Capítulo 10
Tópicos de Álgebra Linear. I

Conteúdo

O principal objetivo deste capítulo é apresentar a demonstração do Teorema Espectral para matrizes diagonalizáveis, em particular, para matrizes autoadjuntas (resultado de grande relevância para a Mecânica Quântica) e a demonstração do Teorema de Decomposição de Jordan. Salvo menção contrária, sempre trabalharemos no contexto de espaços vetoriais de dimensão finita $\mathbb{C}^n$ sobre o corpo dos complexos. A leitura deste capítulo pressupõe que alguns conceitos básicos de Álgebra Linear, tais como o conceito de operador linear, de matriz, de produto de matrizes, de determinante de uma matriz, suas propriedades e métodos de cálculo, sejam familiares ao leitor, mas uma breve revisão é apresentada na Seção 10.1. Na Seção 10.2, página 472, apresentamos a noção de espectro e a de polinômio característico de uma matriz. Na Seção 10.5, página 506, introduzimos as noções de matrizes autoadjuntas, normais e unitárias, noções de importância, por exemplo, na Mecânica Quântica. Na Seção 10.8, página 540, apresentamos algumas representações de matrizes de interesse em diversos contextos (por exemplo, na teoria de grupos). Na Seção 10.9, página 562, estudamos a chamada pseudoinversa de Moore-Penrose, de interesse, por exemplo, em problemas de optimização linear.

Este capítulo será continuado no Capítulo 11, página 584, onde outros aspectos de álgebras de matrizes serão explorados.

Há uma vastíssima literatura sobre Álgebra Linear e sobre a teoria dos espaços vetoriais de dimensão finita, com abordagens e objetivos diferentes, e onde partes do material deste e do Capítulo 11 podem ser encontradas. Uma lista limitada de leituras complementares recomendadas incluiria [157], [257] , [349], [124].

10.1 Propriedades Básicas de Determinantes e Inversas de Matrizes

A presente seção desenvolve a teoria básica de inversas e determinantes de matrizes. Sua leitura pode, provavelmente, ser dispensada por aqueles que julgam dispor desses conhecimentos básicos, mas a notação que aqui introduzimos será empregada alhures. Propriedades mais avançadas de determinantes serão estudadas na Seção 10.11, página 575. Na Seção 11.7, página 623, com base no Teorema de Hadamard, Teorema 10.40, página 576, demonstraremos que o determinante de matrizes é uma função contínua.

• Fatos elementares sobre matrizes e alguma notação

O conjunto de todas as matrizes $m\times n$ (m linhas e n colunas) com entradas complexas será denotado por $\mathrm{Mat}(\mathbb{C},\ m,\ n)$. O conjunto de todas as matrizes quadradas $n \times n$ com entradas complexas será denotado simplesmente por $\mathrm{Mat}(\mathbb{C},\ n)$. Uma matriz $A \in \mathrm{Mat}(\mathbb{C},\ m,\ n)$ é frequentemente representada na forma de um arranjo como

$$A = \begin{pmatrix} A_{11} & \ldots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \ldots & A_{mn} \end{pmatrix} .$$

Uma matriz de $\mathrm{Mat}(\mathbb{C},\ m,\ n)$ possui, portanto, m linhas e n colunas e é também dita ser uma matriz $m \times n$.

$\mathrm{Mat}(\mathbb{C},\ m,\ n)$ é um espaço vetorial complexo, com a operação de soma definida por

$$(A_1 + A_2)_{ij} := (A_1)_{ij} + (A_2)_{ij} ,$$

$A_1,\ A_2 \in \mathrm{Mat}(\mathbb{C},\ m,\ n)$, $i \in \{1,\ \ldots,\ m\}$, $j \in \{1,\ \ldots,\ n\}$, e a operação de multiplicação por escalares (complexos) definida por

$$(\alpha A)_{ij} := \alpha A_{ij}$$

$\alpha \in \mathbb{C}$, $A \in \mathrm{Mat}(\mathbb{C},\ m,\ n)$ e $i \in \{1,\ \ldots,\ m\}$, $j \in \{1,\ \ldots,\ n\}$.

Sejam $m, n, p \in \mathbb{N}$ e sejam $A \in \text{Mat}(\mathbb{C}, m, n)$ e $B \in \text{Mat}(\mathbb{C}, n, p)$. Denotamos por AB a matriz de $\text{Mat}(\mathbb{C}, m, p)$ cujos elementos são dados por

$$\big(AB\big)_{ij} := \sum_{k=1}^{n} A_{ik}B_{kj} \tag{10.1}$$

para todos $i \in \{1, \ldots, m\}$, $j \in \{1, \ldots, p\}$. A expressão (10.1) é denominada *regra de produto de matrizes*. Tem-se, assim, que o produto de uma matriz $m \times n$ por uma matriz $n \times p$ resulta em uma matriz $m \times p$.

É fácil constatar (faça-o!) que valem as propriedades distributivas

$$(\alpha_1 A_1 + \alpha_2 A_2)B = \alpha_1 A_1 B + \alpha_2 A_2 B ,$$

$$A(\beta_1 B_1 + \beta_2 B_2) = \beta_1 AB_1 + \beta_2 AB_2 ,$$

para todos $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{C}$, todas $A, A_1, A_2 \in \text{Mat}(\mathbb{C}, m, n)$ e todas $B, B_1, B_2 \in \text{Mat}(\mathbb{C}, n, p)$.

É também fácil constatar (faça-o!) que se $m, n, p, q \in \mathbb{N}$ valem para todas $A \in \text{Mat}(\mathbb{C}, m, n)$, $B \in \text{Mat}(\mathbb{C}, n, p)$ e $C \in \text{Mat}(\mathbb{C}, p, q)$ a relação

$$(AB)C = A(BC) .$$

Essa importante propriedade é por vezes denominada *associatividade do produto de matrizes*.

Para cada $n \in \mathbb{N}$, e com a operação de produto definida acima, $\text{Mat}(\mathbb{C}, n)$ é uma álgebra associativa, não comutativa (exceto se $n = 1$) e unital, com a unidade sendo dada pela *matriz identidade*, que denotaremos por $\mathbb{1}$ neste texto:

$$\mathbb{1} := \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix} . \tag{10.2}$$

Note-se que $\mathbb{1}_{ij} = \delta_{ij}$, $i, j \in \{1, \ldots, n\}$.

Dada uma matriz $A \in \text{Mat}(\mathbb{C}, m, n)$ denotamos por A^T a matriz de $\text{Mat}(\mathbb{C}, n, m)$ cujos elementos são dados por $(A^T)_{ij} = A_{ji}$ para todos $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, m\}$. A matriz A^T é dita ser a *matriz transposta* de A. É evidente que $(A^T)^T = A$. Para todos $m, n, p \in \mathbb{N}$ vale, pela regra de produto de matrizes, a relação $(AB)^T = B^T A^T$ para quaisquer $A \in \text{Mat}(\mathbb{C}, m, n)$ e $B \in \text{Mat}(\mathbb{C}, n, p)$.

Dado um conjunto de n números complexos $\alpha_1, \ldots, \alpha_n$, denotaremos por $\text{diag}(\alpha_1, \ldots, \alpha_n)$ a matriz $A \in \text{Mat}(\mathbb{C}, n)$ cujos elementos A_{ij} são definidos da seguinte forma:

$$A_{ij} = \begin{cases} \alpha_i, & \text{se } i = j \\ 0, & \text{se } i \neq j \end{cases} .$$

Uma tal matriz é dita ser *diagonal* pois apenas os elementos de sua diagonal principal são eventualmente não nulos. Na representação usual

$$A = \begin{pmatrix} \alpha_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \alpha_n \end{pmatrix} .$$

A mais popular dentre as matrizes diagonais é a matriz identidade (10.2): $\mathbb{1} = \text{diag}(1, \ldots, 1)$.

Denotaremos por $\mathbb{0}_{a,\,b} \in \text{Mat}(\mathbb{C}, m, n)$ a matriz $a \times b$ cujos elementos de matriz são todos nulos. Denotaremos por $\mathbb{1}_l \in \text{Mat}(\mathbb{C}, l)$ a matriz identidade $l \times l$. Por vezes, quando não houver perigo de confusão, poderemos omitir os subíndices e escrever $\mathbb{0}_{a,\,b}$ simplesmente como $\mathbb{0}$ e $\mathbb{1}_l$ simplesmente como $\mathbb{1}$.

Vamos também empregar as seguintes definições. Para $m, n \in \mathbb{N}$, sejam $I_{m,\, m+n} \in \mathrm{Mat}(\mathbb{C},\, m,\, m+n)$ e $J_{m+n,\, n} \in \mathrm{Mat}(\mathbb{C},\, m+n,\, n)$ dadas por

$$I_{m,\, m+n} := \begin{pmatrix} \mathbb{1}_m & \mathbb{0}_{m,\, n} \end{pmatrix} \qquad \text{e} \qquad J_{m+n,\, n} := \begin{pmatrix} \mathbb{1}_n \\ \\ \mathbb{0}_{m,\, n} \end{pmatrix}, \tag{10.3}$$

cujas transpostas são dadas por

$$(I_{m,\, m+n})^T := \begin{pmatrix} \mathbb{1}_m \\ \\ \mathbb{0}_{n,\, m} \end{pmatrix} = J_{m+n,\, m} \qquad \text{e} \qquad (J_{m+n,\, n})^T := \begin{pmatrix} \mathbb{1}_n & \mathbb{0}_{n,\, m} \end{pmatrix} = I_{n,\, m+n} . \tag{10.4}$$

As seguintes identidades úteis serão usadas mais adiante e sua demonstração (fácil) é deixada como exercício ao leitor:

$$I_{m,\, m+n}\,(I_{m,\, m+n})^T = I_{m,\, m+n}\, J_{m+n,\, m} = \mathbb{1}_m , \tag{10.5}$$

$$(J_{m+n,\, n})^T J_{m+n,\, n} = I_{n,\, m+n} J_{m+n,\, n} = \mathbb{1}_n . \tag{10.6}$$

Para cada $A \in \mathrm{Mat}(\mathbb{C},\, m,\, n)$ podemos associar uma matriz quadrada $A' \in \mathrm{Mat}(\mathbb{C},\, m+n)$ dada por

$$A' := J_{m+n,\, m} A I_{n,\, m+n} = \begin{pmatrix} A & \mathbb{0}_{m,\, m} \\ \\ \mathbb{0}_{n,\, n} & \mathbb{0}_{n,\, m} \end{pmatrix} . \tag{10.7}$$

Obtemos das relações (10.5)–(10.6) que

$$A = I_{m,\, m+n} A' J_{m+n,\, n} . \tag{10.8}$$

Sejam $x^1, \ldots, x^n$ vetores, representados na base canônica por vetores-coluna

$$x^a = \begin{pmatrix} x_1^a \\ \vdots \\ x_n^a \end{pmatrix} .$$

Denotaremos por $\left[\!\left[x^1, \ldots, x^n \right]\!\right]$ a matriz $n \times n$ construída de forma que sua a-ésima coluna seja o vetor-coluna x^a, ou seja

$$\left[\!\left[x^1, \ldots, x^n \right]\!\right] = \begin{pmatrix} x_1^1 & \cdots & x_1^n \\ \vdots & \ddots & \vdots \\ x_n^1 & \cdots & x_n^n \end{pmatrix} . \tag{10.9}$$

Considerando os vetores da base canônica

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad \ldots, \qquad \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \tag{10.10}$$

é também evidente que

$$\mathbb{1} = \left[\!\!\left[\mathbf{e}_1, \ldots, \mathbf{e}_n\right]\!\!\right] . \tag{10.11}$$

A notação acima é útil por permitir a seguinte observação. Seja B uma matriz qualquer. Então,

$$B\left[\!\!\left[x^1, \ldots, x^n\right]\!\!\right] = \left[\!\!\left[Bx^1, \ldots, Bx^n\right]\!\!\right] . \tag{10.12}$$

Essa relação é provada observando-se a regra de multiplicação de matrizes: a a-ésima coluna de $B\left[\!\!\left[x^1, \ldots, x^n\right]\!\!\right]$ é

$$\begin{matrix} B_{11}x_1^a + \cdots + B_{1n}x_n^a \\ \vdots \\ B_{n1}x_1^a + \cdots + B_{nn}x_n^a \end{matrix} , \tag{10.13}$$

que vem a ser as componentes de Bx^a, representado como vetor-coluna na base canônica.

É útil observar que se A é uma matriz $n \times n$ temos a regra

$$A\mathbf{e}_i = \sum_{j=1}^{n} A_{ji}\, \mathbf{e}_j , \tag{10.14}$$

onde A_{ji} são os elementos de matriz de A respectivas na base canônica. Verifique!

Ainda sobre essa notação, vale a seguinte identidade útil, cuja demonstração (elementar) deixamos como exercício: se $D = \operatorname{diag}(d_1, \ldots, d_n)$ é uma matriz diagonal, então

$$\left[\!\!\left[x^1, \ldots, x^n\right]\!\!\right] D = \left[\!\!\left[d_1x^1, \ldots, d_nx^n\right]\!\!\right] . \tag{10.15}$$

Seja V um espaço vetorial dotado de um produto escalar $\langle \cdot, \cdot \rangle$. Dizemos que dois vetores u e v são perpendiculares (em relação ao produto escalar $\langle \cdot, \cdot \rangle$) se $\langle u, v \rangle = 0$.

Se $v_1, \ldots, v_k$ são vetores em um espaço vetorial V, denotamos por $[v_1, \ldots, v_k]$ o subespaço gerado pelos vetores $v_1, \ldots, v_k$, ou seja, a coleção de todos os vetores que são combinações lineares dos vetores $v_1, \ldots, v_k$:

$$[v_1, \ldots, v_k] = \Big\{\alpha_1v_1 + \cdots + \alpha_kv_k, \;\; \alpha_1, \ldots, \alpha_k \in \mathbb{C}\Big\} .$$

Denotamos por $[v_1, \ldots, v_k]^\perp$ o subespaço de todos os vetores perpendiculares a todos os vetores de $[v_1, \ldots, v_k]$:

$$[v_1, \ldots, v_k]^\perp = \Big\{w \in V \Big| \; \big\langle w, (\alpha_1v_1 + \cdots + \alpha_kv_k)\big\rangle = 0 \text{ para todos } \alpha_1, \ldots, \alpha_k \in \mathbb{C}\Big\} .$$

• Matrizes bijetoras e a noção de inversa de uma matriz

Uma matriz $A \in \operatorname{Mat}(\mathbb{C}, n)$ define uma aplicação linear de $\mathbb{C}^n$ sobre si mesmo. Se essa aplicação for bijetora, então existe uma aplicação inversa, denotada por $A^{-1} : \mathbb{C}^n \to \mathbb{C}^n$, tal que $A^{-1}\big(Ax\big) = x$ para todo $x \in \mathbb{C}^n$. A proposição seguinte reúne fatos elementares sobre a aplicação inversa A^{-1}:

Proposição 10.1 *Se $A \in \operatorname{Mat}(\mathbb{C}, n)$ é bijetora, então A^{-1} é igualmente uma aplicação linear de $\mathbb{C}^n$ sobre si mesmo, ou seja, $A^{-1} \in \operatorname{Mat}(\mathbb{C}, n)$. Fora isso, A^{-1} é única e $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$. Por fim, vale afirmar que A é inversível se e somente se A^T o for.* □

Prova. É fácil constatar que A^{-1} é também uma aplicação linear e, portanto, é também um elemento de $\operatorname{Mat}(\mathbb{C}, n)$. De fato, sejam v_1, v_2 elementos arbitrários de $\mathbb{C}^n$ e $\alpha_1, \alpha_2 \in \mathbb{C}$, igualmente arbitrários. Como A é bijetora, existem

$u_1, u_2 \in \mathbb{C}^n$, únicos, tais que $Au_1 = v_1$ e $Au_2 = v_2$, ou seja, tais que $u_1 = A^{-1}(v_1)$ e $u_2 = A^{-1}(v_2)$. Assim, usando a linearidade de A, tem-se

$$A^{-1}\big(\alpha_1 v_1 + \alpha_2 v_2\big) \;=\; A^{-1}\big(\alpha_1 A u_1 + \alpha_2 A u_2\big) \;=\; A^{-1}\Big(A\big(\alpha_1 u_1 + \alpha_2 u_2\big)\Big) \;=\; \alpha_1 u_1 + \alpha_2 u_2 \;=\; \alpha_1 A^{-1}(v_1) + \alpha_2 A^{-1}(v_2)\;,$$

o que prova que A^{-1} é também linear e, portanto $A^{-1} \in \mathrm{Mat}\,(\mathbb{C},\ n)$. Com isso, podemos afirmar que $A^{-1}Ax = x$ para todo $x \in \mathbb{C}^n$ e, portanto, $AA^{-1}Ax = Ax$. Como A é sobrejetora, isso diz-nos que $AA^{-1}y = y$ para todo $y \in \mathbb{C}^n$. Assim, estabelecemos que $A^{-1}A = AA^{-1} = \mathbb{1}$. A unicidade é facilmente estabelecida, pois se $B \in \mathrm{Mat}\,(\mathbb{C},\ n)$ é tal que $BA = AB = \mathbb{1}$, então multiplicando-se $AB = \mathbb{1}$ à esquerda por A^{-1} obtém-se $B = A^{-1}$. Por fim, observemos que do fato que $(MN)^T = N^T M^T$ para quaisquer matrizes $M,\ N \in \mathrm{Mat}\,(\mathbb{C},\ n)$, segue de $A^{-1}A = AA^{-1} = \mathbb{1}$ que $A^T\big(A^{-1}\big)^T = \big(A^{-1}\big)^T A^T = \mathbb{1}$, o que implica $\big(A^T\big)^{-1} = \big(A^{-1}\big)^T$. A última relação implica que se A é inversível, então A^T também o é. Como $(A^T)^T = A$, vale também a recíproca. ■

Mais adiante indicaremos como a matriz A^{-1} pode ser calculada a partir de A. Vide para tal a expressão (10.18) ("regra de Laplace") do Teorema 10.1, página 467, e também as expressões (10.47), página 489, e (10.211), página 576.

Em parte do que segue estaremos implicitamente usando a seguinte proposição:

Proposição 10.2 *Uma matriz $A \in \mathrm{Mat}\,(\mathbb{C},\ n)$ é bijetora (ou seja, é inversível) se e somente se $Av = 0$ valer apenas para $v = 0$.* □

Prova. Se A é bijetora, então existe A^{-1}. Logo, aplicando-se A^{-1} à esquerda na igualdade $Av = 0$, obtém-se $v = 0$. Vamos agora provar a recíproca: vamos supor que $Av = 0$ vale apenas para $v = 0$ e provar que A é injetora e sobrejetora e, portanto, bijetora.

Prova-se que A é injetora por absurdo. Se A não é injetora, então, existem vetores x e y com $x \neq y$ mas com $Ax = Ay$. Como A é linear, isso implica $A(x - y) = 0$. Pela hipótese que $Av = 0$ vale apenas para $v = 0$, segue que $x = y$, uma contradição.

Para provarmos que A é sobrejetora procedemos da seguinte forma. Seja $\{\mathbf{b}_1,\ \ldots,\ \mathbf{b}_n\}$ uma base em $\mathbb{C}^n$. Vamos primeiramente mostrar que $\{A\mathbf{b}_1,\ \ldots,\ A\mathbf{b}_n\}$ é um conjunto linearmente independente de vetores em $\mathbb{C}^n$ (e, portanto, uma base em $\mathbb{C}^n$). Suponhamos que assim não o seja e que existam números complexos $\alpha_1,\ \ldots,\ \alpha_n$, não todos nulos, tais que $\alpha_1 A\mathbf{b}_1 + \cdots + \alpha_n A\mathbf{b}_n = 0$. Pela linearidade de A, segue que $A\left(\alpha_1\mathbf{b}_1 + \cdots + \alpha_n\mathbf{b}_n\right) = 0$. Novamente, pela hipótese que $Av = 0$ vale apenas para $v = 0$, segue que $\alpha_1\mathbf{b}_1 + \cdots + \alpha_n\mathbf{b}_n = 0$. Isso, porém, diz que os vetores $\{\mathbf{b}_1,\ \ldots,\ \mathbf{b}_n\}$ são linearmente dependentes, o que é absurdo.

Logo, $\{A\mathbf{b}_1,\ \ldots,\ A\mathbf{b}_n\}$ é um conjunto de n vetores linearmente independente em $\mathbb{C}^n$ e, portanto, é uma base nesse espaço. Assim, qualquer $x \in \mathbb{C}^n$ pode ser escrito como uma combinação linear tal como $x = \beta_1 A\mathbf{b}_1 + \cdots + \beta_n A\mathbf{b}_n = A\left(\beta_1\mathbf{b}_1 + \cdots + \beta_n\mathbf{b}_n\right)$. Isso mostra que x está na imagem de A. Como x é arbitrário, segue que A é sobrejetora. ■

Um corolário evidente é o seguinte:

Corolário 10.1 *Uma matriz $A \in \mathrm{Mat}\,(\mathbb{C},\ n)$ é não bijetora (ou seja, não possui inversa) se e somente se existir um vetor não-nulo v tal que $Av = 0$.* □

O seguinte corolário indica uma maneira prática, necessária e suficiente de se constatar se uma matriz $A \in \mathrm{Mat}\,(\mathbb{C},\ n)$ tem inversa.

Corolário 10.2 *Seja $A \in \mathrm{Mat}\,(\mathbb{C},\ n)$ da forma $A = \llbracket a_1,\ \ldots,\ a_n \rrbracket$ para o conjunto de vetores $a_1,\ \ldots,\ a_n$ que representam suas colunas. Então, A é inversível se e somente se os vetores $a_1,\ \ldots,\ a_n$ forem linearmente independentes. Vale também a afirmação que A é inversível se e somente se suas linhas forem linearmente independentes.* □

Prova. Se $v \in \mathbb{C}^n$ é o vetor coluna $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$, então é fácil constatar (pela regra de produto de matrizes. Faça-o!) que $Av = v_1 a_1 + \ldots + v_n a_n$. Com isso, vemos que a afirmação que existe v não-nulo tal que $Av = 0$ equivale à afirmação que os vetores-coluna $a_1,\ \ldots,\ a_n$ são linearmente dependentes.

Como A é inversível se e somente se A^T o for (Proposição 10.1, página 465), vale afirmar que A é inversível se e somente se suas linhas forem linearmente independentes. ■

• Propriedades básicas de determinantes de matrizes

Seja $A \in \mathrm{Mat}(\mathbb{C},\ n)$ da forma $A = [\![a_1,\ \ldots,\ a_n]\!]$ para o conjunto de vetores $a_1,\ \ldots,\ a_n$ que representam suas colunas. O *determinante* de A, $\det(A)$, foi definido em (3.7), página 225, como

$$\det(A) \ := \ \omega_{det}(a_1,\ \ldots,\ a_n) \ , \tag{10.16}$$

onde ω_{det} é a forma alternante maximal em n dimensões, normalizada de sorte que $\omega_{det}(\mathbf{e}_1,\ \ldots,\ \mathbf{e}_n) = 1$. Com isso, vale $\det(\mathbb{1}) = 1$. Assim, se S_n denota o conjunto de todas as bijeções de $\{1,\ \ldots,\ n\}$ em si mesmo (o chamado grupo de permutações de n elementos, estudado na Seção 21.2, página 1033), tem-se $\omega_{det}(\mathbf{e}_{j(1)},\ \ldots,\ \mathbf{e}_{j(n)}) = \mathrm{sinal}\,(j)$ para todo $j \in S_n$ e, portanto, vale a expressão (3.8), página 225:

$$\det(A) \ = \ \sum_{j\in S_n} \mathrm{sinal}\,(j)\, A_{1j(1)} \cdots A_{nj(n)} \ , \tag{10.17}$$

frequentemente denominada *fórmula de Leibniz*[1] *para o determinante* de uma matriz. Acima, $\mathrm{sinal}\,(j)$ denota o sinal, ou paridade, da permutação $j \in S_n$, noção detalhada na Seção 21.2.1.1, página 1037, e corresponde a $(-1)^{k_j}$, onde k_j é o número de permutações de vizinhos necessárias para levar a sequência finita $j(1)j(2)j(3)\cdots j(n)$ à sequência $123\cdots n$.

O teorema a seguir reúne todas as propriedades fundamentais do determinante de matrizes.

Teorema 10.1 *Para toda matriz* $A \in \mathrm{Mat}(\mathbb{C},\ n)$ *valem:*

1. $\det(\lambda A) = \lambda^n \det(A)$ *para todo* $\lambda \in \mathbb{C}$.
2. $\det(A) = \det\left(A^T\right)$. *Consequentemente, o determinante de uma matriz troca de sinal quando da permuta de duas de suas colunas ou linhas.*
3. $\det(AB) = \det(A)\det(B) = \det(BA)$ *para qualquer* $B \in \mathrm{Mat}(\mathbb{C},\ n)$.
4. $\det(A) = \det(SAS^{-1})$ *para qualquer* $S \in \mathrm{Mat}(\mathbb{C},\ n)$*, inversível.*
5. *Se* $\det(A) = 0$*, então* A *não tem inversa.*
6. *Se* $\det(A) \neq 0$*, então* A *tem inversa e vale a chamada* regra de Laplace[2]*:*

$$A^{-1} \ = \ \frac{1}{\det(A)}\mathrm{Cof}(A)^T \ , \tag{10.18}$$

onde $\mathrm{Cof}(A) \in \mathrm{Mat}(\mathbb{C},\ n)$*, denominada* matriz dos cofatores *de* A*, é a matriz cujos elementos são*

$$\mathrm{Cof}(A)_{jk} \ = \ \omega_{det}(a_1,\ \ldots,\ a_{k-1},\ \mathbf{e}_j,\ a_{k+1},\ \ldots,\ a_n) \ = \ \det [\![a_1,\ \ldots,\ a_{k-1},\ \mathbf{e}_j,\ a_{k+1},\ \ldots,\ a_n]\!] \ . \tag{10.19}$$

Em palavras, $\mathrm{Cof}(A)_{jk}$ *é o determinante da matriz obtida substituindo a* k*-ésima coluna de* A *pelo vetor* $\mathbf{e}_j$*. No próximo item veremos outra caracterização da matriz dos cofatores* $\mathrm{Cof}(A)$.

Conjuntamente com o item 5, concluímos que A *tem inversa se e somente se* $\det(A) \neq 0$.

7. *Os elementos de matriz de* $\mathrm{Cof}(A)$ *são dados por*

$$\mathrm{Cof}(A)_{ij} \ = \ (-1)^{i+j}\mathrm{Men}(A)_{ij} \ ,$$

onde $\mathrm{Men}(A)$*, chamada de* matriz dos menores *de* A*, é a matriz de* $\mathrm{Mat}(\mathbb{C},\ n)$ *definida de sorte que cada elemento* $\mathrm{Men}(A)_{ij}$ *seja o determinante da matriz* $(n-1)\times(n-1)$ *obtida eliminando-se a* i*-ésima linha e a* j*-ésima coluna de* A*. Se* $n = 1$*, convenciona-se definir* $\mathrm{Men}(A) = 1$*. Assim, para* $\det(A) \neq 0$*, a regra de Laplace escreve-se*

$$\left(A^{-1}\right)_{ij} \ = \ \frac{1}{\det(A)}\mathrm{Cof}(A)_{ji} \ = \ \frac{(-1)^{i+j}}{\det(A)}\mathrm{Men}(A)_{ji} \ . \tag{10.20}$$

[1] Gottfried Wilhelm von Leibniz (1646–1716).
[2] Pierre-Simon Laplace (1749–1827).

8. *Para qualquer $k \in \{1, \ldots, n\}$ valem a expansão em linhas do determinante*

$$\det(A) \;=\; \sum_{j=1}^{n} A_{kj}\,\mathrm{Cof}(A)_{kj} \;=\; \sum_{j=1}^{n}(-1)^{j+k}A_{kj}\,\mathrm{Men}(A)_{kj} \tag{10.21}$$

e a expansão em colunas do determinante

$$\det(A) \;=\; \sum_{j=1}^{n} A_{jk}\,\mathrm{Cof}(A)_{jk} \;=\; \sum_{j=1}^{n}(-1)^{j+k}A_{jk}\,\mathrm{Men}(A)_{jk}\;. \tag{10.22}$$

□

Em (10.211), página 576, apresentaremos outra fórmula explícita para o cômputo da inversa de matrizes baseada no Teorema de Hamilton-Cayley (Teorema 10.4, página 486).

Demonstração do Teorema 10.1. *Prova de 1*. Pela fórmula de Leibniz (10.17),

$$\det(\lambda A) \;=\; \sum_{j\in S_n} \mathrm{sinal}\,(j)\,(\lambda A_{1j(1)})\cdots(\lambda A_{nj(n)}) \;=\; \lambda^n\det(A)\;.$$

Prova de 2. Observemos a fórmula de Leibniz (10.17). Usando o fato elementar que um produto de números complexos não depende da ordem dos fatores, podemos escrever $A_{1j(1)}\cdots A_{nj(n)} = A_{l(1)j(l(1))}\cdots A_{l(n)j(l(n))}$ para qualquer $l \in S_n$. Em particular, escolhendo $l = j^{-1}$ obtemos $A_{1j(1)}\cdots A_{nj(n)} = A_{j^{-1}(1)1}\cdots A_{j^{-1}(n)n}$. Assim, pela fórmula de Leibniz (10.17), e usando o fato que $\mathrm{sinal}\,(j) = \mathrm{sinal}\,(j^{-1})$ para todo $j \in S_n$ (justifique!), vale

$$\det(A) \;=\; \sum_{j\in S_n} A_{j^{-1}(1)1}\cdots A_{j^{-1}(n)n}\,\mathrm{sinal}\,(j^{-1}) \;=\; \sum_{j^{-1}\in S_n}\mathrm{sinal}\,(j^{-1})\,A_{j^{-1}(1)1}\cdots A_{j^{-1}(n)n}$$

$$=\; \sum_{j\in S_n}\mathrm{sinal}\,(j)\,A_{j(1)1}\cdots A_{j(n)n} \;=\; \det\left(A^T\right)\;.$$

Quando da permuta de duas linhas ou colunas de A seu determinante troca de sinal devido à alternância da forma ω_{det}. A igualdade $\det(A) = \det\left(A^T\right)$ ensina que isso também ocorre quando da permuta de linhas.

E. 10.1 *Exercício*. Justifique todas as passagens acima. ✠

Prova de 3. Sejam $A = [\![a_1,\ \ldots,\ a_n]\!]$ e $B = [\![b_1,\ \ldots,\ b_n]\!]$. Temos que $AB = [\![Ab_1,\ \ldots,\ Ab_n]\!]$ (vide (10.12)). Agora,

$$(Ab_j)_i \;=\; \sum_{k=1}^{n} A_{ik}(b_j)_k \;=\; \sum_{k=1}^{n}(a_k)_i(b_j)_k\;, \qquad \text{ou seja,} \qquad Ab_j \;=\; \sum_{k=1}^{n}(b_j)_k a_k\;.$$

Assim,

$$\begin{aligned}
\det(AB) &= \omega_{det}(Ab_1, \ldots, Ab_n) \\
&= \omega_{det}\left(\sum_{k_1=1}^{n}(b_1)_{k_1}a_{k_1}, \ldots, \sum_{k_n=1}^{n}(b_n)_{k_n}a_{k_n}\right) \\
&\overset{\text{multilinearidade}}{=} \sum_{k_1=1}^{n}\cdots\sum_{k_n=1}^{n}(b_1)_{k_1}\cdots(b_n)_{k_n}\ \omega_{det}\left(a_{k_1}, \ldots, a_{k_n}\right) \\
&= \sum_{k\in S_n}(b_1)_{k(1)}\cdots(b_n)_{k(n)}\ \omega_{det}\big(a_{k(1)}, \ldots, a_{k(n)}\big) \\
&= \sum_{k\in S_n}\mathrm{sinal}\,(k)\,(b_1)_{k(1)}\cdots(b_n)_{k(n)}\,\omega_{det}\left(a_1, \ldots, a_n\right) \\
&= \left(\sum_{k\in S_n}\mathrm{sinal}\,(k)\,(b_1)_{k(1)}\cdots(b_n)_{k(n)}\right)\det(A) \\
&= \det(B)\det(A)\ .
\end{aligned}$$

Acima, na passagem da terceira para a quarta linha usamos o fato que $\omega_{det}\left(a_{k_1}, \ldots, a_{k_n}\right)$ anula-se a menos que a $k_1, \ldots, k_n$ sejam distintos, o que somente ocorre se forem da forma $k(1), \ldots, k(n)$, respectivamente, para algum $k \in S_n$. Na passagem da quarta para a quinta linha usamos que $\omega_{det}\big(a_{k(1)}, \ldots, a_{k(n)}\big) = \mathrm{sinal}\,(k)\,\omega_{det}\left(a_1, \ldots, a_n\right)$, pois ω_{det} é uma forma alternante.

Estabelecemos, portanto, que $\det(AB) = \det(A)\det(B) = \det(BA)$.

Prova de 4. Do item 3 segue que, para quaisquer $A, S \in \mathrm{Mat}\,(\mathbb{C}, n)$, com S inversível, vale $\det(A) = \det((AS^{-1})S) = \det(SAS^{-1})$.

Prova de 5. Se $\det(A) = 0$ então A não pode ter inversa, pois se existisse A^{-1} teríamos $1 = \det(\mathbb{1}) = \det(AA^{-1}) = \det(A)\det(A^{-1}) = 0$, absurdo.

Prova de 6. É bastante claro que podemos escrever

$$a_k = \sum_{j=1}^{n} A_{jk}\,\mathbf{e}_j\ . \tag{10.23}$$

Logo, para qualquer $k \in \{1, \ldots, n\}$ vale

$$\det(A) = \omega_{det}(a_1, \ldots, a_{k-1}, a_k, a_{k+1}, \ldots, a_n) = \sum_{j=1}^{n} A_{jk}\ \omega_{det}(a_1, \ldots, a_{k-1}, \mathbf{e}_j, a_{k+1}, \ldots, a_n)\ .$$

Note que $\mathbf{e}_j$ ocorre na k-ésima posição. Provamos assim que

$$\det(A) = \sum_{j=1}^{n} A_{jk}\,\mathrm{Cof}(A)_{jk}\ , \tag{10.24}$$

onde a matriz $\mathrm{Cof}(A)$ foi definida em (10.19). Mostremos agora que para $l \neq k$ a expressão $\sum_{j=1}^{n} A_{jl}\ \mathrm{Cof}(A)_{jk}$ é nula. De fato,

$$\sum_{j=1}^{n} A_{jl}\ \mathrm{Cof}(A)_{jk} \ = \ \sum_{j=1}^{n} A_{jl}\ \omega_{det}(a_1,\ \ldots,\ a_{k-1},\ \mathbf{e}_j,\ a_{k+1},\ \ldots,\ a_n)$$

$$\stackrel{(10.23)}{=} \quad \omega_{det}(a_1,\ \ldots,\ a_{k-1},\ a_l,\ a_{k+1},\ \ldots,\ a_n) \ = \ 0 \ ,$$

pois em $\omega_{det}(a_1,\ \ldots,\ a_{k-1},\ a_l,\ a_{k+1},\ \ldots,\ a_n)$ o vetor a_l aparece na l-ésima e na k-ésima posição o que faz ω_{det} anular-se, por ser uma forma alternante. Provamos, assim, que

$$\sum_{j=1}^{n} A_{jl}\ \mathrm{Cof}(A)_{jk} \ = \ \delta_{kl}\ \det(A) \ . \tag{10.25}$$

Vamos supor que $\det(A) \neq 0$. Defina-se a matriz $G = \det(A)^{-1}\mathrm{Cof}(A)^T$, cujos elementos de matriz são $G_{kj} = \det(A)^{-1}\mathrm{Cof}(A)_{jk}$. Então, (10.25) diz-nos que

$$\sum_{j=1}^{n} G_{kj} A_{jl} \ = \ \delta_{kl} \ , \qquad \text{ou seja,} \qquad GA \ = \ \mathbb{1} \ .$$

Isso significa que A é inversível com $A^{-1} = G$.

Prova de 7. Observemos primeiramente que, supondo provisoriamente $k > 1$,

$$\omega_{det}(a_1,\ \ldots,\ a_{k-1},\ \mathbf{e}_j,\ a_{k+1},\ \ldots,\ a_n) \ = \ \omega_{det}\left(a_1 - A_{j1}\mathbf{e}_j,\ \ldots,\ a_{k-1},\ \mathbf{e}_j,\ a_{k+1},\ \ldots,\ a_n\right)$$

devido à linearidade e ao fato que $\omega_{det}(\mathbf{e}_j,\ \ldots,\ a_{k-1},\ \mathbf{e}_j,\ a_{k+1},\ \ldots,\ a_n) = 0$, pelo fato de ω_{det} ser alternante. Agora, a j-ésima linha do vetor-coluna $a_1 - A_{j1}\mathbf{e}_j$ é nula. Repetindo esse argumento podemos anular j-ésima linha de todas as colunas da matriz $[\![a_1,\ \ldots,\ a_{k-1},\ \mathbf{e}_j,\ a_{k+1},\ \ldots,\ a_n]\!]$, exceto a k-ésima coluna, sem alterar seu determinante. Um pouco de meditação nos convence que a matriz resultante é obtida da matriz A anulando-se a k-ésima coluna e a j-ésima linha, exceto no cruzamento das duas, onde o elemento de matriz vale 1 (elemento jk). O determinante dessa matriz é $\mathrm{Cof}(A)_{jk}$.

Pelo item 2 e pela propriedade de alternância, sabemos que o determinante de uma matriz troca de sinal quando permutamos a posição de duas colunas ou duas linhas quaisquer. Com esse tipo de operação podemos transportar o 1 do elemento jk até a posição nn da matriz, ao preço de realizar $n-k$ transposições de colunas vizinhas e $n-j$ de linhas vizinhas, as quais alteram o determinante por fatores $(-1)^{n-k}$ e $(-1)^{n-j}$, respectivamente. Temos com isso que

$$\mathrm{Cof}(A)_{jk} \ = \ (-1)^{k+j}\det\left(\mathcal{A}^{[jk]}\right) \qquad \text{com} \qquad \mathcal{A}^{[jk]} \ := \ \det\begin{pmatrix} \boxed{\begin{matrix} & & \\ & A^{[jk]} & \\ & & \end{matrix}} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \begin{matrix} 0 & \cdots & 0 \end{matrix} & 1 \end{pmatrix} ,$$

onde $A^{[jk]}$ é a matriz de $\mathrm{Mat}\,(\mathbb{C},\ n-1)$ obtida eliminando a j-ésima linha e a k-ésima coluna da matriz A. Pela fórmula de Leibniz (10.17),

$$\det\left(\mathcal{A}^{[jk]}\right) \ = \ \sum_{l \in S_n} \mathrm{sinal}\,(l)\ \left(\mathcal{A}^{[jk]}\right)_{1l(1)} \cdots \left(\mathcal{A}^{[jk]}\right)_{nl(n)} \quad .$$

Como $\left(\mathcal{A}^{[jk]}\right)_{nl(n)} = \delta_{l(n),\,n}$ (justifique!), segue que

$$\begin{aligned}
\det\left(\mathcal{A}^{[jk]}\right) &= \sum_{l'\in S_{n-1}} \text{sinal}\,(l')\left(\mathcal{A}^{[jk]}\right)_{1l'(1)}\cdots\left(\mathcal{A}^{[jk]}\right)_{(n-1)l'(n-1)} \\
&= \sum_{l'\in S_{n-1}} \text{sinal}\,(l')\left(A^{[jk]}\right)_{1l'(1)}\cdots\left(A^{[jk]}\right)_{(n-1)l'(n-1)} \\
&= \det\left(A^{[jk]}\right) = \text{Men}(A)_{jk}\;.
\end{aligned}$$

(Justifique por que a soma no lado direito da primeira linha acima é sobre S_{n-1} e não mais sobre S_n). Provamos, portanto, que

$$\text{Cof}(A)_{jk} = (-1)^{k+j}\,\text{Men}(A)_{jk}\;.$$

A relação (10.20) é imediata por (10.18).

Prova de 8. Eq. (10.22) é imediata por (10.24) e pelo item 7. Eq. (10.21) segue facilmente de (10.22) usando o item 2. ■

• Menores e cofatores de uma matriz. Propriedades adicionais

E. 10.2 *Exercício*. Seja $\Sigma \in \text{Mat}\,(\mathbb{C},\ n)$, $\Sigma = \text{diag}\big(+1,\ -1,\ +1,\ \ldots,\ (-1)^{n+1}\big)$, a matriz diagonal cujos elementos são alternadamente $+1$ e -1, ou seja, $\Sigma_{ij} = (-1)^{i+1}\delta_{ij}$. Mostre que

$$\text{Cof}(A) = \Sigma\text{Men}(A)\Sigma^{-1}$$

para toda matriz $A \in \text{Mat}\,(\mathbb{C},\ n)$. ✠

Para uma matriz $M \in \text{Mat}\,(\mathbb{C},\ n)$, a transformação de similaridade $M \mapsto \Sigma M\Sigma^{-1}$ é denominada "*chessboard transformation*", pois com ela os sinais são trocados em M como alternam-se as cores das casas em um tabuleiro de xadrez.

E. 10.3 *Exercício*. Usando a regra de Laplace (10.18), mostre que para toda matriz $A \in \text{Mat}\,(\mathbb{C},\ n)$ valem as relações

$$\text{Men}\big(\Sigma A\Sigma^{-1}\big) = \Sigma\text{Men}(A)\Sigma^{-1}\,, \qquad \text{Cof}\big(\Sigma A\Sigma^{-1}\big) = \Sigma\text{Cof}(A)\Sigma^{-1}\,,$$
$$\text{Cof}(A) = \text{Men}\big(\Sigma A\Sigma^{-1}\big)\,, \qquad \text{Men}(A) = \text{Cof}\big(\Sigma A\Sigma^{-1}\big)\;.$$

✠

Se $A \in \text{Mat}\,(\mathbb{C},\ n)$ é inversível, segue da regra de Laplace (10.18) que $\det\big(A^{-1}\big) = \frac{1}{\det(A)^n}\det\big(\text{Cof}(A)\big)$ e, portanto,

$$\det\big(\text{Cof}(A)\big) = \det(A)^{n-1}\;. \tag{10.26}$$

Do Exercício E. 10.3, conclui-se também que

$$\det\big(\text{Men}(A)\big) = \det(A)^{n-1}\;. \tag{10.27}$$

E. 10.4 *Exercício*. Mostre que para toda matriz $A \in \text{Mat}\,(\mathbb{C},\ n)$, $n \geq 2$, vale

$$\text{Cof}\big(\text{Cof}(A)\big) = \big(\det(A)\big)^{n-2}A\;.$$

Do Exercício E. 10.3, obtém-se também

$$\text{Men}\big(\text{Men}(A)\big) = \big(\det(A)\big)^{n-2}A\;.$$

Assim, para toda matriz $A \in \text{Mat}\,(\mathbb{C},\ n)$ vale

$$\text{Cof}\big(\text{Cof}(A)\big) = \text{Men}\big(\text{Men}(A)\big)\;.$$

Portanto, se $\det(A) = 1$ e $n \geq 2$, vale $\text{Cof}\big(\text{Cof}(A)\big) = \text{Men}\big(\text{Men}(A)\big) = A$. ✠

• **Um resultado útil**

Mais abaixo, usaremos o seguinte fato:

Proposição 10.3 *Seja* $M \in \mathrm{Mat}(\mathbb{C},\ n)$ *uma matriz da seguinte forma*

$$M = \begin{pmatrix} A & \mathbb{0}_{k,\,n-k} \\ B & C \end{pmatrix},$$

onde A *é uma matriz* $k \times k$ *(com* $k < n$*),* B *é uma matriz* $(n-k) \times k$ *e* C *é uma matriz* $(n-k) \times (n-k)$*. Então,*

$$\det(M) = \det(A)\det(C).$$

□

Prova. O primeiro ingrediente da prova é a constatação que

$$\begin{pmatrix} A & \mathbb{0}_{k,\,n-k} \\ B & C \end{pmatrix} = \begin{pmatrix} A & \mathbb{0}_{k,\,n-k} \\ \mathbb{0}_{n-k,\,k} & \mathbb{1}_{n-k} \end{pmatrix} \begin{pmatrix} \mathbb{1}_k & \mathbb{0}_{k,\,n-k} \\ B & \mathbb{1}_{n-k} \end{pmatrix} \begin{pmatrix} \mathbb{1}_k & \mathbb{0}_{k,\,n-k} \\ \mathbb{0}_{n-k,\,k} & C \end{pmatrix}.$$

E. 10.5 *Exercício*. Verifique! ✣

Com isso, temos pela regra do determinante de um produto de matrizes que

$$\det(M) = \det\begin{pmatrix} A & \mathbb{0}_{k,\,n-k} \\ \mathbb{0}_{n-k,\,k} & \mathbb{1}_{n-k} \end{pmatrix} \det\begin{pmatrix} \mathbb{1}_k & \mathbb{0}_{k,\,n-k} \\ B & \mathbb{1}_{n-k} \end{pmatrix} \det\begin{pmatrix} \mathbb{1}_k & \mathbb{0}_{k,\,n-k} \\ \mathbb{0}_{n-k,\,k} & C \end{pmatrix}.$$

Agora, pelas regras (10.21)–(10.22) de cálculo de determinantes, é fácil constatar (faça-o!) que

$$\det\begin{pmatrix} A & \mathbb{0}_{k,\,n-k} \\ \mathbb{0}_{n-k,\,k} & \mathbb{1}_{n-k} \end{pmatrix} = \det(A), \qquad \det\begin{pmatrix} \mathbb{1}_k & \mathbb{0}_{k,\,n-k} \\ \mathbb{0}_{n-k,\,k} & C \end{pmatrix} = \det(C) \quad \text{e} \quad \det\begin{pmatrix} \mathbb{1}_k & \mathbb{0}_{k,\,n-k} \\ B & \mathbb{1}_{n-k} \end{pmatrix} = 1. \tag{10.28}$$

Cada uma das igualdades acima pode ser provada usando-se a expansão em linhas (10.21) para o determinante. Essa regra nos diz, por exemplo, que o último determinante em (10.1), o da matriz $\begin{pmatrix} \mathbb{1}_k & \mathbb{0}_{k,\,n-k} \\ B & \mathbb{1}_{n-k} \end{pmatrix}$, é igual ao determinante da matriz obtida eliminando-se a primeira linha e a primeira coluna: $\begin{pmatrix} \mathbb{1}_{k-1} & \mathbb{0}_{k-1,\,n-k} \\ B_1 & \mathbb{1}_{n-k} \end{pmatrix}$, com B_1 sendo a matriz obtida de B eliminando-se sua primeira coluna. Mas essa é uma matriz do mesmo tipo da anterior e podemos continuar eliminando a primeira linha e a primeira coluna. Após k repetições desse procedimento, resta apenas a matriz $\mathbb{1}_{n-k}$, cujo determinante vale 1. Para o segundo determinante em (10.21) procede-se analogamente. Para o primeiro, começa-se eliminando a última linha e a última coluna. Isso completa a prova. ■

10.2 Noções Básicas sobre o Espectro de uma Matriz

• **O espectro de uma matriz**

Seja $A \in \mathrm{Mat}(\mathbb{C},\ n)$ uma matriz $n \times n$ com entradas complexas. No estudo das propriedades de A é de grande importância saber para quais números complexos λ a matriz $\lambda\mathbb{1} - A$ é inversível e para quais não é. Essa questão conduz às seguintes importantes definições:

Definição. O *espectro* de $A \in \mathrm{Mat}(\mathbb{C},\ n)$, denotado por $\sigma(A)$, é definido como sendo o conjunto de todos os $\lambda \in \mathbb{C}$ para os quais a matriz $\lambda\mathbb{1} - A$ não tem inversa. Assim, um número complexo λ é dito ser um elemento do *espectro* de $A \in \mathrm{Mat}(\mathbb{C},\ n)$ se a matriz $\lambda\mathbb{1} - A$ não possuir uma inversa. ♠

Definição. O *conjunto resolvente* de $A \in \mathrm{Mat}(\mathbb{C},\ n)$, denotado por $\rho(A)$, é definido como sendo o conjunto de todos os $\lambda \in \mathbb{C}$ para os quais a matriz $\lambda\mathbb{1} - A$ tem inversa. Assim, um número complexo λ é dito ser um elemento do *conjunto resolvente* de $A \in \mathrm{Mat}(\mathbb{C},\ n)$ se a matriz $\lambda\mathbb{1} - A$ possuir uma inversa. ♠

É evidente que $\sigma(A)$ e $\rho(A)$ são conjuntos complementares, ou seja, $\sigma(A) \cap \rho(A) = \emptyset$ mas $\sigma(A) \cup \rho(A) = \mathbb{C}$.

Um fato importante é que $\lambda\mathbb{1} - A$ é não inversível se e somente se $\det(\lambda\mathbb{1} - A) = 0$ (vide Teorema 10.1, página 467). Assim, um número complexo λ é um elemento do espectro de uma matriz A se e somente se for tal que $\det(\lambda\mathbb{1} - A) = 0$.

Essa observação conduz-nos adiante ao importante conceito de polinômio característico de uma matriz.

10.2.1 Autovalores e Polinômios Característicos de Matrizes

• **O polinômio característico de uma matriz**

Seja $A \in \mathrm{Mat}(\mathbb{C},\ n)$ uma matriz cujos elementos de matriz são A_{ij}. Para $z \in \mathbb{C}$ a expressão

$$p_A(z) \ := \ \det(z\mathbb{1} - A) \ = \ \det \begin{pmatrix} z - A_{11} & -A_{12} & \cdots & -A_{1n} \\ -A_{21} & z - A_{22} & \cdots & -A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -A_{n1} & -A_{n2} & \cdots & z - A_{nn} \end{pmatrix} \tag{10.29}$$

define, um polinômio de grau n na variável z, com coeficientes complexos, os quais dependem dos elementos de matriz A_{ij} de A. Isso se constata facilmente pelos métodos usuais de cálculo de determinantes (por exemplo, as expansões em linha ou coluna de (10.21) e (10.22)),

Esse polinômio é denominado *polinômio característico* de A e desempenha um papel muito importante no estudo de propriedades de matrizes. O leitor poderá encontrar na Seção 10.11.1, página 575, uma expressão mais explícita para o polinômio característico em termos dos elementos de matriz A_{ij} de A (vide (10.210), página 575), mas por ora não precisaremos de maiores detalhes sobre esse polinômio.

Como todo polinômio complexo de grau n, p_A possui n raízes, não necessariamente distintas no plano complexo (Teorema Fundamental da Álgebra). As raízes do polinômio característico p_A são denominadas *autovalores* da matriz A. Assim, o espectro de uma matriz A coincide com o conjunto de seus autovalores. O estudo de autovalores de matrizes é de grande importância na Álgebra Linear e em suas aplicações à Teoria das Equações Diferenciais, à Geometria, à Teoria dos Sistemas Dinâmicos e à Física, especialmente à Física Quântica.

Seja $A \in \mathrm{Mat}(\mathbb{C},\ n)$ uma matriz e sejam $\alpha_1,\ \ldots,\ \alpha_r$, $1 \leq r \leq n$, seus autovalores distintos, cada qual com multiplicidade $a_1,\ \ldots,\ a_r$, respectivamente, ou seja, cada α_i é uma raiz de ordem $a_i \in \mathbb{N}$ do polinômio característico de A:

$$p_A(z) \ = \ \det(z\mathbb{1} - A) \ = \ \prod_{i=1}^{r}(z - \alpha_i)^{a_i} \ .$$

A quantidade a_i é um número inteiro positivo e é denominado *multiplicidade algébrica* do autovalor α_i.

Note-se que como o número de raízes de p_A (contando as multiplicidades) é exatamente igual a seu grau, segue facilmente que a seguinte relação é válida:

$$\sum_{i=1}^{r} a_i \ = \ n \ , \tag{10.30}$$

ou seja, a soma das multiplicidades algébricas dos autovalores de uma matriz $A \in \mathrm{Mat}(\mathbb{C},\ n)$ é n. Uma consequência elementar disso é a seguinte proposição útil:

Proposição 10.4 *Seja $A \in \text{Mat}(\mathbb{C},\ n)$ uma matriz e sejam $\alpha_1,\ \ldots,\ \alpha_r$, $1 \leq r \leq n$, seus autovalores distintos, cada qual com multiplicidade algébrica $a_1,\ \ldots,\ a_r$, respectivamente. Então,*

$$\det(A) \;=\; \prod_{k=1}^{r}(\alpha_k)^{a_k} \; . \tag{10.31}$$

□

Prova. Por definição, o polinômio característico de A é $p_A(z) = \det(z\mathbb{1} - A) = \prod_{k=1}^{r}(z - \alpha_k)^{a_k}$. Tomando $z = 0$ e usando (10.30), teremos $\det(-A) = (-1)^n \prod_{k=1}^{r}(\alpha_k)^{a_k}$. Porém, $\det(-A) = (-1)^n \det(A)$ e a proposição está demonstrada. ■

• Matrizes similares. Transformações de similaridade

Duas matrizes $A \in \text{Mat}(\mathbb{C},\ n)$ e $B \in \text{Mat}(\mathbb{C},\ n)$ são ditas *matrizes similares* se existir uma matriz inversível $P \in \text{Mat}(\mathbb{C},\ n)$ tal que $P^{-1}AP = B$. Para uma matriz inversível $P \in \text{Mat}(\mathbb{C},\ n)$ fixa, a transformação que leva cada matriz $A \in \text{Mat}(\mathbb{C},\ n)$ à matriz $P^{-1}AP$ é denominada *transformação de similaridade*.

Sabemos que o determinante é invariante por transformações de similaridade, pois para toda matriz A vale $\det(A) = \det(P^{-1}AP)$, mas não é o único objeto associado a uma matriz que é invariante por tais transformações. O polinômio característico e, portanto, o conjunto de seus autovalores (incluindo as multiplicidades algébricas), também o é. Isso é o conteúdo da seguinte afirmação.

Proposição 10.5 *Sejam A e $B \in \text{Mat}(\mathbb{C},\ n)$ duas matrizes similares, ou seja, tais que existe $P \in \text{Mat}(\mathbb{C},\ n)$, inversível, com $B = P^{-1}AP$. Então, os polinômios característicos de A e de B coincidem: $p_A = p_B$.*

Consequentemente, se A e $B \in \text{Mat}(\mathbb{C},\ n)$ são similares, seus autovalores são iguais (e, portanto, seus espectros: $\sigma(A) = \sigma(B)$), incluindo suas multiplicidades algébricas. □

Prova. O polinômio característico de A é $p_A(z) = \det(z\mathbb{1} - A)$ e o de B é $p_B(z) = \det(z\mathbb{1} - B)$. Logo,

$$p_A(z) \;=\; \det(z\mathbb{1} - A) \;=\; \det(P^{-1}(z\mathbb{1} - A)P) \;=\; \det(z\mathbb{1} - P^{-1}AP) \;=\; \det(z\mathbb{1} - B) \;=\; p_B(z) \; , \tag{10.32}$$

para todo $z \in \mathbb{C}$. Acima usamos o fato que para P inversível e para qualquer matriz M vale $\det(P^{-1}MP) = \det(P^{-1})\det(M)\det(P) = \det(P^{-1}P)\det(M) = \det(\mathbb{1})\det(M) = \det(M)$. ■

• Comentários sobre matrizes inversíveis e sobre matrizes não inversíveis

Proposição 10.6 *Seja $A \in \text{Mat}(\mathbb{C},\ n)$ uma matriz arbitrária e $B \in \text{Mat}(\mathbb{C},\ n)$ uma matriz inversível. Então, existem constantes M_1 e M_2 (dependentes de A e de B) com $0 < M_1 \leq M_2$ tais que a matriz $A + \mu B$ é inversível para todo $\mu \in \mathbb{C}$ com $0 < |\mu| < M_1$ e para todo $\mu \in \mathbb{C}$ com $|\mu| > M_2$.* □

Prova. Como B tem inversa, podemos escrever $A + \mu B = \left(\mu\mathbb{1} + AB^{-1}\right)B$. Assim, $A + \mu B$ será inversível se e somente se $\mu\mathbb{1} + AB^{-1}$ o for.

Seja $C \equiv -AB^{-1}$ e sejam $\{\lambda_1,\ \ldots,\ \lambda_n\} \subset \mathbb{C}$ as n raízes (não necessariamente distintas) do polinômio característico p_C da matriz C. Se todos as raízes forem nulas, tomemos $M_1 = M_2 > 0$, arbitrários. De outra forma, definamos M_1 como sendo o menor valor de $|\lambda_k|$ dentre as raízes não-nulas de p_C: $M_1 := \min\{|\lambda_k|,\ \lambda_k \neq 0\}$ e definimos M_2 como sendo o maior valor de $|\lambda_k|$ para todos os k's: $M_2 := \max\{|\lambda_k|,\ k = 1,\ \ldots,\ n\}$. Então, o conjunto $\{\mu \in \mathbb{C}|\ 0 < |\mu| < M_1\}$ e o conjunto $\{\mu \in \mathbb{C}|\ |\mu| > M_2\}$ não contêm raízes do polinômio característico de C e, portanto, para μ nesses conjuntos a matriz $\mu\mathbb{1} - C = \mu\mathbb{1} + AB^{-1}$ é inversível. ■

Uma consequência evidente da Proposição 10.6 é a seguinte afirmação:

Corolário 10.3 *Seja $B \in \text{Mat}(\mathbb{C},\ n)$ uma matriz inversível e $A \in \text{Mat}(\mathbb{C},\ n)$ uma matriz arbitrária. Então, existem constantes $0 < N_1 \leq N_2$ (dependentes de A e de B) tais que para toda $\nu \in \mathbb{C}$ com $|\nu| < N_1$ ou com $|\nu| > N_2$ a matriz $B + \nu A$ é também inversível.* □

Prova. Para $\nu = 0$ a afirmação é evidente. Para $\nu \neq 0$ a afirmação segue Proposição 10.6 escrevendo-se $B + \nu A = \nu\left(A + \frac{1}{\nu}B\right)$ e tomando-se $\mu = 1/\nu$, $N_1 = 1/M_2$ e $N_2 = 1/M_1$. ■

O interesse pelo Corolário 10.3 é devido ao fato de este afirmar que se $B \in \mathrm{Mat}(\mathbb{C},\ n)$ uma matriz inversível então toda matriz próxima o suficiente da mesma é também inversível. O estudante mais avançado há de reconhecer que essa afirmação ensina-nos que o conjunto da matrizes inversíveis em $\mathrm{Mat}(\mathbb{C},\ n)$ é um conjunto aberto (em uma topologia métrica adequada). Essa afirmação será generalizada (a saber, para álgebras de Banach com unidade) no Corolário 39.6, página 2067.

A Proposição 10.6 afirma também que é sempre possível encontrar uma matriz inversível "próxima" a uma matriz não inversível. De fato, se $A \in \mathrm{Mat}(\mathbb{C},\ n)$ não tem inversa a Proposição 10.6 garante que a matriz $A + \mu\mathbb{1}$, por exemplo, será inversível para todo $\mu \in \mathbb{C}$ com $|\mu|$ pequeno o suficiente, mas não-nulo.

Uma forma geométrica de compreender as afirmações acima é lembrar que conjunto $\mathrm{Mat}(\mathbb{C},\ n)$ é um espaço vetorial n^2-dimensional complexo e as matrizes inversíveis são um subconjunto (n^2-1)-dimensional do mesmo, pois são caracterizados pela condição de terem determinante nulo, uma condição polinomial sobre os n^2 coeficientes das matrizes que define, portanto, uma união finita de superfícies algébricas (n^2-1)-dimensionais fechadas em $\mathrm{Mat}(\mathbb{C},\ n)$. Desse ponto de vista geométrico, fica claro que o conjunto das matrizes inversíveis é aberto (por ser o complementar das superfícies fechadas mencionadas acima) e fica claro que é sempre possível encontrar uma matriz inversível próxima a uma matriz não inversível, pois estas últimas residem em superfícies algébricas de dimensão menor que a dimensão de $\mathrm{Mat}(\mathbb{C},\ n)$.

• Uma propriedade dos polinômios característicos

A seguinte proposição, a qual contém uma afirmação em nada evidente, é uma consequência da Proposição 10.5, página 474, e da Proposição 10.6, página 474:

Proposição 10.7 *Sejam $A,\ B \in \mathrm{Mat}(\mathbb{C},\ n)$. Então, o polinômio característico de AB é igual ao polinômio característico de BA, ou seja, $p_{AB} = p_{BA}$.*

Consequentemente, se $A,\ B \in \mathrm{Mat}(\mathbb{C},\ n)$ então as matrizes AB e BA têm os mesmos autovalores (e, portanto, os mesmos espectros: $\sigma(AB) = \sigma(BA)$), com as mesmas multiplicidades algébricas. □

O estudante mais avançado poderá interessar-se em encontrar na Proposição 39.30, página 2068, uma versão dos resultados da Proposição 10.7 para o caso de álgebras de Banach com unidade.

Prova da Proposição 10.7. Se A ou B são inversíveis (ou ambas), então AB e BA são similares, pois no primeiro caso teremos $AB = A(BA)A^{-1}$ e no segundo teremos $AB = B^{-1}(BA)B$. Nesses casos a afirmação segue da Proposição 10.5, página 474. O único caso que resta considerar é aquele no qual nem A nem B são inversíveis. Nesse caso, porém, temos pela Proposição 10.6, página 474, que existe $M > 0$ tal que a matriz $A + \mu\mathbb{1}$ é inversível para todo $\mu \in \mathbb{C}$ pertencente ao aberto $0 < |\mu| < M$. Assim, para tais valores de μ valerá, pelo raciocínio acima $p_{(A+\mu\mathbb{1})B} = p_{B(A+\mu\mathbb{1})}$. Agora, os coeficientes de $p_{(A+\mu\mathbb{1})B}$ e de $p_{B(A+\mu\mathbb{1})}$ são polinômios em μ e, portanto, são funções contínuas de μ. Logo, a igualdade $p_{(A+\mu\mathbb{1})B} = p_{B(A+\mu\mathbb{1})}$ permanece válida no limite $\mu \to 0$, fornecendo $p_{AB} = p_{BA}$, como desejávamos demonstrar. ■

A Proposição 10.7 pode ser generalizada para matrizes não quadradas, como indicado no exercício que segue:

E. 10.6 *Exercício*. Sejam $A \in \mathrm{Mat}(\mathbb{C},\ m,\ n)$ e $B \in \mathrm{Mat}(\mathbb{C},\ n,\ m)$, de sorte que $AB \in \mathrm{Mat}(\mathbb{C},\ m)$ e $BA \in \mathrm{Mat}(\mathbb{C},\ n)$. Mostre que $x^n p_{AB}(x) = x^m p_{BA}(x)$. *Sugestão:* Considere as matrizes $(m+n)\times(m+n)$ definidas por

$$A' := \begin{pmatrix} A & \mathbb{0}_{m,\ m} \\ \mathbb{0}_{n,\ n} & \mathbb{0}_{n,\ m} \end{pmatrix} \qquad \text{e} \qquad B' := \begin{pmatrix} B & \mathbb{0}_{n,\ n} \\ \mathbb{0}_{m,\ m} & \mathbb{0}_{m,\ n} \end{pmatrix} .$$

(Vide (10.7), página 464). Mostre que

$$A'B' = \begin{pmatrix} AB & \mathbb{0}_{m,\ n} \\ \mathbb{0}_{n,\ m} & \mathbb{0}_{n,\ n} \end{pmatrix} \qquad \text{e que} \qquad B'A' = \begin{pmatrix} BA & \mathbb{0}_{n,\ m} \\ \mathbb{0}_{m,\ n} & \mathbb{0}_{m,\ m} \end{pmatrix} .$$

Em seguida, prove que $p_{A'B'}(x) = x^n p_{AB}(x)$ e que $p_{B'A'}(x) = x^m p_{BA}(x)$. Pela Proposição 10.7, tem-se $p_{A'B'}(x) = p_{B'A'}(x)$, de onde segue que $x^n p_{AB}(x) = x^m p_{BA}(x)$.

