Capitolo 2

AUTOVALORI E AUTOVETTORI

1. Definizioni

Siano $A \in \mathbf{C}^{n\times n}, \lambda \in \mathbf{C}$ e $\mathbf{x} \in \mathbf{C}^n, \mathbf{x} \neq \mathbf{0}$, tali che valga la relazione

$$A\mathbf{x} = \lambda\mathbf{x}. \tag{1}$$

Allora λ è detto *autovalore* di A ed $\mathbf{x}$ è detto *autovettore* corrispondente a λ. L'insieme degli autovalori di una matrice A costituisce lo *spettro* di A e il modulo massimo $\rho(A)$ degli autovalori di A è detto *raggio spettrale* di A.

Il sistema (1), che si può scrivere anche nella forma

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

ammette soluzioni non nulle se e solo se

$$\det(A - \lambda I) = 0. \tag{3}$$

Sviluppando $\det(A - \lambda I)$ risulta

$$\det(A - \lambda I) = P(\lambda) = a_0\lambda^n + a_1\lambda^{n-1} + \ldots + a_n,$$

in cui

$$a_0 = (\text{-}1)^n, \qquad a_i = (\text{-}1)^{n-i}\sigma_i, \quad i = 1, \ldots, n,$$

dove σ_i è la somma dei determinanti delle $\binom{n}{i}$ sottomatrici principali di A di ordine i. In particolare risulta:

$$a_1 = (\text{-}1)^{n-1}\operatorname{tr} A, \qquad a_n = \det A,$$

in cui si è indicato con $\operatorname{tr} A$ la *traccia* di A, cioè la somma degli elementi principali di A.

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

$$\sum_{i=1}^{n} \lambda_i = \operatorname{tr} A \quad \text{e} \quad \prod_{i=1}^{n} \lambda_i = \det A. \tag{4}$$

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

Per il teorema fondamentale dell'algebra l'equazione caratteristica ha nel campo complesso n radici, tenendo conto della loro molteplicità. Quindi una matrice di ordine n ha, tenendo conto della loro molteplicità, n autovalori nel campo complesso.

Poiché gli autovettori sono soluzioni non nulle del sistema lineare omogeneo (2), un autovettore corrispondente ad un autovalore λ risulta determinato a meno di una costante moltiplicativa $\alpha \neq 0$, cioè se $\mathbf{x}$ è autovettore di A, anche $\alpha\mathbf{x}$ è autovettore di A, corrispondente allo stesso autovalore.

2.1 Esempio. Il polinomio caratteristico della matrice

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

si ricava dal determinante

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

L'equazione caratteristica corrispondente

$$\lambda^2 - 2\lambda - 8 = 0$$

ha come radici $\lambda_1 = -2$ e $\lambda_2 = 4$, che sono gli autovalori della matrice A. L'autovettore corrispondente a $\lambda_1 = -2$ si calcola risolvendo il sistema (2) che in questo caso diventa

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

Dalla prima equazione si ottiene

$$x_1 + x_2 = 0, \quad \text{da cui} \quad x_1 = -x_2,$$

da cui segue che qualunque vettore

$$\mathbf{x}_1 = \alpha \begin{bmatrix} 1 \\ \text{-1} \end{bmatrix}$$

con $\alpha \neq 0$, è autovettore della matrice A corrispondente all'autovalore $\lambda_1 = -2$. L'autovettore corrispondente a $\lambda_2 = 4$ si determina risolvendo il sistema

$$\begin{bmatrix} \text{-3} & 3 \\ 3 & \text{-3} \end{bmatrix} \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{0}.$$

Dalla prima equazione si ottiene

$$-x_1 + x_2 = 0, \quad \text{da cui} \quad x_1 = x_2,$$

da cui segue che qualunque vettore

$$\mathbf{x}_1 = \alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

con $\alpha \neq 0$, è autovettore della matrice A corrispondente all'autovalore $\lambda_2 = 4$. ∎

2. Proprietà degli autovalori

– Gli autovalori di una matrice A diagonale o triangolare (superiore o inferiore) sono uguali agli elementi principali. Infatti la matrice $A - \lambda I$ è ancora diagonale o triangolare e quindi il suo determinante è dato dal prodotto degli elementi principali.

– Se λ è un autovalore di una matrice A non singolare e $\mathbf{x}$ un autovettore corrispondente, allora risulta $\lambda \neq 0$ e $1/\lambda$ è autovalore di A^{-1} con $\mathbf{x}$ autovettore corrispondente. Infatti da

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

si ha

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

e quindi

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

– Se λ è un autovalore di una matrice A, allora $\overline{\lambda}$ è autovalore di A^H e λ è autovalore di A^T. Infatti nel primo caso, poiché

$$\det A^H = \overline{\det A},$$

si ha

$$0 = \det(A - \lambda I) = \overline{\det(A - \lambda I)^H} = \overline{\det(A^H - \overline{\lambda} I)},$$

da cui

$$\det(A^H - \overline{\lambda} I) = 0.$$

Si procede in modo analogo per il secondo caso.

– Se λ è autovalore di una matrice unitaria A, cioè tale che $A^H A = AA^H = I$, allora risulta $|\lambda| = 1$. Infatti dalla relazione $A\mathbf{x} = \lambda\mathbf{x}$ si ha

$$(A\mathbf{x})^H = (\lambda\mathbf{x})^H$$

e quindi

$$\mathbf{x}^H A^H = \overline{\lambda}\mathbf{x}^H,$$

da cui si ha

$$\mathbf{x}^H A^H A\mathbf{x} = \overline{\lambda}\lambda\, \mathbf{x}^H\mathbf{x}.$$

Poiché A è unitaria, risulta

$$\mathbf{x}^H\mathbf{x} = \overline{\lambda}\lambda\mathbf{x}^H\mathbf{x},$$

e quindi, essendo $\mathbf{x}^H\mathbf{x} \neq 0$, segue

$$\overline{\lambda}\lambda = |\lambda|^2 = 1.$$

2.2 Esempio. La matrice G dell'esempio 1.1:

$$G = \begin{bmatrix} \cos\phi & \text{-}\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}, \qquad \phi \in \mathbf{R},$$

è unitaria e quindi ha autovalori di modulo 1. Infatti dall'equazione caratteristica

$$\lambda^2 - 2\lambda\cos\phi + 1 = 0$$

risulta

$$\lambda_1 = \cos\phi + \mathbf{i}\sin\phi \quad \text{e} \quad \lambda_2 = \cos\phi - \mathbf{i}\sin\phi. \qquad \blacksquare$$

Particolarmente interessanti sono i polinomi di matrici. Sia

$$p(x) = \alpha_0 x^k + \alpha_1 x^{k-1} + \ldots + \alpha_k,$$

dove $\alpha_0, \alpha_1, \ldots, \alpha_k \in \mathbf{C}$, un polinomio di grado k nella variabile x e sia $A \in \mathbf{C}^{n\times n}$. Un *polinomio della matrice* A è una matrice della forma

$$p(A) = \alpha_0 A^k + \alpha_1 A^{k-1} + \ldots + \alpha_k I$$

(si veda l'esercizio 1.3). Se λ è un autovalore di A e $\mathbf{x}$ è un autovettore corrispondente, allora $p(\lambda)$ è autovalore di $p(A)$ e $\mathbf{x}$ è un autovettore corrispondente. Risulta infatti che

$$A^i\mathbf{x} = A^{i-1}A\mathbf{x} = A^{i-1}\lambda\mathbf{x} = \lambda A^{i-1}\mathbf{x} = \lambda A^{i-2}A\mathbf{x} = \ldots = \lambda^i\mathbf{x},$$

e quindi

$$\begin{aligned} p(A)\mathbf{x} &= \alpha_0 A^k\mathbf{x} + \alpha_1 A^{k-1}\mathbf{x} + \ldots + \alpha_k\mathbf{x} = \alpha_0\lambda^k\mathbf{x} + \alpha_1\lambda^{k-1}\mathbf{x} + \ldots + \alpha_k\mathbf{x} \\ &= p(\lambda)\mathbf{x}. \end{aligned}$$

2.3 Esempio. La matrice

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

ha gli autovalori

$$\lambda_1 = 1 - \sqrt{2}, \quad \lambda_2 = 1, \quad \lambda_3 = 1 + \sqrt{2},$$

e i corrispondenti autovettori sono

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ -\sqrt{2} \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 1 \\ \sqrt{2} \\ 1 \end{bmatrix}.$$

La matrice

$$B = 3A^2 - A + 2I = \begin{bmatrix} 7 & 5 & 3 \\ 5 & 10 & 5 \\ 3 & 5 & 7 \end{bmatrix},$$

ha gli autovalori

$$\mu_i = 3\lambda_i^2 - \lambda_i + 2, \quad i = 1, 2, 3,$$

cioè

$$\mu_1 = 10 - 5\sqrt{2}, \quad \mu_2 = 4, \quad \mu_3 = 10 + 5\sqrt{2};$$

gli autovettori di B sono gli stessi di A. ■

2.4 Esempio. La matrice

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

verifica la relazione

$$A^2 - 5A + 4I = 0. \tag{5}$$

Quindi, se λ è autovalore di A, $\lambda^2 - 5\lambda + 4$ è autovalore della matrice nulla. Perciò deve essere $\lambda^2 - 5\lambda + 4 = 0$, da cui segue che gli autovalori di A sono $\lambda_1 = 1$ e $\lambda_2 = 4$. Poiché uno di questi autovalori ha molteplicità 2, e la somma degli autovalori, che è uguale alla traccia della matrice, è 6, λ_1 risulta di molteplicità 2.

La relazione (5) può essere utilizzata per calcolare la matrice inversa di A. Si ha infatti

$$A(5I - A) = 4I$$

da cui

$$A^{-1} = \frac{1}{4}\,(5I - A) = \frac{1}{4}\begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix}. \qquad \blacksquare$$

Più in generale è possibile dimostrare il seguente

2.5 Teorema (di Cayley-Hamilton). *Sia $A \in \mathbf{C}^{n\times n}$ e sia $P(\lambda)$ il suo polinomio caratteristico. Allora*

$$P(A) = 0.$$

Dim. Per un $\lambda \in \mathbf{C}$ sia $C = A - \lambda I$. Sia B la matrice aggiunta di C, per cui vale (si veda l'esercizio 1.48)

$$CB = (\det C)I. \tag{6}$$

Gli elementi di B sono determinanti di sottomatrici di ordine $n-1$ della matrice $A - \lambda I$, e quindi sono polinomi in λ di grado al più $n-1$. La matrice B può essere così espressa

$$B = \lambda^{n-1}B_0 + \lambda^{n-2}B_1 + \ldots + B_{n-1},$$

dove B_j, $j = 0, 1, \ldots, n-1$ sono matrici di ordine n. Dalla (6) si ha

$$\begin{aligned}(\det C)I &= (A - \lambda I)(\lambda^{n-1}B_0 + \lambda^{n-2}B_1 + \ldots + B_{n-1}) \\ &= -\lambda^n B_0 + \lambda^{n-1}(AB_0 - B_1) + \lambda^{n-2}(AB_1 - B_2) + \ldots + AB_{n-1}.\end{aligned}$$

D'altra parte

$$\det C = \det(A - \lambda I) = P(\lambda) = a_0\lambda^n + a_1\lambda^{n-1} + \ldots + a_n,$$

per cui uguagliando termine a termine si ha:

$$\begin{aligned} a_0 I &= -B_0 \\ a_1 I &= AB_0 - B_1 \\ a_2 I &= AB_1 - B_2 \\ &\;\;\vdots \\ a_n I &= AB_{n-1}. \end{aligned}$$

Moltiplicando queste relazioni rispettivamente per $A^n, A^{n-1}, \ldots, I$ e sommando, si ottiene

$$\begin{aligned} & a_0 A^n + a_1 A^{n-1} + a_2 A^{n-2} + \ldots + a_n I \\ & \quad = -A^n B_0 + A^{n-1}(AB_0 - B_1) + A^{n-2}(AB_1 - B_2) + \ldots + AB_{n-1} = 0. \ \blacksquare \end{aligned}$$

Dal teorema di Cayley-Hamilton segue quindi che una qualsiasi matrice A annulla il suo polinomio caratteristico $P(\lambda)$ e anche ogni altro polinomio di cui $P(\lambda)$ sia fattore.

