CAPITOLO 5

Autovalori e autovettori

1. Generalità su autovalori e autospazi

Siano $\mathbf{A}$ una matrice $n \times n$ a coefficienti in $K \in \{\mathbb{R}, \mathbb{C}\}$ ed $f_{\mathbf{A}}\colon K^n \to K^n$ l'applicazione lineare indotta da $\mathbf{A}$: $f_{\mathbf{A}}(\mathbf{v}) = \mathbf{A}\mathbf{v}$ per ogni $\mathbf{v} \in K^n$.

chiaro che maggiore è il numero dei coefficienti nulli di $\mathbf{A}$, minore è la difficoltà di calcolo di $f_{\mathbf{A}}(\mathbf{v}) = \mathbf{A}\mathbf{v}$ per un generico vettore $\mathbf{v} \in K^n$. In particolare, come abbiamo già osservato nel Capitolo I, se $\mathbf{A} = \mathbf{D} = \mathbf{Diag}(d_1, \dots, d_n)$ allora

$$f_{\mathbf{D}}\left(\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}\right) = \begin{bmatrix} d_1 v_1 \\ \vdots \\ d_n v_n \end{bmatrix}$$

per ogni $[v_1 \ \dots \ v_n]^T \in K^n$ e se $\mathbf{A} = \lambda \mathbf{I}_n$ allora $f_{\mathbf{A}}(\mathbf{v}) = \lambda \mathbf{v}$.

Vedremo ora che anche nel caso in cui $\mathbf{A}$ non sia scalare, oltre al vettore nullo possono esistere (ed esistono quando $K = \mathbb{C}$) dei vettori di K^n per cui la loro premoltiplicazione per $\mathbf{A}$ equivale alla moltiplicazione per un opportuno scalare.

Svilupperemo una teoria che ci permetterà di trovare condizioni necessarie e sufficienti su $\mathbf{A}$ affinché esista una base $\mathcal{B}$ di K^n costituita da vettori di tale tipo. Nel caso in cui ciò avvenga, la matrice associata a $f_{\mathbf{A}}$ rispetto a $\mathcal{B}$ è una matrice diagonale $\mathbf{D}$, per cui $f_{\mathbf{A}}$ ammette una fattorizzazione del tipo $f_{\mathbf{A}} = g f_{\mathbf{D}} g^{-1}$ per un opportuno isomorfismo g di K^n.

Cominciamo con il risolvere i due seguenti problemi:

(i) stabilire per quali $\lambda \in K$ esistano sottospazi non nulli U di K^n tali che la restrizione di $f_{\mathbf{A}}$ a U coincida con la moltiplicazione per λ;

(ii) in corrispondenza di ciascuno di tali $\lambda \in K$ trovare il massimo sottospazio di K^n su cui $f_{\mathbf{A}}$ coincide con la moltiplicazione per λ.

Osserviamo innanzitutto che se U è un sottospazio di K^n su cui la restrizione di $f_{\mathbf{A}}$ coincide con la moltiplicazione per $\lambda \in K$, allora $\mathbf{A}\mathbf{u} = \lambda \mathbf{u} \in U$ per ogni $\mathbf{u} \in U$. Più in generale si pone la seguente

Definizione 1.1. Siano $\mathbf{A}$ una matrice $n \times n$ a coefficienti in K e U un sottospazio di K^n. Il sottospazio U si dice $\mathbf{A}$-*invariante* (o anche $f_{\mathbf{A}}$-*invariante*) se $f_{\mathbf{A}}(\mathbf{u}) = \mathbf{A}\mathbf{u} \in U$ per ogni $\mathbf{u} \in U$.

Esempio 1.2. (a) $\{\mathbf{0}\}$ e K^n sono sottospazi $\mathbf{A}$-invarianti.

(b) Lo spazio nullo $\mathrm{N}(\mathbf{A})$ e lo spazio delle colonne $\mathrm{C}(\mathbf{A})$ di $\mathbf{A}$ sono sottospazi $\mathbf{A}$-invarianti.

Lasciamo da verificare come esercizio (si veda l'Esercizio 2.1) che la somma e l'intersezione di sottospazi $\mathbf{A}$-invarianti di K^n sono sottospazi $\mathbf{A}$-invarianti. $\square$

Si osservi che uno scalare $\lambda \in K$ è soluzione del problema (i) se e solo se esiste un sottospazio unidimensionale U di K^n tale che la restrizione di $f_{\mathbf{A}}$ a U coincida con la moltiplicazione per λ, e quindi se e solo se esiste $\mathbf{0} \neq \mathbf{u} \in K^n$ tale che $\mathbf{A}\mathbf{u} = \lambda \mathbf{u}$.

Si pone allora la seguente definizione.

Definizione 1.3. Uno scalare $\lambda \in K$ tale che esista $\mathbf{0} \neq \mathbf{v} \in K^n$ per cui $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ si dice un *autovalore* della matrice $\mathbf{A}$. L'insieme degli autovalori distinti di $\mathbf{A}$ si dice lo *spettro* di $\mathbf{A}$.

Dunque gli autovalori della matrice $\mathbf{A}$ sono le soluzioni del problema (i). Ciascuno di essi genera un sottospazio $\mathbf{A}$-invariante non nullo minimale, ossia privo di sottospazi $\mathbf{A}$-invarianti propri.

fondamentale ricordare che il vettore $\mathbf{v}$ di cui si richiede l'esistenza nella Definizione 1.3 deve essere non nullo: $\mathbf{A}\mathbf{0} = \lambda\mathbf{0}$ per ogni $\lambda \in K$, ma non tutti gli elementi di K sono autovalori di $\mathbf{A}$.

Esempio 1.4. Sia $\mathbf{P}$ una matrice di proiezione complessa (per cui $K = \mathbb{C}$) di ordine n. Abbiamo visto (Capitolo 3, Paragrafo 6) che $\mathbb{C}^n = \mathrm{N}(\mathbf{P}) \oplus \mathrm{C}(\mathbf{P})$ e che la restrizione di $f_{\mathbf{P}}$ a $\mathrm{C}(\mathbf{P})$ è l'identità.

Se $\mathbf{P} = \mathbf{O}$ allora $\mathbb{C}^n = \mathrm{N}(\mathbf{P})$ e l'unico autovalore di $\mathbf{P}$ è $\lambda = 0$ (e quindi lo spettro di $\mathbf{P}$ è $\{0\}$); se $\mathbf{P}$ è non singolare allora $\mathbb{C}^n = \mathrm{C}(\mathbf{P})$ e l'unico autovalore di $\mathbf{P}$ è $\lambda = 1$ (e quindi lo spettro di $\mathbf{P}$ è $\{1\}$); infine se $\mathrm{N}(\mathbf{P}) \neq \{\mathbf{0}\} \neq \mathrm{C}(\mathbf{P})$, sia 0 che 1 sono autovalori di $\mathbf{P}$. Mostriamo che $\mathbf{P}$ non ha altri autovalori, ossia che lo spettro di $\mathbf{P}$ è $\{0, 1\}$. Siano dunque $\lambda \in \mathbb{C}$ un autovalore di $\mathbf{P}$ e $\mathbf{0} \neq \mathbf{v} \in \mathbb{C}^n$ tale che $\mathbf{P}\mathbf{v} = \lambda\mathbf{v}$. Da $\mathbf{P} = \mathbf{P}^2$ segue che $\lambda\mathbf{v} = \mathbf{P}\mathbf{v} = \mathbf{P}^2\mathbf{v} = \mathbf{P}(\lambda\mathbf{v}) = \lambda\mathbf{P}\mathbf{v} = \lambda^2\mathbf{v}$, per cui $(\lambda - \lambda^2)\mathbf{v} = \mathbf{0}$. Da $\mathbf{v} \neq \mathbf{0}$ si ottiene allora $\lambda - \lambda^2 = 0$, e quindi $\lambda \in \{0, 1\}$. □

Il seguente esempio mostra che il problema (i) può non avere soluzione se $K = \mathbb{R}$.

Esempio 1.5. Si consideri la matrice *reale* 2×2

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

e si osservi che $\mathbf{A}^2 = -\mathbf{I}_2$. Se $\mathbf{A}$ avesse un autovalore *reale* λ, esisterebbe un vettore $\mathbf{0} \neq \mathbf{v} \in \mathbb{R}^2$ tale che $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Premoltiplicando per $\mathbf{A}$ si otterrebbe

$$-\mathbf{v} = -\mathbf{I}_2\mathbf{v} = \mathbf{A}^2\mathbf{v} = \mathbf{A}(\mathbf{A}\mathbf{v}) = \mathbf{A}\lambda\mathbf{v} = \lambda\mathbf{A}\mathbf{v} = \lambda^2\mathbf{v},$$

e quindi, essendo $\mathbf{v} \neq \mathbf{0}$, $\lambda^2 = -1$, che non è possibile per alcun numero $\lambda \in \mathbb{R}$. □

Si usa la seguente terminologia relativamente ai sottospazi di K^n minimali e massimali (rispetto all'inclusione) su cui la premoltiplicazione per $\mathbf{A}$ coincide con la moltiplicazione per lo scalare λ.

Definizione 1.6. Se $\lambda \in K$ è un autovalore di $\mathbf{A}$, ogni vettore $\mathbf{0} \neq \mathbf{v} \in K^n$ per cui si ha $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ si dice un *autovettore di* $\mathbf{A}$ *relativo a* λ.

Definizione 1.7. Se $\lambda \in K$ è un autovalore di $\mathbf{A}$, il massimo sottospazio di K^n su cui $f_{\mathbf{A}}$ equivale alla moltiplicazione per λ si dice l'*autospazio di* $\mathbf{A}$ *relativo a* λ e si indica con il simbolo $E_{\mathbf{A}}(\lambda)$. Risulta pertanto:

$$E_{\mathbf{A}}(\lambda) = \{\mathbf{v} \in K^n : \mathbf{A}\mathbf{v} = \lambda\mathbf{v}\}.$$

La dimensione dello spazio vettoriale $E_{\mathbf{A}}(\lambda)$ si chiama *molteplicità geometrica* di λ e si indica con il simbolo $d(\lambda)$. L'insieme degli autovalori di $\mathbf{A}$ e dei corrispondenti autospazi si chiama l'*autosistema di* $\mathbf{A}$.

Pertanto gli autospazi della matrice $\mathbf{A}$ sono le soluzioni del problema (ii); inoltre per ogni autovalore λ di $\mathbf{A}$, l'autospazio $E_{\mathbf{A}}(\lambda)$ è costituito dal vettore nullo e da tutti gli autovettori di $\mathbf{A}$ relativi a λ.