Segue disso que o conjunto de autovalores não nulos de AB coincide com o conjunto de autovalores não nulos de BA: $\sigma(AB)\setminus\{0\} = \sigma(BA) \setminus \{0\}$ e, portanto, $\sigma(AB)$ e $\sigma(BA)$ podem não ter em comum apenas o elemento 0. ✣

10.2.2 Autovetores

• **Autovetores**

Pela definição, um número $\lambda_0 \in \mathbb{C}$ é um autovalor de uma matriz A se e somente se $\lambda_0 \mathbb{1} - A$ não tem inversa e, portanto (pelo Corolário 10.1, página 466) se e somente se existir um menos um vetor não-nulo v tal que $(\lambda_0 \mathbb{1} - A)v = 0$, ou seja, tal que $Av = \lambda_0 v$. Chegamos a mais uma importante definição:

Definição. Um vetor não-nulo v é dito ser um *autovetor* de uma matriz A se houver $\lambda_0 \in \mathbb{C}$ tal que

$$Av \; = \; \lambda_0 v \; .$$

Note-se que se um tal λ_0 satisfaz a relação acima para algum $v \neq 0$ então $\lambda_0 \mathbb{1} - A$ não tem inversa. λ_0 é então um elemento do espectro de A, ou seja, um autovalor. λ_0 é dito ser o autovalor associado ao autovetor v. ♠

Uma observação importante é a seguinte. Sejam v_1 e v_2 dois autovetores aos quais está associado o mesmo autovalor, ou seja, $Av_1 = \lambda_0 v_1$ e $Av_2 = \lambda_0 v_2$. Então, para quaisquer números complexos c_1 e c_2 o vetor $v = c_1 v_1 + c_2 v_2$ também satisfaz $Av = \lambda_0 v$. De fato,

$$Av \; = \; A(c_1 v_1 + c_2 v_2) \; = \; c_1 A v_1 + c_2 A v_2 \; = \; c_1 \lambda_0 v_1 + c_2 \lambda_0 v_2 \; = \; \lambda_0 (c_1 v_1 + c_2 v_2) \; = \; \lambda_0 v \; .$$

A conclusão é que, para cada autovalor α_i de uma matriz A, a coleção formada pelo vetor nulo e todos os autovetores de A com autovalor α_i é um subespaço vetorial. Vamos denotar esse subespaço por $\mathcal{E}(\alpha_i)$ ou simplesmente $\mathcal{E}_i$.

Se α_i e α_j são autovalores distintos de A então os subespaços de autovetores $\mathcal{E}(\alpha_i)$ e $\mathcal{E}(\alpha_j)$ têm em comum apenas o vetor nulo, ou seja, $\mathcal{E}(\alpha_i) \cap \mathcal{E}(\alpha_j) = \{0\}$. Isso é fácil de provar, pois se w é tal que $Aw = \alpha_i w$ e $Aw = \alpha_j w$ então, subtraindo-se uma relação da outra teríamos $0 = (\alpha_i - \alpha_j)w$, que implica $w = 0$, já que $\alpha_i \neq \alpha_j$.

Essas considerações nos levam a mais um conceito importante: o de multiplicidade geométrica de um autovalor.

• **A multiplicidade geométrica de um autovalor**

Além do conceito de multiplicidade algébrica de um autovalor, há também o conceito de multiplicidade geométrica de um autovalor, do qual trataremos agora.

Como antes seja $A \in \text{Mat}\,(\mathbb{C}, \; n)$ uma matriz e sejam $\alpha_1, \; \ldots, \; \alpha_r$, $1 \leq r \leq n$, seus autovalores distintos, cada qual com multiplicidade algébrica $a_1, \; \ldots, \; a_r$, respectivamente.

Acima introduzimos os subespaços $\mathcal{E}_i = \mathcal{E}(\alpha_i)$, definidos como sendo os subespaços gerados por todos os autovetores que têm α_i como autovalor. A *multiplicidade geométrica* de um autovalor α_i é definida como sendo a dimensão do subespaço $\mathcal{E}_i$, ou seja, como sendo o número máximo de autovetores linearmente independentes com autovalor α_i.

É importante advertir de imediato o leitor do fato que a multiplicidade algébrica e multiplicidade geométrica de autovalores nem sempre coincidem. Isso é bem ilustrado no seguinte exemplo simples. Seja

$$A \; = \; \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \; .$$

Seu polinômio característico é

$$p_a(\lambda) \; = \; \det(\lambda \mathbb{1} - A) \; = \; \det \begin{pmatrix} \lambda & -1 \\ 0 & \lambda \end{pmatrix} \; = \; \lambda^2 \; .$$

Assim, seu (único) autovalor é 0 com multiplicidade algébrica 2. Quais os seus autovetores? São aqueles vetores que satisfazem $Av = 0$. Denotando v como um vetor coluna $v = \begin{pmatrix} a \\ b \end{pmatrix}$, a relação $Av = 0$ significa $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix} = 0$. Logo, $b = 0$ e todos os autovetores são da forma $v = \begin{pmatrix} a \\ 0 \end{pmatrix}$, $a \in \mathbb{C}$. É evidente que o subespaço gerado pelos autovetores com autovalor zero tem dimensão 1. Assim, a multiplicidade algébrica do autovalor zero é 2 mas a sua multiplicidade geométrica é 1.

• A multiplicidade algébrica e a multiplicidade geométrica

Apesar de a multiplicidade algébrica e a multiplicidade geométrica de um autovalor nem sempre coincidirem, há uma relação de ordem entre eles. A saber, é possível mostrar que a multiplicidade geométrica de um autovalor é sempre menor ou igual à sua multiplicidade algébrica.

Isso segue das seguintes considerações. Seja λ_0 um autovalor de $A \in \text{Mat}(\mathbb{C},\ n)$ e $\mathcal{E}(\lambda_0)$ o subespaço gerado pelos autovetores com autovalor λ_0, e cuja dimensão denotaremos por d. Vamos escolher uma base $v_1, \ \ldots, \ v_d, \ v_{d+1}, \ \ldots, \ v_n$ onde os primeiros d vetores são elementos de $\mathcal{E}(\lambda_0)$. Nessa base a matriz A tem a forma

$$\begin{pmatrix} D & \mathbb{0}_{d,\,n-d} \\ A_3 & A_4 \end{pmatrix} ,$$

onde D é uma matriz $d \times d$ diagonal $D = \text{diag}\left(\underbrace{\lambda_0, \ \ldots, \ \lambda_0}_{d\,\text{vezes}}\right)$, A_4 é uma matriz $(n-d) \times (n-d)$ e A_3 é uma matriz $(n-d) \times d$. Alguns segundos (minutos?) de meditação, usando a Proposição 10.3 da página 472, nos levam a concluir que o polinômio característico de A é dado por

$$\det(\lambda \mathbb{1} - A) \ = \ (\lambda - \lambda_0)^d \det(\lambda \mathbb{1} - A_4) \ .$$

Isso mostra que a multiplicidade algébrica de λ_0 é pelo menos igual a d, sua multiplicidade geométrica.

E. 10.7 *Exercício*. Realize a meditação sugerida acima. ✠

• Matrizes simples

O que foi exposto acima leva-nos naturalmente ao conceito de matriz simples que, como veremos mais adiante, está intimamente ligado ao problema da diagonalizabilidade de matrizes.

Definição. Uma matriz $A \in \text{Mat}(\mathbb{C},\ n)$ é dita ser uma *matriz simples* se cada autovalor de A tiver uma multiplicidade algébrica igual à sua multiplicidade geométrica. ♠

Deixamos para o leitor provar o seguinte fato: toda matriz diagonal é simples.

E. 10.8 *Exercício*. Prove isso. ✠

Adiante faremos uso da seguinte proposição.

Proposição 10.8 *Se $A \in \text{Mat}(\mathbb{C},\ n)$ é uma matriz simples e $P \in \text{Mat}(\mathbb{C},\ n)$ é inversível então $P^{-1}AP$ é também simples.* □

Prova. Já vimos na Proposição 10.5, página 474, que A e $P^{-1}AP$ têm o mesmo polinômio característico e, portanto, os mesmos autovalores, incluindo suas multiplicidades algébricas. Seja λ_0 um desses autovalores com multiplicidade algébrica

d e sejam $v_1, \ldots, v_d$ um conjunto de d autovetores linearmente independentes de A. Os vetores $P^{-1}v_1, \ldots, P^{-1}v_d$ são autovetores de $P^{-1}AP$ com autovalor λ_0. De fato, $\left(P^{-1}AP\right)P^{-1}v_i = P^{-1}Av_i = \lambda_0 P^{-1}v_i$. Fora isso os d vetores $P^{-1}v_1, \ldots, P^{-1}\, v_d$ são também linearmente independentes. Para ver isso, suponha houvesse constantes $c_1, \ldots, c_d$ tais que

$$c_1 P^{-1}v_1 + \cdots + c_d P^{-1}v_d \;=\; 0 \;.$$

Multiplicando-se à esquerda por P teríamos $c_1v_1 + \cdots + c_dv_d = 0$. Como $v_1, \ldots, v_d$ são linearmente independentes as constantes c_i têm que ser todas nulas, provando que os vetores $P^{-1}v_1, \ldots, P^{-1}\, v_d$ são também linearmente independentes.

Isso prova que a multiplicidade geométrica do autovalor λ_0 é pelo menos igual a d. Como ela não pode ser maior que d (página 477), conclui-se que é igual a d provando a proposição. ■

A seguinte proposição elementar é por vezes útil para verificar se uma matriz é simples.

Proposição 10.9 *Se todos os n autovalores de uma matriz $A \in \mathrm{Mat}(\mathbb{C}, n)$ forem distintos então A é simples.* □

Prova. Se os autovalores de A são $\alpha_1, \ldots, \alpha_n$, todos distintos, então cada um tem multiplicidade algébrica igual a 1. Forçosamente, sua multiplicidade geométrica é também igual a 1, já que a multiplicidade geométrica não pode ser maior que a algébrica. ■

Ressaltemos que a recíproca da proposição acima não é verdadeira: uma matriz pode ser simples e possuir autovalores com multiplicidade algébrica maior que 1.

10.2.3 O Traço de uma Matriz

• **O traço de uma matriz**

Seja $A \in \mathrm{Mat}(\mathbb{C}, n)$, cujos elementos de matriz são A_{ij}, $i, j = 1, \ldots n$. Sejam $\lambda_1, \ldots, \lambda_n$ seus n autovalores (não necessariamente distintos e repetidos conforme sua multiplicidade).

Definimos o traço de A como sendo a soma de seus n autovalores:

$$\mathrm{Tr}(A) \;:=\; \sum_{a=1}^{n} \lambda_a \;.$$

Uma conclusão que se tira dessa definição é que se duas matrizes são similares, então ambas têm o mesmo traço, ou seja, para qualquer matriz inversível P e qualquer matriz A vale

$$\mathrm{Tr}\left(P^{-1}AP\right) \;=\; \mathrm{Tr}(A) \;. \tag{10.33}$$

A razão reside na observação feita acima que duas matrizes similares têm o mesmo conjunto de autovalores e, portanto, o mesmo traço.

Temos a seguinte e importante proposição:

Proposição 10.10 *O traço de uma matriz $A \in \mathrm{Mat}(\mathbb{C}, n)$ é igual a soma dos elementos de sua diagonal principal, ou seja,*

$$\mathrm{Tr}(A) \;:=\; \sum_{a=1}^{n} \lambda_a \;=\; \sum_{a=1}^{n} A_{aa} \;. \tag{10.34}$$

□

Prova. A demonstração consistirá em se calcular o coeficiente de λ^{n-1} no polinômio característico $p(\lambda)$ de A de dois modos diferentes. O polinômio característico $p_A(\lambda)$ de A é dado por (10.29). As técnicas de cálculo de determinantes (e.g., (10.21) e (10.22)) dizem-nos que o coeficiente de λ^{n-1} é $-\sum_{i=1}^{n} A_{ii}$. Por exemplo, para o caso $n = 2$

$$p(\lambda) \;=\; \det\begin{pmatrix} \lambda - A_{11} & -A_{12} \\ \\ -A_{21} & \lambda - A_{22}\end{pmatrix} \;=\; \lambda^2 - \lambda(A_{11} + A_{22}) + A_{11}A_{22} - A_{12}A_{21} \; .$$

E. 10.9 *Exercício*. Convença-se da veracidade da afirmativa acima para o caso de n arbitrário. Sugestão: use a expansão em cofatores (10.21)–(10.22) ou leia a Seção 10.11.1, página 575. ✠

Por outro lado, os autovalores de A, $\lambda_1, \ldots, \lambda_n$, são por definição as raízes do polinômio característico. Logo,

$$p(\lambda) \;=\; (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n) \; .$$

Expandindo-se essa expressão, conclui-se que o coeficiente de λ^{n-1} é

$$-(\lambda_1 + \;\cdots\; + \lambda_n) \;=\; -\mathrm{Tr}(A) \; .$$

E. 10.10 *Exercício*. Certo? ✠

Do exposto acima, conclui-se que o coeficiente de λ^{n-1} no polinômio característico de A é

$$-\sum_{i=1}^{n} A_{ii} \;=\; -(\lambda_1 + \;\cdots\; + \lambda_n) \;=\; -\mathrm{Tr}(A) \; ,$$

o que termina a prova. ■

Essa proposição leva a duas outras propriedades igualmente importantes: a *linearidade do traço* e a chamada *propriedade cíclica do traço*.

Proposição 10.11 (A Linearidade do Traço) *Sejam $A, B \in \mathrm{Mat}(\mathbb{C}, n)$ e $\alpha, \beta \in \mathbb{C}$. Então,*

$$\mathrm{Tr}(\alpha A + \beta B) \;=\; \alpha\mathrm{Tr}(A) + \beta\mathrm{Tr}(B) \; .$$

□

Prova. A prova é imediata por (10.34). ■

É curioso notar que a linearidade do traço vista acima é evidente por (10.34), mas não é nem um pouco evidente pela definição do traço de uma matriz como soma de seus autovalores, pois os autovalores individuais de $\alpha A + \beta B$ **não** são em geral combinações lineares dos autovalores de A e de B, especialmente no caso em que A e B não comutam!

Proposição 10.12 (A Propriedade Cíclica do Traço) *Sejam $A, B \in \mathrm{Mat}(\mathbb{C}, n)$. Então,*

$$\mathrm{Tr}(AB) \;=\; \mathrm{Tr}(BA) \; .$$

□

Prova. Pelo que vimos acima, tem-se

$$\mathrm{Tr}(AB) \;=\; \sum_{i=1}^{n}(AB)_{ii} \;=\; \sum_{i=1}^{n}\left(\sum_{j=1}^{n} A_{ij}B_{ji}\right) \;=\; \sum_{j=1}^{n}\left(\sum_{i=1}^{n} B_{ji}A_{ij}\right) \;=\; \sum_{j=1}^{n}(BA)_{jj} \;=\; \mathrm{Tr}(BA) \; .$$

Na segunda e quarta igualdades usamos a regra de produto de matrizes. Na terceira igualdade apenas trocamos a ordem das somas. ■

A propriedade cíclica expressa na Proposição 10.12 pode ser provada diretamente da definição do traço de uma matriz como soma de seus autovalores (incluindo multiplicidades algébricas) se recordarmos a Proposição 10.7, página 475, que afirma que AB e BA têm os mesmos autovalores com as mesmas multiplicidades algébricas.

10.2.3.1 Algumas Relações entre Determinantes e Traços de Matrizes

Proposição 10.13 (Fórmula de Jacobi) *Seja $A(\alpha) \in \mathrm{Mat}(\mathbb{C},\ n)$ uma matriz que depende de forma diferenciável de uma variável α (que pode ser real ou complexa) em um certo domínio. Então, vale*

$$\frac{d}{d\alpha}\Big(\det\big(A(\alpha)\big)\Big) \;=\; \mathrm{Tr}\left(\mathrm{Cof}\big(A(\alpha)\big)^T \frac{d}{d\alpha}A(\alpha)\right) . \tag{10.35}$$

Se $A(\alpha)$ for invertível para todos os valores de α no domínio considerado, vale também

$$\frac{1}{\det\big(A(\alpha)\big)}\frac{d}{d\alpha}\Big(\det\big(A(\alpha)\big)\Big) \;=\; \mathrm{Tr}\left(A(\alpha)^{-1}\frac{d}{d\alpha}A(\alpha)\right) . \tag{10.36}$$

Tanto a relação (10.35) quanto a relação (10.36) são por vezes denominadas fórmula de Jacobi[3]. □

Prova. Por (10.17), tem-se

$$\begin{aligned}\frac{d}{d\alpha}\Big(\det\big(A(\alpha)\big)\Big) &= \sum_{\pi\in S_n}\mathrm{sinal}\,(\pi)\left(\frac{d}{d\alpha}A_{1\pi(1)}(\alpha)\right)\cdots A_{n\pi(n)}(\alpha)+\cdots+\sum_{\pi\in S_n}\mathrm{sinal}\,(\pi)A_{1\pi(1)}(\alpha)\cdots\left(\frac{d}{d\alpha}A_{n\pi(n)}(\alpha)\right)\\ &= \sum_{k=1}^{n}\det\big(B_k(\alpha)\big) ,\end{aligned}$$

onde $B_k(\alpha)$ é a matriz obtida substituindo a k-ésima linha da matriz $A(\alpha)$ pela linha $\begin{pmatrix}\frac{d}{d\alpha}A_{k1}(\alpha) & \cdots & \frac{d}{d\alpha}A_{kn}(\alpha)\end{pmatrix}$. Usando a expansão em linha do determinante, expressão (10.21), temos

$$\det\big(B_k(\alpha)\big) \;=\; \sum_{j=1}^{n}\left(\frac{d}{d\alpha}A_{kj}(\alpha)\right)\mathrm{Cof}\big(A(\alpha)\big)_{kj} .$$

Logo,

$$\frac{d}{d\alpha}\Big(\det\big(A(\alpha)\big)\Big) \;=\; \sum_{k=1}^{n}\sum_{j=1}^{n}\left(\frac{d}{d\alpha}A_{kj}(\alpha)\right)\mathrm{Cof}\big(A(\alpha)\big)_{kj} \;=\; \mathrm{Tr}\left(\mathrm{Cof}\big(A(\alpha)\big)^T\frac{d}{d\alpha}A(\alpha)\right) ,$$

estabelecendo (10.35). A relação (10.36) segue de (10.35) com uso de (10.18). ■

A expressão (10.36) é útil até mesmo no contexto da Geometria Riemanniana. Para uma aplicação naquele contexto, vide expressão (34.113), página 1700. Uma das consequências de (10.36) é o seguinte resultado, também muito útil:

Proposição 10.14 *Seja $A \in \mathrm{Mat}(\mathbb{C},\ n)$. Então, vale que*

$$\det\big(e^A\big) \;=\; e^{\mathrm{Tr}(A)} . \tag{10.37}$$

□

<u>*Nota para o estudante*</u>. A noção de exponencial de uma matriz será apresentada em (11.21), página 590. É fácil ver de (11.21) que $Ae^A = e^A A$ para qualquer matriz $A \in \mathrm{Mat}(\mathbb{C},\ n)$. Da Proposição 11.6, página 592, segue facilmente que e^A é invertível e que sua inversa é e^{-A} também para qualquer $A \in \mathrm{Mat}(\mathbb{C},\ n)$. ♣

Prova da Proposição 10.14. Tome-se $A(\alpha) := e^{\alpha A}$. Então, $\frac{d}{d\alpha}e^{\alpha A} = Ae^{\alpha A} = e^{\alpha A}A$ (por (11.21)) e, portanto, $\big(e^{\alpha A}\big)^{-1}\frac{d}{d\alpha}e^{\alpha A} = A$. Dessa forma, (10.36) fica $\frac{d}{d\alpha}\ln\det\big(A(\alpha)\big) = \mathrm{Tr}(A)$. Integrando-se em α entre 0 e 1 e lembrando que $A(1) = e^A$ e que $A(0) = \mathbb{1}$, teremos $\ln\det\big(e^A\big) = \mathrm{Tr}(A)$, que é o que queríamos provar. ■

Uma segunda demonstração da Proposição 10.14 será encontrada na Proposição 11.7, página 594.

[3] Carl Gustav Jacob Jacobi (1804–1851).

10.2.4 Localização dos Autovalores. Os Discos de Gershgorin

Em muitos problema teóricos e práticos é importante possuirmos uma noção suficientemente precisa da posição no plano complexo dos autovalores de uma matriz $A \in \mathrm{Mat}(\mathbb{C},\ n)$. Um dos teoremas mais simples, úteis e elegantes nessa direção é um resultado obtido por Gershgorin[4] em 1931[5], conhecido como *Teorema dos Discos de Gershgorin*, que passamos a apresentar.

Na Seção 10.5.3, página 521, apresentaremos outros resultados sobre a localização do espectro de matrizes autoadjuntas.

Aqui, nosso primeiro resultado preparatório é o seguinte lema, devido a Lévy[6], que afirma que se o valor absoluto de cada elemento da diagonal de A for maior que a soma dos valores absolutos dos demais elementos da linha[7] correspondente na matriz, então A possui uma inversa.

Lema 10.1 *Se* $A \in \mathrm{Mat}(\mathbb{C},\ n)$ *satisfaz*

$$|A_{ii}| \ > \ \varrho_i(A) \ , \qquad \textit{para todo } i \in \{1,\ \ldots,\ n\} \ , \tag{10.38}$$

onde

$$\varrho_i(A) \ := \ \sum_{\substack{j=1 \\ j\neq i}}^{n} |A_{ij}| \ , \tag{10.39}$$

então $\det(A) \neq 0$. □

<u>Nota</u>. A propriedade (10.38) é por vezes denominada *dominância diagonal estrita* da matriz A. ♣

Prova do Lema 10.1. Vamos supor, por contradição, que (10.38) é satisfeita mas $\det(A) = 0$. Então, A não possui inversa e, pelo Corolário 10.1, página 466, existe um vetor não nulo $v \in \mathbb{C}^n$ tal que $Av = 0$. Sejam $v_1,\ \ldots,\ v_n$ as componentes de v e seja v_r a componente de v com maior módulo: $|v_r| = \max\big\{|v_1|,\ \ldots,\ |v_n|\big\} > 0$. A equação $Av = 0$ se escreve em componentes $\sum_{j=1}^n A_{ij}v_j = 0$ para todo $i \in \{1,\ \ldots,\ n\}$. Assim, para $i = r$ tem-se $A_{rr}v_k = -\sum\limits_{\substack{j=1 \\ j\neq r}}^{n} A_{rj}v_j$. Portanto,

$$|A_{rr}|\,|v_r| \ \leq \ \sum_{\substack{j=1 \\ j\neq r}}^{n} |A_{rj}|\,|v_j| \ \leq \ |v_r| \sum_{\substack{j=1 \\ j\neq r}}^{n} |A_{rj}| \ = \ |v_r|\varrho_k(A) \ ,$$

implicando que $|A_{rr}| \leq \varrho_k(A)$, o que contradiz a hipótese (10.38). Portanto, sob (10.38) devemos ter $\det(A) \neq 0$. ■

Para $A \in \mathrm{Mat}(\mathbb{C},\ n)$, denotamos por $D_i(A)$, $i \in \{1,\ \ldots,\ n\}$ o disco fechado em $\mathbb{C}$ centrado em A_{ii} e de raio $\varrho_i(A)$:

$$D_i(A) \ := \ \Big\{z \in \mathbb{C} \Big|\ |z - A_{ii}| \leq \varrho_i(A)\Big\} \ .$$

Esses discos são denominados *discos de Gershgorin*[8].

Nosso segundo resultado afirma que se λ é um autovalor de A, então λ está contido em algum dos discos de Gershgorin.

Lema 10.2 *Seja* λ *um autovalor de* $A \in \mathrm{Mat}(\mathbb{C},\ n)$. *Então,* $\lambda \in D_j(A)$ *para algum disco de Gershgorin* $D_j(A)$ *de* A. □

---

[4] Semyon Aronovich Gershgorin (1901–1933). Seu nome é também transcrito como Gerčgorin.

[5] S. Gerschgorin, "Über die Abgrenzung der Eigenwerte einer Matrix", Izv. Akad. Nauk. USSR Otd. Fiz.-Mat. Nauk, **6**: 749–754. (1931).

[6] Lucien Lévy (1853–1912). O trabalho original é: L. Lévy. "Sur la possibilité de l'équilibre électrique". Comptes Rendus des Seances de l'Académie des Sciences. Paris. **93**, 706–708. t. XCIII (1881). Lucien Lévy foi o pai do matemático Paul Lévy, conhecido por suas contribuições à Teoria das Probabilidades e Processos Estocásticos.

[7] Como o determinante e o espectro de uma matriz são iguais a de sua transposta, todos os resultados da corrente seção permanecem válidos trocando-se linhas por colunas.

[8] Impropriamente, eles são também denominados *círculos de Gershgorin* por alguns autores.

Prova. Em primeiro lugar, note-se o fato trivial que para todo $\mu \in \mathbb{C}$ tem-se para cada $i \in \{1, \ldots, n\}$ que $\varrho_i(A + \mu\mathbb{1}) = \varrho_i(A)$, pois ambas as matrizes $A + \mu\mathbb{1}$ e A coincidem fora da diagonal principal.

Se λ é um autovalor de $A \in \mathrm{Mat}(\mathbb{C}, n)$, então $\det(A - \lambda\mathbb{1}) = 0$. Logo, pelo Lema 10.1, página 481, não pode valer para todo $i \in \{1, \ldots, n\}$ que $\left|(A - \lambda\mathbb{1})_{ii}\right| > \varrho_i(A + \mu\mathbb{1}) = \varrho_i(A)$. Portanto, para algum índice j deve valer $\left|A_{jj} - \lambda\right| = \left|(A - \lambda\mathbb{1})_{jj}\right| \le \varrho_j(A)$. ■

A questão que se coloca é se todo disco de Gershgorin de $A \in \mathrm{Mat}(\mathbb{C}, n)$ contém um autovalor de A. De modo geral, a resposta é não, como mostra o seguinte exemplo:

Exemplo 10.1 Seja $B = \left(\begin{smallmatrix} 0 & 9 \\ 1 & 0 \end{smallmatrix}\right)$. Uma simples conta revela que seus autovalores são $\lambda = \pm 3$. Seus discos de Gershgorin são $D_1(B) = \{z \in \mathbb{C} |\; |z| \le 9\}$ e $D_2(B) = \{z \in \mathbb{C} |\; |z| \le 1\}$. Vemos que ambos os autovalores estão contidos em $D_1(B)$ mas nenhum em $D_2(B)$. ⌑

Considere-se, em contraste, o seguinte exemplo:

Exemplo 10.2 Seja $C = \left(\begin{smallmatrix} 10 & 1 \\ -1 & -10 \end{smallmatrix}\right)$. Uma simples conta revela que seus autovalores são $\lambda = \pm\sqrt{99} \approx \pm 9,95$. Seus discos de Gershgorin são $D_1(C) = \{z \in \mathbb{C} |\; |z - 10| \le 1\}$ e $D_2(C) = \{z \in \mathbb{C} |\; |z + 10| \le 1\}$. Vemos que o autovalor $+\sqrt{99}$ está contido em $D_1(C)$ e o autovalor $-\sqrt{99}$ está contido em $D_2(C)$. ⌑

O que o Teorema 10.2 a seguir revelará é que a diferença relevante entre os dois exemplos acima é que no primeiro, D_1 e D_2 não são disjuntos, mas no segundo o são.

Precisamos de mais uma definição. Para $A \in \mathrm{Mat}(\mathbb{C}, n)$, dizemos que dois discos de Gershgorin distintos $D_i(A)$ e $D_j(A)$, com $i \ne j$, são *conectados* se $D_i(A) \cap D_j(A) \ne \emptyset$ e que são *desconectados* se $D_i(A) \cap D_j(A) = \emptyset$.

A união de todos os discos de Gershgorin $D_1(A) \cup \cdots D_n(A)$ quebra-se em m componentes conexas (para algum $1 \le m \le n$) que denotaremos por $H_1(A), \ldots, H_m(A)$. Assim, $D_1(A) \cup \cdots \cup D_n(A) = H_1(A) \cup \cdots \cup H_m(A)$, sendo $H_i(A) \cap H_j(A) = \emptyset$ sempre que $i \ne j$. Para cada $j = 1, \ldots, m$, seja $h_j(A)$, $j = 1, \ldots, m$, o número de discos de Gershgorin que compõem $H_j(A)$. É claro que $h_1(A) + \cdots + h_m(A) = n$. A Figura 10.1, página 482 ilustra essas definições.

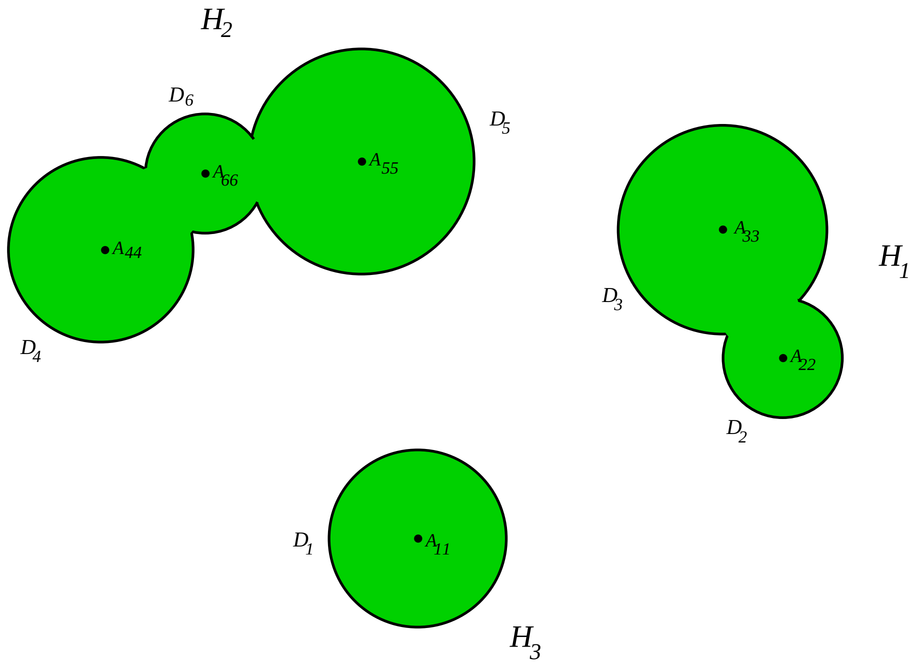

Figura 10.1: No plano complexo, os discos de Gershgorin $D_1, \ldots, D_6$ de uma matriz A complexa 6x6 e as correspondentes regiões conexas $H_1 = D_2 \cup D_3$, $H_2 = D_4 \cup D_5 \cup D_6$ e H_3 e $H_3 = D_1$. Os centros dos discos são os elementos diagonais da matriz: $A_{11}, \ldots, A_{66}$ e seus raios são as grandezas $\varrho_1, \ldots, \varrho_6$. Neste exemplo, $h_1 = 2$, $h_2 = 3$ e $h_3 = 1$.

Vale, então o seguinte:

Teorema 10.2 (Teorema dos Discos de Gershgorin) *Seja $A \in \mathrm{Mat}(\mathbb{C}, n)$. Então, cada componente conexa $H_j(A)$, $j = 1, \ldots, m$, composta por $h_j(A)$ discos de Gershgorin conectados, contém exatamente $h_j(A)$ autovalores de A (contando eventuais multiplicidades).* □

Prova. Seja $H_l(A)$, com $l \in \{1, \ldots, m\}$, uma das regiões conexas supradefinidas e vamos supor que $H_l(A)$ seja a união dos $h_l(A) \equiv h_l$ discos de Gershgorin conectados $D_{i_1}(1), \ldots, D_{i_{h_l}}(A)$.

Seja $A' \in \mathrm{Mat}(\mathbb{C}, n)$ a matriz obtida de A anulando-se todos os elementos não diagonais de A contidos nas linhas $i_1, \ldots, i_{h_l}$:

$$A'_{ij} = \begin{cases} 0, & \text{se } i \in \{i_1, \ldots, i_{h_l}\} \text{ mas } i \neq j , \\ A_{ij}, & \text{de outra forma.} \end{cases}$$

Defina-se também, para $t \in [0, 1]$ a matriz $A_t := tA + (1-t)A'$. Os elementos de matriz de A_t são iguais ao de A, exceto os elementos não diagonais das linhas $i_1, \ldots, i_{h_l}$ que valem t vezes os correspondentes elementos de A:

$$A'_{ij} = \begin{cases} tA_{ij}, & \text{se } i \in \{i_1, \ldots, i_{h_l}\} \text{ mas } i \neq j , \\ A_{ij}, & \text{de outra forma.} \end{cases}$$

As matrizes A_t interpolam continuamente a matriz A' (para $t = 0$) e a matriz A (para $t = 1$).

Tem-se também $\varrho_i(A_t) = t\varrho_i(A)$ para todo $i \in \{i_1, \ldots, i_{h_l}\}$ e todo $t \in [0, 1]$, pois os elementos não diagonais das linhas $i_1, \ldots, i_{h_l}$ de A_t são iguais a t vezes os correspondentes elementos em A.

Segue disso, que cada disco de Gershgorin $D_i(A_t)$, com $i \in \{i_1, \ldots, i_{h_l}\}$, está contido no correspondente disco $D_i(A)$ (ou seja $D_i(A_t) \subset D_i(A)$ para $t \in [0, 1]$), pois eles são concêntricos (centrados no mesmo ponto A_{ii}) e o raio de $D_i(A_t)$ é t vezes o raio de $D_i(A)$, com $0 \leq t \leq 1$.

Isso significa também que

$$D_{i_1}(A_t) \cup \cdots \cup D_{i_{h_l}}(A_t) \subset H_l(A) . \tag{10.40}$$

O leitor pode facilmente verificar que os elementos diagonais das linhas $i_1, \ldots, i_{h_l}$ são autovalores de $A' = A_0$. Eles são os números $A_{i_1 i_1}, \ldots, A_{i_{h_l} i_{h_l}}$. Quando t cresce a partir do valor $t = 0$, esses h_l autovalores variam continuamente com t e permanecem nos respectivos conjuntos $D_{i_j}(A_t)$, $j \in \{1, \ldots, h_l\}$, até o momento em que esses discos de Gershgorin começam a conectar-se, quando então eles podem passar a um novo disco que tenha se conectado a seu disco anterior. De qualquer forma, eles não podem sair fora de $H_l(A)$, devido a (10.40).

Assim, cada $H_l(A)$ conterá ao menos h_l autovalores de A. Em verdade, esse número deve ser exatamente h_l, pois o mesmo argumento pode ser carregado para os demais regiões conexas: cada região $H_i(A)$ deve conter ao menos $h_i(A)$ autovalores. Como $h_1(A) + \cdots + h_m(A) = n$, cada $H_i(A)$ não pode ter mais de $h_i(A)$ autovalores. Essas regiões são desconectas entre si e, portanto, um autovalor de A_t não pode passar continuamente de uma a outra. Isso completa a demonstração. ■

Comentário à demonstração acima. A dependência contínua dos autovalores de A_t, evocada na demonstração acima, é consequência de dois fatos: do Teorema 32.37, página 1596 (que garante a dependência contínua do conjunto de raízes de um polinômio com seus coeficientes), e da dependência contínua (em verdade, polinomial) dos coeficientes do polinômio característico de uma matriz com seus elementos de matriz, fato esse claramente explicitado nas expressões apresentadas na Seção 10.11.1, página 575. ♣

• Aprimoramentos do Teorema dos Discos de Gershgorin

Como o espectro de uma matriz coincide com o de sua transposta, as quantidades $\varrho_i(A)$, $i = \{1, \ldots, n\}$, definidas em (10.39), página 481, podem ser substituídas no Teorema dos Discos de Gershgorin pelas quantidades

$$\varrho'_i(A) := \sum_{\substack{j=1 \\ j\neq i}}^{n} |A_{ji}| , \qquad i = \{1, \ldots, n\} , \tag{10.41}$$

que envolvem somas dos módulos dos elementos não diagonais das *colunas* da matriz A. Essa substituição pode em certos casos conduzir a estimativas mais precisas quanto à localização dos autovalores de A, especialmente se calhar de os valores de $\varrho_i'(A)$ serem estritamente menores que os de $\varrho_i(A)$. Como o espectro de A também é idêntico ao de uma outra matriz que lhe é similar, ou seja, de uma matriz da forma $S^{-1}AS$, com S inversível, há também outras possibilidades de obtenção de discos de Gershgorin de raio menor e que forneçam localizações mais precisas do espectro de A[9]. Em casos concretos, cabe ao engenho e arte do pesquisador explorar a contento essas possibilidades.

Ao leitor interessado cabe dizer que há ainda diversos outros refinamentos ao Teorema dos Discos de Gershgorin sobre a localização de autovalores de matrizes. Há resultados fazendo uso de ovais de Cassini[10], em lugar de discos, e há teoremas devidos a Ostrowski[11], entre outros, que também fornecem informações mais precisas sobre a localização dos autovalores de uma matriz no plano complexo. Vide, *e.g.*, [263] e referências lá citadas. Vide também [349].

10.3 Polinômios de Matrizes

• **Polinômios de matrizes**

Seja p um polinômio de grau m: $p(x) = a_m x^m + \cdots + a_1 x + a_0$ com $x \in \mathbb{C}$, $a_j \in \mathbb{C}$ e $a_m \neq 0$. Para uma matriz $A \in \mathrm{Mat}(\mathbb{C}, n)$ definimos o *polinômio matricial* $p(A)$ por

$$p(A) = a_m A^m + \cdots + a_1 A + a_0 \mathbb{1} .$$

Obviamente $p(A)$ é também uma matriz $n \times n$ com entradas complexas.

Se as raízes do polinômio p forem $\alpha_1, \ldots, \alpha_r$, com multiplicidades $m_1, \ldots, m_r$, respectivamente, então

$$p(x) = a_m \prod_{j=1}^{r} (x - \alpha_j)^{m_j} ,$$

para todo $x \in \mathbb{C}$. É fácil provar, então, que

$$p(A) = a_m \prod_{j=1}^{r} (A - \alpha_j \mathbb{1})^{m_j} .$$

E. 10.11 *Exercício*. Justifique isso. ✠

E. 10.12 *Exercício*. Mostre que se $D = \mathrm{diag}(d_1, \ldots, d_n)$ e q é um polinômio então

$$q(D) = \mathrm{diag}\big(q(d_1), \ldots, q(d_n)\big) .$$

✠

E. 10.13 *Exercício*. Suponha que $A = P^{-1}DP$, onde $D = \mathrm{diag}(d_1, \ldots, d_n)$. Se q é um polinômio mostre que

$$q(A) = P^{-1}q(D)P = P^{-1}\mathrm{diag}\big(q(d_1), \ldots, q(d_n)\big)P .$$

✠

• **O polinômio mínimo**

Vamos mostrar que para cada matriz $A \in \mathrm{Mat}(\mathbb{C}, n)$ sempre existe pelo menos um polinômio p com a propriedade que $p(A) = \mathbb{0}$. Para tal notemos primeiramente que $\mathrm{Mat}(\mathbb{C}, n)$ é um espaço vetorial complexo de dimensão n^2. De fato

[9] Tal se dá no caso trivial em que A é diagonalizável, se S for escolhida de sorte a diagonalizar A, ou seja, se $S^{-1}AS = D$, uma matriz diagonal. Os raios dos discos de Gershgorin da matriz diagonalizada são, evidentemente, todos nulos e, pelo Teorema dos Discos de Gershgorin, o espectro de A coincide com os elementos da diagonal de D, como esperado.

[10] Giovanni Domenico Cassini (1625–1712).

[11] Alexander Markowich Ostrowski (1893–1986).

toda a matriz $A \in \text{Mat}(\mathbb{C}, n)$, cujos elementos de matriz são $A_{ij} \in \mathbb{C}$ pode ser trivialmente escrita na forma

$$A = \sum_{a=1}^{n}\sum_{b=1}^{n} A_{ab}\mathscr{E}^{ab}$$

onde $\mathscr{E}^{ab} \in \text{Mat}(\mathbb{C}, n)$ são matrizes cujos elementos de matriz são $(\mathscr{E}^{ab})_{ij} = \delta_{i,a}\delta_{j,b}$, ou seja, todos os elementos de matriz de $\mathscr{E}^{ab}$ são nulos, exceto o elemento a, b, que vale 1.

E. 10.14 *Exercício*. Certo? ✠

Assim, vemos que as matrizes $\{\mathscr{E}^{ab},\ a = 1, \ldots, n,\ b = 1, \ldots, n\}$ formam uma base em $\text{Mat}(\mathbb{C}, n)$, mostrando que $\text{Mat}(\mathbb{C}, n)$ é um espaço vetorial de dimensão n^2. Isto posto, temos que concluir que qualquer conjunto de mais de n^2 matrizes não-nulas em $\text{Mat}(\mathbb{C}, n)$ é linearmente dependente.

Se uma das matrizes A^k, $k = 1, \ldots, n^2$, for nula, digamos $A^q = \mathbb{0}$, então o polinômio $p(x) = x^q$ tem a propriedade que $p(A) = 0$, que é o que desejamos provar. Se, por outro lado, as matrizes A^k, $k = 1, \ldots, n^2$, são todas não-nulas, então o conjunto $\{\mathbb{1}, A, A^2, \ldots, A^{n^2}\}$ é linearmente dependente, pois possui $n^2 + 1$ elementos. Portanto, existem constantes $c_0, \ldots, c_{n^2}$, nem todas nulas, tais que

$$c_0\mathbb{1} + c_1A + c_2A^2 + \cdots + c_{n^2}A^{n^2} = \mathbb{0} .$$

Como o lado esquerdo é um polinômio em A, fica provada nossa afirmação que toda matriz possui um polinômio que a anula. Chegamos às seguintes definições:

Definição. Polinômio Mônico. Um polinômio $p : \mathbb{R} \to \mathbb{C}$ de grau n é dito ser um *polinômio mônico* se for da forma

$$p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 ,$$

ou seja, se o coeficiente do monômio de maior grau (no caso, x^n) for igual a 1. Note-se que polinômios mônicos nunca são identicamente nulos. ♠

Definição. Polinômio Mínimo de uma Matriz. Dada uma matriz $A \in \text{Mat}(\mathbb{C}, n)$, o *polinômio mínimo* de A é o polinômio mônico de menor grau que é anulado em A, ou seja, é o polinômio não-nulo de menor grau da forma

$$M(x) = x^m + a_{m-1}x^{m-1} + \cdots + a_1x + a_0$$

para o qual $M(A) = \mathbb{0}$. ♠

As considerações acima mostram que um tal polinômio sempre existe e que tem grau no máximo igual a n^2. Essa é, no entanto, uma estimativa exagerada para o grau do polinômio mínimo de uma matriz $A \in \text{Mat}(\mathbb{C}, n)$ pois, como veremos abaixo, o polinômio mínimo de uma matriz $A \in \text{Mat}(\mathbb{C}, n)$ tem, na verdade, grau menor ou igual a n. Isso é um corolário de um teorema conhecido como Teorema de Hamilton-Cayley , que demonstraremos abaixo (Teorema 10.4, página 486).

Finalizamos com um teorema básico que garante a unicidade do polinômio mínimo e estabelece sua relação com outros polinômios que anulam A.

Teorema 10.3 *O polinômio mínimo M de uma matriz $A \in \text{Mat}(\mathbb{C}, n)$ é único. Fora isso se P é um polinômio não identicamente nulo que também se anula em A, ou seja, $P(A) = \mathbb{0}$, então P é divisível por M, ou seja, existe um polinômio F tal que $P(x) = F(x)M(x)$ para todo $x \in \mathbb{C}$.* □

Demonstração. Dada uma matriz $A \in \text{Mat}(\mathbb{C}, n)$, o polinômio mínimo de A é o polinômio de menor grau da forma

$$M(x) = x^m + a_{m-1}x^{m-1} + \cdots + a_1x + a_0$$

para o qual $M(A) = \mathbb{0}$. Vamos supor que haja outro polinômio N da forma

$$N(x) = x^m + b_{m-1}x^{m-1} + \cdots + b_1x + b_0$$

para o qual $N(A) = \mathbb{0}$. Subtraindo um do outro teríamos o polinômio

$$(M - N)(x) \ = \ (a_{m-1} - b_{m-1})x^{m-1} + \cdots + (a_1 - b_1)x + (a_0 - b_0) \ ,$$

que tem grau menor ou igual a $m-1$ e para o qual vale $(M - N)(A) = M(A) - N(A) = \mathbb{0} - \mathbb{0} = \mathbb{0}$. Como, por hipótese, não há polinômios não nulos com grau menor que o de M que anulam A, isso é uma contradição, a menos que $M = N$. Isso prova a unicidade.

Seja P um polinômio não identicamente nulo para o qual valha $P(A) = \mathbb{0}$. Se p é o grau de P, deve-se ter $p \geq m$, onde m é o grau do polinômio mínimo de A. Logo, pelos bem conhecidos fatos sobre divisão de polinômios, podemos encontrar dois polinômios F e R, cujos graus são, respectivamente $p - m$ e r com $0 \leq r < m$, tais que

$$P(x) \ = \ F(x)M(x) + R(x) \ ,$$

para todo $x \in \mathbb{C}$. Ora, isso diz que

$$P(A) \ = \ F(A)M(A) + R(A) \ .$$

Como $P(A) = \mathbb{0}$ e $M(A) = \mathbb{0}$, isso implica $R(A) = \mathbb{0}$. Como, porém, o grau de R é menor que m, tem-se que R deve ser identicamente nulo. Isso completa a prova. ■

10.3.1 O Teorema de Hamilton-Cayley

Vamos aqui demonstrar um teorema sobre matrizes que será usado mais adiante de várias formas, em particular no Teorema Espectral, o chamado Teorema de Hamilton[12]-Cayley[13], o qual afirma que toda matriz de $\mathrm{Mat}(\mathbb{C}, \ n)$ anula seu próprio polinômio característico. Esse teorema fornece também, como veremos, um método eficiente para o cálculo da inversa de matrizes. Cayley e Hamilton demonstraram casos particulares do teorema para matrizes 2×2, 3×3 (Cayley) e 4×4 (Hamilton). A primeira demonstração geral é devida a Frobenius[14]. Cayley, Hamilton e Sylvester[15] estão entre os fundadores modernos da teoria das matrizes[16].

Teorema 10.4 (Teorema de Hamilton-Cayley) *Seja $A \in \mathrm{Mat}(\mathbb{C}, \ n)$ e seja $p_A(x) = \det(x\mathbb{1} - A)$ o polinômio característico de A (e que tem grau n). Então, $p_A(A) = \mathbb{0}$.* □

Comentário. No caso particular de matrizes diagonalizáveis o Teorema 10.4 pode ser provado elementarmente usando o Teorema Espectral, como indicado no Exercício E. 10.21, página 497. ♣

Prova do Teorema 10.4. Desejamos mostrar que para todo vetor $y \in \mathbb{C}^n$ vale $p_A(A)y = 0$. Se $y = 0$ isso é trivial. Se $y \neq 0$ mas com $Ay = 0$ então

$$p_A(A)y \ = \ (-1)^n \lambda_1 \cdots \lambda_n y \ ,$$

onde $\lambda_1, \ \cdots \ , \ \lambda_n$ são os autovalores de A. Mas a própria relação $Ay = 0$ indica que um dos autovalores é igual a zero. Logo $p_A(A)y = 0$. Mais genericamente, se $y \neq 0$ e $\{y, \ Ay\}$ não for um conjunto de vetores linearmente independentes, então Ay e y são proporcionais, ou seja, existe um autovalor, digamos, λ_n tal que $Ay = \lambda_n y$. Nesse caso também tem-se

$$p_A(A)y \ = \ \left(\prod_{i=1}^{n-1}(A - \lambda_i\mathbb{1})\right)(A - \lambda_n\mathbb{1})y \ = \ 0 \ ,$$

pois $(A - \lambda_n\mathbb{1})y = Ay - \lambda_n y = 0$.

Seja então y daqui por diante um vetor fixado, não-nulo e tal que $\{y, \ Ay\}$ é um conjunto de dois vetores não nulos e linearmente independentes.

[12]Sir William Rowan Hamilton (1805–1865).
[13]Arthur Cayley (1821–1895).
[14]Ferdinand Georg Frobenius (1849–1917)
[15]James Joseph Sylvester (1814–1897).
[16]Muitos certamente se surpreenderão muitíssimo em saber que, apesar de suas diversas e importantes contribuições à Matemática, Cayley e Sylvester eram originalmente advogados.

Como o espaço $\mathbb{C}^n$ tem dimensão n, nem todos os conjuntos de vetores da forma

$$\{y,\ Ay,\ A^2y,\ \ldots,\ A^jy\}$$

são formados por vetores não nulos linearmente independentes. Por exemplo, se $j \geq n$, o conjunto $\{y,\ Ay,\ A^2y,\ \ldots,\ A^jy\}$ não pode ser formado por vetores não nulos linearmente independentes pois seu número excede a dimensão do espaço.

Seja k o maior número tal que $\{y,\ Ay,\ A^2y,\ \ldots\ A^{k-1}y\}$ é um conjunto de vetores não nulos e linearmente independentes. É claro que $1 < k \leq n$.

É claro também, pela definição de k, que

$$A^ky \ = \ h_ky + h_{k-1}Ay + \cdots + h_1A^{k-1}y \ , \tag{10.42}$$

para constantes $h_1, \ldots,\ h_k$.

Vamos denominar $z_1 = A^{k-1}y$, $z_2 = A^{k-2}y$, $\ldots$, $z_k = y$, ou seja, $z_j = A^{k-j}y$, $j = 1,\ \ldots,\ k$, todos não nulos por hipótese. Caso $k < n$, escolhamos ainda vetores $z_{k+1}, \ldots, z_n$ de modo que o conjunto $\{z_1,\ \ldots,\ z_n\}$ forme uma base em $\mathbb{C}^n$.

Coloquemo-nos agora a seguinte questão: qual é a forma da matriz A nessa base? No subespaço gerado pelos vetores $\{z_1,\ \ldots,\ z_k\}$ tem-se o seguinte: para $i = 2, \ldots, k$ vale $Az_i = z_{i-1}$. Além disso, por (10.42), $Az_1 = h_1z_1 + h_2z_2 + \cdots + h_kz_k$. Isso mostra que o subespaço gerado pelos vetores $\{z_1,\ \ldots,\ z_k\}$ é invariante pela ação de A e o operador linear A, no mesmo subespaço, tem a forma

$$\begin{pmatrix} h_1 & 1 & 0 & \ldots & 0 & 0 \\ h_2 & 0 & 1 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ h_{k-2} & 0 & 0 & \ddots & 1 & 0 \\ h_{k-1} & 0 & 0 & \ldots & 0 & 1 \\ h_k & 0 & 0 & \ldots & 0 & 0 \end{pmatrix} . \tag{10.43}$$

E. 10.15 *Exercício*. Justifique isso. ✠

Se designarmos por P o operador que realiza essa mudança de base, o operador linear A na base $\{z_1,\ \ldots,\ z_n\}$ tem, portanto, a forma $A' = P^{-1}AP$, onde

$$A' \ = \ \begin{pmatrix} A_1 & \mathbb{0}_{k,\,n-k} \\ A_2 & A_3 \end{pmatrix} ,$$

onde A_1 é a matriz $k \times k$ definida em (10.43), A_2 é uma matriz $(n-k) \times k$ e A_3 é uma matriz $(n-k) \times (n-k)$. Não nos será necessário especificar os elementos das matrizes A_2 e A_3.

Outros segundos (minutos?) de meditação, usando a Proposição 10.3 da página 472, nos levam a concluir que o polinômio característico p_A pode ser escrito como

$$p_A(x) \ = \ \det(x\mathbb{1} - A') \ = \ \det(x\mathbb{1} - A_1)\det(x\mathbb{1} - A_3) \ .$$

O estudante deve recordar-se que as matrizes A e A', por serem similares, têm o mesmo polinômio característico (Proposição 10.5, página 474).

Vamos denominar $q_k(x) = \det(x\mathbb{1} - A_1)$ e $r_k(x) = \det(x\mathbb{1} - A_3)$. Claramente, $p_A(x) = q_k(x)r_k(x)$. Não será necessário, no que segue, calcular r_k, mas precisaremos calcular q_k. Como esse pequeno resultado tem interesse independente, vamos formulá-lo como um lema, para futura referência.

Lema 10.3 *Para $h_1, \ldots, h_k \in \mathbb{C}$, tem-se*

$$q_k(x) := \det\begin{pmatrix} x-h_1 & -1 & 0 & \ldots & 0 & 0 \\ -h_2 & x & -1 & \ddots & 0 & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ -h_{k-2} & 0 & 0 & \ddots & -1 & 0 \\ -h_{k-1} & 0 & 0 & \ldots & x & -1 \\ -h_k & 0 & 0 & \ldots & 0 & x \end{pmatrix} = x^k - \left(h_1 x^{k-1} + \cdots + h_{k-1}x + h_k\right) . \tag{10.44}$$

□

Prova. A prova é feita por indução. Para $k = 2$ vale

$$q_2(x) = \det\begin{pmatrix} x-h_1 & -1 \\ -h_2 & x \end{pmatrix} = x^2 - h_1 x - h_2 .$$

Para $k > 2$, tem-se, pelas bem conhecidas regras de cálculo de determinantes,

$$\begin{aligned} q_k(x) &= x\det\begin{pmatrix} x-h_1 & -1 & & 0 & 0 \\ -h_2 & x & \ddots & 0 & 0 \\ \vdots & & \ddots & \ddots & \\ -h_{k-2} & 0 & & x & -1 \\ -h_{k-1} & 0 & \ldots & 0 & x \end{pmatrix}_{(k-1)\times(k-1)} + 1\det\begin{pmatrix} x-h_1 & -1 & & 0 & 0 \\ -h_2 & x & \ddots & 0 & 0 \\ \vdots & & \ddots & \ddots & \\ -h_{k-2} & 0 & & x & -1 \\ -h_k & 0 & \ldots & 0 & 0 \end{pmatrix}_{(k-1)\times(k-1)} \\ &= xq_{k-1}(x) + (-1)^{k-1+1}(-h_k)\det\begin{pmatrix} -1 & 0 & \ldots & 0 & 0 \\ x & -1 & \ddots & 0 & 0 \\ & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & -1 & 0 \\ 0 & 0 & \ldots & x & -1 \end{pmatrix}_{(k-2)\times(k-2)} \\ &= xq_{k-1}(x) + (-1)^{k+1}h_k(-1)^{k-2} \\ &= xq_{k-1}(x) - h_k . \end{aligned} \tag{10.45}$$

E. 10.16 *Exercício*. Complete os detalhes. ✠

Assim, se pela hipótese indutiva q_{k-1} é da forma

$$q_{k-1}(x) = x^{k-1} - \left(h_1 x^{k-2} + \cdots + h_{k-2}x + h_{k-1}\right) ,$$

segue de (10.45) que

$$
\begin{aligned}
q_k(x) &= x(x^{k-1} - (h_1x^{k-2} + \cdots + h_{k-2}x + h_{k-1})) - h_k \\
&= x^k - (h_1x^{k-1} + \cdots + h_{k-2}x^2 + h_{k-1}x + h_k) \ ,
\end{aligned}
\tag{10.46}
$$

como queríamos provar. ■

Retomando, temos que $p_A(A)y = q_k(A)r_k(A)y = r_k(A)q_k(A)y$. Sucede, porém, que $q_k(A)y = 0$. De fato, pelo cômputo acima,

$$
q_k(A)y \ = \ A^ky - h_1A^{k-1}y - \cdots - h_{k-2}A^2y - h_{k-1}Ay - h_ky \ ,
$$

que é igual a zero por (10.42). Logo $p_A(A)y = 0$. Como y foi escolhido arbitrário, segue que $p_A(A) = \mathbb{0}$, demonstrando o Teorema de Hamilton-Cayley, Teorema 10.4. ■

• O Teorema de Hamilton-Cayley e a inversa de matrizes

O Teorema de Hamilton-Cayley fornece-nos um método de calcular a inversa de matrizes não singulares. De fato, se $p_A(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ é o polinômio característico de uma matriz não singular A, então o Teorema de Hamilton-Cayley afirma que

$$
A^n + a_{n-1}A^{n-1} + \cdots + a_1A + a_0\mathbb{1} \ = \ \mathbb{0} \ ,
$$

ou seja,

$$
A\left(A^{n-1} + a_{n-1}A^{n-2} + \cdots + a_2A + a_1\mathbb{1}\right) \ = \ -a_0\mathbb{1} \ .
$$

Isso tem por implicação

$$
A^{-1} \ = \ -\frac{1}{a_0}\left(A^{n-1} + a_{n-1}A^{n-2} + \cdots + a_2A + a_1\mathbb{1}\right) \ . \tag{10.47}
$$

Vide (10.211), página 576, para uma expressão mais explícita.

Nota. Usando a definição de polinômio característico $p_A(x) = \det(x\mathbb{1} - A)$, é evidente (tomando-se $x = 0$) que $a_0 = (-1)^n \det(A)$. Assim, $a_0 \neq 0$ se e somente se A for não singular. ♣

Em muitos casos a fórmula (10.47) é bastante eficiente para calcular A^{-1}, pois a mesma envolve poucas operações algébricas em comparação com outros métodos, o que é uma vantagem para valores grandes de n. Compare, por exemplo, com a regra de Laplace, expressão (10.20), página 467, para o cálculo de A^{-1}, que envolve o cômputo de $n^2 + 1$ determinantes de submatrizes de ordem $n - 1$ de A.

E. 10.17 *Exercício*. Use esse método para calcular a inversa das suas matrizes não singulares favoritas. ✠

• De volta ao polinômio mínimo

O Teorema 10.3, página 485, e o Teorema de Hamilton-Cayley, juntos, permitem-nos precisar algo a respeito da forma geral do polinômio mínimo de uma matriz.

Se $A \in \mathrm{Mat}\,(\mathbb{C},\ n)$ tem r autovalores distintos $\alpha_1, \ \ldots, \ \alpha_r$, cada qual com multiplicidade algébrica $a_1, \ \ldots, \ a_r$, respectivamente, então seu polinômio característico p_A é da forma

$$
p_A(x) \ = \ \prod_{k=1}^{r}(x - \alpha_k)^{a_k} \ .
$$

Pelo Teorema de Hamilton-Cayley, $p_A(A) = 0$ e, portanto, pelo Teorema 10.3, M, o polinômio mínimo de A, divide q. Logo, M deve ser da forma

$$
M(x) \ = \ \prod_{l=1}^{s}(x - \alpha_{k_l})^{b_l} \ , \tag{10.48}
$$

onde $s \leq r$, $\{\alpha_{k_1}, \ \ldots, \ \alpha_{k_s}\} \subset \{\alpha_1, \ \ldots, \ \alpha_r\}$ e onde $0 < b_l \leq a_{k_l}$ para todo $1 \leq l \leq s$. Seja agora, porém, $v_m \neq 0$ um autovetor de A com autovalor α_m Segue do fato que $M(A) = 0$ que

$$
0 \ = \ M(A)v_m \ = \ \prod_{l=1}^{s}(A - \alpha_{k_l}\mathbb{1})^{b_l}\ v_m \ = \ \prod_{l=1}^{s}(\alpha_m - \alpha_{k_l})^{b_l}\ v_m \ .
$$

Logo, $\prod_{l=1}^{s} (\alpha_m - \alpha_{k_l})^{b_l} = 0$ e isso implica que $\alpha_m \in \{\alpha_{k_1}, \ldots, \alpha_{k_s}\}$. Como isso vale para todo $1 \leq m \leq r$, segue que $\{\alpha_1, \ldots, \alpha_r\} \subset \{\alpha_{k_1}, \ldots, \alpha_{k_s}\}$ e, portanto, $\{\alpha_1, \ldots, \alpha_r\} = \{\alpha_{k_1}, \ldots, \alpha_{k_s}\}$. Nossa conclusão é resumida no seguinte:

Proposição 10.15 *Seja $A \in \mathrm{Mat}(\mathbb{C}, n)$ com r autovalores distintos $\alpha_1, \ldots, \alpha_r \in \mathbb{C}$, cada qual com multiplicidade algébrica $a_1, , \ldots, a_r$, sendo $1 \leq r \leq n$. Então, M, o polinômio mínimo de A, é da forma*

$$M(x) = \prod_{k=1}^{r} (x - \alpha_k)^{b_k} , \tag{10.49}$$

$\forall x \in \mathbb{C}$, onde $0 < b_l \leq a_l$ para todo $1 \leq l \leq r$. Em particular, se $A \in \mathrm{Mat}(\mathbb{C}, n)$ tiver exatamente n autovalores distintos, teremos que $b_l = a_l = 1$ para todo $1 \leq l \leq n$, e

$$M(x) = p_A(x) = \prod_{k=1}^{n} (x - \alpha_k) ,$$

$\forall x \in \mathbb{C}$. □

• Usos do Teorema de Hamilton-Cayley para matrizes 2×2

E. 10.18 *Exercício*. Usando o Teorema de Hamilton-Cayley, mostre que toda matriz 2×2 complexa $A \in \mathrm{Mat}(\mathbb{C}, 2)$ satisfaz

$$A^2 - \mathrm{Tr}(A)A + \det(A)\mathbb{1} = 0 . \tag{10.50}$$

Sugestão: se $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, mostre que seu polinômio característico é $p_A(x) = x^2 - \mathrm{Tr}(A)x + \det(A)$.

Se $A \in \mathrm{Mat}(\mathbb{C}, 2)$ for inversível, mostre com uso de (10.50) que vale a simples relação

$$A^{-1} = \frac{1}{\det(A)}\big[\mathrm{Tr}(A)\mathbb{1} - A\big] . \tag{10.51}$$

Assim, se $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ tem inversa (ou seja, se $ad - bc \neq 0$), então $A^{-1} = \frac{1}{ad-bc}\left[\left(\begin{smallmatrix} a+d & 0 \\ 0 & a+d \end{smallmatrix}\right) - \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\right] = \frac{1}{ad-bc}\left(\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right)$, resultado esse bem conhecido e que pode ser obtido por diversos outros métodos.

A identidade (10.50) tem emprego importante na Mecânica Quântica de sistemas desordenados unidimensionais e na Mecânica Estatística. ✠

10.3.1.1 O Teorema da Aplicação Espectral para Matrizes

Vamos nesta seção demonstrar um importante teorema sobre o espectro de matrizes. Esse teorema e sua demonstração deixam-se generalizar com toda literalidade a uma situação mais geral, a saber, a de álgebras de Banach. Vide Seção 39.3.5.1, página 2072.

Como acima, denotamos por $\sigma(M)$ o espectro (conjunto de autovalores) de uma matriz $M \in \mathrm{Mat}(\mathbb{C}, m)$. Seja a álgebras de matrizes $\mathrm{Mat}(\mathbb{C}, m)$ e seja um polinômio $p(z) = a_0 + a_1 z + \cdots + a_n z^n$ definido para $z \in \mathbb{C}$. Para $A \in \mathrm{Mat}(\mathbb{C}, n)$ definimos, como antes, $p(A) := a_0\mathbb{1} + a_1 A + \cdots + a_n A^n \in \mathrm{Mat}(\mathbb{C}, m)$. O Teorema da Aplicação Espectral, que demonstraremos logo abaixo consiste na afirmação que $\sigma\big(p(A)\big) = p\big(\sigma(A)\big)$, onde

$$p\big(\sigma(A)\big) := \big\{p(\lambda), \ \lambda \in \sigma(A)\big\} .$$

Para demonstrá-lo, usaremos o seguinte resultado:

Lema 10.4 *Seja $A \in \mathrm{Mat}(\mathbb{C}, m)$. Então, se $\lambda \in \sigma(A)$, a matriz $(A - \lambda\mathbb{1})q(A)$ não tem inversa para nenhum polinômio q.* □

Prova. Seja $p(z) := (z - \lambda)q(z)$. Então, $p(A) = (A - \lambda\mathbb{1})q(A)$. É evidente que $q(A)$ e $p(A)$ comutam com A: $q(A)A = Aq(A)$ e $p(A)A = Ap(A)$. Desejamos provar que $p(A)$ não tem inversa e, para tal, vamos supor o oposto, a saber, vamos supor que exista $W \in \mathrm{Mat}(\mathbb{C}, m)$ tal que $Wp(A) = p(A)A = \mathbb{1}$.

Vamos primeiramente provar que A e W comutam. Seja $C := WA - AW$. Então, multiplicando-se à esquerda por $p(A)$, teremos $p(A)C = A - p(A)AW = A - Ap(A)W = A - A = 0$. Assim, $p(A)C = 0$ e multiplicando-se essa igualdade à esquerda por W teremos $C = 0$, estabelecendo que $WA = AW$. Naturalmente, isso implica também que $q(A)W = Wq(A)$.

Agora, por hipótese, A satisfaz $p(A)W = Wp(A) = \mathbb{1}$, ou seja, $(A - \lambda\mathbb{1})q(A)W = \mathbb{1}$ e $W(A - \lambda\mathbb{1})q(A) = \mathbb{1}$. Usando a comutatividade de $q(A)$ com A e com W, essa última relação pode ser reescrita como $q(A)W(A - \lambda\mathbb{1}) = \mathbb{1}$. Assim, estabelecemos que

$$(A - \lambda\mathbb{1})\Big(q(A)W\Big) = \mathbb{1} \qquad \text{e} \qquad \Big(q(A)W\Big)(A - \lambda\mathbb{1}) = \mathbb{1}$$

o que significa que $A - \lambda\mathbb{1}$ tem inversa, sendo $(A - \lambda\mathbb{1})^{-1} = q(A)W$, uma contradição com a hipótese que $\lambda \in \sigma(A)$. Logo, $p(A)$ não pode ter inversa. ■

Passemos agora ao nosso objetivo.