Il polinomio *monico* (cioè con primo coefficiente uguale a 1) $\psi(\lambda)$ di grado minimo che è annullato da A è detto *polinomio minimo* di A ed è un fattore di $P(\lambda)$ e di ogni altro polinomio $p(\lambda)$ che sia annullato da A. Si ha infatti

$$p(\lambda) = \psi(\lambda)s(\lambda) + r(\lambda),$$

dove il grado di $r(\lambda)$ è minore di quello di $\psi(\lambda)$. Poiché

$$0 = p(A) = \psi(A)s(A) + r(A)$$

e $\psi(A) = 0$, ne segue che $r(A) = 0$, ed essendo $\psi(\lambda)$ il polinomio di grado minimo annullato da A, ne segue che $r(\lambda)$ è identicamente nullo.

Poiché $\psi(\lambda)$ è fattore di $P(\lambda)$, gli zeri di $\psi(\lambda)$ devono essere autovalori di A. Viceversa ogni autovalore di A è uno zero di $\psi(\lambda)$. Infatti se μ è autovalore di A, $\psi(\mu)$ è autovalore di $\psi(A)$, ed essendo $\psi(A) = 0$, ne segue che $\psi(\mu) = 0$.

Risulta perciò che $\psi(\lambda)$ è della forma

$$\psi(\lambda) = (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_p)^{n_p},$$

dove $\lambda_1, \lambda_2, \ldots, \lambda_p$ sono gli autovalori distinti di A, e $n_1 + n_2 + \ldots + n_p \leq n$. Se la matrice A ha tutti gli autovalori distinti, allora è

$$P(\lambda) = (-1)^n \psi(\lambda).$$

2.6 Esempio. La matrice A dell'esempio 2.4 annulla il polinomio

$$\psi(\lambda) = \lambda^2 - 5\lambda + 4,$$

che è il suo polinomio minimo. Infatti non esiste alcuna costante α tale che $A + \alpha I = 0$ e quindi nessun polinomio di grado 1 che sia annullato da A. $\blacksquare$

3. Proprietà degli autovettori

2.7 Teorema. *Autovettori corrispondenti ad autovalori distinti sono linearmente indipendenti.*

Dim. Siano $\lambda_1, \ldots, \lambda_m$, $m \leq n$, m autovalori di $A \in \mathbf{C}^{n \times n}$ a due a due distinti, e siano $\mathbf{x}_1, \ldots, \mathbf{x}_m$ i corrispondenti autovettori. Si procede per induzione su m.

Per $m = 1, \mathbf{x}_1 \neq 0$, quindi $\mathbf{x}_1$ è linearmente indipendente.

Per $m > 1$, si supponga per assurdo che esista una combinazione lineare dei vettori $\mathbf{x}_1, \ldots, \mathbf{x}_m$ tale che

$$\sum_{i=1}^{m} \alpha_i \, \mathbf{x}_i = \mathbf{0}, \tag{7}$$

in cui non tutti gli α_i siano nulli e sia j tale che $\alpha_j \neq 0$. In tal caso esiste almeno un altro indice $k \neq j$, per cui $\alpha_k \neq 0$; altrimenti, se ciò non fosse, seguirebbe che $\mathbf{x}_j = 0$. Moltiplicando entrambi i membri della (7) per A, si ottiene

$$\mathbf{0} = A \sum_{i=1}^{m} \alpha_i \, \mathbf{x}_i = \sum_{i=1}^{m} \alpha_i \, A\mathbf{x}_i = \sum_{i=1}^{m} \alpha_i \, \lambda_i \, \mathbf{x}_i, \tag{8}$$

moltiplicando entrambi i membri della (7) per λ_j, si ottiene

$$\mathbf{0} = \lambda_j \sum_{i=1}^{m} \alpha_i \, \mathbf{x}_i = \sum_{i=1}^{m} \alpha_i \, \lambda_j \, \mathbf{x}_i. \tag{9}$$

Sottraendo membro a membro la (9) dalla (8), si ha:

$$\mathbf{0} = \sum_{i=1}^{m} \alpha_i (\lambda_i - \lambda_j) \mathbf{x}_i = \sum_{\substack{i=1 \\ i \neq j}}^{m} \alpha_i (\lambda_i - \lambda_j) \mathbf{x}_i.$$

Si ottiene così una combinazione lineare nulla degli $m-1$ autovettori $\mathbf{x}_i \neq \mathbf{0}$, $\quad i = 1, \ldots, m$, $i \neq j$, in cui $\lambda_i - \lambda_j \neq 0$ per $i \neq j$ e gli α_i per $i \neq j$ non sono tutti nulli, essendo $\alpha_k \neq 0$, ciò che è assurdo perché per l'ipotesi induttiva gli $m-1$ vettori sono linearmente indipendenti. ■

Dal teorema 2.7 risulta che se una matrice A di ordine n ha n autovalori tutti distinti, allora A ha n autovettori linearmente indipendenti. Se la matrice A non ha n autovalori distinti, A può non avere n autovettori linearmente indipendenti, come risulta nell'esempio seguente.

2.8 Esempio. La matrice

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

il cui polinomio caratteristico è $(\lambda-3)^2$, ha 3 come autovalore di molteplicità 2. Poiché ad esso corrispondono solo autovettori della forma

$$\alpha \begin{bmatrix} 1 \\ \text{-1} \end{bmatrix}, \quad \alpha \neq 0,$$

la matrice A non può avere due autovettori linearmente indipendenti. ∎

Le matrici con n autovalori distinti non sono le sole ad avere n autovettori linearmente indipendenti. Ad esempio la matrice identica I_n che ha 1 come autovalore di molteplicità n, ha i vettori $\mathbf{e}_i \in \mathbf{C}^n$, $i = 1, \ldots, n$, della base canonica di $\mathbf{C}^n$ come autovettori.

Nel caso in cui ad uno stesso autovalore corrispondano più autovettori linearmente indipendenti, essi generano un sottospazio lineare i cui vettori non nulli sono tutti autovettori della matrice corrispondenti allo stesso autovalore. Vale infatti il seguente

2.9 Teorema. *Sia $A \in \mathbf{C}^{n\times n}$, e siano $\mathbf{x}_1, \mathbf{x}_2 \ldots, \mathbf{x}_k$ k autovettori linearmente indipendenti, corrispondenti ad uno stesso autovalore λ di A. Allora un vettore $\mathbf{y} \in \mathbf{C}^n$, $\mathbf{y} \neq \mathbf{0}$, della forma*

$$\mathbf{y} = \sum_{j=1}^{k} \alpha_j \, \mathbf{x}_j$$

è autovettore di A.

Dim. Si ha infatti

$$A\mathbf{y} = A \sum_{j=1}^{k} \alpha_j \, \mathbf{x}_j = \sum_{j=1}^{k} \alpha_j A\mathbf{x}_j = \sum_{j=1}^{k} \alpha_j \lambda \mathbf{x}_j = \lambda \sum_{j=1}^{k} \alpha_j \, \mathbf{x}_j = \lambda \, \mathbf{y}. \quad ∎$$

2.10 Definizione. La molteplicità di un autovalore λ come radice dell'equazione caratteristica, è indicata con $\sigma(\lambda)$, ed è detta *molteplicità algebrica* di λ. Il massimo numero di autovettori linearmente indipendenti corrispondenti a λ è indicato con $\tau(\lambda)$ ed è detto *molteplicità geometrica* di λ. ∎

La molteplicità geometrica $\tau(\lambda)$ è uguale alla dimensione del sottospazio vettoriale generato dagli autovettori corrispondenti a λ, e quindi uguale alla dimensione dello spazio

$$N(A - \lambda I) = \{ \, \mathbf{x} \in \mathbf{C}^n \ : \ (A - \lambda I)\mathbf{x} = \mathbf{0} \, \},$$

nucleo di $A - \lambda I$. È evidente che

$$1 \leq \sigma(\lambda) \leq n \quad \text{e} \quad 1 \leq \tau(\lambda) \leq n.$$

2.11 Teorema. *Vale la disuguaglianza*

$$\tau(\lambda) \le \sigma(\lambda).$$

Dim. Sia μ un autovalore di A con molteplicità algebrica $\sigma = \sigma(\mu)$ e geometrica $\tau = \tau(\mu)$. Per la (7) del capitolo 1 si ha

$$\text{rango di } (A - \mu I) = n - \ \dim N(A - \mu I) = n - \tau,$$

e quindi tutte le sottomatrici principali di ordine superiore a $n - \tau$ della matrice $A - \mu I$ sono singolari. Poiché il coefficiente del termine di grado i del polinomio caratteristico di una matrice è dato, a meno del segno, dalla somma dei determinanti delle sue sottomatrici principali di ordine $n - i$, ne segue che il polinomio caratteristico di $A - \mu I$ risulta della forma

$$p(\lambda) = \det[(A - \mu I) - \lambda I] = a_0\lambda^n + a_1\lambda^{n-1} + \ldots + a_k\lambda^{n-k},$$

dove $k \le n - \tau$. Perciò l'equazione $p(\lambda) = 0$ ha la radice $\lambda = 0$ di molteplicità $n - k \ge \tau$. Posto $x = \lambda + \mu$, si ha

$$\begin{aligned}\det[(A - \mu I) - \lambda I] &= \det(A - xI) \\ &= a_0(x - \mu)^n + a_1(x - \mu)^{n-1} + \ldots + a_k(x - \mu)^{n-k},\end{aligned}$$

per cui la molteplicità di μ come radice dell'equazione caratteristica è $\sigma \ge \tau$. ■

2.12 Esempio. La matrice $A \in \mathbf{R}^{n\times n}$, definita da

$$a_{ij} = \begin{cases} 1 & \text{per } j = i, \\ 1 & \text{per } j = i + 1, \\ 0 & \text{altrimenti,} \end{cases}$$

cioè

$$A = \begin{bmatrix} 1 & 1 & & & \\ & 1 & 1 & & \\ & & \ddots & \ddots & \\ & & & 1 & 1 \\ & & & & 1 \end{bmatrix},$$

ha 1 come autovalore di molteplicità algebrica n a cui corrispondono solo autovettori del tipo $\mathbf{x} = \alpha\mathbf{e}_1, \alpha \neq 0$. In questo caso risulta quindi $\tau(1) = 1$ e $\sigma(1) = n$. ■

Nel caso della matrice identica I_n, che ha 1 come autovalore di molteplicità algebrica n, risulta $\tau(1) = \sigma(1) = n$.