Esempio 1.8. Tornando all'Esempio 1.4 si ha che se $\mathbf{P} = \mathbf{O}$ ogni vettore di $\mathbb{C}^n \setminus \{\mathbf{0}\}$ è un autovettore di $\mathbf{P}$ relativo all'autovalore 0 ed $E_{\mathbf{P}}(0) = \mathrm{N}(\mathbf{P}) = \mathbb{C}^n$. Se $\mathbf{P}$ è non singolare ogni vettore di $\mathbb{C}^n \setminus \{\mathbf{0}\}$ è un autovettore di $\mathbf{P}$ relativo all'autovalore 1 ed $E_{\mathbf{P}}(1) = \mathbb{C}^n$. Se infine $\mathrm{N}(\mathbf{P}) \neq \{\mathbf{0}\} \neq \mathrm{C}(\mathbf{P})$, ogni vettore di $\mathrm{N}(\mathbf{P}) \setminus \{\mathbf{0}\}$ è un autovettore di $\mathbf{P}$ relativo all'autovalore 0 e ogni vettore di $\mathrm{C}(\mathbf{P}) \setminus \{\mathbf{0}\}$ è un autovettore di $\mathbf{P}$ relativo all'autovalore 1.

Mostriamo che $E_{\mathbf{P}}(1)$, l'autospazio di $\mathbf{P}$ relativo all'autovalore 1, coincide con $\mathrm{C}(\mathbf{P})$.

Se $U \gneqq \mathrm{C}(\mathbf{P})$ è un sottospazio $\mathbf{P}$-invariante di $\mathbb{C}^n$, da $\mathbb{C}^n = \mathrm{N}(\mathbf{P}) \oplus \mathrm{C}(\mathbf{P})$ segue che $U = (U \cap \mathrm{N}(\mathbf{P})) \oplus \mathrm{C}(\mathbf{P})$ e $U \cap \mathrm{N}(\mathbf{P}) \neq \{\mathbf{0}\}$, per cui la restrizione di $f_{\mathbf{P}}$ a U non può coincidere con la moltiplicazione per 1. In modo analogo si prova che $E_{\mathbf{P}}(0) = \mathrm{N}(\mathbf{P})$.

Si osservi che se 1 è autovalore di $\mathbf{P}$, la sua molteplicità geometrica è uguale a $\mathrm{rk}(\mathbf{P})$. □

Abbiamo visto che se $\mathbf{P}$ è una matrice idempotente ed hermitiana di ordine n allora $\mathbb{C}^n$ si decompone in somma diretta di autospazi di $\mathbf{P}$. Il seguente esempio mostra che esistono matrici $n \times n$ per cui $\mathbb{C}^n$ non è somma diretta di loro autospazi.

Esempio 1.9. Siano
$$\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$
λ un autovalore di $\mathbf{B}$ e $[x\ y]^T$ un autovettore di $\mathbf{B}$ relativo a λ. Allora
$$\begin{bmatrix} x \\ y \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} x+y \\ y \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \lambda x \\ \lambda y \end{bmatrix}.$$
Ne segue che l'unico autovalore di $\mathbf{B}$ è $\lambda = 1$ ed $E_{\mathbf{B}}(\lambda) = \langle \mathbf{e}_1 \rangle$, per cui $\mathbb{C}^2$ non è somma di autospazi di $\mathbf{B}$. □

Vedremo che "essere hermitiana" è già una condizione sufficiente per una matrice $n \times n$ affinché $\mathbb{C}^n$ sia somma diretta di suoi autospazi, ma, come mostra il seguente esempio, non necessaria.

Esempio 1.10. Sia
$$\mathbf{C} = \begin{bmatrix} \lambda & 0 \\ \lambda & 0 \end{bmatrix}$$
con $\lambda \in \mathbb{C} \setminus \{0\}$. Da $\mathbf{C}\mathbf{e}_2 = \mathbf{0}$ e $\mathbf{C}(\mathbf{e}_1 + \mathbf{e}_2) = \mathbf{C}\mathbf{e}_1 + \mathbf{C}\mathbf{e}_2 = \lambda(\mathbf{e}_1 + \mathbf{e}_2) + \mathbf{0} = \lambda(\mathbf{e}_1 + \mathbf{e}_2)$ segue che $\mathbb{C}^n = E_{\mathbf{C}}(0) \oplus E_{\mathbf{C}}(\lambda)$ dove $E_{\mathbf{C}}(0) = \langle \mathbf{e}_2 \rangle$ ed $E_{\mathbf{C}}(\lambda) = \langle \mathbf{e}_1 + \mathbf{e}_2 \rangle$. Dunque $\mathbb{C}^2$ è somma diretta di autospazi di $\mathbf{C}$ (ed effettivamente degli autospazi di $\mathbf{C}$) pur non essendo $\mathbf{C}$ hermitiana. □

Dalla Definizione 1.3 si ottiene che uno scalare $\lambda \in K$ è un autovalore di $\mathbf{A}$ se e solo se

$$(*) \qquad \text{esiste } \mathbf{v} \neq \mathbf{0} \text{ tale che } (\mathbf{A} - \lambda \mathbf{I}_n)\mathbf{v} = \mathbf{0},$$

e un vettore $\mathbf{v} \neq \mathbf{0}$ è un autovettore di $\mathbf{A}$ relativo a λ se e solo se $\mathbf{v}$ soddisfa (*). Pertanto
$$\{\mathbf{0}\} \neq \mathrm{N}(\mathbf{A} - \lambda \mathbf{I}_n) = E_{\mathbf{A}}(\lambda).$$
Poiché
$$\mathrm{N}(\mathbf{A} - \lambda \mathbf{I}_n) \neq \{\mathbf{0}\} \iff \mathbf{A} - \lambda \mathbf{I}_n \text{ è non singolare} \iff \det(\mathbf{A} - \lambda \mathbf{I}_n) = 0,$$
gli autovalori di una matrice $\mathbf{A}$ sono tutte e sole le soluzioni dell'equazione

$$(**) \qquad \det(\mathbf{A} - X\mathbf{I}_n) = 0$$

nell'incognita X, o, equivalentemente, tutte le radici del polinomio $p_{\mathbf{A}}(X) = \det(\mathbf{A} - X\mathbf{I}_n)$ nell'indeterminata X. L'equazione (**) e il polinomio $p_{\mathbf{A}}(X)$ si chiamano rispettivamente l'*equazione caratteristica di* $\mathbf{A}$ e il *polinomio caratteristico di* $\mathbf{A}$. Abbiamo provato allora il seguente risultato.

Proposizione 1.11. *Uno scalare $\lambda \in K$ è un autovalore di $\mathbf{A}$ se e solo se λ è soluzione dell'equazione caratteristica di $\mathbf{A}$; inoltre se λ è una autovalore di $\mathbf{A}$ allora l'autospazio di $\mathbf{A}$ relativo a λ risulta*

$$E_{\mathbf{A}}(\lambda) = \mathrm{N}(\mathbf{A} - \lambda \mathbf{I}_n) \neq \{\mathbf{O}\}.$$

Esempio 1.12. Riprendiamo in considerazione gli Esempi 1.9, 1.10 e 1.5.

Il polinomio caratteristico della matrice

$$\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

dell'Esempio 1.9 è

$$\begin{aligned} p_{\mathbf{B}}(X) &= \det\left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} - X\mathbf{I}_2\right) = \det\left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix}\right) \\ &= \det\begin{bmatrix} 1-X & 1 \\ 0 & 1-X \end{bmatrix} = (1-X)^2, \end{aligned}$$

la sua equazione caratteristica è $(1-X)^2 = 0$, e quindi, come avevamo già osservato, la matrice $\mathbf{B}$ ha come unico autovalore 1 (contato due volte, dal momento che l'equazione $(1-X)^2 = 0$ è di secondo grado). Inoltre l'autospazio di $\mathbf{B}$ relativo a 1 è

$$E_{\mathbf{B}}(1) = \mathrm{N}\left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} - 1 \cdot \mathbf{I}_2\right) = \mathrm{N}\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) = \langle \mathbf{e}_1 \rangle.$$

Il polinomio caratteristico della matrice

$$\mathbf{C} = \begin{bmatrix} \lambda & 0 \\ \lambda & 0 \end{bmatrix}$$

dell'Esempio1.10, dove $\lambda \in \mathbb{C} \setminus \{0\}$, è

$$\begin{aligned} p_{\mathbf{C}}(X) &= \det\left(\begin{bmatrix} \lambda & 0 \\ \lambda & 0 \end{bmatrix} - X\mathbf{I}_2\right) = \det\left(\begin{bmatrix} \lambda & 0 \\ \lambda & 0 \end{bmatrix} - \begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix}\right) \\ &= \det\begin{bmatrix} \lambda - X & 0 \\ \lambda & -X \end{bmatrix} = -X(\lambda - X), \end{aligned}$$

la sua equazione caratteristica è $-X(\lambda - X) = 0$, e quindi, come avevamo già osservato, la matrice $\mathbf{C}$ ha come autovalori i numeri 0 e λ. Gli autospazi di $\mathbf{C}$ relativi a 0 e 1 sono

$$\begin{aligned} &E_{\mathbf{C}}(0) = \mathrm{N}(\mathbf{C}) = \mathrm{N}\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right) = \langle \mathbf{e}_2 \rangle \\ &E_{\mathbf{C}}(\lambda) = \mathrm{N}(\mathbf{C} - \lambda\mathbf{I}_2) = \mathrm{N}\left(\begin{bmatrix} 0 & 0 \\ \lambda & -\lambda \end{bmatrix}\right) = \mathrm{N}\left(\begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}\right) = \langle \mathbf{e}_1 + \mathbf{e}_2 \rangle. \end{aligned}$$

Il polinomio caratteristico della matrice

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

considerata nell'Esempio 1.5 è

$$\begin{aligned} p_{\mathbf{A}}(X) &= \det\left(\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\right) - X\mathbf{I}_2) = \det\left(\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} - \begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix}\right) \\ &= \det\begin{bmatrix} -X & -1 \\ 1 & -X \end{bmatrix} = X^2 + 1. \end{aligned}$$

Poiché $X^2 + 1$ non ha radici reali, $\mathbf{A}$ non ha autovalori reali. $\square$

2. Proprietà del polinomio caratteristico

Dal punto di vista teorico la Proposizione 1.11 risolve entrambi i problemi posti all'inizio del paragrafo. Ma dal punto di vista computazionale, anche quando si ha un'espressione esplicita del polinomio caratteristico di una matrice, il calcolo delle sue radici può presentare problemi, dal momento che non sappiamo risolvere una generica equazione polinomiale di grado qualsiasi.

Osservazione 2.1. La ricerca degli autovalori può risultare più accessibile se si individuano delle proprietà della matrice esprimibili tramite "equazioni" coinvolgenti somme di sue potenze e loro prodotti per scalari. In tal caso, infatti, da esse si ottengono corrispondenti equazioni di cui ogni autovalore deve essere soluzione, e che in generale hanno grado inferiore dell'equazione caratteristica e quindi sono più facilmente trattabili.