Teorema 10.5 (Teorema da Aplicação Espectral (para matrizes)) *Seja* $A \in \mathrm{Mat}\,(\mathbb{C},\ m)$. *Então,*

$$\sigma\big(p(A)\big) = p\big(\sigma(A)\big) := \Big\{p(\lambda),\ \lambda \in \sigma(A)\Big\} \tag{10.52}$$

para todo polinômio p. □

Prova. Vamos supor que $p(z) = a_0 + a_1 z + \cdots + a_n z^n$ seja de grau $n \geq 1$, pois no caso de um polinômio constante a afirmativa é trivial. Naturalmente, $a_n \neq 0$.

Tomemos $\mu \in \sigma\big(p(A)\big)$, que é não-vazio, como sabemos, e sejam $\alpha_1,\ \ldots,\ \alpha_n$ as n raízes do polinômio $p(z) - \mu$ em $\mathbb{C}$. Então, $p(z) - \mu = a_n(z - \alpha_1)\cdots(z - \alpha_n)$, o que implica $p(A) - \mu\mathbb{1} = a_n(A - \alpha_1\mathbb{1})\cdots(A - \alpha_n\mathbb{1})$. Se nenhum dos α_i pertencesse a $\sigma(A)$, então cada fator $(A - \alpha_j\mathbb{1})$ seria inversível, assim como o produto $a_n(A - \alpha_1\mathbb{1})\cdots(A - \alpha_n\mathbb{1})$, contrariando o fato de $\mu \in \sigma\big(p(A)\big)$. Logo, algum dos α_i pertence a $\sigma(A)$. Como $p(\alpha_i) = \mu$, isso diz que $\sigma\big(p(A)\big) \subset \{p(\lambda),\ \lambda \in \sigma(A)\}$.

Provemos agora a recíproca. Já sabemos que $\sigma(A)$ é não-vazio. Para $\lambda \in \sigma(A)$ tem-se evidentemente que o polinômio $p(z) - p(\lambda)$ tem λ como raiz. Logo, $p(z) - p(\lambda) = (z - \lambda)q(z)$, onde q é um polinômio de grau $n - 1$. Portanto, $p(A) - p(\lambda)\mathbb{1} = (A - \lambda\mathbb{1})q(A)$ e como $(A - \lambda\mathbb{1})$ não é inversível, $p(A) - p(\lambda)\mathbb{1}$ também não pode sê-lo (pelo Lema 10.4, página 490), o que diz-nos que $p(\lambda) \in \sigma(p(A))$. Isso significa que $\{p(\lambda),\ \lambda \in \sigma(A)\} \subset \sigma\big(p(A)\big)$, estabelecendo que $\sigma\big(p(A)\big) = \{p(\lambda),\ \lambda \in \sigma(A)\}$. ■

10.4 Matrizes Diagonalizáveis e o Teorema Espectral

• **Matrizes diagonalizáveis**

Vamos agora apresentar uma noção intimamente ligada à de matriz simples introduzida acima (página 477), mas de importância maior.

Definição. Uma matriz $A \in \mathrm{Mat}\,(\mathbb{C},\ n)$ é dita ser uma *matriz diagonalizável* se existir uma matriz inversível $P \in \mathrm{Mat}\,(\mathbb{C},\ n)$ tal que $P^{-1}AP$ é uma matriz diagonal, ou seja,

$$P^{-1}AP = D = \mathrm{diag}\,(d_1,\ \ldots,\ d_n) = \begin{pmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_n \end{pmatrix} .$$

♠

É fácil de se ver que os elementos da diagonal de D são os autovalores de A. De fato, se A é diagonalizável por P, vale para seu polinômio característico

$$
\begin{aligned}
p(\lambda) &= \det(\lambda\mathbb{1} - A) = \det(P^{-1}(\lambda\mathbb{1} - A)P) = \det(\lambda\mathbb{1} - P^{-1}AP) = \det(\lambda\mathbb{1} - D) \\
&= \det\begin{pmatrix} \lambda - d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda - d_n \end{pmatrix} = (\lambda - d_1)\cdots(\lambda - d_n) ,
\end{aligned}
$$

o que mostra que os d_i são as raízes do polinômio característico de A e, portanto, seus autovalores.

E. 10.19 *Exercício*. Justifique todas as passagens acima. ✠

• **Diagonalização de matrizes**

O próximo teorema é fundamental no estudo de matrizes diagonalizáveis.

Teorema 10.6 *Uma matriz $A \in \mathrm{Mat}(\mathbb{C},\ n)$ é diagonalizável se e somente se possuir um conjunto de n autovetores linearmente independentes, ou seja, se e somente se o subespaço gerado pela coleção de todos os autovetores de A possuir dimensão n.* □

Prova. Vamos primeiro provar que se $A \in \mathrm{Mat}(\mathbb{C},\ n)$ possui um conjunto de n autovetores linearmente independentes então A é diagonalizável. Para tal vamos construir a matriz P que diagonaliza A.

Seja $\{v^1,\ \ldots,\ v^n\}$ um conjunto de n autovetores linearmente independentes de A, cujos autovalores são $\{d_1,\ \ldots,\ d_n\}$, respectivamente. Vamos denotar as componentes de v^i na base canônica por v^i_j, $j = 1, \ldots, n$. Seja a matriz P definida por $P = \left[\!\left[v^1,\ \ldots,\ v^n \right]\!\right]$, ou seja,

$$
P = \begin{pmatrix} v^1_1 & \cdots & v^n_1 \\ \vdots & \ddots & \vdots \\ v^1_n & \cdots & v^n_n \end{pmatrix} .
$$

Como se vê pela construção, a a-ésima coluna de P é formada pelas componentes do vetor v^a. Por (10.12), segue que

$$
AP = \left[\!\left[Av^1,\ \ldots,\ Av^n \right]\!\right] = \left[\!\left[d_1v^1,\ \ldots,\ d_nv^n \right]\!\right] .
$$

Por (10.15) vale, porém, que

$$
\left[\!\left[d_1v^1,\ \ldots,\ d_nv^n \right]\!\right] = \begin{pmatrix} v^1_1 & \cdots & v^n_1 \\ \vdots & \ddots & \vdots \\ v^1_n & \cdots & v^n_n \end{pmatrix} \begin{pmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_n \end{pmatrix} = PD .
$$

E. 10.20 *Exercício*. Verifique. ✠

Portanto, $AP = PD$. Como, por hipótese, as colunas de P são formadas por vetores linearmente independentes, tem-se que $\det(P) \neq 0$ (por quê?). Logo, P é inversível e, portanto, $P^{-1}AP = D$, como queríamos demonstrar.

Vamos provar agora a afirmação recíproca que se A é diagonalizável, então possui n autovetores linearmente independentes. Suponha que exista P tal que

$$P^{-1}AP = D = \begin{pmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_n \end{pmatrix} .$$

É evidente que os vetores da base canônica

$$e^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e^2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad e^n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

são autovetores de D com $De^a = d_a e^a$. Logo, $v^a = Pe^a$ são autovetores de A, pois

$$Av^a = APe^a = PDe^a = P(d_a e^a) = d_a Pe^a = d_a v^a .$$

Para provar que os vetores v^a são linearmente independentes, suponha que existam números complexos $\alpha_1, \ldots, \alpha_n$ tais que $\alpha_1 v^1 + \cdots + \alpha_n v^n = 0$. Multiplicando-se à esquerda por P^{-1} teríamos $\alpha_1 e^1 + \cdots + \alpha_n e^n = 0$. Como os e^a são obviamente linearmente independentes, segue que $\alpha_1 = \cdots = \alpha_n = 0$. ■

• Matrizes diagonalizáveis e matrizes simples

Vamos agora discutir a relação entre os conceitos de matriz diagonalizável e o de matriz simples, conceito esse introduzido à página 477. Tem-se a saber o seguinte fato:

Proposição 10.16 *Uma matriz $A \in \mathrm{Mat}(\mathbb{C}, n)$ é diagonalizável se e somente se for simples, ou seja, se e somente se a multiplicidade algébrica de cada um dos seus autovalores coincidir com sua multiplicidade geométrica.* □

Prova. Se A é diagonalizável existe P tal que $P^{-1}AP = D$, diagonal. Como toda matriz diagonal, D é simples. Escrevamos D na forma

$$D = \mathrm{diag}\left(\underbrace{\alpha_1, \ldots, \alpha_1}_{a_1 \text{ vezes}}, \ldots, \underbrace{\alpha_r, \ldots, \alpha_r}_{a_r \text{ vezes}}\right) .$$

Um conjunto de n-autovetores de D linearmente independentes é fornecido pelos vetores da base canônica:

$$e^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e^2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad e^n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} .$$

Os vetores $e^1, \ldots, e^{a_1}$ geram o subespaço de autovetores com autovalor α_1 de D etc.

Para a matriz A, os vetores $Pe^1, \ldots, Pe^{a_1}$ geram o subespaço de autovetores com autovalor α_1 etc. É claro que a dimensão desse subespaço é a_1, pois $Pe^1, \ldots, Pe^{a_1}$ são linearmente independentes, já que os vetores da base canônica $e^1, \ldots, e^{a_1}$ o são. Como isso também vale para os demais autovalores concluímos que A é simples.

Resta-nos agora mostrar que se $A \in \text{Mat}(\mathbb{C}, n)$ é simples então A é diagonalizável. Como antes, sejam $\alpha_1, \ldots, \alpha_r$, $1 \leq r \leq n$, seus autovalores distintos, cada qual com multiplicidade algébrica $a_1, \ldots, a_r$, respectivamente, e seja $\mathcal{E}(\alpha_i)$ o subespaço gerado pelos autovetores com autovalor α_i. Como A é simples, tem-se que a dimensão de $\mathcal{E}(\alpha_i)$ é a_i. Já observamos (página 476) que subespaços $\mathcal{E}(\alpha_i)$ associados a autovalores distintos têm em comum apenas o vetor nulo. Assim, se em cada $\mathcal{E}(\alpha_i)$ escolhermos a_i vetores independentes, teremos ao todo um conjunto de $\sum_{i=1}^{r} a_i = n$ autovetores (vide (10.30)) linearmente independentes de A. Pelo Teorema 10.6, A é diagonalizável, completando a prova. ■

• Projetores

Uma matriz $E \in \text{Mat}(\mathbb{C}, n)$ é dita ser um *projetor* se satisfizer

$$E^2 = E .$$

Projetores são também denominados *matrizes idempotentes*.

Discutiremos várias propriedades importantes de projetores adiante, especialmente de uma classe especial de projetores denominados *projetores ortogonais*. Por ora, vamos mostrar duas propriedades que usaremos logo abaixo quando discutirmos o teorema espectral.

A primeira propriedade é a afirmação que se λ é um autovalor de um projetor E então ou λ é igual a zero ou a um. De fato se v é um autovetor associado a um autovalor λ de E, tem-se que $Ev = \lambda v$ e $E^2 v = \lambda^2 v$. Como $E^2 = E$, segue que $\lambda^2 v = \lambda v$. Logo $\lambda(\lambda - 1) = 0$ e, portanto, $\lambda = 0$ ou $\lambda = 1$.

A segunda propriedade é uma consequência da primeira: o traço de um projetor $E \in \text{Mat}(\mathbb{C}, n)$ é um número inteiro positivo ou nulo, mas menor ou igual a n. De fato, pela definição, o traço de um projetor E é a soma de seus autovalores. Como os mesmos valem zero ou um a soma é um inteiro positivo ou nulo. Como há no máximo n autovalores a soma não pode exceder n. Na verdade, o único projetor cujo traço vale exatamente n é a identidade $\mathbb{1}$ e o único projetor cujo traço vale exatamente 0 é a matriz nula (por quê?).

Essas observações têm a seguinte consequência que usaremos adiante. Se $E_1, \ldots, E_r$ são r projetores não nulos com a propriedade que

$$\mathbb{1} = \sum_{a=1}^{r} E_a$$

então $r \leq n$. Para ver isso, basta tomar o traço de ambos os lados dessa expressão:

$$\text{Tr}(\mathbb{1}) = \sum_{a=1}^{r} \text{Tr}(E_a) . \tag{10.53}$$

O lado esquerdo vale n enquanto que o lado direito é uma soma de r inteiros positivos. Obviamente isso só é possível se $r \leq n$.

Uma outra observação útil é a seguinte: se E e E' são dois projetores satisfazendo $EE' = E'E = \mathbb{0}$, então $E + E'$ é igualmente um projetor, como facilmente se constata.

• O Teorema Espectral

O chamado Teorema Espectral é um dos mais importantes teoremas de toda a Álgebra Linear e, em verdade, de toda Análise Funcional, já que o mesmo possui generalizações para operadores limitados e não limitados (autoadjuntos) agindo em espaços de Hilbert. Dessas generalizações trataremos na Seção 39.8.2, página 2155, para o caso dos chamados operadores compactos e na Seção 39.9, página 2161, para o caso geral de operadores limitados autoadjuntos. Nessa versão mais geral o teorema espectral é de importância fundamental para a interpretação probabilística da Física Quântica. Vide discussão da Seção 46.3, página 2508.

Teorema 10.7 (Teorema Espectral para Matrizes) *Uma matriz $A \in \text{Mat}(\mathbb{C}, n)$ é diagonalizável se e somente se existirem $r \in \mathbb{N}$, $1 \leq r \leq n$, escalares distintos $\alpha_1, \ldots, \alpha_r$ e projetores não nulos distintos $E_1, \ldots, E_r \in \text{Mat}(\mathbb{C}, n)$*

tais que

$$A = \sum_{a=1}^{r} \alpha_a E_a \,, \tag{10.54}$$

$$\mathbb{1} = \sum_{a=1}^{r} E_a \tag{10.55}$$

e

$$E_i E_j = \delta_{i,\,j} E_j \,.$$

Os escalares $\alpha_1, \ldots, \alpha_r$ vêm a ser os autovalores distintos de A. □

Adiante demonstraremos uma versão um pouco mais detalhada desse importante teorema (Teorema 10.9, abaixo). Os projetores E_a que surgem em (10.54) são denominados *projetores espectrais* de A. A decomposição (10.54) é frequentemente denominada *decomposição espectral* de A. Na Proposição 10.18, página 497 mostraremos como os projetores espectrais E_a de A podem ser expressos em termos de polinômios em A. Na Proposição 10.19, página 497, provaremos a unicidade da decomposição espectral de uma matriz diagonalizável.

Prova do Teorema 10.7. Se $A \in \mathrm{Mat}(\mathbb{C}, n)$ é diagonalizável existe $P \in \mathrm{Mat}(\mathbb{C}, n)$ tal que $P^{-1}AP = D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, onde $\lambda_1, \ldots, \lambda_n$ são os autovalores de A. Como pode haver autovalores repetidos, vamos denotar por $\{\alpha_1, \ldots, \alpha_r\}$, $1 \leq r \leq n$, o conjunto de autovalores distintos de A.

É bem claro que podemos escrever

$$D = \sum_{a=1}^{r} \alpha_a K_a \,,$$

onde as matrizes K_a são todas matrizes diagonais, cujos elementos diagonais são ou 0 ou 1 e tais que

$$\sum_{a=1}^{r} K_a = \mathbb{1} \,. \tag{10.56}$$

As matrizes K_a são simplesmente definidas de modo a terem elementos de matriz iguais a 1 nas posições da diagonal ocupadas pelo autovalor α_a em D e zero nos demais. Formalmente,

$$(K_a)_{ij} = \begin{cases} 1, & \text{se } i = j \text{ e } (D)_{ii} = \alpha_a \\[2ex] 0, & \text{se } i = j \text{ e } (D)_{ii} \neq \alpha_a \\[2ex] 0, & \text{se } i \neq j \end{cases} .$$

Por exemplo, se

$$D = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \qquad \text{teremos} \qquad D = 2\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + 3\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + 4\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} .$$

É fácil constatar que as matrizes K_a têm a seguinte propriedade:

$$K_a K_b = \delta_{a,\,b}\, K_a . \tag{10.57}$$

De fato, é evidente que $(K_a)^2 = K_a$ para todo a, pois K_a é diagonal com zeros ou uns na diagonal. Analogamente, se $a \neq b$ $K_a K_b = 0$, pois os zeros ou uns aparecem em lugares distintos das diagonais das duas matrizes.

Como $A = PDP^{-1}$, tem-se que

$$A = \sum_{a=1}^{r} \alpha_a E_a ,$$

onde $E_a := PK_aP^{-1}$. É fácil agora provar que $\mathbb{1} = \sum_{a=1}^{r} E_a$ e que $E_iE_j = \delta_{i,\,j}E_j$. De fato, por (10.56),

$$\sum_{a=1}^{r} E_a = \sum_{a=1}^{r} PK_aP^{-1} = P\left(\sum_{a=1}^{r} K_a\right)P^{-1} = P\mathbb{1}P^{-1} = \mathbb{1} .$$

Analogamente, tem-se por (10.57),

$$E_aE_b = PK_aP^{-1}PK_bP^{-1} = PK_aK_bP^{-1} = \delta_{a,\,b}\, PK_aP^{-1} = \delta_{a,\,b}\, E_a .$$

Vamos agora provar a recíproca. Vamos supor que A possua a representação (10.54), onde os E_a's satisfazem as propriedades enunciadas.

Notemos primeiramente que para $x \in \mathbb{C}^n$, e para $k \in \{1, \ldots, r\}$, tem-se por (10.54)

$$AE_kx = \sum_{j=1}^{r} \alpha_j E_j E_k x = \alpha_k E_k x .$$

Logo, ou $E_kx = 0$ ou E_kx é autovetor de A. Assim, o subespaço $\mathcal{S}$ gerado pelos conjunto de vetores $\{E_kx,\ x \in \mathbb{C}^n,\ k = 1, \ldots, r\}$ é um subespaço do espaço $\mathcal{A}$ gerado pelos autovetores de A. Agora, por (10.55), temos, para todo $x \in \mathbb{C}^n$,

$$x = \mathbb{1}x = \sum_{k=1}^{r} E_k x$$

e este fato revela que $\mathbb{C}^n = \mathcal{S} \subset \mathcal{A}$ e, portanto, que $\mathcal{A} = \mathbb{C}^n$. Assim, pelo Teorema 10.6, página 492, A é diagonalizável. Isso completa a demonstração. ■

No Teorema 10.9, página 500, apresentaremos uma segunda demonstração do Teorema Espectral para Matrizes, a qual lança luz sobre outras condições de diagonalizabilidade de matrizes. Antes, exploremos algumas das consequências do Teorema Espectral.

• O Cálculo Funcional para matrizes diagonalizáveis

O Teorema Espectral tem o seguinte corolário, muitas vezes conhecido como *cálculo funcional*.

Teorema 10.8 (Cálculo Funcional) *Seja $A \in \mathrm{Mat}(\mathbb{C},\ n)$ uma matriz diagonalizável e seja*

$$A = \sum_{a=1}^{r} \alpha_a E_a$$

sua decomposição espectral, de acordo com o Teorema Espectral, o Teorema 10.7, onde os α_a, $a = 1, \ldots, r$, com $0 < r \leq n$ são os autovalores distintos de A e os E_a's são os correspondentes projetores espectrais. Então, para qualquer polinômio p vale

$$p(A) = \sum_{a=1}^{r} p(\alpha_a) E_a . \tag{10.58}$$

□

Prova. Tem-se, pelas propriedades dos E_a's, $A^2 = \sum_{a,\,b=1}^{r} \alpha_a\alpha_b E_aE_b = \sum_{a,\,b=1}^{r} \alpha_a\alpha_b\delta_{a,\,b}E_a = \sum_{a=1}^{r}(\alpha_a)^2E_a$. Analogamente, mostra-se que $A^m = \sum_{a=1}^{r}(\alpha_a)^mE_a$, para qualquer $m \in \mathbb{N}$. O resto da prova é trivial. ■

E. 10.21 *Exercício*. Usando (10.58) demonstre novamente o Teorema de Hamilton-Cayley (Teorema 10.4, página 486), agora apenas para o caso particular de matrizes diagonalizáveis. ✠

Por simples constatação verifica-se também facilmente a validade do seguinte resultado, que usaremos diversas vezes:

Proposição 10.17 *Seja $A \in \mathrm{Mat}(\mathbb{C},\ n)$ uma matriz diagonalizável e inversível e seja $A = \sum_{a=1}^{r} \alpha_a E_a$ sua decomposição espectral, de acordo com o Teorema Espectral, o Teorema 10.7. Então, $A^{-1} = \sum_{a=1}^{r} \frac{1}{\alpha_a} E_a$.* □

• Obtendo os projetores espectrais

O Cálculo Funcional para matrizes, Teorema 10.8, tem diversas consequências práticas, uma delas sendo a seguinte proposição, que permite expressar os projetores espectrais de uma matriz A diretamente em termos de A e seus autovalores.

Proposição 10.18 *Seja $A \in \mathrm{Mat}(\mathbb{C},\ n)$, não-nula e diagonalizável, e seja $A = \alpha_1 E_1 + \cdots + \alpha_r E_r$, com os α_k's distintos, sua representação espectral, descrita no Teorema 10.7. Sejam os polinômios p_j, $j = 1, \ldots, r$, definidos por*

$$p_j(x) \ := \ \prod_{\substack{l=1 \\ l \neq j}}^{r} \left(\frac{x - \alpha_l}{\alpha_j - \alpha_l} \right) . \tag{10.59}$$

Então,

$$E_j \ = \ p_j(A) \ = \ \left(\prod_{\substack{k=1 \\ k \neq j}}^{r} \frac{1}{\alpha_j - \alpha_k} \right) \prod_{\substack{l=1 \\ l \neq j}}^{r} \Big(A - \alpha_l \mathbb{1} \Big) \tag{10.60}$$

para todo $j = 1, \ldots, r$. □

Prova. Pela definição dos polinômios p_j, é evidente que $p_j(\alpha_k) = \delta_{j,\,k}$. Logo, pelo Cálculo Funcional para matrizes,

$$p_j(A) \ = \ \sum_{k=1}^{r} p_j(\alpha_k) E_k \ = \ E_j \ .$$

■

• O Teorema Espectral para matrizes. Unicidade

Proposição 10.19 *A representação espectral de uma matriz diagonalizável $A \in \mathrm{Mat}(\mathbb{C},\ n)$ descrita no Teorema 10.7 é única.* □

Demonstração. Seja $A \in \mathrm{Mat}(\mathbb{C},\ n)$ diagonalizável e seja $A = \sum_{k=1}^{r} \alpha_k E_k$ a representação espectral de A descrita no Teorema 10.7, onde α_k, $k = 1, \ldots, r$, com $1 \leq r \leq n$ são os autovalores distintos de A, Seja $A = \sum_{k=1}^{r'} \alpha'_k E'_k$ uma segunda representação espectral para A, onde os α'_k's são distintos e onde os E'_k's são não nulos e satisfazem $E'_j E'_l = \delta_{j,\,l} E'_l$ e $\mathbb{1} = \sum_{k=1}^{r'} E'_k$. Por essa última propriedade segue que para um dado vetor $x \neq 0$ vale $x = \sum_{k=1}^{r'} E'_k x$, de modo que nem todos os vetores $E'_k x$ são nulos. Seja $E'_{k_0} x$ um desses vetores não nulos. Tem-se que $A E'_{k_0} x = \sum_{k=1}^{r'} \alpha'_k E'_k E'_{k_0} x = \alpha'_{k_0} E'_{k_0} x$. Isso mostra que α'_{k_0} é um dos autovalores de A e, portanto, $\{\alpha'_1, \ldots, \alpha'_{r'}\} \subset \{\alpha_1, \ldots, \alpha_r\}$. Isso, em particular ensina-nos que $r' \leq r$. Podemos sem perda de generalidade considerar que os dois conjuntos sejam ordenados de modo que $\alpha'_k = \alpha_k$ para todo $1 \leq k \leq r'$. Assim,

$$A = \sum_{k=1}^{r} \alpha_k E_k = \sum_{k=1}^{r'} \alpha_k E'_k . \tag{10.61}$$

Sejam agora os polinômios p_j, $j = 1, \ldots, r$, definidos em (10.59), os quais satisfazem $p_j(\alpha_j) = 1$ e $p_j(\alpha_k) = 0$ para todo $k \neq j$. Pelo Cálculo Funcional descrito acima, segue de (10.61) que, com $1 \leq j \leq r'$,

$$p_j(A) = \underbrace{\sum_{k=1}^{r} p_j(\alpha_k) E_k}_{=E_j} = \underbrace{\sum_{k=1}^{r'} p_j(\alpha_k) E'_k}_{=E'_j} , \qquad \therefore \quad E_j = E'_j .$$

(A igualdade $p_j(A) = \sum_{k=1}^{r'} p_j(\alpha_k) E'_k$ segue do fato que os E'_k's satisfazem as mesmas relações algébricas que os E_k's e, portanto, para a representação espectral de A em termos dos E'_k's vale também o Cálculo Funcional). Como $\mathbb{1} = \sum_{k=1}^{r} E_k = \sum_{k=1}^{r'} E'_k$ e como $E_j = E'_j$ para $1 \leq j \leq r'$, tem-se $\sum_{k=r'+1}^{r} E_k = \mathbb{0}$. Multiplicando isso por E_l com $r'+1 \leq l \leq r$, segue que $E_l = \mathbb{0}$ para todo $r'+1 \leq l \leq r$. Isso só é possível se $r = r'$, pois os $E'k$'s são não nulos. Isso completa a demonstração. ■

• Algumas outras identidades decorrentes do Teorema Espectral

Proposição 10.20 *Seja $A \in \mathrm{Mat}(\mathbb{C}, n)$ uma matriz diagonalizável e invertível, cujos autovalores distintos sejam $\{\alpha_1, \ldots, \alpha_r\}$, para algum $1 \leq r \leq n$. Então, vale a identidade*

$$\left(\prod_{\substack{j=1 \\ j\neq k}}^{r} \frac{1}{\alpha_k - \alpha_j} \right) \prod_{\substack{l=1 \\ l\neq k}}^{r} \Big(A - \alpha_l \mathbb{1} \Big) = \left(\prod_{\substack{j=1 \\ j\neq k}}^{r} \frac{1}{\alpha_k^{-1} - \alpha_j^{-1}} \right) \prod_{\substack{l=1 \\ l\neq k}}^{r} \Big(A^{-1} - \alpha_l^{-1} \mathbb{1} \Big) , \tag{10.62}$$

para cada $k \in \{1, \ldots, r\}$. □

Observe-se também que (10.62) implica

$$\prod_{\substack{l=1 \\ l\neq k}}^{r} \Big(\alpha_k^{-1} - \alpha_l^{-1} \Big) \Big(A - \alpha_l \mathbb{1} \Big) = \prod_{\substack{l=1 \\ l\neq k}}^{r} \Big(\alpha_k - \alpha_l \Big) \Big(A^{-1} - \alpha_l^{-1} \mathbb{1} \Big) \tag{10.63}$$

e

$$\frac{(-1)^{r-1}}{\alpha_k^{r-2} \prod_{j=1}^{r} \alpha_j} \prod_{\substack{l=1 \\ l\neq k}}^{r} \Big(A - \alpha_l \mathbb{1} \Big) = \prod_{\substack{l=1 \\ l\neq k}}^{r} \Big(A^{-1} - \alpha_l^{-1} \mathbb{1} \Big) . \tag{10.64}$$

Prova da Proposição 10.20. Pelo Teorema Espectral e por (10.60) podemos escrever A em sua representação espectral:

$$A = \sum_{k=1}^{r} \alpha_k \left(\prod_{\substack{j=1 \\ j\neq k}}^{r} \frac{1}{\alpha_k - \alpha_j} \right) \prod_{\substack{l=1 \\ l\neq k}}^{r} \Big(A - \alpha_l \mathbb{1} \Big) . \tag{10.65}$$

Se A é também invertível, a Proposição 10.17, página 497, informa-nos que

$$A^{-1} = \sum_{k=1}^{r} \frac{1}{\alpha_k} \left(\prod_{\substack{j=1 \\ j\neq k}}^{r} \frac{1}{\alpha_k - \alpha_j} \right) \prod_{\substack{l=1 \\ l\neq k}}^{r} \Big(A - \alpha_l \mathbb{1} \Big) .$$

Por outro lado, se A é invertível, A^{-1} é diagonalizável (justifique!), e seus autovalores distintos são $\{\alpha_1^{-1}, \ldots, \alpha_r^{-1}\}$ (justifique!). Logo, a representação espectral de A^{-1} é

$$A^{-1} = \sum_{k=1}^{r} \alpha_k^{-1} \left(\prod_{\substack{j=1 \\ j\neq k}}^{r} \frac{1}{\alpha_k^{-1} - \alpha_j^{-1}} \right) \prod_{\substack{l=1 \\ l\neq k}}^{r} \left(A^{-1} - \alpha_l^{-1}\mathbb{1} \right) ,$$

onde as matrizes $\left(\prod_{\substack{j=1 \\ j\neq k}}^{r} \frac{1}{\alpha_k^{-1} - \alpha_j^{-1}} \right) \prod_{\substack{l=1 \\ l\neq k}}^{r} \left(A^{-1} - \alpha_l^{-1}\mathbb{1} \right)$ são os projetores espectrais de A^{-1}. Aplicando novamente a Proposição 10.17, obtemos

$$A = \sum_{k=1}^{r} \alpha_k \left(\prod_{\substack{j=1 \\ j\neq k}}^{r} \frac{1}{\alpha_k^{-1} - \alpha_j^{-1}} \right) \prod_{\substack{l=1 \\ l\neq k}}^{r} \left(A^{-1} - \alpha_l^{-1}\mathbb{1} \right) . \tag{10.66}$$

Comparando (10.65) a (10.66) e evocando a unicidade da representação espectral de A, concluímos pela validade de (10.62) para cada $k \in \{1, \ldots, r\}$. ■

E. 10.22 *Exercício*. Seja $A \in \mathrm{Mat}(\mathbb{C}, n)$ uma matriz diagonalizável e invertível com apenas dois autovalores distintos, α_1 e α_2. Usando (10.62) ou (10.64) mostre que

$$A^{-1} = \frac{1}{\alpha_1\alpha_2}\Big((\alpha_1 + \alpha_2)\mathbb{1} - A \Big) . \tag{10.67}$$

Essa relação não é geralmente válida para matrizes não diagonalizáveis e invertíveis com apenas dois autovalores distintos. A matriz $\left(\begin{smallmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{smallmatrix}\right)$ tem autovalores $+1$ e -1, é invertível, não é diagonalizável e não satisfaz (10.67). Verifique! Prove (10.67) diretamente do Teorema Espectral. ✠

• O Teorema Espectral para matrizes. Uma segunda visita

O Teorema Espectral, Teorema 10.7, pode ser formulado de um modo mais detalhado (Teorema 10.9). A principal utilidade dessa outra formulação é a de fornecer mais informações sobre os projetores espectrais E_a (vide expressão (10.70), abaixo). Obtém-se também nessa nova formulação mais condições necessárias e suficientes à diagonalizabilidade e que podem ser úteis, como veremos, por exemplo, no Teorema 10.22 provado adiante (página 503). No teorema a seguir e em sua demonstração seguimos parcialmente [124].

Teorema 10.9 (Teorema Espectral para Matrizes. Versão Detalhada) *Seja $A \in \mathrm{Mat}(\mathbb{C}, n)$. São equivalentes as seguintes afirmações:*

1. *A possui n autovetores linearmente independentes, ou seja, o subespaço gerado pelos autovetores de A tem dimensão n.*
2. *A é diagonalizável, ou seja, existe uma matriz $P \in \mathrm{Mat}(\mathbb{C}, n)$ inversível tal que $P^{-1}AP$ é uma matriz diagonal $\mathrm{diag}(d_1, \ldots, d_n)$, onde os d_i's são autovalores de A.*
3. *Para todo vetor $x \in \mathbb{C}^n$ e todo escalar $\lambda \in \mathbb{C}$ tais que $(A - \lambda\mathbb{1})^2 x = 0$, vale que $(A - \lambda\mathbb{1})x = 0$.*
4. *Se x é um vetor não-nulo tal que $(A - \lambda\mathbb{1})x = 0$ para algum $\lambda \in \mathbb{C}$ então não existe nenhum vetor y com a propriedade que $(A - \lambda\mathbb{1})y = x$.*
5. *Todas as raízes do polinômio mínimo de A têm multiplicidade 1.*
6. *Existem $r \in \mathbb{N}$, escalares distintos $\alpha_1, \ldots, \alpha_r$ e projetores distintos $E_1, \ldots, E_r \in \mathrm{Mat}(\mathbb{C}, n)$, denominados* projetores espectrais *de A, tais que*

$$A = \sum_{a=1}^{r} \alpha_a E_a .$$

Além disso, as matrizes E_a satisfazem

$$\mathbb{1} = \sum_{a=1}^{r} E_a \tag{10.68}$$

e

$$E_i E_j = \delta_{i,\,j} E_j . \tag{10.69}$$

Os projetores espectrais E_k do item 6, acima, podem ser expressos em termos de polinômios da matriz A:

$$E_k \;=\; \frac{1}{m_k(\alpha_k)} m_k(A) \;, \tag{10.70}$$

para todo k, $1 \leq k \leq r$, onde os polinômios m_k são definidos por

$$M(x) \;=\; (x - \alpha_k) m_k(x) \;,$$

M sendo o polinômio mínimo de A. □

Demonstração. A prova da equivalência será feita demonstrando-se sucessivamente as seguintes implicações: $1 \to 2$, $2 \to 3$, $3 \to 4$, $4 \to 5$, $5 \to 6$, $6 \to 1$. Que 1 implica 2 já foi demonstrado no Teorema 10.6, página 492.

$2 \to 3$. Seja $D = P^{-1}AP$ diagonal. $D = \mathrm{diag}\,(d_1, \ldots, d_n)$. Seja $(A - \lambda\mathbb{1})^2 x = 0$. Segue que

$$P^{-1}(A - \lambda\mathbb{1})^2 P y \;=\; 0$$

onde $y = P^{-1}x$. Logo,

$$(D - \lambda\mathbb{1})^2 y \;=\; 0 \;,$$

ou seja, $(d_j - \lambda)^2 y_j = 0$, $j = 1, \ldots, n$, onde y_j são as componentes de y: $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$. Agora, é evidente que se $(d_a - \lambda)^2 y_a = 0$ então $(d_a - \lambda) y_a = 0$. Logo

$$(D - \lambda\mathbb{1}) y \;=\; 0 \;.$$

Usando-se $y = P^{-1}x$ e multiplicando-se à direita por P, concluímos que

$$0 \;=\; P(D - \lambda\mathbb{1})P^{-1}x \;=\; (PDP^{-1} - \lambda\mathbb{1})x \;=\; (A - \lambda\mathbb{1})x \;,$$

que é o que queríamos provar.

$3 \to 4$. A prova é feita por contradição. Vamos supor que para algum vetor $x \neq 0$ exista $\lambda \in \mathbb{C}$ tal que $(A - \lambda\mathbb{1})x = 0$. Suponhamos também que exista vetor y tal que $(A - \lambda\mathbb{1})y = x$. Teríamos

$$(A - \lambda\mathbb{1})^2 y = (A - \lambda\mathbb{1})x = 0 \;.$$

Pelo item 3 isso implica $(A - \lambda\mathbb{1})y = 0$. Mas isso diz que $x = 0$, uma contradição.

$4 \to 5$. Seja M o polinômio mínimo de A, ou seja, o polinômio mônico[17] de menor grau tal que $M(A) = 0$. Vamos mostrar que todas as raízes de M têm multiplicidade 1. Vamos, por contradição, supor que haja uma raiz, λ_0, com multiplicidade maior ou igual a 2. Teríamos, para $x \in \mathbb{C}$,

$$M(x) \;=\; p(x)(x - \lambda_0)^2 \;.$$

Assim, $M(A) = p(A)(A - \lambda_0\mathbb{1})^2 = 0$. Como M é, por definição, o polinômio de menor grau que zera em A, segue que

$$p(A)(A - \lambda_0\mathbb{1}) \;\neq\; 0 \;.$$

Assim, existe pelo menos um vetor z tal que $p(A)(A - \lambda_0\mathbb{1})z \neq 0$. Vamos definir um vetor x por $x := p(A)(A - \lambda_0\mathbb{1})z$. Então,

$$(A - \lambda_0\mathbb{1})x \;=\; (A - \lambda_0\mathbb{1})p(A)(A - \lambda_0\mathbb{1})z \;=\; p(A)(A - \lambda_0\mathbb{1})^2 z \;=\; M(A)z \;=\; 0 \;,$$

pois $M(A) = 0$. Agora, pela definição,

$$x \;=\; (A - \lambda_0\mathbb{1})y \;,$$

onde $y = p(A)z$. Pelo item 4, porém, isso é impossível.

[17] A definição de polinômio mônico está à página 485.

5 → 6. Pela hipótese que as raízes de M são simples segue da expressão (10.49) da Proposição 10.15, página 490, que para $x \in \mathbb{C}$,

$$M(x) \;=\; \prod_{j=1}^{r}(x-\alpha_j) \;,$$

onde α_j são as raízes de M e que coincidem com os r autovalores distintos de A. Para $k = 1, \ldots, r$ defina-se os polinômios m_k por

$$M(x) \;=:\; (x-\alpha_k)m_k(x) \;,$$

ou seja,

$$m_k(x) \;:=\; \prod_{\substack{j=1\\ j\neq k}}^{r}(x-\alpha_j) \;.$$

É claro que $m_k(\alpha_j) = 0 \iff j \neq k$ (por quê?).

Vamos agora definir mais um polinômio, g, da seguinte forma:

$$g(x) \;=\; 1 - \sum_{k=1}^{r}\frac{1}{m_k(\alpha_k)}m_k(x) \;.$$

Como os polinômios m_k têm grau $r-1$, o polinômio g tem grau menor ou igual a $r-1$. Porém, observe-se que, para todos os α_j, $j = 1, \ldots, r$, vale

$$g(\alpha_j) \;=\; 1 - \sum_{k=1}^{r}\frac{1}{m_k(\alpha_k)}m_k(\alpha_j) = 1 - \frac{m_j(\alpha_j)}{m_j(\alpha_j)} \;=\; 0 \;.$$

Assim, g tem pelo menos r raízes distintas! O único polinômio de grau menor ou igual a $r-1$ que tem r raízes distintas é o polinômio nulo. Logo, concluímos que

$$g(x) \;=\; 1 - \sum_{k=1}^{r}\frac{1}{m_k(\alpha_k)}m_k(x) \;\equiv\; 0$$

para todo $x \in \mathbb{C}$. Isso significa que todos os coeficientes de g são nulos. Assim, para qualquer matriz B tem-se $g(B) = 0$. Para a matriz A isso diz que

$$\mathbb{1} \;=\; \sum_{k=1}^{r}\frac{1}{m_k(\alpha_k)}m_k(A) \;.$$

Definindo-se

$$E_k \;:=\; \frac{1}{m_k(\alpha_k)}m_k(A) \;, \tag{10.71}$$

concluímos que

$$\mathbb{1} \;=\; \sum_{k=1}^{r}E_k \;. \tag{10.72}$$

Para todo k vale $0 = M(A) = (A - \alpha_k\mathbb{1})m_k(A)$, ou seja, $Am_k(A) = \alpha_k m_k(A)$. Pela definição de E_k isso significa

$$AE_k \;=\; \alpha_k E_k \;.$$

Assim, multiplicando-se ambos os lados de (10.72) por A, segue que

$$A \;=\; \sum_{k=1}^{r}\alpha_k E_k \;.$$

Para completar a demonstração de 6, resta-nos provar que $E_iE_j = \delta_{i,\,j}E_j$.

Para $i \neq j$ tem-se pela definição dos E_k's que

$$\begin{aligned} E_i E_j &= \frac{1}{m_i(\alpha_i)m_j(\alpha_j)} m_i(A)m_j(A) \\ &= \frac{1}{m_i(\alpha_i)m_j(\alpha_j)} \left[\prod_{\substack{k=1 \\ k\neq i}}^{r} (A-\alpha_k \mathbb{1})\right] \left[\prod_{\substack{l=1 \\ l\neq j}}^{r} (A-\alpha_l \mathbb{1})\right] \\ &= \frac{1}{m_i(\alpha_i)m_j(\alpha_j)} \left[\prod_{\substack{k=1 \\ k\neq i,\ k\neq j}}^{r} (A-\alpha_k \mathbb{1})\right] \left[\prod_{l=1}^{r} (A-\alpha_l \mathbb{1})\right] \\ &= \frac{1}{m_i(\alpha_i)m_j(\alpha_j)} \left[\prod_{\substack{k=1 \\ k\neq i,\ k\neq j}}^{r} (A-\alpha_k \mathbb{1})\right] M(A) \\ &= 0\,, \end{aligned}$$

pois $M(A) = 0$. Resta-nos provar que $E_j^2 = E_j$ para todo j. Multiplicando-se ambos os lados de (10.72) por E_j teremos

$$E_j \;=\; \sum_{k=1}^{r} E_j E_k \;=\; E_j E_j \;,$$

já que $E_j E_k = 0$ quando $j \neq k$. Isso completa a demonstração do item 6.

$6 \to 1$. Notemos primeiramente que para todo vetor x, os vetores $E_k x$ ou são nulos ou são autovetores de A. De fato, por 6,

$$AE_k x \;=\; \sum_{j=1}^{r} \alpha_j E_j E_k x \;=\; \alpha_k E_k x \;.$$

Logo, ou $E_k x = 0$ ou $E_k x$ é autovetor de A. O espaço gerado pelos autovetores de A obviamente tem dimensão menor ou igual a n. Por (10.72), porém, vale para todo vetor x que

$$x \;=\; \mathbb{1}x \;=\; \sum_{k=1}^{r} E_k x \;.$$

Assim, todo vetor x pode ser escrito como uma combinação linear de autovetores de A, o que significa que o espaço gerado pelos autovetores tem dimensão exatamente igual a n.

Isso completa a demonstração do Teorema 10.9. ■

Destacamos ao leitor o fato de que a expressão (10.70) permite representar os projetores espectrais diretamente em termos da matriz diagonalizável A.

• Diagonalizabilidade de projetores

A proposição abaixo é uma aplicação simples do Teorema 10.9 a projetores. A mesma será usada abaixo quando falarmos de diagonalização simultânea de matrizes.

Proposição 10.21 *Seja $E \in \mathrm{Mat}(\mathbb{C},\ n)$ um projetor, ou seja, tal que $E^2 = E$. Então, E é diagonalizável.* □

Prova. Seja $E \in \mathrm{Mat}(\mathbb{C},\ n)$ um projetor. Definamos $E_1 = E$ e $E_2 = \mathbb{1} - E$. Então, E_2 é também um projetor, pois

$$(E_2)^2 \;=\; (\mathbb{1}-E)^2 \;=\; \mathbb{1} - 2E + E^2 \;=\; \mathbb{1} - 2E + E \;=\; \mathbb{1} - E \;=\; E_2 \;.$$

Tem-se também que $E_1E_2 = 0$, pois $E_1E_2 = E(\mathbb{1} - E) = E - E^2 = E - E = 0$. Fora isso, é óbvio que $\mathbb{1} = E_1 + E_2$ e que $E = \alpha_1 E_1 + \alpha_2 E_2$, com $\alpha_1 = 1$ e $\alpha_2 = 0$. Ora, isso tudo diz que E satisfaz precisamente todas as condições do item 6 do Teorema 10.9. Portanto, pelo mesmo teorema, E é diagonalizável. ■

• **Uma condição suficiente para diagonalizabilidade**

Até agora estudamos condições necessárias e suficientes para que uma matriz seja diagonalizável. Vimos que uma matriz $A \in \mathrm{Mat}\,(\mathbb{C},\ n)$ é diagonalizável se e somente se for simples ou se e somente se tiver n autovetores linearmente independentes ou se e somente se puder ser representada na forma espectral, como em (10.54). Nem sempre, porém, é imediato verificar essas hipóteses, de modo que é útil saber de condições mais facilmente verificáveis e que sejam pelo menos suficientes para garantir diagonalizabilidade. Veremos abaixo que é, por exemplo, suficiente que uma matriz seja autoadjunta ou normal para garantir que ela seja diagonalizável.

Uma outra condição útil é aquela contida na seguinte proposição.

Proposição 10.22 *Se $A \in \mathrm{Mat}\,(\mathbb{C},\ n)$ tem n autovalores distintos, então A é diagonalizável.* □

Prova. Isso é imediato pelas Proposições 10.9 e 10.16, das páginas 478 e 493, respectivamente. ■

Observação. A condição mencionada na última proposição é apenas suficiente, pois há obviamente matrizes diagonalizáveis que não têm autovalores todos distintos. ♣

Outra forma de provar a Proposição 10.22 é a seguinte. Seja $\{\lambda_1,\ \ldots,\ \lambda_n\}$ o conjunto dos n autovalores de A, todos distintos. O polinômio característico de A é $q(x) = (x - \lambda_1)\cdots(x - \lambda_n)$. Como as raízes de q têm, nesse caso, multiplicidade 1, segue pela Proposição 10.15, página 490, que o polinômio mínimo de A, M, coincide com o polinômio característico de A: $q(x) = M(x)$, $\forall x \in \mathbb{C}$. Logo, o polinômio mínimo M de A tem também raízes com multiplicidade 1. Assim, pelo item 5 do Teorema 10.9, página 500, A é diagonalizável.

E. 10.23 Exercício. Demonstre a seguinte afirmação: se os autovalores de uma matriz A são todos iguais, então A é diagonalizável se e somente se for um múltiplo de $\mathbb{1}$. Sugestão: use o Teorema Espectral ou a forma geral do polinômio mínimo (10.49). ✠

Segue da afirmativa desse exercício que matrizes triangulares superiores com diagonal principal constante, ou seja, da forma

$$A = \begin{pmatrix} \alpha & A_{12} & \ldots & A_{1(n-1)} & A_{1n} \\ 0 & \alpha & \ldots & A_{2(n-1)} & A_{2n} \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \ldots & \alpha & A_{(n-1)n} \\ 0 & 0 & \ldots & 0 & \alpha \end{pmatrix},$$

só são diagonalizáveis se todos os elementos acima da diagonal principal forem nulos, ou seja, se $A_{ij} = 0$, $\forall j > i$. Naturalmente, a mesma afirmativa é válida para matrizes da forma A^T, triangulares inferiores com diagonal principal constante.

10.4.1 Diagonalização Simultânea de Matrizes

Uma matriz $A \in \mathrm{Mat}\,(\mathbb{C},\ n)$ é dita ser diagonalizada por uma matriz $P \in \mathrm{Mat}\,(\mathbb{C},\ n)$ se $P^{-1}AP$ for uma matriz diagonal.

Uma questão muito importante é saber quando duas matrizes diagonalizáveis podem ser diagonalizadas por uma mesma matriz P. A resposta é fornecida no próximo teorema.

Teorema 10.10 (Diagonalização Simultânea de Matrizes) *Duas matrizes diagonalizáveis A e $B \in \mathrm{Mat}(\mathbb{C}, n)$ podem ser diagonalizadas pela mesma matriz $P \in \mathrm{Mat}(\mathbb{C}, n)$ se e somente se $AB = BA$, ou seja, se e somente se comutarem entre si.* □

Prova. A parte fácil da demonstração é provar que se A e B podem ser diagonalizadas pela mesma matriz P então A e B comutam entre si. De fato $P^{-1}(AB - BA)P = (P^{-1}AP)(P^{-1}BP) - (P^{-1}BP)(P^{-1}AP) = 0$, pois $P^{-1}AP$ e $P^{-1}BP$ são ambas diagonais e matrizes diagonais sempre comutam entre si (por quê?). Assim, $P^{-1}(AB-BA)P = 0$ e, portanto, $AB = BA$.

Vamos agora passar a mostrar que se $AB = BA$ então ambas são diagonalizáveis por uma mesma matriz P. Sejam $\alpha_1, \ldots, \alpha_r$ os r autovalores distintos de A e $\beta_1, \ldots, \beta_s$ os s autovalores distintos de B. Evocando o teorema espectral, A e B podem ser escritos de acordo com suas decomposições espectrais como

$$A = \sum_{i=1}^{r} \alpha_i E_i^A \qquad \text{e} \qquad B = \sum_{j=1}^{s} \beta_j E_j^B ,$$

onde, de acordo com (10.70),

$$E_i^A = \left\{ \prod_{\substack{k=1 \\ k\neq i}}^{r} (\alpha_i - \alpha_k) \right\}^{-1} \left[\prod_{\substack{k=1 \\ k\neq i}}^{r} (A - \alpha_k \mathbb{1}) \right] , \qquad i = 1, \ldots, r \tag{10.73}$$

e

$$E_j^B = \left\{ \prod_{\substack{k=1 \\ k\neq j}}^{s} (\beta_j - \beta_k) \right\}^{-1} \left[\prod_{\substack{k=1 \\ k\neq j}}^{s} (B - \beta_k \mathbb{1}) \right] , \qquad j = 1, \ldots, s . \tag{10.74}$$

Como A e B comutam entre si e como E_i^A e E_j^B, dados em (10.73)–(10.74), são polinômios em A e B, respectivamente, segue que E_i^A e E_j^B também comutam entre si para todo i e todo j.

Com isso, vamos definir

$$Q_{i,j} = E_i^A E_j^B = E_j^B E_i^A$$

para $i = 1, \ldots, r$ e $j = 1, \ldots, s$.

Note-se que os $Q_{i,j}$'s são projetores pois

$$Q_{i,j}^2 = (E_i^A E_j^B)(E_i^A E_j^B) = (E_i^A)^2 (E_j^B)^2 = E_i^A E_j^B = Q_{i,j} .$$

Fora isso, é fácil ver que,

$$Q_{i,j} Q_{k,l} = \delta_{i,k} \delta_{j,l} \, Q_{i,j} . \tag{10.75}$$

E. 10.24 *Exercício*. Mostre isso. ✠

Note-se também que

$$\mathbb{1} = \sum_{i=1}^{r} \sum_{j=1}^{s} Q_{i,j} , \tag{10.76}$$

pois

$$\sum_{i=1}^{r} \sum_{j=1}^{s} Q_{i,j} = \sum_{i=1}^{r} \sum_{j=1}^{s} E_i^A E_j^B = \left(\sum_{i=1}^{r} E_i^A \right) \left(\sum_{j=1}^{s} E_j^B \right) = \mathbb{1}\mathbb{1} = \mathbb{1} .$$

Afirmamos que podemos escrever

$$A = \sum_{i=1}^{r} \sum_{j=1}^{s} \gamma_{i,j}^A Q_{i,j} \tag{10.77}$$

e

$$B = \sum_{i=1}^{r}\sum_{j=1}^{s} \gamma_{i,\,j}^{B} Q_{i,\,j} , \tag{10.78}$$

onde $\gamma_{i,\,j}^A = \alpha_i$ e $\gamma_{i,\,j}^B = \beta_j$. De fato, com essas definições,

$$\sum_{i=1}^{r}\sum_{j=1}^{s} \gamma_{i,\,j}^{A} Q_{i,\,j} = \sum_{i=1}^{r}\sum_{j=1}^{s} \alpha_i E_i^A E_j^B = \left(\sum_{i=1}^{r} \alpha_i E_i^A\right)\left(\sum_{j=1}^{s} E_j^B\right) = A\mathbb{1} = A .$$

Para B a demonstração é análoga.

Nas relações (10.77) e (10.78) é possível fazer simplificações em função do fato de que nem todos os projetores $Q_{i,\,j}$ são não nulos. Seja $\mathcal{Q}_1 \ \ldots,\ \mathcal{Q}_t$ a lista dos projetores $Q_{i,\,j}$ não nulos, ou seja,

$$\{\mathcal{Q}_1 \ \ldots,\ \mathcal{Q}_t\} = \{Q_{i,\,j} |\ Q_{i,\,j} \neq 0,\ i = 1, \ldots, r \text{ e } j = 1, \ldots, s\} .$$

É evidente por (10.75) que os $\mathcal{Q}_k$'s são projetores e que

$$\mathcal{Q}_k \mathcal{Q}_l = \delta_{k,\,l}\, \mathcal{Q}_k .$$

Por (10.76), tem-se

$$\mathbb{1} = \sum_{k=1}^{t} \mathcal{Q}_k \tag{10.79}$$

e por (10.77) e (10.78)

$$A = \sum_{k=1}^{t} \chi_k^A\, \mathcal{Q}_k \tag{10.80}$$

$$B = \sum_{k=1}^{t} \chi_k^B\, \mathcal{Q}_k \tag{10.81}$$

onde as constantes χ_k^A e χ_k^B estão relacionadas de modo óbvio com $\gamma_{i,\,j}^A$ e $\gamma_{i,\,j}^B$, respectivamente.

Em (10.80) e (10.81) vemos que A e B, por serem diagonalizáveis e por comutarem entre si, têm decomposições espectrais com os mesmos projetores espectrais. Note-se também que, pela observação feita no tópico **Projetores**, à página 494 (vide equação (10.53)), tem-se $1 \leq t \leq n$.

Vamos agora completar a demonstração que A e B podem ser diagonalizados por uma mesma matriz inversível P.

Seja $\mathcal{E}_k$ o subespaço dos autovetores de $\mathcal{Q}_k$ com autovalor 1. Subespaços $\mathcal{E}_k$'s diferentes têm em comum apenas o vetor nulo. De fato, se $k \neq l$ e w é um vetor tal que $\mathcal{Q}_k w = w$ e $\mathcal{Q}_l w = w$ então, como $\mathcal{Q}_k\mathcal{Q}_l = 0$ segue que

$$0 = (\mathcal{Q}_k\mathcal{Q}_l)w = \mathcal{Q}_k(\mathcal{Q}_l w) = \mathcal{Q}_k w = w .$$

Seja d_k a dimensão do subespaço $\mathcal{E}_k$ e seja $u_k^1, \ \ldots,\ u_k^{d_k}$ um conjunto de d_k vetores linearmente independentes em $\mathcal{E}_k$. Notemos que d_k coincide com a multiplicidade algébrica do autovalor 1 de $\mathcal{Q}_k$, pois, conforme diz a Proposição 10.21, o projetor $\mathcal{Q}_k$ é diagonalizável e, portanto, é uma matriz simples (Proposição 10.16). Como $\mathbb{1} = \sum_{k=1}^{t} \mathcal{Q}_k$, tem-se, tomando-se o traço, que $n = \sum_{k=1}^{t} d_k$. Pelas definições, temos que

$$\mathcal{Q}_l u_k^a = \delta_{k,\,l}\, u_k^a , \tag{10.82}$$

pois $\mathcal{Q}_k u_k^a = u_k^a$ e, portanto, $\mathcal{Q}_l u_k^a = \mathcal{Q}_l(\mathcal{Q}_k u_k^a) = (\mathcal{Q}_l\mathcal{Q}_k)u_k^a = 0$ para $k \neq l$.

Afirmamos que o conjunto de vetores

$$u_1^1, \ \ldots,\ u_1^{d_1},\ u_2^1, \ \ldots,\ u_2^{d_2}, \ \ldots u_t^1, \ \ldots,\ u_t^{d_t} \tag{10.83}$$

é formado por n vetores linearmente independentes. De fato, suponha que existam constantes $c_{k,\,j}$ tais que

$$\sum_{k=i}^{t}\sum_{j=1}^{d_k} c_{k,\,j}\, u_k^j = 0 .$$

Aplicando-se à direita $\mathcal{Q}_l$ teríamos $\sum_{j=1}^{d_l} c_{l,\,j} u_l^j = 0$, o que só é possível se $c_{l,\,j} = 0$ para todo j pois $u_l^1, \ldots, u_l^{d_l}$, foram escolhidos linearmente independentes. Como l é arbitrário, concluímos que $c_{l,\,j} = 0$ para todo l e todo j, o que mostra que o conjunto de vetores em (10.83) é linearmente independente.

Seja então a matriz $P \in \text{Mat}\,(\mathbb{C},\, n)$ definida por

$$P \;=\; \Big[\!\!\Big[u_1^1,\, \ldots,\, u_1^{d_1},\, u_2^1,\, \ldots,\, u_2^{d_2},\, \ldots u_t^1,\, \ldots,\, u_t^{d_t} \Big]\!\!\Big] \,.$$

P é inversível pois o conjunto (10.83) é linearmente independente (e, portanto, $\det(P) \neq 0$).

Tem-se,

$$AP \;=\; \Big[\!\!\Big[Au_1^1,\, \ldots,\, Au_1^{d_1},\, Au_2^1,\, \ldots,\, Au_2^{d_2},\, \ldots,\, Au_t^1,\, \ldots,\, Au_t^{d_t} \Big]\!\!\Big] \,.$$

Escrevendo $A = \sum_{l=1}^{t} \chi_l^A \, \mathcal{Q}_l$ (10.80) e usando (10.82), temos

$$Au_k^a \;=\; \sum_{l=1}^{t} \chi_l^A \, \mathcal{Q}_l u_k^a \;=\; \chi_k^A u_k^a \,.$$

Assim,

$$AP \;=\; \Big[\!\!\Big[\chi_1^A u_1^1,\, \ldots,\, \chi_1^A u_1^{d_1},\, \chi_2^A u_1^1,\, \ldots,\, \chi_2^A u_1^{d_1},\, \ldots,\, \chi_t^A u_t^1,\, \ldots,\, \chi_t^A u_t^{d_t} \Big]\!\!\Big] \;=\; PD_A \,,$$

onde

$$D_A \;=\; \text{diag}\left(\underbrace{\chi_1^A,\, \ldots,\, \chi_1^A}_{d_1 \text{ vezes}},\; \underbrace{\chi_2^A,\, \ldots,\, \chi_2^A}_{d_2 \text{ vezes}},\; \ldots,\; \underbrace{\chi_t^A,\, \ldots,\, \chi_t^A}_{d_t \text{ vezes}} \right) .$$

Portanto, $P^{-1}AP = D_A$. Analogamente,

$$BP \;=\; \Big[\!\!\Big[Bu_1^1,\, \ldots,\, Bu_1^{d_1},\, Bu_2^1,\, \ldots,\, Bu_2^{d_2},\, \ldots Bu_t^1,\, \ldots,\, Bu_t^{d_t} \Big]\!\!\Big] \,.$$

Escrevendo $B = \displaystyle\sum_{l=1}^{t} \chi_l^B \, \mathcal{Q}_l$ (10.81) temos,

$$BP \;=\; \Big[\!\!\Big[\chi_1^B u_1^1,\, \ldots,\, \chi_1^B u_1^{d_1},\, \chi_2^B u_2^1,\, \ldots,\, \chi_2^B u_2^{d_2},\, \ldots,\, \chi_t^B u_t^1,\, \ldots,\, \chi_t^B u_t^{d_t} \Big]\!\!\Big] \;=\; PD_B \,,$$

onde

$$D_B \;=\; \text{diag}\left(\underbrace{\chi_1^B,\, \ldots,\, \chi_1^B}_{d_1 \text{ vezes}},\; \underbrace{\chi_2^B,\, \ldots,\, \chi_2^B}_{d_2 \text{ vezes}},\; \ldots,\; \underbrace{\chi_t^B,\, \ldots,\, \chi_t^B}_{d_t \text{ vezes}} \right) .$$

Portanto, $P^{-1}BP = D_B$. Isso provou que A e B são diagonalizáveis pela mesma matriz inversível P. A demonstração do Teorema 10.10 está completa. ■

10.5 Matrizes Autoadjuntas, Normais e Unitárias

• **A adjunta de uma matriz**

Seja V um espaço vetorial dotado de um produto escalar $\langle \cdot,\, \cdot \rangle$ e seja $A : V \to V$ um operador linear. Um operador linear A^* que para todos $u,\, v \in V$ satisfaça

$$\langle u,\, Av \rangle \;=\; \langle A^* u,\, v \rangle$$

é dito ser o *operador adjunto* de A. Em espaços vetoriais gerais não é óbvio (e nem sempre verdadeiro!) que sempre exista o adjunto de um operador linear A dado. Há muitos casos, porém, nos quais isso pode ser garantido[18]. Aqui trataremos do caso dos espaços $V = \mathbb{C}^n$ com o produto escalar usual.

[18] Tal é o caso dos chamados operadores lineares limitados agindo em espaços de Hilbert, para os quais sempre é possível garantir a existência do adjunto.

Sejam $u = (u_1, \ldots, u_n)$ e $v = (v_1, \ldots, v_n)$ dois vetores de $\mathbb{C}^n$ para os quais define-se o produto escalar usual

$$\langle u, v\rangle = \sum_{k=1}^{n} \overline{u_k} v_k .$$

Um operador linear A é representado (na base canônica) por uma matriz cujos elementos de matriz são A_{ij}, com $i, j \in \{1, \ldots, n\}$.

É um exercício simples (faça!) verificar que o operador adjunto A^* de A é representado (na base canônica) por uma matriz cujos elementos de matriz são $(A^*)_{ij} = \overline{A_{ji}}$, com $i, j \in \{1, \ldots, n\}$. Ou seja, a matriz adjunta de A é obtida (na base canônica!) transpondo-se A e tomando-se o complexo conjugado de seus elementos.

Os seguintes fatos são importantes:

Proposição 10.23 *Se A e B são dois operadores lineares agindo em $\mathbb{C}^n$ então*

$$(\alpha A + \beta B)^* = \overline{\alpha} A^* + \overline{\beta} B^*$$

para todos $\alpha, \beta \in \mathbb{C}$. Fora isso,

$$(AB)^* = B^* A^* .$$

Por fim, vale para todo A que $(A^*)^* = A$. □

Deixamos a demonstração como exercício para o leitor.

A operação $\mathrm{Mat}(\mathbb{C}, n) \ni A \mapsto A^* \in \mathrm{Mat}(\mathbb{C}, n)$ é denominada *operação de adjunção de matrizes*. Como vimos na Proposição 10.23, a operação de adjunção é antilinear e é um anti-homomorfismo algébrico.

• Os espectro e a operação de adjunção

Seja $A \in \mathrm{Mat}(\mathbb{C}, n)$. Como já vimos, o espectro de A, $\sigma(A)$, é o conjunto de raízes de seu polinômio característico, definido por $p_A(z) = \det(z\mathbb{1} - A)$, $z \in \mathbb{C}$. Como para toda $B \in \mathrm{Mat}(\mathbb{C}, n)$ vale $\det(B^*) = \overline{\det(B)}$ (por quê?), segue que $\overline{p_A(\overline{z})} = \overline{\det(\overline{z}\mathbb{1} - A)} = \det(z\mathbb{1} - A^*) = p_{A^*}(z)$, ou seja, $p_{A^*}(z) = \overline{p_A(\overline{z})}$. Com isso, provamos a seguinte afirmação:

Proposição 10.24 *Seja $A \in \mathrm{Mat}(\mathbb{C}, n)$. Então, $\lambda \in \sigma(A)$ se e somente se $\overline{\lambda} \in \sigma(A^*)$, ou seja, λ é um autovalor de A se e somente se $\overline{\lambda}$ é um um autovalor de A^*.* ■

Em símbolos, as afirmações acima são expressas pela igualdade $\sigma(A) = \overline{\sigma(A^*)}$.

• Matrizes Hermitianas, normais e unitárias

Vamos agora a algumas definições muito importantes.

Definição. Um operador linear em $\mathbb{C}^n$ é dito ser *simétrico, Hermitiano ou autoadjunto* se $A = A^*$, ou seja, se para todos $u, v \in V$ satisfizer

$$\langle u, Av\rangle = \langle Au, v\rangle .$$

♠

Advertência. Em espaços vetoriais de dimensão finita as noções de operador simétrico, Hermitiano ou autoadjunto são sinônimas. Em espaços vetoriais de dimensão infinita, porém, há uma distinção entre essas noções relativa a problemas com o domínio de definição de operadores.

Definição. Um operador linear em $\mathbb{C}^n$ é dito ser *normal* se $AA^* = A^*A$. Ou seja, A é normal se comuta com seu adjunto. ♠

Definição. Um operador linear em $\mathbb{C}^n$ é dito ser *unitário* se $A^*A = AA^* = \mathbb{1}$. É claro que todo operador unitário é normal e que um operador é unitário em $\mathbb{C}^n$ se e somente se $A^* = A^{-1}$. Note que se A é unitário então, para todos $u, v \in V$, tem-se

$$\langle Au, Av\rangle = \langle u, v\rangle .$$

♠

Definição. Se A é um operador linear em $\mathbb{C}^n$ define-se a *parte real* de A por

$$\text{Re}\,(A) = \frac{1}{2}(A + A^*)$$

e a *parte imaginária* de A por

$$\text{Im}\,(A) = \frac{1}{2i}(A - A^*).$$

É claro que essas definições foram inspiradas nas relações análogas para números complexos. Note também que

$$A \;=\; \text{Re}\,(A) + i\text{Im}\,(A)\;.$$

♠

E. 10.25 *Exercício*. Por quê? ✠

É importante notar que para qualquer operador linear A em $\mathbb{C}^n$ sua parte real e imaginária são ambas operadores Hermitianos: $(\text{Re}\,(A))^* = \text{Re}\,(A)$ e $(\text{Im}\,(A))^* = \text{Im}\,(A)$.