4. Trasformazioni per similitudine

Data una base $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ di $\mathbf{C}^n$ e una matrice $A \in \mathbf{C}^{n\times n}$, viene individuata univocamente l'applicazione lineare $\mathcal{L} : \mathbf{C}^n \to \mathbf{C}^n$, definita sugli elementi della base da

$$\mathcal{L}(\mathbf{u}_j) = \sum_{i=1}^{n} a_{ij}\, \mathbf{u}_i. \tag{10}$$

L'applicazione $\mathcal{L}$ risulta naturalmente estesa per linearità a tutti i vettori $\mathbf{x} \in \mathbf{C}^n$. Si considerino le matrici U e W le cui colonne sono rispettivamente i vettori $\mathbf{u}_j$ e $\mathcal{L}(\mathbf{u}_j)$. Allora la (10) può essere rappresentata nella forma:

$$W = UA. \tag{11}$$

Una stessa applicazione lineare $\mathcal{L}$ può essere rappresentata su due basi diverse $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ e $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ e quindi da due matrici A e B, in generale diverse. Se V e Z sono le matrici le cui colonne sono rispettivamente formate dai vettori $\mathbf{v}_j$ e $\mathcal{L}(\mathbf{v}_j)$, vale la relazione analoga alla (11):

$$Z = VB. \tag{12}$$

Si vuole ora determinare la relazione che lega le due matrici A e B. Si supponga che i vettori $\mathbf{u}_i$ e $\mathbf{v}_i$, $i = 1, \ldots, n$, siano legati dalla relazione

$$\mathbf{v}_j = \sum_{i=1}^{n} s_{ij}\, \mathbf{u}_i, \quad j = 1, 2, \ldots, n, \tag{13}$$

che in notazione matriciale è rappresentata da

$$V = US,$$

dove la matrice non singolare S è la matrice del *cambiamento di base.* Sostituendo quest'ultima relazione nella (12) si ha

$$Z = USB. \tag{14}$$

D'altra parte, per la linearità dell'applicazione $\mathcal{L}$, dalla (13) si ha:

$$\mathcal{L}(\mathbf{v}_j) = \mathcal{L}\,(\sum_{i=1}^{n} s_{ij}\, \mathbf{u}_i) = \sum_{i=1}^{n} s_{ij}\;\mathcal{L}(\mathbf{u}_i)$$

e in notazione matriciale

$$Z = WS. \tag{15}$$

Sostituendo la (11) nella (15) si ha:

$$Z = UAS$$

da cui, per confronto con la (14), poiché U è non singolare, si ha:

$$AS = SB$$

e quindi

$$A = SBS^{-1}.$$

Se le due basi $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ e $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ sono basi ortonormali, le matrici U e V sono unitarie, e allora anche la matrice S è unitaria; infatti:

$$I = V^H V = (US)^H (US) = S^H U^H US = S^H S.$$

In tal caso le due matrici A e B soddisfano la relazione

$$A = SBS^H.$$

2.13 Definizione. Due matrici $A, B \in \mathbf{C}^{n\times n}$ si dicono *simili* se esiste una matrice non singolare S per cui

$$A = SBS^{-1}.$$

La trasformazione che associa la matrice A alla matrice B viene detta *trasformazione per similitudine.* Se la matrice S è unitaria, la trasformazione viene detta *trasformazione per similitudine unitaria.* ■

Si noti che la trasformazione per similitudine è una relazione di equivalenza, in quanto gode delle proprietà riflessiva, simmetrica e transitiva.

Data una matrice A, si consideri l'applicazione lineare $\mathcal{L}_A$ individuata dalla A e dalla base canonica $\mathbf{e}_i,\ i = 1, \ldots, n$ di $\mathbf{C}^n$; allora per ogni vettore $\mathbf{x} \in \mathbf{C}^n$ risulta

$$\mathcal{L}_A(\mathbf{x}) = A\mathbf{x}.$$

Quindi dalla relazione (1), riformulata in termini di applicazioni lineari, risulta che gli autovettori $\mathbf{x}$ di A sono i vettori che vengono trasformati dall'applicazione $\mathcal{L}_A$ in vettori proporzionali a se stessi. Cioè ogni autovettore individua una retta che è invariante per la trasformazione lineare $\mathcal{L}_A$. Le proprietà degli autovalori e autovettori sono dunque proprietà intrinseche dell'applicazione lineare e non legate solamente alla matrice che la rappresenta in una particolare base. Vale infatti il seguente teorema.

2.14 Teorema. *Due matrici simili hanno gli stessi autovalori con le stesse molteplicità algebriche e geometriche.*

Dim. Siano A e B matrici simili, cioè tali che $A = SBS^{-1}$. Si ha:

$$\begin{aligned}\det(A - \lambda I) &= \det(SBS^{-1} - \lambda SS^{-1}) = \det\left[S(B - \lambda I)S^{-1}\right] \\ &= \det S \det(B - \lambda I) \det(S^{-1}) = \det(B - \lambda I)\end{aligned}$$

per cui le due matrici hanno lo stesso polinomio caratteristico e quindi hanno gli stessi autovalori con le stesse molteplicità algebriche. Se $\mathbf{x}$ è autovettore di A corrispondente all'autovalore λ, risulta:

$$SBS^{-1}\mathbf{x} = \lambda \mathbf{x}$$

e quindi

$$BS^{-1}\mathbf{x} = \lambda S^{-1}\mathbf{x}.$$

Perciò il vettore $\mathbf{y} = S^{-1}\mathbf{x}$ è autovettore di B corrispondente a λ. Inoltre, essendo S^{-1} non singolare, se $\mathbf{x}_i$, $i = 1, \ldots, \tau(\lambda)$, sono autovettori linearmente indipendenti di A, anche $\mathbf{y}_i = S^{-1}\mathbf{x}_i$, $i = 1, \ldots, \tau(\lambda)$ sono linearmente indipendenti. Quindi A e B hanno gli stessi autovalori con le stesse molteplicità geometriche. ■

Da questo teorema risulta che se due matrici sono simili, hanno uguali la traccia e il determinante.

2.15 Definizione. Una matrice A simile ad una matrice diagonale D si dice *diagonalizzabile*. ■

2.16 Teorema. *Una matrice A di ordine n è diagonalizzabile se e solo se ha n autovettori linearmente indipendenti. Inoltre le colonne della matrice S, per cui $S^{-1}AS$ è diagonale, sono gli autovettori di A.*

Dim. Si suppone dapprima che A abbia n autovettori linearmente indipendenti $\mathbf{x}_1, \ldots, \mathbf{x}_n$, corrispondenti agli autovalori $\lambda_1, \ldots, \lambda_n$. Siano D la matrice diagonale avente λ_i come i-esimo elemento principale, e S la matrice la cui i-esima colonna è uguale a $\mathbf{x}_i$. Dalla relazione

$$A\mathbf{x}_i = \lambda_i \mathbf{x}_i, \quad i = 1, 2, \ldots, n,$$

si ha anche che

$$AS = SD. \tag{16}$$

Essendo S non singolare, perché formata da colonne linearmente indipendenti, esiste S^{-1}; quindi dalla (16) si ha

$$A = SDS^{-1}.$$

Viceversa, sia $A = SDS^{-1}$, D matrice diagonale con gli autovalori di A come elementi principali. Allora risulta $AS = SD$. Indicando con $\mathbf{s}_1, \ldots, \mathbf{s}_n$ le colonne di S, si ha:

$$A\,[\mathbf{s}_1 \,|\, \mathbf{s}_2 \,|\, \ldots \,|\, \mathbf{s}_n] = [\lambda_1\mathbf{s}_1 \,|\, \lambda_2\mathbf{s}_2 \,|\, \ldots \,|\, \lambda_n\mathbf{s}_n]$$

e quindi

$$A\mathbf{s}_i = \lambda_i\mathbf{s}_i, \quad i = 1, 2, \ldots, n.$$

Perciò le colonne di S sono n autovettori di A, che risultano linearmente indipendenti, perché S è non singolare. ■

2.17 Esempio. Le matrici A e $B = 3A^2 - A + 2I$ dell'esempio 2.3, che hanno tre autovalori distinti, sono diagonalizzabili dalla stessa trasformazione per similitudine, in quanto hanno gli stessi tre autovettori linearmente indipendenti. Posto

$$S = \begin{bmatrix} 1 & 1 & 1 \\ -\sqrt{2} & 0 & \sqrt{2} \\ 1 & -1 & 1 \end{bmatrix},$$

si ha

$$S^{-1} = \begin{bmatrix} 1/4 & -\sqrt{2}/4 & 1/4 \\ 1/2 & 0 & -1/2 \\ 1/4 & \sqrt{2}/4 & 1/4 \end{bmatrix}$$

e

$$A = S \begin{bmatrix} 1-\sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1+\sqrt{2} \end{bmatrix} S^{-1},$$

$$B = S \begin{bmatrix} 10-5\sqrt{2} & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 10+5\sqrt{2} \end{bmatrix} S^{-1},$$

Anche la matrice A dell'esempio 2.4, pur non avendo 3 autovalori distinti ha 3 autovettori linearmente indipendenti, ed è quindi diagonalizzabile. Posto

$$S = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ -1 & -1 & 1 \end{bmatrix},$$

si ha

$$S^{-1} = \frac{1}{3} \begin{bmatrix} -1 & 2 & -1 \\ 2 & -1 & -1 \\ 1 & 1 & 1 \end{bmatrix}$$

e

$$A = S \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix} S^{-1}. \qquad \blacksquare$$

5. Forme canoniche

Dai teoremi 2.7 e 2.16 segue che se una matrice ha tutti gli autovalori distinti, allora è diagonalizzabile, in quanto ha n autovettori linearmente indipendenti. Se una matrice non ha tutti gli autovalori distinti, può non essere diagonalizzabile e questo accade se per almeno un autovalore di A la molteplicità geometrica è minore della corrispondente molteplicità algebrica. A tale proposito vale il seguente teorema (per la dimostrazione si veda [3]).

2.18 Teorema (Forma canonica o normale di Jordan). *Sia $A \in \mathbf{C}^{n\times n}$ e siano λ_i, $i = 1, \ldots, p$, i suoi autovalori distinti, con molteplicità algebrica $\sigma(\lambda_i)$ e molteplicità geometrica $\tau(\lambda_i)$. Allora A è simile ad una matrice diagonale a blocchi*

$$J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_p \end{bmatrix},$$

in cui il blocco J_i, relativo all'autovalore λ_i, ha ordine $\sigma(\lambda_i)$ ed è a sua volta diagonale a blocchi

$$J_i = \begin{bmatrix} C_i^{(1)} & & & \\ & C_i^{(2)} & & \\ & & \ddots & \\ & & & C_i^{(\tau(\lambda_i))} \end{bmatrix}, \quad i = 1, 2, \ldots, p,$$

e ognuno dei $\tau(\lambda_i)$ blocchi è della forma

$$C_i^{(j)} = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i & 1 \\ & & & \lambda_i \end{bmatrix} \in \mathbf{C}^{\nu_i^{(j)} \times \nu_i^{(j)}}, \quad j = 1, 2, \ldots, \tau(\lambda_i),$$

dove gli interi $\nu_i^{(j)}$ sono tali che

$$\sum_{j=1}^{\tau(\lambda_i)} \nu_i^{(j)} = \sigma(\lambda_i).$$

La matrice J è detta forma canonica (o normale) di Jordan della matrice A, ed è unica, a meno dell'ordinamento dei blocchi che la compongono. ■

Se gli autovalori λ_i di A sono tutti distinti, i blocchi J_i hanno tutti ordine 1, e quindi la matrice è diagonalizzabile. Se invece gli autovalori non sono tutti distinti, ma A ha n autovettori linearmente indipendenti, allora i blocchi J_i sono diagonali, e anche in questo caso la matrice è diagonalizzabile.