Un'illustrazione di questo nell'Esempio 1.4, dove abbiamo espresso il fatto che una matrice $\mathbf{A}$ sia idempotente tramite "l'equazione" $\mathbf{A}^2-\mathbf{A}=\mathbf{O}$, e trovato che gli autovalori di $\mathbf{A}$ devono essere soluzioni della corrispondente equazione $X^2-X=0$, che è un'equazione facilmente risolvibile, avendo grado 2. Vedremo più avanti che, invece, l'equazione caratteristica di una matrice di ordine n, in particolare di una matrice idempotente di ordine n, ha grado n.

Inversamente, dimostreremo nel Paragrafo 6 il teorema di Hamilton-Cayley che, partendo dal polinomio caratteristico di una generica matrice $\mathbf{A}$, e quindi da una particolare equazione soddisfatta dagli autovalori di $\mathbf{A}$, fornisce una "equazione" che deve essere soddisfatta da $\mathbf{A}$.

C'è una sostanziale differenza tra la fattorizzazione di polinomi a coefficienti reali e quella di polinomi a coefficienti complessi, in particolare di polinomi caratteristici di matrici. Tale differenza è espressa dal seguente risultato.

Teorema Fondamentale dell'Algebra. *Ogni polinomio a coefficienti complessi si fattorizza in fattori di grado* 1 *a coefficienti complessi.*

Nell'Esempio 1.5 abbiamo visto che esistono matrici reali senza autovalori reali. Invece, per il Teorema Fondamentale dell'Algebra, le matrici complesse (o quelle reali pensate come complesse) hanno sempre autovalori complessi. Più precisamente, gli autovalori complessi di una matrice complessa sono, contati con la loro molteplicità, tanti quant'è il suo ordine.

Trovarli tutti dipende esclusivamente dalla nostra abilità tecnica nel fattorizzare in fattori lineari il suo polinomio caratteristico. In particolare, come abbiamo preannunciato all'inizio del paragrafo, per una matrice complessa $\mathbf{A}$ di ordine n esistono sempre vettori non nulli di $\mathbb{C}^n$ la cui premoltiplicazione per $\mathbf{A}$ equivale alla moltiplicazione per un opportuno scalare.

Per questo motivo preferiremo trattare con matrici complesse, piuttosto che con quelle reali.

Vogliamo evidenziare alcune proprietà del polinomio caratteristico di una matrice. Si ricordi che per convenzione si assume che il polinomio nullo abbia grado $-\infty$, ossia minore di qualsiasi numero, in particolare di qualsiasi intero non negativo.

Cominciamo con il fare la seguente osservazione.

Osservazione 2.2. Se $\mathbf{B}(X)$ è una matrice quadrata di ordine m i cui elementi sono polinomi di grado ≤ 1, allora il grado del polinomio $\det(\mathbf{B}(X))$ è al più m. La dimostrazione di questo fatto, per induzione su m, è lasciata come esercizio (si veda l'Esercizio ??). Come conseguenza di ciò si ottiene in particolare che se tutti gli elementi di una colonna di $\mathbf{B}(X)$ sono elementi di K, allora il grado del polinomio $\det(\mathbf{B}(X))$ è al più $m-1$.

Lemma 2.3. *Sia* $\mathbf{A} = [a_{ij}]$ *una matrice* $n \times n$ *con* $n \geq 2$. *Allora esiste un polinomio* $f(X)$ *di grado al più* $n-2$ *tale che*

$$p_{\mathbf{A}}(X) = f(X) + \prod_{i=1,\dots,n} (a_{ii} - X).$$

Dimostrazione. Procediamo per induzione su $n \geq 2$.

Se $n = 2$ (primo passo dell'induzione) allora

$$p_{\mathbf{A}}(X) = \det(\mathbf{A} - X\mathbf{I}_2) = \det \begin{bmatrix} a_{11} - X & a_{12} \\ a_{21} & a_{22} - X \end{bmatrix} = (a_{11} - X)(a_{22} - X) + f(X)$$

dove $f(X) = -a_{12}a_{21}$. Il grado di $f(X)$ è minore o uguale a $0 = n - 2$.

Sia $n > 2$ e si supponga (ipotesi induttiva) la proposizione vera per matrici $(n-1) \times (n-1)$. Sia $\mathbf{A} = [a_{ij}]$ una matrice $n \times n$. Sviluppando

$$p_{\mathbf{A}}(X) = \det(\mathbf{A} - X\mathbf{I}_n) = \det \begin{bmatrix} a_{11} - X & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - X & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} - X \end{bmatrix}$$

rispetto alla prima riga si ottiene

$$p_{\mathbf{A}}(X) = (a_{11} - X) \det \begin{bmatrix} a_{22} - X & \dots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n2} & \dots & a_{nn} - X \end{bmatrix} + \sum_{j \geq 2} (-1)^{1+j} a_{1j} \det(\mathbf{A}_{1j}(X))$$

dove $\mathbf{A}_{1j}(X)$, per $j \geq 2$, è la matrice $(n-1) \times (n-1)$ che si ottiene da $\mathbf{A} - X\mathbf{I}_n$ sopprimendo la prima riga e la j-esima colonna. Poiché i suoi elementi sono polinomi di grado ≤ 1, e inoltre tutti gli elementi della sua prima colonna sono elementi di K, dall'Osservazione 2.2 si ottiene che $\det(\mathbf{A}_{1j}(X))$ ha grado al più $n-2$, e quindi anche $g(X) = \sum_{j \geq 2} (-1)^{1+j} a_{1j} \det(\mathbf{A}_{1j}(X))$ ha grado al più $n-2$. Per ipotesi induttiva

$$\det \begin{bmatrix} a_{22} - X & \dots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n2} & \dots & a_{nn} - X \end{bmatrix} = h(X) + \prod_{i=2,\dots,n} (a_{ii} - X),$$

dove $h(X)$ ha grado al più $n-3$. Allora $f(X) = (a_{11} - X)h(X) + g(X)$ ha grado al più $n-2$ e

$$p_{\mathbf{A}}(X) = f(X) + \prod_{i=1,\dots,n} (a_{ii} - X). \qquad \square$$

Possiamo ora agevolmente descrivere le principali proprietà cui soddisfa il polinomio caratteristico di una matrice quadrata. Ricordiamo (si veda l'Esercizio 1.17 del Capitolo I??) che se $\mathbf{A} = [a_{ij}]$ è una matrice $n \times n$, la traccia di $\mathbf{A}$ è il numero $\mathrm{Tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$.

Proposizione 2.4. *Sia* $\mathbf{A} = [a_{ij}]$ *una matrice* $n \times n$. *Allora*

(1) $p_{\mathbf{A}}(X)$ *ha grado* n;
(2) *il coefficiente di* X^n *in* $p_{\mathbf{A}}(X)$ *è* $(-1)^n$;
(3) *il coefficiente di* X^{n-1} *in* $p_{\mathbf{A}}(X)$ *è* $(-1)^{n-1}\,\mathrm{Tr}(\mathbf{A}) = (-1)^{n-1} \sum_{i=1,\dots,n} a_{ii}$;
(4) *il termine noto di* $p_{\mathbf{A}}(X)$ *è* $\det(\mathbf{A})$.

Dimostrazione. Se $\mathbf{A} = [a_{11}]$, allora $p_{\mathbf{A}}(X) = a_{11} - X$ ha grado 1, coefficiente di X uguale a -1 e termine noto $a_{11} = \det(\mathbf{A})$.

Si supponga quindi $n \geq 2$. Poiché $\prod_{i=1,\dots,n}(a_{ii} - X)$ ha grado n, dal Lemma 2.3 si ottiene (1), e inoltre che il coefficiente di X^n e il coefficiente di X^{n-1} di $p_{\mathbf{A}}(X)$ coincidono con il coefficiente di X^n e il coefficiente di X^{n-1} di $\prod_{i=1,\dots,n}(a_{ii} - X)$.

Sviluppando $\prod_{i=1,\dots,n}(a_{ii}-X)$ si ottiene che il coefficiente di X^n è il prodotto, per $i=1,\dots,n$, dei coefficienti di X in $a_{ii}-X$, per cui è $(-1)^n$. Ciò prova (2).

Il coefficiente di X^{n-1} coincide con il coefficiente di X^{n-1} di $\sum_{i=1,\dots,n} a_{ii}\prod_{j\neq i}(a_{jj}-X)$. Poiché per ogni $i=1,\dots,n$ si ha che il coefficiente di X^{n-1} di $\prod_{j\neq i}(a_{jj}-X)$ è $(-1)^{n-1}$, allora il coefficiente di X^{n-1} di $\sum_{i=1,\dots,n} a_{ii}\prod_{j\neq i}(a_{jj}-X)$ è $(-1)^{n-1}\sum_{i=1,\dots,n} a_{ii} = (-1)^{n-1}\operatorname{Tr}(\mathbf{A})$, e anche (3) è provato.

Per provare (4) si osservi che attribuendo il valore 0 all'indeterminata X si ottiene che il termine noto di $p_{\mathbf{A}}(X)$ è

$$p_{\mathbf{A}}(0) = \det(\mathbf{A} - 0\cdot\mathbf{I}) = \det(\mathbf{A}). \qquad \square$$

Siano $\mathbf{A}$ una matrice quadrata complessa di ordine n e $p_{\mathbf{A}}(X)$ il suo polinomio caratteristico. Per il Teorema Fondamentale dell'Algebra $p_{\mathbf{A}}(X)$ si fattorizza in fattori di grado 1, per cui esistono numeri complessi distinti $\lambda_1,\dots,\lambda_r$, gli autovalori di $\mathbf{A}$, e numeri naturali $m_1,\dots,m_r$ tali che

$$p_{\mathbf{A}}(X) = c(X-\lambda_1)^{m_1}(X-\lambda_2)^{m_2}\dots(X-\lambda_r)^{m_r}$$

per un opportuno numero complesso c. Il numero naturale m_i è detto la *molteplicità algebrica dell'autovalore* λ_i, per ogni $i=1,\dots,r$.

Essa può non coincidere con la molteplicità geometrica. Per esempio, l'autovalore 1 della matrice $\mathbf{B}$ nell'Esempio 1.12 ha molteplicità algebrica uguale a 2 e molteplicità geometrica uguale a 1.

Per il punto (1) della Proposizione 2.4 $p_{\mathbf{A}}(X)$ ha grado n, per cui si ha

$$n = m_1 + m_2 + \dots + m_r.$$

Abbiamo quindi provato il seguente risultato.

PROPOSIZIONE 2.5. *La somma delle molteplicità algebriche degli autovalori di una matrice quadrata è uguale all'ordine della matrice.* $\square$

Inoltre per il punto (2) della Proposizione 2.4 il coefficiente di X^n di $p_{\mathbf{A}}(X)$ è $(-1)^n$, quindi $c=(-1)^n$ e si ha:

$$p_{\mathbf{A}}(X) = (\lambda_1 - X)^{m_1}(\lambda_2 - X)^{m_2}\dots(\lambda_r - X)^{m_r}.$$

In particolare $p_{\mathbf{A}}(X) = p_{\mathbf{D}}(X)$, dove $\mathbf{D}$ è la matrice diagonale di ordine n i cui elementi diagonali sono gli autovalori di $\mathbf{A}$ ciascuno contato con la sua molteplicità algebrica. Applicando la Proposizione 2.4 a $\mathbf{D}$ e $\mathbf{A}$ si ottiene allora il seguente risultato.