E. 10.26 *Exercício*. Mostre isso. ✠

Para operadores normais tem-se a seguinte proposição, que será útil adiante e serve como caracterização alternativa do conceito de operador normal.

Proposição 10.25 *Um operador linear agindo em $\mathbb{C}^n$ é normal se e somente se sua parte real comuta com sua parte imaginária.* □

Deixamos a demonstração (elementar) como exercício para o leitor.

A importância das definições acima reside no seguinte fato, que demonstraremos adiante: matrizes Hermitianas e matrizes normais são diagonalizáveis. Antes de tratarmos disso, vamos discutir algumas propriedades do espectro de matrizes Hermitianas e de matrizes unitárias.

• **Os autovalores de matrizes Hermitianas e de matrizes unitárias**

Os seguintes teoremas têm importância fundamental para o estudo de propriedades de matrizes Hermitianas e de matrizes unitárias.

Teorema 10.11 *Os autovalores de uma matriz Hermitiana são sempre números reais.* □

Prova. Seja A Hermitiana, λ um autovalor de A e $v \neq 0$ um autovetor de A com autovalor λ. Como A é Hermitiana tem-se

$$\langle v,\; Av\rangle \;=\; \langle Av,\; v\rangle\;.$$

Como v é um autovetor, o lado esquerdo vale $\lambda\langle v,\; v\rangle$ e o lado direito vale $\overline{\lambda}\langle v,\; v\rangle$. Logo, $(\lambda - \overline{\lambda})\langle v,\; v\rangle = 0$. Como $v \neq 0$ isso implica $\lambda = \overline{\lambda}$, ou seja, λ é real. ■

Note-se que a recíproca desse teorema é falsa. A matriz $\begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix}$ tem autovalores reais (2 e 3) mas não é Hermitiana.

Para matrizes unitárias temos

Teorema 10.12 *Os autovalores de uma matriz unitária são sempre números complexos de módulo* 1. □

Prova. Seja A unitária, λ um autovalor de A e $v \neq 0$ um autovetor de A com autovalor λ. Como A é unitária tem-se $\langle Av, \, Av\rangle = \langle v, \, v\rangle$. Como v é um autovetor, o lado esquerdo vale $\overline{\lambda}\lambda\langle v, \, v\rangle$. Assim, $(|\lambda|^2 - 1)\langle v, \, v\rangle = 0$. Como $v \neq 0$ isso implica $|\lambda| = 1$. ■

• Operadores simétricos e unitários. Ortogonalidade de autovetores

Teorema 10.13 *Os autovetores associados a autovalores distintos de uma matriz simétrica são ortogonais entre si.* □

Prova. Seja A simétrica e λ_1, λ_2 dois de seus autovalores, que suporemos distintos. Seja v_1 autovetor de A com autovalor λ_1 e v_2 autovetor de A com autovalor λ_2. Temos, por A ser simétrico, $\langle v_1, \, Av_2\rangle = \langle Av_1, \, v_2\rangle$. O lado esquerdo vale $\lambda_2\langle v_1, \, v_2\rangle$ e o lado direito $\lambda_1\langle v_1, \, v_2\rangle$ (lembre-se que λ_1 é real). Assim $(\lambda_2 - \lambda_1)\langle v_1, \, v_2\rangle = 0$. Como $\lambda_2 \neq \lambda_1$, segue que $\langle v_1, \, v_2\rangle = 0$, que é o que se queria provar. ■

Teorema 10.14 *Os autovetores associados a autovalores distintos de uma matriz unitária são ortogonais entre si.* □

Prova. Seja U unitária e sejam λ_1, λ_2 dois de seus autovalores, sendo que suporemos $\lambda_1 \neq \lambda_2$. Seja v_1 autovetor de U com autovalor λ_1 e v_2 autovetor de U com autovalor λ_2. Temos, por U ser unitário, $\langle Uv_1, \, Uv_2\rangle = \langle v_1, \, U^*Uv_2\rangle = \langle v_1, \, v_2\rangle$. O lado esquerdo vale $\lambda_2\overline{\lambda_1}\langle v_1, \, v_2\rangle = \frac{\lambda_2}{\lambda_1}\langle v_1, \, v_2\rangle$ (lembre-se que λ_1 é um número complexo de módulo 1 e, portanto $\overline{\lambda_1} = \lambda_1^{-1}$). Assim

$$\left(\frac{\lambda_2}{\lambda_1} - 1\right)\langle v_1, \, v_2\rangle \; = \; 0 \; .$$

Como $\lambda_2 \neq \lambda_1$, segue que $\langle v_1, \, v_2\rangle = 0$, que é o que se queria provar. ■

• Projetores ortogonais

Um operador linear E agindo em $\mathbb{C}^n$ é dito ser um *projetor ortogonal* se $E^2 = E$ e se $E^* = E$.

Projetores ortogonais são importantes na decomposição espectral de matrizes autoadjuntas, como veremos.

Note-se que nem todo projetor é ortogonal. Por exemplo $E = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ é um projetor ($E^2 = E$) mas não é ortogonal ($E^* \neq E$). O mesmo vale para $E = \begin{pmatrix} 1 & 0 \\ 2 & 0 \end{pmatrix}$.

E. 10.27 *Exercício*. Mostre que uma matriz complexa 2×2 é um projetor ortogonal se e somente se ou for a matriz identidade $\mathbb{1}$ ou se for da forma $\frac{1}{2}\big(\mathbb{1} + \vec{a} \cdot \vec{\sigma}\big)$, com $\vec{a} \equiv (a_1, \, a_2, \, a_3) \in \mathbb{R}^3$ e $\big\|\vec{a}\big\| = 1$. Aqui, $\vec{a} \cdot \vec{\sigma} := a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3$, com σ_k sendo as matrizes de Pauli, cuja definição e cujas propriedades básicas encontram-se no Exercício E. 11.26, página 625. ✠

Um exemplo importante de projetor ortogonal é representado por projetores sobre subespaços unidimensionais gerados por vetores. Seja v um vetor cuja norma assumiremos ser 1, ou seja, $\|v\| = \sqrt{\langle v, \, v\rangle} = 1$. Definimos o projetor P_v sobre o subespaço gerado por v por

$$P_v u \; := \; \langle v, \, u\rangle \, v \; , \tag{10.84}$$

para todo vetor u. Provemos que P_v é um projetor ortogonal. Por um lado, tem-se

$$P_v^2 u \; = \; \langle v, \, u\rangle \, P_v v \; = \; \langle v, \, u\rangle \, \langle v, \, v\rangle \, v \; = \; \langle v, \, u\rangle \, v \; = \; P_v u \; ,$$

o que mostra que $P_v^2 = P_v$. Por outro lado, para quaisquer vetores a e b, usando as propriedades de linearidade, antilinearidade e conjugação complexa do produto escalar, tem-se

$$\langle a, \, P_v b\rangle \; = \; \Big\langle a, \, \langle v, \, b\rangle \, v\Big\rangle \; = \; \langle v, \, b\rangle \, \langle a, \, v\rangle \; = \; \Big\langle \overline{\langle a, \, v\rangle} \, v, \, b\Big\rangle \; = \; \Big\langle \langle v, \, a\rangle \, v, \, b\Big\rangle \; = \; \langle P_v a, \, b\rangle \; ,$$

provando que $P_v^* = P_v$. Isso mostra que P_v é um projetor ortogonal.

Um fato crucial sobre projetores como P_v é o seguinte. Se u e v são dois vetores ortogonais, ou seja, se $\langle u, v\rangle = 0$ então $P_u P_v = P_v P_u = 0$. Para provar isso notemos que para qualquer vetor a vale

$$P_u(P_v a) = P_u\big(\langle v, a\rangle v\big) = \langle v, a\rangle P_u v = \langle v, a\rangle \langle u, v\rangle u = 0 .$$

O mesmo se passa para $P_v(P_u a)$.

• Matrizes autoadjuntas e diagonalizabilidade

Vamos aqui demonstrar a seguinte afirmação importante: toda matriz autoadjunta é diagonalizável. Uma outra demonstração (eventualmente mais simples) dessa afirmação pode ser encontrada na Seção 10.8.3, página 549. Vide Teorema 10.30, página 551.

Teorema 10.15 (Teorema Espectral para Matrizes Autoadjuntas) *Se $A \in \mathrm{Mat}(\mathbb{C}, n)$ é autoadjunta, então A possui n autovetores mutuamente ortonormais $v_1, \ldots, v_n$, com autovalores reais $\lambda_1, \ldots, \lambda_n$, respectivamente, e pode ser representada na forma espectral*

$$A = \lambda_1 P_{v_1} + \cdots + \lambda_n P_{v_n} . \tag{10.85}$$

Os projetores P_{v_k} satisfazem $P_{v_k}^* = P_{v_k}$ para todo k e valem também $P_{v_j} P_{v_k} = \delta_{j\,k} P_{v_k}$, sendo que $\sum_{k=1}^n P_{v_k} = \mathbb{1}$.

Portanto, se A é autoadjunta, então A é diagonalizável, sendo que é possível encontrar uma matriz unitária P que diagonaliza A, ou seja, tal que $P^{-1}AP$ é diagonal e $P^{-1} = P^*$.

Note-se que se $\alpha_1, \ldots, \alpha_r$ com $1 \leq r \leq n$ são os autovalores distintos de A, então (10.85) pode ser reescrita como $A = \alpha_1 P_1 + \cdots + \alpha_r P_r$, onde cada P_k é o projetor ortogonal dado pela soma dos P_{v_j}'s de mesmo autovalor α_k. A Proposição 10.19, página 497, garante a unicidade dessa representação para A. □

Prova do Teorema 10.15. A demonstração que A é diagonalizável será feita construindo-se a representação espectral (10.85) para A. Seja λ_1 um autovalor de A e v_1 um autovetor de A com autovalor λ_1 normalizado de tal forma que $\|v_1\| = 1$. Vamos definir um operador A_1 por

$$A_1 := A - \lambda_1 P_{v_1} .$$

Como A e P_{v_1} são autoadjuntos e λ_1 é real, segue que A_1 é igualmente autoadjunto.

Afirmamos que $A_1 v_1 = 0$ e que $[v_1]^\perp$ é um subespaço invariante por A_1. De fato,

$$A_1 v_1 = A v_1 - \lambda_1 P_{v_1} v_1 = \lambda_1 v_1 - \lambda_1 v_1 = 0 .$$

Fora isso, se $w \in [v_1]^\perp$ tem-se

$$\langle A_1 w, v_1\rangle = \langle w, A_1 v_1\rangle = 0 ,$$

mostrando que $A_1 w$ é também elemento de $[v_1]^\perp$.

O operador A_1 restrito a $[v_1]^\perp$ é também autoadjunto (por que?). Seja λ_2 um de seus autovalores com autovetor $v_2 \in [v_1]^\perp$, que escolhemos com norma 1. Seja

$$A_2 := A_1 - \lambda_2 P_{v_2} = A - \lambda_1 P_{v_1} - \lambda_2 P_{v_2} .$$

Como λ_2 também é real A_2 é igualmente autoadjunto. Fora isso afirmamos que A_2 anula os vetores do subespaço $[v_1, v_2]$ e mantém $[v_1, v_2]^\perp$ invariante. De fato,

$$A_2 v_1 = A v_1 - \lambda_1 P_{v_1} v_1 - \lambda_2 P_{v_2} v_1 = \lambda_1 v_1 - \lambda_1 v_1 - \lambda_2 \langle v_2, v_1\rangle v_2 = 0 ,$$

pois $\langle v_2, v_1\rangle = 0$. Analogamente,

$$A_2 v_2 = A_1 v_2 - \lambda_2 P_{v_2} v_2 = \lambda_2 v_2 - \lambda_2 v_2 = 0 .$$

Por fim, para quaisquer $\alpha, \beta \in \mathbb{C}$ e $w \in [v_1, v_2]^\perp$ tem-se

$$\big\langle A_2 w, (\alpha v_1 + \beta v_2)\big\rangle = \big\langle w, A_2(\alpha v_1 + \beta v_2)\big\rangle = 0 ,$$

que é o que queríamos provar.

Prosseguindo indutivamente, construiremos um conjunto de vetores $v_1, \ldots, v_n$, todos com norma 1 e com $v_a \in [v_1, \ldots, v_{a-1}]^\perp$ e um conjunto de números reais $\lambda_1, \ldots, \lambda_n$ tais que

$$A_n := A - \lambda_1 P_{v_1} - \cdots - \lambda_n P_{v_n}$$

anula-se no subespaço $[v_1, \ldots, v_n]$. Ora, como estamos em um espaço de dimensão n e os vetores v_k são mutuamente ortogonais, segue que $[v_1, \ldots, v_n]$ deve ser o espaço todo, ou seja, $A_n = 0$. Provamos então que

$$A = \lambda_1 P_{v_1} + \cdots + \lambda_n P_{v_n} . \tag{10.86}$$

Vamos provar agora que essa é a representação espectral de A. Como os v_k's são mutuamente ortogonais, é evidente que $P_{v_k} P_{v_l} = \delta_{k,l} P_{v_k}$. Resta-nos provar que $P_{v_1} + \cdots + P_{v_n} = \mathbb{1}$. Como $v_1, \ldots, v_n$ formam uma base, todo vetor x pode ser escrito como uma combinação linear

$$x = \alpha_1 v_1 + \cdots + \alpha_n v_n . \tag{10.87}$$

Tomando-se o produto escalar com v_a, e usando o fato que os v_k's são mutuamente ortogonais, tem-se $\alpha_a = \langle v_a, x\rangle$.

E. 10.28 *Exercício*. Verifique. ✠

Assim, (10.87) pode ser escrita como

$$x = \langle v_1, x\rangle v_1 + \cdots + \langle v_n, x\rangle v_n = P_{v_1}x + \cdots + P_{v_n}x = (P_{v_1} + \cdots + P_{v_n})\, x .$$

Como isso vale para todo vetor x, segue que $P_{v_1} + \cdots + P_{v_n} = \mathbb{1}$. Assim, A possui uma representação espectral como (10.54). Pelo Teorema Espectral 10.7, A é diagonalizável.

Por (10.86), vemos que $Av_a = \lambda_a v_a$ (verifique!). Logo os λ_a's são autovalores de A e os v_a's seus autovetores. Assim, se A é autoadjunto, podemos encontrar n autovetores de A mutuamente ortogonais, mesmo que sejam autovetores com o mesmo autovalor. Isso generaliza o Teorema 10.13.

Pelo que já vimos A é diagonalizada por $P^{-1}AP$, onde podemos escolher $P = [\![v^1, \ldots, v^n]\!]$. É fácil verificar, porém, que P é unitária. De fato, é um exercício simples (faça!) mostrar que

$$P^*P = \begin{pmatrix} \langle v_1, v_1\rangle & \cdots & \langle v_1, v_n\rangle \\ \vdots & \ddots & \vdots \\ \langle v_n, v_1\rangle & \cdots & \langle v_n, v_n\rangle \end{pmatrix} .$$

Como $\langle v_a, v_b\rangle = \delta_{a,b}$, a matriz do lado direito é igual a $\mathbb{1}$, mostrando que $P^*P = PP^* = \mathbb{1}$ e que, portanto, P é unitária. ■

Para concluir essa discussão, temos:

Proposição 10.26 *Uma matriz $A \in \mathrm{Mat}(\mathbb{C}, n)$ é autoadjunta, se e somente se for diagonalizável por uma transformação de similaridade unitária e se seus autovalores forem reais.* □

Prova. Se $A \in \mathrm{Mat}(\mathbb{C}, n)$ é diagonalizável por uma transformação de similaridade unitária e seus autovalores são reais, ou seja, existe P unitária e D diagonal real com $P^*AP = D$, então $A = PDP^*$ e $A^* = PD^*P^*$. Como D é diagonal e real, vale $D^* = D$ e, portanto, $A^* = PDP^* = A$, provando que A é autoadjunta. A recíproca já foi provada acima. ■

• Matrizes normais e diagonalizabilidade

O teorema que afirma que toda matriz simétrica é diagonalizável tem a seguinte consequência:

Teorema 10.16 *Se $A \in \mathrm{Mat}(\mathbb{C},\ n)$ é normal então A é diagonalizável.* □

Prova. Já vimos que toda matriz A pode ser escrita na forma $A = \mathrm{Re}(A) + i\mathrm{Im}(A)$ onde $\mathrm{Re}(A)$ e $\mathrm{Im}(A)$ são autoadjuntas. Vimos também que se A é normal $\mathrm{Re}(A)$ e $\mathrm{Im}(A)$ comutam entre si (Proposição 10.25). Pelo Teorema 10.10, $\mathrm{Re}(A)$ e $\mathrm{Im}(A)$ podem ser simultaneamente diagonalizados. ■

Observação. Como no caso autoadjunto, o operador que faz a diagonalização pode ser escolhido unitário. De fato, vale uma afirmativa ainda mais forte. ♣

Teorema 10.17 *Uma matriz $A \in \mathrm{Mat}(\mathbb{C},\ n)$ é normal se e somente se for diagonalizável por um operador unitário.* □

Prova. Resta provar apenas que se A é diagonalizável por um operador unitário P então A é normal. Seja $D = P^*AP$. Tem-se $D^* = P^*A^*P$ (por quê?). Assim,

$$A^*A - AA^* \ = \ PD^*P^*PDP^* - PDP^*PD^*P^* \ = \ P(D^*D - DD^*)P^* \ = \ 0 \ ,$$

já que D^* e D comutam por serem diagonais (duas matrizes diagonais quaisquer sempre comutam. Por quê?). Isso completa a prova que A é normal. ■

Uma outra demonstração (eventualmente mais simples) dessa afirmação pode ser encontrada na Seção 10.8.3, página 549. Vide Teorema 10.31, página 551.

10.5.1 Matrizes Positivas

Uma matriz $A \in \mathrm{Mat}(\mathbb{C},\ n)$ é dita ser uma *matriz positiva* se $\langle w,\ Aw\rangle \geq 0$ para todo vetor $w \in \mathbb{C}^n$. A seguinte proposição é relevante[19]:

Proposição 10.27 *Se $A \in \mathrm{Mat}(\mathbb{C},\ n)$ é positiva, então A é Hermitiana e tem autovalores não negativos. Reciprocamente, se A é Hermitiana e tem autovalores não negativos, então A é positiva.* □

Prova. A expressão $\omega(u,\ v) := \langle u,\ Av\rangle$, $u,\ v \in \mathbb{C}^n$, define uma forma sesquilinear que, por hipótese, é positiva, ou seja, satisfaz $\omega(u,\ u) \geq 0$ para todo $u \in \mathbb{C}^n$. Pelo Teorema 3.1, página 227, ω é Hermitiana, ou seja, $\omega(u,\ v) \ = \ \overline{\omega(v,\ u)}$, para todos os vetores u e v. Mas isso significa que $\langle u,\ Av\rangle = \overline{\langle v,\ Au\rangle}$, ou seja, $\langle u,\ Av\rangle = \langle Au,\ v\rangle$ para todos os vetores u e v e assim provou-se que $A = A^*$. Uma outra forma de demonstrar isso usa a identidade de polarização. Se A é positiva então, para quaisquer vetores $u,\ v \in \mathbb{C}^n$ vale $\langle (u + i^n v),\ A(u + i^n v)\rangle \geq 0$ para todo $n \in \mathbb{Z}$ e, portanto, $\langle (u + i^n v),\ A(u + i^n v)\rangle$ é um número real. Usando a identidade de polarização, eqs. (3.34)–(3.35), página 237, vale, para

[19] Vários dos resultados que seguem podem ser generalizados para operadores lineares positivos agindo em espaços de Hilbert. Vide Teorema 39.30, página 2133.

quaisquer vetores $u,\ v \in \mathbb{C}^n$,

$$
\begin{aligned}
\langle Av,\ u\rangle \;=\; \overline{\langle u,\ Av\rangle} \;&\stackrel{(3.34)}{=}\; \overline{\frac{1}{4}\sum_{n=0}^{3} i^{-n}\langle (u+i^n v),\ A(u+i^n v)\rangle} \;=\; \frac{1}{4}\sum_{n=0}^{3} i^{n}\langle (u+i^n v),\ A(u+i^n v)\rangle \\
&=\; \frac{1}{4}\sum_{n=0}^{3} i^{-n}i^n i^n\langle (u+i^n v),\ A(u+i^n v)\rangle \\
&\stackrel{\text{sesquilin.}}{=}\; \frac{1}{4}\sum_{n=0}^{3} i^{-n}\langle i^{-n}(u+i^n v),\ Ai^n(u+i^n v)\rangle \\
&=\; \frac{1}{4}\sum_{n=0}^{3} i^{-n}\langle (v+i^{-n} u),\ A((-1)^n v+i^n u)\rangle \\
&=\; \frac{1}{4}\sum_{n=0}^{3} (-1)^n i^{-n}\langle (v+i^{-n} u),\ A(v+i^{-n} u)\rangle \\
&=\; \frac{1}{4}\sum_{n=0}^{3} i^{n}\langle (v+i^{-n} u),\ A(v+i^{-n} u)\rangle \;\stackrel{(3.35)}{=}\; \langle v,\ Au\rangle\ .
\end{aligned}
$$

Assim, $\langle Av,\ u\rangle = \langle v,\ Au\rangle$ para todos $u,\ v \in \mathbb{C}^n$, o que significa que A é Hermitiana. Portanto, por (10.85), podemos escrever $A = \lambda_1 P_{v_1} + \cdots + \lambda_n P_{v_n}$, onde $v_1, \ldots, v_n$ são autovetores mutuamente ortonormais de A com autovalores $\lambda_1, \ldots, \lambda_n$, respectivamente. Disso segue que $\langle v_j,\ Av_j\rangle = \lambda_j$ para todo $j = 1, \ldots, n$. Como o lado esquerdo é ≥ 0, por hipótese, segue que $\lambda_j \geq 0$ para todo $j = 1, \ldots, n$.

Se, reciprocamente, A for autoadjunta com autovalores não negativos, segue de (10.85) e da definição de P_{v_j} em (10.84) que $\langle w,\ Aw\rangle = \sum_{j=1}^{n} \lambda_j |\langle w,\ v_j\rangle|^2 \geq 0$, para todo $w \in \mathbb{C}^n$, provando que A é positiva. ■

O seguinte corolário é imediato.

Corolário 10.4 *Uma matriz $A \in \mathrm{Mat}(\mathbb{C},\ n)$ é positiva se somente se existe uma matriz positiva B (unívoca!) tal que $A = B^2$. As matrizes A e B comutam: $AB = BA$.* □

Demonstração. Se $A = B^2$ com B positiva, então, como B é autoadjunta (pela Proposição 10.27), segue que para todo $w \in \mathbb{C}^n$ vale $\langle w,\ Aw\rangle = \langle w,\ B^2 w\rangle = \langle Bw,\ Bw\rangle = \|Bw\|^2 \geq 0$, provando que A é positiva. Provemos agora a recíproca.

Se A é positiva então, como comentamos na demonstração da Proposição 10.27, A é autoadjunta com representação espectral $A = \lambda_1 P_{v_1} + \cdots + \lambda_n P_{v_n}$, onde $v_1, \ldots, v_n$ são autovetores mutuamente ortonormais de A com autovalores $\lambda_1, \ldots, \lambda_n$, respectivamente, todos não negativos. Defina-se a matriz

$$
B \;:=\; \sqrt{\lambda_1} P_{v_1} + \cdots + \sqrt{\lambda_n} P_{v_n}\ . \tag{10.88}
$$

Como, pela ortonormalidade dos v_j's, vale $P_{v_j} P_{v_k} = \delta_{j,\,k} P_{v_j}$, é fácil ver que $B^2 = \lambda_1 P_{v_1} + \cdots + \lambda_n P_{v_n} = A$. A unicidade de B segue da unicidade da decomposição espectral, Proposição 10.19, página 497. A igualdade $(B^2)B = B(B)^2$ significa $AB = BA$, provando que A e B comutam. ■

Definição. Se A é uma matriz positiva, a (única!) matriz positiva B satisfazendo $B^2 = A$ é frequentemente denotada por $\sqrt{A}$ e denominada *raiz quadrada da matriz A*. Como vimos, $A\sqrt{A} = \sqrt{A}A$. ♠

Lema 10.5 *Se $A \in \mathrm{Mat}(\mathbb{C},\ n)$ é uma matriz positiva e $C \in \mathrm{Mat}(\mathbb{C},\ n)$ satisfaz $CA = AC$ então $C\sqrt{A} = \sqrt{A}C$.* □

Prova. Se C comuta com A, então C comuta com qualquer polinômio em A. Vimos na Proposição 10.18, página 497, que os projetores espectrais de A podem ser escritos como polinômios em A. Assim, C comuta com os projetores espectrais de A e, portanto, com $\sqrt{A}$, devido a (10.88). ■

Uma consequência interessante das considerações acima é a seguinte proposição:

Proposição 10.28 *Toda matriz Hermitiana pode ser escrita como combinação linear de até duas matrizes unitárias. Toda matriz pode ser escrita como combinação linear de até quatro matrizes unitárias.* □

Demonstração. Seja $A \in \mathrm{Mat}(\mathbb{C},\ n)$. Se A é Hermitiana (vamos supor que $A \neq \mathbb{0}$, pois de outra forma não há o que se provar), então, para todo $w \in \mathbb{C}^n$, o produto escalar $\langle w\ A^2 w\rangle$ é um número real e, pela desigualdade de Cauchy-Schwarz, $|\langle w\ A^2 w\rangle| \leq \|A^2\|\,\|w\|^2_{\mathbb{C}^n}$. Assim, $-\|A^2\|\,\|w\|^2_{\mathbb{C}^n} \leq \langle w,\ A^2 w\rangle \leq \|A^2\|\,\|w\|^2_{\mathbb{C}^n}$ Logo, a matriz $\mathbb{1} - A^2/\|A^2\|$ é positiva, pois $\langle w,\ (\mathbb{1} - A^2/\|A^2\|)w\rangle = \|w\|^2_{\mathbb{C}^n} - \langle w,\ A^2 w\rangle/\|A^2\| \geq \|w\|^2_{\mathbb{C}^n} - \|w\|^2_{\mathbb{C}^n} = 0$. Consequentemente, $\sqrt{\mathbb{1} - A^2/\|A^2\|}$ existe e é positiva e Hermitiana. Trivialmente, podemos escrever

$$A \ = \ \frac{\sqrt{\|A^2\|}}{2}\left(\frac{A}{\sqrt{\|A^2\|}} + i\sqrt{\mathbb{1} - \frac{A^2}{\|A^2\|}}\right) + \frac{\sqrt{\|A^2\|}}{2}\left(\frac{A}{\sqrt{\|A^2\|}} - i\sqrt{\mathbb{1} - \frac{A^2}{\|A^2\|}}\right) . \tag{10.89}$$

Agora, as matrizes $\frac{A}{\sqrt{\|A^2\|}} \pm i\sqrt{\mathbb{1} - \frac{A^2}{\|A^2\|}}$ são unitárias. Para ver isso, notemos que

$$\left(\frac{A}{\sqrt{\|A^2\|}} + i\sqrt{\mathbb{1} - \frac{A^2}{\|A^2\|}}\right)^* \ = \ \left(\frac{A}{\sqrt{\|A^2\|}} - i\sqrt{\mathbb{1} - \frac{A^2}{\|A^2\|}}\right)$$

e que

$$\left(\frac{A}{\sqrt{\|A^2\|}} + i\sqrt{\mathbb{1} - \frac{A^2}{\|A^2\|}}\right)\left(\frac{A}{\sqrt{\|A^2\|}} - i\sqrt{\mathbb{1} - \frac{A^2}{\|A^2\|}}\right) \ = \ \mathbb{1} .$$

Para provar a última igualdade basta expandir o produto e notar que, pelo Lema 10.5, A e $\sqrt{\mathbb{1} - \frac{A^2}{\|A^2\|}}$ comutam, já que A e $\mathbb{1} - \frac{A^2}{\|A^2\|}$ comutam.

Assim, vemos de (10.89) que uma matriz Hermitiana A é combinação linear de até duas unitárias, provando a primeira parte da Proposição 10.28. Para provar a segunda parte, basta notar que se $M \in \mathrm{Mat}(\mathbb{C},\ n)$ é uma matriz qualquer, podemos escrever

$$M \ = \ \left(\frac{M + M^*}{2}\right) + i\left(\frac{M - M^*}{2i}\right) .$$

Ambas as matrizes entre parênteses são Hermitianas e, portanto, podem cada uma ser escritas como combinação linear de até duas unitárias, totalizando até quatro unitárias para M. ■

A Proposição 10.28 é válida não apenas para álgebras de matrizes. Vide Proposição 39.47, página 2084.

10.5.1.1 Matrizes Pseudoautoadjuntas e Quaseautoadjuntas

Uma matriz $A \in \mathrm{Mat}(\mathbb{C},\ n)$ é dita ser uma *matriz pseudoautoadjunta*, ou *pseudo-Hermitiana*, se existir uma matriz inversível $S \in \mathrm{Mat}(\mathbb{C},\ n)$, não necessariamente única, tal que

$$A^* \ = \ S^{-1}AS ,$$

ou seja, se for similar à sua adjunta. Se A for pseudoautoadjunta, A e A^* possuem o mesmo espectro por serem similares (Proposição 10.5, página 474). Como o espectro de A^* é sempre o complexo conjugado do espectro de A (Proposição

10.24, página 507), concluímos que para uma matriz pseudoautoadjunta o espectro consiste em autovalores reais ou de pares de números complexo-conjugados.

Uma matriz $A \in \mathrm{Mat}(\mathbb{C},\ n)$ é dita ser uma *matriz quaseautoadjunta*, ou *quase-Hermitiana* , se for pseudoautoadjunta e se a matriz inversível $S \in \mathrm{Mat}(\mathbb{C},\ n)$, acima, puder ser escolhida positiva.

As matrizes quaseautoadjunta são importantes pois seus autovalores são todos reais. Para ver isso, seja $S^{1/2}$ a raiz quadrada positiva de S. Defina $B := S^{-1/2}AS^{1/2}$. Por definição, A, A^* e B são similares e, portanto, possuem o mesmo espectro. Agora, pela definição, $B^* = S^{1/2}A^*S^{-1/2} = S^{1/2}S^{-1}ASS^{-1/2} = S^{-1/2}AS^{1/2} = B$, provando que B é autoadjunta e, assim, tem espectro real.

É conveniente aqui notar que a condição de uma matriz ser quaseautoadjunta é condição suficiente, mas não é condição necessária, para garantir a realidade dos seus autovalores.

Se A for quaseautoadjunta, então A possui algo similar à decomposição espectral (vide (10.54), página 495). Como B é autoadjunta, sua decomposição espectral é $B = \sum_{a=1}^r \alpha_a P_a$, com $1 \leq r \leq n$, com os α_a's sendo os autovalores distintos de B (todos reais) e com P_a sendo seus projetores espectrais, satisfazendo: $P_aP_b = \delta_{ab}P_a$, $\sum_{a=1}^r P_a = \mathbb{1}$ e $P_a^* = P_a$. Como $A = S^{1/2}BS^{-1/2}$, podemos escrever

$$A = \sum_{a=1}^{r} \alpha_a Q_a ,$$

com $Q_aQ_b = \delta_{ab}Q_a$ e $\sum_{a=1}^r Q_a = \mathbb{1}$, onde $Q_a := S^{1/2}P_aS^{-1/2}$. Verifique! Os operadores Q_a não são necessariamente projetores ortogonais[20], por não serem necessariamente autoadjuntos (como os operadores P_a o são), mas satisfazem $Q_a^* = S^{-1}Q_aS$, ou seja, são também matrizes quaseautoadjunta. Como $Q_a^2 = Q_a$, os autovalores de cada Q_a são 0 ou 1.

E. 10.29 *Exercício*. Seja A a matriz real 2×2 dada por $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, com a, b, c e d reais, sendo $b \neq 0$ e $c \neq 0$. Seja $S = \begin{pmatrix} 1 & 0 \\ 0 & \frac{c}{b} \end{pmatrix}$, com $S^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{b}{c} \end{pmatrix}$. Verifique explicitamente que $S^{-1}AS = A^T$ e, portanto, que A é pseudoautoadjunta. Verifique que S é positiva caso $\frac{c}{b} > 0$ e, portanto, nessa situação A é quaseautoadjunta.

Verifique que os autovalores de A são $\lambda_\pm = \frac{1}{2}\Big(a + d \pm \sqrt{(a-d)^2 + 4bc}\Big)$, o que deixa claro que se $c/b > 0$ (e, portanto, $bc > 0$) os dois autovalores são reais. A condição mais geral, no entanto, para que os autovalores sejam reais é, naturalmente, $4bc \geq -(a-d)^2$. ✠

Esse exemplo se deixa generalizar nas chamadas matrizes tridiagonais reais, as quais são relevantes na Mecânica Quântica de sistemas unidimensionais em um espaço discretizado (como no modelo de localização de Anderson[21] unidimensional).

• Matrizes tridiagonais

Uma matriz $T \in \mathrm{Mat}(\mathbb{C},\ n)$ é dita ser uma *matriz tridiagonal* se seus elementos de matriz satisfizerem $T_{ij} = 0$ sempre que $|i-j| > 1$. Em palavras, uma matriz é tridiagonal se todos os seus elementos de matriz forem nulos fora da diagonal principal, da primeira supradiagonal e da primeira infradiagonal.

Algumas matrizes tridiagonais são exemplos relevantes de matrizes pseudoautoadjuntas e quaseautoadjuntas.

E. 10.30 *Exercício*. Seja A a matriz tridiagonal real 3×3 dada por $A = \begin{pmatrix} a_1 & b_1 & 0 \\ c_1 & a_2 & b_2 \\ 0 & c_2 & a_3 \end{pmatrix}$, com todos os a_i's, b_j's e c_k's reais, sendo $b_j \neq 0$ e $c_k \neq 0$ para todos $j,\ k \in \{1,\ 2\}$. Seja $S = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{c_1}{b_1} & 0 \\ 0 & 0 & \frac{c_1c_2}{b_1b_2} \end{pmatrix}$, com $S^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{b_1}{c_1} & 0 \\ 0 & 0 & \frac{b_1b_2}{c_1c_2} \end{pmatrix}$. Verifique explicitamente que $S^{-1}AS = A^T$ e, portanto, que A é pseudoautoadjunta. Verifique que S é positiva caso $\frac{c_1}{b_1} > 0$ e $\frac{c_2}{b_2} > 0$ e, portanto, nessa situação A é quaseautoadjunta. ✠

O exercício acima pode ser generalizado.

[20]Exceto, é claro, no caso trivial em que $S = \mathbb{1}$, ou seja, quando A é autoadjunta.
[21]Philip Warren Anderson (1923–).

E. 10.31 *Exercício*. Seja A a matriz tridiagonal real $n \times n$ dada por

$$A = \begin{pmatrix} a_1 & b_1 & 0 & 0 & 0 & \cdots & 0 \\ c_1 & a_2 & b_2 & 0 & 0 & \cdots & 0 \\ 0 & c_2 & a_3 & b_3 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & c_{n-2} & a_{n-1} & b_{n-1} \\ 0 & 0 & \dots & 0 & 0 & c_{n-1} & a_n \end{pmatrix} .$$

Os elementos da diagonal principal são $a_1, \ldots, a_n$, os elementos da primeira supradiagonal são $b_1, \ldots, b_{n-1}$ e os elementos da primeira infradiagonal são $c_1, \ldots, c_{n-1}$. Assumimos que todos os a_i's, b_j's e c_k's são reais, sendo $b_j \neq 0$ e $c_k \neq 0$ para todos $j, k \in \{1, , \ldots, n-1\}$. Seja

$$S = \text{diag}\left(1, \frac{c_1}{b_1}, \frac{c_1 c_2}{b_1 b_2}, \ldots, \prod_{j=1}^{n-1} \frac{c_j}{b_j}\right) \quad \text{com} \quad S^{-1} = \text{diag}\left(1, \frac{b_1}{c_1}, \frac{b_1 b_2}{c_1 c_2}, \ldots, \prod_{j=1}^{n-1} \frac{b_j}{c_j}\right) .$$

Verifique que $S^{-1}AS = A^T$ e, portanto, que A é pseudoautoadjunta. Verifique que S é positiva caso $\frac{c_j}{b_j} > 0$ para todo $j \in \{1, \ldots, n-1\}$ e, portanto, nessa situação A é quaseautoadjunta. ✠

*** ** * ** ***

Para um tratamento de operadores quaseautoadjuntos que inclua o caso de operadores não limitados, vide [96].

10.5.2 O Teorema de Inércia de Sylvester. Superfícies Quadráticas

• **Transformações de congruência em** $\text{Mat}(\mathbb{C}, n)$

Seja $M \in \text{Mat}(\mathbb{C}, n)$. Se $P \in \text{Mat}(\mathbb{C}, n)$ é inversível, a transformação $M \mapsto P^* M P$ é dita ser uma *transformação de congruência*. Uma transformação de congruência representa a transformação de uma matriz por uma mudança de base (justifique essa afirmação!).

Se M for autoadjunta, $P^* M P$ é também autoadjunta e, portanto, ambas têm autovalores reais. Em geral, o conjunto dos autovalores de M é distinto do conjunto dos autovalores de $P^* M P$ (exceto, por exemplo, se P for unitária). Porém, um teorema devido a Sylvester, frequentemente denominado *Lei de Inércia de Sylvester*, afirma que uma propriedade do conjunto dos autovalores é preservada em uma transformação de congruência, a saber, o número de autovalores, positivos, de autovalores negativos e de autovalores nulos (contando-se as multiplicidades). Enunciaremos e demonstraremos esse teorema logo adiante.

Dada uma matriz autoadjunta $M \in \text{Mat}(\mathbb{C}, n)$, a tripla de números (m, m', m_0), onde m é o número de autovalores positivos de M, m' é o número de autovalores negativos de M, m_0 é o número de autovalores nulos de M, (em todos os casos contando-se as multiplicidades) é denominada (por razões históricas obscuras) a *inércia* da matriz M. Naturalmente, vale $m + m' + m_0 = n$. A Lei de Inércia de Sylvester afirma, portanto, que a inércia de uma matriz é preservada por transformações de congruência.

Dizemos que duas matrizes A e $B \in \text{Mat}(\mathbb{C}, n)$ são congruentes se existir $P \in \text{Mat}(\mathbb{C}, n)$ inversível tal que $A = P^* B P$. É muito fácil provar que a relação de congruência é uma relação de equivalência.

E. 10.32 *Exercício*. Demonstre essa afirmação! ✠

Dessa forma, a Lei de Inércia de Sylvester afirma que a inércia de matrizes é constante nas classes de equivalência (pela relação de congruência). Assim, é legítimo perguntar se as classes de equivalência são univocamente determinadas pela inércia de seus elementos. A resposta é negativa (exceto no caso trivial $n = 1$), como mostra a argumentação do parágrafo que segue.

Se $A \in \text{Mat}(\mathbb{C}, n)$, com $n > 1$, é uma matriz positiva, A é da forma P^*P (Corolário 10.4, página 513). Assim, $\det A = |\det P|^2$ e concluímos que A é inversível se e somente se P o for. Conclui-se disso que a classe de equivalência (por relações de congruência) que contém a matriz identidade contém todas as matrizes positivas e inversíveis. Pela Proposição 10.27, página 512, esse conjunto coincide com o conjunto de todas as matrizes autoadjuntas com autovalores positivos, ou seja, que possuem inércia $(n, 0, 0)$. Entretanto, existem também matrizes não autoadjuntas com inércia $(n, 0, 0)$ (por exemplo, matrizes triangulares superiores[22] com elementos positivos na diagonal e alguns elementos não nulos acima da diagonal). Como tais matrizes não podem ser equivalentes à identidade (toda matriz da forma $P^*\mathbb{1}P$ é autoadjunta), concluímos que as classes de equivalência não são determinadas univocamente pela inércia das matrizes que as compõem.

• A Lei de Inércia de Sylvester

A Lei de Inércia de Sylvester é importante para a classificação de formas quadráticas e sua relevância estende-se até à classificação de equações diferenciais parciais de segunda ordem. Tratemos de seu enunciado e demonstração.

Teorema 10.18 (Lei de Inércia de Sylvester) *Sejam A e $B \in \text{Mat}(\mathbb{C}, n)$ duas matrizes autoadjuntas. Denotemos por $\mathcal{A}_+$, $\mathcal{A}_-$, $\mathcal{A}_0$ os subespaços gerados, respectivamente, pelos autovetores com autovalores positivos, negativos e nulos de A (e analogamente para B).*

Suponhamos que exista $P \in \text{Mat}(\mathbb{C}, n)$, inversível, tal que $A = P^*BP$. Então, $\dim \mathcal{A}_+ = \dim \mathcal{B}_+$, $\dim \mathcal{A}_- = \dim \mathcal{B}_-$ e $\dim \mathcal{A}_0 = \dim \mathcal{B}_0$, onde $\dim \mathcal{C}$ denota a dimensão de um subespaço $\mathcal{C} \subset \mathbb{C}^n$. Assim, concluímos também que A e B têm o mesmo número de autovalores positivos, o mesmo número de autovalores negativos e o mesmo número de autovalores nulos (em todos os casos, contando-se as multiplicidades). $\square$

Prova. Sejam $\alpha_1, \ldots, \alpha_a$ os autovalores positivos (não necessariamente distintos) e $\alpha_{a+1}, \ldots, \alpha_{a+a'}$ os autovalores negativos (não necessariamente distintos) de A. Analogamente, sejam $\beta_1, \ldots, \beta_b$ os autovalores positivos (não necessariamente distintos) e $\beta_{b+1}, \ldots, \beta_{b+b'}$ os autovalores negativos (não necessariamente distintos) de B. Naturalmente, valem $0 \leq a + a' \leq n$ e $0 \leq b + b' \leq n$.

Se A e B forem nulos não há o que demonstrar, de modo que podemos supor que ambos têm pelo menos um autovalor não-nulo. Nesse caso, podemos sempre, sem perder em generalidade, supor que A tem pelo menos um autovalor positivo, pois se tal não for verdade para A será verdadeiro para $-A$.

O Teorema Espectral, Teorema 10.7, página 494, permite-nos escrever

$$A = \sum_{k=1}^{a} \alpha_k A_k - \sum_{l=a+1}^{a+a'} |\alpha_l| A_l$$

e

$$B = \sum_{k=1}^{b} \beta_k B_k - \sum_{l=b+1}^{b+b'} |\beta_l| B_l , \tag{10.90}$$

onde A_j e B_j são os projetores espectrais de A e B, respectivamente. Defina-se

$$A_+ := \sum_{k=1}^{a} A_k , \qquad A_- := \sum_{l=a+1}^{a+a'} A_l \qquad \text{e} \qquad A_0 := \mathbb{1} - A_+ - A_-$$

e, analogamente,

$$B_+ := \sum_{k=1}^{b} B_k , \qquad B_- := \sum_{l=b+1}^{b+b'} B_l \qquad \text{e} \qquad B_0 := \mathbb{1} - B_+ - B_- .$$

A_+, A_- e A_0 são, respectivamente, o projetor sobre o subespaço de autovetores com autovalores positivos, negativos e nulos de A. Analogamente para B. Esses subespaços são

$$\mathcal{A}_\pm = A_\pm \mathbb{C}^n , \quad \mathcal{A}_0 = A_0 \mathbb{C}^n , \quad \mathcal{B}_\pm = B_\pm \mathbb{C}^n , \quad \mathcal{B}_0 = B_0 \mathbb{C}^n .$$

[22] Para a definição, vide página 523

Seja x um vetor não-nulo de $\mathcal{A}_+$. Tem-se que $A_l x = 0$ para todo $l > a$ e $A_k x \neq 0$ para pelo menos um $k = 1, \ldots, a$. Logo, como $\alpha_k > 0$ para todo $k = 1, \ldots, a$, segue que

$$\langle x, Ax\rangle_{\mathbb{C}} = \sum_{k=1}^{a} \alpha_k \langle x, A_k x\rangle_{\mathbb{C}} = \sum_{k=1}^{a} \alpha_k \langle A_k x, A_k x\rangle_{\mathbb{C}} = \sum_{k=1}^{a} \alpha_k \|A_k x\|^2 > 0 . \tag{10.91}$$

Porém, para um tal x vale também

$$\langle x, Ax\rangle_{\mathbb{C}} = \langle x, P^* BPx\rangle_{\mathbb{C}} = \langle Px, BPx\rangle_{\mathbb{C}} \overset{(10.90)}{=} \sum_{k=1}^{b} \beta_k \|B_k Px\|^2 - \sum_{l=b+1}^{b+b'} |\beta_k| \left\|B_k Px\right\|^2 .$$

Vamos agora supor que $\mathcal{B}_+ < \dim \mathcal{A}_+$ (ou seja, que $b < a$). Afirmamos que podemos encontrar ao menos um $x_+ \in \mathcal{A}_+$, não-nulo, tal que $B_k Px_+ = 0$ para todo $k = 1, \ldots, b$. Se assim não fosse, não existiria $x \in \mathcal{A}_+$ não-nulo satisfazendo $B_+ Px = 0$, ou seja, valeria $B_+ Px \neq 0$ para todo $x \in \mathcal{A}_+$ com $x \neq 0$. Logo, $(P\mathcal{A}_+) \cap (\mathcal{B}_+)^\perp = \{0\}$, o que implica que $P\mathcal{A}_+ \subset \mathcal{B}_+$. Isso, por sua vez, significa que dimensão do subespaço $P\mathcal{A}_+$ é menor ou igual à dimensão de $\mathcal{B}_+$ e, como P é inversível, isso implica, $\dim \mathcal{A}_+ \leq \dim \mathcal{B}_+$, uma contradição.

Assim, para um tal x_+ teríamos

$$\langle x_+, Ax_+\rangle_{\mathbb{C}} = -\sum_{l=b+1}^{b+b'} |\beta_k| \left\|B_k Px_+\right\|^2 \leq 0 ,$$

contradizendo (10.91). Concluímos disso que $\dim \mathcal{B}_+ \geq \dim \mathcal{A}_+$. Como $B = (P^*)^{-1}AP^{-1}$, um raciocínio análogo trocando A e B e trocando $P \to P^{-1}$ implica que $\dim \mathcal{A}_+ \geq \dim \mathcal{B}_+$. Assim, $\dim \mathcal{B}_+ = \dim \mathcal{A}_+$.

Também de forma totalmente análoga prova-se que $\dim \mathcal{B}_- = \dim \mathcal{A}_-$ (isso também pode ser visto imediatamente trocando $A \mapsto -A$ e $B \mapsto -B$). Isso implica ainda que $\dim \mathcal{B}_0 = \dim \mathcal{A}_0$, completando a demonstração. ■

• Transformações de congruência em $\mathrm{Mat}(\mathbb{R}, n)$

Para matrizes reais agindo no espaço $\mathbb{R}^n$ valem afirmações análogas às obtidas acima. Seja $M \in \mathrm{Mat}(\mathbb{R}, n)$. Se $P \in \mathrm{Mat}(\mathbb{R}, n)$ é inversível, a transformação $M \mapsto P^T MP$ é dita ser uma *transformação de congruência real*, ou simplesmente *transformação de congruência*. Uma transformação de congruência representa a transformação de uma matriz por uma mudança de base (justifique essa afirmação!). Para transformações de congruência reais vale também a Lei de Inércia de Sylvester: se $A \in \mathrm{Mat}(\mathbb{R}, n)$ é simétrica (ou seja, se $A = A^T$) sua inércia é preservada por transformações de congruência $A \mapsto P^T AP$ com $P \in \mathrm{Mat}(\mathbb{R}, n)$ inversível. Como essa afirmação é um mero caso particular do anterior, omitimos a demonstração e convidamos o estudante a completá-la.

• Classificação de matrizes simétricas em $\mathbb{R}^n$

Matrizes simétricas em $\mathbb{R}^n$ podem ser classificadas de acordo com o tipo de inércia que possuem, classificação essa invariante por transformações de congruência. Uma matriz simétrica $A \in \mathrm{Mat}(\mathbb{R}, n)$, $n > 1$, é dita ser

1. *Parabólica*, se ao menos um dos seus autovalores for nulo, ou seja, se sua inércia for da forma (a, a', a_0) com $a_0 \geq 1$;

2. *Elíptica*, se todos os seus autovalores forem positivos ou se todos forem negativos, ou seja, se sua inércia for da forma $(a, a', 0)$ com $a \geq 1$ e $a' = 0$ ou com $a' \geq 1$ e $a = 0$;

3. *Hiperbólica*, se um de seus autovalores for positivo e os demais negativos, ou o oposto: se um de seus autovalores for negativo e os demais positivos, ou seja, se sua inércia for da forma $(1, a', 0)$ com $a' \geq 1$ $(a, 1, 0)$ com $a \geq 1$;

4. *Ultra-hiperbólica*, se ao menos dois de seus autovalores forem positivos e ao menos dois forem negativos, nenhum sendo nulo, ou seja, se sua inércia for da forma $(a, a', 0)$ com $a \geq 2$ e $a' \geq 2$. Esse caso só se dá se $n \geq 4$.

Essa nomenclatura que classifica as matrizes em *parabólicas*, *elípticas*, *hiperbólicas* e *ultra-hiperbólicas* tem uma motivação geométrica relacionada à classificação de superfícies quadráticas em $\mathbb{R}^n$, assunto que ilustraremos abaixo.

• **Superfícies quadráticas $\mathbb{R}^n$**

Sejam $x_1, \ldots, x_n$ são n variáveis reais. A forma mais geral de um polinômio real de segundo grau nessas variáveis é

$$p(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_ix_j + \sum_{k=1}^{n} c_kx_k + d ,$$

onde $A_{ij} \in \mathbb{R}$, $c_k \in \mathbb{R}$ e $d \in \mathbb{R}$. A expressão acima para p pode ser escrita como $p(x) = \langle x,\ Ax\rangle_{\mathbb{R}} + \langle c,\ x\rangle_{\mathbb{R}} + d$, onde, naturalmente, A é a matriz cujos elementos são A_{ij}, $c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$ e $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. A matriz A pode ser sempre, sem perda de generalidade, escolhida como simétrica. Para ver isso, notemos que, A pode sempre ser escrita como soma de uma matriz simétrica e uma antissimétrica: $A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T)$. Contudo,

$$\langle x,\ (A - A^T)x\rangle_{\mathbb{R}} = \sum_{i=1}^{n}\sum_{j=1}^{n} (A_{ij} - A_{ji})\, x_ix_j = 0$$

como facilmente se constata. Assim, a parte antissimétrica de A, ou seja, $\frac{1}{2}(A - A^T)$, não contribui em $\langle x,\ Ax\rangle_{\mathbb{R}}$, apenas a parte simétrica $\frac{1}{2}(A + A^T)$. Portanto, A será doravante considerada simétrica.

Estamos agora interessados em classificar as superfícies em $\mathbb{R}^n$ definidas por $p(x) = \alpha$, com α constante. Há primeiramente dois casos a considerar: *1)* A é inversível e *2)* A não é inversível.

1. Se A é inversível, podemos escrever

$$p(x) = \langle x,\ Ax\rangle_{\mathbb{R}} + \langle c,\ x\rangle_{\mathbb{R}} + d = \left\langle \left(x + \frac{1}{2}A^{-1}c\right),\ A\left(x + \frac{1}{2}A^{-1}c\right)\right\rangle_{\mathbb{R}} - \frac{1}{4}\left\langle c,\ A^{-1}c\right\rangle_{\mathbb{R}} + d .$$

Verifique! Assim, a equação $p(x) = \alpha$ fica $\left\langle \left(x + \frac{1}{2}A^{-1}c\right),\ A\left(x + \frac{1}{2}A^{-1}c\right)\right\rangle_{\mathbb{R}} = \beta$, onde β é a constante $\alpha + \frac{1}{4}\left\langle c,\ A^{-1}c\right\rangle_{\mathbb{R}} - d$. A matriz simétrica A pode ser diagonalizada por uma matriz ortogonal, ou seja, podemos escrever $A = O^TDO$, com $D = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n)$, com λ_k sendo os autovalores de A e O sendo ortogonal. Podemos sempre escolher O de sorte que os primeiros m autovalores $\lambda_1, \ldots, \lambda_m$ são positivos e os demais $\lambda_{m+1}, \ldots, \lambda_n$ são negativos (não há autovalores nulos, pois A foi suposta inversível).

Com isso, $\left\langle \left(x + \frac{1}{2}A^{-1}c\right),\ A\left(x + \frac{1}{2}A^{-1}c\right)\right\rangle_{\mathbb{R}} = \langle y,\ Dy\rangle_{\mathbb{R}}$, onde $y = O\left(x + \frac{1}{2}A^{-1}c\right)$. A equação $p(x) = \alpha$ fica, então, $\sum_{k=1}^{n} \lambda_k\ y_k^2 = \beta$ ou seja,

$$\sum_{k=1}^{m} \lambda_k\ y_k^2 - \sum_{l=m+1}^{n} |\lambda_l|\ y_l^2 = \beta . \tag{10.92}$$

Temos os seguintes subcasos a tratar:

(a) Se todos os autovalores de A são positivos e $\beta > 0$, a equação (10.92) descreve um *elipsoide* em $\mathbb{R}^n$ (se $\beta < 0$ não há soluções e se $\beta = 0$ a equação descreve apenas o ponto $y = 0$ em $\mathbb{R}^n$). O mesmo vale, reciprocamente, se todos os autovalores de A forem negativos e $\beta < 0$ (se $\beta > 0$ não há soluções e se $\beta = 0$ a equação descreve apenas o ponto $y = 0$ em $\mathbb{R}^n$).

(b) Se um dos autovalores de A é positivo e os demais $n - 1$ são negativos, ou se ocorre o oposto, ou seja, se um dos autovalores de A é negativo e os demais $n - 1$ são positivos, então a equação (10.92) descreve um *hiperboloide* $(n - 1)$-dimensional em $\mathbb{R}^n$ no caso $\beta \neq 0$.

Se $\beta > 0$ o hiperboloide tem duas folhas (i.e., possui duas componentes conexas) e no caso $\beta < 0$ apenas uma. A Figura 10.2, página 521, exibe hiperboloides com uma e duas folhas em $\mathbb{R}^3$.

Devido a sua estabilidade, hiperboloides de uma folha são frequentemente encontrados em estruturas arquitetônicas. A bem conhecida *catedral de Brasília*, de Niemeyer[23], é um exemplo. A estabilidade estrutural desse formato decorre do fato que por qualquer ponto de um hiperboloide de uma folha passam duas linhas retas inteiramente contidas dentro do mesmo (prove isso!).

Se $\beta = 0$ a equação (10.92) descreve um *cone* $(n - 1)$-dimensional em $\mathbb{R}^n$.

[23]Oscar Niemeyer Soares Filho (1907–).

(c) Este caso ocorre apenas se $n \geq 4$. Se ao menos dois autovalores de A é positivo e ao menos dois são positivos a equação (10.92) descreve, no caso $\beta \neq 0$, uma superfície $(n-1)$-dimensional em $\mathbb{R}^n$ denominada *ultra-hiperboloide*. Se $\beta = 0$ a equação (10.92) descreve uma $(n-1)$-dimensional em $\mathbb{R}^n$ denominada *ultracone*.

2. Se A não é inversível temos que proceder de modo ligeiramente diferente. Como antes, a matriz simétrica A pode ser diagonalizada por uma matriz ortogonal, ou seja, podemos escrever $A = O^T DO$, com $D = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n)$, com λ_k sendo os autovalores de A e O sendo ortogonal. Como A não tem inversa, alguns de seus autovalores são nulos. Podemos sempre escolher O de sorte que os primeiros m autovalores $\lambda_1, \ldots, \lambda_m$ são positivos, os m' autovalores seguintes $\lambda_{m+1}, \ldots, \lambda_{m+m'}$ são negativos e os demais $\lambda_{m+m'+1}, \ldots, \lambda_n$ são nulos. Naturalmente, $0 \leq m+m' < n$. Podemos, então, escrever $p(x) = \langle x, Ax\rangle_{\mathbb{R}} + \langle c, x\rangle_{\mathbb{R}} + d = \langle y, Dy\rangle_{\mathbb{R}} + \langle Oc, y\rangle_{\mathbb{R}} + d$ onde $y = Ox$.

Assim, se $c \neq 0$ a equação $p(x) = \alpha$ fica

$$y_{Oc} \;=\; \gamma + \frac{1}{\|c\|}\left(\sum_{l=m+1}^{m+m'} |\lambda_l|\, y_l^2 - \sum_{k=1}^{m} \lambda_k\, y_k^2\right) , \tag{10.93}$$

onde $\gamma = (\alpha - d)/\|c\|$ e y_{Oc} é a projeção de y na direção do vetor Oc. Se a dimensão do subespaço dos autovalores nulos $\mathcal{A}_0$ for maior que 1 a equação (10.93) descreverá cilindros de diversos tipos, dependendo do número de autovalores positivos e negativos e de Oc ter uma projeção ou não em $\mathcal{A}_0$. Não descreveremos os todos os detalhes aqui, mas um exemplo de interesse se dá em $\mathbb{R}^3$, se $\mathcal{A}_0$ tiver dimensão 2 e Oc for um vetor não-nulo de $\mathcal{A}_0$. Nesse caso equação (10.93) descreve um *cilindro parabólico*. Vide Figura 10.4, página 522.

Para o caso em que $\mathcal{A}_0$ tem dimensão 1 e Oc é um elemento não-nulo desse subespaço, a equação (10.93) descreve diversos tipos de paraboloides $(n-1)$-dimensionais. Temos os seguintes casos:

(a) a equação (10.93) descreve um *paraboloide elíptico* $(n-1)$-dimensional caso todos os autovalores não nulos de A forem positivos ou se todos os autovalores não nulos de A forem negativos. Vide Figura 10.3, página 10.3.

(b) A equação (10.93) descreve um *paraboloide hiperbólico* $(n-1)$-dimensional caso um autovalor de A seja negativo e os demais autovalores não nulos de A sejam positivos (ou o contrário: caso um autovalor de A seja positivo e os demais autovalores não nulos de A sejam negativos). Vide Figura 10.3, página 10.3.

(c) A equação (10.93) descreve um *paraboloide ultra-hiperbólico* $(n-1)$-dimensional caso pelo menos dois dos autovalores não nulos de A sejam positivos e pelo menos dois dos autovalores não nulos de A sejam negativos. Esse caso só pode ocorrer se $n \geq 5$.

Para $c \neq 0$ diversas situações acima podem também descrever cilindros, por exemplo, se Oc encontra-se no subespaço dos autovetores com autovalores não nulos.

Se $c = 0$ e $\dim A_0 \geq 1$, equação $p(x) = \alpha$ fica

$$\sum_{k=1}^{m} \lambda_k\, y_k^2 - \sum_{l=m+1}^{m+m'} |\lambda_l|\, y_l^2 \;=\; \beta , \tag{10.94}$$

com $\beta = \alpha - d$. A equação (10.94) descreve diversos tipo de cilindros $(n-1)$-dimensionais.

(a) Caso $c = 0$ a equação (10.94) descreve um *cilindro elíptico* $(n-1)$-dimensional caso todos os autovalores não nulos de A forem positivos ou se todos os autovalores não nulos de A forem negativos. Vide Figura 10.4, página 522.

(b) Caso $c = 0$ a equação (10.94) descreve um *cilindro hiperbólico* $(n-1)$-dimensional caso um autovalor de A seja negativo e os demais autovalores não nulos de A sejam positivos (ou o contrário: caso um autovalor de A seja positivo e os demais autovalores não nulos de A sejam negativos). Vide Figura 10.4, página 522.

(c) Caso $c = 0$ a equação (10.94) descreve um *cilindro ultra-hiperbólico* $(n-1)$-dimensional caso pelo menos dois dos autovalores não nulos de A sejam positivos e pelo menos dois dos autovalores não nulos de A sejam negativos. Esse caso só pode ocorrer se $n \geq 5$ (lembrar que pelo menos um dos autovalores de A é nulo).

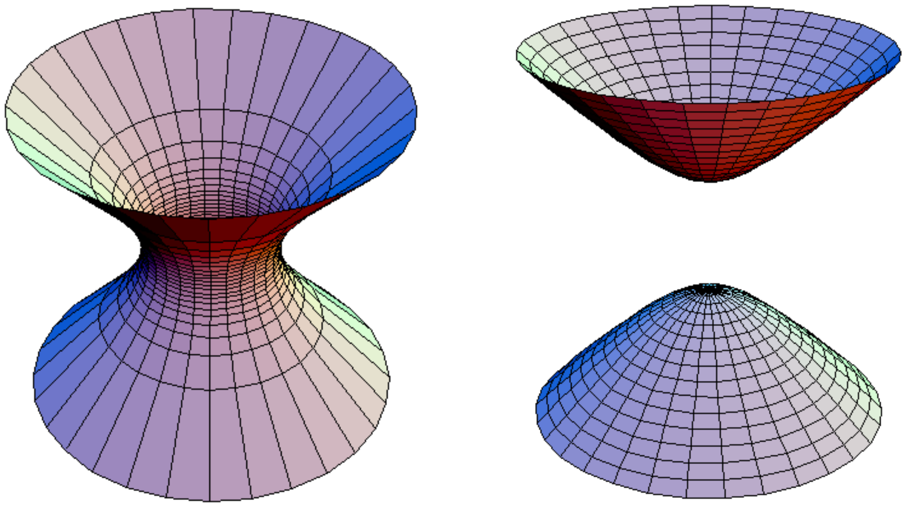

Figura 10.2: Hiperboloides com uma e duas folhas em $\mathbb{R}^3$.

10.5.3 Um Resultado Sobre Localização do Espectro de Matrizes Autoadjuntas

Apresentaremos aqui alguns resultados sobre a localização do espectro de matrizes autoadjuntas complexas, empregando as chamadas *desigualdades de Samuelson* e suas generalizações, estudadas na Seção 6.3.1.2, página 330.

Seja $A \in \mathrm{Mat}(\mathbb{C},\ m)$ autoadjunta e sejam $\lambda_1,\ \ldots,\ \lambda_m$ seus autovalores (não necessáriamente distintos), os quais sabidamente são todos reais. Seu polinômio característico é $q_A(x) = (x-\lambda_1)\cdots(x-\lambda_m)$. Escrevamos q_A na forma $x^m + a_{m-1}x^{m-1} + \cdots + a_1 x + a_0$.

Pelas duas primeiras fórmulas de Viète em (6.47), página 327, temos

$$a_{m-1} \;=\; -\big(\lambda_1 + \cdots + \lambda_m\big) \;=\; -\mathrm{Tr}(A)\;, \tag{10.95}$$

$$a_{m-2} \;=\; \sum_{\substack{i,\,j=1\\ i<j}}^{m} \lambda_i\lambda_j\;. \tag{10.96}$$

Agora,

$$\big(a_{m-1}\big)^2 \stackrel{(10.95)}{=} \big(\lambda_1+\cdots+\lambda_m\big)^2 \;=\; 2\sum_{\substack{i,\,j=1\\ i<j}}^{m}\lambda_i\lambda_j + \sum_{k=1}^{m}\lambda_k^2 \stackrel{(10.96)}{=} 2a_{m-2} + \sum_{k=1}^{m}\lambda_k^2 \;=\; 2a_{m-2} + \mathrm{Tr}\big(A^2\big)$$

e concluímos que

$$a_{m-2} \;=\; \frac{1}{2}\Big(\big(\mathrm{Tr}(A)\big)^2 - \mathrm{Tr}\big(A^2\big)\Big)\;.$$

Como as raízes de q_A (ou seja, os autovalores de A) são reais, podemos evocar a Proposição 6.3, página 328, e afirmar que os autovalores de A estão localizados no intervalo $\big[\alpha_m^-,\ \alpha_m^+\big]$, onde

$$\begin{aligned}\alpha_m^{\pm} \;&:=\; \frac{-a_{m-1} \pm \sqrt{\big((m-1)a_{m-1}\big)^2 - 2m(m-1)a_{m-2}}}{m}\\[1ex] &=\; \frac{\mathrm{Tr}(A) \pm \sqrt{(m-1)\Big(m\mathrm{Tr}\big(A^2\big) - \big(\mathrm{Tr}(A)\big)^2\Big)}}{m}\;.\end{aligned}$$

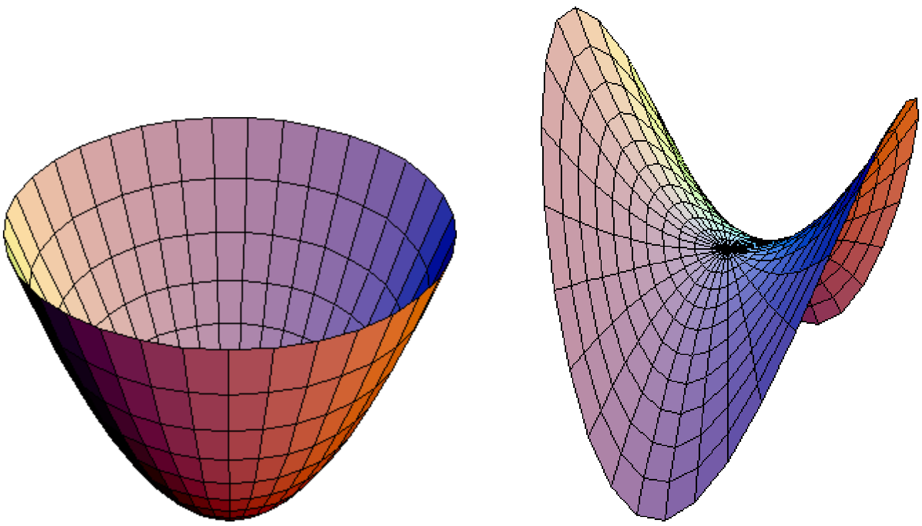

Figura 10.3: Um paraboloide elíptico (esquerda) e um paraboloide hiperbólico (direita) em $\mathbb{R}^3$.

