consideri la matrice $A \in \mathbf{R}^{4\times 4}$

$$A = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 1 & 4 & 3 & 2 \\ 1 & 1 & 4 & 3 \\ 1 & 1 & 1 & 4 \end{bmatrix}.$$

Applicando il metodo di Householder, al primo passo si ottiene

$$\beta_1 = 0.2113248$$

$$\mathbf{u}_1 = [0,\ 2.732051,\ 1,\ 1]^T,$$

e quindi

$$A^{(2)} = \begin{bmatrix} 4 & \text{-}3.464098 & \text{-}0.3660240 & \text{-}1.366024 \\ \text{-}1.732050 & 7.666641 & \text{-}0.5446615 & \text{-}1.122009 \\ 0 & 0.08931351 & 1.877991 & 1.032692 \\ 0 & 1.244013 & \text{-}0.6993599 & 2.455341 \end{bmatrix};$$

al secondo passo si ottiene

$$\beta_2 = 0.5999027$$

$$\mathbf{u}_2 = [0,\ 0,\ 1.336528,\ 1.244013]^T,$$

e quindi

$$A^{(3)} = \begin{bmatrix} 4 & \text{-}3.464098 & 1.388726 & 0.2672625 \\ \text{-}1.732050 & 7.666641 & 1.158131 & 0.4629154 \\ 0 & \text{-}1.247213 & 2.476189 & \text{-}0.7423077 \\ 0 & 0 & 0.9897442 & 1.857142 \end{bmatrix},$$

che è in forma di Hessenberg superiore. Applicando invece il metodo di Givens, al primo passo si pone $r = 1,\ j = 2,\ p = 3$ e si ottiene

$$c = s = 0.7071069$$

e quindi

$$A^{(2)} = \begin{bmatrix} 4 & 3.535534 & -0.7071069 & 1 \\ 1.414213 & 6 & 1 & 3.535534 \\ 0 & -1 & 1 & 0.7071069 \\ 1 & 1.414213 & 0 & 4 \end{bmatrix};$$

al secondo passo si pone $r = 1,\ j = 2,\ p = 4$ e si ottiene

$$c = 0.8164967, \quad s = 0.5773506$$

e quindi

$$A^{(3)} = \begin{bmatrix} 4 & 3.464101 & -0.7071069 & -1.224745 \\ 1.732049 & 7.666669 & 0.8164975 & 0.9428082 \\ 0 & -0.4082471 & 2 & 1.154700 \\ 0 & -1.178512 & -0.5773511 & 2.333335 \end{bmatrix};$$

al terzo passo si pone $r = 2,\ j = 3,\ p = 4$ e si ottiene

$$c = 0.3273256, \quad s = 0.9449114$$

e quindi

$$A^{(4)} = \begin{bmatrix} 4 & 3.464101 & -1.388729 & 0.2672636 \\ 1.732049 & 7.666669 & 1.158131 & -0.4629126 \\ 0 & -1.247218 & 2.476188 & 0.7423077 \\ 0 & 0 & 0.9897417 & 1.857139 \end{bmatrix},$$

che risulta, non tenendo conto degli errori di arrotondamento, uguale, a meno di una matrice di fase reale, a quella ottenuta con il metodo di Householder. ∎

Per ridurre una matrice in forma di Hessenberg superiore attraverso trasformazioni per similitudine, si possono anche utilizzare le matrici elementari di Gauss. Per questioni di stabilità, analoghe a quelle già viste per il caso dei sistemi lineari, è necessario applicare il metodo con la tecnica del massimo pivot.

Al primo passo, posto

$$A^{(1)} = A = \begin{bmatrix} a_{11}^{(1)} & \mathbf{b}_1^H \\ \\ \mathbf{a}_1 & B^{(1)} \end{bmatrix} \begin{matrix} \} & 1 \text{ riga} \\ \\ \} & n-1 \text{ righe,} \end{matrix}$$

se $\mathbf{a}_1 = \mathbf{0}$, si pone $A^{(2)} = A^{(1)}$ e $T_1 = I$, altrimenti sia $\Pi_1 \in \mathbf{R}^{(n-1)\times(n-1)}$ una matrice di permutazione tale che il vettore $\mathbf{a}'_1 = \Pi_1 \mathbf{a}_1$ abbia come prima componente una componente di $\mathbf{a}_1$ di modulo massimo, e si consideri la matrice elementare di Gauss $M_1 \in \mathbf{C}^{(n-1)\times(n-1)}$ tale che il vettore

$$M_1 \mathbf{a}'_1 = \mathbf{a}''_1$$

abbia nulle tutte le componenti, esclusa la prima. Allora nella matrice

$$A^{(2)} = T_1 A^{(1)} T_1^{-1}, \quad T_1 = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & M_1 \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \Pi_1 \end{bmatrix},$$

sono nulli tutti gli elementi della prima colonna con indice di riga maggiore di due.

Al k-esimo passo, si supponga che $A^{(k)}$ abbia la struttura seguente

$$A^{(k)} = \begin{bmatrix} C^{(k)} & \mathbf{b}_k & D^{(k)} \\ \mathbf{c}_k^H & a_{kk}^{(k)} & \mathbf{d}_k^H \\ O & \mathbf{a}_k & B^{(k)} \end{bmatrix} \begin{matrix} \} \ k-1 \text{ righe} \\ \} \ 1 \text{ riga} \\ \} \ n-k \text{ righe,} \end{matrix}$$

dove $C^{(k)} \in \mathbf{C}^{(k-1)\times(k-1)}$ è in forma di Hessenberg superiore e $\mathbf{c}_k \in \mathbf{C}^{k-1}$ ha tutte le componenti nulle eccetto al più l'ultima. Se $\mathbf{a}_k = \mathbf{0}$, si pone $A^{(k+1)} = A^{(k)}$ e $T_k = I$, altrimenti sia $\Pi_k \in \mathbf{R}^{(n-k)\times(n-k)}$ una matrice di permutazione tale che il vettore $\mathbf{a}'_k = \Pi_k \mathbf{a}_k$ abbia come prima componente una componente di $\mathbf{a}_k$ di modulo massimo, e si consideri la matrice elementare di Gauss $M_k \in \mathbf{C}^{(n-k)\times(n-k)}$ tale che il vettore

$$M_k \mathbf{a}'_k = \mathbf{a}''_k$$

abbia nulle tutte le componenti, esclusa la prima. Allora la matrice

$$A^{(k+1)} = T_k A^{(k)} T_k^{-1}, \quad T_k = \begin{bmatrix} I_k & O \\ O & M_k \end{bmatrix} \begin{bmatrix} I_k & O \\ O & \Pi_k \end{bmatrix},$$

ha la struttura

$$A^{(k+1)} = \begin{bmatrix} C^{(k+1)} & \mathbf{b}_{k+1} & D^{(k+1)} \\ \mathbf{c}_{k+1}^H & a_{k+1,k+1}^{(k+1)} & \mathbf{d}_{k+1}^H \\ O & \mathbf{a}_{k+1} & B^{(k+1)} \end{bmatrix} \begin{matrix} \} \ k \text{ righe} \\ \} \ 1 \text{ riga} \\ \} \ n-k-1 \text{ righe.} \end{matrix}$$

Al termine del procedimento $A^{(n-1)}$ è in forma di Hessenberg superiore.

Per moltiplicare la matrice M_k per $\Pi_k B^{(k)}$ sono richieste $(n-k)^2$ operazioni moltiplicative, per moltiplicare la matrice $M_k \Pi_k B^{(k)} \Pi_k^T$ per M_k^{-1} sono richieste ancora $(n-k)^2$ operazioni moltiplicative e per moltiplicare la matrice $D^{(k)} \Pi_k^T$ per M_k^{-1} sono richieste $k(n-k)$ operazioni moltiplicative. Per trasformare la matrice $A^{(k)}$ nella matrice $A^{(k+1)}$ sono quindi richieste $(n-k)(2n-k)$ operazioni moltiplicative. Perciò per trasformare una matrice in forma di Hessenberg superiore, sono richieste $5n^3/6$ operazioni moltiplicative. Quindi il costo computazionale di questo metodo è inferiore a quello dei metodi di Householder e di Givens. Però con questo metodo possono presentarsi problemi di instabilità numerica, in particolare quando gli elementi delle matrici $A^{(k)}$ hanno modulo molto elevato rispetto agli elementi di A, come accade nel caso dei sistemi lineari. Può accadere infatti che il massimo dei moduli degli elementi di $A^{(k)}$ sia una funzione esponenziale di k. Se ciò accade, conviene utilizzare metodi di riduzione che fanno uso di matrici ortogonali (Householder e Givens) e che risultano più stabili.

6.25 Esempio. Facendo uso delle matrici elementari di Gauss, la matrice

$$A = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 1 & 4 & 3 & 2 \\ 1 & 1 & 4 & 3 \\ 1 & 1 & 1 & 4 \end{bmatrix}$$

dell'esempio 6.24 è trasformata successivamente nelle matrici

$$A^{(2)} = \begin{bmatrix} 4 & 6 & 2 & 1 \\ 1 & 9 & 3 & 2 \\ 0 & -1 & 1 & 1 \\ 0 & -3 & -2 & 2 \end{bmatrix}$$

$$A^{(3)} = \begin{bmatrix} 4 & 6 & \frac{5}{3} & 2 \\ 1 & 9 & 3 & 3 \\ 0 & -3 & \frac{4}{3} & -2 \\ 0 & 0 & \frac{8}{9} & \frac{5}{3} \end{bmatrix}.$$

La matrice $A^{(3)}$ è in forma di Hessenberg superiore e differisce naturalmente dalle due matrici ottenute con i metodi di Householder e di Givens. ∎

Anche per le matrici in forma di Hessenberg superiore è possibile calcolare il valore assunto in un punto dal polinomio caratteristico senza determinarne effettivamente i coefficienti, con il seguente metodo di *Hyman*.

Sia $A \in \mathbf{C}^{n\times n}$ in forma di Hessenberg superiore e irriducibile. Fissato un punto λ, si determinano un vettore $\mathbf{x}$ con l'ultima componente $x_n = 1$ e uno scalare γ, dipendenti da λ, tali che

$$(A - \lambda I)\mathbf{x} = \gamma \mathbf{e}_1, \tag{24}$$

nel modo seguente: si ricava x_{n-1} dall'ultima equazione e procedendo mediante sostituzioni all'indietro, si ricava infine x_1 dalla seconda equazione e γ dalla prima equazione. Il costo computazionale di questo procedimento è di $n^2/2$ operazioni moltiplicative. Poiché per la regola di Cramer risulta

$$x_n = \frac{(-1)^{n+1}}{\det(A - \lambda I)}\ \gamma a_{21}a_{32}\ldots a_{n,n-1},$$

essendo $x_n = 1$, si ha

$$P(\lambda) = \det(A - \lambda I) = (-1)^{n+1}\gamma a_{21}a_{32}\ldots a_{n,n-1}. \tag{25}$$

È possibile, in modo analogo, calcolare anche la derivata prima

$$P'(\lambda) = \frac{\mathrm{d}}{\mathrm{d}\lambda}\ \det(A - \lambda I).$$

Derivando entrambi i membri della (24) si ha

$$(A - \lambda I)\mathbf{x}' - \mathbf{x} = \gamma' \mathbf{e}_1,$$

da cui, attraverso il processo di sostituzione all'indietro, è possibile ricavare $\mathbf{x}'$ e γ' dopo avere calcolato $\mathbf{x}$ dal sistema (24). Dalla (25) si ha poi:

$$P'(\lambda) = (-1)^{n+1}\gamma' a_{21}a_{32}\ldots a_{n,n-1}.$$

8. Metodo QR per il calcolo degli autovalori

Il metodo QR è il metodo più usato per calcolare tutti gli autovalori di una matrice, in quanto è il più efficiente e può essere applicato anche a matrici non hermitiane. Il metodo è assai complicato, sia come descrizione che come implementazione, anche se il principio su cui si basa è semplice. Il metodo richiede tutta una serie di accorgimenti, senza i quali non potrebbe essere efficiente: riduzione preliminare della matrice in forma tridiagonale o di Hessenberg superiore, per ridurre il costo computazionale ad ogni iterazione; utilizzazione di una tecnica di traslazione per aumentare la velocità di convergenza; riduzione dell'ordine della matrice quando un autovalore è

stato approssimato con sufficiente precisione, per calcolare un altro autovalore.