PROPOSIZIONE 2.6. *Sia A una matrice complessa.*

(1) *La traccia di $\mathbf{A}$ coincide con la somma degli autovalori di $\mathbf{A}$ autovalori ciascuno contato con la sua molteplicità algebrica.*
(2) *Il determinante di $\mathbf{A}$ coincide con il prodotto degli autovalori di $\mathbf{A}$ ciascuno contato con la sua molteplicità algebrica.* $\square$

3. Proprietà degli autospazi

Sia $\mathbf{A}$ una matrice complessa $n\times n$. Abbiamo già accennato nel paragrafo precedente dal fatto che l'idea di fondo per lo studio dell'applicazione lineare $f_{\mathbf{A}}\colon \mathbb{C}^n \to \mathbb{C}^n$ è quella di cercare, ed eventualmente costruire effettivamente, un'opportuna base $\mathcal{B}$ di $\mathbb{C}^n$ rispetto alla quale la matrice $\mathbf{B}$ associata a $f_{\mathbf{A}}$ abbia molti coefficienti nulli. In tal caso, come abbiamo già osservato, il calcolo di $f_{\mathbf{B}}$ è meno problematico ed $f_{\mathbf{A}}$ ammette una fattorizzazione del tipo $f_{\mathbf{A}} = g f_{\mathbf{B}} g^{-1}$ con g isomorfismo di $\mathbb{C}^n$. Dal punto di vista matriciale, se $\mathbf{S}$ è la matrice che ha come colonne i vettori della base $\mathcal{B}$, la relazione che intercorre tra $\mathbf{A}$ e $\mathbf{B}$ è $\mathbf{A} = \mathbf{S}\mathbf{B}\mathbf{S}^{-1}$ (si veda il Capitolo II).

Definizione 3.1. Due matrici quadrate $\mathbf{A}$ e $\mathbf{B}$ si dicono *simili* se esiste una matrice non singolare $\mathbf{S}$ tale che $\mathbf{A} = \mathbf{S}\mathbf{B}\mathbf{S}^{-1}$.

La relazione di similitudine è una relazione riflessiva, simmetrica e transitiva (si veda l'Esercizio ??), ossia è una relazione di equivalenza, e, come proviamo nella prossima proposizione, conserva i polinomi caratteristici e le molteplicità algebriche e geometriche degli autovalori.

Proposizione 3.2. *Due matrici simili hanno lo stesso polinomio caratteristico e gli stessi autovalori con le stesse molteplicità algebriche e geometriche.*

Dimostrazione. Siano $\mathbf{A}$ e $\mathbf{B}$ due matrici simili di ordine n ed $\mathbf{S}$ una matrice non singolare tale che $\mathbf{A} = \mathbf{S}\mathbf{B}\mathbf{S}^{-1}$. Allora

$$\begin{aligned} p_{\mathbf{A}}(X) &= \det(\mathbf{A} - X\mathbf{I}_n) = \det(\mathbf{S}\mathbf{B}\mathbf{S}^{-1} - X\mathbf{I}_n) = \det(\mathbf{S}(\mathbf{B} - X\mathbf{I}_n)\mathbf{S}^{-1}) \\ &= \det(\mathbf{S})\det(\mathbf{B} - X\mathbf{I}_n)(\det(\mathbf{S}))^{-1} = \det(\mathbf{B} - X\mathbf{I}_n) = p_{\mathbf{B}}(X) \end{aligned}$$

per cui $\mathbf{A}$ e $\mathbf{B}$ hanno lo stesso polinomio caratteristico, e quindi anche gli stessi autovalori con le stesse molteplicità algebriche.

Sia ora λ un autovalore di $\mathbf{A}$ e $\mathbf{B}$. Da

$$\mathbf{v} \in E_{\mathbf{B}}(\lambda) \iff \mathbf{B}\mathbf{v} = \lambda\mathbf{v} \iff \mathbf{S}^{-1}\mathbf{A}\mathbf{S}\mathbf{v} = \lambda\mathbf{v} \iff \mathbf{A}\mathbf{S}\mathbf{v} = \lambda\mathbf{S}\mathbf{v} \iff \mathbf{S}\mathbf{v} \in E_{\mathbf{A}}(\lambda),$$

segue che la restrizione a $E_{\mathbf{B}}(\lambda)$ della premoltiplicazione per $\mathbf{S}$ è un'applicazione lineare da $E_{\mathbf{B}}(\lambda)$ a $E_{\mathbf{A}}(\lambda)$. Tale applicazione lineare è un isomorfismo: la sua inversa è la restrizione a $E_{\mathbf{A}}(\lambda)$ della premoltiplicazione per $\mathbf{S}^{-1}$. In particolare gli spazi vettoriali $E_{\mathbf{A}}(\lambda)$ ed $E_{\mathbf{B}}(\lambda)$ hanno la stessa dimensione, ossia la molteplicità geometrica di λ in quanto autovalore di $\mathbf{A}$ coincide con la molteplicità geometrica di λ in quanto autovalore di $\mathbf{B}$. $\square$

Nel teorema che segue, impiegando la relazione di similitudine tra matrici, proviamo che la molteplicità geometrica di un autovalore è minore o uguale a quella algebrica. Questo fatto ci permetterà nel prossimo paragrafo di caratterizzare in modo soddisfacente le matrici diagonalizzabili.

Teorema 3.3. *Siano $\mathbf{A}$ una matrice $n \times n$ e λ un autovalore di $\mathbf{A}$ con molteplicità geometrica e algebrica d ed m rispettivamente. Allora $d \leq m$.*

Dimostrazione. Si estenda una base $\{\mathbf{v}_1; \mathbf{v}_2; \dots; \mathbf{v}_d\}$ dell'autospazio $E_{\mathbf{A}}(\lambda)$ a una base $\mathcal{B} = \{\mathbf{v}_1; \mathbf{v}_2; \dots; \mathbf{v}_d; \mathbf{v}_{d+1}; \dots; \mathbf{v}_n\}$ di $\mathbb{C}^n$, e si consideri la matrice $\mathbf{B}$ associata all'applicazione lineare $f_{\mathbf{A}}\colon \mathbb{C}^n \to \mathbb{C}^n$ rispetto alla base $\mathcal{B}$ (su dominio e codominio). Per ogni $i = 1, \dots, d$ si ha che

$$\mathbf{B}\mathbf{e}_i = \mathrm{C}_{\mathcal{B}}(\mathbf{A}\mathbf{v}_i) = \mathrm{C}_{\mathcal{B}}(\lambda\mathbf{v}_i) = \lambda\mathbf{e}_i,$$

dove $\mathbf{e}_1, \dots, \mathbf{e}_d$ sono le prime d colonne della matrice identica $\mathbf{I}_n$. Pertanto $\mathbf{B}$ è una matrice triangolare superiore a blocchi del tipo

$$\mathbf{B} = \begin{bmatrix} \lambda\mathbf{I}_d & \mathbf{C} \\ \mathbf{O} & \mathbf{D} \end{bmatrix}$$

con $\mathbf{D}$ matrice quadrata. Quindi $p_{\mathbf{B}}(X) = (x - \lambda)^d p_{\mathbf{D}}(X)$.

Poiché $\mathbf{A}$ è la matrice associata all'applicazione lineare $f_{\mathbf{A}}$ rispetto alla base canonica di $\mathbb{C}^n$ (su dominio e codominio), $\mathbf{B}$ è simile ad $\mathbf{A}$, e dalla Proposizione 3.2 segue che $p_{\mathbf{B}}(X) = p_{\mathbf{A}}(X)$. Dunque il polinomio $(X - \lambda)^d$ divide il polinomio caratteristico di $\mathbf{A}$, per cui la molteplicità geometrica d di λ è minore o uguale a quella algebrica m. $\square$

Ricordiamo (si veda il Paragrafo ?? del capitolo II) che se U e Z sono due sottospazi vettoriali dello spazio vettoriale V, la somma di U e Z

$$U + Z = \{\mathbf{u} + \mathbf{z} : \mathbf{u} \in U, \mathbf{z} \in Z\}$$

è il più piccolo sottospazio di V contenente U e Z. Inoltre, se ogni elemento di $U+Z$ si scrive in modo unico come somma di un vettore di U e un vettore di Z, allora la somma di U e Z si chiama somma diretta e si indica con il simbolo $U \oplus Z$. Quanto detto si può estendere a un numero finito di sottospazi di V.

DEFINIZIONE 3.4. Siano $U_1, \ldots, U_n$ sottospazi vettoriali dello spazio vettoriale V. La *somma*

$$\sum_{i=1}^{n} U_i = \{\mathbf{u}_1 + \cdots + \mathbf{u}_n : \mathbf{u}_i \in U_i, i = 1, \ldots, n\}$$

è il più piccolo sottospazio di V contenente $U_1, \ldots, U_n$. Se ogni elemento di $\sum_{i=1}^{n} U_i$ si scrive in modo unico come somma di vettori di $U_1, \ldots, U_n$, la somma $\sum_{i=1}^{n} U_i$ si chiama *diretta* e si indica con il simbolo $\bigoplus_{i=1}^{n} U_i$.

Si osservi che per verificare che $\sum_{i=1}^{n} U_i = \bigoplus_{i=1}^{n} U_i$, è sufficiente verificare che i sottospazi $U_1, \ldots, U_n$ sono *indipendenti*, nel senso che

$$\sum_{i=1}^{n} \mathbf{u}_i = \mathbf{0}, \quad \mathbf{u}_i \in U_i \quad \text{per ogni } i = 1, \ldots, n \implies \mathbf{u}_i = \mathbf{0} \quad \text{per ogni } i = 1, \ldots, n.$$

Si osservi inoltre che se $\sum_{i=1}^{n} U_i = \bigoplus_{i=1}^{n} U_i$, e $\mathcal{B}_i$ è una base di U_i per ogni $i = 1, \ldots, n$, allora $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_n$ è una base di $\bigoplus_{i=1}^{n} U_i$. Pertanto il risultato che segue esprime il fatto che se $\{\lambda_1, \ldots, \lambda_k\}$ è lo spettro di una matrice quadrata $\mathbf{A}$, esiste una base di $\sum_{i=1}^{k} E_{\mathbf{A}}(\lambda_i)$ rispetto alla quale la matrice associata alla restrizione di $f_{\mathbf{A}}$ a $\sum_{i=1}^{k} E_{\mathbf{A}}(\lambda_i)$ è diagonale. Infatti, ciò vale per ciascun autospazio, e la base cercata si ottiene dall'unione delle basi scelte nei diversi autospazi.