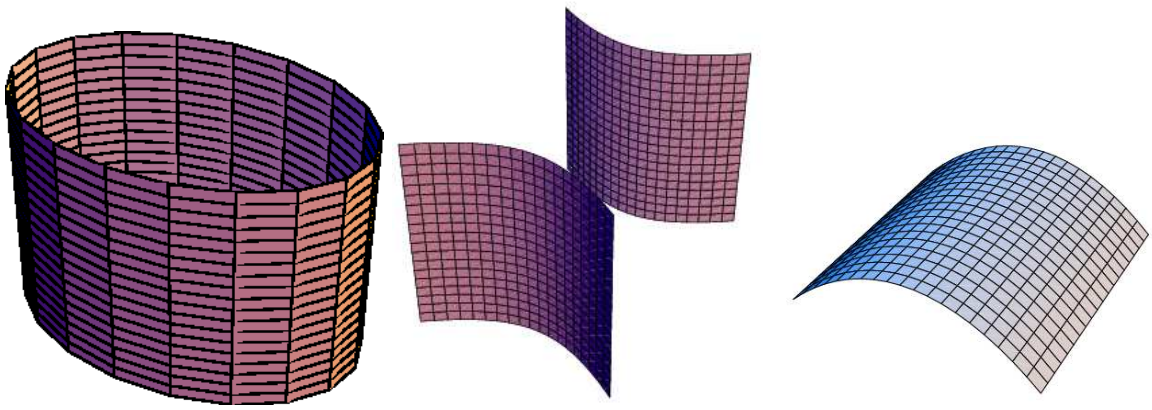

Figura 10.4: Um cilindro elíptico (esquerda), um cilindro hiperbólico (centro) e um cilindro parabólico (direita) em $\mathbb{R}^3$.

Essa relação é mais elegantemente expressa em termos do traço normalizado $\omega_{\mathrm{tr}}(B) := \frac{1}{m}\mathrm{Tr}(B)$, para $B \in \mathrm{Mat}(\mathbb{C},\ m)$, que vem a ser um estado na álgebra C*definida por $\mathrm{Mat}(\mathbb{C},\ m)$. Temos,

$$\alpha_m^{\pm} \ = \ \omega_{\mathrm{tr}}(A) \pm \sqrt{(m-1)}\sqrt{\omega_{\mathrm{tr}}\left(A^2\right) - \left(\omega_{\mathrm{tr}}(A)\right)^2}\ . \tag{10.97}$$

Vale observar que $\omega_{\mathrm{tr}}\left(A^2\right) - \left(\omega_{\mathrm{tr}}(A)\right)^2$ é a variância de A no estado ω_{tr}.

Para referência futura, coletemos o resultado demonstrado acima na seguinte proposição:

Proposição 10.29 *Seja $A \in \mathrm{Mat}(\mathbb{C},\ m)$, autoadjunta. Então, $\sigma(A) \subset \left[\alpha_m^-,\ \alpha_m^+\right]$, onde*

$$\alpha_m^{\pm} \ = \ \frac{\mathrm{Tr}(A) \pm \sqrt{(m-1)\Big(m\mathrm{Tr}\big(A^2\big) - \big(\mathrm{Tr}(A)\big)^2\Big)}}{m} \ = \ \omega_{\mathrm{tr}}(A) \pm \sqrt{m-1}\sqrt{\omega_{\mathrm{tr}}\left(A^2\right) - \left(\omega_{\mathrm{tr}}(A)\right)^2}\ , \tag{10.98}$$

sendo que, para $B \in \mathrm{Mat}(\mathbb{C},\ m)$, definimos $\omega_{\mathrm{tr}}(B) := \frac{1}{m}\mathrm{Tr}(B)$, que é um estado na álgebra C* definida por $\mathrm{Mat}(\mathbb{C},\ m)$. Vale observar que $\omega_{\mathrm{tr}}\left(A^2\right) - \left(\omega_{\mathrm{tr}}(A)\right)^2 \equiv \mathrm{Var}_{\omega_{\mathrm{tr}}}(A)$, a variância de A no estado ω_{tr}. □

No caso $A = \mu\mathbb{1}_m$ com $\mu \in \mathbb{R}$, por exemplo, é fácil ver que $\alpha_m^{\pm} = \mu$ e, portanto, $\sigma(A) = \{\mu\}$, como deveria ser. Para $m = 1$ é trivial verificar que o resultado é igualmente ótimo. Para $m = 2$ e para uma matriz autoadjunta geral

$A = \left(\begin{smallmatrix} a & b \\ \overline{b} & c \end{smallmatrix}\right)$, com $a, \ c \in \mathbb{R}$ e $b \in \mathbb{C}$, tem-se $\sigma(A) = \left\{\frac{a+c-\Delta}{2}, \ \frac{a+c+\Delta}{2}\right\}$ com $\Delta := \sqrt{(a-c)^2 + 4|b|^4}$, enquanto que $\alpha_2^{\pm} = \frac{a+c\pm\Delta}{2}$, mostrando que o intervalo $\left[\alpha_2^-, \ \alpha_2^+\right]$ é ótimo, o que geralmente não ocorre se $m \geq 3$. Esse fato sugere procurar-se aprimoramentos das desigualdades empregadas acima.

Para $m \in \mathbb{N}$ e p um número par, $p > 1$, a generalização das desigualdedes de Samuelson obtida em (6.68), página 333, pode ser aplicada aqui, fornecendo para uma matriz autoadjunta $A \in \operatorname{Mat}(\mathbb{C}, \ m)$,

$$\sigma(A) \ \subset \ \left[\beta_{m,\,p}^-, \ \beta_{m,\,p}^+\right] ,$$

onde,

$$\beta_{m,\,p}^{\pm} \ := \ \omega_{\mathrm{tr}}(A) \pm f_p(m)\left[\omega_{\mathrm{tr}}\Big(\big(A - \omega_{\mathrm{tr}}(A)\mathbb{1}_m\big)^p\Big)\right]^{1/p} ,$$

onde

$$f_p(m) \ := \ \left[\frac{m}{1 + (m-1)^{1-p}}\right]^{1/p} .$$

Para $p = 2$ recupera-se o resultado anterior, pois $\beta_{m,\,2}^{\pm} = \alpha_m^{\pm}$, como facilmente se verifica.

Com p par, a expressão $\omega_{\mathrm{tr}}\Big(\big(A - \omega_{\mathrm{tr}}(A)\mathbb{1}_m\big)^p\Big)$ representa o p-ésimo momento central de A no estado ω_{tr}.

10.6 Matrizes Triangulares

Uma matriz $S \in \operatorname{Mat}(\mathbb{C}, \ n)$ é dita ser uma *matriz triangular superior* se forem nulos os elementos abaixo da diagonal principal, ou seja, se $S_{ij} = 0$ sempre que $i > j$. Note que esses não precisam ser necessariamente os únicos elementos nulos de S.

Uma matriz $I \in \operatorname{Mat}(\mathbb{C}, \ n)$ é dita ser uma *matriz triangular inferior* se forem nulos os elementos acima da diagonal principal, ou seja, se $I_{ij} = 0$ sempre que $i < j$. Note que esses não precisam ser necessariamente os únicos elementos nulos de I.

Proposição 10.30 *Matrizes triangulares superiores possuem as seguintes propriedades:*

1. *A matriz identidade $\mathbb{1}$ é uma matriz triangular superior.*
2. *O produto de duas matrizes triangulares superiores é novamente uma matriz triangular superior.*
3. *O determinante de uma matriz triangular superior é o produto dos elementos da sua diagonal. Assim, uma matriz triangular superior é inversível se e somente se não tiver zeros na diagonal.*
4. *Se uma matriz triangular superior é inversível, sua inversa é novamente uma matriz triangular superior.*

As afirmações acima permanecem verdadeiras trocando "matriz triangular superior" por "matriz triangular inferior". □

Prova. Os três primeiros itens são elementares. Para provar o item 4, usa-se a regra de Laplace, expressão (10.20), página 467. Como é fácil de se ver, $\operatorname{Cof}(S)_{ji} = 0$ se $i > j$. Logo, S^{-1} é triangular superior, se existir. ■

As propriedades acima atestam que o conjunto das matrizes $n \times n$ triangulares superiores inversíveis forma um grupo, denominado por alguns autores *Grupo de Borel*[24] de ordem n e denotado por $\mathrm{GB}_n(\mathbb{C})$. Vide Seção 21.3.2, página 21.3.2. Como discutido naquela seção, um tipo de grupo de Borel particularmente relevante é o chamado grupo de Heisenberg.

O seguinte resultado sobre matrizes triangulares superiores será usado diversas vezes adiante.

Lema 10.6 *Uma matriz triangular superior $S \in \operatorname{Mat}(\mathbb{C}, \ n)$ é normal (ou seja, satisfaz $SS^* = S^*S$) se e somente se for diagonal.* □

[24] Armand Borel (1923–2003).

Prova. Se S é diagonal, S é obviamente normal pois S^* é também diagonal e matrizes diagonais sempre comutam entre si. Provaremos a recíproca, o que será feito por indução. Para $n = 1$ não há o que provar. Se $n = 2$, S é da forma $S = \left(\begin{smallmatrix} a & b \\ 0 & c \end{smallmatrix}\right)$, com $a,\ b,\ c \in \mathbb{C}$. A condição $SS^* = S^*S$ significa

$$\begin{pmatrix} |a|^2 + |b|^2 & b\overline{c} \\ \\ c\overline{b} & |c|^2 \end{pmatrix} = \begin{pmatrix} |a|^2 & b\overline{a} \\ \\ a\overline{b} & |b|^2 + |c|^2 \end{pmatrix} ,$$

o que implica $b = 0$, provando que S é diagonal. Procedemos agora por indução, supondo $n > 2$ e que o lema seja válido para matrizes $(n-1) \times (n-1)$ triangulares superiores normais. Se $S \in \mathrm{Mat}\,(\mathbb{C},\ n)$ é triangular superior, S é da forma

$$S = \begin{pmatrix} a & b^T \\ \\ \mathbb{0} & C \end{pmatrix} , \quad \text{sendo } a \in \mathbb{C}\ , \quad b\ = \begin{pmatrix} b_1 \\ \vdots \\ b_{n-1} \end{pmatrix} , \quad \mathbb{0}\ = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} ,$$

ambas b e $\mathbb{0}$ com $n-1$ linhas, sendo C uma matriz $(n-1) \times (n-1)$ triangular superior. A condição $SS^* = S^*S$ significa

$$\begin{pmatrix} |a|^2 + b^T\overline{b} & b^TC^* \\ \\ C\overline{b} & CC^* \end{pmatrix} = \begin{pmatrix} |a|^2 & \overline{a}b^T \\ \\ a\overline{b} & B + C^*C \end{pmatrix} ,$$

sendo B a matriz cujos elementos são $B_{ij} = \overline{b}_i b_j$. Disso extraímos que $b^T\overline{b} = 0$, ou seja, $|b_1|^2 + \cdots + |b_{n-1}|^2 = 0$ e, portanto, $b = \mathbb{0}$. Com isso, ficamos com $CC^* = C^*C$, ou seja, C é normal. Como C é triangular superior então, pela hipótese indutiva, C é diagonal. Isso, mais o fato provado que b é nulo, implica que S é diagonal, provando o lema. ■

10.7 O Teorema de Decomposição de Jordan e a Forma Canônica de Matrizes

Nas seções anteriores demonstramos condições que permitem diagonalizar certas matrizes. Nem todas as matrizes, porém, podem ser diagonalizadas. Podemos nos perguntar, no entanto, quão próximo podemos chegar de uma matriz diagonal.

Mostraremos nesta seção que toda matriz A pode ser levada (por uma transformação de similaridade) à uma forma próxima à diagonal, denominada *forma canônica de Jordan*[25]. Resumidamente (a afirmação precisa será apresentada mais adiante), mostraremos que existe uma matriz P tal que $P^{-1}AP$ tem a seguinte forma:

$$\begin{pmatrix} \lambda_1 & \gamma_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 & \gamma_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \lambda_3 & \gamma_3 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \lambda_4 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \lambda_{n-1} & \gamma_{n-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & \lambda_n \end{pmatrix} , \tag{10.99}$$

[25] Marie Ennemond Camille Jordan (1838–1922). A forma canônica de matrizes foi originalmente descoberta por Weierstrass (Karl Theodor Wilhelm Weierstrass (1815–1897)) e redescoberta por Jordan em 1870.

onde $\lambda_1, \ldots, \lambda_n$ são os autovalores de A e onde os γ_i valem 1 ou 0, mas que forma que a matriz diagonal

$$\begin{pmatrix} \lambda_1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \lambda_4 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \lambda_{n-1} & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & \lambda_n \end{pmatrix}, \tag{10.100}$$

e a matriz supradiagonal

$$\begin{pmatrix} 0 & \gamma_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \gamma_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \gamma_3 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & \gamma_{n-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \tag{10.101}$$

comutam entre si.

O resultado central que provaremos, e do qual as afirmativas feitas acima seguirão, diz que toda matriz A pode ser levada por uma transformação do tipo $P^{-1}AP$ a uma matriz da forma $D + N$, onde D é diagonal e N é nilpotente (ou seja, tal que $N^q = 0$ para algum q) e tais que D e N comutam: $DN = ND$. Essa é a afirmativa principal do célebre "Teorema da Decomposição de Jordan", que demonstraremos nas páginas que seguem.

Esse Teorema da Decomposição de Jordan generaliza os teoremas sobre diagonalizabilidade de matrizes: para matrizes diagonalizáveis tem-se simplesmente $N = 0$ para um P conveniente.

Antes de nos dedicarmos à demonstração desses fatos precisaremos de alguma preparação.

10.7.1 Resultados Preparatórios

• **Somas diretas de subespaços**

Seja V um espaço vetorial e V_1 e V_2 dois de seus subespaços. Dizemos que V é a *soma direta* de V_1 e V_2 se todo vetor v de V puder ser escrito de modo único da forma $v = v_1 + v_2$ com $v_1 \in V_1$ e $v_2 \in V_2$.

Se V é a soma direta de V_1 e V_2 escrevemos $V = V_1 \oplus V_2$.

• **Subespaços invariantes**

Um subespaço $\mathcal{E}$ de $\mathbb{C}^n$ é dito ser *invariante pela ação de uma matriz* A, se $Av \in \mathcal{E}$ para todo $v \in \mathcal{E}$.

Se $V = V_1 \oplus V_2$ e tanto V_1 quanto V_2 são invariantes pela ação de A, escrevemos $A = A_1 \oplus A_2$ onde A_i é A restrita a V_i. Se escolhermos uma base em V da forma $\{v_1, \ldots, v_m, v_{m+1}, \ldots, v_n\}$, onde $\{v_1, \ldots, v_m\}$ é uma base em V_1 e

$\{v_{m+1}, \ldots, v_n\}$ é uma base em V_2, então nessa base A terá a forma

$$A = \begin{pmatrix} A_1 & \mathbb{0}_{m,\, n-m} \\ \mathbb{0}_{n-m,\, m} & A_2 \end{pmatrix} . \tag{10.102}$$

onde $A_1 \in \mathrm{Mat}(\mathbb{C},\ m)$ e $A_2 \in \mathrm{Mat}(\mathbb{C},\ n-m)$.

E. 10.33 *Exercício*. Justifique a forma (10.102). ✠

A representação (10.102) é dita ser uma *representação em blocos diagonais* de A, os blocos sendo as submatrizes A_1 e A_2.

Um fato relevante que decorre imediatamente de (10.102) e da Proposição 10.3, página 472, e que usaremos frequentemente adiante, é que se $A = A_1 \oplus A_2$ então

$$\det(A) = \det(A_1)\det(A_2) .$$

• Operadores nilpotentes

Seja V um espaço vetorial e $N : V \to V$ um operador linear agindo em V. O operador N é dito ser um *operador nilpotente* se existir um inteiro positivo q tal que $N^q = 0$. O menor $q \in \mathbb{N}$ para o qual $N^q = 0$ é dito ser o *índice* de N.

Vamos a alguns exemplos.

E. 10.34 *Exercício*. Verifique que $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ e $\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$ são matrizes nilpotentes de índice 3. ✠

E. 10.35 *Exercício*. Verifique que $\begin{pmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix}$ com $a \neq 0$ e $b \neq 0$ é uma matriz nilpotente de índice 3. ✠

E. 10.36 *Exercício*. Verifique que $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ e $N = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ são matrizes nilpotentes de índice 2. ✠

O seguinte fato sobre os autovalores de operadores nilpotentes será usado adiante.

Proposição 10.31 *Se $N \in \mathrm{Mat}(\mathbb{C},\ n)$ é nilpotente, então seus autovalores são todos nulos. Isso implica que seu polinômio característico é $q_N(x) = x^n$, $x \in \mathbb{C}$. Se o índice de N é q então o polinômio mínimo de N é $m_N(x) = x^q$, $x \in \mathbb{C}$.* □

No Corolário 10.5, página 531, demonstraremos que uma matriz é nilpotente se e somente se seus autovalores forem todos nulos.

Prova da Proposição 10.31. Se $N = 0$ o índice é $q = 1$ e tudo é trivial. Seja $N \neq 0$ com índice $q > 1$. Seja $v \neq 0$ um autovetor de N com autovalor λ: $Nv = \lambda v$. Isso diz que $0 = N^q v = \lambda^q v$. Logo $\lambda^q = 0$ e, obviamente, $\lambda = 0$. É claro então que $q_N(x) = x^n$. Que o polinômio mínimo é $m_N(x) = x^q$ segue do fato que $m_N(x)$ deve ser um divisor de $q_n(x)$ (isso segue do Teorema 10.3 junto com o Teorema de Hamilton-Cayley, Teorema 10.4), página 486). Logo $m_N(x)$ é da forma x^k para algum $k \leq n$. Mas o menor k tal que $m_N(N) = N^k = 0$ é, por definição, igual a q. Isso completa a prova. ■

Mais sobre matrizes nilpotentes será estudado na Seção 10.7.3 onde, em particular, discutiremos a chamada *forma canônica de matrizes nilpotentes.*

• O núcleo e a imagem de um operador linear

Seja V um espaço vetorial e $A : V \to V$ um operador linear agindo em V.

O *núcleo* de A é definido como o conjunto de todos os vetores que são anulados por A:

$$\mathcal{N}(A) := \{x \in V \mid Ax = 0\} .$$

A *imagem* de A é definida por

$$\mathcal{R}(A) \;:=\; \{x \in V \mid \exists\, y \in V \text{ tal que } x = Ay\} \;.$$

Afirmamos que $\mathcal{N}(A)$ e $\mathcal{R}(A)$ são dois subespaços de V. Note-se primeiramente que $0 \in \mathcal{N}(A)$ e $0 \in \mathcal{R}(A)$ (por quê?). Fora isso, se x e $y \in \mathcal{N}(A)$ então, para quaisquer escalares α e β,

$$A(\alpha x + \beta y) \;=\; \alpha Ax + \beta Ay \;=\; 0 \;,$$

provando que combinações lineares $\alpha x + \beta x'$ também pertencem a $\mathcal{N}(A)$. Analogamente se x e $x' \in \mathcal{R}(A)$ então existem y e $y' \in V$ com $x = Ay$, $x' = Ay'$. Logo

$$\alpha x + \beta x' \;=\; A(\alpha y + \beta y') \;,$$

provando que combinações lineares $\alpha x + \beta y$ também pertencem a $\mathcal{R}(A)$.

Para um operador A fixado, e $k \in \mathbb{N}$, vamos definir

$$\mathcal{N}_k \;=\; \mathcal{N}(A^k) \qquad \text{e} \qquad \mathcal{R}_k \;=\; \mathcal{R}(A^k) \;.$$

Esses subespaços $\mathcal{N}_k$ e $\mathcal{R}_k$ são invariantes por A. De fato, se $x \in \mathcal{N}_k$, então $A^k(Ax) = A(A^k x) = A0 = 0$, mostrando que $Ax \in \mathcal{N}_k$. Analogamente, se $x \in \mathcal{R}_k$ então $x = A^k y$ para algum vetor y. Logo, $Ax = A(A^k y) = A^k(Ay)$, mostrando que $Ax \in \mathcal{R}_k$.

Afirmamos que

$$\mathcal{N}_k \subset \mathcal{N}_{k+1} \tag{10.103}$$

e que

$$\mathcal{R}_k \supset \mathcal{R}_{k+1} \;.$$

As demonstrações dessas afirmativas são quase banais. Se $x \in \mathcal{N}_k$ então $A^k x = 0$. Isso obviamente implica $A^{k+1}x = 0$. Logo $x \in \mathcal{N}_{k+1}$ e, portanto, $\mathcal{N}_k \subset \mathcal{N}_{k+1}$. Analogamente, se $x \in \mathcal{R}_{k+1}$ então existe y tal que $x = A^{k+1}y$. Logo $x = A^k(Ay)$, o que diz que $x \in \mathcal{R}_k$. Portanto $\mathcal{R}_{k+1} \subset \mathcal{R}_k$.

Isso diz que os conjuntos $\mathcal{N}_k$ formam uma cadeia crescente de conjuntos:

$$\{0\} \;\subset\; \mathcal{N}_1 \;\subset\; \mathcal{N}_2 \;\subset\; \cdots \;\subset\; \mathcal{N}_k \;\subset\; \cdots \;\subset\; V \;, \tag{10.104}$$

e os $\mathcal{R}_k$ formam uma cadeia decrescente de conjuntos:

$$V \;\supset\; \mathcal{R}_1 \;\supset\; \mathcal{R}_2 \;\supset\; \cdots \;\supset\; \mathcal{R}_k \;\supset\; \cdots \;\supset\; \{0\} \;. \tag{10.105}$$

Consideremos a cadeia crescente (10.104). Como os conjuntos $\mathcal{N}_k$ são subespaços de V, é claro que a cadeia não pode ser estritamente crescente se V for um espaço de dimensão finita, ou seja, deve haver um inteiro positivo p tal que $\mathcal{N}_p = \mathcal{N}_{p+1}$. Seja p o menor número inteiro para o qual isso acontece. Afirmamos que para todo $k \geq 1$ vale $\mathcal{N}_p = \mathcal{N}_{p+k}$.

Vamos provar isso. Se $x \in \mathcal{N}_{p+k}$ então $A^{p+k}x = 0$, ou seja, $A^{p+1}(A^{k-1}x) = 0$. Logo, $A^{k-1}x \in \mathcal{N}_{p+1}$. Dado que $\mathcal{N}_p = \mathcal{N}_{p+1}$, isso diz que $A^{k-1}x \in \mathcal{N}_p$, ou seja, $A^p(A^{k-1}x) = 0$. Isso, por sua vez, afirma que $x \in \mathcal{N}_{p+k-1}$. O que fizemos então foi partir de $x \in \mathcal{N}_{p+k}$ e concluir que $x \in \mathcal{N}_{p+k-1}$. Se repetirmos a argumentação k vezes concluiremos que $x \in \mathcal{N}_p$. Logo, $\mathcal{N}_{p+k} \subset \mathcal{N}_p$. Por (10.103) tem-se, porém, que $\mathcal{N}_p \subset \mathcal{N}_{p+k}$ e, assim, $\mathcal{N}_{p+k} = \mathcal{N}_p$.

Assim, a cadeia (10.104) tem, no caso de V ter dimensão finita, a forma

$$\{0\} \;\subset\; \mathcal{N}_1 \;\subset\; \mathcal{N}_2 \;\subset\; \cdots \;\subset\; \mathcal{N}_p \;=\; \mathcal{N}_{p+1} \;=\; \cdots \;= \mathcal{N}_{p+k} \;=\; \cdots \;\subset\; V \;. \tag{10.106}$$

Como dissemos, p será daqui por diante o menor inteiro para o qual $\mathcal{N}_p = \mathcal{N}_{p+1}$. O lema e o teorema que seguem têm grande importância na demonstração do Teorema de Decomposição de Jordan.

Lema 10.7 *Com as definições acima, $\mathcal{N}_p \cap \mathcal{R}_p = \{0\}$, ou seja, os subespaços $\mathcal{N}_p$ e $\mathcal{R}_p$ têm em comum apenas o vetor nulo.* □

Demonstração. Seja x tal que $x \in \mathcal{N}_p$ e $x \in \mathcal{R}_p$. Isso significa que $A^p x = 0$ e que existe y tal que $x = A^p y$. Logo, $A^{2p} y = A^p x = 0$, ou seja, $y \in \mathcal{N}_{2p}$. Pela definição de p tem-se que $\mathcal{N}_{2p} = \mathcal{N}_p$. Assim, $y \in \mathcal{N}_p$. Logo $A^p y = 0$. Mas, pela própria definição de y valia que $A^p y = x$. Logo $x = 0$. ■

Esse lema tem a seguinte consequência importante.

Teorema 10.19 *Com as definições acima vale que $V = \mathcal{N}_p \oplus \mathcal{R}_p$, ou seja, cada $x \in V$ pode ser escrito de modo único na forma $x = x_n + x_r$, onde $x_n \in \mathcal{N}_p$ e $x_r \in \mathcal{R}_p$.* □

Demonstração. Seja m a dimensão de $\mathcal{N}_p$ e seja $\{u_1, \ldots, u_m\}$ uma base em $\mathcal{N}_p$. Vamos estender essa base, incluindo vetores $\{v_{m+1}, \ldots, v_n\}$ de modo que $\{u_1, \ldots, u_m, v_{m+1}, \ldots, v_n\}$ seja uma base em V. Afirmamos que $\{A^p v_{m+1}, \ldots, A^p v_n\}$ é uma base em $\mathcal{R}_p$. Seja $x \in \mathcal{R}_p$ e seja $y \in V$ tal que $x = A^p y$. Como todo vetor de V, y pode ser escrito como combinação linear de elementos da base $\{u_1, \ldots, u_m, v_{m+1}, \ldots, v_n\}$:

$$y = \sum_{i=1}^{m} \alpha_i u_i + \sum_{i=m+1}^{n} \alpha_i v_i .$$

Logo,

$$x = \sum_{i=1}^{m} \alpha_i A^p u_i + \sum_{i=m+1}^{n} \alpha_i A^p v_i = \sum_{i=m+1}^{n} \alpha_i A^p v_i . \tag{10.107}$$

Os vetores $\{A^p v_{m+1}, \ldots, A^p v_n\}$ são linearmente independentes. Isso se mostra com o seguinte argumento. Se existirem escalares $\beta_{m+1}, \ldots, \beta_n$ tais que $\sum_{i=m+1}^{n} \beta_i A^p v_i = 0$, então teríamos $A^p \left(\sum_{i=m+1}^{n} \beta_i v_i \right) = 0$, ou seja, $\sum_{i=m+1}^{n} \beta_i v_i \in \mathcal{N}_p$. Isso implica que existem constantes $\gamma_1, \ldots, \gamma_m$ tais que $\sum_{i=m+1}^{n} \beta_i v_i = \sum_{i=1}^{m} \gamma_i u_i$, pois os vetores $\{u_1, \ldots, u_m\}$ são uma base em $\mathcal{N}_p$. Ora, como $\{u_1, \ldots, u_m, v_{m+1}, \ldots, v_n\}$ são linearmente independentes, segue que os β_i's e os γ_j's são todos nulos. Isso prova que $\{A^p v_{m+1}, \ldots, A^p v_n\}$ são linearmente independentes e, portanto, por (10.107), formam uma base em $\mathcal{R}_p$.

Isso incidentalmente provou que a dimensão de $\mathcal{R}_p$ é $n - m$. Temos, portanto, que $\dim(\mathcal{N}_p) + \dim(\mathcal{R}_p) = \dim(V)$.

Para $i = m+1, \ldots, n$ defina-se $u_i = A^p v_i$. Afirmamos que o conjunto de vetores

$$\{u_1, \ldots, u_m, u_{m+1}, \ldots, u_n\} = \{u_1, \ldots, u_m, A^p v_{m+1}, \ldots, A^p v_n\}$$

é também linearmente independente e, portanto, forma uma base em V. Suponhamos que haja constantes escalares $\alpha_1, \ldots, \alpha_n$ tais que

$$0 = \sum_{i=1}^{n} \alpha_i u_i = \sum_{i=1}^{m} \alpha_i u_i + A^p \left(\sum_{i=m+1}^{n} \alpha_i v_i \right) .$$

Isso implica, obviamente,

$$\sum_{i=1}^{m} \alpha_i u_i = -A^p \left(\sum_{i=m+1}^{n} \alpha_i v_i \right) .$$

O lado esquerdo dessa igualdade é um elemento de $\mathcal{N}_p$ (pois $u_1, \ldots, u_m$ são uma base em $\mathcal{N}_p$), enquanto que o lado esquerdo é obviamente um elemento da imagem de A^p, ou seja, de $\mathcal{R}_p$. Contudo, já vimos (Lema 10.7) que o único vetor que $\mathcal{N}_p$ e $\mathcal{R}_p$ têm em comum é o vetor nulo. Logo,

$$\sum_{i=1}^{m} \alpha_i u_i = 0 \tag{10.108}$$

e

$$\sum_{i=m+1}^{n} \alpha_i A^p v_i = 0 . \tag{10.109}$$

A relação (10.108) implica $\alpha_1 = \cdots = \alpha_m = 0$, pois $\{u_1, \ldots, u_m\}$ é uma base em $\mathcal{N}_p$. A relação (10.109) implica $\alpha_{m+1} = \cdots = \alpha_n = 0$, pois $\{A^p v_1, \ldots, A^p v_m\}$ é uma base em $\mathcal{R}_p$. Assim, todos os α_i's são nulos, provando que $\{u_1, \ldots, u_m, u_{m+1}, \ldots, u_n\} = \{u_1, \ldots, u_m, A^p v_{m+1}, \ldots, A^p v_n\}$ é um conjunto de n vetores linearmente independentes.

Consequentemente, todo $x \in V$ pode ser escrito na forma

$$x = \sum_{i=1}^{n} \alpha_i u_i = \underbrace{\sum_{i=1}^{m} \alpha_i u_i}_{x_n \in \mathcal{N}_p} + \underbrace{A^p \left(\sum_{i=m+1}^{n} \alpha_i v_i \right)}_{x_r \in \mathcal{R}_p} .$$

Provar a unicidade dessa decomposição fica como exercício. Isso completa a demonstração. ■

Uma das coisas que o teorema que acabamos de demonstrar diz é que, dado um operador A, o espaço V pode ser decomposto em uma soma direta de dois subespaços, invariantes por A: um onde A é nilpotente, $\mathcal{N}_p$, e outro onde A é inversível, $\mathcal{R}_p$. A é nilpotente em $\mathcal{N}_p$ pois $A^p x = 0$ para todo elemento x de $\mathcal{N}_p$. A é inversível em $\mathcal{R}_p$ pois se $x \in \mathcal{R}_p$ é tal que $Ax = 0$ isso implica $x \in \mathcal{N}_1 \subset \mathcal{N}_p$. Mas x só pode pertencer a $\mathcal{N}_p$ e a $\mathcal{R}_p$ se for nulo. Logo, em $\mathcal{R}_p$, $Ax = 0$ se e somente se $x = 0$, provando que A é inversível[26]. Para referência futura formulemos essa afirmativa na forma de um teorema:

Teorema 10.20 *Se A é um operador linear não-nulo agindo em um espaço vetorial $V = \mathbb{C}^n$ então é possível decompor V em dois subespaços invariantes por A, $V = \mathcal{S} \oplus \mathcal{T}$, de forma que A restrito a $\mathcal{S}$ é nilpotente, enquanto que A restrito a $\mathcal{T}$ é inversível.* □

Esse será o teorema básico do qual extrairemos a demonstração do Teorema de Decomposição de Jordan.

10.7.2 O Teorema da Decomposição de Jordan

Chegamos agora ao resultado mais importante desta seção, o Teorema da Decomposição de Jordan[27], um importante teorema estrutural sobre matrizes de importância em vários campos, por exemplo na teoria das equações diferenciais ordinárias. Para tais aplicações, vide Capítulo 14, página 668.

O Teorema da Decomposição de Jordan também tem certa relevância na Teoria de Grupos, e o usaremos para provar que toda matriz $n \times n$ complexa inversível (ou seja, todo elemento do grupo $\mathrm{GL}(\mathbb{C}, n)$) pode ser escrita como exponencial de outra matriz (Proposição 11.12, página 599). No Capítulo 11 usaremos o Teorema da Decomposição de Jordan para provar a identidade útil $\det(e^A) = e^{\mathrm{Tr}(A)}$, válida para qualquer matriz $n \times n$ real ou complexa (Proposição 11.7, página 594). Vide também Proposição 10.14, página 480.

• **Enunciado e demonstração do Teorema da Decomposição de Jordan**

Teorema 10.21 (Teorema da Decomposição de Jordan) *Seja A um operador linear agindo no espaço $V = \mathbb{C}^n$ e seja $\{\alpha_1, \ldots, \alpha_r\}$ o conjunto de seus autovalores distintos. Então, existem r subespaços $\mathcal{S}_1, \ldots, \mathcal{S}_r$ tais que $V = \mathcal{S}_1 \oplus \ldots \oplus \mathcal{S}_r$ e tais que cada $\mathcal{S}_i$ é invariante por A. Ou seja, $A = A_1 \oplus \ldots \oplus A_r$, onde A_i é A restrita a $\mathcal{S}_i$. Fora isso, cada A_i, é da forma $A_i = \alpha_i \mathbb{1}_i + N_i$, onde $\mathbb{1}_i$ é a matriz identidade em $\mathcal{S}_i$ e onde N_i é nilpotente. Por fim, a dimensão s_i de cada subespaço $\mathcal{S}_i$ é igual à multiplicidade algébrica do autovalor α_i.* □

Demonstração. Seja $\{\alpha_1, \ldots, \alpha_r\}$ o conjunto dos autovalores distintos de A e seja n_i a multiplicidade algébrica do autovalor α_i. Seja $A_1 = A - \alpha_1 \mathbb{1}$. Pelo Teorema 10.20, página 529, V pode ser escrito como $V = \mathcal{S}_1 \oplus \mathcal{T}_1$, onde $\mathcal{S}_1$ e $\mathcal{T}_1$ são invariantes por A_1, sendo A_1 nilpotente em $\mathcal{S}_1$ e inversível em $\mathcal{T}_1$. Assim, A_1 é da forma $A_1 = N_1 \oplus M_1$ com N_1 nilpotente e M_1 inversível. Logo

$$A = \alpha_1 \mathbb{1} + A_1 = (\alpha_1 \mathbb{1}_{\mathcal{S}_1} + N_1) \oplus (\alpha_1 \mathbb{1}_{\mathcal{T}_1} + M_1) , \tag{10.110}$$

[26] Lembre-se que esse argumento só funciona em espaços vetoriais V que tenham dimensão finita, o que estamos supondo aqui.

[27] Marie Ennemond Camille Jordan (1838–1922). A forma canônica de matrizes (que será discutida mais adiante) foi originalmente descoberta por Weierstrass (Karl Theodor Wilhelm Weierstrass (1815–1897)) e redescoberta por Jordan em 1870.

onde $\mathbb{1}_{\mathcal{S}_1}$ é a matriz identidade em $\mathcal{S}_1$ etc. Vamos mostrar que a dimensão de $\mathcal{S}_1$ é igual à multiplicidade algébrica de α_1. Por (10.110) o polinômio característico de A é

$$q_A(\lambda) \; = \; \det(\lambda\mathbb{1} - A) \; = \; \det((\lambda - \alpha_1)\mathbb{1}_{\mathcal{S}_1} - N_1) \; \det((\lambda - \alpha_1)\mathbb{1}_{\mathcal{T}_1} - M_1) \; .$$

Se q_{N_1} denota o polinômio característico de N_1, tem-se

$$\det((\lambda - \alpha_1)\mathbb{1}_{\mathcal{S}_1} - N_1) \; = \; q_{N_1}(\lambda - \alpha_1) \; = \; (\lambda - \alpha_1)^{s_1} \; ,$$

onde, na última igualdade, usamos a Proposição 10.31, página 526, sobre a forma do polinômio característico de uma matriz nilpotente. Daí, segue que $q_A(\lambda) = (\lambda - \alpha_1)^{s_1} \, q_{M_1}(\lambda - \alpha_1)$, sendo q_{M_1} o polinômio característico de M_1. Como M_1 é inversível, M_1 não tem o zero como autovalor. Logo, $q_{M_1}(0) \neq 0$. Portanto s_1 é igual à multiplicidade de α_1 como raiz de q_A, ou seja, é igual a n_1, a multiplicidade algébrica de α_1.

A ideia agora é prosseguir decompondo agora o operador $\alpha_1\mathbb{1}_{\mathcal{T}_1} + M_1$ que aparece em (10.110) da mesma maneira como fizermos acima com A.

Seja $A' = \alpha_1\mathbb{1}_{\mathcal{T}_1} + M_1$ e que age em $\mathcal{T}_1$, que é um espaço de dimensão $n - n_1$. Definimos $A_2 = A' - \alpha_2\mathbb{1}_{\mathcal{T}_1}$.

Evocando novamente o Teorema 10.20, página 529, $\mathcal{T}_1$ pode ser escrito como $\mathcal{T}_1 = \mathcal{S}_2 \oplus \mathcal{T}_2$, onde $\mathcal{S}_2$ e $\mathcal{T}_2$ são invariantes por A_2, sendo A_2 nilpotente em $\mathcal{S}_2$ e inversível em $\mathcal{T}_2$. Assim, $V = \mathcal{S}_1 \oplus \mathcal{S}_2 \oplus \mathcal{T}_2$. Agindo em $\mathcal{T}_1 = \mathcal{S}_2 \oplus \mathcal{T}_2$, A_2 é da forma $A_2 = N_2 \oplus M_2$ com N_2 nilpotente e M_2 inversível. Logo

$$A' \; = \; \alpha_2\mathbb{1}_{\mathcal{T}_1} + A_2 \; = \; (\alpha_2\mathbb{1}_{\mathcal{S}_2} + N_2) \oplus (\alpha_2\mathbb{1}_{\mathcal{T}_2} + M_2) \; . \tag{10.111}$$

Vamos, como acima, mostrar que a dimensão de $\mathcal{S}_2$ é igual à multiplicidade algébrica de α_2.

Pela definição,

$$A = (\alpha_1\mathbb{1}_{\mathcal{S}_1} + N_1) \oplus A' \; = \; (\alpha_1\mathbb{1}_{\mathcal{S}_1} + N_1) \oplus (\alpha_2\mathbb{1}_{\mathcal{S}_2} + N_2) \oplus (\alpha_2\mathbb{1}_{\mathcal{T}_2} + M_2) \; .$$

Logo,

$$q_A(\lambda) \; = \; \det\left((\lambda - \alpha_1)\mathbb{1}_{\mathcal{S}_1} - N_1\right) \; \det\left((\lambda - \alpha_2)\mathbb{1}_{\mathcal{S}_2} - N_2\right) \; \det\left((\lambda - \alpha_2)\mathbb{1}_{\mathcal{T}_2} - M_2\right) \; .$$

Portanto, pelos mesmos argumentos usados acima,

$$q_A(\lambda) \; = \; (\lambda - \alpha_1)^{n_1} \, (\lambda - \alpha_2)^{s_2} \, q_{M_2}(\lambda - \alpha_2) \; .$$

Como M_2 é inversível, M_2 não tem autovalor zero e, assim, $q_{M_2}(0) \neq 0$. Logo, $s_2 = n_2$. $\mathcal{T}_2$ é assim um subespaço de dimensão $n - n_1 - n_2$.

Prosseguindo nas mesmas linhas, após r passos chegaremos a um subespaço $\mathcal{T}_r$ de dimensão $n - n_1 - \cdots - n_r = 0$ (por (10.30), página 473). Aí, teremos $V = \mathcal{S}_1 \oplus \cdots \oplus \mathcal{S}_r$, onde cada $\mathcal{S}_i$ tem dimensão n_i e

$$A = (\alpha_1\mathbb{1}_{\mathcal{S}_1} + N_1) \oplus \cdots \oplus (\alpha_r\mathbb{1}_{\mathcal{S}_r} + N_r) \; ,$$

onde os N_i's são todos nilpotentes. Isso completa a demonstração. ■

Um corolário importante do Teorema de Decomposição de Jordan é o seguinte:

Teorema 10.22 *Para toda matriz $A \in \mathrm{Mat}(\mathbb{C}, \; n)$ existe uma matriz inversível $P \in \mathrm{Mat}(\mathbb{C}, \; n)$ tal que $P^{-1}AP = D + N$, onde D é uma matriz diagonal formada pelos autovalores de A e N é uma matriz nilpotente e de tal forma que D e N comutam: $DN = ND$.*

Consequentemente, toda matriz $A \in \mathrm{Mat}(\mathbb{C}, \; n)$ pode ser escrita na forma $A = A_d + A_n$ com $A_d A_n = A_n A_d$, sendo A_d diagonalizável e A_n nilpotente, a saber, $A_d = PDP^{-1}$ e $A_n = PNP^{-1}$, com D e N dados acima. □

Demonstração do Teorema 10.22. O Teorema 10.21 está dizendo que, numa base conveniente, A tem a forma de blocos diagonais

$$A \;=\; \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_r \end{pmatrix} \;=\; \begin{pmatrix} \alpha_1 \mathbb{1}_{s_1} + N_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 \mathbb{1}_{s_2} + N_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_r \mathbb{1}_{s_r} + N_r \end{pmatrix} , \tag{10.112}$$

ou seja,

$$A \;=\; D + N \, ,$$

onde

$$D \;=\; \begin{pmatrix} \alpha_1 \mathbb{1}_{s_1} & 0 & \cdots & 0 \\ 0 & \alpha_2 \mathbb{1}_{s_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_r \mathbb{1}_{s_r} \end{pmatrix} \;=\; \mathrm{diag}\left(\underbrace{\alpha_1, \, \ldots, \, \alpha_1}_{s_1 \text{ vezes}}, \, \ldots, \, \underbrace{\alpha_r, \, \ldots, \, \alpha_r}_{s_r \text{ vezes}} \right)$$

e

$$N \;=\; \begin{pmatrix} N_1 & 0 & \cdots & 0 \\ 0 & N_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & N_r \end{pmatrix} . \tag{10.113}$$

Acima s_i é a dimensão do subespaço $\mathcal{S}_i$.

É fácil de se ver que N é uma matriz nilpotente, pois se o k_i é o índice de N_i (ou seja, k_i é o menor inteiro positivo para o qual $N_i^{k_i} = 0$), então para $k := \max(k_1, \, \ldots, \, k_r)$ tem-se

$$N^k \;=\; \begin{pmatrix} (N_1)^k & 0 & \cdots & 0 \\ 0 & (N_2)^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (N_r)^k \end{pmatrix} \;=\; 0 \, .$$

Em verdade, $k = \max(k_1, \, \ldots, \, k_r)$ é o índice de N (por quê?).

Por fim, como cada N_i comuta com $\alpha_i \mathbb{1}_{s_i}$, fica claro que D e N comutam. Isso completa a demonstração. ■

Corolário 10.5 *Uma matriz $M \in \mathrm{Mat}(\mathbb{C}, n)$ é nilpotente se e somente se todos os seus autovalores forem nulos.* □

Prova. A Proposição 10.31, página 526, afirma que se M é nilpotente todos os seus autovalores são nulos. O Teorema 10.22, página 530, afirma que se os autovalores de M são nulos, então existe P tal que $P^{-1}MP = N$, nilpotente. Isso implica que M é nilpotente. ■

10.7.3 Matrizes Nilpotentes e sua Representação Canônica

Os teoremas que estudamos acima nesta seção revelam a importância de matrizes nilpotentes. Um fato relevante é que elas podem ser representadas de uma forma especial, denominada forma canônica, da qual traremos logo abaixo. Antes, alguma preparação se faz necessária.

Seja $N \in \mathrm{Mat}(\mathbb{C},\ n)$ uma matriz nilpotente de índice q, ou seja, $N^q = 0$, mas $N^{q-1} \neq 0$. Para uso futuro, provemos o seguinte lema:

Lema 10.8 *Seja N uma matriz nilpotente de índice q. Estão existe um vetor $v \neq 0$ tal que os q vetores*

$$v, \qquad Nv, \qquad N^2v, \qquad \ldots, \qquad N^{q-1}v\ , \tag{10.114}$$

são linearmente independentes. Fora isso, o subespaço q-dimensional $J_{v,\,q} := \langle v,\ Nv,\ N^2v,\ \ldots,\ N^{q-1}v\rangle$ de V gerado por esses q vetores é invariante por N. □

Prova. Se $q = 1$, então $N = 0$ e não há nada a provar, pois a afirmação é trivialmente verdadeira para qualquer $v \neq 0$. Seja então $q > 1$ (em cujo caso $N \neq 0$, trivialmente). Sabemos, por hipótese, que a matriz N^{q-1} é não-nula. Isso significa que existe pelo menos um vetor $v \neq 0$ tal que $N^{q-1}v \neq 0$. Fixemos um tal vetor. É imediato que os vetores $Nv,\ N^2v,\ \ldots,\ N^{q-1}v$ são todos não nulos pois, se tivéssemos $N^jv = 0$ para algum $1 \leq j < q-1$, então, aplicando-se N^{q-1-j} à esquerda, teríamos $N^{q-1}v = 0$, uma contradição.

Sejam agora $\alpha_1,\ \ldots, \alpha_q$ escalares tais que

$$\alpha_1 v + \alpha_2 Nv + \alpha_3 N^2 v + \cdots + \alpha_q N^{q-1}v \ = \ 0\ . \tag{10.115}$$

Aplicando-se N^{q-1} nessa igualdade e lembrando que $N^q = 0$, concluímos que $\alpha_1 N^{q-1}v = 0$. Como $N^{q-1}v \neq 0$, segue que $\alpha_1 = 0$ e, com isso, (10.115) fica

$$\alpha_2 Nv + \alpha_3 N^2 v + \cdots + \alpha_q N^{q-1}v \ = \ 0\ . \tag{10.116}$$

Aplicando agora N^{q-2} nessa igualdade concluímos que $\alpha_2 = 0$. Prosseguindo, concluímos depois de q passos que todos os escalares α_j são nulos. Isso prova que os q vetores de (10.114) são linearmente independentes.

Que o subespaço $J_{v,\,q}$ definido acima é invariante por N é evidente pois, para quaisquer escalares $\beta_1,\ \ldots,\ \beta_q$, tem-se

$$N\left(\beta_1 v + \beta_2 Nv + \cdots + \beta_q N^{q-1}v\right) \ = \ \beta_1 Nv + \beta_2 N^2 v + \cdots + \beta_{q-1}N^{q-1}v \ \in \ J_{v,\,q}\ .$$

■

O seguinte teorema é central para o que segue.

Teorema 10.23 *Se N é uma matriz nilpotente de índice q agindo em V e v um vetor com a propriedade que $N^{q-1}v \neq 0$, então existe um subespaço K de V tal que $J_{v,\,q} \cap K = \{0\}$, tal que $V = J_{v,\,q} \oplus K$ e tal que K é também invariante por N.* □

Prova.[28] A prova é feita por indução em q. Note-se que se $q = 1$, então $N = 0$ e a afirmativa é trivial, pois podemos tomar como v qualquer vetor não-nulo, $J_{v,\,q}$ seria o subespaço gerado por esse v e K o subespaço complementar a v, que é trivialmente invariante por N, pois $N = 0$.

[28] Extraída, com modificações, de [157].

Vamos supor então que a afirmação seja válida para matrizes nilpotentes de índice $q-1$ e provar que a mesma é válida para matrizes nilpotentes de índice q. O que desejamos é construir um subespaço K com as propriedades desejadas, ou seja, tal que $V = J_{v,\, q} \oplus K$, sendo K invariante por N.

Seja $V_0 = \mathcal{R}(N)$ o conjunto imagem de N. Sabemos que V_0 é um subespaço de V e que é invariante por N. Fora isso, N é nilpotente de índice $q-1$ agindo em V_0 (por quê?)

Seja $v_0 = Nv \in V_0$. É claro que $N^{q-2}v_0 = N^{q-1}v \neq 0$. Assim, pelo Lema 10.8, o subespaço $(q-1)$-dimensional

$$J_{v_0,\, q-1} \;=\; \langle v_0,\; Nv_0,\; \ldots,\; N^{q-2}v_0\rangle \;=\; \langle Nv,\; N^2v,\; \ldots,\; N^{q-1}v\rangle \;=\; J_{Nv,\, q-1}\;,$$

que é um subespaço de V_0, é invariante por N e, da hipótese indutiva, concluímos que existe um subespaço K_0 de V_0 que é invariante por N tal que $J_{Nv,\, q-1} \cap K_0 = \{0\}$ e tal que $V_0 = J_{Nv,\, q-1} \oplus K_0$.

Seja agora $K_1 := \{x \in V |\; Nx \in K_0\}$. Vamos provar a seguinte afirmação:

I. Todo vetor x de V pode ser escrito na forma $x = y + z$ onde $y \in J_{v,\, q}$ e $z \in K_1$.

Para provar isso, notemos que para qualquer $x \in V$ vale certamente que $Nx \in V_0$. Portanto, como pela hipótese indutiva $V_0 = J_{Nv,\, q-1} \oplus K_0$, podemos escrever $Nx = y' + z'$, com $y' \in J_{Nv,\, q-1}$ e $z' \in K_0$. Como $y' \in J_{Nv,\, q-1}$, y' é da forma de uma combinação linear $y' = \alpha_1 Nv + \cdots + \alpha_{q-1}N^{q-1}v = Ny$, onde $y := \alpha_1 v + \alpha_2 Nv + \cdots + \alpha_{q-1}N^{q-2}v$ é um elemento de $J_{v,\, q}$. Logo, $z' = N(x-y)$. Como $z' \in K_0$, segue que $z := x - y \in K_1$. Assim, $x = y + z$, com $y \in J_{v,\, q}$ e $z \in K_1$. Isso provou **I**.

Note que a afirmação feita em **I** não significa que $V = J_{v,\, q} \oplus K_1$, pois os subespaços $J_{v,\, q}$ e K_1 podem ter uma intersecção não-trivial. Tem-se, porém, o seguinte:

II. $J_{v,\, q} \cap K_0 = \{0\}$.

Provemos essa afirmação. Seja $x \in J_{v,\, q} \cap K_0$. Como $x \in J_{v,\, q}$, x é da forma $x = \alpha_1 v + \alpha_2 Nv + \cdots + \alpha_q N^{q-1}v$. Logo $Nx = \alpha_1 Nv + \alpha_2 N^2 v + \cdots + \alpha_{q-1}N^{q-1}v \in J_{Nv,\, q-1}$. Agora, como $x \in K_0$ e, por hipótese, K_0 é invariante por N, segue que $Nx \in K_0$. Logo, $Nx \in J_{Nv,\, q-1} \cap K_0$. Todavia, mencionamos acima que $J_{Nv,\, q-1} \cap K_0 = \{0\}$. Logo, $Nx = 0$, ou seja, $0 = Nx = \alpha_1 Nv + \alpha_2 N^2 v + \cdots + \alpha_{q-1}N^{q-1}v$. Como os vetores $Nv,\; \ldots,\; N^{q-1}v$ são linearmente independentes, concluímos que $\alpha_1 = \cdots \alpha_{q-1} = 0$. Logo, $x = \alpha_q N^{q-1}v$. Isso significa que $x \in J_{Nv,\, q-1}$. Demonstramos, então, que se $x \in J_{v,\, q} \cap K_0$ então $x \in J_{Nv,\, q-1} \cap K_0$ mas, como $J_{Nv,\, q-1} \cap K_0 = \{0\}$, segue que $x = 0$. Isso conclui a prova de **II**.

III. K_0 e $J_{v,\, q} \cap K_1$, são dois subespaços disjuntos de K_1.

A demonstração é muito simples. É evidente que $J_{v,\, q} \cap K_1$ é subespaço de K_1. Como K_0 é invariante pela ação de N, segue que se $x \in K_0$ então $Nx \in K_0$. Pela definição, isso diz que $x \in K_1$ e concluímos que K_0 é um subespaço e K_1.

Que K_0 e $J_{v,\, q} \cap K_1$ são subespaços disjuntos, segue do fato que

$$K_0 \cap (J_{v,\, q} \cap K_1) \;=\; K_1 \cap (J_{v,\, q} \cap K_0) \;\stackrel{II}{=}\; K_1 \cap \{0\} = \{0\}\;.$$

A afirmação **III** implica que $K_1 = (J_{v,\, q} \cap K_1) \oplus K_0 \oplus K_0'$ para algum subespaço K_0' de K_1 (não necessariamente único). Seja agora $K := K_0 \oplus K_0'$. Note que $K_1 = (J_{v,\, q} \cap K_1) \oplus K$ e, portanto,

$$(J_{v,\, q} \cap K_1) \cap K \;=\; \{0\}\;. \tag{10.117}$$

Provaremos que esse K possui as propriedades desejadas, ou seja, que $V = J_{v,\, q} \oplus K$, sendo K invariante por N. Isso é feito em três passos.

1. $J_{v,\, q}$ e K são subespaços disjuntos, ou seja, $J_{v,\, q} \cap K = \{0\}$, pois, como $K \subset K_1$, segue que $K = K \cap K_1$ e, portanto,

$$J_{v,\, q} \cap K \;=\; J_{v,\, q} \cap (K \cap K_1) \;=\; (J_{v,\, q} \cap K_1) \cap K \;\stackrel{(10.117)}{=}\; \{0\}\;.$$

2. $J_{v,\, q} \oplus K$ contém os vetores de $J_{v,\, q}$ e de $(J_{v,\, q} \cap K_1) \oplus K = K_1$. Por **I**, isso implica que $J_{v,\, q} \oplus K = V$.

3. K é invariante por N, pois o fato que $K \subset K_1$, implica, pela definição de K_1, que $NK \subset NK_1 \subset K_0 \subset K$.

A prova do Teorema 10.23 está completa ■

A principal consequência do Teorema 10.23 é a seguinte.

Proposição 10.32 *Seja $N \in \mathrm{Mat}(\mathbb{C},\, n)$ uma matriz nilpotente de índice q. Então, existem*

1. *um inteiro positivo r, com $1 \leq r \leq n$,*
2. *r números inteiros positivos $n \geq q_1 \geq q_2 \geq \cdots \geq q_r \geq 1$, com $q_1 + \cdots + q_r = n$,*
3. *r vetores $v_1,\, \ldots,\, v_r$ satisfazendo $N^{q_j} v_j = 0$ mas $N^{q_j - 1} v_j \neq 0$, $j = 1,\, \ldots,\, r$,*

tais que

$$V \;=\; J_{v_1,\, q_1} \oplus \cdots \oplus J_{v_r,\, q_r} \;.$$

□

Prova. Se $q = 1$ então $N = 0$. Basta tomar $r = n$ e escolher $v_1,\, \ldots,\, v_n$ uma base qualquer em V. Os q_j's são todos iguais a 1.

Consideremos então $q > 1$ com $N \neq 0$. Tomemos $q_1 = q$. Pelo Teorema 10.23, existem um vetor $v_1 \neq 0$ e um subespaço K^1, invariante por N tais que

$$V \;=\; J_{v_1,\, q_1} \oplus K^1 \;.$$

Como K^1 é invariante por N, podemos também dizer que a matriz N é nilpotente quando restrita a K^1 (já que é nilpotente em todo V). Denotemos por q_2 o índice de N quando restrita a K^1. É claro que $q_2 \leq q = q_1$.

Assim, podemos aplicar o Teorema 10.23 para a matriz N restrita a K^1 e concluir que existe $v_2 \neq 0$ em K^1 e um subespaço K^2 de K^1, invariante por N, tais que $K^1 = J_{v_2,\, q_2} \oplus K^2$. Note que $N^{q_2} v_2 = 0$, pois $v_2 \in K^1$.

Com isso, temos

$$V \;=\; J_{v_1,\, q_1} \oplus J_{v_2,\, q_2} \oplus K^2 \;.$$

Novamente K^2 é invariante por N e, como K^2 é um subespaço de K^1. O índice de N em K^2 será $q_3 \leq q_2 \leq q_1$.

O espaço V tem dimensão finita. Assim, a prova se concluí repetindo o procedimento acima um número finito r de vezes. Note que $N^{q_j} v_j = 0$, pois $N^{q_1} v_1 = 0$, e $v_j \in K^{j-1}$ para todo $j = 2,\, \ldots, r$. ■

Pela construção acima, é claro que $q_1 + \cdots + q_r = n$, a dimensão de V, e que os n vetores

$$v_1,\; Nv_1,\; \ldots,\; N^{q_1-1}v_1,\; v_2,\; Nv_2,\; \ldots,\; N^{q_2-1}v_2, \ldots,\; v_r,\; Nv_r,\; \ldots,\; N^{q_r-1}v_r$$

são linearmente independentes e formam uma base em V. Vamos denotá-los (na ordem em que aparecem acima) por $b_1,\, \ldots,\, b_n$.

Note agora que, pela construção, $Nb_j = b_{j+1}$, para j em cada um dos conjuntos

$$\{1,\, \ldots,\, q_1 - 1\}, \quad \{1 + q_1,\, \ldots,\, q_1 + q_2 - 1\}, \quad \{1 + q_1 + q_2,\, \ldots,\, q_1 + q_2 + q_3 - 1\}\;, \quad \ldots \quad \{1 + q_1 + \cdots + q_{r-1},\, \ldots,\, q_1 + \cdots + q_r - 1\}\;, \tag{10.118}$$

com $l = 0,\, \ldots,\, r-1$, sendo que $Nb_j = 0$ para todo j na forma $q_1 + \cdots + q_l$, $l = 1,\, \ldots,\, r$.

E. 10.37 *Exercício importante para compreender o que segue.* Justifique as últimas afirmações. ✠

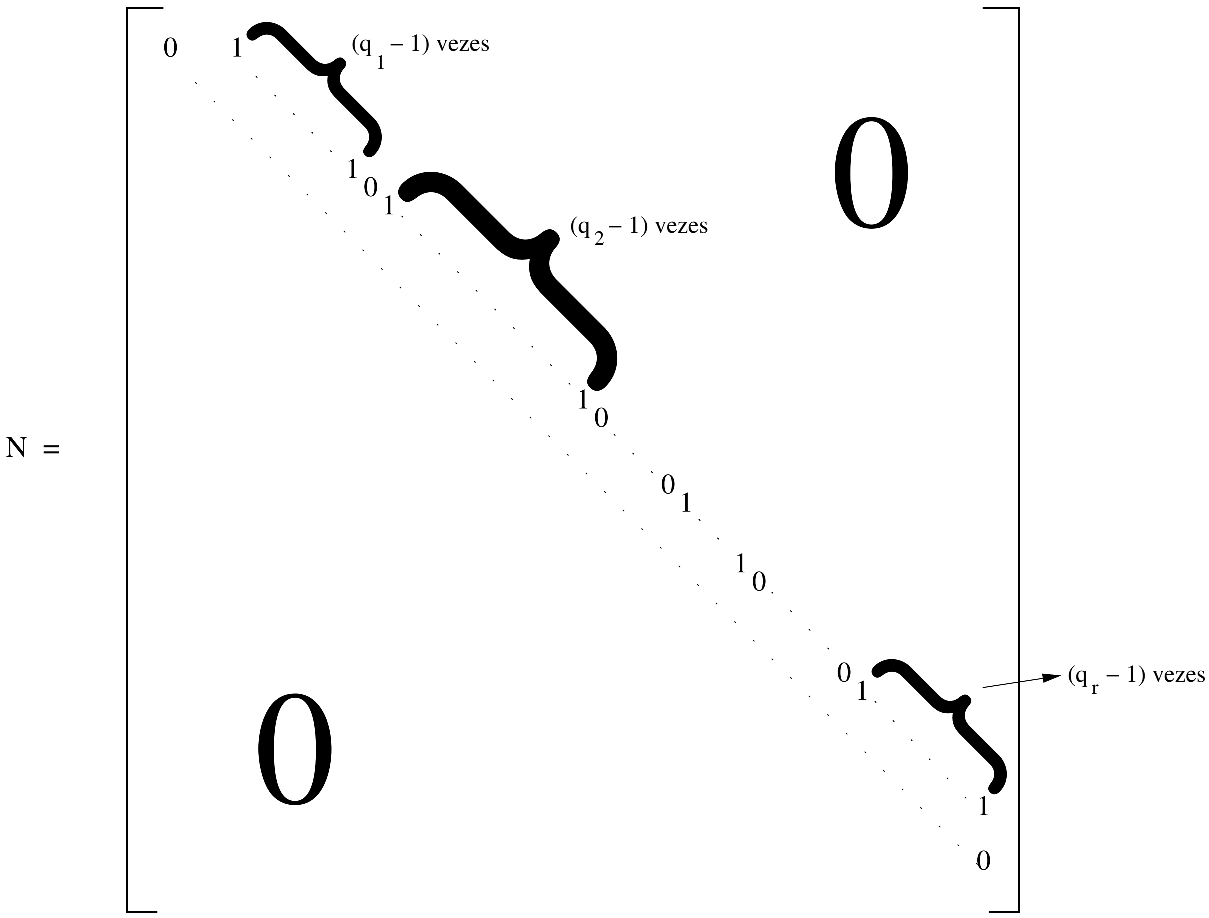

Figura 10.5: Forma canônica típica de uma matriz nilpotente N. Os elementos da primeira supradiagonal podem valer 0 ou 1. Todos os demais elementos de matriz são nulos.

Isso significa que na base $b_1, \ldots, b_n$ os elementos de matriz de N são todos nulos exceto aqueles na forma $N_{j,\,j+1}$ com j em algum dos conjuntos listados em (10.118), em cujo caso $N_{j,\,j+1} = 1$. Pictoricamente, isso diz-nos que na base $b_1, \ldots, b_n$ a matriz N assume uma forma genericamente ilustrada na Figura 10.5. Essa é a denominada *forma canônica da matriz nilpotente* N ou *representação canônica da matriz nilpotente* N, que descrevemos mais detalhadamente no que segue.

Os elementos da diagonal principal são todos nulos. Os únicos elementos não nulos da matriz podem estar localizados apenas na diagonal imediatamente acima da principal, ou seja, aquela diagonal formada por elementos de matriz do tipo $N_{j,\,j+1}$ com $j = 1, \ldots, n-1$. Chamaremos essa diagonal de *primeira supradiagonal*. Os elementos da primeira supradiagonal podem ser 0 ou 1, da forma seguinte: a primeira supradiagonal possuirá r fileiras. As primeiras $r-1$ fileiras são formadas por q_j elementos, $j = 1, \ldots, n-1$, sendo os primeiros $q_j - 1$ elementos iguais a 1 e o último igual a 0. A última fileira terá $q_r - 1$ elementos iguais a 1. Assim, se $q_r = 1$, o último elemento da primeira supradiagonal será nulo, proveniente da $(r-1)$-ésima fileira (essa é a única forma de aparecer um zero no último elemento da primeira supradiagonal).

Note que zeros consecutivos podem ocorrer, se tivermos alguns q_j's iguais a 1. Note também que os elementos da primeira supradiagonal podem ser todos nulos (o que valerá se $r = n$, em cujo caso $q_1 = \cdots = r_n = 1$. Isso só pode ocorrer se $N = 0$ e, nesse caso, $q = 1$) ou todos iguais a 1 (o que valerá se $r = 1$, em cujo caso $q_1 = n$).