Il metodo QR, che è stato descritto da Francis nel 1961, utilizza la fattorizzazione QR di una matrice; esso deriva da un precedente metodo, detto *metodo LR*, proposto da Rutishauser nel 1958, che utilizza la fattorizzazione LU di una matrice.

La descrizione del metodo si articola nei seguenti punti:

a) algoritmo di base,

b) teorema di convergenza,

c) costo computazionale e stabilità,

d) convergenza in ipotesi più deboli,

e) condizioni di arresto e riduzione dell'ordine della matrice,

f) tecnica di traslazione,

g) calcolo degli autovettori.

a) Algoritmo di base

Nel metodo QR viene generata una successione $\{A_k\}$ di matrici nel modo seguente: posto

$$A_1 = A,$$

per $k = 1, 2, \ldots$, si calcola una fattorizzazione QR di A_k

$$A_k = Q_k R_k, \tag{26}$$

dove Q_k è unitaria e R_k è triangolare superiore, e si definisce la matrice A_{k+1} per mezzo della relazione

$$A_{k+1} = R_k Q_k. \tag{27}$$

Da (26) e (27) risulta che

$$A_{k+1} = Q_k^H A_k Q_k, \tag{28}$$

e quindi le matrici della successione $\{A_k\}$ sono tutte simili fra di loro. Sotto opportune ipotesi la successione converge ad una matrice triangolare superiore (diagonale se A è hermitiana) che ha come elementi diagonali gli autovalori di A.

6.26 Esempio. Il metodo QR viene applicato alla matrice

$$A = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 3 & 4 & 3 & 2 \\ 2 & 3 & 4 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix}.$$

dell'esempio 6.17. Si ottiene

$$A_2 = \begin{bmatrix} 9.733320 & 2.834947 & 0.8783645 & -0.2318690 \\ 2.834947 & 3.783903 & 1.539931 & -0.6027983 \\ 0.8783645 & 1.539931 & 1.515015 & -0.5091640 \\ -0.2318690 & -0.6027983 & -0.5091640 & 0.9677416 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} 10.95491 & 1.039794 & 0.8115560\ 10^{-1} & 0.1726981\ 10^{-1} \\ 1.039794 & 3.494645 & 0.3880183 & 0.1244645 \\ 0.8115560\ 10^{-1} & 0.3880183 & 0.8236489 & 0.1754468 \\ 0.1726981\ 10^{-1} & 0.1244645 & 0.1754468 & 0.7267331 \end{bmatrix},$$

$$\vdots$$

Gli elementi non principali formano successioni decrescenti in modulo, e dopo 9 iterazioni la matrice A_{10} è data da

$$\begin{bmatrix} 11.09831 & 0.2762247\ 10^{-3} & 0.1729076\ 10^{-8} & -0.2617231\ 10^{-10} \\ 0.2762247\ 10^{-3} & 3.414135 & 0.3305371\ 10^{-4} & -0.7052673\ 10^{-6} \\ 0.1729076\ 10^{-8} & 0.3305372\ 10^{-4} & 0.9003896 & -0.1320110\ 10^{-1} \\ -0.2617231\ 10^{-10} & -0.7052673\ 10^{-6} & -0.1320110\ 10^{-1} & 0.5863345 \end{bmatrix}$$

L'elemento non principale di massimo modulo è dell'ordine di 10^{-2}. Ripetendo il procedimento fino a quando il massimo modulo degli elementi non principali è minore di 10^{-4}, gli elementi sulla diagonale principale alla 21-esima iterazione sono

$$11.09720 \quad 3.414135 \quad 0.9008932 \quad 0.5857800,$$

che si assumono come approssimazioni degli autovalori di A. ∎

b) Teorema di convergenza

Il seguente teorema di convergenza viene dato con ipotesi piuttosto restrittive, allo scopo di renderne più semplice la dimostrazione. La convergenza del metodo si può dimostrare anche con ipotesi assai più deboli, che verranno esaminate in seguito.

6.27 Teorema. *Sia $A \in \mathbf{C}^{n\times n}$ tale che i suoi autovalori λ_i, $i = 1, 2, \ldots, n$, abbiano moduli tutti distinti, cioè*

$$|\lambda_1| > |\lambda_2| > \ldots > |\lambda_n| > 0. \tag{29}$$

Indicata con X la matrice degli autovettori di A, tale che

$$A = XDX^{-1}, \tag{30}$$

in cui D è la matrice diagonale il cui i-esimo elemento principale è λ_i, si supponga che la matrice X^{-1} ammetta la fattorizzazione LU. Allora esistono delle matrici di fase S_k tali che

$$\lim_{k\to\infty} S_k^H R_k S_{k-1} = \lim_{k\to\infty} S_{k-1}^H A_k S_{k-1} = T, \tag{31}$$

e

$$\lim_{k\to\infty} S_{k-1}^H Q_k S_k = I,$$

dove T è triangolare superiore con gli elementi principali uguali a $\lambda_1, \lambda_2, \ldots, \lambda_n$. Quindi gli elementi principali di A_k tendono agli autovalori di A. Se A è una matrice hermitiana, allora T è diagonale.

Dim. Il teorema viene dimostrato confrontando due fattorizzazioni QR della matrice A^k ottenute in due modi diversi. Una prima fattorizzazione è data dalla seguente relazione

$$A^k = H_k U_k, \tag{32}$$

dove

$$H_k = Q_1 Q_2 \ldots Q_k$$

è una matrice unitaria e

$$U_k = R_k R_{k-1} \ldots R_1$$

è una matrice triangolare superiore. Per dimostrare la (32) si procede per induzione: per $k = 1$ risulta $A = A_1 = H_1 U_1$. Per $k > 1$, supposta valida la (32), da (26) e (27) si ottiene

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

da cui

$$Q_1 \ldots Q_{k-1} Q_k A_{k+1} = Q_1 \ldots Q_{k-1} A_k Q_k = \ldots = A Q_1 \ldots Q_{k-1} Q_k \tag{33}$$

e quindi

$$\begin{aligned} H_{k+1} U_{k+1} &= Q_1 \ldots Q_k Q_{k+1} R_{k+1} R_k \ldots R_1 \\ &= Q_1 \ldots Q_{k-1} Q_k A_{k+1} R_k R_{k-1} \ldots R_1 \\ &= A Q_1 \ldots Q_{k-1} Q_k R_k R_{k-1} \ldots R_1 = A H_k U_k = A^{k+1}, \end{aligned}$$

cioè $A^{k+1} = H_{k+1} U_{k+1}$.

Una seconda fattorizzazione QR della matrice A^k viene ottenuta dalla relazione (30). Sia $X^{-1} = LU$ la fattorizzazione LU di X^{-1}. Allora

$$A^k = XD^kX^{-1} = XD^kLU = XD^kLD^{-k}D^kU.$$

Poiché gli elementi della matrice D^kLD^{-k} sono dati da

$$\begin{cases} l_{ij}\left(\dfrac{\lambda_i}{\lambda_j}\right)^k & \text{per } i > j, \\ 1 & \text{per } i = j, \\ 0 & \text{per } i < j, \end{cases} \tag{34}$$

e $|\lambda_i| < |\lambda_j|$ per $i > j$, si può porre

$$D^kLD^{-k} = I + E_k,$$

dove

$$\lim_{k\to\infty} E_k = 0,$$

e quindi è

$$A^k = X(I + E_k)D^kU.$$

Indicata con

$$X = QR$$

una fattorizzazione QR della matrice X, si ha

$$A^k = QR(I + E_k)D^kU = Q(I + RE_kR^{-1})RD^kU,$$

e indicata con

$$I + RE_kR^{-1} = P_kT_k \tag{35}$$

una fattorizzazione QR della matrice $I + RE_kR^{-1}$, si ha

$$A^k = (QP_k)\ (T_kRD^kU). \tag{36}$$

La (36) dà una seconda fattorizzazione QR di A^k: infatti QP_k è unitaria e T_kRD^kU è triangolare superiore.

Poiché la fattorizzazione QR di una matrice è unica a meno di una matrice di fase, confrontando le due fattorizzazioni di A^k ottenute, cioè la (32) e la (36) segue che esiste una matrice di fase $\hat{S}_k$ tale che

$$H_k = QP_k\hat{S}_k^H \quad \text{e} \quad U_k = \hat{S}_kT_kRD^kU.$$

Risulta

$$Q_k = (H_{k-1})^{-1}H_k = \hat{S}_{k-1}P_{k-1}^HQ^HQP_k\hat{S}_k^H = \hat{S}_{k-1}P_{k-1}^HP_k\hat{S}_k^H,$$

da cui

$$\hat{S}_{k-1}^H Q_k \hat{S}_k = P_{k-1}^H P_k,$$

e

$$\begin{aligned} R_k = U_k (U_{k-1})^{-1} &= \hat{S}_k T_k R D^k U U^{-1} D^{-k+1} R^{-1} T_{k-1}^{-1} \hat{S}_{k-1}^H \\ &= \hat{S}_k T_k R D R^{-1} T_{k-1}^{-1} \hat{S}_{k-1}^H, \end{aligned}$$

e quindi

$$\hat{S}_k^H R_k \hat{S}_{k-1} = T_k R D R^{-1} T_{k-1}^{-1}.$$

Poiché $\lim_{k\to\infty} E_k = 0$, per la (35) risulta

$$\lim_{k\to\infty} (I + R E_k R^{-1}) = \lim_{k\to\infty} P_k T_k = I,$$

e quindi (si veda l'esercizio 6.30) esiste una matrice di fase $\check{S}_k$ tale che

$$\lim_{k\to\infty} P_k \check{S}_k = \lim_{k\to\infty} \check{S}_k^H T_k = I.$$

Allora posto $S_k = \hat{S}_k \check{S}_k$, è

$$\lim_{k\to\infty} S_{k-1}^H Q_k S_k = \lim_{k\to\infty} P_{k-1}^H P_k = I,$$

$$\lim_{k\to\infty} S_k^H R_k S_{k-1} = \lim_{k\to\infty} T_k R D R^{-1} T_{k-1}^{-1} = R D R^{-1},$$

e

$$\begin{aligned} \lim_{k\to\infty} S_{k-1}^H A_k S_{k-1} &= \lim_{k\to\infty} S_{k-1}^H Q_k R_k S_{k-1} = \lim_{k\to\infty} S_{k-1}^H Q_k S_k S_k^H R_k S_{k-1} \\ &= \lim_{k\to\infty} S_k^H R_k S_{k-1} = R D R^{-1}. \end{aligned}$$

La matrice $T = RDR^{-1}$ è triangolare superiore e quindi per gli elementi diagonali di A_k vale

$$\lim_{k\to\infty} a_{jj}^{(k)} = \lambda_j.$$

Se A è hermitiana, dalla (28) segue che le matrici A_k, e quindi le matrici $S_{k-1}^H A_k S_{k-1}$, sono hermitiane. Dalla (31) segue allora che T è hermitiana e quindi diagonale. ∎

c) Costo computazionale e stabilità

Il metodo QR applicato a una matrice di ordine n ha ad ogni passo un costo computazionale dell'ordine di n^3 operazioni moltiplicative (per calcolare la fattorizzazione $A_k = Q_k R_k$ e per moltiplicare la matrice triangolare R_k per le matrici elementari della fattorizzazione). Per abbassare il costo

computazionale globale conviene prima trasformare la matrice A in forma di Hessenberg superiore. Questa trasformazione viene eseguita una sola volta perché il metodo QR, applicato a matrici in forma di Hessenberg superiore produce matrici A_k in forma di Hessenberg superiore. Infatti se A_k è in forma di Hessenberg superiore, la matrice Q_k è data dal prodotto di $n-1$ matrici elementari di Householder (o di Givens) che sono in forma di Hessenberg superiore e quindi la matrice A_{k+1}, prodotto di una matrice triangolare superiore R_k per una matrice Q_k in forma di Hessenberg superiore, risulta ancora in forma di Hessenberg superiore. Se la matrice A è hermitiana, la matrice in forma di Hessenberg superiore, ottenuta applicando ad A i metodi di Householder o di Givens, è ancora hermitiana, e quindi risulta tridiagonale. Inoltre anche tutte le matrici A_k generate dal metodo QR sono hermitiane e quindi tridiagonali.