TEOREMA 3.5. *Siano* $\mathbf{A}$ *una matrice* $n \times n$ *e* $\lambda_1, \ldots, \lambda_k$ *autovalori distinti di* $\mathbf{A}$. *Allora si ha* $\sum_{i=1}^{k} E_{\mathbf{A}}(\lambda_i) = \bigoplus_{i=1}^{k} E_{\mathbf{A}}(\lambda_i)$.

DIMOSTRAZIONE. Abbiamo osservato che è sufficiente provare che se $\mathbf{v}_i \in E_{\mathbf{A}}(\lambda_i)$ allora

$$\sum_{i=1}^{k} \mathbf{v}_i = \mathbf{0} \implies \mathbf{v}_i = \mathbf{0} \quad \text{per ogni } i = 1, \ldots, k.$$

Procediamo per induzione su k. Per $k = 1$ non c'è nulla da dimostrare. Sia allora $k > 1$ e si supponga che la somma degli autospazi relativi a $k-1$ autovalori distinti di $\mathbf{A}$ sia diretta.

Premoltiplicando ambo i membri dell'uguaglianza $\sum_{i=1}^{k} \mathbf{v}_i = \mathbf{0}$ per λ_1 si ottiene

$$(*) \qquad \sum_{i=1}^{k} \lambda_1 \mathbf{v}_i = \mathbf{0},$$

mentre premoltiplicandoli per $\mathbf{A}$ si ottiene

$$(**) \qquad \mathbf{0} = \mathbf{A}\mathbf{0} = \mathbf{A} \sum_{i=1}^{k} \mathbf{v}_i = \sum_{i=1}^{k} \mathbf{A}\mathbf{v}_i = \sum_{i=1}^{k} \lambda_i \mathbf{v}_i.$$

Sottraendo ora membro a membro (**) da (*), si ricava che

$$\sum_{i=2}^{k} (\lambda_i - \lambda_1)\mathbf{v}_i = \mathbf{0},$$

da cui, per ipotesi induttiva, $(\lambda_i - \lambda_1)\mathbf{v}_i = \mathbf{0}$ per ogni $i \geq 2$. Poiché gli autovalori sono distinti, $\lambda_i - \lambda_1 \neq 0$ per ogni $i \geq 2$, per cui $\mathbf{v}_i = \mathbf{0}$ per ogni $i \geq 2$. Da ciò segue che anche $\mathbf{v}_1 = \mathbf{0}$, essendo $\sum_{i=1}^{k} \mathbf{v}_i = \mathbf{0}$. La dimostrazione è cos completata. $\square$

Come conseguenza del Teorema 3.5 si ha il seguente risultato.

COROLLARIO 3.6. *Autovettori relativi ad autovalori distinti sono linearmente indipendenti.*

DIMOSTRAZIONE. Siano $\mathbf{A}$ una matrice quadrata, $\lambda_1, \dots, \lambda_k$ suoi autovalori distinti, $\mathbf{v}_1, \dots, \mathbf{v}_k$ autovettori a essi corrispondenti, e $\alpha_1 \mathbf{v}_1 + \cdots + \alpha_k \mathbf{v}_k = \mathbf{0}$ una loro combinazione lineare nulla. Poiché $E_{\mathbf{A}}(\lambda_i)$ è uno spazio vettoriale per ogni $i = 1, \dots, k$, allora $\alpha_i \mathbf{v}_i \in E_{\mathbf{A}}(\lambda_i)$, per cui dal Teorema 3.5 si ottiene che $\alpha_i \mathbf{v}_i = \mathbf{0}$ per ogni $i = 1, \dots, k$. Dal momento che ciascun $\mathbf{v}_i$, in quanto autovettore di $\mathbf{A}$, è non nullo, da ciò segue che $\alpha_i = 0$ per ogni $i = 1, \dots, k$, ossia che i vettori $\mathbf{v}_1, \dots, \mathbf{v}_k$ sono linearmente indipendenti. □

4. Matrici diagonalizzabili e matrici triangolarizzabili

Nel Paragrafo 2 abbiamo già espresso il concetto di matrice diagonalizzabile in relazione all'applicazione lineare da essa indotta. Riprendiamo ora quel concetto.

DEFINIZIONE 4.1. Si dice che una matrice $\mathbf{A}$ è *diagonalizzabile* se è simile a una matrice diagonale, ossia se esistono una matrice diagonale $\mathbf{D}$ e una matrice non singolare $\mathbf{S}$ tali che $\mathbf{A} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1}$. In tal caso si dice che $\mathbf{S}$ diagonalizza $\mathbf{A}$.

Oltre che per lo studio dell'applicazione lineare indotta, una situazione in cui risulta conveniente sapere se una matrice è diagonalizzabile, è quando si vogliono calcolare le sue potenze (si veda il Paragrafo 1), dal momento che se $\mathbf{A} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1}$, allora per ogni numero naturale k si ha che $\mathbf{A}^k = \mathbf{S}\mathbf{D}^k\mathbf{S}^{-1}$, e se $\mathbf{D} = \mathbf{Diag}(d_1, \dots, d_n)$, allora

$$\mathbf{D}^k = \mathbf{Diag}(d_1^k, \dots, d_n^k).$$

Il problema del calcolo delle potenze di una matrice è trattabile più facilmente non solo nel caso di matrici diagonali o diagonalizzabili, ma anche nel caso di matrici triangolari: sappiamo infatti che se $\mathbf{T}$ è una matrice triangolare, per ogni numero naturale k anche $\mathbf{T}^k$ è triangolare e i coefficienti diagonali di $\mathbf{T}^k$ sono le potenze k-esime dei coefficienti diagonali di $\mathbf{T}$.

Possiamo allora dare una definizione analoga.

DEFINIZIONE 4.2. Si dice che una matrice $\mathbf{A}$ è *triangolarizzabile* se è simile a una matrice triangolare, ossia se esistono una matrice triangolare $\mathbf{T}$ e una matrice non singolare $\mathbf{S}$ tali che $\mathbf{A} = \mathbf{S}\mathbf{T}\mathbf{S}^{-1}$. In tal caso si dice che $\mathbf{S}$ triangolarizza $\mathbf{A}$.

Abbiamo ora tutti gli strumenti per caratterizzare le matrici diagonalizzabili.

TEOREMA 4.3. *Siano $\mathbf{A}$ una matrice complessa $n \times n$ e $\lambda_1, \dots, \lambda_k$ i suoi autovalori distinti. Sono equivalenti i seguenti fatti:*

(1) $\mathbf{A}$ *è diagonalizzabile;*
(2) $\mathbb{C}^n$ *ha una base costituita da autovettori di* $\mathbf{A}$*;*
(3) $\mathbb{C}^n = \bigoplus_{i=1}^k E_{\mathbf{A}}(\lambda_i)$*;*
(4) *la molteplicità geometrica di ciascun autovalore λ_i coincide con la sua molteplicità algebrica.*

DIMOSTRAZIONE. (1) $\Longrightarrow$ (2) Poiché

$$\mathbf{A} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1} \iff \mathbf{A}\mathbf{S} = \mathbf{S}\mathbf{D} \iff \mathbf{A}\mathbf{S}\mathbf{e}_i = \mathbf{S}\mathbf{D}\mathbf{e}_i \quad \text{per ogni } i = 1, \dots, n,$$

allora $\mathbf{A}$ è diagonalizzabile se e solo se esistono $\mathbf{D} = \mathbf{Diag}(d_1, \dots, d_n)$ ed $\mathbf{S}$ invertibile tali che

$$\mathbf{A}(\mathbf{S}\mathbf{e}_i) = \mathbf{S}d_i\mathbf{e}_i = d_i(\mathbf{S}\mathbf{e}_i) \quad \text{per ogni } i = 1, \dots, n,$$

ossia se e solo se l'insieme $\{\mathbf{S}\mathbf{e}_1; \dots; \mathbf{S}\mathbf{e}_n\}$ delle colonne di $\mathbf{S}$, che essendo $\mathbf{S}$ invertibile è una base di $\mathbb{C}^n$, è costituito da autovettori di $\mathbf{A}$. Si osservi che in tal caso gli elementi diagonali di $\mathbf{D}$ sono gli autovalori di $\mathbf{A}$, ripetuti tante volte quanto è la loro molteplicità algebrica.

(2) $\Longrightarrow$ (4) Sia $\mathcal{B}$ una base di $\mathbb{C}^n$ costituita da autovettori di $\mathbf{A}$. Per ogni $i = 1, \ldots, k$ sia s_i il numero degli elementi di $\mathcal{B}$ appartenenti a $E_{\mathbf{A}}(\lambda_i)$. Allora $s_i \le d_i$ e, per il Teorema 3.3, $d_i \le m_i$ per ogni $i = 1, \ldots, k$. Poiché $\mathcal{B}$ ha n elementi, allora $\sum_{i=1}^k s_i = n$. D'altra parte per la Proposizione 2.5 si ha anche che $\sum_{i=1}^k m_i = n$, per cui $s_i = d_i = m_i$ per ogni $i = 1, \ldots, k$.

(4) $\Longrightarrow$ (3) Se ogni autovalore λ_i ha molteplicità geometrica d_i e algebrica m_i uguali, allora $\sum_{i=1}^k m_i = \sum_{i=1}^k d_i$, per cui dalla Proposizione 2.5 si ottiene

$$\dim(\mathbb{C}^n) = n = \sum_{i=1}^k m_i = \sum_{i=1}^k d_i = \dim\Big(\bigoplus_{i=1}^k E_{\mathbf{A}}(\lambda_i)\Big)$$

e quindi $\mathbb{C}^n = \bigoplus_{i=1}^k E_{\mathbf{A}}(\lambda_i)$.

(3) $\Longrightarrow$ (2) Sia $\mathcal{B}_i$ una base di $E_{\mathbf{A}}(\lambda_i)$ per ogni $i = 1, \ldots, k$. Allora $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k$ è una base di $\bigoplus_{i=1}^k E_{\mathbf{A}}(\lambda_i) = \mathbb{C}^n$ costituita da autovettori di $\mathbf{A}$. $\square$

Dal Teorema precedente si deduce che le matrici di proiezione sono matrici diagonalizzabili (si veda l'Esempio 1.4), cos come la matrice dell'Esempio 1.10. Non è invece diagonalizzabile la matrice dell'Esempio 1.9.

Osservando che se $\mathbf{A}$ è reale un autovalore λ di $\mathbf{A}$ è reale se e solo se gli autovettori di $\mathbf{A}$ relativi a λ possono essere scelti a coordinate reali, si ottiene il caso reale del teorema precedente.

Teorema 4.4. *Siano* $\mathbf{A}$ *una matrice reale* $n \times n$ *e* $\lambda_1, \ldots, \lambda_k$ *i suoi autovalori complessi distinti. Sono equivalenti i seguenti fatti:*

(1) $\mathbf{A}$ *è diagonalizzabile con una matrice reale;*
(2) $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ *e* $\mathbb{R}^n$ *ha una base costituita da autovettori di* $\mathbf{A}$*;*
(3) $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ *e* $\mathbb{R}^n = \bigoplus_{i=1}^k (E_{\mathbf{A}}(\lambda_i) \cap \mathbb{R}^n)$*;*
(4) $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ *e per ogni* $i = 1, \ldots, k$ *la dimensione del sottospazio* $E_{\mathbf{A}}(\lambda_i) \cap \mathbb{R}^n$ *di* $\mathbb{R}^n$ *coincide con la molteplicità algebrica* m_i *di* λ_i. $\square$

Abbiamo allora un'immediata conseguenza dei Teoremi 3.3 e 4.3.