10.7.4 A Forma Canônica de Matrizes

Finalizamos esta seção e nossa discussão sobre o Teorema da Decomposição de Jordan e suas consequências reunindo o que descobrimos até aqui.

Se $A \in \mathrm{Mat}(\mathbb{C},\ n)$ o Teorema 10.21, página 529 ensinou-nos que numa base conveniente (ou seja, por uma transformação de similaridade $P_0^{-1}AP_0$), toda matriz A tem a forma de blocos diagonais:

$$P_0^{-1}AP_0 \;=\; \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_r \end{pmatrix} \;=\; \begin{pmatrix} \alpha_1 \mathbb{1}_{n_1} + N_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 \mathbb{1}_{n_2} + N_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_r \mathbb{1}_{n_r} + N_r \end{pmatrix}, \tag{10.119}$$

sendo $\alpha_1, \ldots, \alpha_r$ os autovalores distintos de A. O j-ésimo bloco é de tamanho $n_j \times n_j$, sendo que n_j é a multiplicidade algébrica do autovalor α_j. As matrizes N_j são nilpotentes.

Cada matriz N_j pode ser levada à sua forma canônica N_j^c (tal como explicado na Figura 10.5, página 535, e no que se lhe segue) em uma base conveniente, ou seja, por uma transformação de similaridade $P_j^{-1}N_jP_j$. Assim, definindo

$$P \;=\; \begin{pmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P_r \end{pmatrix}, \tag{10.120}$$

vemos que $P^{-1}(P_0^{-1}AP_0)P = (P_0P)^{-1}A(P_0P)$, sendo que, por (10.119),

$$
\begin{aligned}
P^{-1}(P_0^{-1}AP_0)P &= \begin{pmatrix} P_1^{-1}\left(\alpha_1\mathbb{1}_{n_1}+N_1\right)P_1 & 0 & \cdots & 0 \\ \\ 0 & P_2^{-1}\left(\alpha_2\mathbb{1}_{n_2}+N_2\right)P_1 & \cdots & 0 \\ \\ \vdots & \vdots & \ddots & \vdots \\ \\ 0 & 0 & \cdots & P_r^{-1}\left(\alpha_r\mathbb{1}_{n_r}+N_r\right)P_r \end{pmatrix} \\
&= \begin{pmatrix} \alpha_1\mathbb{1}_{n_1}+N_1^c & 0 & \cdots & 0 \\ \\ 0 & \alpha_2\mathbb{1}_{n_2}+N_2^c & \cdots & 0 \\ \\ \vdots & \vdots & \ddots & \vdots \\ \\ 0 & 0 & \cdots & \alpha_r\mathbb{1}_{n_r}+N_r^c \end{pmatrix} . \qquad (10.121)
\end{aligned}
$$

E. 10.38 *Exercício*. Complete os detalhes. ✠

A matriz final de (10.121) é denominada *forma canônica da matriz* A, ou *forma canônica de Jordan da matriz* A. Como dissemos, toda matriz A assume essa forma numa certa base. Devido ao fato de todos as submatrizes nilpotentes N_j^c terem a forma canônica, os únicos elementos não nulos da forma canônica da matriz A podem estar ou na diagonal principal (sendo estes os autovalores de A, cada um aparecendo em uma fileira de n_j elementos), ou na primeira supradiagonal, sendo que estes valem apenas 0 ou 1 e seguem as regras descritas acima. Isso é ilustrado na Figura 10.6,

A Figura 10.6, mostra a forma canônica de uma matriz que possui 4 autovalores distintos α_1, α_2, α_3 e α_4. A primeira supradiagonal é formada pela sequência de números

$$
\gamma_1^1, \ \ldots, \ \gamma_1^a, \ 0, \ \gamma_1^1, \ \ldots, \ \gamma_1^b, \ 0, \ \gamma_1^1, \ \ldots, \ \gamma_1^c, \ 0, \ \gamma_1^1, \ \ldots, \ \gamma_1^d \ , \qquad (10.122)
$$

sendo que os γ_i^j assumem apenas os valores 0 ou 1, de acordo com as regras explicadas acima quando discutimos a forma canônica de matrizes nilpotentes. Todos os elementos fora da diagonal principal e da primeira supradiagonal são nulos. O primeiro bloco é de dimensão $(a+1)\times(a+1)$, o segundo bloco é de dimensão $(b+1)\times(b+1)$ etc., sendo $a+1$ a multiplicidade algébrica de α_1, $b+1$ a multiplicidade algébrica de α_2 etc.

É interessante notar que na primeira supradiagonal, sempre ocorrem zeros nos pontos localizados fora dos blocos, ou seja, nos pontos onde ocorrem transições entre dois autovalores distintos (indicados por setas na Figura 10.6). Esses são os zeros que ocorrem explicitamente na lista (10.122).

Por fim, comentamos que a forma canônica não é exatamente única, pois é possível ainda fazer transformações de similaridade que permutem os blocos de Jordan da matriz. Além disso, dentro de cada subespaço invariante (onde cada bloco age) é possível fazer certas permutações dos elementos da base, de modo a preservar a diagonal e permutar os γ_i's da primeira supradiagonal.

10.7.5 Mais Alguns Resultados Sobre Matrizes Nilpotentes

O Teorema da Decomposição de Jordan permite-nos demonstrar mais alguns fatos úteis sobre matrizes, particularmente sobre matrizes nilpotentes.

Recordemos que o índice de uma matriz nilpotente N é o menor $q \in \mathbb{N}$ para o qual vale tem-se $N^q = 0$. Um resultado útil sobre matrizes nilpotentes é o seguinte lema:

Lema 10.9 *Sejam N_1 e $N_2 \in \mathrm{Mat}(\mathbb{C},\ n)$ duas matrizes nilpotentes com índices q_1 e q_2, respectivamente. Se N_1 e N_2 comutarem, ou seja, se $N_1N_2 = N_2N_1$, então $\alpha_1 N_1 + \alpha_2 N_2$ é também nilpotente para todos $\alpha_1,\ \alpha_2 \in \mathbb{C}$. O índice de $\alpha_1 N_1 + \alpha_2 N_2$ é menor ou igual a $q_1 + q_2$.* □

Prova. Como N_1 e N_2 comutam, vale o binômio de Newton[29] e, para todo $m \in \mathbb{N}$ tem-se

$$\big(\alpha_1 N_1 + \alpha_2 N_2\big)^m \;=\; \sum_{p=0}^{m} \binom{m}{p} \alpha_1^{m-p}\alpha_2^p \, N_1^{m-p} N_2^p \, .$$

A condição de N_1 ter índice q_1 implica que é suficiente considerar os valores de p com $m - p < q_1$, ou seja, $p > m - q_1$. A condição de N_2 ter índice q_2 implica que é suficiente considerar os valores de p com $p < q_2$. Assim, só podem ser eventualmente não nulos os termos da soma com $m - q_1 < p < q_2$. Se tivermos $m - q_1 \geq q_2$ (ou seja, $m \geq q_1 + q_2$), essa condição é impossível e todos os termos da soma do lado direito são nulos, implicando que $\alpha_1 N_1 + \alpha_2 N_2$ é nilpotente de índice menor ou igual a m. Assim, o índice de $\alpha_1 N_1 + \alpha_2 N_2$ é menor ou igual a $q_1 + q_2$. ■

Um corolário disso é a Proposição 10.33, página 539, a qual indica-nos uma condição suficiente e necessária para que uma matriz seja nilpotente. Antes precisamos apresentar e demonstrar o seguinte lema, o qual tem interesse por si só:

Lema 10.10 *Seja $A \in \mathrm{Mat}(\mathbb{C},\ n)$, sejam $\alpha_1,\ \ldots,\ \alpha_r$ seus autovalores distintos (naturalmente, com $1 \leq r \leq n$) e sejam $m_1,\ \ldots,\ m_r$ suas multiplicidades algébricas respectivas (naturalmente, $m_1 + \cdots + m_r = n$). Então,*

$$\mathrm{Tr}\big(A^k\big) \;=\; \sum_{l=1}^{r} m_l \, \alpha_l^k \tag{10.123}$$

para para todo $k \in \mathbb{N}_0$. □

Prova. Seja $A \in \mathrm{Mat}(\mathbb{C},\ n)$. Para o caso $k = 0$, lembremo-nos da convenção que $A^0 = \mathbb{1}$. Assim, $\mathrm{Tr}(A^0) = \mathrm{Tr}(\mathbb{1}) = n$. Mas no caso $k = 0$ o lado direito de (10.123) fica $\sum_{l=1}^r m_l = n$. Isso estabeleceu (10.123) para $k = 0$. Tomemos doravante $k > 0$.

Seja $P \in \mathrm{Mat}(\mathbb{C},\ n)$ uma matriz inversível que leva A à sua forma de Jordan, ou seja, tal que $PAP^{-1} = D + N$ com D diagonal, N nilpotente e com $DN = ND$. Seja q índice de N. É claro que para cada $k \in \mathbb{N}$ tem-se

$$PA^kP^{-1} \;=\; \big(PAP^{-1}\big)^k \;=\; \big(D+N\big)^k \;=\; \sum_{p=0}^{k}\binom{k}{p} D^{k-p}N^p \;=\; D^k + \sum_{p=1}^{k}\binom{k}{p} D^{k-p}N^p \, . \tag{10.124}$$

Afirmamos que cada termo da última somatória (ou seja, aqueles termos com $1 \leq p \leq k$) é uma matriz nilpotente. De fato, para cada $l \in \mathbb{N}$ tem-se

$$\Big(D^{k-p}N^p\Big)^l \;=\; D^{(k-p)l}N^{pl}$$

e se escolhermos l de sorte que $pl \geq q$ (e isso é sempre possível para cada $p \geq 1$, o fator N^{pl} será nulo, provando que $D^{k-p}N^p$ é nilpotente.

Assim, (10.124) e o Lema 10.9, página 538 informam que $PA^kP^{-1} = D^k + M$ com M nilpotente. Logo, para todo $k \in \mathbb{N}$ tem-se

$$\mathrm{Tr}\big(A^k\big) \;=\; \mathrm{Tr}\big(PA^kP^{-1}\big) \;=\; \mathrm{Tr}\big(D^k\big) + \mathrm{Tr}(M) \;=\; \mathrm{Tr}\big(D^k\big) \, .$$

[29] Sir Isaac Newton (1643–1727).

Na última igualdade usamos o fato de que o traço de uma matriz nilpotente é nulo, pois todos os seus autovalores são nulos (vide Corolário 10.5, página 531 e Proposição 10.31, página 526).

Sejam $\alpha_1, \ldots, \alpha_r$ os autovalores distintos de A (naturalmente, com $1 \leq r \leq n$) e sejam $m_1, \ldots, m_r$ suas multiplicidades algébricas respectivas (naturalmente, $m_1 + \cdots + m_r = n$). Já sabemos que $D = \operatorname{diag}(\alpha_1, \ldots, \alpha_r)$, sendo que cada α_j aparece m_j vezes na diagonal de D. Logo, $D^k = \operatorname{diag}\left(\alpha_1^k, \ldots, \alpha_r^k\right)$. Consequentemente,

$$\operatorname{Tr}\left(A^k\right) = \operatorname{Tr}\left(D^k\right) = \sum_{l=1}^{r} m_l \alpha_l^k ,$$

completando a prova. ■

Vamos agora ao resultado mais desejado.

Proposição 10.33 *Uma condição necessária e suficiente para que uma matriz $A \in \operatorname{Mat}(\mathbb{C}, n)$ seja nilpotente é que valha $\operatorname{Tr}\left(A^k\right) = 0$ para todo $k = 1, \ldots, n$.* □

Prova. Se A é nilpotente, então todos os seus autovalores, são nulos, assim como todos os autovalores de todas as suas potências. Logo, $\operatorname{Tr}\left(A^k\right) = 0$ para todo $k \in \mathbb{N}$.

Vamos agora supor que $\operatorname{Tr}\left(A^k\right) = 0$ para todo $k = 1, \ldots, n$. Sejam $\alpha_1, \ldots, \alpha_r$ os autovalores distintos de A (naturalmente, com $1 \leq r \leq n$) e sejam $m_1, \ldots, m_r$ suas multiplicidades algébricas respectivas (naturalmente, $m_1 + \cdots + m_r = n$).

Vamos agora, por contradição, supor que A não seja nilpotente. Pelo Corolário 10.5, página 531, isso equivale a dizer que ao menos um dos autovalores de A é não nulo. Digamos que este seja o autovalor α_r. Temos, assim que $\alpha_r \neq 0$ e que $m_r \geq 1$.

Note-se que se $r = 1$, então todos os autovalores de A seriam iguais (a α, digamos) e teríamos $m_r = n$. Porém nesse caso teríamos $\operatorname{Tr}(A) = n\alpha$, o que é incompatível com a hipótese que $\operatorname{Tr}(A) = 0$, pois isso implicaria que $\alpha = 0$, ou seja, que todos os autovalores de A são nulos, o que implicaria, pelo Pelo Corolário 10.5, página 531, que A é nilpotente. Podemos, portanto, supor $r > 1$.

Seja $p(x) = \sum_{k=1}^{n} \beta_k x^k$ um polinômio de grau menor ou igual a n e cujo termo constante é nulo. Teremos, pela hipótese que $\operatorname{Tr}\left(A^k\right) = 0$ para todo $k = 1, \ldots, n$, que

$$0 = \sum_{k=1}^{n} \beta_k \operatorname{Tr}\left(A^k\right) \stackrel{(10.123)}{=} \sum_{k=1}^{n} \beta_k \left(\sum_{l=1}^{r} m_l \alpha_l^k\right) = \sum_{l=1}^{r} m_l \left(\sum_{k=1}^{n} \beta_k \alpha_l^k\right) = \sum_{l=1}^{r} m_l p(\alpha_l) . \tag{10.125}$$

Vamos agora escolher $p(x) = x(x - \alpha_1) \cdots (x - \alpha_{r-1})$. Teremos, evidentemente, que: 1º o polinômio p é um polinômio de grau $r \leq n$ cujo termo constante é nulo. 2º $p(\alpha_l) = 0$ para cada $l = 1, \ldots r-1$. 3º $p(\alpha_r) = \alpha_r(\alpha_r - \alpha_1) \cdots (\alpha_r - \alpha_{r-1}) \neq 0$, pois nenhum dos fatores do lado direito é nulo (já que $\alpha_r \neq 0$ e já que os α_j's são distintos). Para esse polinômio a relação (10.125) fica $0 = m_r p(\alpha_r)$. Como $p(\alpha_r) \neq 0$, concluímos que $m_r = 0$, uma contradição com o fato que $m_r \geq 1$ que por sua vez decorria da hipótese de A não ser nilpotente.

Logo, a hipótese que $\operatorname{Tr}\left(A^k\right) = 0$ para todo $k = 1, \ldots, n$, implica que A é nilpotente, completando a prova. ■

Uma consequência evidente da Proposição 10.33, acima, é que se para $A \in \operatorname{Mat}(\mathbb{C}, n)$ valer que $\operatorname{Tr}\left(A^k\right) = 0$ para cada $k = 1, \ldots, n$, então $\operatorname{Tr}\left(A^k\right) = 0$ para todo $k \geq 1$. O próximo exercício apresenta mais um corolário da Proposição 10.33.

E. 10.39 *Exercício*. Demonstre o seguinte corolário da Proposição 10.33, página 539:

Corolário 10.6 *Uma condição necessária e suficiente para que uma matriz $A \in \operatorname{Mat}(\mathbb{C}, n)$ seja nilpotente é que valha $\operatorname{Tr}\left(e^{zA}\right) = n$ para todo z em algum domínio aberto de $\mathbb{C}$.* □

Sugestão: prove que $\operatorname{Tr}\left(e^{zA}\right) = n + \sum_{k=1}^{\infty} \frac{z^k}{k!} \operatorname{Tr}\left(A^k\right)$ é analítica em z e use esse fato. Para a demonstrar a analiticidade, prove (usando (10.123)) que $\left|\operatorname{Tr}\left(A^k\right)\right| \leq n\Big(\max\{|\alpha_1|, \ldots, |\alpha_r|\}\Big)^k$ e use esse fato. ✠

10.8 Algumas Representações Especiais de Matrizes

Nas seções anteriores apresentamos algumas formas especiais de representar matrizes com determinadas características, como aquelas expressas no Teorema Espectral e no Teorema de Jordan. Nesta seção apresentaremos outras representações, relevantes em certos contextos, como a decomposição polar.

10.8.1 A Decomposição Polar de Matrizes

É bem conhecido o fato de que todo número complexo z pode ser escrito na forma polar $z = |z|e^{i\theta}$, onde $|z| \geq 0$ e $\theta \in [-\pi, \pi)$. Tem-se que $|z| = \sqrt{\overline{z}z}$ e $e^{i\theta} = z|z|^{-1}$. Há uma afirmação análoga válida para matrizes $A \in \mathrm{Mat}(\mathbb{C}, n)$, a qual é muito útil, e da qual trataremos nesta seção. Antes de enunciarmos esse resultado de forma mais precisa (o Teorema da Decomposição Polar, Teorema 10.24, abaixo), façamos algumas observações preliminares.

Seja $A \in \mathrm{Mat}(\mathbb{C}, n)$ e seja a matriz A^*A. Notemos primeiramente que $(A^*A)^* = A^*A^{**} = A^*A$, ou seja, A^*A e autoadjunta. Pelo Teorema 10.15, página 510, é possível encontrar um conjunto ortonormal $\{v_k, \ k = 1, \ldots, n\}$ de autovetores de A^*A, com autovalores d_k, $k = 1, \ldots, n$, respectivamente, sendo que a matriz

$$P := \big[\!\big[v_1, \ \ldots, \ v_n \big]\!\big] \tag{10.126}$$

(para a notação, vide (10.9)) é unitária e diagonaliza A^*A, ou seja, $P^*(A^*A)P = D$, sendo D a matriz diagonal $D := \mathrm{diag}(d_1, \ \ldots, \ d_n)$, cujos elementos da diagonal são os autovalores de A^*A. Os autovalores d_k são todos maiores ou iguais a zero. De fato, se $v_k \neq 0$ é um autovetor de A^*A com autovalor d_k, teremos $d_k\|v_k\|^2 = d_k\langle v_k, \ v_k\rangle_{\mathbb{C}} = \langle v_k, \ Bv_k\rangle_{\mathbb{C}} = \langle v_k, \ A^*Av_k\rangle_{\mathbb{C}} = \langle Av_k, \ Av_k\rangle_{\mathbb{C}} = \|Av_k\|^2$. Logo, $d_k = \|Av_k\|^2/\|v_k\|^2 \geq 0$.

Com esses fatos à mão, vamos definir uma matriz diagonal, que denotaremos sugestivamente por $D^{1/2}$, por $D^{1/2} := \mathrm{diag}(\sqrt{d_1}, \ \ldots, \ \sqrt{d_n})$. Tem-se que $\big(D^{1/2}\big)^2 = D$, uma propriedade óbvia[30]. Note-se também que $\big(D^{1/2}\big)^* = D^{1/2}$, pois cada $\sqrt{d_k}$ é real. Os números não negativos $\sqrt{d_1}, \ \ldots, \ \sqrt{d_n}$ são frequentemente denominados *valores singulares* de A.

Definamos agora a matriz $\sqrt{A^*A}$, por

$$\sqrt{A^*A} := PD^{1/2}P^* . \tag{10.127}$$

Essa matriz $\sqrt{A^*A}$ é autoadjunta, pois $\left(\sqrt{A^*A}\right)^* = \big(PD^{1/2}P^*\big)^* = PD^{1/2}P^* = \sqrt{A^*A}$. Observemos que $\left(\sqrt{A^*A}\right)^2 = P(D^{1/2})^2P^* = PDP^* = A^*A$. Disso segue que

$$\left(\det\left(\sqrt{A^*A}\right)\right)^2 = \det\left(\left(\sqrt{A^*A}\right)^2\right) = \det(A^*A) = \det(A^*)\det(A) = \overline{\det(A)}\det(A) = |\det(A)|^2 .$$

Provamos assim que $\det\left(\sqrt{A^*A}\right) = |\det(A)|$ e, portanto, $\sqrt{A^*A}$ é inversível se e somente se A o for.

Alguns autores denotam a matriz $\sqrt{A^*A}$ por $|A|$, por analogia com o módulo de um número complexo. Podemos agora formular e demonstrar o resultado que procuramos:

Teorema 10.24 (Teorema da Decomposição Polar) *Seja $A \in \mathrm{Mat}(\mathbb{C}, n)$. Então, existe uma matriz unitária $U \in \mathrm{Mat}(\mathbb{C}, n)$ tal que*

$$A = U\sqrt{A^*A} . \tag{10.128}$$

Se A é inversível, então U é univocamente determinada. A representação (10.128) é denominada representação polar *de A.* □

Prova. Sejam, como acima, d_k, $k = 1, \ldots, n$ os autovalores de A^*A com autovetores respectivos v_k, $k = 1, \ldots, n$. Sabemos pelo Teorema 10.15, página 510 que podemos escolher os v_k's de forma que $\langle v_k, \ v_l\rangle_{\mathbb{C}} = \delta_{k\,l}$.

Como vimos acima, os autovalores d_k satisfazem $d_k \geq 0$. Sem perda de generalidade, vamos supô-los ordenados de forma que $d_k > 0$ para todo $k = 1, \ldots, r$ e $d_k = 0$ para todo $k = r+1, \ldots, n$. Com essa escolha, tem-se que

$$Av_k = 0 \text{ para todo } k = r+1, \ \ldots, \ n , \tag{10.129}$$

[30]Essa não é a única matriz com essa propriedades, pois qualquer matriz do tipo $\mathrm{diag}(\pm\sqrt{d_1}, \ \ldots, \ \pm\sqrt{d_n})$, com os sinais $\pm$ escolhidos independentemente uns dos outros, também tem como quadrado a matriz D.

pois de $A^*Av_k = 0$, segue que $0 = \langle v_k,\ A^*Av_k\rangle_{\mathbb{C}} = \langle Av_k,\ Av_k\rangle_{\mathbb{C}} = \|Av_k\|^2$.

Para $k = 1, \ldots, r$, sejam w_k os vetores definidos da seguinte forma:

$$w_k \ := \ \frac{1}{\sqrt{d_k}} Av_k \ , \quad k \ = \ 1, \ldots, r \ . \tag{10.130}$$

É fácil ver que

$$\langle w_k,\ w_l\rangle_{\mathbb{C}} \ = \ \frac{1}{\sqrt{d_k d_l}}\langle Av_k,\ Av_l\rangle_{\mathbb{C}} \ = \ \frac{1}{\sqrt{d_k d_l}}\langle A^*Av_k,\ v_l\rangle_{\mathbb{C}} \ = \ \frac{d_k}{\sqrt{d_k d_l}}\langle v_k,\ v_l\rangle_{\mathbb{C}} \ = \ \frac{d_k}{\sqrt{d_k d_l}}\delta_{k\,l} \ = \ \delta_{k\,l} \ ,$$

para todos $k,\ l = 1, \ldots, r$. Assim, o conjunto de vetores $\{w_k,\ k = 1, \ldots, r\}$ forma um conjunto ortonormal. A eles podemos acrescentar um novo conjunto $\{w_k,\ k = r+1, \ldots, n\}$, escolhido arbitrariamente, de vetores ortonormais pertencentes ao complemento ortogonal do subespaço gerado por $\{w_k,\ k = 1, \ldots, r\}$ e construir assim, um conjunto ortonormal $\{w_k,\ k = 1, \ldots, n\}$.

Sejam agora a matriz P, definida em (10.126) e as seguintes matrizes de $\mathrm{Mat}\,(\mathbb{C},\ n)$:

$$Q \ := \ \Big[\!\!\Big[w_1, \ \ldots, \ w_n \Big]\!\!\Big] \ , \qquad U \ := \ QP^*$$

(para a notação, vide (10.9)). Como $\{v_k,\ k = 1, \ldots, n\}$ e $\{w_k,\ k = 1, \ldots, n\}$ são dois conjuntos ortonormais, segue que P e Q são matrizes unitárias (por quê?) e, portanto, U também é unitária.

É fácil ver que $AP = QD^{1/2}$, onde $D^{1/2} := \mathrm{diag}\left(\sqrt{d_1}, \ \ldots, \ \sqrt{d_n}\right)$, De fato,

$$\begin{aligned}
AP \ &\stackrel{(10.126)}{=} \ A\Big[\!\!\Big[v_1, \ \ldots, \ v_n \Big]\!\!\Big] \ \stackrel{(10.12)}{=} \ \Big[\!\!\Big[Av_1, \ \ldots, \ Av_n \Big]\!\!\Big] \\
&\stackrel{(10.129)}{=} \ \Big[\!\!\Big[Av_1, \ \ldots, \ Av_r \ 0, \ \ldots, \ 0 \Big]\!\!\Big] \\
&\stackrel{(10.130)}{=} \ \Big[\!\!\Big[\sqrt{d_1}w_1, \ \ldots, \ \sqrt{d_r}w_r \ 0, \ \ldots, \ 0 \Big]\!\!\Big] \\
&\stackrel{(10.15)}{=} \ \Big[\!\!\Big[w_1, \ \ldots, \ w_n \Big]\!\!\Big] D^{1/2} \ = \ QD^{1/2} \ .
\end{aligned}$$

Agora, de $AP = QD^{1/2}$, segue que $A = QD^{1/2}P^* = UPD^{1/2}P^* \stackrel{(10.127)}{=} U\sqrt{A^*A}$, que é o que queríamos provar.

Para mostrar que U é univocamente determinado se A for inversível, suponhamos que exista U' tal que $A = U\sqrt{A^*A} = U'\sqrt{A^*A}$. Como comentamos acima, $\sqrt{A^*A}$ é inversível se e somente se A o for. Logo, se A é inversível, a igualdade $U\sqrt{A^*A} = U'\sqrt{A^*A}$ implica $U = U'$, estabelecendo a unicidade. Caso A não seja inversível a arbitrariedade de U reside na escolha dos vetores ortonormais $\{w_k,\ k = r+1, \ldots, n\}$. ■

O seguinte corolário é elementar:

Teorema 10.25 *Seja $A \in \mathrm{Mat}\,(\mathbb{C},\ n)$. Então, existe uma matriz unitária $V \in \mathrm{Mat}\,(\mathbb{C},\ n)$ tal que*

$$A \ = \ \sqrt{AA^*}\,V \ . \tag{10.131}$$

Se A é inversível, então V é univocamente determinada. □

Prova. Para a matriz A^*, (10.128) diz-nos que $A^* = U_0\sqrt{(A^*)^*A^*} = U_0\sqrt{AA^*}$ para alguma matriz unitária U_0. Como $\sqrt{AA^*}$ é autoadjunta, segue que $A = \sqrt{AA^*}\,U_0^*$. Identificando $V = U_0^*$, obtemos o que desejamos. ■

O Teorema da Decomposição Polar pode ser generalizado para abranger operadores limitados agindo em espaços de Hilbert (vide Teorema 39.31, página 2136) e mesmo para abranger operadores não limitados agindo em espaços de Hilbert (vide [315]).

10.8.2 A Decomposição em Valores Singulares

O Teorema da Decomposição Polar, Teorema 10.24, página 540, tem um corolário de particular interesse, o Teorema da Decomposição em Valores Singulares, qua afirma que toda matriz pode ser escrita como produto de três matrizes, uma delas envolvendo seus valores singulares.

Na Seção 10.9, página 562, estudaremos uma aplicação do Teorema da Decomposição em Valores Singulares, a saber, ao estudo da chamada Pseudoinversa de Moore-Penrose e suas aplicações em problemas de optimização linear. Outra aplicação do versa do Teorema da Decomposição em Valores Singulares, relevante à Teoria da Informação Quântica é o Teorema de Decomposição de Schmidt, objeto da Seção 10.8.2.1, página 545.

A decomposição em valores singulares admite uma generalização para operadores compactos agindo em espaços de Hilbert de dimensão infinita. Vide Teorema 39.39, página 2160.

Teorema 10.26 (Teorema da Decomposição em Valores Singulares) I. *Seja* $A \in \text{Mat}(\mathbb{C},\ n)$ *com* $n \in \mathbb{N}$. *Então, existem matrizes unitárias* V *e* $W \in \text{Mat}(\mathbb{C},\ n)$ *tais que*

$$A = VSW^*, \tag{10.132}$$

onde

$$S = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} \in \text{Mat}(\mathbb{C},\ n) \tag{10.133}$$

é uma matriz diagonal cujos elementos diagonais são os valores singulares $d_k \geq 0$, $k = 1, \ldots, n$, *de* A, *ou seja, os autovalores de* $\sqrt{A^*A}$.

Essa afirmação pode ser generalizada para matrizes retangulares. Seja $A \in \text{Mat}(\mathbb{C},\ m,\ n)$ *com* $m,\ n \in \mathbb{N}$ *e* $m \neq n$ *(o caso* $m = n$ *corresponde ao caso* **I**, *acima).*

II. *Caso* $m > n$. *Existem matrizes unitárias* $V \in \text{Mat}(\mathbb{C},\ m)$ *e* $W \in \text{Mat}(\mathbb{C},\ n)$ *tais que*

$$A = VSW^*, \tag{10.134}$$

onde $S \in \text{Mat}(\mathbb{C},\ m,\ n)$ *é a matriz*

$$S = \begin{pmatrix} D \\ \\ \mathbb{0}_{m-n,\ n} \end{pmatrix} = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \\ & \mathbb{0}_{m-n,\ n} & \end{pmatrix} \in \text{Mat}(\mathbb{C},\ m,\ n), \tag{10.135}$$

sendo $D \in \text{Mat}(\mathbb{C},\ n)$ *uma matriz diagonal contendo na diagonal os* n *valores singulares* $d_k \geq 0$, $k = 1, \ldots, n$, *de* A, *ou seja, os autovalores de* $\sqrt{A^*A} \in \text{Mat}(\mathbb{C},\ n)$, *incluindo multiplicidades.*

III. *Caso* $m < n$. *Existem matrizes unitárias* $V \in \text{Mat}(\mathbb{C},\ m)$ *e* $W \in \text{Mat}(\mathbb{C},\ n)$ *tais que*

$$A = VSW^*, \tag{10.136}$$

onde neste caso $S \in \text{Mat}(\mathbb{C},\ m,\ n)$ *é a matriz*

$$S = \begin{pmatrix} L & \mathbb{0}_{m,\ n-m} \end{pmatrix} = \begin{pmatrix} l_1 & & & \\ & \ddots & & \mathbb{0}_{m,\ n-m} \\ & & l_m & \end{pmatrix} \in \text{Mat}(\mathbb{C},\ m,\ n), \tag{10.137}$$

sendo $L \in \text{Mat}(\mathbb{C},\ m)$ *uma matriz diagonal contendo na diagonal os* m *valores singulares* $l_k \geq 0$, $k = 1, \ldots, m$, *de* A^*, *ou seja, os autovalores de* $\sqrt{AA^*} \in \text{Mat}(\mathbb{C},\ m)$, *incluindo multiplicidades.* □

Comentário. É apropriado aqui recordarmos que os conjuntos de valores singulares $\{d_1, \ldots, d_n\}$ e $\{l_1, \ldots, l_m\}$ de A e A^*, respectivamente, são iguais a menos de multiplicidades e da eventual presença de 0's em algum deles. Isso é consequência da Proposição 10.7, página 475 e, especialmente, de sua generalização apresentada no Exercício E. 10.6, página 475, onde se afirma que $\sigma(A^*A) \setminus \{0\} = \sigma(AA^*) \setminus \{0\}$. ♣

Prova do Teorema 10.26. **Prova de I.** A afirmação de (10.132) segue imediatamente de (10.128) e de (10.127) tomando $V = UP$, $W = P$ e $S = D^{1/2}$.

Prova de II. Temos $m > n$ e seja $A_e \in \mathrm{Mat}(\mathbb{C},\ m)$ a matriz quadrada $m \times m$ dada por

$$A_e \ := \ \begin{pmatrix} A & \mathbb{0}_{m,\ m-n} \end{pmatrix} . \tag{10.138}$$

Temos que

$$A_e^* A_e \ = \ \begin{pmatrix} A^*A & \mathbb{0}_{n,\ m-n} \\ \mathbb{0}_{m-n,\ n} & \mathbb{0}_{m-n,\ m-n} \end{pmatrix} ,$$

sendo $A_e^* A_e \in \mathrm{Mat}(\mathbb{C},\ m)$ e $A^*A \in \mathrm{Mat}(\mathbb{C},\ n)$. Naturalmente, segue disso que

$$\sqrt{A_e^* A_e} \ = \ \begin{pmatrix} \sqrt{A^*A} & \mathbb{0}_{n,\ m-n} \\ \mathbb{0}_{m-n,\ n} & \mathbb{0}_{m-n,\ m-n} \end{pmatrix} .$$

Para a matriz quadrada A_e aplica-se o Teorema da Decomposição Polar, Teorema 10.24, página 540 , e podemos escrever $A_e = \widetilde{U}\sqrt{A_e^* A_e}$ para $\widetilde{U} \in \mathrm{Mat}(\mathbb{C},\ m)$, unitária.

A matriz $\sqrt{A^*A} \in \mathrm{Mat}(\mathbb{C},\ n)$ pode ser diagonalizada por uma matriz unitária $P \in \mathrm{Mat}(\mathbb{C},\ n)$ de sorte que $\sqrt{A^*A} = PDP^*$, onde $D \in \mathrm{Mat}(\mathbb{C},\ n)$ é diagonal e contém na diagonal os autovalores de $\sqrt{A^*A}$. Com isso, podemos escrever

$$\sqrt{A_e^* A_e} \ = \ \widetilde{P}\widetilde{D}\widetilde{P}^* ,$$

onde,

$$\widetilde{P} \ := \ \begin{pmatrix} P & \mathbb{0}_{n,\ m-m} \\ \mathbb{0}_{m-n,\ n} & \mathbb{1}_{m-n} \end{pmatrix} \qquad \text{e} \qquad \widetilde{D} \ := \ \begin{pmatrix} D & \mathbb{0}_{n,\ m-m} \\ \mathbb{0}_{m-n,\ n} & \mathbb{1}_{m-n} \end{pmatrix} ,$$

e, portanto, $A_e = \widetilde{U}\sqrt{A_e^* A_e} = \widetilde{U}\widetilde{P}\widetilde{D}\widetilde{P}^*$. Agora,

$$\begin{pmatrix} A & \mathbb{0}_{m,\ m-n} \end{pmatrix} \ = \ A \begin{pmatrix} \mathbb{1}_n & \mathbb{0}_{n,\ m-m} \end{pmatrix} ,$$

e, portanto, temos

$$A \begin{pmatrix} \mathbb{1}_n & \mathbb{0}_{m,\ m-n} \end{pmatrix} \ = \ \widetilde{U}\widetilde{P}\widetilde{D}\widetilde{P}^* .$$

Multiplicando-se à direita por $\begin{pmatrix} \mathbb{1}_n \\ \mathbb{0}_{m-n,\ m} \end{pmatrix}$ e usando o fato que

$$\begin{pmatrix} \mathbb{1}_n & \mathbb{0}_{m,\ m-n} \end{pmatrix} \begin{pmatrix} \mathbb{1}_n \\ \mathbb{0}_{m-n,\ m} \end{pmatrix} \ = \ \mathbb{1}_n$$

(verifique!), obtemos

$$A \ = \ \widetilde{U}\widetilde{P}\widetilde{D}\widetilde{P}^* \begin{pmatrix} \mathbb{1}_n \\ \mathbb{0}_{m-n,\ m} \end{pmatrix} \ = \ \widetilde{U}\widetilde{P}\widetilde{D} \begin{pmatrix} P \\ \mathbb{0}_{m-n,\ n} \end{pmatrix} .$$

Agora,

$$\widetilde{D} \begin{pmatrix} P \\ \mathbb{0}_{m-n,\ n} \end{pmatrix} \ = \ \begin{pmatrix} D & \mathbb{0}_{n,\ m-m} \\ \mathbb{0}_{m-n,\ n} & \mathbb{1}_{m-n} \end{pmatrix} \begin{pmatrix} P^* \\ \mathbb{0}_{m-n,\ n} \end{pmatrix} \ = \ \begin{pmatrix} DP^* \\ \mathbb{0}_{m-n,\ n} \end{pmatrix} \ = \ \begin{pmatrix} D \\ \mathbb{0}_{m-n,\ n} \end{pmatrix} P^*$$

e temos finalmente

$$A = \widetilde{U}\widetilde{P}\begin{pmatrix} D \\ \\ \mathbb{0}_{m-n,\, n}\end{pmatrix} P^* .$$

Adotando-se $V = \widetilde{U}\widetilde{P}$, que é uma matriz unitária de $\mathrm{Mat}(\mathbb{C},\ m)$ (pois $\widetilde{U}$ e $\widetilde{P}$ o são) e $W = P$, que é uma matriz unitária de $\mathrm{Mat}(\mathbb{C},\ n)$ e $S = \begin{pmatrix} D \\ \\ \mathbb{0}_{m-n,\, n}\end{pmatrix}$ a demonstração de (10.134) está completa.

Prova de III. Esse caso segue diretamente do caso **II**. Seja $A \in \mathrm{Mat}(\mathbb{C},\ m,\ n)$ agora com $m < n$. A matriz $A^* \in \mathrm{Mat}(\mathbb{C},\ n,\ m)$ tem, portanto, as mesmas características das matrizes tratadas na parte **II** (o número de linhas é maior que o de colunas) e, portanto, vale a afirmação (da parte **II**, trocando-se $m \leftrightarrow n$) que existem operadores unitários $W \in \mathrm{Mat}(\mathbb{C},\ n)$ e $V \in \mathrm{Mat}(\mathbb{C},\ m)$ tais que

$$A^* = WRV^* , \tag{10.139}$$

(por conveniência trocamos as letras V e W e substituímos a letra S or R em (10.134)), onde $R \in \mathrm{Mat}(\mathbb{C},\ n,\ m)$ é a matriz

$$R = \begin{pmatrix} L \\ \\ \mathbb{0}_{n-m,\, m}\end{pmatrix} = \begin{pmatrix} l_1 & & \\ & \ddots & \\ & & l_m \\ & \mathbb{0}_{n-m,\, m} & \end{pmatrix} \in \mathrm{Mat}(\mathbb{C},\ n,\ m) ,$$

sendo $L \in \mathrm{Mat}(\mathbb{C},\ m)$ uma matriz diagonal contendo na diagonal os m valores singulares l_k, $k = 1, \ldots, m$, de A^*, ou seja, os autovalores de $\sqrt{AA^*} \in \mathrm{Mat}(\mathbb{C},\ m)$, incluindo multiplicidades.

Tomando-se o adjunto de (10.139), obtemos

$$A = VSW^* ,$$

onde

$$S = R^* = \begin{pmatrix} L & \mathbb{0}_{m,\, n-m}\end{pmatrix} = \begin{pmatrix} l_1 & & & \\ & \ddots & & \mathbb{0}_{m,\, n-m} \\ & & l_m & \end{pmatrix} \in \mathrm{Mat}(\mathbb{C},\ m,\ n) .$$

Isso demonstrou (10.136), completando a prova do Teorema 10.26. ∎

• Decomposição em valores singulares. Uma segunda abordagem

Vamos agora apresentar uma segunda versão do Teorema da Decomposição em Valores Singulares com consequências diferentes da anterior, mas similares às mesmas, especialmente no caso de matrizes não quadradas. Essa versão do Teorema da Decomposição em Valores Singulares pode ser relevante em certos problemas.

No que segue, $m,\ n \in \mathbb{N}$. Aqui não se faz necessário distinguir os casos $m < n$ e $m > n$ ou $m = n$. Usaremos as definições (10.3), (10.7) e a relação (10.8) (vide página 464) que permitem mapear injetivamente matrizes retangulares em certas matrizes quadradas.

Teorema 10.27 (Teorema da Decomposição em Valores Singulares. II) *Seja $A \in \mathrm{Mat}(\mathbb{C},\ m,\ n)$. Então, existem matrizes unitárias V e $W \in \mathrm{Mat}(\mathbb{C},\ m+n)$ tais que*

$$A = I_{m,\, m+n} VSW^* J_{m+n,\, n} , \tag{10.140}$$

onde $S \in \mathrm{Mat}(\mathbb{C},\ m+n)$ é uma matriz diagonal cujos elementos diagonais são os valores singulares de

$$A' := J_{m+n,\, m} A I_{n,\, m+n} = \begin{pmatrix} A & \mathbb{0}_{m,\, m} \\ \\ \mathbb{0}_{n,\, n} & \mathbb{0}_{n,\, m}\end{pmatrix} \in \mathrm{Mat}(\mathbb{C},\ m+n) \tag{10.141}$$

(definida em (10.7)), ou seja, os autovalores de $\sqrt{(A')^*A'}$. *Mais especificamente,*

$$S \;=\; \begin{pmatrix} d_1 & & & \\ & \ddots & & \mathbb{0}_{n,\,m} \\ & & d_n & \\ & \mathbb{0}_{m,\,n} & & \mathbb{0}_{m,\,m} \end{pmatrix} , \tag{10.142}$$

onde $d_1, \ldots, d_n \in [0, \infty)$ *são os autovalores de* $\sqrt{A^*A} \in \mathrm{Mat}(\mathbb{C}, n)$. □

Comentário. O leitor deve observar que as matrizes V, W e S dos enunciado do Teorema 10.27 são distintas das das matrizes V, W e S do enunciado do Teorema 10.26, página 542, entre outras coisas, por terem dimensões diferentes. Observe-se também que, como $(A')^*A' = \begin{pmatrix} A^*A & \mathbb{0}_{n,\,m} \\ \mathbb{0}_{m,\,n} & \mathbb{0}_{m,\,m}\end{pmatrix} \in \mathrm{Mat}(\mathbb{C}, m+n)$, concluímos que, a menos de 0's e de multiplicidades, os valores singulares de A' coincidem com os de A e A^*. ♣

Prova do Teorema 10.27. A matriz $A' \in \mathrm{Mat}(\mathbb{C}, m+n)$ é uma matriz quadrada e, pela parte **I** do Teorema 10.26, possui uma decomposição em valores singulares $A' = VSW^*$ com V e $W \in \mathrm{Mat}(\mathbb{C}, m+n)$, unitárias, e $S \in \mathrm{Mat}(\mathbb{C}, m+n)$ sendo diagonal cujos elementos diagonais são os valores singulares de A'. Com isso, (10.140) segue de (10.8). ■

10.8.2.1 Uma Aplicação: a Decomposição de Schmidt

O Teorema de Decomposição em Valores Singulares, Teorema 10.26, páginas 542, tem um corolário relevante para a Teoria da Informação Quântica, o chamado Teorema de Decomposição de Schmidt[31].

Teorema 10.28 *Sejam* $\mathcal{H}_{\mathrm{I}}$ *e* $\mathcal{H}_{\mathrm{II}}$ *dois espaços de Hilbert complexos de dimensão finita, com dimensões* m *e* n*, respectivamente (concretamente podemos assumir* $\mathcal{H}_{\mathrm{I}} = \mathbb{C}^m$ *e* $\mathcal{H}_{\mathrm{II}} = \mathbb{C}^n$*). Considere-se o produto tensorial* $\mathcal{H}_{\mathrm{I}} \otimes \mathcal{H}_{\mathrm{II}}$*. Seja* $l = \min\{m, n\}$*. Então, para cada* $u \in \mathcal{H}_{\mathrm{I}} \otimes \mathcal{H}_{\mathrm{II}}$ *existem números* $d_a \geq 0$ *com* $a = 1, \ldots, l$ *e conjuntos de vetores* $\{\mathbf{v}_1, \ldots, \mathbf{v}_l\} \subset \mathcal{H}_{\mathrm{I}}$ *e* $\{\mathbf{w}_1, \ldots, \mathbf{w}_l\} \subset \mathcal{H}_{\mathrm{II}}$*, ortonormais em seus respectivos espaços, tais que*

$$u \;=\; \sum_{a=0}^{l} d_a \big(\mathbf{v}_a \otimes \mathbf{w}_a\big) . \tag{10.143}$$

O número de termos da soma é l*, a menor dimensão dentre os espaços* $\mathcal{H}_{\mathrm{I}}$ *e* $\mathcal{H}_{\mathrm{II}}$*. A expressão* (10.143) *é denominada* Decomposição de Schmidt *do vetor* $u \in \mathcal{H}_{\mathrm{I}} \otimes \mathcal{H}_{\mathrm{II}}$*.* □

Comentários. Há algumas observações a se fazer. Os vetores $\mathbf{v}_a$ e $\mathbf{w}_a$, assim como os números d_a, $a \in \{1, \ldots, l\}$, dependem de u e, portanto, os ingredientes do lado direito da Decomposição de Schmidt (10.143) mudam com u. Não está sendo dito em (10.143) que os vetores $\{\mathbf{v}_1 \otimes \mathbf{w}_1, \ldots, \mathbf{v}_l \otimes \mathbf{w}_l\}$ componham uma base em $\mathcal{H}_{\mathrm{I}} \otimes \mathcal{H}_{\mathrm{II}}$ (um erro infelizmente encontrado em alguns textos da literatura em Física), até por que não são em número suficiente para tal (a dimensão de $\mathcal{H}_{\mathrm{I}} \otimes \mathcal{H}_{\mathrm{II}}$ é mn, que é maior que $l = \min\{m, n\}$, exceto no caso trivial em que $m = n = 1$).

No caso em que $l = m < n$, o conjunto de vetores ortonormais $\{\mathbf{v}_1, \ldots, \mathbf{v}_l\}$ compõe uma base em $\mathcal{H}_{\mathrm{I}}$, mas o conjunto de vetores ortonormais $\{\mathbf{w}_1, \ldots, \mathbf{w}_l\}$ não compõe uma base em $\mathcal{H}_{\mathrm{II}}$, por não serem em número suficiente para tal. *Mutatis mutandis*, se $l = n < n$ as afirmações opostas são válidas. No caso $l = m = n$ ambos os conjuntos compõem bases em seus respectivos espaços. ♣

Prova do Teorema 10.28. Seja $\{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$ uma base ortonormal em $\mathcal{H}_{\mathrm{I}}$ e seja $\{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ uma base ortonormal em $\mathcal{H}_{\mathrm{II}}$. Então, $\Big\{\mathbf{e}_i \otimes \mathbf{f}_j, \;\; i \in \{1, \ldots, m\} \text{ e } j \in \{1, \ldots, n\}\Big\}$ compõe uma base ortonormal em $\mathcal{H}_{\mathrm{I}} \otimes \mathcal{H}_{\mathrm{II}}$.

Assim, se $u \in \mathcal{H}_{\mathrm{I}} \otimes \mathcal{H}_{\mathrm{II}}$, podemos escrever

$$u \;=\; \sum_{i=1}^{m} \sum_{j=1}^{n} u_{ij} \big(\mathbf{e}_i \otimes \mathbf{f}_j\big) ,$$

com $u_{ij} \in \mathbb{C}$ para todos i, j.

[31] Erhard Schmidt (1876–1959).

Seja agora $A \in \mathrm{Mat}(\mathbb{C}, m, n)$ a matriz cujos elementos de matriz são precisamente os coeficientes u_{ij}, ou seja, $A_{ij} := u_{ij}$ para cada $i \in \{1, \ldots, m\}$ e $j \in \{1, \ldots, n\}$.

Vamos considerar o caso $m \geq n$ (portanto, com $l := \min\{m, n\} = n$) fazendo uso das partes **I** e **II** do Teorema 10.26. O caso $m < n$ é similar, fazendo uso do caso **III** do mesmo Teorema e sua prova é deixada ao leitor. Segundo o Teorema 10.26, podemos escrever $A = VSW^*$, onde $V \in \mathrm{Mat}(\mathbb{C}, m)$ e $W \in \mathrm{Mat}(\mathbb{C}, n)$ são ambas unitárias (e, portanto, satisfazem $V^*V = \mathbb{1}_m$ e $W^*W = \mathbb{1}_n$). Assim, para os elementos de matriz de A, temos

$$u_{ij} = A_{ij} = \sum_{a=1}^{m}\sum_{b=1}^{n} V_{ia}S_{ab}(W^*)_{bj} = \sum_{a=1}^{n}\sum_{b=1}^{n} V_{ia}d_a\delta_{ab}(W^*)_{bj} = \sum_{a=1}^{n} d_a V_{ia}(W^*)_{aj} ,$$

onde, na segunda igualdade usamos o fato que $S_{ab} = 0$ caso $a > m$ e $S_{ab} = d_a\delta_{ab}$ de outra forma. Vide (10.133) ou (10.135), página 542. Assim,

$$u = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}\big(\mathbf{e}_i \otimes \mathbf{f}_j\big) = \sum_{a=1}^{m} d_a \left(\sum_{i=1}^{m} V_{ia}\mathbf{e}_i\right) \otimes \left(\sum_{j=1}^{n} (W^*)_{aj}\mathbf{f}_j\right) = \sum_{a=1}^{n} d_a\big(\mathbf{v}_a \otimes \mathbf{w}_a\big) ,$$

onde, para $a = 1, \ldots, n$,

$$\mathbf{v}_a := \sum_{i=1}^{m} V_{ia}\mathbf{e}_i \qquad \text{e} \qquad \mathbf{w}_a := \sum_{j=1}^{n} (W^*)_{aj}\mathbf{f}_j .$$

Isso provou (10.143). Agora, para $a, b \in \{1, \ldots, n\}$, tem-se

$$\big\langle \mathbf{v}_a, \mathbf{v}_b \big\rangle_{\mathcal{H}_{\mathrm{I}}} = \sum_{i=1}^{m}\sum_{i'=1}^{m} \overline{V_{ia}}V_{i'b} \underbrace{\langle \mathbf{e}_i, \mathbf{e}_{i'} \rangle_{\mathcal{H}_{\mathrm{I}}}}_{\delta_{i,\, i'}} = \sum_{i=1}^{m} \overline{V_{ia}}V_{ib} = \sum_{i=1}^{m} (V^*)_{ai}V_{ib} = \big(V^*V\big)_{ab} = \delta_{a,\, b}$$

e, analogamente,

$$\big\langle \mathbf{w}_a, \mathbf{w}_b \big\rangle_{\mathcal{H}_{\mathrm{II}}} = \sum_{j=1}^{m}\sum_{j'=1}^{m} \overline{W_{ja}}W_{j'b} \underbrace{\langle \mathbf{f}_j, \mathbf{f}_{j'} \rangle_{\mathcal{H}_{\mathrm{II}}}}_{\delta_{j,\, j'}} = \sum_{j=1}^{n} \overline{W_{ja}}W_{jb} = \sum_{j=1}^{n} (W^*)_{aj}W_{jb} = \big(W^*W\big)_{ab} = \delta_{a,\, b} ,$$

mostrando que $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset \mathcal{H}_{\mathrm{I}}$ e $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\} \subset \mathcal{H}_{\mathrm{II}}$ são conjuntos ortonormais em seus respectivos espaços. ■

A decomposição de Schmidt possui um colorário de interesse sobre matrizes definidas em um produto tensorial de espaços de dimensão finita.

Corolário 10.7 *Sejam os espaços de Hilbert $\mathbb{C}^m$ e $\mathbb{C}^n$ e seja A um operador linear definido em $\mathbb{C}^m \otimes \mathbb{C}^n \simeq \mathbb{C}^{mn}$. Então, A pode ser escrito na forma*

$$A = \sum_{k=1}^{l} d_k\big(A_k \otimes B_k\big) , \tag{10.144}$$

onde $l = \min\{m^2, n^2\}$, onde $d_k \geq 0$ e onde $\{A_1, \ldots, A_l\} \subset \mathrm{Mat}(\mathbb{C}, m)$ e $\{B_1, \ldots, B_l\} \subset \mathrm{Mat}(\mathbb{C}, n)$ são conjuntos ortonormais de matrizes em relação aos produtos escalares $\langle D, F\rangle_1 := \mathrm{Tr}_{\mathbb{C}^m}\big(D^*F\big)$ e, respectivamente, $\langle G, H\rangle_2 := \mathrm{Tr}_{\mathbb{C}^n}\big(G^*H\big)$, para $D, F \in \mathrm{Mat}(\mathbb{C}, m)$ e $G, H \in \mathrm{Mat}(\mathbb{C}, n)$.

A decomposição (10.144) *é denominada* decomposição de Schmidt para matrizes. □

Prova. A demonstração é imediata pelo Teorema 10.28, página 545, cabendo apenas notar que as dimensões de $\mathrm{Mat}(\mathbb{C}, m)$ e $\mathrm{Mat}(\mathbb{C}, n)$ são m^2 e n^2, respectivamente, daí tomarmos $l = \min\{m^2, n^2\}$. ■

10.8.2.2 A Noção de Traço Parcial de Matrizes

A decomposição de Schmidt para matrizes (10.144), página 546, é útil na tarefa de definir-se a noção de *traço parcial* de uma matriz agindo em um produto tensorial de espaços de dimensão finita. A noção de traço parcial é empregada na Teoria da Informação Quântica. Uma extensão da noção de traço parcial para o caso de operadores agindo em espaços de Hilbert de dimensão infinita, tecnicamente mais elaborado, é apresentada na Seção 39.11, página 2199.

Sejam $\mathcal{H}'$ e $\mathcal{H}''$ dois espaços de Hilbert de dimensões m e n, respectivamente, (concretamente podemos assumir $\mathcal{H}' = \mathbb{C}^m$ e $\mathcal{H}'' = \mathbb{C}^n$) e sejam $\{\mathbf{e}_1, \ldots, \mathbf{e}_m\} \subset \mathcal{H}'$ e $\{\mathbf{f}_1, \ldots, \mathbf{f}_n\} \subset \mathcal{H}''$ bases ortonormais nesses espaços. Defina-se $\mathcal{H} := \mathcal{H}' \otimes \mathcal{H}'' = \mathbb{C}^m \otimes \mathbb{C}^n \simeq \mathbb{C}^{nm}$.

Para um operador A agindo em $\mathcal{H}$ (ou seja $A \in \mathrm{Mat}(\mathbb{C}, mn)$) a expressão

$$\mathcal{H}' \times \mathcal{H}' \ni (\psi, \phi) \longmapsto \omega(\psi, \phi) := \sum_{j=1}^{n} \big\langle (\psi \otimes \mathbf{f}_j), A(\phi \otimes \mathbf{f}_j) \big\rangle_{\mathcal{H}} \in \mathbb{C}$$

define uma forma sesquilinear em $\mathcal{H}'$ e que é contínua (uma afirmação elementar em dimensão finita[32]). Assim, existe (pela Proposição 39.11, página 2045.) um operador linear que denotamos por $\mathrm{Ptr}_{\mathcal{H}''}(A)$ agindo em $\mathcal{H}'$ (ou seja, $\mathrm{Ptr}_{\mathcal{H}''}(A) \in \mathrm{Mat}(\mathbb{C}, n)$), denominado o *traço parcial* de A em relação a $\mathcal{H}''$, tal que $\omega(\psi, \phi) = \big\langle \psi, \mathrm{Ptr}_{\mathcal{H}''}(A)\phi \big\rangle_{\mathcal{H}'}$, ou seja, tal que

$$\big\langle \psi, \mathrm{Ptr}_{\mathcal{H}''}(A)\phi \big\rangle_{\mathcal{H}'} = \sum_{j=1}^{n} \big\langle (\psi \otimes \mathbf{f}_j), A(\phi \otimes \mathbf{f}_j) \big\rangle_{\mathcal{H}} . \tag{10.145}$$

Podemos obter expressões mais concretas para $\mathrm{Ptr}_{\mathcal{H}''}(A)$ se usarmos a decomposição de Schmidt para matrizes (10.144), página 546. Segundo ela podemos escrever

$$A = \sum_{k=1}^{l} d_k \big(A_k \otimes B_k \big) . \tag{10.146}$$

com $\mathrm{Tr}_{\mathbb{C}^m}\big(A_i^* A_j\big) = \delta_{i,j}$ e $\mathrm{Tr}_{\mathbb{C}^n}\big(B_i^* B_j\big) = \delta_{i,j}$, sendo $d_k \geq 0$ para todo k e $l = \min\{m^2, n^2\}$. Assim, vemos facilmente que

$$\big\langle \psi, \mathrm{Ptr}_{\mathcal{H}''}(A)\phi \big\rangle_{\mathcal{H}'} = \sum_{k=1}^{l} d_k \langle \psi, A_k \phi \rangle_{\mathcal{H}'} \left(\sum_{j=1}^{n} \big\langle \mathbf{f}_j, B_k \mathbf{f}_j \big\rangle_{\mathbb{C}^n} \right)$$

para todos $\psi, \phi \in \mathcal{H}'$ e assim,

$$\mathrm{Ptr}_{\mathcal{H}''}(A) = \sum_{k=1}^{l} d_k \mathrm{Tr}_{\mathbb{C}^n}\big(B_k\big) A_k , \tag{10.147}$$

sendo $\mathrm{Tr}_{\mathbb{C}^n}\big(B\big)$ o traço usual de $B \in \mathrm{Mat}(\mathbb{C}, n)$ em $\mathbb{C}^n$: $\mathrm{Tr}_{\mathbb{C}^n}\big(B_k\big) = \sum_{j=1}^{n} \big\langle \mathbf{f}_j, B_k \mathbf{f}_j \big\rangle_{\mathbb{C}^n}$.

A expressão (10.147) mostra que $\mathrm{Ptr}_{\mathcal{H}''}(A)$ é independe da base ortonormal $\{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ escolhida em $\mathcal{H}''$, pois é bem sabido que $\mathrm{Tr}_{\mathbb{C}^n}$ independe dessa escolha (vide Seção 10.2.3, página 478, e, em particular (10.33)).

• Algumas propriedades do traço parcial

Três outros resultados seguem facilmante de (10.147). O primeiro é a afirmação que se $A \in \mathrm{Mat}(\mathbb{C}, mn)$, $D \in \mathrm{Mat}(\mathbb{C}, n)$ e $E \in \mathrm{Mat}(\mathbb{C}, m)$, então

$$\mathrm{Ptr}_{\mathcal{H}''}\big(A(D \otimes E)\big) = \mathrm{Ptr}_{\mathcal{H}''}\big(A(\mathbb{1}_m \otimes E)\big) D , \tag{10.148}$$

e, em particular,

$$\mathrm{Ptr}_{\mathcal{H}''}\big(D \otimes E\big) = \mathrm{Ptr}_{\mathcal{H}''}\big(\mathbb{1}_m \otimes E\big) D = \mathrm{Tr}_{\mathbb{C}^n}(E) D . \tag{10.149}$$

De fato, por (10.146), $A(D \otimes E) = \sum_{k=1}^{l} d_k \big(A_k D \otimes B_k E\big)$ e, portanto

$$\mathrm{Ptr}_{\mathcal{H}''}\big(A(D \otimes E)\big) = \sum_{k=1}^{l} d_k A_k D \mathrm{Tr}_{\mathbb{C}^n}\big(B_k E\big) = \left(\sum_{k=1}^{l} d_k A_k \mathrm{Tr}_{\mathbb{C}^n}\big(B_k E\big) \right) D = \mathrm{Ptr}_{\mathcal{H}''}\big(A(\mathbb{1}_m \otimes E)\big) D , \tag{10.150}$$

provando (10.148).

[32] Esse ponto não é trivial em dimensão infinita e obriga-nos a restringirmo-nos a operadores de tipo traço. Vide Seção 39.11, página 2199

E. 10.40 *Exercício*. Seguindo os passos acima, mostre também que

$$\mathrm{Ptr}_{\mathcal{H}''}\big((D\otimes E)A\big) \;=\; D\mathrm{Ptr}_{\mathcal{H}''}\big((\mathbb{1}_m\otimes E)A\big)\;. \tag{10.151}$$

✠

O segundo resultado é uma versão particular da propriedade cíclica. A propriedade cíclica do traço em $\mathbb{C}^n$ permite-nos escrever, tomando $D=\mathbb{1}_m$ na primeira igualdade de (10.150),

$$\mathrm{Ptr}_{\mathcal{H}''}\big(A(\mathbb{1}_m\otimes E)\big) \;=\; \left(\sum_{k=1}^{l} d_k A_k \mathrm{Tr}_{\mathbb{C}^n}\big(B_k E\big)\right) \;=\; \left(\sum_{k=1}^{l} d_k A_k \mathrm{Tr}_{\mathbb{C}^n}\big(E B_k\big)\right) \;=\; \mathrm{Ptr}_{\mathcal{H}''}\big((\mathbb{1}_m\otimes E)A\big)\;,$$

demonstrando a propriedade cíclica particular

$$\mathrm{Ptr}_{\mathcal{H}''}\big(A(\mathbb{1}_m\otimes E)\big) \;=\; \mathrm{Ptr}_{\mathcal{H}''}\big((\mathbb{1}_m\otimes E)A\big)\;. \tag{10.152}$$

Propriedades cíclicas que igualem $\mathrm{Ptr}_{\mathcal{H}''}\big(AB\big)$ a $\mathrm{Ptr}_{\mathcal{H}''}\big(BA\big)$, para $A,\ B\in \mathrm{Mat}\,(\mathbb{C},\ mn)$ arbitrários, **não** são geralmente válidas para o traço parcial. Justifique!

O terceiro resultado é a afirmação que se $A\in\mathrm{Mat}\,(\mathbb{C},\ mn)$, então

$$\mathrm{Tr}_{\mathbb{C}^{mn}}\big(A\big) \;=\; \mathrm{Tr}_{\mathbb{C}^m}\Big(\mathrm{Ptr}_{\mathcal{H}''}(A)\Big)\;. \tag{10.153}$$

De fato,

$$\mathrm{Tr}_{\mathbb{C}^{mn}}\big(A\big) \;=\; \sum_{i=1}^{m}\sum_{j=1}^{n}\big\langle(\mathbf{v}_i\otimes\mathbf{w}_j),\ A(\mathbf{v}_i\otimes\mathbf{w}_j)\big\rangle_{\mathcal{H}} \overset{(10.146)}{=} \sum_{k=1}^{l} d_k \mathrm{Tr}_{\mathbb{C}^m}\big(A_k\big)\mathrm{Tr}_{\mathbb{C}^n}\big(B_k\big)$$

$$=\; \mathrm{Tr}_{\mathbb{C}^m}\left(\sum_{k=1}^{l} d_k A_k \mathrm{Tr}_n\big(B_k\big)\right) \;=\; \mathrm{Tr}_{\mathbb{C}^m}\Big(\mathrm{Ptr}_{\mathcal{H}''}(A)\Big)\;,$$

estabelecendo (10.153).

10.8.2.3 Purificação

Seja $\mathcal{H}_{\mathrm{I}}=\mathbb{C}^m$, um espaço de Hilbert de dimensão m, e seja ρ uma matriz densidade agindo em $\mathcal{H}_{\mathrm{I}}$, ou seja, um operador autoadjunto e positivo (portanto, com autovalores não-negativos) e satisfazendo $\mathrm{Tr}_{\mathbb{C}^m}(\rho)=1$.

Afirmamos que existe um espaço de Hilbert $\mathcal{H}_{\mathrm{II}}=\mathbb{C}^n$ (com $n\geq m$) e um vetor $\Psi\in\mathcal{H}:=\mathcal{H}_{\mathrm{I}}\otimes\mathcal{H}_{\mathrm{II}}\simeq\mathbb{C}^{nm}$ tais que

$$\mathrm{Tr}_{\mathbb{C}^m}(\rho D) \;=\; \big\langle\Psi,\ (D\otimes\mathbb{1}_n)\Psi\big\rangle_{\mathcal{H}}\;. \tag{10.154}$$

Sejam $\rho_1,\ \ldots,\ \rho_m$ os autovalores de ρ (incluindo multiplicidades), com $\rho_k\geq 0$ para todo k, e sejam $\mathbf{v}_1,\ \ldots,\ \mathbf{v}_m$ os correspondentes autovetores compondo uma base ortonormal em $\mathbb{C}^m$: $\big\langle\mathbf{v}_i,\ \mathbf{v}_j\big\rangle_{\mathbb{C}^m}=\delta_{i,j}$. A decomposição espectral de ρ é $\rho=\sum_{i=1}^m\rho_i P_{\mathbf{v}_i}$, onde $P_{\mathbf{v}_i}$ é o projetor ortogonal sobre o subespaço unidimensional gerado por $\mathbf{v}_i$.