Il metodo QR applicato a una matrice A in forma di Hessenberg superiore ha ad ogni passo un costo computazionale di $2n^2$ operazioni moltiplicative (che è il costo computazionale per calcolare la fattorizzazione $A_k = Q_k R_k$, infatti il numero delle operazioni richieste per moltiplicare la matrice triangolare R_k per le matrici elementari della fattorizzazione è di ordine inferiore al secondo). Se A è una matrice tridiagonale, il costo computazionale di ogni passo del metodo è lineare in n.

In [28] viene dimostrato che il metodo QR gode delle stesse proprietà di stabilità di cui gode la fattorizzazione QR di una matrice.

6.28 Esempio. Il metodo QR viene applicato alla matrice tridiagonale

$$A_1 = \begin{bmatrix} 4 & 3.741654 & 0 & 0 \\ 3.741654 & 8.285707 & 2.602977 & 0 \\ 0 & 2.602977 & 3.039581 & 0.2254009 \\ 0 & 0 & 0.2254009 & 0.6746972 \end{bmatrix},$$

ottenuta con il metodo di Givens nell'esempio 6.17 dalla matrice A di cui sono stati approssimati gli autovalori nell'esempio 6.26. Si ottiene

$$A_2 = \begin{bmatrix} 9.733309 & 2.976943 & 0 & 0 \\ 2.976943 & 4.547497 & 0.7894507 & 0 \\ 0 & 0.7894507 & 1.094460 & \text{-}0.1081211 \\ 0 & 0 & \text{-}0.1081211 & 0.6246958 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} 10.95485 & 1.043100 & 0 & 0 \\ 1.043100 & 3.542464 & 0.2006968 & 0 \\ 0 & 0.2006968 & 0.8992355 & 0.7246423\ 10^{-1} \\ 0 & 0 & 0.7246423\ 10^{-1} & 0.6033282 \end{bmatrix},$$

$$\vdots$$

Ripetendo il procedimento fino a quando il massimo modulo degli elementi non principali è minore di 10^{-4}, alla 18-esima iterazione gli elementi principali sono

$$11.09809 \quad 3.414161 \quad 0.9008972 \quad 0.5857840,$$

che si assumono come approssimazioni degli autovalori di A. ∎

d) Convergenza in ipotesi più deboli

La dimostrazione della convergenza del metodo QR è stata fatta nell'ipotesi che la matrice X^{-1} fosse fattorizzabile nella forma LU. In questo caso gli elementi principali di T coincidono, nell'ordine, con $\lambda_1, \lambda_2, \ldots, \lambda_n$. Se X^{-1} non ammette fattorizzazione LU, si può dimostrare [28] che il metodo QR è ancora convergente. In questo caso gli elementi principali di T coincidono ancora con i λ_i, ma non sono più in ordine di modulo decrescente.

Se l'ipotesi (29) del teorema 6.27, che tutti gli autovalori abbiano modulo distinto, non è verificata, la successione formata dagli elementi diagonali di A_k non converge. Questa ipotesi è troppo restrittiva, e non consente di utilizzare il metodo QR in casi particolarmente importanti nelle applicazioni, come quelli in cui la matrice A ha elementi reali e autovalori non reali. Però anche in questo caso il metodo QR può essere applicato con opportune varianti. Sia ad esempio

$$|\lambda_1| > \ldots > |\lambda_r| = |\lambda_{r+1}| > \ldots > |\lambda_n| > 0,$$

dove λ_r e λ_{r+1} sono due numeri complessi coniugati, oppure due numeri reali. Allora nella (34) la successione degli elementi

$$l_{r+1,r}\Big(\frac{\lambda_{r+1}}{\lambda_r}\Big)^k$$

non converge a zero per $k \to \infty$. Ne segue che le matrici P_k, e quindi le matrici $S_{k-1}^H Q_k S_k$, non convergono alla matrice I per la presenza nella posizione $(r+1, r)$ di elementi che non tendono a zero. Sia

$$A_r^{(k)} = \begin{bmatrix} a_{rr}^{(k)} & a_{r,r+1}^{(k)} \\ a_{r+1,r}^{(k)} & a_{r+1,r+1}^{(k)} \end{bmatrix}$$

la sottomatrice principale di ordine 2 di A_k formata dalle righe e colonne di indici r e $r+1$. La successione $\{A_r^{(k)}\}$ non converge, ma gli autovalori delle sottomatrici $A_r^{(k)}$ convergono a λ_r e λ_{r+1} [28]. Gli elementi principali di A_k di indice diverso da r e $r+1$ convergono agli altri autovalori. Situazioni analoghe si presentano quando la matrice A ha più autovalori di modulo

uguale e in questo caso il metodo QR genera matrici R_k con struttura triangolare a blocchi, in cui gli autovalori dei blocchi diagonali convergono ad autovalori di A.

6.29 Esempio. Si applica il metodo QR alla matrice

$$A_1 = \begin{bmatrix} 4 & 3.464101 & -1.388729 & 0.2672636 \\ 1.732049 & 7.666669 & 1.158131 & -0.4629126 \\ 0 & -1.247218 & 2.476188 & 0.7423077 \\ 0 & 0 & 0.9897417 & 1.857139 \end{bmatrix},$$

in forma di Hessenberg superiore ottenuta nell'esempio 6.24. Si ottiene

$$A_2 = \begin{bmatrix} 6.473673 & 4.062625 & 0.3739953\ 10^{-1} & -0.7850719\ 10^{-1} \\ 2.302595 & 4.968325 & 2.265884 & -0.9043741\ 10^{-1} \\ 0 & -0.6322317 & 2.717972 & -0.9186863 \\ 0 & 0 & 0.6584795 & 1.840011 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} 8.314364 & 2.673093 & 1.266714 & 0.3786074 \\ 1.132446 & 2.832693 & 2.120341 & 0.3386071 \\ 0 & -0.5868980 & 2.906489 & 1.142451 \\ 0 & 0 & -0.4178842 & 1.946413 \end{bmatrix},$$

$$\vdots$$

$$A_{10} = \begin{bmatrix} 8.783114 & -1.744292 & -1.348724 & -0.9030869 \\ 0.3378619\ 10^{-3} & 2.480618 & 1.925506 & 0.7184988 \\ 0 & -0.7168258 & 2.609033 & 0.9730009 \\ 0 & 0 & 0.6583786\ 10^{-1} & 2.126765 \end{bmatrix},$$

$$A_{11} = \begin{bmatrix} 8.783008 & -1.301571 & -1.799058 & 0.8643191 \\ 0.9932932\ 10^{-4} & 2.168211 & 1.806849 & -0.3811607 \\ 0 & -0.8444027 & 2.947040 & -1.116467 \\ 0 & 0 & -0.4549125\ 10^{-1} & 2.101224 \end{bmatrix}.$$

Come si può notare, le successioni degli elementi di indici (2,1) e (4,3) sono decrescenti in modulo, più rapidamente la prima, più lentamente la seconda, mentre questo non accade nella successione delle sottomatrici principali formate dagli elementi delle righe e colonne di indici 2 e 3. Ripetendo il procedimento fino a quando l'elemento di indici (4,3) risulta inferiore in modulo a 10^{-4}, alla 30-esima iterazione gli elementi principali di indici 1 e 4 sono

$$a_{11}^{(31)} = 8.782016, \quad a_{44}^{(31)} = 2.089449,$$

che sono delle buone approssimazioni degli autovalori λ_1 e λ_4 di massimo e minimo modulo. La sottomatrice principale $A_2^{(31)}$ di ordine 2 risulta

$$A_2^{(31)} = \begin{bmatrix} 2.576344 & 1.940763 \\ -0.6846419 & 2.550408 \end{bmatrix},$$

da cui si ricavano per λ_2 e λ_3 le approssimazioni

$$\lambda_2 = 2.563376 + \mathbf{i}\, 1.152632\ , \quad \lambda_3 = 2.563376 - \mathbf{i}\, 1.152632\ . \qquad \blacksquare$$

e) Condizioni di arresto e riduzione dell'ordine della matrice

Fissato un valore ϵ di tolleranza, si procede applicando il metodo QR alla matrice A in forma di Hessenberg superiore fino a quando per un indice p, $1 \le p < n$, l'elemento $a_{p+1,p}^{(k)}$ diventa sufficientemente piccolo. Un criterio utilizzato è il seguente

$$|a_{p+1,p}^{(k)}| < \epsilon(|a_{pp}^{(k)}| + |a_{p+1,p+1}^{(k)}|). \tag{37}$$

Quando la condizione (37) è verificata, nella matrice A_k

$$A_k = \begin{bmatrix} B_k & D_k \\ & \\ E_k & C_k \end{bmatrix} \begin{matrix} \} & p \text{ righe} \\ & \\ \} & n-p \text{ righe} \end{matrix}$$

dove $B_k \in \mathbf{C}^{p\times p}, C_k \in \mathbf{C}^{(n-p)\times(n-p)}$, la sottomatrice E_k ha un elemento di modulo piccolo e gli altri tutti nulli. Si procede quindi operando separatamente con le matrici B_k e C_k. Se la matrice A è hermitiana, gli autovalori di B_k e C_k sono delle buone approssimazioni degli autovalori di A_k (si veda l'esercizio 7.3).

f) Tecnica di traslazione

La velocità di convergenza del metodo QR dipende per la (34) dai rapporti $|\lambda_i/\lambda_j|$ per $i > j$, e quindi per l'ipotesi (29) dal numero

$$\max_{1\le i\le n-1} \left|\frac{\lambda_{i+1}}{\lambda_i}\right|. \tag{38}$$

Se tale rapporto è vicino ad 1, la convergenza può essere lenta. In questo caso per accelerare la convergenza si utilizza una tecnica di traslazione dello spettro degli autovalori di A, detta *di shift*.

Sia μ un numero che approssima un autovalore λ meglio degli altri autovalori. Le matrici Q_k e R_k, generate dal metodo QR a partire dalla

matrice $A - \mu I$ possono essere costruite anche per mezzo delle seguenti relazioni (*metodo QR con shift*)

$$\left.\begin{aligned} A_k - \mu I &= Q_k R_k, \\ A_{k+1} &= R_k Q_k + \mu I, \end{aligned}\right\} \quad \text{per } k = 1, 2, \ldots$$

e risulta

$$Q_k A_{k+1} = A_k Q_k - \mu Q_k + \mu Q_k = A_k Q_k.$$

Tenendo presente che gli autovalori di $A - \mu I$ sono $\lambda_i - \mu$ e che la velocità di convergenza è regolata dalla (38), è possibile scegliere un parametro μ in modo da accelerare la convergenza del metodo QR con shift. È conveniente scegliere per μ un valore che approssima λ_n. Ciò può essere ottenuto applicando il metodo QR inizialmente senza shift per un certo numero p di iterazioni, e scegliendo $\mu = a_{nn}^{(p)}$ per le successive iterazioni con shift.

Poiché μ può essere modificato ad ogni iterazione è più conveniente scegliere

$$\mu_k = a_{nn}^{(k)}, \quad k = 1, 2, \ldots. \tag{39}$$

Nel caso delle matrici hermitiane è possibile dimostrare [29] che con questa strategia la convergenza a zero dell'elemento $a_{n,n-1}^{(k)}$ è del terzo ordine (si veda anche l'esercizio 6.31).

Quando la (37) è verificata per $p = n-1$, si passa a operare sulla matrice B_k di ordine $n - 1$ ottenuta dalla matrice A_k eliminando l'ultima riga e l'ultima colonna. Per l'approssimazione degli altri autovalori si procede in modo analogo.