2.19 Esempio. Le matrici

$$A_1 = \begin{bmatrix} -4 & 7 & 4 & -8 & 6 & -3 \\ -5 & 7 & 5 & -8 & 6 & -3 \\ -4 & 4 & 6 & -7 & 6 & -3 \\ -3 & 3 & 3 & -4 & 6 & -3 \\ -2 & 2 & 2 & -4 & 6 & -2 \\ -1 & 1 & 1 & -2 & 2 & 1 \end{bmatrix}$$

e

$$A_2 = \begin{bmatrix} -4 & 12 & -10 & 8 & -6 & 4 \\ -5 & 12 & -9 & 8 & -6 & 4 \\ -4 & 8 & -6 & 8 & -6 & 4 \\ -3 & 6 & -16 & 8 & -5 & 4 \\ -2 & 4 & -4 & 4 & -2 & 4 \\ -1 & 2 & -2 & 2 & -2 & 4 \end{bmatrix}$$

hanno entrambe l'autovalore $\lambda = 2$ di molteplicità algebrica 6 e geometrica 3, e risulta

$$A_1 = SJ'S^{-1} = S \begin{bmatrix} 2 & 1 & 0 & & & \\ 0 & 2 & 1 & & & \\ 0 & 0 & 2 & & & \\ & & & 2 & 1 & \\ & & & 0 & 2 & \\ & & & & & 2 \end{bmatrix} S^{-1}$$

e

$$A_2 = SJ''S^{-1} = S \begin{bmatrix} 2 & 1 & & & & \\ 0 & 2 & & & & \\ & & 2 & 1 & & \\ & & 0 & 2 & & \\ & & & & 2 & 1 \\ & & & & 0 & 2 \end{bmatrix} S^{-1}.$$

In entrambi i casi si ha

$$S = \begin{bmatrix} 6 & 5 & 4 & 3 & 2 & 1 \\ 5 & 5 & 4 & 3 & 2 & 1 \\ 4 & 4 & 4 & 3 & 2 & 1 \\ 3 & 3 & 3 & 3 & 2 & 1 \\ 2 & 2 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \qquad S^{-1} = \begin{bmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & -1 & 2 & -1 & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix}.$$

■

Se la matrice A ha elementi reali, allora esiste una *forma normale reale di Jordan* di A, analoga a quella definita nel teorema 2.18, in cui i blocchi $C_i^{(j)}$ relativi ad autovalori reali hanno la stessa forma che nel teorema 2.18, mentre i blocchi $C_i^{(j)}$ relativi ad autovalori complessi risultano così modificati: in corrispondenza ad ogni coppia $\lambda_i = a_i + \mathbf{i}b_i$ e $\overline{\lambda}_i = a_i - \mathbf{i}b_i$ di autovalori complessi e coniugati di A le sottomatrici $C_i^{(j)}$ sono bidiagonali a blocchi della forma

$$C_i^{(j)} = \begin{bmatrix} E_i & I_2 & & & \\ & E_i & I_2 & & \\ & & \ddots & \ddots & \\ & & & \ddots & I_2 \\ & & & & E_i \end{bmatrix},$$

dove

$$E_i = \begin{bmatrix} a_i & -b_i \\ b_i & a_i \end{bmatrix}.$$

2.20 Esempio. La matrice

$$A = \begin{bmatrix} 8 & -16 & 13 & -3 \\ 6 & -12 & 10 & -2 \\ 4 & -9 & 9 & -3 \\ 2 & -5 & 5 & -1 \end{bmatrix}$$

ha la forma normale di Jordan

$$A = SJS^{-1} = S \begin{bmatrix} 1+\mathbf{i} & 1 & 0 & 0 \\ 0 & 1+\mathbf{i} & 0 & 0 \\ 0 & 0 & 1-\mathbf{i} & 1 \\ 0 & 0 & 0 & 1-\mathbf{i} \end{bmatrix} S^{-1},$$

dove

$$S = \begin{bmatrix} 4-3\mathbf{i} & 2-\mathbf{i} & 4+3\mathbf{i} & 2+\mathbf{i} \\ 3-3\mathbf{i} & 2-\mathbf{i} & 3+3\mathbf{i} & 2+\mathbf{i} \\ 2-2\mathbf{i} & 2-\mathbf{i} & 2+2\mathbf{i} & 2+\mathbf{i} \\ 1-\mathbf{i} & 1-\mathbf{i} & 1+\mathbf{i} & 1+\mathbf{i} \end{bmatrix}$$

e

$$S^{-1} = \frac{1}{2}\begin{bmatrix} 1-\mathbf{i} & -1+2\mathbf{i} & -\mathbf{i} & 0 \\ 0 & -1 & 2-\mathbf{i} & -1+2\mathbf{i} \\ 1+\mathbf{i} & -1-2\mathbf{i} & \mathbf{i} & 0 \\ 0 & -1 & 2+\mathbf{i} & -1-2\mathbf{i} \end{bmatrix}.$$

La matrice A, avendo elementi reali, può essere rappresentata anche nella forma normale reale di Jordan

$$A = ZJ_RZ^{-1} = Z\begin{bmatrix} 1 & -1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 \end{bmatrix} Z^{-1},$$

dove

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

■

Dalla forma normale di Jordan si può ricavare il polinomio minimo di A. Infatti posto $A = SJS^{-1}$, dove J è la forma normale di Jordan di A, si ha per ogni intero k

$$A^k = \underbrace{SJS^{-1}SJS^{-1}\ldots SJS^{-1}}_{k \text{ volte}} = SJ^kS^{-1}$$

e quindi per ogni polinomio $P(A)$ si ha

$$P(A) = SP(J)S^{-1}.$$

In particolare per il polinomio minimo $\psi(\lambda)$ di A, risulta

$$\psi(A) = S\psi(J)S^{-1},$$

e quindi il polinomio minimo di A e quello di J coincidono. Per la struttura diagonale a blocchi di J si ha

$$\psi(J) = \begin{bmatrix} \psi(J_1) & & & \\ & \psi(J_2) & & \\ & & \ddots & \\ & & & \psi(J_p) \end{bmatrix}.$$

Perciò

$$\psi(\lambda) = (\lambda-\lambda_1)^{n_1}(\lambda-\lambda_2)^{n_2}\ldots(\lambda-\lambda_p)^{n_p},$$

deve avere gli esponenti $n_i,\ i = 1, 2, \ldots, p$, tali che $(\lambda-\lambda_i)^{n_i}$ sia il polinomio minimo di J_i, cioè gli n_i devono essere gli interi più piccoli per cui il polinomio $(\lambda-\lambda_i)^{n_i}$ sia annullato contemporaneamente da tutte le sottomatrici $C_i^{(j)},\ j = 1, 2, \ldots, \tau(\lambda_i)$. Ciò è vero se e solo se n_i è la dimensione massima delle sottomatrici $C_i^{(j)}$, per $j = 1, 2, \ldots, \tau(\lambda_i)$.

2.21 Esempio. Il polinomio minimo della matrice A_1 dell'esempio 2.19 è dato da

$$\psi(\lambda) = (\lambda - 2)^3,$$

e quello della matrice A_2 è dato da

$$\psi(\lambda) = (\lambda - 2)^2.$$

■

Fra le trasformazioni per similitudine che associano alla matrice B la matrice $A = SBS^{-1}$, hanno particolare importanza quelle per cui S è unitaria, cioè $S^H S = SS^H = I$. Il teorema che segue mostra come sia possibile, mediante una trasformazione per similitudine unitaria, ricondurre una qualsiasi matrice a una forma triangolare superiore.

2.22 Teorema (forma canonica o normale di Schur). *Sia $A \in \mathbf{C}^{n\times n}$ e siano $\lambda_1, \ldots, \lambda_n$ i suoi autovalori. Allora esiste una matrice unitaria U e una matrice triangolare superiore T i cui elementi principali sono i λ_i, tali che*

$$A = UTU^H.$$

Dim. Si procede per induzione. Per $n = 1$ la tesi vale con $U = [1]$. Per $n > 1$, sia $\mathbf{x}_1$ l'autovettore normalizzato corrispondente all'autovalore λ_1 e sia S lo spazio generato da $\mathbf{x}_1$. Indicata con $\mathbf{y}_2, \ldots, \mathbf{y}_n$ una base ortonormale dello spazio $S^\perp$, la matrice

$$Q = [\, \mathbf{x}_1 \,|\, \mathbf{y}_2 \,|\, \ldots \,|\, \mathbf{y}_n \,]$$

è unitaria e $Q^H\mathbf{x}_1 = \mathbf{e}_1$. Si considera la matrice

$$B = Q^H AQ$$

la cui prima colonna è

$$B\mathbf{e}_1 = Q^H AQ\mathbf{e}_1 = Q^H A\mathbf{x}_1 = Q^H \lambda_1\mathbf{x}_1 = \lambda_1 Q^H\mathbf{x}_1 = \lambda_1\mathbf{e}_1$$

e quindi B può essere partizionata nel modo seguente:

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

dove $\mathbf{c} \in \mathbf{C}^{n-1}$ e $A_1 \in \mathbf{C}^{(n-1)\times(n-1)}$. Per l'ipotesi induttiva esiste una matrice unitaria $U_1 \in \mathbf{C}^{(n-1)\times(n-1)}$ tale che

$$A_1 = U_1 A_2 U_1^H,$$

dove $A_2 \in \mathbf{C}^{(n-1)\times(n-1)}$ è triangolare superiore. Allora risulta

$$A = QBQ^H = Q \begin{bmatrix} \lambda_1 & \mathbf{c}^H \\ & \\ \mathbf{0} & A_1 \end{bmatrix} Q^H = Q \begin{bmatrix} \lambda_1 & \mathbf{c}^H \\ & \\ \mathbf{0} & U_1 A_2 U_1^H \end{bmatrix} Q^H.$$

Indicando con $U_2 \in \mathbf{C}^{n\times n}$ la matrice unitaria

$$U_2 = \begin{bmatrix} 1 & \mathbf{0}^H \\ & \\ \mathbf{0} & U_1 \end{bmatrix},$$

si ha:

$$A = QU_2 \begin{bmatrix} \lambda_1 & \mathbf{c}^H U_1 \\ & \\ \mathbf{0} & A_2 \end{bmatrix} U_2^H Q^H.$$

Poiché la matrice $U = QU_2$ è ancora unitaria in quanto prodotto di matrici unitarie, risulta

$$A = U \begin{bmatrix} \lambda_1 & \mathbf{c}^H U_1 \\ & \\ \mathbf{0} & A_2 \end{bmatrix} U^H.$$

da cui la tesi, essendo A_2 matrice triangolare superiore. ∎

2.23 Esempio. La matrice

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

ha l'autovalore $\lambda_1 = \mathbf{i}$ con il corrispondente autovettore normalizzato

$$\mathbf{x}_1 = \frac{1}{2}\,[\,1\,,\,1\,,\,\mathbf{i}\,,\,-\mathbf{i}\,]^T.$$

Si considerano altri tre vettori $\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4 \in \mathbf{C}^4$ linearmente indipendenti fra di loro e da $\mathbf{x}_1$:

$$\mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

Si costruiscono poi, a partire dai vettori $\mathbf{x}_i$, $i = 1, ..., 4$, con il metodo di Gram-Schmidt, tre vettori $\mathbf{y}_2, \mathbf{y}_3, \mathbf{y}_4$

$$\mathbf{y}_2 = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ \text{-}\mathbf{i} \\ \mathbf{i} \end{bmatrix}, \quad \mathbf{y}_3 = \frac{1}{2}\begin{bmatrix} \text{-}1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{y}_4 = \frac{1}{2}\begin{bmatrix} 1 \\ \text{-}1 \\ 1 \\ 1 \end{bmatrix},$$

tali che la matrice

$$Q = [\mathbf{x}_1 \mid \mathbf{y}_2 \mid \mathbf{y}_3 \mid \mathbf{y}_4] = \frac{1}{2}\begin{bmatrix} 1 & 1 & \text{-}1 & 1 \\ 1 & 1 & 1 & \text{-}1 \\ \mathbf{i} & \text{-}\mathbf{i} & 1 & 1 \\ \text{-}\mathbf{i} & \mathbf{i} & 1 & 1 \end{bmatrix}$$