Corollario 4.5. *Una matrice* $n \times n$ *con* n *autovalori distinti è diagonalizzabile.* $\square$

Naturalmente le matrici diagonali, e quindi in particolare le matrici scalari, sono diagonalizzabili, per cui esistono matrici diagonalizzabili che non hanno autovalori distinti.

Esempio 4.6. La matrice dell'Esempio 1.5 è una matrice reale con autovalori complessi distinti e per il Corollario 4.5 è diagonalizzabile come matrice complessa. Dal momento però che i suoi autovalori, i e $-i$, non sono reali, essa non è diagonalizzabile con una matrice reale. $\square$

Per passare da una matrice a una a essa simile, occorre calcolare l'inversa della matrice $\mathbf{S}$ che realizza la similitudine. In alcuni casi questo calcolo può essere particolarmente semplice, come quando, per esempio, si riduce a una trasposizione o a una H-trasposizione.

Definizione 4.7. Una matrice $\mathbf{A}$ si dice *ortogonale* se $\mathbf{A}^{-1} = \mathbf{A}^T$; si dice *unitaria* se $\mathbf{A}^{-1} = \mathbf{A}^H$.

Le matrici reali ortogonali vengono chiamate in questo modo perché, come anche le matrici complesse unitarie, sono caratterizzate dal fatto che l'insieme delle loro colonne è un insieme ortonormale di vettori di $\mathbb{R}^n$, rispettivamente di $\mathbb{C}^n$.

Ovviamente una matrice reale è ortogonale se e solo se è unitaria, ma in generale esistono matrici ortogonali che non sono unitarie ed esistono matrici unitarie che non sono ortogonali.

Definizione 4.8. Se due matrici $\mathbf{A}$ e $\mathbf{B}$ sono simili e la matrice $\mathbf{S}$ che realizza la similitudine è ortogonale (rispettivamente unitaria), allora $\mathbf{A}$ e $\mathbf{B}$ si dicono *ortogonalmente simili* (rispettivamente *unitariamente simili*). Si parla anche di *similitudini ortogonali* e di *similitudini unitarie.*

Abbiamo già osservato che se $\mathbf{A}$ e $\mathbf{B}$ sono simili e $\mathbf{S}$ è la matrice che realizza la similitudine (ossia se $\mathbf{A} = \mathbf{S}\mathbf{B}\mathbf{S}^{-1}$), allora $\mathbf{B}$ è la matrice associata a $f_{\mathbf{A}}$ rispetto alla base ordinata $\mathcal{B}$ di $\mathbb{C}^n$ che ha come i-esimo vettore la i-esima colonna di $\mathbf{S}$. Se $\mathbf{A}$ e $\mathbf{B}$ sono matrici reali e $\mathbf{S}$ è una matrice reale ortogonale, ossia se $\mathbf{A}$ e $\mathbf{B}$ sono matrici reali ortogonalmente simili (e analogamente se $\mathbf{S}$ è una matrice unitaria, ossia se $\mathbf{A}$ e $\mathbf{B}$ sono unitariamente simili), allora la base $\mathcal{B}$ di $\mathbb{R}^n$ (rispettivamente di $\mathbb{C}^n$) è una base ortonormale.

Diamo ora una caratterizzazione delle matrici unitariamente simili a matrici triangolari, ottenendo in particolare che ogni matrice complessa è unitariamente simile a una matrice triangolare. Questo fatto verrà usato nel Capitolo VI per caratterizzare le matrici che sono unitariamente diagonalizzabili, ossia unitariamente simili a una matrice diagonale.

Non è invece vero che ogni matrice reale è ortogonalmente simile a una matrice triangolare. Tale diversità dipende dal fatto che per il Teorema fondamentale dell'Algebra ogni polinomio complesso si fattorizza in fattori lineari, e quindi ogni matrice complessa di ordine n ha n autovalori, se contati con la loro molteplicità, mentre le matrici reali di ordine n non necessariamente hanno n autovalori reali, anzi possono addirittura non averne alcuno (si veda l'Esempio 1.5).

Sia $K \in \{\mathbb{R}, \mathbb{C}\}$.

Teorema 4.9 (Schur). *Sia $\mathbf{A}$ una matrice in $M_n(K)$. $\mathbf{A}$ è unitariamente simile in $M_n(K)$ a una matrice triangolare superiore se e solo se tutti gli n autovalori di $\mathbf{A}$ sono elementi di K.*

Dimostrazione. Sia $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^H$ con $\mathbf{U}$ unitaria e $\mathbf{T} = [t_{ij}] \in \mathrm{M}_n(K)$ triangolare. Allora per la Proposizione 3.2 si ha

$$p_{\mathbf{A}}(X) = p_{\mathbf{T}}(X) = (t_{11} - X)(t_{22} - X)\dots(t_{nn} - X)$$

per cui gli autovalori di $\mathbf{A}$ sono i coefficienti diagonali di $\mathbf{T}$, e pertanto elementi di K.

Viceversa si supponga che tutti gli n autovalori di $\mathbf{A} \in \mathrm{M}_n(K)$ siano elementi di K.

Procediamo per induzione su n per provare che $\mathbf{A}$ è unitariamente simile a una matrice triangolare superiore. Se $n = 1$ non c'è nulla da dimostrare. Supponiamo allora $n > 1$ e che ogni matrice $(n-1) \times (n-1)$ a coefficienti in K soddisfacente all'ipotesi sia unitariamente simile in $\mathrm{M}_{n-1}(K)$ a una matrice triangolare superiore.

Siano λ_1 un autovalore di $\mathbf{A}$ e $\mathbf{v}_1$ un autovettore di $\mathbf{A}$ relativo a λ_1 con $\|\mathbf{v}_1\|_2 = 1$. Completando $\mathbf{v}_1$ a una base ortonormale $\{\mathbf{v}_1; \dots; \mathbf{v}_n\}$ di K^n si ha che $\mathbf{U}_1 = [\mathbf{v}_1 \ \dots \ \mathbf{v}_n]$ è una matrice unitaria e che

$$\mathbf{U}_1^{-1}\mathbf{A}\mathbf{U}_1\mathbf{e}_1 = \mathbf{U}_1^{-1}\mathbf{A}\mathbf{v}_1 = \mathbf{U}_1^{-1}\lambda_1\mathbf{v}_1 = \lambda_1\mathbf{U}_1^{-1}\mathbf{v}_1 = \lambda_1\mathbf{e}_1,$$

per cui

$$\mathbf{A}' = \mathbf{U}_1^{-1}\mathbf{A}\mathbf{U}_1 = \begin{bmatrix} \lambda_1 & \mathbf{w}^T \\ \mathbf{0} & \mathbf{C} \end{bmatrix}$$

per opportune matrici $\mathbf{w}$ e $\mathbf{C}$. Ne segue che

$$p_{\mathbf{A}}(X) = p_{\mathbf{A}'}(X) = (\lambda_1 - X)p_{\mathbf{C}}(X),$$

per cui gli autovalori di $\mathbf{C}$ sono autovalori di $\mathbf{A}$ e quindi, per ipotesi, elementi di K.

Per ipotesi induttiva esistono una matrice unitaria $\mathbf{V} \in \mathrm{M}_{n-1}(K)$ e una matrice triangolare superiore $\mathbf{T}_1 \in \mathrm{M}_{n-1}(K)$ tali che

$$\mathbf{C} = \mathbf{V}\mathbf{T}_1\mathbf{V}^H.$$

Posto $\mathbf{U}_2 = \mathbf{Diag}(1, \mathbf{V})$ e $\mathbf{U} = \mathbf{U}_1\mathbf{U}_2$, si ha che $\mathbf{U}_2$ è unitaria essendolo $\mathbf{V}$ e $\mathbf{U}$ è unitaria in quanto prodotto di due matrici unitarie. Inoltre

$$\begin{aligned}\mathbf{U}^{-1}\mathbf{A}\mathbf{U} &= \mathbf{U}_2^{-1}\mathbf{U}_1^{-1}\mathbf{A}\mathbf{U}_1\mathbf{U}_2 = \mathbf{U}_2^{-1}\mathbf{A}'\mathbf{U}_2 = \mathbf{U}_2^{-1}\begin{bmatrix}\lambda_1 & \mathbf{w}^T\\ \mathbf{0} & \mathbf{C}\end{bmatrix}\mathbf{U}_2\\ &= \begin{bmatrix}1 & \mathbf{0}^T\\ \mathbf{0} & \mathbf{V}^H\end{bmatrix}\begin{bmatrix}\lambda_1 & \mathbf{w}^T\\ \mathbf{0} & \mathbf{C}\end{bmatrix}\begin{bmatrix}1 & \mathbf{0}^T\\ \mathbf{0} & \mathbf{V}\end{bmatrix} = \begin{bmatrix}\lambda_1 & \mathbf{w}^T\mathbf{V}\\ \mathbf{0} & \mathbf{V}^H\mathbf{C}\mathbf{V}\end{bmatrix} = \begin{bmatrix}\lambda_1 & \mathbf{w}^T\mathbf{V}\mathbf{0} & \mathbf{T}_1\end{bmatrix}\end{aligned}$$

è triangolare superiore essendolo $\mathbf{T}_1$. Ciò completa la dimostrazione. □

Osserviamo che $\mathbf{A}$ è anche unitariamente simile a una matrice triangolare inferiore. Infatti, per quanto appena dimostrato, esistono una matrice triangolare superiore $\mathbf{T}_1$ e una matrice unitaria $\mathbf{U}$ tali che $\mathbf{A}^H = \mathbf{U}\mathbf{T}_1\mathbf{U}^H$. Passando alla H-trasposta, dall'uguaglianza precedente si ottiene $\mathbf{A} = \mathbf{U}\mathbf{T}_1^H\mathbf{U}^H$ con $\mathbf{T}_1^H$ triangolare inferiore.

Dal teorema precedente segue un corollario.

Corollario 4.10. *Ogni matrice complessa è unitariamente simile a una matrice triangolare.* □

Data una matrice $\mathbf{A}$ con polinomio caratteristico che si fattorizza in fattori lineari, la dimostrazione del teorema di Schur fornisce un procedimento effettivo per costruire una matrice triangolare $\mathbf{T}$ e una matrice unitaria $\mathbf{U}$ tali che $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^H$.

Dal punto di vista computazionale, però, tale costruzione dipende dalla possibilità di calcolare gli autovalori di $\mathbf{A}$, e tale calcolo, come abbiamo già sottolineato, può essere estremamente problematico, se non impossibile.