6.30 Esempio. Si applica il metodo QR con lo shift (39) alla matrice tridiagonale

$$A_1 = \begin{bmatrix} 4 & 3.741654 & 0 & 0 \\ 3.741654 & 8.285707 & 2.602977 & 0 \\ 0 & 2.602977 & 3.039581 & 0.2254009 \\ 0 & 0 & 0.2254009 & 0.6746972 \end{bmatrix},$$

ottenuta con il metodo di Givens nell'esempio 6.17, di cui sono stati approssimati gli autovalori negli esempi 6.26 e 6.28. Si ha

$$A_2 = \begin{bmatrix} 10.11023 & 2.576269 & 0 & 0 \\ 2.576269 & 4.380260 & 0.2509962 & 0 \\ 0 & 0.2509962 & 0.9084192 & 0.6795645\ 10^{-1} \\ 0 & 0 & 0.6795645\ 10^{-1} & 0.6010619 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} 11.01885 & 0.7804608 & 0 & 0 \\ 0.7804608 & 3.494054 & 0.2471317\ 10^{-1} & 0 \\ 0 & 0.2471317\ 10^{-1} & 0.9011788 & 0.3621320\ 10^{-2} \\ 0 & 0 & 0.3621320\ 10^{-2} & 0.5858268 \end{bmatrix},$$

$$A_4 = \begin{bmatrix} 11.09302 & 0.2120095 & 0 & 0 \\ 0.2120095 & 3.420040 & 0.2740411\ 10^{-2} & 0 \\ 0 & 0.2740411\ 10^{-2} & 0.9009814 & 0.4774449\ 10^{-6} \\ 0 & 0 & 0.4774449\ 10^{-6} & 0.5857852 \end{bmatrix}.$$

Poiché l'elemento $a_{43}^{(4)}$ soddisfa alla condizione (37) con $\epsilon = 10^{-6}$, si passa a operare sulla sottomatrice di ordine 3 ottenuta eliminando l'ultima riga e colonna. Dopo altre 3 iterazioni, si ottiene la matrice

$$A_7 = \begin{bmatrix} 11.09874 & 0.4131589\ 10^{-2} & 0 \\ 0.4131589\ 10^{-2} & 3.414158 & \text{-}0.3791176\ 10^{-5} \\ 0 & \text{-}0.3791176\ 10^{-5} & 0.9009783 \end{bmatrix},$$

che può essere a sua volta ridotta. Gli altri due autovalori possono essere calcolati direttamente dalla sottomatrice principale di testa di ordine 2. Si ottengono così le approssimazioni degli autovalori di A

$$\lambda_1 = 11.09835, \quad \lambda_2 = 3.414142, \quad \lambda_3 = 0.9009783, \quad \lambda_4 = 0.5857852. \quad \blacksquare$$

Il metodo QR con shift può essere applicato anche ai casi in cui esistono più autovalori con lo stesso modulo. In particolare, se $|\lambda_{n-1}| = |\lambda_n|$, allora conviene scegliere come $\mu^{(k)}$, $k = 1, 2, \ldots$, l'autovalore della sottomatrice

$$A_{n-1}^{(k)} = \begin{bmatrix} a_{n-1,n-1}^{(k)} & a_{n-1,n}^{(k)} \\ a_{n,n-1}^{(k)} & a_{nn}^{(k)} \end{bmatrix}$$

che è più vicino ad $a_{nn}^{(k)}$.

In questo caso, anche se la matrice A ha elementi reali, l'utilizzazione dello shift può portare ad una matrice A_k ad elementi complessi, con conseguente aumento del costo computazionale. Questo può essere evitato eseguendo due iterazioni successive

$$A_k \to A_{k+1} \to A_{k+2},$$

e usando come costanti di traslazione $\mu^{(k)} = \alpha$ e $\mu^{(k+1)} = \beta$, dove α e β sono i due autovalori della sottomatrice $A_{n-1}^{(k)}$. Si ha infatti

$$A_k - \alpha I = Q_k R_k,$$

$$A_{k+1} = R_k Q_k + \alpha I, \tag{40}$$
$$A_{k+1} - \beta I = Q_{k+1} R_{k+1},$$
$$A_{k+2} = R_{k+1} Q_{k+1} + \beta I,$$

e quindi

$$Q_k Q_{k+1} R_{k+1} R_k = Q_k (A_{k+1} - \beta I) R_k = Q_k (R_k Q_k + \alpha I - \beta I) R_k$$
$$= Q_k R_k (Q_k R_k + \alpha I - \beta I) = (A_k - \alpha I)\ (A_k - \beta I). \tag{41}$$

La matrice $M = (A_k - \alpha I)(A_k - \beta I)$ ha elementi reali se A_k è reale, perché α e β sono radici di un'equazione di secondo grado a coefficienti reali. Ne segue che ponendo

$$Z = Q_k Q_{k+1} \quad \text{e} \quad S = R_{k+1} R_k,$$

dalla (41) si ricava che ZS è una fattorizzazione QR della matrice reale M e quindi Z e S sono, a meno di moltiplicazione per una matrice di fase, matrici reali rispettivamente ortogonale e triangolare superiore. D'altra parte dalle (40) risulta che

$$Z A_{k+2} = Q_k Q_{k+1} A_{k+2} = Q_k Q_{k+1} R_{k+1} Q_{k+1} + \beta Q_k Q_{k+1} = Q_k A_{k+1} Q_{k+1}$$
$$= Q_k R_k Q_k Q_{k+1} + \alpha Q_k Q_{k+1} = A_k Q_k Q_{k+1} = A_k Z,$$

da cui

$$A_{k+2} = Z^H A_k Z.$$

È quindi possibile ricavare A_{k+2} direttamente da A_k utilizzando la fattorizzazione QR della matrice reale M. Questo modo di procedere però ha un consistente costo computazionale, in quanto la sola costruzione della matrice M, che non è in forma di Hessenberg superiore anche se lo è la A_k, richiede $n^3/6$ operazioni moltiplicative.

Per superare questo inconveniente si utilizza il seguente procedimento suggerito da Francis, che richiede $6n^2$ operazioni moltiplicative:

1. si costruisce la prima colonna $\mathbf{m}_1$ della matrice M e la matrice elementare di Householder P_0 tale che

$$P_0 \mathbf{m}_1 = \gamma \mathbf{e}_1, \quad \text{dove} \quad |\gamma| = \|\mathbf{m}_1\|_2.$$

2. si costruiscono le matrici di Householder $P_1, P_2, \ldots, P_{n-2}$ tali che, posto $Z' = P_0 P_1 \ldots P_{n-2}$, la matrice $(Z')^H A_k Z'$ sia in forma di Hessenberg superiore. È possibile dimostrare che le matrici

$$A_{k+2} = Z^H A_k Z \quad \text{e} \quad A'_{k+2} = (Z')^H A_k Z'$$

sono "essenzialmente" uguali, cioè uguali nel senso che esiste una matrice di fase reale D, tale che

$$A_{k+2} = D^{-1} A'_{k+2} D$$

(si vedano come esempio di matrici essenzialmente uguali le due matrici $A^{(3)}$ e $H^{(3)}$ ottenute nell'esempio 6.17).

6.31 Esempio. Applicando il metodo QR con lo shift (39) alla matrice

$$A_1 = \begin{bmatrix} 4 & 3.464101 & -1.388729 & 0.2672636 \\ 1.732049 & 7.666669 & 1.158131 & -0.4629126 \\ 0 & -1.247218 & 2.476188 & 0.7423077 \\ 0 & 0 & 0.9897417 & 1.857139 \end{bmatrix},$$

in forma di Hessenberg superiore ottenuta nell'esempio 6.24, si ottiene (il metodo senza shift è stato applicato alla matrice A_1 nell'esempio 6.29)

$$A_2 = \begin{bmatrix} 7.989221 & 2.992575 & 1.020255 & -0.6292015 \\ 1.667174 & 3.078217 & 2.177103 & -0.5580172 \\ 0 & -0.7987912 & 2.882763 & -1.142143 \\ 0 & 0 & 0.1381161 & 2.049773 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} 8.842974 & -1.227221 & 1.630806 & 0.8734073 \\ 0.2213716 & 2.623399 & 0.7493563 & -0.9821870 \\ 0 & -1.915713 & 2.440961 & 0.6386327 \\ 0 & 0 & -0.3020395\ 10^{-2} & 2.092627 \end{bmatrix}.$$

Dopo altre due iterazioni l'elemento $a_{43}^{(5)}$ soddisfa alla condizione (37) con $\epsilon = 10^{-10}$. Il valore

$$a_{44}^{(5)} = 2.089537$$

viene assunto come approssimazione di λ_4 e si passa a operare sulla sottomatrice di ordine 3 ottenuta eliminando l'ultima riga e l'ultima colonna. In questa matrice gli autovalori di minimo modulo sono due e sono complessi, per cui si applica il procedimento suggerito da Francis, ottenendo dopo altre 2 iterazioni la matrice

$$A_9 = \begin{bmatrix} 8.783265 & 1.081047 & -1.955143 \\ 0.2728484\ 10^{-11} & 2.064380 & -1.694178 \\ 0 & 0.9313574 & 3.062660 \end{bmatrix},$$

in cui l'elemento $a_{21}^{(9)}$ è in modulo minore di 10^{-11}. L'elemento $a_{11}^{(9)}$ viene assunto come approssimazione dell'autovalore λ_1, mentre gli autovalori λ_2 e

λ_3 vengono approssimati calcolando gli autovalori della sottomatrice principale $A_2^{(9)}$. ∎

g) Calcolo degli autovettori

Con il metodo QR si ottiene la forma normale di Schur della matrice A. Infatti dalla (33) si ha

$$H_k A_{k+1} = A H_k$$

e per la (31) è

$$\lim_{k\to\infty} S_k^H H_k^H A H_k S_k = T,$$

in cui T è una matrice triangolare superiore e $H_k S_k$ è una matrice unitaria. Se la matrice A è normale, è facile dimostrare (si veda l'esercizio 6.29) che esiste una sottosuccessione $\{ H_{k_i} S_{k_i} \}$ della successione $\{ H_k S_k \}$ che converge alla matrice unitaria le cui colonne sono gli autovettori di A. Il costo computazionale del calcolo degli autovettori è elevato perché ad ogni passo è richiesta la costruzione e la memorizzazione della matrice $H_k = H_{k-1} Q_k$. Per il calcolo degli autovettori conviene ricorrere al metodo delle potenze con la variante di Wielandt, descritto più avanti.

9. Metodo di Jacobi

Il *metodo di Jacobi* è un metodo classico per calcolare gli autovalori e gli autovettori di matrici hermitiane. Per la semplicità con cui può essere implementato viene ancora oggi usato nel caso di matrici di piccole dimensioni, quando sono richiesti tutti gli autovalori. Inoltre questo metodo si presta bene per una utilizzazione in ambiente di calcolo parallelo, dove si assume che ad ogni passo possano essere effettuate simultaneamente p operazioni aritmetiche, con $p > 1$.

Il metodo di Jacobi è un metodo iterativo che utilizza trasformazioni della forma (16)

$$A^{(1)} = A, \quad A^{(k+1)} = T_k^{-1} A^{(k)} T_k, \quad k = 1, 2, \ldots,$$

in cui le matrici T_k sono matrici di Givens. T_k viene scelta in modo da rendere nullo un opportuno elemento non principale di $A^{(k+1)}$. La successione $\{A^{(k)}\}$, se è convergente, converge ad una matrice diagonale D.