è unitaria. Si ha poi

$$B = Q^H A Q = \begin{bmatrix} \mathbf{i} & 0 & \text{-}2 & 3+\mathbf{i} \\ 0 & \text{-}\mathbf{i} & \text{-}2 & 3-\mathbf{i} \\ 0 & 0 & 0 & 1 \\ 0 & 0 & \text{-}1 & 0 \end{bmatrix} = \begin{bmatrix} T_1 & C \\ O & A_1 \end{bmatrix},$$

in cui $T_1 \in \mathbf{C}^{2\times 2}$ è triangolare superiore (più precisamente in questo caso T_1 risulta diagonale). Si deve quindi riapplicare il procedimento alla matrice

$$A_1 = \begin{bmatrix} 0 & 1 \\ \text{-}1 & 0 \end{bmatrix}$$

che ha ancora l'autovalore $\mathbf{i}$, con l'autovettore normalizzato

$$\mathbf{z}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} \text{-}\mathbf{i} \\ 1 \end{bmatrix}.$$

Con il metodo di Gram-Schmidt si determina il vettore

$$\mathbf{z}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{i} \\ 1 \end{bmatrix}.$$

tale che la matrice

$$Q_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} \text{-}\mathbf{i} & \mathbf{i} \\ 1 & 1 \end{bmatrix}$$

è unitaria e si ha

$$B_1 = Q_1^H A_1 Q_1 = \begin{bmatrix} \mathbf{i} & 0 \\ 0 & \text{-}\mathbf{i} \end{bmatrix}.$$

La forma normale di Schur della matrice A risulta quindi

$$A = QBQ^H = Q \begin{bmatrix} I_2 & O \\ O & Q_1 \end{bmatrix} \begin{bmatrix} T_1 & CQ_1 \\ O & Q_1^H A_1 Q_1 \end{bmatrix} \begin{bmatrix} I_2 & O \\ O & Q_1 \end{bmatrix}^H Q^H$$

$$= U \begin{bmatrix} \mathbf{i} & 0 & (3+3\mathbf{i})/\sqrt{2} & (3-\mathbf{i})/\sqrt{2} \\ 0 & \text{-}\mathbf{i} & (3+\mathbf{i})/\sqrt{2} & (3-3\mathbf{i})/\sqrt{2} \\ 0 & 0 & \mathbf{i} & 0 \\ 0 & 0 & 0 & \text{-}\mathbf{i} \end{bmatrix} U^H,$$

dove U è la matrice unitaria

$$U = Q \begin{bmatrix} I_2 & O \\ O & Q_1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 & (1+\mathbf{i})/\sqrt{2} & (1-\mathbf{i})/\sqrt{2} \\ 1 & 1 & (\text{-}1-\mathbf{i})/\sqrt{2} & (\text{-}1+\mathbf{i})/\sqrt{2} \\ \mathbf{i} & \text{-}\mathbf{i} & (1-\mathbf{i})/\sqrt{2} & (1+\mathbf{i})/\sqrt{2} \\ \text{-}\mathbf{i} & \mathbf{i} & (1-\mathbf{i})/\sqrt{2} & (1+\mathbf{i})/\sqrt{2} \end{bmatrix}.$$

■

Come nel caso della forma canonica di Jordan, anche nel caso della forma normale di Schur, se la matrice A ha elementi reali esiste la *forma normale reale di Schur.*

2.24 Teorema. *Se $A \in \mathbf{R}^{n\times n}$, esiste una matrice ortogonale $U \in \mathbf{R}^{n\times n}$ e una matrice $T \in \mathbf{R}^{n\times n}$ triangolare superiore a blocchi, della forma*

$$T = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ & R_{22} & \cdots & R_{2m} \\ & & \ddots & \vdots \\ & & & R_{mm} \end{bmatrix},$$

dove i blocchi R_{jj} per $j = 1, 2, \ldots, m$ hanno ordine 1 o 2. Se λ_j è autovalore reale di A, allora R_{jj} ha ordine 1 e coincide con $[\lambda_j]$, se λ_j è complesso, allora il blocco R_{jj} ha ordine 2 ed ha come autovalori λ_j e $\overline{\lambda}_j$. La somma delle dimensioni dei blocchi R_{jj}, $j = 1, 2, \ldots, m$ è pari ad n.

Dim. La dimostrazione viene fatta per induzione, in modo analogo a quella del teorema 2.22. Se l'autovalore λ_1 è reale, si ripete il ragionamento fatto per il caso complesso. Se invece $\lambda_1 = \mu_1 + \mathbf{i}\nu_1$, $\mu_1, \nu_1 \in \mathbf{R}$, $\nu_1 \neq 0$, si considera il corrispondente autovettore $\mathbf{x}_1 + \mathbf{i}\mathbf{y}_1$, $\mathbf{x}_1, \mathbf{y}_1 \in \mathbf{R}^n$, in cui si può supporre che il vettore $\mathbf{x}_1$ sia normalizzato. Poiché

$$A(\mathbf{x}_1 + \mathbf{i}\mathbf{y}_1) = A\mathbf{x}_1 + \mathbf{i}A\mathbf{y}_1 = (\mu_1\mathbf{x}_1 - \nu_1\mathbf{y}_1) + \mathbf{i}(\mu_1\mathbf{y}_1 + \nu_1\mathbf{x}_1),$$

risulta

$$A[\,\mathbf{x}_1\,|\,\mathbf{y}_1\,] = [\,\mathbf{x}_1\,|\,\mathbf{y}_1\,]\begin{bmatrix}\mu_1 & \nu_1\\ \text{-}\nu_1 & \mu_1\end{bmatrix}. \tag{17}$$

I due vettori $\mathbf{x}_1$ e $\mathbf{y}_1$ sono linearmente indipendenti: infatti, se ciò non fosse, esisterebbe una costante $\alpha \neq 0$ per cui $\mathbf{y}_1 = \alpha\mathbf{x}_1$, e quindi

$$\mathbf{x}_1 + \mathbf{i}\mathbf{y}_1 = \mathbf{x}_1 + \mathbf{i}\alpha\mathbf{x}_1 = (1+\mathbf{i}\alpha)\,\mathbf{x}_1$$

e il vettore reale $\mathbf{x}_1$ risulterebbe essere un autovettore reale di A corrispondente all'autovalore complesso λ_1, ciò che è assurdo perché A ha elementi reali.

Si costruisce un vettore normalizzato $\mathbf{z}_1$, ortogonale al vettore $\mathbf{x}_1$, ponendo

$$\mathbf{z}_1 = \beta\mathbf{x}_1 + \gamma\mathbf{y}_1, \quad \gamma = \frac{1}{\sqrt{\mathbf{y}_1^T\mathbf{y}_1 - (\mathbf{x}_1^T\mathbf{y}_1)^2}}, \quad \beta = \text{-}\gamma(\mathbf{x}_1^T\mathbf{y}_1).$$

Si ha perciò

$$[\,\mathbf{x}_1\,|\,\mathbf{z}_1\,] = [\,\mathbf{x}_1\,|\,\mathbf{y}_1\,]\,W, \quad \text{dove } W = \begin{bmatrix}1 & \beta\\ 0 & \gamma\end{bmatrix}. \tag{18}$$

Si costruisce poi una matrice $Q \in \mathbf{R}^{n\times n}$ ortogonale, le cui prime due colonne siano $\mathbf{x}_1$ e $\mathbf{z}_1$:

$$Q = [\,\mathbf{x}_1\,|\,\mathbf{z}_1\,|\,\mathbf{y}_3\,|\,\ldots\,|\,\mathbf{y}_n\,].$$

Proseguendo la dimostrazione in modo analogo a quanto fatto nel teorema 2.22, per le prime due colonne della matrice $B = Q^TAQ$ si ha dalle (17) e (18):

$$\begin{aligned} B\,[\,\mathbf{e}_1\,|\,\mathbf{e}_2\,] &= Q^TA\,[\,\mathbf{x}_1\,|\,\mathbf{z}_1\,] = Q^TA\,[\,\mathbf{x}_1\,|\,\mathbf{y}_1\,]W\\ &= Q^T[\,\mathbf{x}_1\,|\,\mathbf{y}_1\,]\begin{bmatrix}\mu_1 & \nu_1\\ \text{-}\nu_1 & \mu_1\end{bmatrix}W = Q^T[\,\mathbf{x}_1\,|\,\mathbf{z}_1\,]\,W^{-1}\begin{bmatrix}\mu_1 & \nu_1\\ \text{-}\nu_1 & \mu_1\end{bmatrix}W. \end{aligned}$$

Poiché Q è ortogonale, le prime due colonne di B possono essere scritte nel modo seguente

$$B\,[\,\mathbf{e}_1\,|\,\mathbf{e}_2\,] = \begin{bmatrix} I_2\\ \\ O\end{bmatrix} W^{-1}\begin{bmatrix}\mu_1 & \nu_1\\ \text{-}\nu_1 & \mu_1\end{bmatrix}W = \begin{bmatrix} R_{11}\\ \\ O\end{bmatrix}\begin{matrix}\}\quad 2 \text{ righe}\\ \\ \}\ n-2 \text{ righe},\end{matrix}$$

in cui il blocco

$$R_{11} = W^{-1}\begin{bmatrix}\mu_1 & \nu_1\\ \text{-}\nu_1 & \mu_1\end{bmatrix}W$$

è reale e ha come autovalori λ_1 e $\overline{\lambda}_1$. La dimostrazione prosegue sfruttando l'ipotesi induttiva come nel teorema 2.22. ∎

2.25 Esempio. Si determina la forma normale reale di Schur della matrice

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

dell'esempio 2.23. La matrice A ha l'autovalore $\lambda_1 = \mathbf{i}$ con il corrispondente autovettore

$$\mathbf{x}_1 + \mathbf{i}\mathbf{y}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ \mathbf{i} \\ -\mathbf{i} \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \mathbf{i}\frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}.$$

In questo caso i due vettori $\mathbf{x}_1$ e $\mathbf{y}_1$ sono ortonormali. Si considerano poi gli altri due vettori ortonormali

$$\mathbf{y}_3 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{y}_4 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

La matrice

$$U = [\,\mathbf{x}_1 \,|\, \mathbf{y}_1 \,|\, \mathbf{y}_3 \,|\, \mathbf{y}_4\,]$$

risulta così ortogonale, ed è tale che

$$A = U\begin{bmatrix} 0 & 1 & 5 & 1 \\ -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}U^T.$$

Si è cosí determinata la forma normale reale di Schur di A. ∎

Un caso particolarmente importante è quello in cui le matrici sono hermitiane.

2.26 Teorema. *Sia A una matrice hermitiana di ordine n, cioè $A = A^H$ e siano $\lambda_1, \ldots, \lambda_n$ i suoi autovalori. Allora esiste una matrice unitaria U tale che*

$$A = U \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} U^H,$$

cioè la matrice A è diagonalizzabile. Inoltre i λ_i, $i = 1, \ldots, n$ sono reali e le colonne di U costituiscono un insieme di autovettori ortonormali.

Dim. Per il teorema 2.22 si ha $T = U^H AU$, dove T è una matrice triangolare superiore e U è unitaria. Poiché $A = A^H$, si ha

$$T^H = (U^H AU)^H = U^H A^H U = U^H AU = T,$$

cioè la matrice triangolare T risulta essere una matrice diagonale con gli elementi pricipali reali e per il teorema 2.16 le colonne di U, che sono ortonormali perché U è unitaria, risultano essere gli autovettori di A. ∎

Se la matrice A è reale e simmetrica, la matrice U risulta reale, e quindi è ortogonale.