Ma dal punto di vista teorico il teorema di Schur ha applicazioni notevoli, permettendo di ricondursi al caso di matrici triangolari ogniqualvolta si vogliano studiare delle proprietà delle matrici che siano invarianti rispetto alla relazione di similitudine.

Il corollario che segue è un esempio di applicazione teorica del teorema di Schur.

Corollario 4.11. *Sia $\mathbf{A}$ una matrice $n \times n$ con n autovalori (ripetuti con le loro molteplicità) $\lambda_1, \dots, \lambda_n$. Per ogni intero positivo k gli autovalori di $\mathbf{A}^k$ sono i numeri $\lambda_1^k, \dots, \lambda_n^k$.*

Dimostrazione. Per il teorema di Schur esistono una matrice triangolare $\mathbf{T}$ e una matrice unitaria $\mathbf{U}$ tali che $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^{-1}$, per cui si ha anche $\mathbf{A}^k = \mathbf{U}\mathbf{T}^k\mathbf{U}^{-1}$. Dalla Proposizione 3.2 segue allora che gli autovalori di $\mathbf{A}$ coincidono con quelli di $\mathbf{T}$, e gli autovalori di $\mathbf{A}^k$ coincidono con quelli di $\mathbf{T}^k$. Poiché anche $\mathbf{T}^k$ è triangolare, essendolo $\mathbf{T}$, e gli autovalori di una matrice triangolare sono i suoi coefficienti diagonali, allora $\lambda_1, \dots, \lambda_n$ sono i coefficienti diagonali di $\mathbf{T}$, e gli autovalori di $\mathbf{A}^k$ sono i coefficienti diagonali di $\mathbf{T}^k$. La conclusione segue quindi dal fatto che i coefficienti diagonali di $\mathbf{T}^k$ sono le potenze k-esime dei coefficienti diagonali di $\mathbf{T}$, essendo $\mathbf{T}$ triangolare. □

5. I teoremi di Hamilton-Cayley e di Gerschgorin

Dei molti risultati sul polinomio caratteristico di una matrice quadrata e sui suoi autovalori, due sono di basilare importanza: il Teorema di Hamilton-Cayley e il Teorema "dei cerchi" di Gerschgorin.

Il primo teorema venne dimostrato per la prima volta per una particolare classe di matrici nel 1853 da W. R. Hamilton, matematico di Dublino famoso anche per la scoperta dei quaternioni (numeri che generalizzano i numeri complessi) e per i suoi studi in dinamica fisica. Il matematico di Cambridge A. Cayley dimostrò il teorema in tutta generalità cinque anni dopo.

Del Teorema di Hamilton-Cayley esistono svariate dimostrazioni, la cui difficoltà decresce con l'aumentare della portata degli strumenti teorici a disposizione. La dimostrazione che viene qui presentata usa il Teorema di Schur 4.9; una dimostrazione più semplice fa uso della forma di Jordan (si veda l'Esercizio ??).

Teorema 5.1 (Hamilton-Cayley). *Sia* $\mathbf{A}$ *una matrice* $n \times n$ *con polinomio caratteristico* $p_{\mathbf{A}}(X) = (-1)^n X^n + a_{n-1}X^{n-1} + \dots + a_1 X + a_0$. *Allora* $p_{\mathbf{A}}(\mathbf{A}) = \mathbf{O}$, *cioè*

$$(-1)^n \mathbf{A}^n + a_{n-1}\mathbf{A}^{n-1} + \dots + a_1\mathbf{A} + a_0\mathbf{I}$$

è la matrice nulla.

Dimostrazione. Per il Teorema di Schur, la matrice $\mathbf{A}$ è unitariamente simile a una matrice triangolare superiore $\mathbf{T} = [t_{ij}]$:

$$\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^{-1},$$

dove $\mathbf{U}$ è una matrice unitaria ($\mathbf{U}^{-1} = \mathbf{U}^H$). Per la Proposizione 3.2, $p_{\mathbf{A}}(X) = p_{\mathbf{T}}(X)$, perciò basta provare che $p_{\mathbf{T}}(\mathbf{A}) = \mathbf{O}$. Osserviamo che:

$$\begin{aligned} p_{\mathbf{A}}(\mathbf{A}) &= (-1)^n \mathbf{A}^n + a_{n-1}\mathbf{A}^{n-1} + \dots + a_1\mathbf{A} + a_0\mathbf{I} \\ &= (-1)^n (\mathbf{U}\mathbf{T}\mathbf{U}^{-1})^n + a_{n-1}(\mathbf{U}\mathbf{T}\mathbf{U}^{-1})^{n-1} + \dots + a_1(\mathbf{U}\mathbf{T}\mathbf{U}^{-1}) + a_0\mathbf{I} \\ &= (-1)^n \mathbf{U}\mathbf{T}^n\mathbf{U}^{-1} + a_{n-1}\mathbf{U}^{n-1}\mathbf{T}\mathbf{U}^{-1} + \dots + a_1(\mathbf{U}\mathbf{T}\mathbf{U}^{-1}) + a_0\mathbf{U}\mathbf{U}^{-1} \\ &= (-1)^n \mathbf{U}(\mathbf{T}^n + a_{n-1}\mathbf{T}^{n-1} + \dots + a_1\mathbf{T} + a_0\mathbf{I})\mathbf{U}^{-1}, \end{aligned}$$

quindi $p_{\mathbf{T}}(\mathbf{A}) = \mathbf{U}p_{\mathbf{T}}(\mathbf{T})\mathbf{U}^{-1}$; pertanto è sufficiente provare che $p_{\mathbf{T}}(\mathbf{T}) = \mathbf{O}$, ovvero, basta provare il Teorema per la matrice triangolare $\mathbf{T}$. Essendo gli autovalori di $\mathbf{T}$ i suoi elementi diagonali, si ha:

$$p_{\mathbf{T}}(X) = (t_{11} - X)(t_{22} - X)\dots(t_{nn} - X),$$

quindi siamo ricondotti a provare l'uguaglianza:

$$(\mathbf{T} - t_{11}\mathbf{I})(\mathbf{T} - t_{22}\mathbf{I})\dots(\mathbf{T} - t_{nn}\mathbf{I}) = \mathbf{O}.$$

A tal fine, dimostriamo che la generica colonna j-esima della matrice $(\mathbf{T}-t_{11}\mathbf{I})\dots(\mathbf{T}-t_{nn}\mathbf{I})$ coincide con il vettore nullo. Tenuto conto del fatto che gli n fattori di tale matrice commutano tra di loro, basta provare che, per ogni $j \geq n$, si ha:

$$(\mathbf{T} - t_{11}\mathbf{I})(\mathbf{T} - t_{22}\mathbf{I})\dots(\mathbf{T} - t_{jj}\mathbf{I})\mathbf{e}_j = \mathbf{0} \tag{1}$$

Per $j = 1$ l'uguaglianza in (1) è banale, perché la prima colonna della matrice $\mathbf{T} - t_{11}\mathbf{I}$ è ovviamente nulla. Sia allora (1) vera per $j = 1, 2, \dots, k-1$, e proviamo che è vera per k. Si ha infatti:

$$(\mathbf{T} - t_{kk}\mathbf{I})\mathbf{e}_k = \mathbf{T}\mathbf{e}_k - t_{kk}\mathbf{e}_k = t_{1k}\mathbf{e}_1 + t_{2k}\mathbf{e}_2 + \dots + t_{k-1,k}\mathbf{e}_{k-1}.$$

Da ciò si ricava, usando il fatto che $(\mathbf{T} - t_{ii}\mathbf{I})(\mathbf{T} - t_{jj}\mathbf{I}) = (\mathbf{T} - t_{jj}\mathbf{I})(\mathbf{T} - t_{ii}\mathbf{I})$ per ogni i e j:

$$(\mathbf{T}-t_{11}\mathbf{I})\dots(\mathbf{T}-t_{k-1,k-1}\mathbf{I})(\mathbf{T}-t_{kk}\mathbf{I})\mathbf{e}_k = (\mathbf{T}-t_{11}\mathbf{I})\dots(\mathbf{T}-t_{k-1,k-1}\mathbf{I})(t_{1k}\mathbf{e}_1+t_{2k}\mathbf{e}_2+\dots+t_{k-1,k}\mathbf{e}_{k-1}) = \mathbf{0},$$

giacché $(\mathbf{T} - t_{11}\mathbf{I})\dots(\mathbf{T} - t_{jj}\mathbf{I})\mathbf{e}_j = \mathbf{0}$ per $j = 1, 2, \dots, k-1$. □

Una immediata applicazione del Teorema di Hamilton-Cayley si dà al calcolo delle potenze di una matrice quadrata e al calcolo della sua inversa, nel caso esista (si vedano gli Esercizi ??).

Il secondo Teorema di questo paragrafo, dimostrato circa 80 anni dopo di quello di Hamilton-Cayley, è dovuto al matematico russo Gerschgorin (il cui nome trascritto dal cirillico ha spesso forme diverse). Si chiama anche Teorema "dei cerchi", perché localizza gli autovalori di una matrice $n \times n$ in una regione del piano complesso racchiusa in n cerchi, con centri e raggi immediatamente deducibili dai coefficienti della matrice. Questo fatto costituisce la prima parte del Teorema; la seconda parte dice qualcosa di più preciso nel caso in cui i cerchi siano disposti in modo opportuno, e ha una dimostrazione più complessa, di cui si dà traccia negli Esercizi ??.

Premettiamo una definizione e una notazione.

Data la matrice $\mathbf{A} = [a_{ij}] \in \mathrm{M}_n(\mathbb{C})$, si chiama *$i$-esimo cerchio di Gerschgorin* di $\mathbf{A}$ la regione del piano complesso:

$$C_i(\mathbf{A}) = \{z \in \mathbb{C} : |z - a_{ii}| \leq R_i\},$$

dove

$$R_i = \sum_{j \neq i} |a_{ij}|.$$

Pertanto $C_i(\mathbf{A})$ è costituito esattamente dai numeri complessi che stanno nel cerchio che ha centro nell'i-esimo elemento diagonale di $\mathbf{A}$, e ha come raggio R_i la somma dei moduli degli elementi nella i-esima riga di $\mathbf{A}$ diversi da a_{ii}.