Considerando per semplicità il caso in cui $A^{(k)}$ è reale e seguendo la notazione del paragrafo 5, in cui si indicano con a_{rj} gli elementi di $A^{(k)}$ e con $\hat{a}_{rj}$ gli elementi di $A^{(k+1)}$, la matrice $T_k = G_{pq}$, viene determinata in modo che risulti $\hat{a}_{pq} = 0$ (se $a_{pq} = 0$, basta porre $T_k = I$). Dalla (17), posto $t = \text{tg}\ \phi$, risulta

$$(1 - t^2) a_{pq} - t(a_{pp} - a_{qq}) = 0,$$

da cui si ottiene l'equazione

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

dove

$$m = \frac{a_{pp} - a_{qq}}{2a_{pq}}.$$

Fra le due soluzioni dell'equazione si sceglie quella di minimo modulo, per cui $|\phi| \leq \frac{\pi}{4}$, calcolata nella forma

$$t = \frac{\text{sgn}(m)}{|m| + \sqrt{1 + m^2}}$$

per evitare possibili fenomeni di cancellazione. Si calcola poi

$$c = \frac{1}{\sqrt{1 + t^2}} \quad \text{e} \quad s = tc.$$

Con questa trasformazione si annulla quindi un elemento non principale della matrice, che in generale può essere modificato al passo successivo. Lo scopo del metodo è quello di ridurre ad ogni passo la quantità

$$S(A^{(k)}) = \sum_{\substack{r,j=1 \\ r \neq j}}^{n} |a_{rj}^{(k)}|^2.$$

Tenendo conto delle relazioni, riportate nel paragrafo 5, che legano gli elementi di $A^{(k)}$ e di $A^{(k+1)}$ e del fatto che $c^2 + s^2 = 1$, si ricava che

$$|\hat{a}_{pp}|^2 + |\hat{a}_{qq}|^2 + 2|\hat{a}_{pq}|^2 = |a_{pp}|^2 + |a_{qq}|^2 + 2|a_{pq}|^2$$

e

$$|\hat{a}_{rp}|^2 + |\hat{a}_{rq}|^2 = |a_{rp}|^2 + |a_{rq}|^2, \text{ per } r \neq p,\ q.$$

Poiché gli altri elementi delle due matrici $A^{(k)}$ e $A^{(k+1)}$ non cambiano, si ha

$$\sum_{r,j=1}^{n} |\hat{a}_{rj}|^2 = \sum_{r,j=1}^{n} |a_{rj}|^2$$

e quindi, avendo imposto la condizione che $\hat{a}_{pq} = 0$, se $a_{pq} \neq 0$ si ha

$$\begin{aligned} S(A^{(k+1)}) &= \sum_{r,j=1}^{n} |\hat{a}_{rj}|^2 - \sum_{r=1}^{n} |\hat{a}_{rr}|^2 \\ &= \sum_{r,j=1}^{n} |a_{rj}|^2 - \sum_{r=1}^{n} |a_{rr}|^2 - 2|a_{pq}|^2 = S(A^{(k)}) - 2|a_{pq}|^2 \\ &< S(A^{(k)}). \end{aligned} \tag{42}$$

La successione dei numeri positivi $\{S(A^{(k)})\}$ risulta allora decrescente. Si può dimostrare che questa successione tende a zero solo individuando ad ogni passo un'opportuna strategia per la scelta degli elementi da azzerare. Nella *strategia classica* al k-esimo passo si sceglie un elemento non principale di massimo modulo di $A^{(k)}$. Il procedimento si arresta quando tale modulo risulta inferiore ad una quantità prefissata che dipende dalla precisione che si vuole ottenere.

6.32 Teorema. *Sia* $\{A^{(k)}\}$ *la successione ottenuta applicando il metodo di Jacobi alla matrice hermitiana* $A \in \mathbf{C}^{n\times n}$ *secondo la strategia classica. Allora* $\lim_{k\to\infty} S(A^{(k)}) = 0$ *e quindi il* $\lim_{k\to\infty} A^{(k)}$ *è una matrice diagonale.*

Dim. Poiché a_{pq} è un elemento non principale di massimo modulo di $A^{(k)}$, risulta

$$a_{pq}^2 \geq \frac{S(A^{(k)})}{n(n-1)} \, .$$

Dalla (42) si ha allora

$$S(A^{(k+1)}) \leq S(A^{(k)}) - \frac{2S(A^{(k)})}{n(n-1)} = \gamma S(A^{(k)}) \tag{43}$$

dove $\gamma = 1 - \dfrac{2}{n(n-1)} < 1$ per $n \geq 2$. Applicando in modo ricorrente la (43) si ha

$$S(A^{(k+1)}) \leq \gamma^k S(A^{(1)}),$$

da cui la tesi. ∎

Per individuare nella matrice $A^{(k)}$ un elemento non principale di massimo modulo si devono confrontare fra di loro $\dfrac{n(n-1)}{2}$ elementi. Perciò la strategia classica ha un costo computazionale elevato, e per questo motivo è preferibile adottare una *strategia ciclica* in cui la scelta della successione degli indici (p, q) avviene nel modo seguente

$$\begin{array}{cccc} (1,2) & (1,3) & \dots & (1,n) \\ & (2,3) & \dots & (2,n) \\ & & \ddots & \vdots \\ & & & (n-1,n) \end{array}$$

e tale successione viene ripetuta ciclicamente saltando gli indici (p, q) corrispondenti a elementi che in modulo sono minori di una quantità prefissata.

Anche per il metodo di Jacobi applicato con la strategia ciclica si può dimostrare un teorema di convergenza analogo al 6.32. Inoltre sia per la strategia classica che per quella ciclica è possibile dimostrare [27] che se A è hermitiana con autovalori distinti $\lambda_i,\ i = 1, 2, \ldots, n$, allora da un certo passo k in poi risulta

$$S(A^{(k+N)}) \leq \frac{2S(A^{(k)})^2}{\delta^2} \ ,$$

dove $N = \dfrac{n(n-1)}{2}$ e $\delta = \min\limits_{i \neq j} |\lambda_i - \lambda_j|$, cioè il metodo di Jacobi ha convergenza localmente quadratica.

6.33 Esempio. Applicando il metodo di Jacobi con la strategia classica alla matrice

$$A = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 3 & 4 & 3 & 2 \\ 2 & 3 & 4 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix} .$$

dell'esempio 6.17, si ha

k	p	q	$S(A^{(k)})$
1	1	2	72.00000
2	2	3	53.99997
3	2	4	28.99997
4	3	4	5.540541
5	1	4	2.681776
6	1	2	1.627234
7	2	4	1.058018
8	2	3	0.5056292
9	1	3	$0.5161846\ 10^{-1}$
10	3	4	$0.2429459\ 10^{-2}$
11	1	4	$0.1281016\ 10^{-3}$
12	1	2	$0.5892318\ 10^{-4}$
13	2	3	$0.2793697\ 10^{-4}$
14	2	4	$0.1687663\ 10^{-5}$
15	1	3	$0.1030698\ 10^{-7}$

Gli elementi principali della matrice $A^{(15)}$ sono

$$0.9009814 \quad 11.09902 \quad 0.5857867 \quad 3.414210,$$

che si assumono come approssimazioni degli autovalori della matrice A. Con la strategia ciclica, per ottenere le stesse approssimazioni degli autovalori, sono richiesti due passi in più. ∎

Posto $Q_k = T_k T_{k-1} \dots T_1$, la matrice

$$Q = \lim_{k \to \infty} Q_k$$

ha per colonne gli autovettori di A.

10. Metodo delle potenze

Il *metodo delle potenze* è un classico metodo iterativo per approssimare l'autovalore di modulo massimo di una matrice e il corrispondente autovettore. Sulla base di questo metodo sono stati sviluppati altri metodi che sono particolarmente adatti per approssimare gli autovalori di matrici sparse di grosse dimensioni. È facile dimostrare la convergenza del metodo nel caso che la matrice sia diagonalizzabile e abbia un solo autovalore di modulo massimo.

Sia $A \in \mathbf{C}^{n \times n}$, con n autovettori $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ linearmente indipendenti e autovalori $\lambda_1, \lambda_2, \dots, \lambda_n$ tali che

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

cioè l'autovalore di modulo massimo ha molteplicità algebrica 1 e non esistono altri autovalori con lo stesso modulo.

Fissato un vettore $\mathbf{t}_0 \in \mathbf{C}^n$, si genera la successione $\{\mathbf{y}_k\},\ k = 1, 2, \dots$, così definita

$$\begin{aligned} &\mathbf{y}_0 = \mathbf{t}_0, \\ &\mathbf{y}_k = A\mathbf{y}_{k-1}, \quad k = 1, 2, \dots \end{aligned} \tag{44}$$

Poiché i vettori $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ sono linearmente indipendenti, il vettore $\mathbf{t}_0$ può essere espresso per mezzo della combinazione lineare

$$\mathbf{t}_0 = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i,$$

e si supponga scelto in modo tale che $\alpha_1 \neq 0$; risulta quindi

$$\mathbf{y}_k = A^k \mathbf{t}_0 = \sum_{i=1}^{n} \alpha_i A^k \mathbf{x}_i = \sum_{i=1}^{n} \alpha_i \lambda_i^k \mathbf{x}_i = \lambda_1^k \left[\alpha_1 \mathbf{x}_1 + \sum_{i=2}^{n} \alpha_i \Big(\frac{\lambda_i}{\lambda_1} \Big)^k \mathbf{x}_i \right]. \tag{45}$$

Indicate con $y_r^{(k)}$ e con $x_r^{(i)}$ le r-esime componenti dei vettori $\mathbf{y}_k$ e $\mathbf{x}_i$, per gli indici j per cui $y_j^{(k)} \neq 0$ e $x_j^{(1)} \neq 0$, si ha

$$\frac{y_j^{(k+1)}}{y_j^{(k)}} = \lambda_1 \frac{\alpha_1 x_j^{(1)} + \sum_{i=2}^{n} \alpha_i \Big(\frac{\lambda_i}{\lambda_1}\Big)^{k+1} x_j^{(i)}}{\alpha_1 x_j^{(1)} + \sum_{i=2}^{n} \alpha_i \Big(\frac{\lambda_i}{\lambda_1}\Big)^{k} x_j^{(i)}} \ , \tag{46}$$

e poiché $|\lambda_i/\lambda_1| < 1$ per $i \geq 2$ si ha

$$\lim_{k\to\infty} \frac{y_j^{(k+1)}}{y_j^{(k)}} = \lambda_1.$$

Quindi da un certo indice k in poi l'autovalore λ_1 può essere approssimato mediante uno dei rapporti $y_j^{(k+1)}/y_j^{(k)}$.

Con questo metodo si può approssimare anche l'autovettore $\mathbf{x}_1$. Dalla (45) risulta infatti

$$\lim_{k\to\infty} \frac{\mathbf{y}_k}{\lambda_1^k} = \alpha_1 \mathbf{x}_1,$$

e quindi per $j = 1, \ldots, n$, è

$$\lim_{k\to\infty} \frac{y_j^{(k)}}{\lambda_1^k} = \alpha_1 \mathbf{x}_j^{(1)},$$

e

$$\lim_{k\to\infty} \frac{\mathbf{y}_k}{y_j^{(k)}} = \frac{\mathbf{x}_1}{x_j^{(1)}} \ , \tag{47}$$

per tutti gli indici j per cui $x_j^{(1)} \neq 0$. Poiché per k sufficientemente elevato l'indice m di una componente di massimo modulo di $\mathbf{y}_k$ rimane costante, la successione $\mathbf{y}_k/y_m^{(k)}$ converge all'autovettore $\mathbf{x}_1$ normalizzato in norma ∞.

Questo metodo richiede ad ogni passo il calcolo del prodotto di una matrice A per un vettore: se A non è sparsa ogni passo richiede n^2 operazioni moltiplicative, mentre se A è sparsa ogni passo richiede θ operazioni moltiplicative, dove $\theta << n^2$ è il numero di elementi non nulli di A (ad esempio se A è tridiagonale, il numero degli elementi non nulli di A è $3n-2$).

Però operando in aritmetica finita, con la (44) dopo pochi passi si possono presentare condizioni di overflow o di underflow. Per evitare che ciò accada è necessario eseguire ad ogni passo una normalizzazione del vettore ottenuto, costruendo una successione $\mathbf{t}_k$, $k = 1, 2, \ldots$ così definita

$$\left.\begin{aligned} \mathbf{u}_k &= A\mathbf{t}_{k-1}, \\ \mathbf{t}_k &= \frac{1}{\beta_k}\ \mathbf{u}_k, \end{aligned}\right\}, \quad k = 1, 2, \ldots, \tag{48}$$

dove β_k è uno scalare tale che $\|\mathbf{t}_k\| = 1$, per qualche norma vettoriale $\| \, . \, \|$.