2.27 Esempio. La matrice

$$A = \begin{bmatrix} 1 & \mathbf{i} & 0 \\ -\mathbf{i} & 2 & -\mathbf{i} \\ 0 & \mathbf{i} & 1 \end{bmatrix}$$

ha autovalori $\lambda_1 = 0$, $\lambda_2 = 1$ e $\lambda_3 = 3$, con i corrispondenti autovettori

$$\mathbf{x}_1 = \alpha_1 \begin{bmatrix} 1 \\ \mathbf{i} \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \alpha_2 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad \mathbf{x}_3 = \alpha_3 \begin{bmatrix} 1 \\ -2\mathbf{i} \\ 1 \end{bmatrix}, \quad \alpha_1,\ \alpha_2,\ \alpha_3 \neq 0,$$

che sono fra loro ortogonali e che risultano normalizzati ponendo $\alpha_1 = 1/\sqrt{3}$, $\alpha_2 = 1/\sqrt{2}$ e $\alpha_3 = 1/\sqrt{6}$. Per cui, in questo caso, la matrice

$$U = [\,\mathbf{x}_1 \,|\, \mathbf{x}_2 \,|\, \mathbf{x}_3\,],$$

data da

$$U = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & \sqrt{3} & 1 \\ \mathbf{i}\sqrt{2} & 0 & -2\mathbf{i} \\ \sqrt{2} & -\sqrt{3} & 1 \end{bmatrix},$$

è unitaria e si ha

$$A = U \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix} U^H.$$ ∎

Una classe più ampia di matrici che comprende, come casi particolari, le matrici hermitiane e le matrici unitarie, è quella delle matrici normali, cioè tali che $A^H A = AA^H$. Questa classe di matrici è particolarmente importante perché è quella che comprende tutte e sole le matrici diagonalizzabili con trasformazioni per similitudine unitarie. Vale infatti il

2.28 Teorema. *Una matrice $A \in \mathbf{C}^{n\times n}$ è normale, cioè $A^H A = AA^H$, se e solo se esiste una matrice unitaria U tale che*

$$A = U \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} U^H,$$

in cui $\lambda_1, \ldots, \lambda_n$ sono gli autovalori di A. La matrice U ha per colonne gli autovettori della matrice A, che quindi sono a due a due ortonormali.

Dim. Si supponga dapprima che A sia normale. Per il teorema 2.22 esiste una matrice U unitaria tale che

$$T = U^H AU,$$

con T matrice triangolare superiore e si ha:

$$\begin{aligned} T^H T &= U^H A^H UU^H AU = U^H A^H AU,\\ TT^H &= U^H AUU^H A^H U = U^H AA^H U. \end{aligned}$$

Poiché A è normale, ne segue che

$$T^H T = TT^H, \tag{19}$$

e quindi anche T è normale. Si dimostra per induzione su n che T è diagonale. Se $n = 1$ questo è ovvio. Se $n > 1$, poiché T è triangolare superiore, per l'elemento p_{11} della matrice $P = T^H T = TT^H$, si ha

$$p_{11} = \bar{t}_{11} t_{11} = |\lambda_1|^2 \quad \text{e} \quad p_{11} = \sum_{j=1}^{n} t_{1j}\bar{t}_{1j} = |\lambda_1|^2 + \sum_{j=2}^{n} |t_{1j}|^2,$$

da cui

$$t_{1j} = 0, \quad \text{per} \quad j = 2, \ldots, n,$$

cioè la prima riga di T ha tutti gli elementi nulli eccetto quello principale. Indicata con T_{n-1} la sottomatrice ottenuta da T cancellando la prima riga e la prima colonna, dalla (19) segue che

$$T_{n-1}^H T_{n-1} = T_{n-1} T_{n-1}^H.$$

Per l'ipotesi induttiva T_{n-1} è diagonale, quindi anche T risulta diagonale.

Viceversa, sia A diagonalizzabile con una trasformazione per similitudine unitaria:

$$A = UDU^H,$$

con D diagonale. Si ha:

$$A^H A = UD^H U^H UDU^H = UD^H DU^H,$$

$$AA^H = UDU^H UD^H U^H = UDD^H U^H.$$

Poiché D è diagonale, $D^H D$ e DD^H sono diagonali, con elementi principali uguali a $\overline{\lambda}_i \lambda_i$; risulta allora $D^H D = DD^H$ e quindi

$$A^H A = AA^H. \qquad \blacksquare$$

Nel caso in cui la matrice A è normale e ha elementi reali, la sua forma normale reale di Schur risulta

$$A = UTU^T,$$

dove T e U sono matrici reali, U è ortogonale e T è diagonale a blocchi di ordine 1 o 2.

2.29 Esempio. La matrice $A \in \mathbf{R}^{4\times 4}$

$$A = \begin{bmatrix} 4 & -5 & 0 & 3 \\ 0 & 4 & -3 & -5 \\ 5 & -3 & 4 & 0 \\ 3 & 0 & 5 & 4 \end{bmatrix}$$

è normale, perché risulta $A^T A = AA^T$, e quindi è diagonalizzabile con trasformazioni per similitudine unitarie. Posto

$$U = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -\mathbf{i} & \mathbf{i} & 1 \\ 1 & -\mathbf{i} & \mathbf{i} & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix},$$

risulta

$$A = U \begin{bmatrix} 12 & & & \\ & 1+5\mathbf{i} & & \\ & & 1-5\mathbf{i} & \\ & & & 2 \end{bmatrix} U^H.$$

A può essere rappresentata anche nella forma normale reale di Schur. Poiché

$$\begin{bmatrix} 1+5\mathbf{i} & 0 \\ 0 & 1-5\mathbf{i} \end{bmatrix} = V \begin{bmatrix} 1 & -5 \\ 5 & 1 \end{bmatrix} V^H,$$

dove V è la matrice unitaria

$$V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -\mathbf{i} \\ 1 & \mathbf{i} \end{bmatrix},$$

si ha

$$A = Z \begin{bmatrix} 12 & 0 & 0 & 0 \\ 0 & 1 & -5 & 0 \\ 0 & 5 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} Z^T,$$

dove la matrice ortogonale Z è data da

$$Z = U \begin{bmatrix} 1 & \mathbf{0}^H & 0 \\ \mathbf{0} & V & \mathbf{0} \\ 0 & \mathbf{0}^H & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & \sqrt{2} & 0 & 1 \\ -1 & 0 & -\sqrt{2} & 1 \\ 1 & 0 & -\sqrt{2} & -1 \\ 1 & -\sqrt{2} & 0 & 1 \end{bmatrix}.$$

■

Per il teorema di Schur è possibile caratterizzare completamente gli autovalori dei polinomi di matrici. È infatti immediato dimostrare il seguente teorema.

2.30 Teorema. *Sia $A = UTU^H$ la forma normale di Schur della matrice A. Se $p(x)$ è un polinomio in x, allora $p(A) = Up(T)U^H$ e gli autovalori di $p(A)$ sono tutti e soli i numeri $p(\lambda)$, dove λ è autovalore di A.* ■

Siano $p(x)$ e $q(x)$ due polinomi nella variabile x, tali che $q(\lambda) \neq 0$ per ogni autovalore λ di A, e si consideri la funzione razionale $f(x) = p(x)/q(x)$. Per il teorema 2.30, la matrice $q(A)$ risulta non singolare ed è quindi possibile definire

$$f(A) = [q(A)]^{-1} p(A).$$

Per la matrice $f(A)$ vale un risultato analogo a quello del teorema 2.30:

$$f(A) = Uf(T)U^H. \tag{20}$$

6. Alcune proprietà delle matrici definite positive

2.31 Teorema. *Sia A una matrice hermitiana di ordine n e siano $\lambda_1, \ldots, \lambda_n$ i suoi autovalori. Allora A è definita positiva se e solo se $\lambda_i > 0,\ i = 1, \ldots, n$.*

Dim. Si dimostra prima che se A è definita positiva, allora i suoi autovalori sono positivi. Poiché A è hermitiana, i suoi autovalori sono tutti reali. Se λ è un autovalore e $\mathbf{x} \neq \mathbf{0}$ è un autovettore corrispondente, da $A\mathbf{x} = \lambda\mathbf{x}$, premoltiplicando per $\mathbf{x}^H$, si ottiene

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

Poiché A è definita positiva, il primo membro è positivo, ed essendo $\mathbf{x}^H\mathbf{x} > 0$, risulta $\lambda > 0$.

Viceversa, poiché A è hermitiana, risulta $A = UDU^H$, con U matrice unitaria e D matrice diagonale avente come elementi principali gli autovalori $\lambda_i,\ i = 1, \ldots n$, di A. Se $\mathbf{x} \in \mathbf{C}^n, \mathbf{x} \neq \mathbf{0}$, si ha:

$$\mathbf{x}^H A\mathbf{x} = \mathbf{x}^H UDU^H \mathbf{x} = \mathbf{y}^H D\mathbf{y}, \tag{21}$$

dove il vettore $\mathbf{y} = U^H\mathbf{x}$ non può essere uguale a $\mathbf{0}$, perché U è non singolare. Dalla (21) si ha:

$$\begin{aligned}\mathbf{x}^H A\mathbf{x} &= (\overline{y}_1, \ldots, \overline{y}_n) \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \\ &= \lambda_1 \overline{y}_1 y_1 + \ldots + \lambda_n \overline{y}_n y_n = \lambda_1 |y_1|^2 + \cdots + \lambda_n |y_n|^2 > 0\end{aligned}$$

poiché gli autovalori λ_i sono tutti positivi e gli $|y_i|$ non sono tutti nulli. ∎

Poiché il prodotto degli autovalori di una matrice è uguale al determinante, dal teorema 2.31 segue che il determinante di una matrice definita positiva è positivo. Inoltre dal teorema 2.31 segue che l'inversa di una matrice definita positiva è ancora definita positiva. Infatti l'inversa A^{-1} di una matrice hermitiana A è hermitiana, e gli autovalori di A^{-1} sono positivi in quanto reciproci di quelli di A.

2.32 Esempio. La matrice hermitiana

$$A = \begin{bmatrix} 1 & \mathbf{i} & 0 \\ -\mathbf{i} & 2 & -2\mathbf{i} \\ 0 & 2\mathbf{i} & 5 \end{bmatrix}$$

è definita positiva, infatti per ogni vettore $\mathbf{x} \neq \mathbf{0}$ si ha

$$\mathbf{x}^H A\mathbf{x} = |x_1 + \mathbf{i}x_2|^2 + |x_2 - 2\mathbf{i}x_3|^2 + |x_3|^2 > 0,$$

e il suo polinomio caratteristico è dato da

$$P(\lambda) = -\lambda^3 + 8\lambda^2 - 12\lambda + 1. \tag{22}$$

Esaminando il grafico di $P(\lambda)$ riportato nella figura 2.1 e calcolando i valori che questo polinomio assume nei punti 0, 1, 2, 7:

$$P(0) = 1,\ P(1) = -4,\ P(2) = 1,\ P(7) = -34,$$

risulta che il polinomio ha 3 zeri reali compresi negli intervalli

$$(0, 1),\ (1, 2) \text{ e } (2, 7),$$

e quindi gli autovalori di A sono tutti positivi. ∎

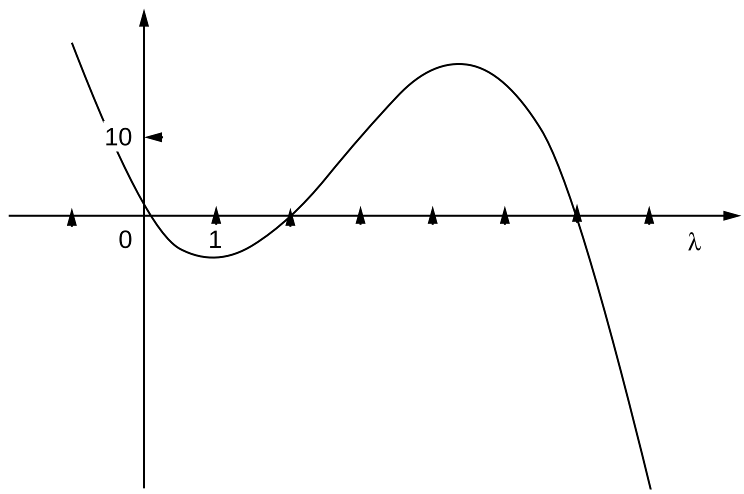

Fig. 2.1 - Grafico del polinomio(22).