TEOREMA 5.2 (Gerschgorin). *Sia $\mathbf{A} \in \mathrm{M}_n(\mathbb{C})$. Allora:*

(i) *ogni autovalore di $\mathbf{A}$ appartiene ad almeno un cerchio di Gerschgorin,*

(ii) *se l'unione di $k \leq n$ cerchi di Gerschgorin forma una regione connessa e disgiunta dall'unione dei restanti $n - k$ cerchi, essa contiene esattamente k autovalori di $\mathbf{A}$ (contati con la loro molteplicità algebrica).*

DIMOSTRAZIONE. (i) Sia $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ per un vettore $\mathbf{v} \neq \mathbf{0}$. Se $\mathbf{v} = [v_1\ v_2\ \dots\ v_n]^T$, per ogni $i \leq n$ risulta:

$$\sum_j a_{ij} v_j = \lambda v_i$$

da cui si ricava

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

e, passando ai moduli,

$$|\lambda - a_{ii}||v_i| \leq \sum_{j \neq i} |a_{ij}||v_j|. \tag{2}$$

Si scelga una coordinata v_k di $\mathbf{v}$ di modulo massimo, tale cioè che $|v_k| \geq |v_j|$ per ogni j. Risulta $|v_k| > 0$, perché $\mathbf{v} \neq \mathbf{0}$. Dividendo ambo i membri della disuguaglianza in (2) dove $i = k$, si ha:

$$|\lambda - a_{kk}| \leq \sum_{j \neq k} |a_{kj}||\frac{v_j}{v_k}| \leq \sum_{j \neq k} |a_{kj}| = R_k,$$

il che dice appunto che $\lambda \in C_k(\mathbf{A})$.

(ii) Per una traccia di dimostrazione si vedano gli Esercizi ??. □

Una utile applicazione del Teorema 5.2 (i) si ha per le *matrici strettamente diagonalmente dominanti*, che sono quelle matrici $\mathbf{A} = [a_{ij}] \in \mathrm{M}_n(\mathbb{C})$ tali che, per ogni $i \leq n$, risulta:

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

Corollario 5.3. *Una matrice* $\mathbf{A} \in \mathrm{M}_n(\mathbb{C})$ *strettamente diagonalmente dominante è invertibile.*

Dimostrazione. Nessuno dei cerchi di Gerschgorin di $\mathbf{A}$ contiene l'origine del piano complesso, perché i centri a_{ii} hanno distanza dall'origine maggiore del raggio R_i. Quindi 0 non è un autovalore di $\mathbf{A}$, cioè l'equazione $\mathbf{Ax} = \mathbf{0}$ ha solo la soluzione nulla. Per il Teorema **??** del Capitolo I, $\mathbf{A}$ risulta invertibile. □

Esercizi

Paragrafo 2

5.1. Sia $\mathbf{A}$ una matrice $n \times n$ a coefficienti in $K \in \{\mathbb{R}, \mathbb{C}\}$. Si provi che la somma e l'intersezione di sottospazi $\mathbf{A}$-invarianti di K^n sono sottospazi $\mathbf{A}$-invarianti di K^n.

5.2. Sia $\mathbf{T}$ una matrice triangolare. Si provi che gli autovalori di $\mathbf{T}$ sono i suoi elementi diagonali.

5.3. Siano $\mathbf{A}$ una matrice non singolare e λ un autovalore di $\mathbf{A}$. Si provi che λ^{-1} è un autovalore di $\mathbf{A}^{-1}$ e che $E_{\mathbf{A}}(\lambda) = E_{\mathbf{A}^{-1}}(\lambda^{-1})$.

5.4. Siano $\mathbf{A}$ una matrice $n \times n$ e λ un autovalore di $\mathbf{A}$. Per ogni $k > 1$, si provi che λ^k è un autovalore di $\mathbf{A}^k$ e che $E_{\mathbf{A}}(\lambda) \subseteq E_{\mathbf{A}^k}(\lambda^k)$.

5.5. Siano $\mathbf{A}$ e $\mathbf{B}$ due matrici $n \times n$. Si provi che se esistono λ e μ autovalori di $\mathbf{A}$ e $\mathbf{B}$ rispettivamente con $E_{\mathbf{A}}(\lambda) \cap E_{\mathbf{B}}(\mu) \neq \{\mathbf{0}\}$, allora $\mathbf{AB} - \mathbf{BA}$ è singolare.

Paragrafo 3

5.6. Sia $\mathbf{A}$ una matrice nilpotente, tale cioè che $\mathbf{A}^k = \mathbf{O}$ per un $k > 0$. Si provi che l'unico autovalore di $\mathbf{A}$ è lo 0.

5.7. Sia $\mathbf{A}$ una matrice idempotente, tale cioè che $\mathbf{A}^{k+1} = \mathbf{A}^k$ per un $k > 0$. Si provi che se $\mathbf{A} \neq \mathbf{O}$ e $\mathbf{A} \neq \mathbf{I}$ allora lo spettro di $\mathbf{A}$ è $\{0, 1\}$.

5.8. Si provi l'Osservazione 2.2.

5.9. Si provi che se λ è un autovalore della matrice quadrata $\mathbf{A}$, allora $\overline{\lambda}$ è un autovalore della coniugata $\overline{\mathbf{A}}$ di $\mathbf{A}$. Si deduca che se $\mathbf{A}$ è reale e λ è un suo autovalore complesso, allora anche $\overline{\lambda}$ è un autovalore di $\mathbf{A}$.

5.10. Si deduca dall'esercizio precedente che una matrice quadrata reale di ordine dispari ha sempre un autovalore reale.

Paragrafo 4

5.11. Si provi che la relazione di similitudine è riflessiva, simmetrica e transitiva.

5.12. Siano $\mathbf{A}$ e $\mathbf{B}$ due matrici simili ed $\mathbf{S}$ tale che $\mathbf{A} = \mathbf{SBS}^{-1}$. Si supponga che $\mathbf{A}$ e $\mathbf{B}$ abbiano un autovalore λ in comune e, relativamente a λ, un autovettore

$\mathbf{v}$ in comune, ossia che esistano uno scalare λ ed un vettore $\mathbf{v}$ tali che $\mathbf{0} \neq \mathbf{v} \in E_{\mathbf{A}}(\lambda) \cap E_{\mathbf{B}}(\lambda)$.

Si provi che sussiste uno dei seguenti fatti:

(a) $\langle \mathbf{v} \rangle \neq E_{\mathbf{A}}(\lambda)$;
(b) $\mathbf{v}$ è un autovettore di $\mathbf{S}$.

5.13. Siano $\mathbf{A}$ e $\mathbf{B}$ due matrici simili e λ un autovalore della trasposta $\mathbf{A}^T$ di $\mathbf{A}$. Si provi che λ è un autovalore della trasposta $\mathbf{B}^T$ di $\mathbf{B}$ e che gli autospazi $E_{\mathbf{A}^T}(\lambda)$ e $E_{\mathbf{B}^T}(\lambda)$ hanno la stessa dimensione.

5.14. Si trovi una matrice 2×2 con un autovalore di molteplicità geometrica 1 e molteplicità algebrica 2.

5.15. Siano λ e μ due autovalori distinti della matrice quadrata $\mathbf{A}$. Si provi che se $\mathbf{v}$ e $\mathbf{w}$ sono due autovettori di $\mathbf{A}$ relativi a λ e μ rispettivamente, allora $\mathbf{v}+\mathbf{w}$ non è un autovettore di $\mathbf{A}$.

5.16. Si provi che una matrice complessa $n \times n$ $\mathbf{A}$ è simile ad una matrice triangolare superiore se e solo se esiste una successione crescente di sottospazi $\mathbf{A}$-invarianti di $\mathbb{C}^n$, $\{\mathbf{0}\} \subset V_1 \subset V_2 \subset \cdots \subset V_{n-1} \subset V_n = \mathbb{C}^n$, dove V_i ha dimensione i per ogni $i = 1, 2, \dots, n$.

Paragrafo 5

5.17. Si provi che se $\mathbf{A}$ è una matrice diagonalizzabile, allora anche la trasposta $\mathbf{A}^T$ di $\mathbf{A}$ è diagonalizzabile.

5.18. Si trovino due matrici diagonalizzabili 2×2 $\mathbf{A}$ e $\mathbf{B}$ tali che $\mathbf{AB}$ non è diagonalizzabile.

5.19. Siano $\mathbf{A}$ una matrice complessa $n \times n$ ed $\alpha \in \mathbb{C}$ tali che $(\mathbf{A} - \alpha\mathbf{I})^m = \mathbf{O}$, per un $m > 0$. Sia $\mathbf{A} = \mathbf{UTU}^H$ una decomposizione di Schur di $\mathbf{A}$. Si provi che $(\mathbf{T} - \alpha\mathbf{I})^m = \mathbf{O}$. Si deduca che, se λ è un autovalore di $\mathbf{A}$, allora $\lambda = \alpha$.

Paragrafo 6

5.20. Si provi il Teorema di Hamilton-Cayley per una matrice diagonale.

5.21. Sia $\mathbf{A} = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ una generica matrice 2×2. Si calcoli espressamente la matrice

$$\mathbf{A}^2 + (a+d)\mathbf{A} + (ad - bc)\mathbf{I}_2.$$

5.22. Si provi il Teorema di Hamilton-Cayley per una matrice $\mathbf{J}$ che è un λ-blocco di Jordan, cioè con elementi diagonali uguali a λ, elementi di posto $(i, i+1)$ uguali ad 1, e nulla altrove.

5.23. Sia $\mathbf{A} \in \mathrm{M}_n(\mathbb{C})$ una matrice invertibile. Si provi che $\mathbf{A}^{-1}$ si scrive come polinomio in $\mathbf{A}$ di grado $n-1$.

5.24. Sia $\mathbf{A} \in \mathrm{M}_n(\mathbb{C})$. Si provi che per ogni intero $k \geq 1$, $\mathbf{A}^k$ si scrive come polinomio in $\mathbf{A}$ di grado minore od uguale a $n-1$.

5.25. Sia $\mathbf{A} = [a_{ij}] \in \mathrm{M}_n(\mathbb{C})$ e si ponga $\mathbf{A} = \mathbf{D} + \mathbf{B}$, dove $\mathbf{D} = \mathbf{Diag}(a_{11}, \dots, a_{nn})$. Per ogni numero reale ϵ nell'intervallo $[0, 1]$, sia $\mathbf{A}_\epsilon = \mathbf{D} + \mathbf{B}_\epsilon$; si provi che l'$i$-esimo cerchio di Gerschgorin di $\mathbf{A}_\epsilon$ è contenuto nell'i-esimo cerchio di $\mathbf{A}$.

5.26. Usando l'Esercizio 5.25 e il fatto che gli autovalori di una matrice sono funzioni continue dei coefficienti, si provi che ogni autovalore di $\mathbf{A}$ è l'estremo di una curva continua avente l'altro estremo in un elemento diagonale a_{ii} e contenuta nell'unione dei cerchi di Gerschgorin di $\mathbf{A}$.

5.27. Si usi l'Esercizio 5.26 per provare che l'unione dei k cerchi considerati nel Teorema 5.2 (ii) contiene al più k autovalori di $\mathbf{A}$. Si ragioni sui cerchi rimanenti per provare che la suddetta unione dei k cerchi contiene esattamente k autovalori.

5.28. Sia $\mathbf{A} \in \mathrm{M}_n(\mathbb{R})$ una matrice reale che ha i cerchi di Gerschgorin a due a due disgiunti. Si provi che $\mathbf{A}$ ha n autovalori distinti reali (si usino il Teorema 5.2 (ii) e il fatto che, se λ è un autovalore di $\mathbf{A}$, anche $\overline{\lambda}$ lo è).