Si ha allora

$$\mathbf{t}_k = \frac{1}{\gamma_k}\, \mathbf{y}_k = \frac{1}{\gamma_k} A^k \mathbf{t}_0, \quad \text{dove } \gamma_k = \prod_{i=1}^{k} \beta_i,$$

e poiché

$$\mathbf{u}_{k+1} = \frac{1}{\gamma_k}\, A^{k+1} \mathbf{t}_0,$$

operando come nella (46) si ha che il rapporto fra le j-esime componenti di $\mathbf{u}_{k+1}$ e $\mathbf{t}_k$, per gli indici j per cui $t_j^{(k)} \neq 0$ e $x_j^{(1)} \neq 0$, è dato da

$$\frac{u_j^{(k+1)}}{t_j^{(k)}} = \lambda_1 \frac{\alpha_1 x_j^{(1)} + \sum_{i=2}^{n} \alpha_i \Big(\frac{\lambda_i}{\lambda_1}\Big)^{k+1} x_j^{(i)}}{\alpha_1 x_j^{(1)} + \sum_{i=2}^{n} \alpha_i \Big(\frac{\lambda_i}{\lambda_1}\Big)^{k} x_j^{(i)}}, \tag{49}$$

e quindi

$$\lim_{k\to\infty} \frac{u_j^{(k+1)}}{t_j^{(k)}} = \lambda_1.$$

Si esaminano ora in dettaglio i casi particolari in cui la normalizzazione sia fatta con la norma ∞ o con la norma 2.

Utilizzando la norma ∞, sia $\|\mathbf{t}_0\|_\infty = 1$ e sia β_k una componente di massimo modulo di $\mathbf{u}_k$, cioè tale che

$$\beta_k = u_m^{(k)}, \quad \text{con} \quad |u_m^{(k)}| = \max_{j=1,\ldots,n} |u_j^{(k)}| = \|\mathbf{u}_k\|_\infty.$$

I vettori $\mathbf{t}_k$ ottenuti con la (48) sono quindi tali che $t_m^{(k)} = 1$. Dalla (49) risulta

$$u_m^{(k+1)} = \lambda_1 \left(1 + O\Big(\frac{\lambda_2}{\lambda_1}\Big)^k\right).$$

Poiché si può assumere che da una certa iterazione in poi l'indice m, corrispondente a una componente di massimo modulo di $\mathbf{u}_k$, resti sempre lo stesso, ne segue che la successione dei β_k converge a λ_1 e che l'errore che si commette approssimando λ_1 con β_k tende a zero come $|\lambda_2/\lambda_1|^k$. Inoltre, poiché $\|\mathbf{t}_k\|_\infty = 1$, dalla (47) risulta

$$\lim_{k\to\infty} \mathbf{t}_k = \frac{\mathbf{x}_1}{x_m^{(1)}},$$

e quindi la successione $\mathbf{t}_k$ converge all'autovettore $\mathbf{x}_1$ normalizzato in norma ∞.

Fissata una tolleranza ϵ, come condizione di arresto del metodo iterativo si può utilizzare una delle condizioni seguenti:

$$|\beta_{k+1} - \beta_k| < \epsilon, \tag{50}$$

o

$$\left|\frac{\beta_{k+1} - \beta_k}{\beta_{k+1}}\right| < \epsilon.$$

6.34 Esempio. Si consideri la matrice

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

che, come si è visto nell'esempio 2.36, ha due autovalori λ_1 e λ_2 in

$$\{\ z \in \mathbf{C} : |z - 15| \leq 3\ \} \cup \{\ z \in \mathbf{C} : |z - 10| \leq 3\ \},$$

e un autovalore λ_3 in

$$\{\ z \in \mathbf{C} : |z| \leq 3\ \}.$$

Il metodo delle potenze, applicato ad A normalizzando rispetto alla norma ∞, a partire dal vettore $\mathbf{t}_0 = [1, 1, 1]^T$, fornisce le seguenti successioni di valori che approssimano l'autovalore λ_1 e l'autovettore corrispondente:

k	β_k	$\mathbf{t}_k^T$		
1	15.00000	1.000000	0.5333333	-0.06666667
2	13.80000	1.000000	0.4734299	-0.1062801
3	13.84058	1.000000	0.4373472	-0.1102966
4	13.90471	1.000000	0.4102466	-0.1123829
⋮	⋮	⋮	⋮	⋮
41	14.10255	1.000000	0.3303270	-0.1183949

Con il criterio di arresto (50) e la tolleranza $\epsilon = 10^{-6}$, il metodo si arresta al 41-esimo passo fornendo i valori approssimati

$$\lambda_1 = 14.10255$$

e

$$\mathbf{x}_1 = [1.000000, 0.3303270, -0.1183949]^T. \qquad \blacksquare$$

Utilizzando la norma 2, sia $\|\mathbf{t}_0\|_2 = 1$ e sia $\beta_k = \|\mathbf{u}_k\|_2$. Questa scelta di β_k è particolarmente conveniente nel caso che la matrice A sia normale, perché si ottiene una successione che converge a λ_1 più velocemente che nel caso precedente. Infatti, tenendo conto che gli autovettori $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ di una matrice normale A possono essere scelti ortonormali, risulta che

$$\begin{aligned}\sigma_k = \mathbf{t}_k^H \mathbf{u}_{k+1} = \frac{\mathbf{t}_k^H A \mathbf{t}_k}{\mathbf{t}_k^H \mathbf{t}_k} &= \frac{(A^k\mathbf{t}_0)^H (A^{k+1}\mathbf{t}_0)}{(A^k\mathbf{t}_0)^H (A^k\mathbf{t}_0)} \\ &= \lambda_1 \, \frac{|\alpha_1|^2 + \sum_{i=2}^{n} |\alpha_i|^2 \left|\frac{\lambda_i}{\lambda_1}\right|^{2k} \left(\frac{\lambda_i}{\lambda_1}\right)}{|\alpha_1|^2 + \sum_{i=2}^{n} |\alpha_i|^2 \left|\frac{\lambda_i}{\lambda_1}\right|^{2k}} \\ &= \lambda_1 \left[1 + O\Big(\Big|\frac{\lambda_2}{\lambda_1}\Big|^{2k}\Big)\right].\end{aligned}$$

La successione dei σ_k converge a λ_1 e l'errore che si commette approssimando λ_1 con σ_k tende a zero con $|\lambda_2/\lambda_1|^{2k}$. Quindi la successione dei σ_k converge più rapidamente della successione dei β_k.

In questo caso, invece della (50), poiché la matrice A è normale, si può utilizzare come criterio di arresto la condizione

$$\|\mathbf{u}_{k+1} - \sigma_k \mathbf{t}_k\|_2 < \epsilon, \tag{51}$$

che oltre ad essere facilmente applicabile, fornisce una maggiorazione dell'errore assoluto: infatti per la (2) risulta che esiste un autovalore λ di A tale che

$$|\lambda - \sigma_k| \leq \frac{\|(A - \sigma_k I)\mathbf{t}_k\|_2}{\|\mathbf{t}_k\|_2} = \|\mathbf{u}_{k+1} - \sigma_k \mathbf{t}_k\|_2 < \epsilon.$$

In modo analogo, se la matrice A non è singolare, una condizione di arresto per l'errore relativo è data da

$$\frac{\|\mathbf{u}_{k+1} - \sigma_k \mathbf{t}_k\|_2}{\|\mathbf{u}_{k+1}\|_2} < \epsilon,$$

infatti per la (3) risulta che esiste un autovalore λ di A tale che

$$\Big|\frac{\lambda_1 - \sigma_k}{\lambda_1}\Big| \leq \frac{\|(A - \sigma_k I)A^{-1}\mathbf{u}_{k+1}\|_2}{\|\mathbf{u}_{k+1}\|_2} = \frac{\|\mathbf{u}_{k+1} - \sigma_k \mathbf{t}_k\|_2}{\|\mathbf{u}_{k+1}\|_2} < \epsilon.$$

Si noti che con la normalizzazione in norma 2 la successione $\{\mathbf{t}_k\}$ può non avere limite, ma per la (47) ogni successione $\{\dfrac{\mathbf{t}_k}{t_j^{(k)}},\ t_j^{(k)} \neq 0\}$ ha per limite l'autovettore $\mathbf{x}_1$ opportunamente normalizzato.

6.35 Esempio. Alla matrice

$$A = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 3 & 4 & 3 & 2 \\ 2 & 3 & 4 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

dell'esempio 6.17 si applica il metodo delle potenze con la normalizzazione di $\mathbf{t}_k$ rispetto alla norma ∞ a partire dal vettore $\mathbf{t}_0 = [1, 1, 1, 1]^T$ e rispetto alla norma 2, a partire dal vettore $\mathbf{t}_0 = [0.5, 0.5, 0.5, 0.5]^T$. Nella figura 6.2 sono riportati gli errori $|\beta_k - \lambda_1|$ (indicati con i quadratini vuoti) e $|\sigma_k - \lambda_1|$ (indicati con i quadratini pieni).

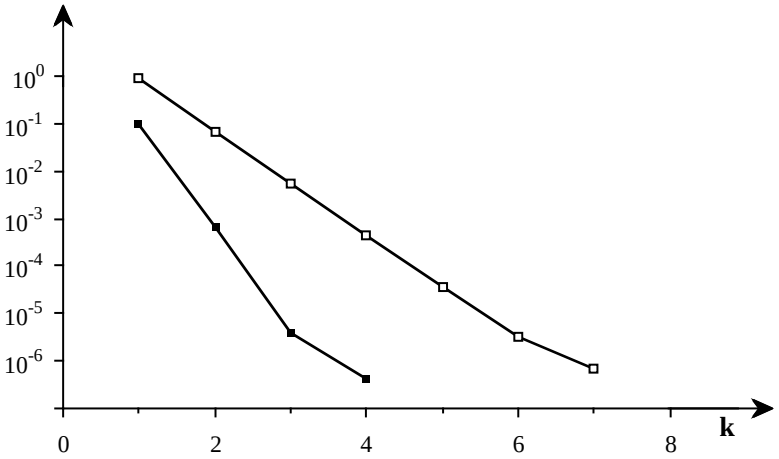

Fig. 6.2 - Errori del metodo delle potenze con la normalizzazione rispetto alla norma ∞ e alla norma 2.

Fissata la tolleranza $\epsilon = 10^{-6}$, il metodo si arresta alla settima iterazione quando la normalizzazione viene fatta rispetto alla norma ∞ e si usa il criterio (50) e alla quarta iterazione quando la normalizzazione viene fatta rispetto alla norma 2 e si usa il criterio (51). Si noti la maggiore velocità di convergenza della successione dei σ_k. ∎

Il metodo delle potenze è convergente anche nel caso in cui l'autovalore di modulo massimo abbia molteplicità algebrica maggiore di 1, cioè $\lambda_1 = \lambda_2 = \ldots = \lambda_r$, con

$$|\lambda_1| = |\lambda_2| = \ldots = |\lambda_r| > |\lambda_{r+1}| \geq \ldots \geq |\lambda_n|.$$

Infatti al posto della (45) si ha

$$\mathbf{y}_k = \lambda_1^k \Big[\sum_{i=1}^{r} \alpha_i \mathbf{x}_i + \sum_{i=r+1}^{n} \alpha_i \Big(\frac{\lambda_i}{\lambda_1} \Big)^k \mathbf{x}_i \Big];$$

l'autovalore λ_1 si approssima con la successione dei β_k o dei σ_k, e l'errore dell'approssimazione tende a zero come $(\lambda_{r+1}/\lambda_1)^k$ o come $|\lambda_{r+1}/\lambda_1|^{2k}$. Inoltre

$$\lim_{k\to\infty} \frac{\mathbf{y}_k}{y_j^{(k)}} = \frac{1}{\theta_j} \sum_{i=1}^{r} \alpha_i \mathbf{x}_i, \quad \text{dove} \quad \theta_j = \sum_{i=1}^{r} \alpha_i x_j^{(i)},$$

e quindi la successione $\{\mathbf{y}_k/y_m^{(k)}\}$, dove m è l'indice di una componente di massimo modulo di $\mathbf{y}_k$, converge ad un autovettore normalizzato in norma ∞ appartenente allo spazio vettoriale generato da $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r$.