2.33 Teorema. *Una matrice hermitiana A è definita positiva se e solo se i determinanti di tutte le sottomatrici principali di testa di A (e quindi anche il determinante di A) sono positivi.*

Dim. Se A è definita positiva, allora la tesi segue direttamente dal teorema 1.14. Viceversa, si suppone che i determinanti di tutte le sottomatrici principali di testa di A siano positivi e si procede per induzione su n. Per $n = 1$ il risultato è banale. Per $n > 1$, siano $\lambda_1, \ldots, \lambda_n$ gli autovalori di A. Poiché per ipotesi il prodotto dei λ_i è positivo, si dimostra che non può esistere un numero pari di autovalori negativi e quindi tutti i λ_i sono positivi. Si supponga, per assurdo, che esistano m autovalori negativi, con $m \geq 2$, m numero pari (si può supporre, senza violare la generalità, che tali autovalori siano i primi m). Sia U la matrice unitaria tale che $A = UDU^H$, in cui D è la matrice diagonale i cui elementi principali sono i λ_i, ordinati come indicato. Allora è possibile costruire due vettori $\mathbf{x}, \mathbf{y} \in \mathbf{C}^n$, tali che

$$\mathbf{x}, \mathbf{y} \neq \mathbf{0},\ x_n = 0,\ y_{m+1} = y_{m+2} = \ldots = y_n = 0,\ \mathbf{y} = U^H \mathbf{x}.$$

Infatti, partizionando la matrice U^H e i vettori $\mathbf{x}, \mathbf{y}$ nel modo seguente:

$$U^H = \begin{bmatrix} V & \mathbf{v} \\ & \\ W & \mathbf{w} \end{bmatrix} \begin{array}{l} \}\quad m \text{ righe} \\ \\ \}\ n-m \text{ righe} \end{array}$$

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \\ 0 \end{bmatrix} \begin{array}{l} \}\ n-1 \text{ componenti} \\ \\ \}\quad 1 \text{ componente} \end{array} \qquad \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \\ \mathbf{0} \end{bmatrix} \begin{array}{l} \}\quad m \text{ componenti} \\ \\ \}\ n-m \text{ componenti} \end{array}$$

dalla condizione $\mathbf{y} = U^H \mathbf{x}$ si ottiene

$$V\mathbf{x}_1 = \mathbf{y}_1$$

$$W\mathbf{x}_1 = \mathbf{0}.$$

Poiché nella matrice W il numero $(n-1)$ di colonne è maggiore del numero di righe $(n-m, m \geq 2)$, è sempre possibile determinare una combinazione lineare nulla a coefficienti non tutti nulli delle colonne di W. Esiste allora un vettore $\mathbf{x}_1 \neq \mathbf{0}$ formato da tali coefficienti, tale che $W\mathbf{x}_1 = \mathbf{0}$ e $\mathbf{y}_1 = V\mathbf{x}_1 \neq \mathbf{0}$, perché altrimenti le prime $n-1$ colonne di U^H risulterebbero linearmente dipendenti. Ne segue la relazione

$$\mathbf{x}^H A\mathbf{x} = \mathbf{x}^H UDU^H \mathbf{x} = \mathbf{y}^H D\mathbf{y} = \sum_{i=1}^{n} \lambda_i\, |y_i|^2 = \sum_{i=1}^{m} \lambda_i\, |y_i|^2 < 0,$$

che è assurda perché

$$\mathbf{x}^H A\mathbf{x} = \mathbf{x}_1^H A_{n-1}\mathbf{x}_1,$$

dove $A_{n-1} \in \mathbf{C}^{(n-1)\times(n-1)}$ è la sottomatrice principale di testa di ordine $n-1$, che è definita positiva per l'ipotesi induttiva. ∎

7. Localizzazione degli autovalori

In questo paragrafo vengono dati tre teoremi che consentono di individuare zone del piano complesso in cui si trovano gli autovalori di una matrice.

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

$$K_i = \{\, z \in \mathbf{C} \ : \ |z - a_{ii}| \le \sum_{\substack{j=1\\ j\ne i}}^{n} |a_{ij}|\,\},\ i = 1, 2, \ldots, n,$$

di centro a_{ii} e raggio $r_i = \sum\limits_{\substack{j=1\\ j\ne i}}^{n} |a_{ij}|$ sono detti *cerchi di Gerschgorin*. Vale il seguente

2.35 Teorema (primo teorema di Gerschgorin). *Gli autovalori della matrice A di ordine n sono tutti contenuti in*

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

Dim. Sia λ un autovalore di A e $\mathbf{x}$ un autovettore corrispondente, ossia

$$A\mathbf{x} = \lambda\,\mathbf{x},\ \ \mathbf{x} \ne \mathbf{0}.$$

Allora si ha:

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

da cui

$$(\lambda - a_{ii})x_i = \sum_{\substack{j=1\\ j\ne i}}^{n} a_{ij}\, x_j,\ \ i = 1, \ldots, n. \tag{23}$$

Sia x_p la componente di $\mathbf{x}$ di massimo modulo, cioè quella per cui

$$|x_p| = \max_{j=1,\ldots,n} |x_j| \ne 0, \tag{24}$$

e, ponendo $i = p$ nella (23), si ha:

$$(\lambda - a_{pp})x_p = \sum_{\substack{j=1\\ j\ne p}}^{n} a_{pj}\, x_j;$$

da cui:

$$|\lambda - a_{pp}|\,|x_p| \le \sum_{\substack{j=1\\ j\ne p}}^{n} |a_{pj}|\,|x_j|, \tag{25}$$

e, per la (24),

$$|\lambda - a_{pp}|\,|x_p| \le \sum_{\substack{j=1\\ j\ne p}}^{n} |a_{pj}|\,|x_p|;$$

infine dividendo per $|x_p| > 0$, si ottiene

$$|\lambda - a_{pp}| \le \sum_{\substack{j=1\\ j\ne p}}^{n} |a_{pj}|, \tag{26}$$

e quindi $\lambda \in K_p$. Si osservi che, poiché a priori non è noto il valore dell'indice p, è possibile solo dire che λ appartiene all'unione di tutti i cerchi K_i. ∎

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

$$H_i = \{\, z \in \mathbf{C} \ :\ |z - a_{ii}| \le \sum_{\substack{j=1\\ j\ne i}}^{n} |a_{ji}|\,\},\ i = 1, 2, \ldots, n,$$

e quindi gli autovalori di A appartengono all'insieme

$$\Big(\bigcup_{i=1,\ldots,n} K_i\Big)\bigcap\Big(\bigcup_{i=1,\ldots,n} H_i\Big).$$

2.36 Esempio. Si consideri la matrice

$$A = \begin{bmatrix} 15 & -2 & 2 \\ 1 & 10 & -3 \\ -2 & 1 & 0 \end{bmatrix} \tag{27}$$

alla quale sono associati i cerchi

$$\begin{aligned} K_1 &= \{z \in \mathbf{C} \ :\ |z - 15| \le 4\}, \\ K_2 &= \{z \in \mathbf{C} \ :\ |z - 10| \le 4\}, \\ K_3 &= \{z \in \mathbf{C} \ :\ |z| \le 3\}, \end{aligned}$$

rappresentati nella figura 2.2.

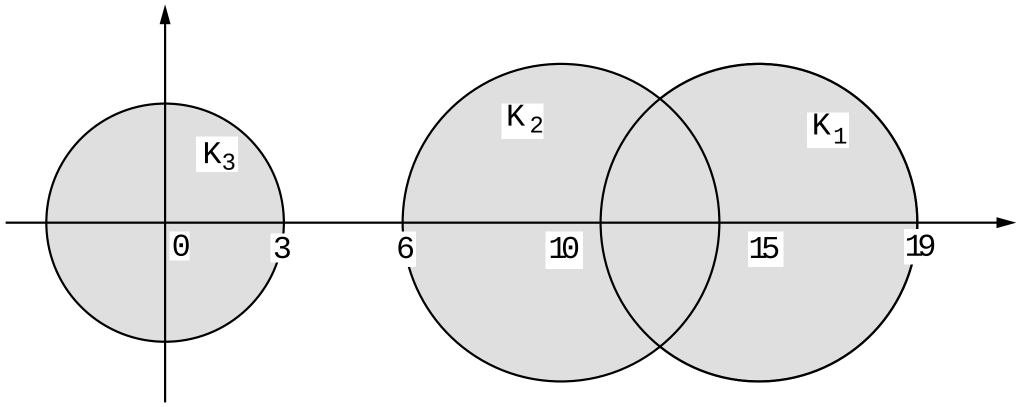

Fig. 2.2 - Cerchi di Gerschgorin associati alla matrice A in (27).

Quindi gli autovalori stanno nelle aree grigie. Si considerino poi i cerchi

$$\begin{aligned} H_1 &= \{z \in \mathbf{C} \ : \ |z - 15| \leq 3\}, \\ H_2 &= \{z \in \mathbf{C} \ : \ |z - 10| \leq 3\}, \\ H_3 &= \{z \in \mathbf{C} \ : \ |z| \leq 5\}, \end{aligned}$$

associati alla matrice A^T e rappresentati nella figura 2.3.

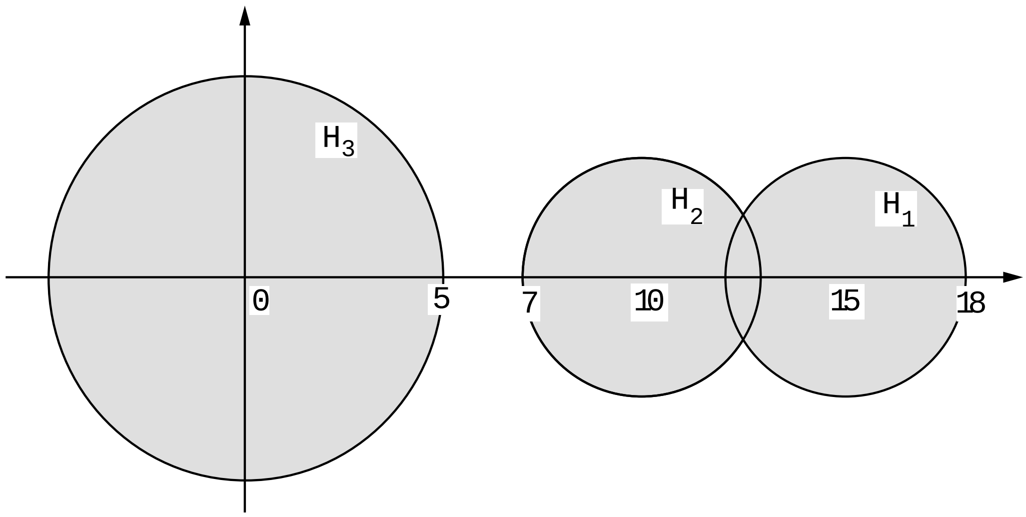

Fig. 2.3 - Cerchi di Gerschgorin associati alla matrice A^T in (27).