Seja $n\geq m$ e $\mathbf{w}_1,\ \ldots,\ \mathbf{w}_n$ uma base ortonormal em $\mathbb{C}^n$ (com $\big\langle\mathbf{w}_i,\ \mathbf{w}_j\big\rangle_{\mathbb{C}^n}=\delta_{i,j}$). Definamos $\Psi\in\mathbb{C}^m\otimes\mathbb{C}^m\simeq\mathbb{C}^{nm}$ por

$$\Psi \;:=\; \sum_{i=1}^{m}\sqrt{\rho_i}\big(\mathbf{v}_i\otimes\mathbf{w}_i\big)\;. \tag{10.155}$$

Temos

$$\big\langle\Psi,\ (D\otimes\mathbb{1}_n)\Psi\big\rangle_{\mathcal{H}} \;=\; \sum_{i=1}^{m}\sum_{j=1}^{m}\sqrt{\rho_i}\sqrt{\rho_j}\Big\langle\mathbf{v}_i\otimes\mathbf{w}_i,\ \big(D\mathbf{v}_j\big)\otimes\mathbf{w}_j\Big\rangle_{\mathcal{H}}$$

$$=\; \sum_{i=1}^{m}\sum_{j=1}^{m}\sqrt{\rho_i}\sqrt{\rho_j}\big\langle\mathbf{v}_i,\ \big(D\mathbf{v}_j\big)\big\rangle_{\mathbb{C}^m}\underbrace{\big\langle\mathbf{w}_i,\ \mathbf{w}_j\big\rangle_{\mathbb{C}^n}}_{\delta_{i,j}} \;=\; \sum_{i=1}^{m}\rho_i\big\langle\mathbf{v}_i,\ \big(D\mathbf{v}_i\big)\big\rangle_{\mathbb{C}^m} \;=\; \mathrm{Tr}_{\mathbb{C}^m}(\rho D)\;,$$

demonstrando que (10.154) é satisfeita, como desejávamos.

E. 10.41 *Exercício*. Justifique por que foi necessário tomar $n \geq m$ na construção do vetor Ψ. ✠

De (10.154) segue também (tomando-se $D = \mathbb{1}_m$) que $\|\Psi\|^2_{\mathcal{H}} = \mathrm{Tr}_{\mathbb{C}^m}(\rho) = 1$.

O vetor Ψ dado em (10.155) depende de ρ (através dos autovalores ρ_i e dos autovetores $\mathbf{v}_i$) e da escolha arbitrária dos vetores ortonormais $\mathbf{w}_j$ de $\mathbb{C}^n$. O lado esquerdo de (10.154), porém, independe da escolha dos $\mathbf{w}_j$'s.

No jargão da Teoria da Informação Quântica, o vetor Ψ é dito ser uma *purificação* de ρ. Essa noção será estendida a espaços de Hilbert de dimensão infinita na Seção 46.6.1, página 2542, quando discutiremos também sua interpretação.

• **Relação com o traço parcial**

Considere-se o vetor $\Psi \in \mathbb{C}^{mn}$ dado em (10.155) e calculemos o traço parcial em relação a $\mathcal{H}_{\mathrm{II}} = \mathbb{C}^n$ do projetor unidimensional P_Ψ (recordar que $\|\Psi\| = 1$). Por (10.145), escolhendo $\mathbf{f}_j = \mathbf{w}_j$ para todo $j \in \mathbb{N}$, temos para todos $\psi, \phi \in \mathcal{H}_{\mathrm{I}} = \mathbb{C}^m$,

$$
\begin{aligned}
\big\langle \psi,\ \mathrm{Ptr}_{\mathcal{H}_{\mathrm{II}}}(P_\Psi)\phi \big\rangle_{\mathcal{H}_{\mathrm{I}}} &= \sum_{j=1}^{n} \big\langle (\psi \otimes \mathbf{w}_j),\ P_\Psi(\phi \otimes \mathbf{w}_j)\big\rangle_{\mathcal{H}} = \sum_{j=1}^{n} \big\langle (\psi \otimes \mathbf{w}_j),\ \Psi\big\rangle_{\mathcal{H}} \big\langle \Psi,\ (\phi \otimes \mathbf{w}_j)\big\rangle_{\mathcal{H}} \\
&\overset{(10.155)}{=} \sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{m} \sqrt{\rho_k}\sqrt{\rho_l} \Big\langle \big(\psi \otimes \mathbf{w}_j\big),\ \big(\mathbf{v}_k \otimes \mathbf{w}_k\big)\Big\rangle_{\mathcal{H}} \Big\langle \big(\mathbf{v}_l \otimes \mathbf{w}_l\big),\ \big(\phi \otimes \mathbf{w}_j\big)\Big\rangle_{\mathcal{H}} \\
&= \sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{m} \sqrt{\rho_k}\sqrt{\rho_l} \big\langle \psi,\ \mathbf{v}_k\big\rangle_{\mathbb{C}^m} \big\langle \mathbf{v}_l,\ \phi\big\rangle_{\mathbb{C}^m} \delta_{j,\,k}\delta_{j,\,l} \\
&= \sum_{j=1}^{m} \rho_j \big\langle \psi,\ \mathbf{v}_j\big\rangle_{\mathbb{C}^m} \big\langle \mathbf{v}_j,\ \phi\big\rangle_{\mathbb{C}^m} = \big\langle \psi,\ \rho\phi\big\rangle_{\mathbb{C}^m} ,
\end{aligned}
$$

provando a interessante relação

$$
\mathrm{Ptr}_{\mathcal{H}_{\mathrm{II}}}(P_\Psi) = \rho . \tag{10.156}
$$

A relação (10.151) para $E = \mathbb{1}_n$ e $A = P_\Psi$ fica

$$
\mathrm{Ptr}_{\mathcal{H}''}\big((D \otimes \mathbb{1}_n)P_\Psi\big) = D\mathrm{Ptr}_{\mathcal{H}''}\big(P_\Psi\big) = D\rho .
$$

Portanto, temos de (10.153),

$$
\mathrm{Tr}_{\mathbb{C}^m}(D\rho) = \mathrm{Tr}_{\mathbb{C}^m}\Big(\mathrm{Ptr}_{\mathcal{H}''}\big((D \otimes \mathbb{1}_n)P_\Psi\big)\Big) \overset{(10.153)}{=} \mathrm{Tr}_{\mathbb{C}^{mn}}\big((D \otimes \mathbb{1}_n)P_\Psi\big) = \big\langle \Psi,\ (D \otimes \mathbb{1}_n)\Psi\big\rangle_{\mathbb{C}^{mn}} ,
$$

que é uma nova forma de se justificar (10.154).

10.8.3 O Teorema da Triangularização de Schur

O teorema que apresentamos abaixo, devido a Schur[33], é semelhante, mas não idêntico, ao Teorema de Jordan: toda matriz de $\mathrm{Mat}(\mathbb{C},\ n)$ pode ser levada por uma transformação de similaridade induzida por uma matriz unitária a uma matriz triangular superior (para a definição, vide Seção 10.6, página 523). Esse teorema é alternativamente denominado *Teorema da Triangularização de Schur* ou *Teorema da Decomposição de Schur*. Como veremos, esse teorema pode ser usado para fornecer uma outra demonstração (eventualmente mais simples) da diagonalizabilidade de matrizes autoadjuntas e de matrizes normais por matrizes unitárias.

Teorema 10.29 (Teorema da Decomposição de Schur) *Seja $A \in \mathrm{Mat}(\mathbb{C},\ n)$. Então, existe $U \in \mathrm{Mat}(\mathbb{C},\ n)$, unitária, e $S \in \mathrm{Mat}(\mathbb{C},\ n)$, triangular superior, tais que $A = U^*SU$. Os elementos da diagonal de S são os autovalores de A.* □

[33] Issai Schur (1875–1941).

Antes de provarmos esse teorema, mencionemos um corolário evidente:

Corolário 10.8 *Seja $A \in \mathrm{Mat}(\mathbb{C},\ n)$. Então, existe $V \in \mathrm{Mat}(\mathbb{C},\ n)$, unitária, e $I \in \mathrm{Mat}(\mathbb{C},\ n)$, triangular inferior, tais que $A = V^* I V$. Os elementos da diagonal de I são os autovalores de A.* □

Prova do Corolário 10.8. Pelo Teorema 10.29, a matriz A^* pode ser escrita da forma $A^* = V^* S V$, com V unitária e S triangular superior. Logo, $A = V^* S^* V$. Porém, $S^* \equiv I$ é triangular inferior.

Também pelo Teorema 10.29, os autovalores de A^* são os elementos diagonais de S, que são o complexo conjugado dos elementos diagonais de $S^* \equiv I$. Mas os autovalores de A são o complexo conjugado dos autovalores de A^* (pela Proposição 10.24, página 507) e, portanto, são os elementos diagonais de I. ■

Prova do Teorema 10.29. Comecemos observando que se $A = U^* S U$ com U unitário, então A e S têm o mesmo polinômio característico e, portanto, os mesmos autovalores, incluindo a multiplicidade (vide a discussão em torno de (10.32), página 474). Mas o polinômio característico de S é $p_S(x) = \det(x\mathbb{1} - S) = \prod_{k=1}^{n}(x - S_{kk})$, pois S é triangular superior e, portanto, os autovalores de S são os elementos de sua diagonal. Passemos à demonstração da afirmativa principal, ou seja, que $A = U^* S U$ com U unitário e S triangular superior.

Seja $n \geq 2$ e v_1 um autovetor de A com autovalor λ_1 e $\|v_1\| = 1$. Seja $U^{(1)}$ uma matriz unitária da forma $U^{(1)} = \left[\!\left[u_1^{(1)},\ \ldots,\ u_n^{(1)}\right]\!\right]$ com $u_1^{(1)} = v_1$, ou seja, cuja primeira coluna é o vetor v_1. Então,

$$
AU^{(1)} \stackrel{(10.12)}{=} \left[\!\left[Au_1^{(1)},\ \ldots,\ Au_n^{(1)}\right]\!\right] = \left[\!\left[\lambda_1 u_1^{(1)},\ Au_2^{(1)},\ \ldots,\ Au_n^{(1)}\right]\!\right] = U^{(1)} \begin{pmatrix} \lambda_1 & b_1^{(1)} & \cdots & b_{n-1}^{(1)} \\ 0 & a_{11}^{(1)} & \cdots & a_{1(n-1)}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{(n-1)1}^{(1)} & \cdots & a_{(n-1)(n-1)}^{(1)} \end{pmatrix},
$$

para certos $b_k^{(1)}$ e $a_{kl}^{(1)}$, $k,\ l = 1,\ \ldots,\ n-1$, onde

$$
Au_k^{(1)} = b_k^{(1)} u_1^{(1)} + \sum_{l=1}^{n-1} a_{lk}^{(1)} u_{l+1}^{(1)}, \qquad k = 2,\ \ldots,\ n. \tag{10.157}
$$

Para simplificar a notação, definimos

$$
b^{(1)} = \begin{pmatrix} b_1^{(1)} \\ \vdots \\ b_{n-1}^{(1)} \end{pmatrix}, \qquad \mathbb{0}_{n-1} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad A^{(1)} = \begin{pmatrix} a_{11}^{(1)} & \cdots & a_{1(n-1)}^{(1)} \\ \vdots & \ddots & \vdots \\ a_{(n-1)1}^{(1)} & \cdots & a_{(n-1)(n-1)}^{(1)} \end{pmatrix},
$$

($\mathbb{0}_{n-1}$ tendo $n-1$ linhas) e escrevemos a identidade (10.157) como

$$
{U^{(1)}}^* A U^{(1)} = \begin{pmatrix} \lambda_1 & {b^{(1)}}^T \\ \mathbb{0}_{n-1} & A^{(1)} \end{pmatrix}. \tag{10.158}
$$

Para $n = 2$ isso demonstra o teorema, pois afirma que

$$
{U^{(1)}}^* A U^{(1)} = \begin{pmatrix} \lambda_1 & b_1^{(1)} \\ 0 & a_{11}^{(1)} \end{pmatrix},
$$

sendo o lado direito uma matriz triangular superior. Para $n > 2$ procedemos por indução. Supondo a afirmação válida para matrizes $(n-1) \times (n-1)$, então existe uma matriz unitária $V \in \mathrm{Mat}(\mathbb{C},\ n-1)$ tal que $V^* A^{(1)} V = S^{(1)}$, sendo $S^{(1)}$ triangular superior. Assim, definindo a matriz unitária $U^{(2)} \in \mathrm{Mat}(\mathbb{C},\ n)$ por $U^{(2)} := \begin{pmatrix} 1 & \mathbb{0}_{n-1}^T \\ \mathbb{0}_{n-1} & V \end{pmatrix}$, teremos por (10.158),

$$
\begin{aligned}
\big(U^{(1)}U^{(2)}\big)^* A U^{(1)} U^{(2)} &= {U^{(2)}}^* {U^{(1)}}^* A U^{(1)} U^{(2)} \\
&= \begin{pmatrix} 1 & \mathbb{0}_{n-1}^T \\ \mathbb{0}_{n-1} & V^* \end{pmatrix} \begin{pmatrix} \lambda_1 & {b^{(1)}}^T \\ \mathbb{0}_{n-1} & A^{(1)} \end{pmatrix} \begin{pmatrix} 1 & \mathbb{0}_{n-1}^T \\ \mathbb{0}_{n-1} & V \end{pmatrix} \\
&= \begin{pmatrix} \lambda_1 & \big(V^T b^{(1)}\big)^T \\ \mathbb{0}_{n-1} & V^* A^{(1)} V \end{pmatrix} \\
&= \begin{pmatrix} \lambda_1 & \big(V^T b^{(1)}\big)^T \\ \mathbb{0}_{n-1} & S^{(1)} \end{pmatrix} ,
\end{aligned}
$$

que é triangular superior, pois $S^{(1)}$ o é. Como $U^{(1)}U^{(2)}$ é unitária (pois $U^{(1)}$ e $U^{(2)}$ o são), o teorema está provado. ■

<u>*Comentário*</u>. Toda matriz triangular superior S pode ser escrita na forma $D+N$, sendo D a matriz diagonal formada pela diagonal de S (ou seja, $D_{ii} = S_{ii}$ para todo $i = 1, \ldots, n$) e N é nilpotente (pois é triangular superior, mas com diagonal nula). Assim, o Teorema 10.29 afirma que toda matriz A pode ser levada à forma $D+N$ por uma transformação de similaridade unitária. Porém, o Teorema 10.29 não garante (nem é verdade, em geral) que D e N comutem. Assim, o Teorema 10.29 é distinto do Teorema de Jordan, Teorema 10.22, página 530. ♣

O Teorema 10.29 tem por corolário o seguinte teorema, já provado anteriormente por outros meios (Teorema 10.15, página 510, e Proposição 10.26, página 511).

Teorema 10.30 *Uma matriz $A \in \mathrm{Mat}(\mathbb{C},\ n)$ é autoadjunta, se e somente se for diagonalizável por uma transformação de similaridade unitária e se seus autovalores forem reais.* □

Prova. Pelo Teorema 10.29, existe uma matriz unitária U tal que $U^*AU = S$, sendo S triangular superior cujos elementos diagonais são os autovalores de A. Assim, se $A = A^*$, segue que $S^* = (U^*AU)^* = U^*A^*U = U^*AU = S$. Mas para uma matriz triangular superior S, a igualdade $S = S^*$ implica que S é diagonal e os elementos da diagonal são reais.

Reciprocamente, se $A \in \mathrm{Mat}(\mathbb{C},\ n)$ é diagonalizável por uma transformação de similaridade unitária e seus autovalores são reais, ou seja, existe U unitária e D diagonal real com $U^*AU = D$, então $A = UDU^*$ e $A^* = UD^*U^*$. Como D é diagonal e real, vale $D^* = D$ e, portanto, $A^* = UDU^* = A$, provando que A é autoadjunta. ■

Pelo Teorema 10.29, se $A \in \mathrm{Mat}(\mathbb{C},\ n)$ é uma matriz normal e $U^*AU = S$, com U unitária e S triangular superior, então S é normal (justifique!). Assim, junto com o Lema 10.6, página 523, provamos o seguinte:

Teorema 10.31 *Uma matriz $A \in \mathrm{Mat}(\mathbb{C},\ n)$ é normal se e somente se for diagonalizável por uma transformação de similaridade unitária.* □

Essas afirmações foram demonstradas por outros meios no Teorema 10.17, página 512.

10.8.4 A Decomposição QR e a Decomposição de Iwasawa ("KAN")

O propósito desta seção é apresentar a chamada *decomposição de Iwasawa*[34], ou *decomposição* KAN[35], de matrizes inversíveis, Teorema 10.33. Esse teorema tem relação com a teoria dos grupos de Lie, como discutiremos brevemente ao final. Os dois primeiros resultados preparatórios abaixo, Proposição 10.34 e Teorema 10.32 (Decomposição QR), têm interesse por si só.

Proposição 10.34 *Seja* $R \in \mathrm{Mat}(\mathbb{C}, n)$ *uma matriz triangular superior cujos elementos diagonais são não nulos (i.e.,* R *é inversível). Então, podemos escrever* $R = AN$*, onde* $A \in \mathrm{Mat}(\mathbb{C}, n)$ *é a matriz diagonal formada com a diagonal de* R*:* $A = \mathrm{diag}(R_{11}, \ldots, R_{nn})$*, e* $N \in \mathrm{Mat}(\mathbb{C}, n)$ *é uma matriz triangular superior cujos elementos diagonais são iguais a* 1. □

Prova. É fácil constatar que (abaixo $m \equiv n-1$)

$$R = \begin{pmatrix} R_{11} & R_{12} & \cdots & \cdots & R_{1n} \\ 0 & R_{22} & \ddots & & R_{2n} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & & \ddots & R_{mm} & R_{mn} \\ 0 & \cdots & \cdots & 0 & R_{nn} \end{pmatrix} = \underbrace{\begin{pmatrix} R_{11} & 0 & \cdots & \cdots & 0 \\ 0 & R_{22} & \ddots & & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & & \ddots & R_{mm} & 0 \\ 0 & \cdots & \cdots & 0 & R_{nn} \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} 1 & \frac{R_{12}}{R_{11}} & \cdots & \cdots & \frac{R_{1n}}{R_{11}} \\ 0 & 1 & \ddots & & \frac{R_{2n}}{R_{22}} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & & \ddots & 1 & \frac{R_{mn}}{R_{mm}} \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}}_{N} .$$

■

O estudante deve comparar as afirmações do teorema a seguir com o Teorema da Decomposição Polar, Teorema 10.24, página 540, e com o Teorema da Decomposição de Schur, Teorema 10.29, página 549.

Teorema 10.32 (Teorema da Decomposição QR) *Seja* $M \in \mathrm{Mat}(\mathbb{C}, n)$ *uma matriz inversível. Então,* M *pode ser escrita na forma* $M = QR$*, onde* $Q \in \mathrm{Mat}(\mathbb{C}, n)$ *é unitária e* $R \in \mathrm{Mat}(\mathbb{C}, n)$ *é triangular superior, sendo que os elementos diagonais de* R *são estritamente positivos.*

Prova do Teorema 10.32. Seja $M = [\![\mathbf{m}_1, \ldots, \mathbf{m}_n]\!]$. Como M é inversível, os vetores $\mathbf{m}_k$, $k = 1, \ldots, n$, são linearmente independentes, ou seja, formam uma base em $\mathbb{C}^n$. Podemos, portanto, usar o procedimento de ortogonalização de Gram-Schmidt (vide Seção 3.3, página 242) e construir uma nova base ortonormal de vetores $\mathbf{q}_j$, $j = 1, \ldots, n$, a partir dos vetores $\mathbf{m}_l$, $l = 1, \ldots, n$. Tais vetores são definidos por

$$\mathbf{q}_1 = \frac{\mathbf{m}_1}{\|\mathbf{m}_1\|}, \qquad \mathbf{q}_j = \frac{\mathbf{m}_j - \sum_{l=1}^{j-1} \langle \mathbf{q}_l, \mathbf{m}_j \rangle_{\mathbb{C}} \mathbf{q}_l}{\left\| \mathbf{m}_j - \sum_{l=1}^{j-1} \langle \mathbf{q}_l, \mathbf{m}_j \rangle_{\mathbb{C}} \mathbf{q}_l \right\|}, \qquad j = 2, \ldots, n .$$

Como é fácil verificar, tem-se $\langle \mathbf{q}_i, \mathbf{q}_j \rangle_{\mathbb{C}} = \delta_{ij}$ para todos $i, j = 1, \ldots, n$. As relações acima implicam trivialmente

$$\mathbf{m}_1 = \mathbf{q}_1 \|\mathbf{m}_1\| , \qquad \mathbf{m}_j = \mathbf{q}_j \left\| \mathbf{m}_j - \sum_{l=1}^{j-1} \langle \mathbf{q}_l, \mathbf{m}_j \rangle_{\mathbb{C}} \mathbf{q}_l \right\| + \sum_{l=1}^{j-1} \mathbf{q}_l \langle \mathbf{q}_l, \mathbf{m}_j \rangle_{\mathbb{C}} , \qquad j = 2, \ldots, n ,$$

[34] Kenkichi Iwasawa (1917–1998).

[35] Infelizmente não há uniformidade na literatura quanto à denominação dessa decomposição. Vamos chamá-la de "decomposição de Iwasawa" pois a mesma é um caso particular (para o grupo $\mathrm{GL}(\mathbb{C}, n)$ das matrizes complexas $n \times n$ inversíveis) de um teorema mais geral da teoria dos grupos de Lie, denominado Teorema da Decomposição de Iwasawa, que afirma que todo elemento g de um grupo de Lie semissimples pode ser escrito como produto de um elemento k de um subgrupo compacto maximal, por um elemento a de um subgrupo Abeliano (real) e por um elemento n de um subgrupo nilpotente (ou seja, cuja álgebra de Lie é nilpotente): $g = kan$. Em Alemão, as palavras compacto, Abeliano e nilpotente são "Kompakt", "Abelsch" e "Nilpotent", daí a denominação "decomposição KAN" para essa decomposição, denominação essa encontrada em alguns textos.

relações estas que podem ser escritas em forma matricial como

$$
\Big[\!\!\Big[\mathbf{m}_1,\,\ldots,\,\mathbf{m}_n\Big]\!\!\Big] \;=\; \Big[\!\!\Big[\mathbf{q}_1,\,\ldots,\,\mathbf{q}_n\Big]\!\!\Big]\,R\,,\quad \text{onde}\quad R\;:=\;\begin{pmatrix} R_{11} & \langle\mathbf{q}_1,\,\mathbf{m}_2\rangle_{\mathbb{C}} & \cdots & \cdots & \langle\mathbf{q}_1,\,\mathbf{m}_n\rangle_{\mathbb{C}} \\ 0 & R_{22} & \ddots & \cdots & \langle\mathbf{q}_2,\,\mathbf{m}_n\rangle_{\mathbb{C}} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & & \ddots & R_{(n-1)(n-1)} & \langle\mathbf{q}_{n-1},\,\mathbf{m}_n\rangle_{\mathbb{C}} \\ 0 & \cdots & \cdots & 0 & R_{nn} \end{pmatrix}\,, \tag{10.159}
$$

com

$$
R_{11} \;=\; \|\mathbf{m}_1\|\,,\qquad R_{jj} \;=\; \left\|\mathbf{m}_j-\sum_{l=1}^{j-1}\langle\mathbf{q}_l,\,\mathbf{m}_j\rangle_{\mathbb{C}}\,\mathbf{q}_l\right\|\,,\qquad j\;=\;2,\,\ldots,\,n\,.
$$

E. 10.42 *Exercício*. Convença-se da validade da relação (10.159). ✣

Definindo $Q := [\![\mathbf{q}_1,\,\ldots,\,\mathbf{q}_n]\!]$, a relação (10.159) diz-nos que $M = QR$, sendo R triangular superior (como se vê) e Q unitária (pois os vetores $\mathbf{q}_l$, $l = 1,\,\ldots,\,n$, são ortonormais). Isso completa a prova do Teorema 10.32. ■

Chegamos assim ao importante Teorema da Decomposição de Iwasawa para matrizes inversíveis:

Teorema 10.33 (Teorema da Decomposição de Iwasawa, ou Decomposição KAN) *Seja $M \in \mathrm{Mat}\,(\mathbb{C},\ n)$ uma matriz inversível. Então, M pode ser escrita de modo único na forma $M = KAN$, onde $K \in \mathrm{Mat}\,(\mathbb{C},\ n)$ é uma matriz unitária, $A \in \mathrm{Mat}\,(\mathbb{C},\ n)$ é a uma matriz diagonal, tendo elementos diagonais estritamente positivos, e $N \in \mathrm{Mat}\,(\mathbb{C},\ n)$ é uma matriz triangular superior cujos elementos diagonais são iguais a 1.* □

Prova. A afirmação que M pode ser escrita na forma $M = KAN$, com K, A e N com as propriedades acima segue imediatamente da Proposição 10.34 e do Teorema 10.32, dispensando demonstração. O único ponto a se demonstrar é a unicidade dessa decomposição.

Vamos então supor que para algum $M \in \mathrm{Mat}\,(\mathbb{C},\ n)$ existam $K,\ K_0 \in \mathrm{Mat}\,(\mathbb{C},\ n)$, matrizes unitárias, $A,\ A_0 \in \mathrm{Mat}\,(\mathbb{C},\ n)$, matrizes diagonais, tendo elementos diagonais estritamente positivos, e $N,\ N_0 \in \mathrm{Mat}\,(\mathbb{C},\ n)$ matrizes triangulares superiores cujos elementos diagonais são iguais a 1, tais que $M = KAN = K_0A_0N_0$.

Segue imediatamente disso que $K_0^{-1}K = A_0N_0N^{-1}A^{-1}$. O lado esquerdo dessa igualdade é uma matriz unitária e, portanto, normal. O lado direito é uma matriz triangular superior (pela Proposição 10.30, página 523). Pelo Lema 10.6, página 523, $A_0N_0N^{-1}A^{-1}$ deve ser uma matriz diagonal D. Assim, temos que $K_0^{-1}K = D$ e $A_0N_0N^{-1}A^{-1} = D$. A primeira dessas relações diz-nos que D é unitária. A segunda diz-nos que $N_0N^{-1} = A_0^{-1}DA$, ou seja, $N_0 = D_0N$, onde $D_0 := A_0^{-1}DA$ é diagonal (por ser o produto de três matrizes diagonais). Agora, N e N_0 são matrizes triangulares superiores cujos elementos diagonais são iguais a 1. Portanto, a relação $N_0 = D_0N$ com D_0 diagonal só é possível se $D_0 = \mathbb{1}$ (de outra forma haveria elementos na diagonal de N ou de N_0 diferentes de 1), estabelecendo que $N = N_0$.

Provamos, assim, que $A_0^{-1}DA = \mathbb{1}$, ou seja, $D = A_0A^{-1}$. Agora, A e A_0 são diagonais, tendo na diagonal números reais positivos. Logo, D também é diagonal e tem na diagonal números reais positivos e, portanto, $D = D^*$. Como D é unitária (como observado linhas acima), segue que $D^2 = \mathbb{1}$. Logo, os elementos D_{kk} da diagonal de D satisfazem $D_{kk} = \pm 1$, para todo $k = 1,\,\ldots,\,n$ (os sinais podendo ser distintos para k's distintos). Agora, como $A_0 = DA$ e como A e A_0 têm na diagonal números reais positivos, não podemos ter $D_{kk} = -1$ para algum k e, portanto, $D = \mathbb{1}$. Consequentemente, $K = K_0$ e $A = A_0$, estabelecendo a unicidade desejada. ■

Note o leitor que o conjunto das matrizes unitárias de $\mathrm{Mat}\,(\mathbb{C},\ n)$ forma um subgrupo de $\mathrm{GL}(\mathbb{C},\ n)$ (o grupo das matrizes complexas $n \times n$ inversíveis). O conjunto das matrizes diagonais de $\mathrm{Mat}\,(\mathbb{C},\ n)$ tendo elementos diagonais estritamente positivos é igualmente um subgrupo de $\mathrm{GL}(\mathbb{C},\ n)$. Por fim, o conjunto das matrizes triangulares superiores de $\mathrm{Mat}\,(\mathbb{C},\ n)$ cujos elementos diagonais são iguais a 1 é também um subgrupo de $\mathrm{GL}(\mathbb{C},\ n)$. Assim, o Teorema 10.33 afirma que cada elemento de $\mathrm{GL}(\mathbb{C},\ n)$ pode ser escrito de modo único como produto de elementos de cada um desses três subgrupos. Esse é um caso particular de um teorema da teoria dos grupos de Lie conhecido como *Teorema da Decomposição de Iwasawa*.

10.8.5 Diagonalização em Blocos de Matrizes Antissimétricas Reais

Vamos aqui estabelecer que matrizes reais antissimétricas podem ser postas em uma forma de blocos diagonais específica. Esse resultado é usado na demonstração de um importante Teorema (Teorema de Williamson, Teorema 10.36, página 561) sobre o grupo simplético. Ele também é relevante em áreas da chamada Geometria Simplética. Começamos com algumas observações preliminares.

• **Alguns resultados preliminares sobre matrizes antissimétricas**

Façamos primeiramente algumas afirmações elementares. Uma matriz real $A \in \mathrm{Mat}\,(\mathbb{R},\ m)$, com $m \in \mathbb{N}$, $m \geq 2$, é dita ser *antissimétrica* se $A^T = -A$. O caso $m = 1$ é trivial, a única matriz antissimétrica nesse caso é 0. Note-se que os elementos da diagonal de uma matriz antissimétrica são nulos.

O espectro de uma matriz antissimétrica é composto por números imaginários puros (e, eventualmente, pelo zero). De fato, se A é antissimétrica, a matriz $-iA$ é autoadjunta e, portanto, tem espectro real, o que leva a concluir que os autovalores de A são múltiplos reais de i.

Por exemplo, a matriz $A = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ tem por polinômio característico $q_A(z) = \det(z\mathbb{1} - A) = z^2 + 1$, cujas raízes, os autovalores de A, são $\pm i$. Os autovetores não nulos correspondentes a esses autovalores são múltiplos de $\left(\begin{smallmatrix} 1 \\ \pm i \end{smallmatrix}\right)$, respectivamente. Verifique!

Assim, mesmo que uma matriz antissimétrica $A \in \mathrm{Mat}\,(\mathbb{R},\ m)$ seja real, seus autovalores e autovetores são complexos e, por essa razão, é conveniente considerarmos $\mathrm{Mat}\,(\mathbb{R},\ m)$ como uma subálgebra real de $\mathrm{Mat}\,(\mathbb{C},\ m)$ agindo em $\mathbb{C}^m$.

Recordemos que os valores singulares de uma matriz $M \in \mathrm{Mat}\,(\mathbb{R},\ m)$ (não necessariamente antissimétrica) são os autovalores de $\sqrt{M^T M}$. (Vide Seção 10.8.2, página 542).

Seja $M \in \mathrm{Mat}\,(\mathbb{R},\ m)$ uma matriz arbitrária e seja a matriz simétrica $M^T M$. Os autovalores de $M^T M$ são reais (pois $M^T M$ é real e simétrica e, portanto, autoadjunta) e não negativos, pois se $\lambda \in \mathbb{R}$ é um autovalor de $M^T M$ e $u \in \mathbb{C}^m$ um correspondente autovetor normalizado (isto é, tal que $\|u\|^2_{\mathbb{C}} = \langle u,\ u\rangle_{\mathbb{C}} = 1$), então $\lambda = \langle u,\ \lambda u\rangle_{\mathbb{C}} = \langle u,\ M^T M u\rangle_{\mathbb{C}} = \langle Mu,\ Mu\rangle_{\mathbb{C}} = \|Mu\|^2_{\mathbb{C}} \geq 0$.

Uma matriz $M \in \mathrm{Mat}\,(\mathbb{R},\ m)$ (não necessariamente antissimétrica) é não-singular se e somente se os autovalores de $M^T M$ são positivos. De fato, se $\lambda = 0$ for autovalor de $M^T M$, então, pelo que acabamos de ver, um correspondente autovetor normalizado $u \in \mathbb{R}^m$ satisfaz $Mu = 0$, o que implica que M não tem inversa, contrariando a hipótese. Reciprocamente, se M não tem inversa, então existe (Corolário 10.1, página 466) $u \in \mathbb{C}^m$ não nulo tal que $Mu = 0$, o que implica trivialmente $M^T M u = 0$, mostrando que $M^T M$ tem autovalor nulo.

Se λ é autovalor de uma matriz antissimétrica $A \in \mathrm{Mat}\,(\mathbb{R},\ m)$, então, por definição, $\det(\lambda\mathbb{1} - A) = 0$. Logo, $0 = \det\big((\lambda\mathbb{1} - A)^T\big) = \det\big(\lambda\mathbb{1} - A^T\big) = \det\big(\lambda\mathbb{1} + A\big) = (-1)^m \det\big(-\lambda\mathbb{1} - A\big)$, o que mostra que $-\lambda$ também é autovalor de A.

Se $A \in \mathrm{Mat}\,(\mathbb{R},\ m)$ é antissimétrica e m é ímpar, então $\det(A) = \det\big(-A^T\big) = -\det\big(A^T\big) = -\det(A)$. Logo, $\det(A) = 0$ e, portanto, A não possui inversa e $A^T A$ possui autovalor nulo.

Como iA é autoadjunta, A é simples, e portanto, se um autovalor λ é degenerado, o subespaço dos autovetores correspondentes possui a mesma degenerescência. Para $\lambda \neq 0$ essa degenerescência se repassa ao autovalor $-\lambda$, como mostra o seguinte argumento. O polinômio característico de A é $q_A(z) = \det(z\mathbb{1} - A) = \prod_{j=1}^m (z - \lambda_j)$, onde λ_j são os autovalores não necessariamente distintos de A. Como A é antissimétrico, vale $q_A(z) = \det\big((z\mathbb{1} - A)^T\big) = \det\big(z\mathbb{1} - A^T\big) =$

$\det(z\mathbb{1} + A) = (-1)^m \det(-z\mathbb{1} - A) = (-1)^m q_A(-z)$. Assim,

$$\prod_{j=1}^{m}(z-\lambda_j) \;=\; (-1)^m \prod_{j=1}^{m}(-z-\lambda_j) \;=\; \prod_{j=1}^{m}(z+\lambda_j)\;. \tag{10.160}$$

Essa igualdade mostra que cada autovalor λ_j possui a mesma degenerescência algébrica (e, portanto, geométrica, pois A é simples), que o autovalor $-\lambda_j$. Os autovetores de $A \in \mathrm{Mat}(\mathbb{R},\ m)$ são também autovetores da matriz autoadjunta iA e, portanto, os subespaços de autovalores distintos são ortogonais em $\mathbb{C}^m$. Segue dessas observações que o posto de A (a dimensão de sua imagem) sempre possui dimensão par, independente de m.

Reunindo os resultados de acima, as conclusões relevantes para nós são as seguintes:

Proposição 10.35 *Seja uma matriz antissimétrica $A \in \mathrm{Mat}(\mathbb{R},\ m)$, $m \geq 2$. Os autovalores de A são múltiplos reais de $i \equiv \sqrt{-1}$ (incluindo eventualmente o zero) e se $\lambda \neq 0$ é um autovalor, $-\lambda$ também o é e com a mesma degenerescência. Se m for ímpar, A sempre possui ao menos um autovalor nulo. Se m for par pode ou não possuir valores singulares nulos, mas (evidentemente) não possuirá se e somente se A for não singular. Em todos os casos o posto de A é sempre par. Os subespaços de autovalores distintos de A são subespaços ortogonais em $\mathbb{C}^m$.*

Podemos refrasear essas afirmações da seguinte forma. Se A tem N autovalores distintos não nulos, então $\mathbb{C}^m = S_1 \oplus \cdots \oplus S_N \oplus Z$, onde cada S_k é um subespaço associado a um autovetor não nulo distinto e Z o núcleo de A. Se m é ímpar, Z é não trivial. Se S_a é o subespaço associado a um autovalor $\lambda \neq 0$, então existe um outro subespaço S_b distinto associado ao autovalor $-\lambda$ e suas dimensões são iguais: $\dim(S_a) = \dim(S_b)$. Com isso, vemos que N é par e que a soma $\dim(S_1) + \cdots + \dim(S_N)$, que é o posto de A, e é também um número par. □

10.8.5.1 Resultado Principal. Enunciado e Demonstração

Começamos com um resultado técnico mas relevante sobre os valores singulares de matrizes antissimétricas.

Teorema 10.34 . *Seja $A \in \mathrm{Mat}(\mathbb{R},\ m)$ uma matriz antissimétrica, e sejam $(\alpha_1,\ \ldots,\ \alpha_r)$, onde $1 \leq r \leq m$, e $\alpha_k \in [0,\ \infty)$ para todo k, seus valores singulares* distintos. *Então, cada valor singular não nulo possui degenerescência par. No caso de m ser par, a degenerescência do valor singular nulo (se o houver) é também par. No caso de m ser ímpar, a degenerescência do valor singular nulo é ímpar.* □

Prova. Seja $A \in \mathrm{Mat}(\mathbb{R},\ m)$ uma matriz antissimétrica real. A matriz A^TA, é não negativa, autoadjunta e real. Os chamados valores singulares de A são os autovalores da matriz $\sqrt{A^TA}$, os quais são, naturalmente, não negativos. Vamos denotar por os valores singulares ***distintos*** de A por $(\alpha_1,\ \ldots,\ \alpha_r)$, onde $1 \leq r \leq m$, e $\alpha_k \in [0,\ \infty)$ para todo k. Vamos denotar por $\epsilon_k \in \mathbb{N}$ a multiplicidade do valor singular α_k. Naturalmente $\sum_{k=1}^{r} \epsilon_k = m$.

Convencionamos que os q' primeiros sejam não nulos e os demais, se os houver, são nulos: $\alpha_k > 0$ para $1 \leq k \leq q'$ e $\alpha_k = 0$ para $k > q'$.

Vamos denotar por $(\beta_1,\ \ldots,\ \beta_m)$ a lista de valores singulares de A, *incluindo degenerescência*. Ou seja, cada α_k, com $1 \leq k \leq r$, comparece nessa lista repetido ϵ_k-vezes.

Convencionamos também que os q primeiros elementos da lista $(\beta_1,\ \ldots,\ \beta_m)$ sejam não nulos e os demais, se os houver, são nulos: $\beta_k > 0$ para $1 \leq k \leq q$ mas $\beta_k = 0$ para $k > q$. Está claro que $q = \sum_{k=1}^{q'} \epsilon_k$.

Vamos denotar por $u_1,\ \ldots,\ u_m$ autovetores normalizados de A^TA correspondentes à lista $(\beta_1,\ \ldots,\ \beta_m)$. Escolhemos esses autovetores compondo uma base ortonormal (o que é sempre possível por A^TA ser autoadjunta). Assim,

$$A^TAu_j \;=\; \beta_j^2 u_j, \qquad j \;=\; 1,\ \ldots,\ m\;.$$

Eles podem ser escolhidos reais, pois A^TA é real e autoadjunta. Pela convenção os q primeiros correspondem a autovalores não nulos de A^TA.

Definimos agora novos q vetores por

$$v_j := \frac{1}{\beta_j} A u_j , \qquad j = 1, \ldots, q .$$

Como escolhemos os u_j's reais, os vetores v_j são também reais, já que A é uma matriz real e β_j é real.

Temos que $A^T A u_j = \beta_j^2 u_j$, $j = 1, \ldots, q$, e, com a definição acima,

$$A^T v_j = \frac{1}{\beta_j} A^T A u_j = \beta_j u_j , \qquad j = 1, \ldots, q . \tag{10.161}$$

A relação $A^T v_j = \beta_j u_j$ implica

$$A A^T v_j = \beta_j A u_j = \beta_j^2 v_j , \qquad j = 1, \ldots, q . \tag{10.162}$$

Assim, temos

$$v_j = \frac{1}{\beta_j} A u_j \qquad \text{e} \qquad u_j = \frac{1}{\beta_j} A^T v_j \overset{\text{antissimetria}}{=} \frac{-1}{\beta_j} A v_j , \qquad j = 1, \ldots, q . \tag{10.163}$$

A primeira é uma definição e a segunda foi obtida em (10.161).

Os vetores v_j, $j = 1, \ldots, q$, são também ortonormais:

$$\langle v_i, v_j \rangle_{\mathbb{C}} = \frac{1}{\beta_i \beta_j} \langle A u_i, A u_j \rangle_{\mathbb{C}} = \frac{1}{\beta_i \beta_j} \langle u_i, A^T A u_j \rangle_{\mathbb{C}} = \frac{\beta_j}{\beta_i} \langle u_i, u_j \rangle_{\mathbb{C}} = \frac{\beta_j}{\beta_i} \delta_{i,j} = \delta_{i,j} , \tag{10.164}$$

$i, j \in \{1, \ldots, q\}$.

Para os vetores u_j com $j = q+1, \ldots, m$, que são autovetores de $A^T A$ com autovalor nulo (se os houver), temos $\|A u_j\|_{\mathbb{C}}^2 = \langle A u_j, A u_j \rangle_{\mathbb{C}} = \langle u_j, A^T A u_j \rangle_{\mathbb{C}} = 0$. Analogamente, para os vetores v_j com $j = q+1, \ldots, m$, que são autovetores de $A A^T$ com autovalor nulo (se os houver), temos $\|A^T v_j\|_{\mathbb{C}}^2 = \langle A^T v_j, A^T v_j \rangle_{\mathbb{C}} = \langle v_j, A A^T v_j \rangle_{\mathbb{C}} = 0$. Como $A = -A^T$, concluímos que

$$A u_j = A^T u_j = A v_j = A^T v_j \quad \text{para todos } j = q+1, \ldots, m . \tag{10.165}$$

A relação (10.162) nos ensinou que v_j, com $j = 1, \ldots, q$, são autovetores de $A A^T$ com autovalores não nulos. Como $A A^T$ é autoadjunto, o complemento ortogonal do subespaço gerado por $v_1, \ldots, v_q$ é o núcleo de $A A^T$. Assim, podemos completar esses vetores $v_1, \ldots, v_q$ com um conjunto adicional de vetores ortonormais $v_{q+1}, \ldots, v_m$ que são autovetores de $A A^T$ com autovalores nulos.

As bases ortonormais $\{u_1, \ldots, u_m\}$ e $\{v_1, \ldots, v_m\}$ são denominadas *bases singulares à esquerda e à direita*, respectivamente.

É importante observar aqui que as relações (10.163) implicam $\langle v_k, u_k \rangle_{\mathbb{C}} = 0$ para todo $k \in \{1, \ldots, q\}$. De fato,

$$\langle v_k, u_k \rangle_{\mathbb{C}} = \frac{-1}{\beta_k} \langle v_k, A v_k \rangle_{\mathbb{C}} = -\frac{1}{\beta_k} \langle A^T v_k, v_k \rangle_{\mathbb{C}} = -\langle u_j, v_k \rangle_{\mathbb{C}} = -\langle v_k, u_k \rangle_{\mathbb{C}} , \tag{10.166}$$

o que estabelece que $\langle v_k, u_k \rangle_{\mathbb{C}} = 0$ para todo $k \in \{1, \ldots, q\}$, como desejado. Acima, usamos que u_j e v_k são vetores reais e, portanto, $\langle u_k, v_k \rangle_{\mathbb{C}} = \overline{\langle v_k, u_k \rangle_{\mathbb{C}}} = \langle v_k, u_k \rangle_{\mathbb{C}}$.

Um outro ponto importante a se observar é que, devido à antissimetria de A, as matrizes $A^T A$ e $A A^T$ são iguais, pois ambas valem $-A^2$. Disso segue automaticamente que, caso $\beta_j \neq \beta_k$, para algum par $j, k \in \{1, \ldots, q\}$, então $\langle v_j, u_k \rangle_{\mathbb{C}} = 0$, pois nesse caso ambos os vetores v_j e u_k seriam autovetores da mesma matriz autoadjunta $-A^2$ com autovalores distintos. De fato,

$$\begin{aligned} \beta_j^2 \langle v_j, u_k \rangle_{\mathbb{C}} &= \Big\langle \big(-A^2\big) v_j, u_k \Big\rangle_{\mathbb{C}} = \Big\langle v_j, \big(-A^2\big) u_k \Big\rangle_{\mathbb{C}} \\ &= \beta_k^2 \langle v_j, u_k \rangle_{\mathbb{C}} \quad \therefore \quad \big(\beta_j^2 - \beta_k^2\big) \langle v_j, u_k \rangle_{\mathbb{C}} = 0 \implies \langle v_j, u_k \rangle_{\mathbb{C}} = 0 . \end{aligned} \tag{10.167}$$

Porém, mesmo no caso em que $\beta_{j_0} = \beta_{k_0}$ para algum par j_0, $k_0 \in \{1, \ldots, q\}$, podemos escolher os vetores $\langle v_j, u_k\rangle_{\mathbb{C}} = 0$,

Como $\langle v_k, u_k\rangle_{\mathbb{C}} = 0$ para todo $k \in \{1, \ldots, q\}$, vemos que cada autovalor de $A^T A = AA^T$ é ao menos duplamente degenerado, pois u_k e v_k são autovetores com o mesmo autovalor β_k e são linearmente independentes, pois $\langle v_k, u_k\rangle_{\mathbb{C}} = 0$.

Afirmamos que a degenerescência de cada autovalor é sempre par. Para provar isso é conveniente fazermos uma escolha específica dos vetores $u_1, \ldots, u_m$ que compõem uma base de autovetores ortonormais de $A^T A = AA^T$.

Vamos supor que um dado autovalor não nulo β_p de $A^T A = AA^T$ tenha uma degenerescência maior que dois e que u_p e v_p sejam dois correspondentes autovetores. Então, como é bem sabido, podemos escolher um vetor, que listamos como u_{p+1}, também autovetor de $A^T A = AA^T$ com autovalor β_p e de sorte que u_{p+1} seja ortogonal a u_p e a v_p e de norma 1. Afirmamos que o correspondente vetor $v_{p+1} := \frac{1}{\beta_{p+1}} Au_{p+1} = \frac{1}{\beta_p} Au_{p+1}$ é igualmente ortogonal a u_p, a v_p e a u_{p+1}. A ortogonalidade entre v_{p+1} e v_p já foi estabelecida em (10.164), enquanto que a ortogonalidade entre v_{p+1} e u_{p+1} foi provada em (10.166). Para provar que v_{p+1} e u_p são ortogonais, observemos que

$$\langle v_{p+1}, u_p\rangle_{\mathbb{C}} = \frac{-1}{\beta_p}\langle v_{p+1}, Av_p\rangle_{\mathbb{C}} = -\frac{1}{\beta_p}\langle A^T v_{p+1}, v_p\rangle_{\mathbb{C}} \stackrel{\beta_p=\beta_{p+1}}{=} -\langle u_{p+1}, v_p\rangle_{\mathbb{C}} = 0 , \tag{10.168}$$

pois u_{p+1} foi escolhido ortogonal a v_p e a u_p. Assim, vemos que u_p, v_p, u_{p+1} e v_{p+1} são mutuamente ortogonais, e, portanto, linearmente independentes, além de serem todos autovalores de $A^T A = AA^T$ com o mesmo autovalor[36].

Esse proceder pode ser estendido a degenerescências maiores, indicando que a cada autovalor de $A^T A = AA^T$ existem autovetores $u_{p_1}, \ldots, u_{p_h}$ e $v_{p_1}, \ldots, v_{p_h}$ para algum $h \in \mathbb{N}$, todos mutuamente ortogonais e normalizados a 1.

Concluímos disso que os vetores $u_1, \ldots, u_q$ todos são ortogonais entre si e ortogonais aos vetores $v_1, \ldots, v_q$, que também são ortogonais entre si (todos têm norma 1). Segue disso também que $2q \leq m$.

A degenerescência do autovalor nulo é exatamente $m - 2q$. Esse valor é sempre par se m for par e sempre ímpar se m for ímpar. No caso de m ser par e $m = 2q$, não ocorre o autovalor nulo, o que não pode se dar caso m seja ímpar. ■

• Lista canônica de valores singulares de uma matriz antissimétrica

Vamos agora introduzir a para nós relevante noção de *lista canônica de valores singulares* de uma matriz antissimétrica.

Consideremos primeiramente o caso de matrizes antissimétricas de ordem par.

Seja $A \in \text{Mat}(\mathbb{R}, m)$ uma matriz antissimétrica de ordem par m e sejam $(\alpha_1, \ldots, \alpha_r)$, para algum $r \in \{1, \ldots, m\}$, os valores singulares **distintos** de A. Pelo Teorema 10.34, página 555, cada α_k tem degenerescência par (inclusive o valor singular nulo, se ocorrer, pois $2n$ é par). Vamos escrever como $\epsilon_k = 2\delta_k$ a degenerescência do valor singular α_k, sendo $\delta_k \in \mathbb{N}$. Tem-se, portanto, que $r \leq m/2$, pois $\sum_{k=1}^{r} 2\delta_k = m$ e cada δ_k é maior ou igual a 1.

Consideremos a lista $(\sigma_1, \ldots, \sigma_{m/2})$ construída repetindo-se cada valor singular distinto α_k um número δ_k de vezes (a metade de sua degenerescência). Com isso a lista dos valores singulares de A é da forma

$$(\sigma_1, \ldots, \sigma_m) = (\sigma_1, \ldots, \sigma_{m/2}, \sigma_1, \ldots, \sigma_{m/2}) ,$$

sendo que cada cada valor singular distinto α_k comparece $2\delta_k$ vezes, sua devida multiplicidade. Note-se que alguns σ_k's podem ser nulos.

A lista $(\sigma_1, \ldots, \sigma_{m/2})$ será denominada aqui a *lista canônica dos valores singulares* de A.

Exemplo 10.3 Se α_1 e α_2 são os valores singulares distintos de uma matriz antissimétrica A, de ordem $m = 6$, com degenerescências $2\delta_1 = 2$ e $2\delta_2 = 4$, respectivamente, então a lista canônica dos valores singulares de A é $(\sigma_1, \sigma_3, \sigma_3) = (\alpha_1, \alpha_2, \alpha_2)$ e a lista dos seis valores singulares de A, repetindo degenerescências, é $(\sigma_1, \sigma_3, \sigma_3, \sigma_4, \sigma_5, \sigma_6) = (\alpha_1, \alpha_2, \alpha_2, \alpha_1, \alpha_2, \alpha_2)$. ⌑

Consideremos agora o caso de matrizes antissimétricas de ordem ímpar. Seja $A \in \text{Mat}(\mathbb{R}, m)$ uma matriz antissimétrica de ordem ímpar m, e sejam $(\alpha_1, \ldots, \alpha_r)$, para algum $r \in \{1, \ldots, m\}$, os valores singulares **distintos** de A. Pelo Teorema 10.34, página 555, cada α_k não nulo tem degenerescência par, mas o autovalor nulo tem degenerescência ímpar. Seja $m - 2q$ a degenerescência do autovalor nulo.

[36]Decorre de (10.162) que v_{p+1} tem o mesmo autovalor de u_{p+1}

Consideremos a lista $(\sigma_1, \ldots, \sigma_{(m-1)/2})$ construída repetindo-se cada valor singular não nulo distinto α_k um número δ_k de vezes (a metade de sua degenerescência) e o autovalor nulo um número $(m-1)/2 - q$ vezes. Com isso, a lista dos valores singulares de A é da forma

$$(\sigma_1, \ldots, \sigma_m) = (\sigma_1, \ldots, \sigma_{(m-1)/2}, \sigma_1, \ldots, \sigma_{(m-1)/2}, 0) ,$$

sendo que cada cada valor singular não nulo distinto α_k comparece $2\delta_k$ vezes e o autovalor nulo $m-2q$ vezes, suas devidas multiplicidades.

Também nesse caso a lista $(\sigma_1, \ldots, \sigma_{(m-1)/2})$ é denominada lista canônica dos valores singulares de A.

É importante notas que tanto no caso par quanto no ímpar a lista $(\sigma_1, \ldots, \sigma_m)$ difere na lista $(\beta_1, \ldots, \beta_m)$ apenas por uma permutação. Ambas contêm os autovalores não nulos degenerados um número par de vezes, os demais sendo os autovalores nulos.

Por uma questão de simplicidade notacional, vamos denotar os autovetores do operador $A^T A = AA^T$ associados a cada σ_k por u_k e v_k.

• Enunciado e demonstração do teorema de diagonalização por blocos de matrizes antissimétricas reais

Passemos agora ao enunciado e à demonstração do resultado principal, o qual é um corolário do Teorema 10.34, página 555.

Seja $A \in \mathrm{Mat}(\mathbb{R}, m)$ uma matriz antissimétrica, ou seja, tal que $A^T = -A$. Afirmamos que A pode ser levada por uma transformações de similaridade produzidas por certas matrizes ortogonais a certas formas em blocos. Mais especificamente, afirmamos:

Teorema 10.35 *Seja $A \in \mathrm{Mat}(\mathbb{R}, m)$ uma matriz antissimétrica e seja $m' := \lfloor m/2 \rfloor$. Seja $(\sigma_1, \ldots, \sigma_{m'})$ a lista canônica dos valores singulares de A.*

Então, existe uma matriz ortogonal $P \in \mathrm{Mat}(\mathbb{R}, m)$ tal que

$$P^T AP = \begin{pmatrix} \mathbb{0}_{m'} & D \\ -D & \mathbb{0}_{m'} \end{pmatrix} \quad \textit{caso } m \textit{ seja par, ou} \quad \begin{pmatrix} \mathbb{0}_{m'} & D & 0 \\ -D & \mathbb{0}_{m'} & \vdots \\ 0 & \ldots & 0 \end{pmatrix} \quad \textit{caso } m \textit{ seja ímpar,} \tag{10.169}$$

onde D é em ambos os casos a matriz diagonal $m' \times m'$

$$D = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_{m'} \end{pmatrix} ,$$

e onde $\mathbb{0}_{m'}$ é a matriz $m' \times m'$ identicamente nula, sendo que σ_j são os valores singulares de A, ou seja, são os autovalores de $\sqrt{A^T A}$ (alguns dos quais podem ser nulos).

Alternativamente, existe matriz ortogonal Q tal que

$$Q^T A Q \quad = \quad \begin{pmatrix} \boxed{A_1} & & & & \\ & \ddots & & & \\ & & \ddots & & \\ & & & \boxed{A_{m'}} & \\ & & & & \boxed{\mathbb{0}_{m-2m'}} \end{pmatrix} ,$$

onde cada A_k, $k = 1, \ldots, m'$, é uma matriz antissimétrica 2×2: $A_k = \left(\begin{smallmatrix} 0 & \sigma_k \\ -\sigma_k & 0 \end{smallmatrix}\right)$, onde novamente σ_k são os valores singulares de A (isto é, os autovalores de $\sqrt{A^T A}$) e $\mathbb{0}_{m-2m'}$ é a matriz $(m-2m') \times (m-2m'))$ identicamente nula. Todos os demais elementos de matriz fora dos blocos são nulos. □

Prova. Seja $q \leq m'$ o número de elementos não nulos na lista canônica $(\sigma_1, \ldots, \sigma_{m'})$. Recordando que $2q \leq m$, definamos $m' = \lfloor m/2 \rfloor$. É claro que $q \leq m'$.

Vamos considerar os dois conjunto $(u_1, \ldots, u_m)$ e $(v_1, \ldots, v_m)$ de autovetores que introduzimos na prova do Teorema 10.34, página 555, mas reordenados de forma que seus autovalores sejam os σ_k's. Mais especificamente,

$$\begin{aligned} A^T A u_j &= \sigma_j u_j, \qquad j = 1, \ldots, q , \\ A^T A v_j &= \sigma_j v_j, \qquad j = 1, \ldots, q , \end{aligned}$$

e os demais tendo autovalor zero.

A ideia é considerarmos bases ortonormais constituídas pelos $2q$ vetores $u_1, \ldots, u_q, v_1, \ldots, v_q$ e mais, digamos, os $m-2q$ vetores $u_{q+1}, \ldots, u_m$, que são autovetores de autovalor nulo de $A^T A = AA^T$ e são ortogonais aos $2q$ anteriormente listados. Como visto na prova do Teorema 10.34, página 555,

Consideremos uma nova base ortonormal

$$\mathbf{f} := (f_1, \ldots, f_m) \equiv \Big(u_1, \ldots, u_{m'}, v_1, \ldots, v_{m'}, u_{m'+1}, \ldots, u_m\Big),$$

ou seja, com

$$f_j = u_j \delta_{1, \ldots, m'}(j) + v_{j-m'+1} \delta_{m'+1, \ldots, 2m'}(j) + u_j \delta_{2m'+1, \ldots, m}(j), \qquad j = 1, \ldots, m$$

onde,

$$\delta_{a_1, \ldots, a_l}(j) := \begin{cases} 1, & \text{se } j \in \{a_1, \ldots, a_l\} , \\ 0, & \text{de outra forma.} \end{cases}$$

A questão agora é saber como serão os elementos de matriz de A essa base. Para isso, observe-se que para $j, k \in \{1, \ldots, q\}$,

$$\begin{aligned} \langle u_j, Au_k \rangle_{\mathbb{C}} &= \sigma_k \langle u_j, v_k \rangle_{\mathbb{C}} = 0 , \\ \langle u_j, Av_k \rangle_{\mathbb{C}} &= -\sigma_k \langle u_j, u_k \rangle_{\mathbb{C}} = -\sigma_k \delta_{j,k} , \\ \langle v_j, Au_k \rangle_{\mathbb{C}} &= \sigma_k \langle v_j, v_k \rangle_{\mathbb{C}} = -\sigma_k \delta_{j,k} , \\ \langle v_j, Av_k \rangle_{\mathbb{C}} &= -\sigma_k \langle v_j, u_k \rangle_{\mathbb{C}} = 0 . \end{aligned} \tag{10.170}$$

Devido a (10.165), $\langle u_j,\ Au_k\rangle_{\mathbb{C}} = \langle v_j,\ Au_k\rangle_{\mathbb{C}} = 0$ para todo j se k for maior que q.

Dessa forma, segundo (10.170),

$$\begin{aligned}\langle f_a,\ Af_b\rangle_{\mathbb{C}} &= \langle u_a,\ Av_{b-m'+1}\rangle_{\mathbb{C}}\delta_{1,\ \ldots,\ m'}(a)\delta_{m'+1,\ \ldots,\ 2m'}(b) + \langle v_{a-m'+1},\ Au_b\rangle_{\mathbb{C}}\delta_{m'+1,\ \ldots,\ 2m'}(a)\delta_{1,\ \ldots,\ m'}(b)\\ &= \sigma_a\delta_{a,b-m'+1}\delta_{1,\ \ldots,\ m'}(a)\delta_{m'+1,\ \ldots,\ 2m'}(b) - \sigma_b\delta_{a-m'+1,b}\delta_{m'+1,\ \ldots,\ 2m'}(a)\delta_{1,\ \ldots,\ m'}(b)\\ &= \sigma_a\delta_{a,b-m'+1}\delta_{1,\ \ldots,\ m'}(a)\delta_{m'+1,\ \ldots,\ 2m'}(a+m'-1) - \sigma_b\delta_{a-m'+1,b}\delta_{m'+1,\ \ldots,\ 2m'}(m'+b-1)\delta_{1,\ \ldots,\ m'}(b)\ ,\end{aligned} \tag{10.171}$$

Agora,

$$\delta_{1,\ \ldots,\ m'}(a)\delta_{m'+1,\ \ldots,\ 2m'}(a+m'-1) = \delta_{1,\ \ldots,\ m'}(a)\delta_{2,\ \ldots,\ m'+1}(a) = \delta_{2,\ \ldots,\ m'}(a)$$

e

$$\delta_{m'+1,\ \ldots,\ 2m'}(m'+b-1)\delta_{1,\ \ldots,\ m'}(b) = \delta_{2,\ \ldots,\ m'+1}(b)\delta_{1,\ \ldots,\ m'}(b) = \delta_{2,\ \ldots,\ m'}(b)\ ,$$

e assim estabelece-se que

$$\langle f_a,\ Af_b\rangle_{\mathbb{C}} = = \sigma_a\delta_{b,a+m'-1}\delta_{2,\ \ldots,\ m'}(a) - \sigma_b\delta_{a,b+m'-1}\delta_{2,\ \ldots,\ m'}(b)\ .$$

Observe-se que os valores singulares σ_j com $j > q$ (sendo $q \leq m'$) são todos nulos, se os houver.

Disso vê-se facilmente que na base $\mathbf{f}$ a matriz antissimétrica A assume a forma em blocos

$$\begin{pmatrix} \mathbb{0}_{m'} & D \\ -D & \mathbb{0}_{m'} \end{pmatrix} \quad \text{caso } m \text{ seja par, ou} \quad \begin{pmatrix} \mathbb{0}_{m'} & D & 0 \\ -D & \mathbb{0}_{m'} & \vdots \\ 0 & \ldots & 0 \end{pmatrix} \quad \text{caso } m \text{ seja ímpar,} \tag{10.172}$$

onde D é em ambos os casos a matriz diagonal $m' \times m'$

$$D = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_{m'} \end{pmatrix} .$$

Observe-se que os elementos diagonais σ_j com $j > q$ (sendo $q \leq m'$) são todos nulos, se os houver.

A mudança de base da base canônica à base ortonormal $\mathbf{f}$ é realizada por uma matriz ortogonal P. portanto, estabelecemos que P^TAP assume a uma das formas de (10.172), dependendo de m ser par ou ímpar.

Se considerarmos alternativamente a base ortonormal

$$\mathbf{g} := (g_1,\ \ldots,\ g_m) \equiv \Big(u_1,\ v_1,\ u_2,\ v_2,\ \ldots,\ u_{m'}\ v_{m'},\ u_{m'+1},\ \ldots,\ u_m\Big),$$

é fácil ver, seguindo os mesmos passos acima, que A assume a forma de blocos ao longo da diagonal principal

$$\begin{pmatrix} \boxed{\begin{matrix} 0 & \sigma_1 \\ -\sigma_1 & 0 \end{matrix}} & & & & \\ & \ddots & & & \\ & & \ddots & & \\ & & & \boxed{\begin{matrix} 0 & \sigma_{m'} \\ -\sigma_{m'} & 0 \end{matrix}} & \\ & & & & \boxed{\mathbb{0}_{m-2m'}} \end{pmatrix} . \tag{10.173}$$

Observe-se que a base **g** é obtida da base **f** por permutações. Assim, a passagem de uma a outra é também implementada por uma matriz ortogonal (que implementa as permutações). Com isso e das observações anteriores, concluímos que existe uma matriz ortogonal Q tal que $Q^T AQ$ assume a forma de blocos (10.173). Isso completa a demonstração. ■

A Seção 10.8.6, página 561, apresenta um relevante corolário do Teorema 10.35, denominado Teorema de Williamson.

10.8.6 O Teorema de Williamson

Para $n \in \mathbb{N}$, seja J_{2n} a matriz antissimétrica definida por

$$J_{2n} := \begin{pmatrix} \mathbb{0}_n & \mathbb{1}_n \\ -\mathbb{1}_n & \mathbb{0}_n \end{pmatrix} .$$

É fácil ver que $J_{2n}^T = -J_{2n} = J_{2n}^{-1}$.

Uma matriz real de ordem $2n$, $S \in \mathrm{Mat}(\mathbb{R},\ 2n)$, é dita ser simplética real se satisfazer $S^T J_{2n} S = J_{2n}$. O conjunto de todas as matrizes simpléticas reais de ordem $2n$ compõe um grupo, denominado *grupo simplético real* e denotado por $\mathrm{Sp}(\mathbb{R},\ 2n)$. Esse grupo é apresentado e discutido nas Seções 21.3.3.3 e 21.9, páginas 1058 e 1150, respectivamente. Pode ser visto facilmente (vide Seção 21.3.3.3) que se $S \in \mathrm{Sp}(\mathbb{R},\ 2n)$, então $S^T \in \mathrm{Sp}(\mathbb{R},\ 2n)$.

Matrizes do grupo simplético podem ser utilizadas para transformar matrizes reais simétricas e positivas de ordem par em uma forma diagonal específica através de uma transformação de congruência (não necessariamente de similaridade!). Esse resultado, conhecido como *Teorema de Williamson*[37], é utilizado na Mecânica Quântica e na Geometria Simplética (*i.e.*, na Mecânica Clássica). Vide *e.g.*, [145].

• **O Teorema de Williamson**

O principal resultado que desejamos provar na corrente seção é o seguinte:

Teorema 10.36 (Teorema de Williamson) *Para $m \in \mathbb{N}$, seja $M \in \mathrm{Mat}(\mathbb{R},\ 2m)$ uma matriz real simétrica positiva. Então, existe uma matriz simplética $S \in \mathrm{Sp}(\mathbb{R},\ 2m)$ e uma matriz $m \times m$ diagonal $D \in \mathrm{Mat}(\mathbb{R},\ m)$ tais que*

$$M = S^T \begin{pmatrix} D & 0 \\ 0 & D \end{pmatrix} S . \tag{10.174}$$

Acima, D é a matriz diagonal $D = \mathrm{diag}\{\sigma_1,\ \ldots,\ \sigma_m\}$, onde $(\sigma_1,\ \ldots,\ \sigma_m)$ é a lista canônica dos valores singulares *(definida à página 557) da matriz antissimétrica de ordem par $A := M^{1/2} J_{2n} M^{1/2}$.*

Assim, toda matriz matriz simétrica positiva de ordem par pode ser diagonalizada por uma transformação de congruência gerada por uma matriz do grupo simplético.