Se invece esistono più autovalori di modulo massimo diversi fra loro, il metodo delle potenze non è convergente (si veda l'esercizio 6.35).

6.36 Esempio. La matrice

$$A = \begin{bmatrix} 8 & -1 & -5 \\ -4 & 4 & -2 \\ 18 & -5 & -7 \end{bmatrix}$$

ha gli autovalori $2 \pm 4\,\mathbf{i}$ e 1. Gli autovalori di modulo massimo sono quelli complessi e quindi sono distinti. Con il metodo delle potenze, applicato a partire dal vettore $\mathbf{t}_0 = [1, 1, 1]^T$, normalizzando $\mathbf{t}_k$ rispetto alla norma ∞, si ottiene la successione:

k	β_k
1	6.000000
2	-4.666667
3	3.714286
4	4.769224
5	-5.096744
⋮	⋮
44	-3.377194
45	5.844138
⋮	⋮

che risulta non convergente. ∎

Il metodo delle potenze può essere modificato in modo da approssimare anche autovalori distinti con lo stesso modulo, come nel caso di autovalori complessi coniugati [28] (si veda anche l'esercizio 6.34).

Come risulta dalle considerazioni precedenti, la condizione $\alpha_1 \neq 0$ è, in teoria, necessaria per la convergenza a λ_1 della successione (46). Se infatti

fosse $\alpha_1 = 0$, $\alpha_2 \neq 0$ e $|\lambda_2| > |\lambda_3|$, allora è possibile dimostrare con argomentazioni analoghe che la successione $\{y_j^{(k+1)}/y_j^{(k)}\}$ tende all'autovalore λ_2. In pratica però, anche se $\mathbf{t}_0$ fosse tale che $\alpha_1 = 0$, per la presenza degli errori di arrotondamento, i vettori $\mathbf{t}_k$ effettivamente calcolati sarebbero comunque rappresentabili come combinazioni lineari degli autovettori con una componente non nulla rispetto a $\mathbf{x}_1$. Perciò la successione effettivamente calcolata convergerebbe ugualmente a λ_1. Inoltre nel caso che le componenti del vettore $\mathbf{t}_0$ vengano scelte casualmente nell'insieme dei numeri complessi, la probabilità di ottenere un vettore per cui $\alpha_1 = 0$ è nulla.

11. Varianti del metodo delle potenze

Varianti del metodo delle potenze consentono di calcolare anche gli altri autovalori e i corrispondenti autovettori.

a) Variante di Wielandt (metodo delle potenze inverse)

Se A è una matrice non singolare, diagonalizzabile, con autovalori $\lambda_i,\ i = 1, \ldots, n$, tali che

$$|\lambda_1| \geq \ldots \geq |\lambda_{n-1}| > |\lambda_n| > 0,$$

la matrice A^{-1} ha autovalori $\dfrac{1}{\lambda_i},\ i = 1, \ldots, n$, tali che

$$\frac{1}{|\lambda_n|} > \frac{1}{|\lambda_{n-1}|} \geq \ldots \geq \frac{1}{|\lambda_1|}.$$

Per calcolare l'autovalore di modulo minimo di A si applica il metodo delle potenze alla matrice A^{-1}, nel modo seguente

$$\left.\begin{aligned} A\mathbf{u}_k &= \mathbf{t}_{k-1}, \\ \mathbf{t}_k &= \frac{1}{\beta_k}\,\mathbf{u}_k, \end{aligned}\right\}, \quad k = 1, 2, \ldots, \tag{52}$$

dove β_k è uno scalare tale che $\|\mathbf{t}_k\| = 1$, per la norma scelta. Ogni passo del metodo richiede la risoluzione del sistema lineare $A\mathbf{u}_k = \mathbf{t}_{k-1}$. Per $k \to \infty$ la successione dei β_k, se si usa la $\| \ . \ \|_\infty$, o dei σ_k, se si usa la $\| \ . \ \|_2$ e la matrice A è normale, tende a $\dfrac{1}{\lambda_n}$ e $\mathbf{t}_k$ tende al corrispondente autovettore della matrice A^{-1} (e quindi di A).

Se di un autovalore λ_j è nota una stima μ, tale che

$$0 < |\mu - \lambda_j| < |\mu - \lambda_i|, \quad j \neq i,$$

questo autovalore può essere calcolato, applicando il metodo delle potenze alla matrice $(A-\mu I)^{-1}$, nel modo seguente

$$\left.\begin{aligned}(A-\mu I)\mathbf{u}_k &= \mathbf{t}_{k-1},\\ \mathbf{t}_k &= \frac{1}{\beta_k}\,\mathbf{u}_k,\end{aligned}\right\},\quad k=1,2,\ldots, \tag{53}$$

dove β_k è uno scalare tale che $\|\mathbf{t}_k\|=1$, per la norma scelta. Per $k\to\infty$ la successione dei β_k o dei σ_k tende a $\dfrac{1}{\lambda_j-\mu}$ e $\mathbf{t}_k$ tende al corrispondente autovettore della matrice $(A-\mu I)^{-1}$ (e quindi di A).

L'autovalore λ_j viene calcolato a meno di un errore che tende a zero con

$$\Big(\frac{|\lambda_j-\mu|}{\min\{\,|\lambda_{j-1}-\mu|\,,\,|\lambda_{j+1}-\mu|\,\}}\Big)^k.$$

Questo metodo è spesso usato per migliorare l'approssimazione di un autovalore ottenuta con altri metodi. Va però rilevato che più μ è vicino a λ_j più rapida è la convergenza del metodo, ma aumentano le difficoltà numeriche nel calcolo di $\mathbf{u}_k$ perché la matrice $A-\mu I$ tende a diventare mal condizionata.

Per il calcolo effettivo della (52) o (53), conviene prima fattorizzare, con un costo computazionale di $n^3/3$ operazioni moltiplicative, la matrice A o la matrice $A-\mu I$ nella forma LU. Poi ad ogni passo si risolvono due sistemi con matrice dei coefficienti triangolare e il costo computazionale di questo metodo è quindi confrontabile ad ogni passo con quello del metodo delle potenze.

6.37 Esempio. Fissata la tolleranza $\epsilon=10^{-6}$, si applica il metodo di Wielandt, normalizzando $\mathbf{t}_k$ rispetto alla norma ∞, alla matrice

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

dell'esempio 6.34, ponendo $\mu=14$ e $\mathbf{t}_0=[1,1,1]^T$. Si ottengono le successioni:

k	β_k	$\mathbf{t}_k^T$		
1	9.399977	1.000000	0.3191492	-0.1276596
2	9.782953	1.000000	0.3305786	-0.1183123
3	9.749693	1.000000	0.3303205	-0.1183960
4	9.750801	1.000000	0.3303275	-0.1183950
5	9.750768	1.000000	0.3303272	-0.1183950
6	9.750769	1.000000	0.3303273	-0.1183950

cioè $\dfrac{1}{\lambda_1 - \mu} = 9.750769$, da cui si ricava $\lambda_1 = 14.10256$.

Con il metodo di Wielandt il risultato viene raggiunto con 6 passi, mentre con il metodo delle potenze (si veda l'esempio 6.34) occorrono 41 passi.

Ponendo $\mu = 13$ oppure $\mu = 15$ e partendo dallo stesso vettore $\mathbf{t}_0$, si ottengono ancora successioni convergenti, ma il numero di iterazioni richieste per ottenere la stessa precisione è maggiore (rispettivamente 17 e 9 iterazioni). Ponendo invece $\mu = 12$ e partendo dallo stesso vettore $\mathbf{t}_0$, la successione dei β_k converge in 50 iterazioni a -0.6193352, da cui si ricava l'autovalore $\lambda_2 = 10.38537$.

Per approssimare l'autovalore λ_3 della matrice A, si pone $\mu = 0$ (in questo caso il metodo di Wielandt coincide con il metodo delle potenze applicato alla matrice A^{-1}) e si ottiene per β_k la successione:

k	β_k
1	2.160000
2	1.959506
3	1.953039
4	1.952810
5	1.952802
6	1.952801

cioè $\dfrac{1}{\lambda_3} = 1.952801$, da cui si ricava l'approssimazione $\lambda_3 = 0.5120849$. ∎

Anche per il metodo di Wielandt valgono le stesse considerazioni fatte per il metodo delle potenze. In particolare, se μ si trova alla stessa distanza da due autovalori distinti, allora il metodo non è convergente, come risulta anche dall'esempio seguente.

6.38 Esempio. La matrice

$$A = \begin{bmatrix} 33 & 16 & 72 \\ -24 & -10 & -57 \\ -8 & -4 & -17 \end{bmatrix}$$

ha gli autovalori $\lambda_1 = 3$, $\lambda_2 = 2$, $\lambda_3 = 1$. Ponendo $\mu = 2.5$, a partire da $\mathbf{t}_0 = [1, 1, 1]^T$, si ottiene la successione:

k	β_k
1	73.99426
2	3.657217
3	0.4363987
4	10.16883
⋮	⋮
97	0.2937573
98	13.61612
99	0.2933033

La successione β_k non è convergente perché il valore scelto per μ è equidistante dai due autovalori λ_1 e λ_2. ∎

b) Metodo delle iterazioni del quoziente di Rayleigh

Questo metodo è una variante del metodo di Wielandt applicato a una matrice hermitiana con la normalizzazione in norma 2. La (53) viene così modificata

$$\left.\begin{aligned} \mu_{k-1} &= \mathbf{t}_{k-1}^H A \mathbf{t}_{k-1}, \\ (A - \mu_{k-1} I)\mathbf{u}_k &= \mathbf{t}_{k-1}, \\ \mathbf{t}_k &= \frac{1}{\|\mathbf{u}_k\|_2}\, \mathbf{u}_k, \end{aligned}\right\}, \quad k = 1, 2, \ldots \tag{54}$$

Si può dimostrare [17], in modo analogo a quanto fatto nel caso del metodo delle potenze, che la successione dei μ_k converge ad un autovalore λ della matrice A e che localmente la convergenza è del terzo ordine (per il caso che la matrice abbia autovalori distinti, si veda l'esercizio 6.30). Però ogni passo del metodo richiede in generale un numero di operazioni moltiplicative dell'ordine di $n^3/6$, perché la matrice del sistema (54) cambia ogni volta ed è hermitiana. Inoltre all'aumentare di k aumenta il numero di condizionamento della matrice $A - \mu_{k-1} I$ e quindi aumentano le difficoltà numeriche del calcolo di $\mathbf{u}_k$.