Quindi gli autovalori di A stanno nell'intersezione dei due insiemi di cerchi, rappresentata nella figura 2.4. ∎

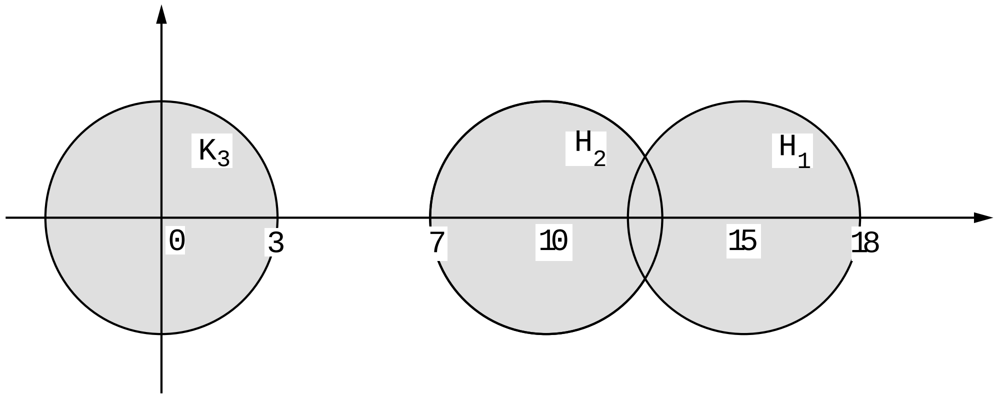

Fig. 2.4 - Intersezione dei cerchi di Gerschgorin delle figure 2.2 e 2.3.

2.37 Teorema (secondo teorema di Gerschgorin). *Se l'unione M_1 di k cerchi di Gerschgorin è disgiunta dall'unione M_2 dei rimanenti $n-k$, allora k autovalori appartengono a M_1 e $n-k$ autovalori appartengono a M_2.*

Dim. Si può supporre, senza ledere la generalità, che i cerchi che costituiscono M_1 siano i primi k, cioè che

$$M_1 = \bigcup_{i=1,\ldots,k} K_i \quad \text{e} \quad M_2 = \bigcup_{i=k+1,\ldots,n} K_i.$$

Siano D e R le matrici di elementi

$$d_{ij} = \begin{cases} a_{ij} & \text{se } i = j, \\ 0 & \text{se } i \neq j, \end{cases} \qquad r_{ij} = \begin{cases} 0 & \text{se } i = j, \\ a_{ij} & \text{se } i \neq j, \end{cases}$$

per cui $A = D + R$. La matrice

$$A(t) = D + tR, \quad t \in [0, 1],$$

i cui elementi sono funzioni continue di t, ha autovalori che sono funzioni continue di t, in quanto zeri del polinomio caratteristico i cui coefficienti sono funzioni continue degli elementi di $A(t)$. Infatti gli zeri di un polinomio, sono funzioni continue dei coefficienti (si veda [5]). Per ogni $t \in [0, 1]$ i primi k cerchi di Gerschgorin di $A(t)$ sono contenuti in M_1 perché hanno gli stessi centri dei cerchi K_i, $1, \ldots, k$, e raggio crescente con t, e analogamente i restanti $n-k$ cerchi di Gerschgorin di $A(t)$ sono contenuti in M_2. Poiché l'unione dei primi k cerchi di Gerschgorin di $A(t)$ è disgiunta dall'unione dei restanti cerchi di Gerschgorin, facendo variare con continuità t nell'intervallo [0,1], gli autovalori di $A(t)$ non possono passare da un insieme all'altro fra loro disgiunti. Per $t = 0$ in M_1 e M_2 stanno rispettivamente k e $n-k$

autovalori di $A(t)$, perché $A(0) = D$ e gli autovalori coincidono con i centri dei cerchi di Gerschgorin (che sono gli elementi principali di A). Quindi per ogni $t \in [0, 1]$, e in particolare per $t = 1$ per cui $A(1) = A$, in M_1 stanno k autovalori e in M_2 stanno $n - k$ autovalori. ∎

Dei tre autovalori della matrice A dell'esempio 2.36, uno è contenuto in K_3, mentre gli altri due appartengono ad $H_1 \cup H_2$. I due autovalori contenuti in $H_1 \cup H_2$ possono essere reali o complessi e hanno modulo compreso fra 7 e 18. L'autovalore contenuto in K_3 è reale; infatti, se avesse parte immaginaria non nulla, anche il suo coniugato dovrebbe essere un autovalore di A, essendo zero di un polinomio a coefficienti reali.

Per le matrici irriducibili vi è poi un altro teorema che precisa ulteriormente la localizzazione degli autovalori.

2.38 Teorema (terzo teorema di Gerschgorin). *Se la matrice A di ordine n è irriducibile, ogni autovalore λ, che sta sulla frontiera dei cerchi di Gerschgorin a cui appartiene, sta sulla frontiera di tutti i cerchi di Gerschgorin. In particolare questo vale per gli autovalori che appartengono alla frontiera dell'unione dei cerchi di Gerschgorin.*

Dim. Sia $\mathbf{x}$ un autovettore corrispondente all'autovalore λ, e sia x_p la sua componente di massimo modulo

$$|x_p| = \max_{j=1,\ldots,n} |x_j|.$$

Procedendo come nella dimostrazione del teorema 2.35, si ha $\lambda \in K_p$. Poiché per ipotesi λ sta sulla frontiera di K_p, la (26) deve valere con il segno di uguaglianza:

$$|\lambda - a_{pp}| = \sum_{\substack{j=1 \\ j \neq p}}^{n} |a_{pj}|.$$

Ne segue che il segno di uguaglianza deve valere anche per la (25) e quindi $|x_j| = |x_p|$, per tutti gli indici j per cui $a_{pj} \neq 0$. Per l'ipotesi di irriducibilità, esiste però almeno un indice $r, r \neq p$, per cui $a_{pr} \neq 0$, e poiché

$$|x_r| = \max_{j=1,\ldots,n} |x_j|,$$

si può riapplicare per l'indice r il procedimento appena seguito per l'indice p. Così procedendo si arriva alla conclusione che λ sta sulla frontiera di K_r ed inoltre che $|x_j| = |x_r|$, per tutti gli indici j per cui $a_{rj} \neq 0$. Il procedimento si può ripetere poi per un altro indice $s, s \neq r$, per cui $a_{rs} \neq 0$, e così via per tutti i rimanenti indici, in quanto per la irriducibilità di A, esiste un cammino orientato che tocca tutti i nodi del grafo di A. ∎

2.39 Esempio. La matrice

$$F = \begin{bmatrix} 0 & \cdot & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ & \ddots & \ddots & \vdots & \vdots \\ & & \ddots & 0 & -a_{n-2} \\ & & & 1 & -a_{n-1} \end{bmatrix}$$

è detta *matrice di Frobenius*. Calcolando il $\det(F - \lambda I)$ con la regola di Laplace applicata all'ultima colonna, risulta che

$$\det(F - \lambda I) = (-1)^n \Big(\lambda^n + \sum_{i=0}^{n-1} a_i \, \lambda^i \Big).$$

Inoltre il polinomio minimo coincide, a meno del segno, con il polinomio caratteristico. Infatti se per assurdo fosse

$$\psi(\lambda) = \lambda^k + \alpha_0 \lambda^{k-1} + \ldots + \alpha_{k-1}, \quad \text{con } k < n,$$

allora, moltiplicando la matrice $\psi(F)$ per il primo vettore della base canonica $\mathbf{e}_1$, risulterebbe

$$\begin{aligned} \psi(F)\mathbf{e}_1 &= F^k \mathbf{e}_1 + \alpha_0 F^{k-1} \mathbf{e}_1 + \ldots + \alpha_{k-1} \mathbf{e}_1 \\ &= \mathbf{e}_{k+1} + \alpha_0 \mathbf{e}_k + \ldots + \alpha_{k-1} \mathbf{e}_1 \end{aligned}$$

e quindi $\psi(F)\mathbf{e}_1$ sarebbe uguale al vettore le cui prime $k+1$ componenti sono nell'ordine

$$\alpha_{k-1}, \ldots, \alpha_0, 1,$$

e ciò è assurdo perché $\psi(F) = 0$.

Il teorema di Gerschgorin, applicato alle matrici F e F^T permette di dare al modulo degli zeri λ_i del polinomio

$$\lambda^n + \sum_{i=0}^{n-1} a_i \, \lambda^i$$

le seguenti limitazioni

$$\begin{aligned} &|\lambda_i| \le \max \{ \, |a_0|, 1 + |a_1|, \ldots, 1 + |a_{n-1}| \, \} \\ &|\lambda_i| \le \max \{ \, 1, \sum_{i=0}^{n-1} |a_i| \, \}. \end{aligned}$$

∎

8. Predominanza diagonale

Un'altra classe importante di matrici è quella delle matrici a predominanza diagonale, che si presentano spesso nella risoluzione numerica di problemi differenziali.

2.40 Definizioni. Una matrice $A \in \mathbf{C}^{n\times n}$ si dice *a predominanza diagonale* se per ogni $i = 1, \ldots, n$ risulta

$$|a_{ii}| \geq \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}|$$

ed esiste almeno un indice s per cui

$$|a_{ss}| > \sum_{\substack{j=1\\ j\neq s}}^{n} |a_{sj}|. \tag{28}$$

Una matrice $A \in \mathbf{C}^{n\times n}$ si dice *a predominanza diagonale in senso stretto* se per ogni $i = 1, \ldots, n$ risulta

$$|a_{ii}| > \sum_{\substack{j=1\\ j\neq i}} |a_{ij}|.$$

Le due definizioni di predominanza diagonale e di predominanza diagonale in senso stretto si possono dare anche per colonne, considerando le somme per colonne anziché per righe e in tal caso si specifica che la *predominanza diagonale (predominanza diagonale in senso stretto) è per colonne.* ■

2.41 Teorema. *Se $A \in \mathbf{C}^{n\times n}$ è una matrice a predominanza diagonale in senso stretto, oppure a predominanza diagonale e irriducibile, allora A è non singolare. Se inoltre A ha elementi principali tutti reali e positivi, allora gli autovalori di A hanno parte reale positiva e se A è anche hermitiana, allora A è definita positiva.*

Dim. Se A è a predominanza diagonale in senso stretto, dal teorema 2.35 risulta che i cerchi di Gerschgorin, avendo raggio minore della distanza del centro dall'origine del piano complesso, non possono includere l'origine, e quindi A non può avere un autovalore nullo.

Se A è a predominanza diagonale ed è irriducibile, è possibile che l'origine appartenga alla frontiera di un cerchio di Gerschgorin. Se però un autovalore di A fosse nullo, allora per il teorema 2.38 l'origine dovrebbe appartenere alla frontiera di tutti i cerchi di Gerschgorin, ma ciò, per la (28), non può essere vero per l's-esimo cerchio.

Inoltre, se A ha predominanza diagonale in senso stretto o ha predominanza diagonale ed è irriducibile, e se gli elementi principali di A sono tutti positivi, nessun cerchio può contenere numeri complessi a parte reale negativa. Quindi se A è anche hermitiana, cioè con autovalori reali, questi devono essere positivi, e per il teorema 2.31 risulta che la matrice è definita positiva. ∎

Poiché gli autovalori della matrice A sono uguali a quelli della matrice A^T, le tesi del teorema 2.41 valgono anche nel caso in cui la predominanza diagonale (la predominanza diagonale in senso stretto) della matrice sia per colonne.

Esercizi proposti

2.1 Si calcolino gli autovalori e gli autovettori delle matrici

$$a)\quad \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \qquad b)\quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

$$c)\quad \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \qquad d)\quad \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & -1 \\ 1 & 1 & 2 \end{bmatrix}.$$

2.2 Si determinino il polinomio caratteristico e il polinomio minimo delle seguenti matrici

$$A_1 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 2 \\ 1 & 1 & 2 \end{bmatrix}.$$

2.3 Si dica quante e quali sono le matrici di ordine 6, non simili fra di loro, il cui polinomio caratteristico è dato da

$$P(\lambda) = (3-\lambda)^4(1-\lambda)^2,$$

e per ogni matrice si scriva il polinomio minimo.

2.4 Si dica se la matrice

$$A = \begin{bmatrix} 17 & 2 & 2 \\ 1 & 10 & 3 \\ 1 & 2 & 0 \end{bmatrix}$$