6.39 Esempio. Si calcola l'autovalore λ_1 della matrice

$$A = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 3 & 4 & 3 & 2 \\ 2 & 3 & 4 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

dell'esempio 6.17 con il metodo di Wielandt con la normalizzazione in norma ∞, $\mu = 10$ e $\mathbf{t}_0 = [1, 1, 1, 1]^T$, il metodo di Wielandt con la normalizzazione

in norma 2, $\mu = 10$ e $\mathbf{t}_0 = [0.5, 0.5, 0.5, 0.5]^T$, e il metodo del quoziente di Rayleigh con $\mu_0 = 10$ e $\mathbf{t}_0 = [0.5, 0.5, 0.5, 0.5]^T$. Nella figura 6.3 sono riportati gli errori assoluti della successione β_k (indicati con quadratini vuoti), ottenuta con il metodo di Wielandt con la normalizzazione in norma ∞, gli errori assoluti della successione σ_k (indicati con quadratini pieni), ottenuta con il metodo di Wielandt con la normalizzazione in norma 2 e gli errori assoluti della successione μ_k (indicati con triangolini), ottenuta con il metodo del quoziente di Rayleigh. Si confrontino questi risultati con quelli ottenuti con metodo delle potenze e riportati nella figura 6.2. ∎

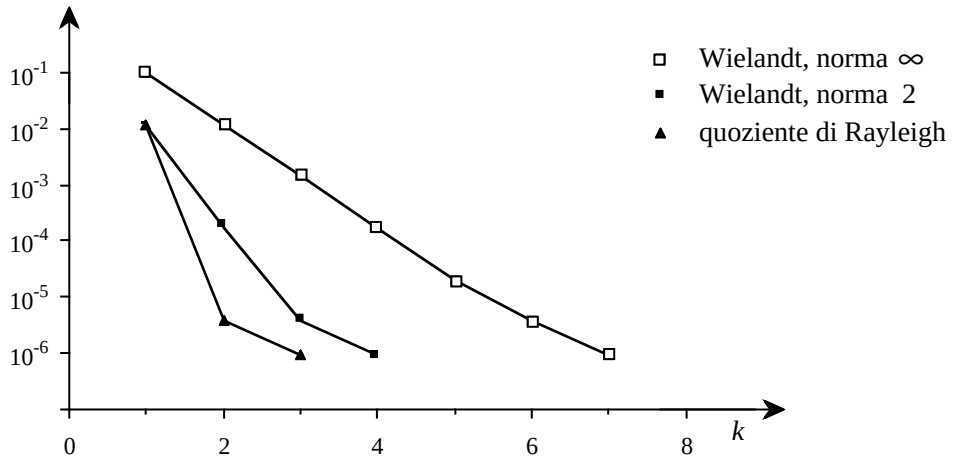

Fig. 6.3 - Errori delle soluzioni ottenute con il metodo di Wielandt con la normalizzazione rispetto alla norma ∞ e alla norma 2 e con il metodo del quoziente di Rayleigh.

c) Variante dell'ortogonalizzazione - 1

Sia A normale e tale che $|\lambda_1| > |\lambda_2| > \ldots > |\lambda_n|$. Dopo aver calcolato λ_1 e $\mathbf{x}_1$, con $\|\mathbf{x}_1\|_2 = 1$, si considera un qualunque vettore $\mathbf{y} \in \mathbf{C}^n$, $\mathbf{y} \neq \mathbf{0}$ e si applica il metodo delle potenze con la normalizzazione rispetto alla norma 2 partendo dal vettore

$$\mathbf{t}_0 = \frac{\mathbf{z}}{\|\mathbf{z}\|_2}\ , \quad \mathbf{z} = \mathbf{y} - (\mathbf{x}_1^H \mathbf{y})\mathbf{x}_1, \tag{55}$$

ortogonale a $\mathbf{x}_1$. Poiché i vettori $\mathbf{t}_k$ generati con il metodo delle potenze sono (in teoria) ortogonali a $\mathbf{x}_1$, il metodo calcola λ_2. In pratica però, per effetto degli errori di arrotondamento, i vettori $\mathbf{t}_k$ effettivamente calcolati hanno una componente diversa da zero lungo la direzione $\mathbf{x}_1$ che si accentua al crescere di k. Quindi per ottenere una successione dei σ_k che non converga nuovamente a λ_1, occorre *riortogonalizzare*, dopo un certo numero di passi, $\mathbf{t}_k$ rispetto a $\mathbf{x}_1$. Cioè ogni m passi, dove m è un intero opportuno, si

sostituisce il vettore $\mathbf{t}_k$ con il vettore

$$\mathbf{t}'_k = \frac{\mathbf{z}}{\|\mathbf{z}\|_2}\ , \quad \mathbf{z} = \mathbf{t}_k - (\mathbf{x}_1^H \mathbf{t}_k)\mathbf{x}_1.$$

In modo analogo, calcolati gli autovalori $\lambda_1, \lambda_2, \ldots, \lambda_j$ e i corrispondenti autovettori $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_j$, tutti di norma 2 unitaria, è possibile calcolare λ_{j+1} scegliendo

$$\mathbf{t}_0 = \frac{\mathbf{z}}{\|\mathbf{z}\|_2}\ , \quad \mathbf{z} = \mathbf{y} - \sum_{i=1}^{j} (\mathbf{x}_i^H \mathbf{y})\mathbf{x}_i,$$

in modo che $\mathbf{t}_0$ risulti ortogonale a $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_j$. Anche in questo caso occorre effettuare ogni m passi il processo di riortogonalizzazione, che richiede $2jn$ operazioni moltiplicative.

d) Variante dell'ortogonalizzazione - 2

Sia A normale e tale che $|\lambda_1| > |\lambda_2| > \ldots > |\lambda_n|$. Si ha

$$A = \sum_{i=1}^{n} \lambda_i \mathbf{x}_i \mathbf{x}_i^H, .$$

dove $\mathbf{x}_1, \ldots, \mathbf{x}_n$ sono autovettori ortonormali. La matrice

$$A_1 = A - \lambda_1 \mathbf{x}_1 \mathbf{x}_1^H$$

ha autovalori $\lambda_2, \lambda_3, \ldots, \lambda_n$, e 0. Quindi calcolati λ_1 e $\mathbf{x}_1$, il metodo delle potenze, applicato ad A_1 approssima λ_2. In generale, calcolati λ_1, $\lambda_2, \ldots, \lambda_j$ e $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_j$, per calcolare λ_{j+1} si applica il metodo delle potenze alla matrice

$$A - \sum_{i=1}^{j} \lambda_i \mathbf{x}_i \mathbf{x}_i^H.$$

Se la matrice A è sparsa, per utilizzare questa proprietà ad ogni passo il metodo delle potenze viene applicato nel modo seguente

$$\mathbf{u}_k = A\mathbf{t}_{k-1} - \sum_{i=1}^{j} \lambda_i (\mathbf{x}_i^H \mathbf{t}_{k-1})\mathbf{x}_i, \quad k = 1, 2, \ldots,$$

con un aumento ad ogni passo di $2jn$ operazioni moltiplicative.

6.40 Esempio. Fissata una tolleranza $\epsilon = 10^{-6}$ si applica il metodo delle potenze alla matrice dell'esempio 6.17

$$A = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 3 & 4 & 3 & 2 \\ 2 & 3 & 4 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

per approssimare tutti gli autovalori (si veda per confronto il calcolo con il metodo QR nell'esempio 6.30). Calcolato il primo autovalore $\lambda_1 = 11.09901$ e il corrispondente autovettore $\mathbf{x}_1$, e posto $\mathbf{y} = [0.5, 0.5, 0.5, 0.5]^T$, si calcola il vettore $\mathbf{t}_0$, ortogonale a $\mathbf{x}_1$, con la (55). La successione dei σ_k che si ottiene è la seguente

k	σ_k
1	0.7694202
2	2.590006
3	3.351716
4	3.410331
⋮	⋮
10	3.414937
11	3.421915
12	3.494808
13	4.188173
14	7.579575
15	10.53008
16	11.04131

Si noti come dopo la decima iterazione per effetto di una progressiva perdita di ortogonalità di $\mathbf{t}_k$ rispetto a $\mathbf{x}_1$, la successione dei σ_k tenda nuovamente a λ_1. Se invece si riortogonalizza ogni 5 passi $\mathbf{t}_k$ rispetto a $\mathbf{x}_1$, si ottiene la successione

k	σ_k
⋮	⋮
4	3.410331
5	3.413966
6	3.414192
7	3.414209
8	3.414209

che converge a $\lambda_2 = 3.414209$. Il calcolo dei successivi autovalori diventa sempre più complicato: riortogonalizzando ogni 5 passi si determina λ_3 solo dopo 18 iterazioni, mentre non si riesce a determinare λ_4. Solamente riortogonalizzando ogni 2 passi si riesce a calcolare $\lambda_4 = 0.5857863$ in 7 iterazioni.

Con la seconda variante, applicando il metodo delle potenze alle matrici

$$A_1 = A - \lambda_1 \mathbf{x}_1 \mathbf{x}_1^T, \quad A_2 = A_1 - \lambda_2 \mathbf{x}_2 \mathbf{x}_2^T, \quad A_3 = A_2 - \lambda_3 \mathbf{x}_3 \mathbf{x}_3^T,$$

si ottengono risultati migliori: il numero di passi richiesti risulta infatti di 9 per λ_2, 4 per λ_3 e 7 per λ_4. ∎

e) Variante della deflazione

Sia $|\lambda_1| > |\lambda_2|$. Calcolati λ_1 e $\mathbf{x}_1$, di norma 2 unitaria, si considera la matrice di Householder P tale che $P\mathbf{x}_1 = \mathbf{e}_1$; risulta

$$PAP^H = \begin{bmatrix} \lambda_1 & \mathbf{0}^H \\ \mathbf{0} & A_1 \end{bmatrix},$$

se A è hermitiana o

$$PAP^H = \begin{bmatrix} \lambda_1 & \mathbf{a}^H \\ \mathbf{0} & A_1 \end{bmatrix},$$

se A non lo è.

Si applica il metodo delle potenze alla matrice A_1 di ordine $n-1$ e si calcolano λ_2 e il corrispondente autovettore $\mathbf{y}_2$ di A_1. L'autovettore $\mathbf{x}_2$ di A corrispondente a λ_2 è dato da

$$\mathbf{x}_2 = P^H \begin{bmatrix} \theta \\ \mathbf{y}_2 \end{bmatrix}, \quad \text{con} \quad \theta = \begin{cases} 0 & \text{se } A \text{ è hermitiana,} \\ \dfrac{\mathbf{a}^H \mathbf{y}_2}{\lambda_2 - \lambda_1} & \text{se } A \text{ non lo è.} \end{cases}$$

Procedendo in questo modo si costruisce la forma di Schur della matrice A. Poiché la trasformazione $A \to PAP^H$ può distruggere la eventuale struttura e sparsità di A, questo procedimento può non essere indicato per matrici sparse.

6.41 Esempio. Fissata una tolleranza $\epsilon = 10^{-6}$, si applica il metodo delle potenze con la variante della deflazione alla matrice

$$A = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 3 & 4 & 3 & 2 \\ 2 & 3 & 4 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

dell'esempio 6.17 (si veda l'esempio 6.40 per la variante dell'ortogonalizzazione). Calcolato il primo autovalore $\lambda_1 = 11.09901$ e il corrispondente autovettore $\mathbf{x}_1$, si ottiene la matrice

$$A_1 = \begin{bmatrix} 0.7225373 & 0.1001084 & \text{-}0.1358454 \\ 0.1001084 & 1.477678 & 1.173688 \\ \text{-}0.1358454 & 1.173688 & 2.700763 \end{bmatrix}.$$

Applicando nuovamente il metodo delle potenze ad A_1, a partire dal vettore

$$\mathbf{t}_0 = \frac{1}{\sqrt{3}} [1, 1, 1]^T,$$

si calcola il secondo autovalore $\lambda_2 = 3.414209$ e il corrispondente autovettore $\mathbf{y}_2$ di A_1 in 9 passi. Si ottiene poi la matrice

$$A_2 = \begin{bmatrix} 0.8919271 & 0.05264682 \\ 0.05264682 & 0.5948396 \end{bmatrix},$$

a cui si riapplica il metodo delle potenze, a partire dal vettore

$$\mathbf{t}_0 = \frac{1}{\sqrt{2}} [1, 1]^T,$$

e occorrono 18 iterazioni per calcolare λ_3. ∎

12. Metodo delle iterazioni ortogonali

Questo metodo, noto anche con il nome di *metodo delle iterazioni di sottospazi*, è un'estensione a blocchi del metodo delle potenze ed è particolarmente conveniente quando la matrice A è sparsa e di grandi dimensioni e sono richiesti solo pochi dei suoi autovalori di maggior modulo. Come il metodo delle potenze, anche questo si basa sul fatto che se $\lambda_1, \lambda_2, \ldots, \lambda_n$ sono autovalori della matrice A, allora $\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k$ sono autovalori di A^k e che se alcuni di essi sono *dominanti* sugli altri, cioè di modulo maggiore degli altri, questa dominanza diventa sempre più grande per gli autovalori di A^k, al crescere di k.

Il metodo può essere applicato a matrici qualsiasi, anche se qui viene presentato solo il caso delle matrici hermitiane con autovalori di modulo distinto, per le quali è possibile applicare una tecnica di accelerazione che rende il metodo molto efficiente.

Sia $A \in \mathbf{C}^{n \times n}$, hermitiana, siano $\lambda_1, \lambda_2, \ldots, \lambda_n$ i suoi autovalori, tali che

$$|\lambda_1| > |\lambda_2| > \ldots > |\lambda_n|,$$