As quantidades $(\sigma_1,\ \ldots,\ \sigma_m)$ são frequentemente denominados autovalores simpléticos *de M.* □

Demonstração. Para $A,\ B \in \mathrm{Mat}(\mathbb{R},\ m)$, denotaremos a matriz $\left(\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right) \in \mathrm{Mat}(\mathbb{R},\ 2m)$ por $A \oplus B$.

Como $M \in \mathrm{Mat}(\mathbb{R},\ 2m)$ é simétrica e positiva, possui uma inversa e também uma raiz quadrada também simétricas e positivas (vide Corolário 10.4, página 513, e definição que se lhe segue). Denotemos essa raiz quadrada por $M^{1/2} \equiv \sqrt{M} \in \mathrm{Mat}(\mathbb{R},\ 2m)$.

Pelos Teoremas 10.35 e 10.34, páginas 558 e 555, respectivamente, exite uma matriz ortogonal $P \in \mathrm{Mat}(\mathbb{R},\ 2m)$ tal

[37]John Williamson (1901–1949). O artigo original é: John Williamson, "On the Algebraic Problem Concerning the Normal Forms of Linear Dynamical Systems". American Journal of Mathematics Vol. **58**, N. 1, 141–163 (1936). doi:10.2307/2371062.

que

$$P^T M^{1/2} J_{2n} M^{1/2} P \;=\; \begin{pmatrix} \mathbb{0}_m & D \\ -D & \mathbb{0}_m \end{pmatrix} , \tag{10.175}$$

onde D é uma matriz diagonal $m \times m$ com $D = \operatorname{diag}\big(\sigma_m, \ldots, \sigma_m\big)$, onde os $(\sigma_1, \ldots, \sigma_m)$ compõe a chamada lista canônica dos valores singulares de $M^{1/2} J_{2n} M^{1/2}$. Vide definição à página 557.

Seja agora

$$S \;:=\; M^{-1/2} P \Big(D^{1/2} \oplus D^{1/2}\Big) .$$

Então,

$$S^T M S \;=\; \Big(D^{1/2} \oplus D^{1/2}\Big) P^T M^{-1/2} M M^{-1/2} P \Big(D^{1/2} \oplus D^{1/2}\Big) \;=\; \Big(D^{1/2} \oplus D^{1/2}\Big)^2 \;=\; D \oplus D .$$

Agora,

$$S J_{2m} S^T \;=\; M^{-1/2} P \Big(D^{1/2} \oplus D^{1/2}\Big) J_{2m} \Big(D^{1/2} \oplus D^{1/2}\Big) P^T M^{-1/2} .$$

Por um lado,

$$\Big(D^{1/2} \oplus D^{1/2}\Big) J_{2m} \Big(D^{1/2} \oplus D^{1/2}\Big) \;=\; \begin{pmatrix} D^{1/2} & \mathbb{0}_m \\ \mathbb{0}_m & D^{1/2} \end{pmatrix} \begin{pmatrix} \mathbb{0}_m & \mathbb{1}_m \\ -\mathbb{1}_m & \mathbb{0}_m \end{pmatrix} \begin{pmatrix} D^{1/2} & \mathbb{0}_m \\ \mathbb{0}_m & D^{1/2} \end{pmatrix} \;=\; \begin{pmatrix} \mathbb{0}_m & D \\ -D & \mathbb{0}_m \end{pmatrix}$$

e, assim,

$$S J_{2m} S^T \;=\; M^{-1/2} \left[P \begin{pmatrix} \mathbb{0}_m & D \\ -D & \mathbb{0}_m \end{pmatrix} P^T \right] M^{-1/2} \;\stackrel{(10.175)}{=}\; M^{-1/2} \Big[M^{1/2} J_{2m} M^{1/2} \Big] M^{-1/2} \;=\; J_{2m} ,$$

estabelecendo que S^T é simplética, ou seja, é um elemento do grupo $\mathrm{Sp}(\mathbb{R}, 2m)$, e, portanto, que S também o é (vide (21.51), página 1059). ■

10.9 A Pseudoinversa de Moore-Penrose

Na presente seção introduziremos uma generalização especial da noção de inversa de matrizes, a qual aplica-se mesmo a matrizes não quadradas. O conceito que descreveremos, a chamada pseudoinversa de Moore-Penrose, é particularmente útil no tratamento de problemas de optimização linear, como discutiremos adiante (Seção 10.9.2, página 571), ou seja, em problemas onde procura-se soluções optimalmente aproximadas de sistemas de equações lineares como $Ax = y$, onde A é uma matriz $m \times n$ dada, y um vetor-coluna, dado, com m componentes e x, a incógnita do problema, é um vetor-coluna com n componentes. Em tais problemas procura-se vetores x tais que a norma de $Ax - y$ seja a menor possível e que representem, portanto, não necessariamente a solução exata do sistema $Ax = y$ (que pode não existir), mas a melhor aproximação em termos de "mínimos quadrados" ao que seria a solução.

• **Inversas generalizadas, ou pseudoinversas**

Sejam $m, n \in \mathbb{N}$ e seja uma matriz (não necessariamente quadrada) $A \in \mathrm{Mat}(\mathbb{C}, m, n)$. Uma matriz $B \in \mathrm{Mat}(\mathbb{C}, n, m)$ é dita ser uma *inversa generalizada*, ou *pseudoinversa*, de A, se satisfizer as seguintes condições:

1. $ABA = A$,
2. $BAB = B$.

O leitor há de notar que se $A \in \mathrm{Mat}(\mathbb{C},\ n)$ é uma matriz quadrada inversível, sua inversa A^{-1} satisfaz trivialmente as propriedades definidoras da inversa generalizada. Provaremos mais adiante que toda matriz $A \in \mathrm{Mat}(\mathbb{C},\ m\ , n)$ possui ao menos uma inversa generalizada, a saber, a pseudoinversa de Moore-Penrose. Com a generalidade da definição acima, porém, não se pode garantir a unicidade da inversa generalizada de A.

Com a amplitude da definição acima, a noção inversa generalizada não é muito útil, mas certos tipos mais específicos de inversas generalizadas são de interesse em certos tipos de problemas. No que segue discutiremos a chamada pseudoinversa de Moore-Penrose e seu emprego em problemas de optimização linear.

• Definição da pseudoinversa de Moore-Penrose de uma matriz

Sejam $m,\ n \in \mathbb{N}$ e seja uma matriz (não necessariamente quadrada) $A \in \mathrm{Mat}(\mathbb{C},\ m\ , n)$. Uma matriz $A^+ \in \mathrm{Mat}(\mathbb{C},\ n,\ m)$ é dita ser uma *pseudoinversa de Moore-Penrose* de A se satisfizer as seguintes condições:

1. $AA^+A = A$,
2. $A^+AA^+ = A^+$,
3. $AA^+ \in \mathrm{Mat}(\mathbb{C},\ m)$ e $A^+A \in \mathrm{Mat}(\mathbb{C},\ n)$ são autoadjuntas.

O leitor há de notar que se $A \in \mathrm{Mat}(\mathbb{C},\ n)$ é uma matriz quadrada inversível, sua inversa A^{-1} satisfaz trivialmente as propriedades definidoras da pseudoinversa de Moore-Penrose.

A noção de pseudoinversa descrita acima foi introduzida por E. H. Moore[38] em 1920 e redescoberta por R. Penrose[39] em 1955. O conceito de pseudoinversa de Moore-Penrose é útil para a resolução de problemas de optimização lineares, ou seja, à determinação da melhor aproximação em termos de "mínimos quadrados" à solução de sistemas lineares. Trataremos desses aspectos mais adiante (vide Teorema 10.39, página 571), após demonstrarmos resultados sobre existência e unicidade. Outros desenvolvimentos da teoria das pseudoinversas de Moore-Penrose e suas aplicações, podem ser encontrados em [43]. Vide também as referências originais: E. H. Moore, *"On the reciprocal of the general algebraic matrix"*. Bulletin of the American Mathematical Society **26**, 394–395 (1920); R. Penrose, *"A generalized inverse for matrices"*, Proceedings of the Cambridge Philosophical Society **51**, 406–413 (1955) e R. Penrose, *"On best approximate solution of linear matrix equations"*, Proceedings of the Cambridge Philosophical Society **52**, 17–19 (1956).

Nas páginas que seguem demonstraremos que toda a matriz $A \in \mathrm{Mat}(\mathbb{C},\ m,\ n)$ possui uma pseudoinversa de Moore-Penrose, a qual é <u>única</u>. Começamos com a questão da unicidade para em seguida tratarmos de propriedades gerais e, posteriormente, da questão da existência. As aplicações em problemas de optimização são discutidas na Seção 10.9.2, página 571.

• A unicidade da pseudoinversa de Moore-Penrose

Demonstremos a unicidade da pseudoinversa de Moore-Penrose de uma matriz $A \in \mathrm{Mat}(\mathbb{C},\ m,\ n)$, caso exista.

Seja $A^+ \in \mathrm{Mat}(\mathbb{C},\ n,\ m)$ uma pseudoinversa de Moore-Penrose de $A \in \mathrm{Mat}(\mathbb{C},\ m,\ n)$ e seja $B \in \mathrm{Mat}(\mathbb{C},\ n,\ m)$ uma outra pseudoinversa de Moore-Penrose de A, ou seja, tal que $ABA = A$, $BAB = B$ com AB e BA autoadjuntas. Seja $M_1 := AB - AA^+ = A(B - A^+) \in \mathrm{Mat}(\mathbb{C},\ m)$. Pelas hipóteses, M_1 é autoadjunta (por ser a diferença de duas matrizes autoadjuntas) e $(M_1)^2 = (AB - AA^+)A(B - A^+) = (ABA - AA^+A)(B - A^+) = (A - A)(B - A^+) = 0$. Como M_1 é autoadjunta, o fato que $(M_1)^2 = 0$ implica $M_1 = 0$, pois para todo $x \in \mathbb{C}^m$ tem-se $\|M_1x\|^2_{\mathbb{C}} = \langle M_1x,\ M_1x\rangle_{\mathbb{C}} = \langle x,\ (M_1)^2x\rangle_{\mathbb{C}} = 0$, o que significa que $M_1 = 0$. Isso provou que $AB = AA^+$. Analogamente, prova-se que $BA = A^+A$ (para tal, considere-se a matriz autoadjunta $M_2 := BA - A^+A \in \mathrm{Mat}(\mathbb{C},\ n)$ e proceda-se como acima). Agora, tudo isso implica que $A^+ = A^+AA^+ = A^+(AA^+) = A^+AB = (A^+A)B = BAB = B$, provando a unicidade.

Como já comentamos, se $A \in \mathrm{Mat}(\mathbb{C},\ n)$ é uma matriz quadrada inversível, sua inversa A^{-1} satisfaz trivialmente as propriedades definidoras da pseudoinversa de Moore-Penrose e, portanto, tem-se nesse caso $A^+ = A^{-1}$, univocamente. É também evidente pela definição que para $\mathbb{0}_{mn}$, a matriz $m \times n$ identicamente nula, vale $(\mathbb{0}_{mn})^+ = \mathbb{0}_{nm}$.

• A existência da pseudoinversa de Moore-Penrose

Apresentaremos no que seguirá duas demonstrações da existência da pseudoinversa de Moore-Penrose de matrizes arbitrárias de $\mathrm{Mat}(\mathbb{C},\ m,\ n)$. Ambas as demonstrações permitem produzir algoritmos para a determinação explícita

[38] Eliakim Hastings Moore (1862–1932).
[39] Sir Roger Penrose (1931–).

da pseudoinversa de Moore-Penrose. Uma primeira demonstração será apresentada na Seção 10.9.1.1, página 567, (vide o Teorema 10.37, página 568, e o Teorema 10.38, página 570) e decorrerá de diversos resultados que estabeleceremos a seguir. Destacamos particularmente as expressões (10.197) e (10.198), as quais permitem calcular a pseudoinversa de Moore-Penrose A^+ de uma matriz $A \in \mathrm{Mat}(\mathbb{C},\ m,\ n)$ diretamente em termos de A, A^* e dos autovalores de AA^* ou de A^*A (ou seja, dos valores singulares de A).

Uma segunda demonstração será apresentada na Seção 10.9.3, página 572, e para a mesma faremos uso da decomposição em valores singulares apresentada no Teorema 10.26, página 542. A essa segunda demonstração da Seção 10.9.3 o leitor interessado poderá passar sem perdas neste ponto. Os resultados da Seção 10.9.3, porém, não serão usados no que segue. Essa segunda demonstração é a mais frequentemente apresentada na literatura, mas cremos que as expressões (10.197) e (10.198), adiante, forneçam um método algoritmicamente mais simples para a determinação da pseudoinversa de Moore-Penrose de uma matriz geral.

• Calculando a pseudoinversa de Moore-Penrose em casos particulares

Se $A \in \mathrm{Mat}(\mathbb{C},\ m,\ n)$, então $A^* \in \mathrm{Mat}(\mathbb{C},\ n,\ m)$ é definida como a matriz cujos elementos $(A^*)_{ij}$ são dados por $\overline{A_{ji}}$ para todos $0 \leq i \leq n$ e $0 \leq j \leq m$. Futuramente obteremos as expressões (10.197) e (10.198), as quais permitem calcular a pseudoinversa de Moore-Penrose $A^+ \in \mathrm{Mat}(\mathbb{C},\ n,\ m)$ de uma matriz $A \in \mathrm{Mat}(\mathbb{C},\ m,\ n)$ diretamente em termos de A, A^* e dos autovalores de AA^* ou de A^*A. Nos exercícios que seguem indicaremos situações especiais mas úteis nas quais a pseudoinversa de Moore-Penrose pode ser calculada de modo relativamente simples.

E. 10.43 *Exercício*. Constate que se $A \in \mathrm{Mat}(\mathbb{C},\ m,\ 1)$, $A = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}$, um vetor-coluna não-nulo, então $A^+ = \frac{1}{\|A\|_{\mathbb{C}}^2} A^* = \frac{1}{\|A\|_{\mathbb{C}}^2} \left(\overline{a_1}\,,\ \ldots,\ \overline{a_m} \right)$, onde $\|A\|_{\mathbb{C}} = \sqrt{|a_1|^2 + \cdots + |a_m|^2}$. ✠

Observe-se que se $z \in \mathbb{C}$, podemos considerar z como uma matriz complexa 1×1, ou seja, como elemento de $\mathrm{Mat}(\mathbb{C},\ 1,\ 1)$ e, com isso, obtemos do exposto acima $(z)^+ = \begin{cases} 0, & z = 0 \\ \frac{1}{z}, & z \neq 0 \end{cases}$.

O resultado do Exercício E. 10.43 pode ser generalizado.

E. 10.44 *Exercício*. Seja $A \in \mathrm{Mat}(\mathbb{C},\ m,\ n)$. Mostre que se $(AA^*)^{-1}$ existe, então $A^+ = A^*(AA^*)^{-1}$. Mostre que se $(A^*A)^{-1}$ existe, então $A^+ = (A^*A)^{-1}A^*$. *Sugestão:* em ambos os casos, verifique que o lado direito satisfaz as propriedades definidoras da pseudoinversa de Moore-Penrose e use a unicidade. ✠

Os resultados do Exercício E. 10.44 podem ser generalizados para situações em que AA^* ou A^*A não são inversíveis pois, como veremos na Proposição 10.37, página 565 valem sempre as relações $A^+ = A^*\big(AA^*\big)^+ = \big(A^*A\big)^+A^*$. Também o Teorema 10.37, página 568, apresentará uma generalização dos resultados do Exercício E. 10.44, mostrando uma outra forma de proceder quando AA^* ou A^*A não forem inversíveis.

Os exercícios que seguem contêm aplicações dos resultados do Exercício E. 10.44.

E. 10.45 *Exercício*. Seja $A = \left(\begin{smallmatrix} 2 & 0 & i \\ 0 & i & 1 \end{smallmatrix}\right)$, com $A^* = \left(\begin{smallmatrix} 2 & 0 \\ 0 & -i \\ -i & 1 \end{smallmatrix}\right)$. Mostre que AA^* possui inversa, mas que A^*A não possui. Usando o Exercício E. 10.44, calcule a pseudoinversa de Moore-Penrose A^+ de A, obtendo $A^+ = \frac{1}{9}\left(\begin{smallmatrix} 4 & -2i \\ 1 & -5i \\ -i & 4 \end{smallmatrix}\right)$. Verifique que essa A^+ satisfaz de fato as propriedades definidoras da pseudoinversa de Moore-Penrose. ✠

E. 10.46 *Exercício*. Seja $A = \left(\begin{smallmatrix} 1 & 2 \\ 0 & i \\ 0 & 3 \end{smallmatrix}\right)$, com $A^* = \left(\begin{smallmatrix} 1 & 0 & 0 \\ 2 & -i & 3 \end{smallmatrix}\right)$. Mostre que AA^* não possui inversa, mas que A^*A possui. Usando o Exercício E. 10.44, calcule a pseudoinversa de Moore-Penrose A^+ de A, obtendo $A^+ = \frac{1}{10}\left(\begin{smallmatrix} 10 & 2i & -6 \\ 0 & -i & 3 \end{smallmatrix}\right)$. Verifique que essa A^+ satisfaz de fato as propriedades definidoras da pseudoinversa de Moore-Penrose. ✠

10.9.1 Outras Propriedades da Pseudoinversa de Moore-Penrose

As seguintes propriedades da pseudoinversa de Moore-Penrose seguem das definições e da unicidade. Suas demonstrações são elementares e são deixadas como exercício: para $A \in \mathrm{Mat}(\mathbb{C}, m, n)$ valem

1. $(A^+)^+ = A$,
2. $(A^+)^T = (A^T)^+$, $\overline{A^+} = (\overline{A})^+$ e, consequentemente, $(A^+)^* = (A^*)^+$,
3. $(zA)^+ = z^{-1}A^+$ para todo $z \in \mathbb{C}$ não-nulo.

É de se observar, porém, que se $A \in \mathrm{Mat}(\mathbb{C}, m, n)$ e $B \in \mathrm{Mat}(\mathbb{C}, n, p)$, nem sempre $(AB)^+$ é dada por B^+A^+, ao contrário do que ocorre com a inversa usual (para o caso $m = n = p$). Uma exceção relevante será encontrada na Proposição 10.37, página 565.

A seguinte proposição lista mais algumas propriedades importantes, algumas das quais usaremos logo adiante:

Proposição 10.36 *A pseudoinversa de Moore-Penrose satisfaz as seguintes relações*

$$A^+ = A^+ (A^+)^* A^* , \qquad (10.176)$$

$$A = A A^* (A^+)^* , \qquad (10.177)$$

$$A^* = A^* A A^+ , \qquad (10.178)$$

$$A^+ = A^* (A^+)^* A^+ , \qquad (10.179)$$

$$A = (A^+)^* A^* A , \qquad (10.180)$$

$$A^* = A^+ A A^* , \qquad (10.181)$$

válidas para toda $A \in \mathrm{Mat}(\mathbb{C}, m, n)$. □

Das relações acima, a mais relevante talvez seja a relação (10.178), pois faremos uso importante dela na demonstração da Proposição 10.39, página 571, que trata da aplicação da pseudoinversa de Moore-Penrose a problemas de optimização linear.

Prova da Proposição 10.36. Por AA^+ ser autoadjunta, vale $AA^+ = (AA^+)^* = (A^+)^*A^*$. Multiplicando-se à esquerda por A^+ obtemos $A^+ = A^+(A^+)^*A^*$, provando (10.176). Substituindo-se $A \to A^+$ e usando o fato que $A = (A^+)^+$, obtém-se de (10.176) que $A = AA^*(A^+)^*$, que é a relação (10.177). Substituindo-se $A \to A^*$ e usando o fato que $(A^*)^+ = (A^+)^*$, obtém-se de (10.177) que $A^* = A^*AA^+$ que é a relação (10.178).

As relações (10.179)–(10.181) podem ser obtidas analogamente a partir do fato de A^+A ser também autoadjunta, mas é mais fácil obtê-las substituindo-se $A \to A^*$ em (10.176)–(10.178) e tomando-se o adjunto das expressões resultantes. ■

Da Proposição 10.36 podem ser obtidos vários resultados de interesse, alguns dos quais encontram-se reunidos na proposição que segue.

Proposição 10.37 *Para a pseudoinversa de Moore-Penrose vale*

$$\big(AA^*\big)^+ = \big(A^*\big)^+ A^+ \qquad (10.182)$$

para todo $A \in \mathrm{Mat}(\mathbb{C}, m, n)$. *Disso obtém-se que*

$$A^+ = A^*\big(AA^*\big)^+ = \big(A^*A\big)^+ A^* , \qquad (10.183)$$

também para todo $A \in \mathrm{Mat}(\mathbb{C}, m, n)$. □

A expressão (10.183) generaliza os resultados do Exercício E. 10.44, página 564 e pode ser empregada para calcular A^+ desde que $\big(AA^*\big)^+$ ou $\big(A^*A\big)^+$ sejam previamente conhecidas.

Prova da Proposição 10.37. Seja $B = \big(A^*\big)^+ A^+$. Tem-se

$$AA^* \stackrel{(10.177)}{=} A\,A^*\,(A^+)^*\,A^* \stackrel{(10.181)}{=} A\,A^*\,(A^+)^*\,A^+\,A\,A^* \;=\; (AA^*)B(AA^*)\;,$$

onde usamos também que $\big(A^*\big)^+ = \big(A^+\big)^*$. Tem-se também que

$$B \;=\; \big(A^*\big)^+ A^+ \stackrel{(10.176)}{=} (A^+)^*\,A^+\,A\,A^+ \stackrel{(10.179)}{=} (A^+)^*\,A^+\,A\,A^*\,(A^+)^*\,A^+ \;=\; B\big(A\,A^*\big)B\;.$$

Observe-se também que

$$\big(A\,A^*\big)B \;=\; \Big(A\,A^*(A^+)^*\Big)A^+ \stackrel{(10.178)}{=} AA^+$$

que é autoadjunto, por definição. Analogamente,

$$B\big(A\,A^*\big) \;=\; (A^+)^*\Big(A^+A\,A^*\Big) \stackrel{(10.180)}{=} (A^*)^+A^*$$

que também é autoadjunto, por definição. Os fatos expostos nas linhas acima provaram que B é a pseudoinversa de Moore-Penrose de AA^*, provando (10.182). Substituindo-se $A \to A^*$ em (10.182) obtém-se também

$$\big(A^*A\big)^+ \;=\; A^+\big(A^*\big)^+\;. \tag{10.184}$$

Observe-se agora que

$$A^*\big(AA^*\big)^+ \stackrel{(10.182)}{=} A^*\big(A^*\big)^+A^+ \stackrel{(10.179)}{=} A^+$$

e que

$$\big(A^*A\big)^+A^* \stackrel{(10.184)}{=} A^+\big(A^*\big)^+A^* \stackrel{(10.176)}{=} A^+\;,$$

provando (10.183). ■

• A pseudoinversa de Moore-Penrose, o núcleo e a imagem de uma matriz

Definimos o núcleo e a imagem (“range”) de uma matriz $A \in \mathrm{Mat}\,(\mathbb{C},\ m,\ n)$ por $\mathrm{Ker}\,(A) := \{u \in \mathbb{C}^n |\ Au = 0\}$ e $\mathrm{Ran}\,(A) := \{Au,\ u \in \mathbb{C}^n\}$, respectivamente. É evidente que $\mathrm{Ker}\,(A)$ é um subespaço linear de $\mathbb{C}^n$ e que $\mathrm{Ran}\,(A)$ é um subespaço linear de $\mathbb{C}^m$.

A seguinte proposição será usada logo adiante, mas é de interesse por si só.

Proposição 10.38 *Seja $A \in \mathrm{Mat}\,(\mathbb{C},\ m,\ n)$ e sejam definidos $P_1 := \mathbb{1}_n - A^+A \in \mathrm{Mat}\,(\mathbb{C},\ n)$ e $P_2 := \mathbb{1}_m - AA^+ \in \mathrm{Mat}\,(\mathbb{C},\ n)$. Então, valem as seguintes afirmações:*

1. *P_1 e P_2 são projetores ortogonais, ou seja, satisfazem $(P_k)^2 = P_k$ e $P_k^* = P_k$, $k = 1,\ 2$.*
2. *$\mathrm{Ker}\,(A) = \mathrm{Ran}\,(P_1)$, $\mathrm{Ran}\,(A) = \mathrm{Ker}\,(P_2)$,
$\mathrm{Ker}\,(A^+) = \mathrm{Ran}\,(P_2)$ e $\mathrm{Ran}\,(A^+) = \mathrm{Ker}\,(P_1)$.*
3. *$\mathrm{Ran}\,(A) = \mathrm{Ker}\,(A^+)^\perp$ e $\mathrm{Ran}\,(A^+) = \mathrm{Ker}\,(A)^\perp$.*
4. *$\mathrm{Ker}\,(A) \oplus \mathrm{Ran}\,(A^+) = \mathbb{C}^n$ e $\mathrm{Ker}\,(A^+) \oplus \mathrm{Ran}\,(A) = \mathbb{C}^m$, ambas somas diretas de subespaços ortogonais.* □

Prova. Que P_1 e P_2 são autoadjuntos segue do fato de AA^+ e A^+A o serem. Tem-se também que $(P_1)^2 = \mathbb{1} - 2A^+A + A^+AA^+A = \mathbb{1} - 2A^+A + A^+A = \mathbb{1} - A^+A = P_1$ e analogamente para P_2. Isso provou o item 1.

Seja $x \in \mathrm{Ker}\,(A)$. Como $\mathrm{Ran}\,(P_1)$ é um subespaço linear fechado de $\mathbb{C}^n$, o Teorema do Melhor Aproximante e o Teorema da Decomposição Ortogonal (que neste texto são apresentados com toda generalidade – no contexto de espaços de Hilbert, como $\mathbb{C}^m$ – na forma do Teorema 38.1, página 1980, e do Teorema 38.2, página 1982, respectivamente) garantem-nos a existência de um <u>único</u> $z_0 \in \mathrm{Ran}\,(P_1)$ tal que $\|x - z_0\|_{\mathbb{C}^m}$ é mínimo. Além disso, $x - z_0$ é ortogonal a

$\mathrm{Ran}\,(P_1)$. Assim, existe ao menos um $y_0 \in \mathbb{C}^m$ tal que $x - P_1 y_0$ é ortogonal a todo elemento da forma $P_1 y$, ou seja, $\langle x - P_1 y_0,\ P_1 y\rangle_{\mathbb{C}} = 0$ para todo $y \in \mathbb{C}^m$, o que implica $\langle P_1(x - P_1 y_0),\ y\rangle_{\mathbb{C}} = 0$ para todo $y \in \mathbb{C}^m$, o que por sua vez implica $P_1(x - P_1 y_0) = 0$. Isso, porém, afirma que $P_1 x = P_1 y_0$. Como $x \in \mathrm{Ker}\,(A)$ vale $P_1 x = x$ (pela definição de P_1). Provamos portanto que se $x \in \mathrm{Ker}\,(A)$ então $x \in \mathrm{Ran}\,(P_1)$, estabelecendo que $\mathrm{Ker}\,(A) \subset \mathrm{Ran}\,(P_1)$. Por outro lado, o fato que $AP_1 = A(\mathbb{1} - A^+A) = A - A = 0$ implica que $\mathrm{Ran}\,(P_1) \subset \mathrm{Ker}\,(A)$, provando que $\mathrm{Ran}\,(P_1) = \mathrm{Ker}\,(A)$.

Se $z \in \mathrm{Ker}\,(P_1)$, então $z = A^+Az$, provando que $z \in \mathrm{Ran}\,(A^+)$. Isso provou que $\mathrm{Ker}\,(P_1) \subset \mathrm{Ran}\,(A^+)$. Por outro lado, se $u \in \mathrm{Ran}\,(A^+)$ então existe $v \in \mathbb{C}^m$ tal que $u = A^+v$. Logo, $P_1 u = (\mathbb{1}_n - A^+A)A^+v = (A^+ - A^+AA^+)v = 0$, provando que $u \in \mathrm{Ker}\,(P_1)$ e que $\mathrm{Ran}\,(A^+) \subset \mathrm{Ker}\,(P_1)$. Isso estabeleceu que $\mathrm{Ker}\,(P_1) = \mathrm{Ran}\,(A^+)$.

P_2 é obtida de P_1 com a substituição $A \to A^+$ (lembrando-se que $(A^+)^+ = A$). Logo, os resultados acima implicam que $\mathrm{Ran}\,(P_2) = \mathrm{Ker}\,(A^+)$ e que $\mathrm{Ker}\,(P_2) = \mathrm{Ran}\,(A)$. Isso provou o item 2.

Se $M \in \mathrm{Mat}\,(\mathbb{C},\ p)$ (com $p \in \mathbb{N}$, arbitrário) é autoadjunta, então $\langle y,\ Mx\rangle_{\mathbb{C}} = \langle My,\ x\rangle_{\mathbb{C}}$ para todos $x,\ y \in \mathbb{C}^p$. Essa relação torna evidente que $\mathrm{Ker}\,(M) = \mathrm{Ran}\,(M)^{\perp}$ (justifique!). Com isso o item 3 segue do item 2 tomando-se $M = P_1$ e $M = P_2$. O item 4 é evidente pelo item 3. ■

E. 10.47 *Exercício*. Calcule P_1 e P_2 para o exemplo do Exercício E. 10.45, página 564, e para o exemplo do Exercício E. 10.46, página 564. ✠

10.9.1.1 A Regularização de Tikhonov. Existência

No Exercício E. 10.44, página 564, vimos que se $(AA^*)^{-1}$ existe, então $A^+ = A^*(AA^*)^{-1}$ e que se $(A^*A)^{-1}$ existe, então $A^+ = (A^*A)^{-1}A^*$. No caso de essas inversas não existirem há um procedimento alternativo que também permite obter A^+. Sabemos da Proposição 10.6, página 474, que mesmo se $(AA^*)^{-1}$ não existir, a matriz $AA^* + \mu\mathbb{1}$ será invertível para todo $\mu \in \mathbb{C}$ não-nulo com $|\mu|$ pequeno o suficiente. Isso permite conjecturar que as expressões $A^*(AA^* + \mu\mathbb{1})^{-1}$ e $(A^*A + \mu\mathbb{1})^{-1}A^*$, que estão bem definidas para $\mu \neq 0$ com $|\mu|$ pequeno, convergem a A^+ quando tomamos o limite $\mu \to 0$. Como veremos no que segue, essa conjectura é correta.

Pelo dito acima, podemos substituir as matrizes AA^* ou A^*A, caso sejam singulares, pelas matrizes inversíveis $AA^* + \mu\mathbb{1}$ ou $A^*A + \mu\mathbb{1}$ com $\mu \neq 0$ com $|\mu|$ pequeno. Esse procedimento de regularização (que envolve a substituição provisória de uma expressão singular por outra regular) é denominado *regularização de Tikhonov*[40], em honra ao matemático que desenvolveu essas ideias no contexto de equações integrais[41].

Nosso primeiro resultado consiste em provar que os limites descritos acima de fato existem e são iguais, o que será feito nos dois lemas que seguem.

Lema 10.11 *Seja $A \in \mathrm{Mat}\,(\mathbb{C},\ m,\ n)$ e seja $\mu \in \mathbb{C}$ tal que $AA^* + \mu\mathbb{1}_m$ e $A^*A + \mu\mathbb{1}_n$ sejam inversíveis (i.e., $\mu \notin \sigma(AA^*) \cup \sigma(A^*A)$, um conjunto finito). Então, $A^*(AA^* + \mu\mathbb{1}_m)^{-1} = (A^*A + \mu\mathbb{1}_n)^{-1}A^*$.* □

Prova. Sejam $B_\mu := A^*(AA^* + \mu\mathbb{1}_m)^{-1}$ e $C_\mu := (A^*A + \mu\mathbb{1}_n)^{-1}A^*$. Temos que

$$A^*AB_\mu = A^*\big[AA^*\big](AA^*+\mu\mathbb{1}_m)^{-1} = A^*\big[AA^*+\mu\mathbb{1}_m-\mu\mathbb{1}_m\big](AA^*+\mu\mathbb{1}_m)^{-1} = A^*\big(\mathbb{1}_m-\mu(AA^*+\mu\mathbb{1}_m)^{-1}\big) = A^*-\mu B_\mu\,.$$

Logo, $(A^*A + \mu\mathbb{1}_n)B_\mu = A^*$, o que implica $B_\mu = (A^*A + \mu\mathbb{1}_n)^{-1}A^* = C_\mu$. ■

Lema 10.12 *Para toda $A \in \mathrm{Mat}\,(\mathbb{C},\ m,\ n)$ os limites $\lim\limits_{\mu\to 0} A^*(AA^* + \mu\mathbb{1}_m)^{-1}$ e $\lim\limits_{\mu\to 0}(A^*A + \mu\mathbb{1}_n)^{-1}A^*$ existem e são iguais (pelo Lema 10.11), definindo um elemento de $\mathrm{Mat}\,(\mathbb{C},\ n,\ m)$.* □

[40]Andrei Nikolaevich Tikhonov (1906–1993). O sobrenome russo "Tikhonov" é por vezes transliterado como "Tykhonov", "Tichonov" ou ainda "Tychonoff".

[41]Para uma referência geral, vide [395]. Para os trabalhos originais, vide: Tikhonov, A. N., 1943, "On the stability of inverse problems", Dokl. Akad. Nauk. USSR, **39**, No. 5, 195–198 (1943); Tikhonov, A. N., "Solution of incorrectly formulated problems and the regularization method", Soviet Math. Dokl. **4**, 1035–1038 (1963), tradução para o inglês de Dokl. Akad. Nauk. USSR **151**, 501–504 (1963).

Prova do Lema 10.12. Notemos primeiramente que A é uma matriz identicamente nula se e somente se AA^* ou A^*A o forem. De fato, se, por exemplo, $A^*A = 0$, valerá para todo vetor x que $0 = \langle x, \ A^*Ax\rangle_{\mathbb{C}} = \langle Ax, \ Ax\rangle_{\mathbb{C}} = \|Ax\|^2$, provando que $A = 0$. Como a afirmação a ser provada é evidente se A for nula, suporemos no que segue que AA^* e A^*A não são nulas.

A matriz $AA^* \in \mathrm{Mat}(\mathbb{C}, \ m)$ é, evidentemente, autoadjunta. Sejam $\alpha_1, \ \ldots, \ \alpha_r$ seus autovalores distintos. Pelo Teorema Espectral para operadores autoadjuntos (vide Teorema 10.7, página 494 e Teorema 10.15, página 510) podemos escrever

$$AA^* \ = \ \sum_{a=1}^{r} \alpha_a E_a \ , \tag{10.185}$$

onde E_a são os projetores espectrais de AA^* e satisfazem $E_a E_b = \delta_{ab} E_a$, $E_a^* = E_a$ e $\sum_{a=1}^{r} E_a = \mathbb{1}_m$. Logo,

$$AA^* + \mu\mathbb{1}_m \ = \ \sum_{a=1}^{r} (\alpha_a + \mu) E_a$$

e, portanto, para $\mu \notin \{\alpha_1, \ \ldots, \ \alpha_r\}$, vale pela Proposição 10.17, página 497,

$$\big(AA^* + \mu\mathbb{1}_m\big)^{-1} \ = \ \sum_{a=1}^{r} \frac{1}{\alpha_a + \mu} E_a \qquad \text{e} \qquad A^*\big(AA^* + \mu\mathbb{1}_m\big)^{-1} \ = \ \sum_{a=1}^{r} \frac{1}{\alpha_a + \mu} A^* E_a \ .$$

Há dois casos a se considerar *1.* AA^* não tem autovalor nulo e *2.* AA^* tem autovalor nulo.

No caso em que AA^* não tem autovalor nulo é claro pela última expressão que o limite $\lim\limits_{\mu\to 0} A^*(AA^* + \mu\mathbb{1}_m)^{-1}$ existe e vale

$$\lim_{\mu\to 0} A^*(AA^* + \mu\mathbb{1}_m)^{-1} \ = \ \sum_{a=1}^{r} \frac{1}{\alpha_a} A^* E_a \ . \tag{10.186}$$

No caso em que AA^* tem autovalor nulo, digamos, $\alpha_1 = 0$, o projetor E_1 projeta sobre o núcleo de AA^*: $\mathrm{Ker}\,(AA^*) := \{u \in \mathbb{C}^n | \ AA^* u = 0\}$. Se $x \in \mathrm{Ker}\,(AA^*)$, então $A^*x = 0$, pois $0 = \langle x, \ AA^*x\rangle_{\mathbb{C}} = \langle A^*x, \ A^*x\rangle_{\mathbb{C}} = \|A^*x\|$. Portanto,

$$A^* E_1 \ = \ 0 \tag{10.187}$$

e, assim, podemos escrever,

$$A^*(AA^* + \mu\mathbb{1}_m)^{-1} \ = \ \sum_{a=2}^{r} \frac{1}{\alpha_a + \mu} A^* E_a \ ,$$

donde obtém-se

$$\lim_{\mu\to 0} A^*(AA^* + \mu\mathbb{1}_m)^{-1} \ = \ \sum_{a=2}^{r} \frac{1}{\alpha_a} A^* E_a \ . \tag{10.188}$$

Isso provou que $\lim\limits_{\mu\to 0} A^*(AA^* + \mu\mathbb{1}_m)^{-1}$ sempre existe.

Pelo Lema 10.11, página 567, o limite $\lim\limits_{\mu\to 0} (A^*A + \mu\mathbb{1}_n)^{-1} A^*$ também existe e coincide com $\lim\limits_{\mu\to 0} A^*(AA^* + \mu\mathbb{1}_m)^{-1}$. ■

A principal consequência é o seguinte resultado:

Teorema 10.37 (Regularização de Tikhonov) *Para toda* $A \in \mathrm{Mat}(\mathbb{C}, \ m, \ n)$ *valem*

$$A^+ \ = \ \lim_{\mu\to 0} A^*\big(AA^* + \mu\mathbb{1}_m\big)^{-1} \tag{10.189}$$

e

$$A^+ \ = \ \lim_{\mu\to 0} \big(A^*A + \mu\mathbb{1}_n\big)^{-1} A^* \ . \tag{10.190}$$

□

Como a existência dos limites acima foi estabelecida para matrizes arbitrárias no Lema 10.12, página 567, o Teorema 10.37 contém uma prova geral de existência da pseudoinversa de Moore-Penrose.

Prova do Teorema 10.37. As afirmações a serem provadas são evidentes caso $A = \mathbb{0}_{mn}$ pois, como já vimos $(\mathbb{0}_{mn})^+ = \mathbb{0}_{nm}$. Assim, assumiremos no que segue que A é não nula, o que equivale, pelo exposto no início da prova do Lema 10.12, a supor que AA^* e A^*A não são nulas.

Pelos Lemas 10.11 e 10.12 é suficiente demonstrar (10.189). Há dois casos a se considerar *1.* AA^* não tem autovalor nulo e *2.* AA^* tem autovalor nulo. No caso *1.*, vimos em (10.186), na prova do Lema 10.12 (e com a notação lá estabelecida), que

$$\lim_{\mu\to 0} A^*\big(AA^* + \mu\mathbb{1}_m\big)^{-1} \;=\; \sum_{a=1}^{r}\frac{1}{\alpha_a}A^*E_a \;=:\; B\,.$$

Note-se agora que

$$AB \;=\; \sum_{a=1}^{r}\frac{1}{\alpha_a}AA^*E_a \;=\; \sum_{a=1}^{r}\frac{1}{\alpha_a}\left(\sum_{b=1}^{r}\alpha_b E_b\right)E_a \;=\; \sum_{a=1}^{r}\sum_{b=1}^{r}\frac{1}{\alpha_a}\alpha_b\,\delta_{ab}E_a \;=\; \sum_{a=1}^{r}E_a \;=\; \mathbb{1}_m\,, \tag{10.191}$$

que é autoadjunta, e que

$$BA \;=\; \sum_{a=1}^{r}\frac{1}{\alpha_a}A^*E_aA\,, \tag{10.192}$$

que é também autoadjunta, pois $\alpha_a \in \mathbb{R}$ para todo a (por serem autovalores de uma matriz autoadjunta) e pelo fato de $(A^*E_aA)^* = A^*E_aA$ para todo a, já que $E_a^* = E_a$.

De (10.191) segue que $ABA = A$. De (10.192) segue que

$$BAB \;=\; \left(\sum_{a=1}^{r}\frac{1}{\alpha_a}A^*E_aA\right)\left(\sum_{b=1}^{r}\frac{1}{\alpha_b}A^*E_b\right) \;=\; \sum_{a=1}^{r}\sum_{b=1}^{r}\frac{1}{\alpha_a\alpha_b}A^*E_a(AA^*)E_b\,.$$

Agora, pela decomposição espectral (10.185) de AA^*, segue que $(AA^*)E_b = \alpha_bE_b$. Logo,

$$BAB \;=\; \sum_{a=1}^{r}\sum_{b=1}^{r}\frac{1}{\alpha_a}A^*E_aE_b \;=\; \left(\sum_{a=1}^{r}\frac{1}{\alpha_a}A^*E_a\right)\underbrace{\left(\sum_{b=1}^{r}E_b\right)}_{\mathbb{1}_m} \;=\; B\,.$$

Isso provou que $A = A^+$ no caso em que AA^* não tem autovalor nulo.

Vamos agora supor que AA^* não autovalor nulo, a saber, α_1. Vimos em (10.188), na prova do Lema 10.12, que

$$\lim_{\mu\to 0} A^*\big(AA^* + \mu\mathbb{1}_m\big)^{-1} \;=\; \sum_{a=2}^{r}\frac{1}{\alpha_a}A^*E_a \;=:\; B\,.$$

Usando o fato que $(AA^*)E_a = \alpha_aE_a$, o qual segue da decomposição espectral (10.185) de AA^*, obtém-se

$$AB \;=\; \sum_{a=2}^{r}\frac{1}{\alpha_a}AA^*E_a \;=\; \sum_{a=2}^{r}\frac{1}{\alpha_a}\alpha_aE_a \;=\; \sum_{a=2}^{r}E_a \;=\; \mathbb{1}_m - E_1\,, \tag{10.193}$$

que é autoadjunta, pois E_1 o é. Tem-se também

$$BA \;=\; \sum_{a=2}^{r}\frac{1}{\alpha_a}A^*E_aA\,, \tag{10.194}$$

que é também autoadjunta, pelos argumentos já expostos.

De (10.193) segue que $ABA = A - E_1A$. Note-se agora que $(E_1A)^* = A^*E_1 = 0$, por (10.187). Isso demonstrou que $E_1A = 0$ e que $ABA = A$. De (10.194) segue que

$$BAB \;=\; \left(\sum_{a=2}^{r}\frac{1}{\alpha_a}A^*E_aA\right)\left(\sum_{b=2}^{r}\frac{1}{\alpha_b}A^*E_b\right) \;=\; \sum_{a=2}^{r}\sum_{b=2}^{r}\frac{1}{\alpha_a\alpha_b}A^*E_a(AA^*)E_b\,.$$

Usando novamente que $(AA^*)E_b = \alpha_b E_b$, obtemos

$$BAB = \sum_{a=2}^{r}\sum_{b=2}^{r}\frac{1}{\alpha_a}A^*E_aE_b = \left(\sum_{a=2}^{r}\frac{1}{\alpha_a}A^*E_a\right)\underbrace{\left(\sum_{b=2}^{r}E_b\right)}_{\mathbb{1}_m - E_1} = B - \sum_{a=2}^{r}\frac{1}{\alpha_a}A^*E_aE_1 = B ,$$

pois $E_aE_1 = 0$ para $a \neq 1$. Isso demonstrou que $BAB = B$. Assim, estabelecemos que $A = A^+$ também no caso em que AA^* tem autovalor nulo, completando a prova de (10.189). ■

10.9.1.2 A Pseudoinversa de Moore-Penrose e o Teorema Espectral

Durante a demonstração do Teorema 10.37 estabelecemos também o seguinte resultado de interesse:

Teorema 10.38 *Seja $A \in \mathrm{Mat}(\mathbb{C},\ m,\ n)$ não-nula e seja $AA^* = \sum_{a=1}^{r}\alpha_aE_a$ a representação espectral de AA^*, onde $\{\alpha_1,\ \ldots,\ \alpha_r\} \subset \mathbb{R}$ é o conjunto dos autovalores distintos de AA^* e E_a são os correspondentes projetores espectrais autoadjuntos. Então, vale*

$$A^+ = \sum_{\substack{a=1\\ \alpha_a\neq 0}}^{r}\frac{1}{\alpha_a}A^*E_a . \tag{10.195}$$

Analogamente, seja $A^*A = \sum_{b=1}^{s}\beta_bF_b$ a representação espectral de A^*A, onde $\{\beta_1,\ \ldots,\ \beta_s\} \subset \mathbb{R}$ é o conjunto dos autovalores distintos de A^*A e F_b os correspondentes projetores espectrais autoadjuntos. Então, vale também

$$A^+ = \sum_{\substack{b=1\\ \beta_b\neq 0}}^{s}\frac{1}{\beta_b}F_bA^* . \tag{10.196}$$

(Vale mencionar aqui que, pelo Exercício E. 10.6, página 475, o conjunto de autovalores não nulos de AA^* coincide com o conjunto de autovalores não nulos de A^*A: $\{\alpha_1,\ \ldots,\ \alpha_r\}\setminus\{0\} = \{\beta_1,\ \ldots,\ \beta_s\}\setminus\{0\}$).

De (10.195) e (10.196) segue que para A não-nula valem

$$A^+ = \sum_{\substack{a=1\\ \alpha_a\neq 0}}^{r}\frac{1}{\alpha_a\prod\limits_{\substack{l=1\\ l\neq a}}^{r}(\alpha_a-\alpha_l)}\ A^*\left[\prod_{\substack{l=1\\ l\neq a}}^{r}\Big(AA^* - \alpha_l\mathbb{1}_m\Big)\right] , \tag{10.197}$$

$$A^+ = \sum_{\substack{b=1\\ \beta_b\neq 0}}^{s}\frac{1}{\beta_b\prod\limits_{\substack{l=1\\ l\neq b}}^{s}(\beta_b-\beta_l)}\left[\prod_{\substack{l=1\\ l\neq b}}^{s}\Big(A^*A - \beta_l\mathbb{1}_n\Big)\right]A^* . \tag{10.198}$$

□

As expressões (10.197) e (10.198) fornecem mais um algoritmo geral para o cômputo da pseudoinversa de Moore-Penrose, o qual pode ser de implementação simples, pois requer apenas a determinação dos autovalores de AA^* ou de A^*A.

Prova do Teorema 10.38. A igualdade (10.195) foi provada durante a demonstração do Teorema 10.37 (vide (10.186) e (10.188)). A relação (10.196) pode ser provada analogamente, mas segue mais facilmente do truque já mencionado de usar (10.195), trocando $A \to A^*$ e tomando-se o adjunto da expressão obtida. As relações (10.197) e (10.198) seguem da Proposição 10.18, página 497, particularmente de (10.60). ■

E. 10.48 *Exercício*. Usando (10.197) ou (10.198) reobtenha as matrizes A^+ dos Exercícios E. 10.43, E. 10.45 e E. 10.46. ✠

10.9.2 A Pseudoinversa de Moore-Penrose e Problemas de Optimização Linear

Tratemos agora de uma das principais aplicações da noção de pseudoinversa de Moore-Penrose, a saber, no tratamento de problemas de optimização linear, que motivaremos e definiremos a seguir.

Sejam $A \in \mathrm{Mat}(\mathbb{C},\ m,\ n)$ e $y \in \mathbb{C}^m$ dados e considere-se o problema de determinar $x \in \mathbb{C}^n$ que satisfaça a equação linear

$$Ax \ = \ y \ . \tag{10.199}$$

No caso em que $m = n$ e A tem inversa, a solução (única) é, naturalmente, $x = A^{-1}y$. Nos demais casos uma solução pode não estar presente ou não ser única. Podemos considerar o problema alternativo de saber para quais $x' \in \mathbb{C}^n$ a norma Euclidiana $\|Ax' - y\|_{\mathbb{C}^m}$ é a menor possível. Tais vetores $x' \in \mathbb{C}^n$ seriam, no sentido da norma Euclidiana $\|\cdot\|_{\mathbb{C}^m}$, ou seja, em termos de "mínimos quadrados", os melhores aproximantes ao que seria a solução de (10.199). Um tal problema é por vezes dito ser um *problema de optimização linear*. Esse problema pode ser tratado com o uso da noção de pseudoinversa de Moore-Penrose, a qual permite caracterizar precisamente o conjunto dos vetores x' que minimizam $\|Ax' - y\|_{\mathbb{C}^m}$. A isso dedicaremos as linhas que seguem, sendo o principal resultado condensado no seguinte teorema:

Teorema 10.39 (Optimização Linear) *Sejam $A \in \mathrm{Mat}(\mathbb{C},\ m,\ n)$ e $y \in \mathbb{C}^m$ dados. Então, a coleção de todos vetores de $\mathbb{C}^n$ para os quais a aplicação $\mathbb{C}^n \ni x \mapsto \|Ax - y\|_{\mathbb{C}^m} \in [0,\ \infty)$ assume um mínimo absoluto coincide com o conjunto*

$$A^+y + \mathrm{Ker}\,(A) = \Big\{A^+y + \big(\mathbb{1}_n - A^+A\big)z,\ z \in \mathbb{C}^n\Big\} \ . \tag{10.200}$$

Esse conjunto é dito ser o conjunto minimizante *do problema de optimização linear em questão. É interessante observar que pela Proposição 10.38, página 566, tem-se também $A^+y + \mathrm{Ker}\,(A) = A^+y + \mathrm{Ran}\,(A^+)^\perp$.* □

Como se vê do enunciado acima, a pseudoinversa de Moore-Penrose fornece a melhor aproximação em termos de "mínimos quadrados" à solução de sistemas lineares. Observe-se que para os elementos x do conjunto minimizante (10.200) vale $\|Ax - y\|_{\mathbb{C}^m} = \|(AA^+ - \mathbb{1}_m)y\|_{\mathbb{C}^m} = \|P_2 y\|_{\mathbb{C}^m}$, que é nulo se e somente se $y \in \mathrm{Ker}\,(P_2) = \mathrm{Ran}\,(A)$ (pela Proposição 10.38, página 566), um fato um tanto óbvio.

Prova do Teorema 10.39. A imagem de A, $\mathrm{Ran}\,(A)$, é um subespaço linear fechado de $\mathbb{C}^m$. O Teorema do Melhor Aproximante e o Teorema da Decomposição Ortogonal (que neste texto são apresentados com toda generalidade – no contexto de espaços de Hilbert, como $\mathbb{C}^m$ – na forma do Teorema 38.1, página 1980, e do Teorema 38.2, página 1982, respectivamente) garantem-nos a existência de um único $y_0 \in \mathrm{Ran}\,(A)$ tal que $\|y_0 - y\|_{\mathbb{C}^m}$ é mínimo, sendo que esse y_0 satisfaz a propriedade de $y_0 - y$ ser ortogonal a $\mathrm{Ran}\,(A)$.

Assim, existe ao menos um $x_0 \in \mathbb{C}^n$ tal que $\|Ax_0 - y\|_{\mathbb{C}^m}$ é mínimo. Tal x_0 não é necessariamente único e, como é fácil ver, $x_1 \in \mathbb{C}^n$ tem as mesmas propriedades se e somente se $x_0 - x_1 \in \mathrm{Ker}\,(A)$ (já que $Ax_0 = y_0$ e $Ax_1 = y_0$, pela unicidade de y_0). Como observamos, $Ax_0 - y$ é ortogonal a $\mathrm{Ran}\,(A)$, ou seja, $\langle (Ax_0 - y),\ Au\rangle_{\mathbb{C}} = 0$ para todo $u \in \mathbb{C}^n$. Isso significa que $\langle (A^*Ax_0 - A^*y),\ u\rangle_{\mathbb{C}} = 0$ para todo $u \in \mathbb{C}^n$ e, portanto, x_0 satisfaz

$$A^*Ax_0 \ = \ A^*y \ . \tag{10.201}$$

Agora, a relação (10.178) mostra-nos que $x_0 = A^+y$ satisfaz (10.201), pois $A^*AA^+y \stackrel{(10.178)}{=} A^*y$. Assim, concluímos que o conjunto de todos $x \in \mathbb{C}^n$ que satisfazem a condição de $\|Ax - y\|_{\mathbb{C}^m}$ ser mínimo é composto por todos os vetores da forma $A^+y + x_1$ com $x_1 \in \mathrm{Ker}\,(A)$. Pela Proposição 10.38, página 566, x_1 é da forma $x_1 = (\mathbb{1}_n - A^+A)z$ para algum $z \in \mathbb{C}^n$, completando a prova. ■

Os exercícios que seguem ilustram a aplicação da pseudoinversa de Moore-Penrose no tratamento de problemas de optimização linear.

E. 10.49 *Exercício*. Usando o Exercício E. 10.45, página 564, determine o conjunto dos melhores aproximantes $x \in \mathbb{C}^3$ à solução da equação linear $Ax = y$ com $A = \left(\begin{smallmatrix} 2 & 0 & i \\ 0 & i & 1 \end{smallmatrix}\right)$ e $y = \left(\begin{smallmatrix} 1 \\ -2i \end{smallmatrix}\right)$. Para tais vetores minimizantes x, calcule $\|Ax - y\|_{\mathbb{C}}$. ✠

O exercício que segue envolve uma situação menos trivial que a do exercício anterior, pois trata de um sistema linear subdeterminado e que não tem solução.

E. 10.50 *Exercício*. Usando o Exercício E. 10.46, página 564, determine o conjunto dos melhores aproximantes $x \in \mathbb{C}^2$ à solução da equação linear $Ax = y$ com $A = \begin{pmatrix} 1 & 2 \\ 0 & i \\ 0 & 3 \end{pmatrix}$ e $y = \begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix}$. Para tais vetores minimizantes x, calcule $\|Ax - y\|_{\mathbb{C}}$. Observe que nesse caso $y \notin \mathrm{Ran}\,(A)$ e, portanto, o sistema $Ax = y$ não tem solução. ✠

10.9.3 Existência e Decomposição em Valores Singulares

Passemos agora a uma segunda demonstração da existência da pseudoinversa de Moore-Penrose de uma matriz $A \in \mathrm{Mat}\,(\mathbb{C},\ m,\ n)$ geral, fazendo uso aqui do Teorema da Decomposição em Valores Singulares, Teorema 10.26, página 542. Trataremos primeiramente de matrizes quadradas para depois passarmos ao caso de matrizes não quadradas.

• **Determinando a pseudoinversa de Moore-Penrose para matrizes quadradas**

Começaremos pelas matrizes diagonais. Se $D \in \mathrm{Mat}\,(\mathbb{C},\ n)$ é uma matriz diagonal, a pseudoinversa de Moore-Penrose de D é dada pela matriz diagonal $D^+ \in \mathrm{Mat}\,(\mathbb{C},\ n)$ cujos elementos diagonais são definidos para todo $i = 1,\ \ldots,\ n$ por

$$\left(D^+\right)_{ii} = \begin{cases} \frac{1}{D_{ii}}\ , & \text{se } D_{ii} \neq 0\ , \\ 0\ , & \text{se } D_{ii} = 0\ . \end{cases}$$

É elementar verificar que $DD^+D = D$, $D^+DD^+ = D^+$ e que DD^+ e D^+D são autoadjuntas. Em verdade, vale $DD^+ = D^+D$ que é uma matriz diagonal com elementos diagonais iguais a 0 ou a 1:

$$\left(DD^+\right)_{ii} = \left(D^+D\right)_{ii} = \begin{cases} 1\ , & \text{se } D_{ii} \neq 0\ , \\ 0\ , & \text{se } D_{ii} = 0\ . \end{cases}$$

Passemos agora à questão da existência da pseudoinversa de Moore-Penrose de uma matriz quadrada geral. Se $A \in \mathrm{Mat}\,(\mathbb{C},\ n)$ tem uma decomposição em valores singulares $A = VSW^*$ (vide Teorema 10.26, página 542), então a pseudoinversa de Moore-Penrose A^+ de A é dada por

$$A^+ = WS^+V^*\ .$$

De fato, $AA^+A = \big(VSW^*\big)\big(WS^+V^*\big)\big(VSW^*\big) = VSS^+SW^+ = VSW^* = A$ e $A^+AA^+ = \big(WS^+V^*\big)\big(VSW^*\big)\big(WS^+V^*\big) = WS^+SS^+V^* = WS^+V^* = A^+$. Além disso, $AA^+ = \big(VSW^*\big)\big(WS^+V^*\big) = V\big(SS^+\big)V^*$ é autoadjunta, pois SS^+ é uma matriz diagonal com elementos diagonais iguais a 0 ou a 1. Analogamente, $A^+A = \big(WS^+V^*\big)\big(VSW^*\big) = W\big(S^+S\big)W^*$ é autoadjunta.

• **Determinando a pseudoinversa de Moore-Penrose para matrizes retangulares**

Consideraremos agora matrizes gerais (não necessariamente quadradas) $A \in \mathrm{Mat}\,(\mathbb{C},\ m,\ n)$.

Seja $A' \in \mathrm{Mat}\,(\mathbb{C},\ m+n)$ a matriz quadrada $(m+n) \times (m+n)$ definida em (10.7), página 464. Como A' é uma matriz quadrada, estabelecemos acima que ela possui uma pseudoinversa de Moore-Penrose $(A')^+$, única, satisfazendo

1. $A'\big(A'\big)^+A' = A'$,
2. $\big(A'\big)^+A'\big(A'\big)^+ = \big(A'\big)^+$,
3. $A'\big(A'\big)^+$ e $\big(A'\big)^+A'$ são autoadjuntas.

No que segue, demonstraremos que $A^+ \in \mathrm{Mat}\,(\mathbb{C},\ n,\ m)$, a pseudoinversa de Moore-Penrose de $A \in \mathrm{Mat}\,(\mathbb{C},\ m,\ n)$, é dada, seguindo as definições (10.3)–(10.4), por

$$A^+ := I_{n,\ m+n}\big(A'\big)^+J_{m+n,\ m}\ , \tag{10.202}$$

ou seja,

$$A^+ \;=\; I_{n,\;m+n}\Big(J_{m+n,\;m}AI_{n,\;m+n}\Big)^+J_{m+n,\;m}\;. \tag{10.203}$$

O ponto de partida é a existência da pseudoinversa de A'. A relação $A'\big(A'\big)^+A' = A'$ significa, usando a definição (10.7),

$$J_{m+n,\;m}A\Big[I_{n,\;m+n}\big(A'\big)^+J_{m+n,\;m}\Big]AI_{n,\;m+n} \;=\; J_{m+n,\;m}AI_{n,\;m+n}$$

e das relações (10.5)–(10.6) segue, multiplicando-se à esquerda por $I_{m,\;m+n}$ e à direita por $J_{m+n,\;n}$ que $AA^+A = A$, uma das relações que desejamos provar.

A relação $\big(A'\big)^+A'\big(A'\big)^+ = \big(A'\big)^+$ significa, usando a definição (10.7),

$$\big(A'\big)^+J_{m+n,\;m}AI_{n,\;m+n}\big(A'\big)^+ \;=\; \big(A'\big)^+\;.$$

Multiplicando à esquerda por $I_{n,\;m+n}$ e à direita por $J_{m+n,\;m}$, isso estabelece a validade de $A^+AA^+ = A^+$.

Como $A'(A')^+$ é autoadjunta, segue da definição a definição (10.7), que $J_{m+n,\;m}AI_{n,\;m+n}\big(A'\big)^+$ é autoadjunta, ou seja,

$$J_{m+n,\;m}AI_{n,\;m+n}\big(A'\big)^+ \;=\; \Big(AI_{n,\;m+n}\big(A'\big)^+\Big)^*I_{m,\;m+n}\;.$$

Logo, multiplicando-se à esquerda por $I_{m,\;m+n}$ e à direita por $J_{m+n,\;m}$, segue de (10.5) que

$$AI_{n,\;m+n}\big(A'\big)^+J_{m+n,\;m} \;=\; I_{m,\;m+n}\Big(AI_{n,\;m+n}\big(A'\big)^+\Big)^* \;=\; \Big(AI_{n,\;m+n}\big(A'\big)^+J_{m+n,\;m}\Big)^*\;,$$

provando que AA^+ é autoadjunta.

Por fim, como $(A')^+A'$ é autoadjunta, segue da definição (10.7) que $\big(A'\big)^+J_{m+n,\;m}AI_{n,\;m+n}$ é autoadjunta, ou seja,

$$(A')^+J_{m+n,\;m}AI_{n,\;m+n} \;=\; J_{m+n,\;n}\Big(\big(A'\big)^+J_{m+n,\;m}A\Big)^*\;.$$

Logo, multiplicando-se à esquerda por $I_{n,\;m+n}$ e à direita por $J_{m+n,\;n}$, segue de (10.6) que

$$I_{n,\;m+n}\big(A'\big)^+J_{m+n,\;m}A \;=\; \big((A')^+J_{m+n,\;m}A\big)^*J_{m+n,\;n} \;=\; \big(I_{n,\;m+n}(A')^+J_{m+n,\;m}A\big)^*\;,$$

estabelecendo que A^+A é autoadjunta. Com isso estabelecemos que A^+ dada em (10.202) é a pseudoinversa de Moore-Penrose de A.

10.10 Produtos Tensoriais de Matrizes

A noção de produto tensorial de espaços vetoriais foi introduzida e desenvolvida na Seção 2.3.5, página 173. Vamos considerar os espaços vetoriais complexos de dimensão finita $\mathbb{C}^m$ e $\mathbb{C}^n$ (o caso de espaços vetoriais reais de dimensão finita é totalmente análogo) e seu produto tensorial $\mathbb{C}^m \otimes \mathbb{C}^n$, que á um espaço vetorial complexo de dimensão mn, isomorfo, portanto, a $\mathbb{C}^{mn}$. Sejam $\big\{\mathbf{e}_1,\;\ldots,\;\mathbf{e}_m\big\}$ e $\big\{\mathbf{f}_1,\;\ldots,\;\mathbf{f}_n\big\}$ as bases canônicas em $\mathbb{C}^m$ e $\mathbb{C}^n$, respectivamente, com a qual podemos constituir a base canônica $\mathcal{B} := \big\{\mathbf{e}_i \otimes \mathbf{f}_j,\; i = 1,\;\ldots m,\; j = 1,\;\ldots,\;n\big\}$ em $\mathbb{C}^m \otimes \mathbb{C}^n$.

Um elemento genérico de $\mathbb{C}^m \otimes \mathbb{C}^n$ é uma soma finita $\sum_{a=1}^N \psi_a \otimes \phi_a$, para algum $N \in \mathbb{N}$ e com $\psi_a \in \mathbb{C}^m$ e $\phi_a \in \mathbb{C}^n$ para todo $a = 1,\;\ldots,\;N$. Se $A \in \mathrm{Mat}\,(\mathbb{C},\;m)$ e $B \in \mathrm{Mat}\,(\mathbb{C},\;n)$, definimos seu produto tensorial, denotado por $A \otimes B$, como a matriz que age em um vetor genérico qualquer de $\sum_{a=1}^N \psi_a \otimes \phi_a$ de $\mathbb{C}^m \otimes \mathbb{C}^n$ de acordo com o seguinte:

$$A \otimes B\left(\sum_{a=1}^N \psi_a \otimes \phi_a\right) \;=\; \sum_{a=1}^N \big(A\psi_a\big) \otimes \big(B\phi_a\big)\;. \tag{10.204}$$

O produto tensorial $A \otimes B$ de duas matrizes é também denominado *produto de Kronecker*[42] de A e B.

[42] Leopold Kronecker (1823–1891).

É elementar constatar que $A \otimes B$, assim definido, é um operador linear em $\mathbb{C}^m \otimes \mathbb{C}^n$. Por convenção, as matrizes A e B agem nos respectivos vetores de base de acordo com a regra estabelecida em (10.14):

$$A\mathbf{e}_i \; = \; \sum_{a=1}^{m} A_{ai}\,\mathbf{e}_a \qquad \text{e} \qquad B\mathbf{f}_j \; = \; \sum_{b=1}^{n} B_{bj}\,\mathbf{f}_b \; ,$$

onde A_{ai} e B_{bj} são os elementos de matriz de A e B nas respectivas bases. Assim, vale

$$A \otimes B\left(\mathbf{e}_i \otimes \mathbf{f}_j\right) \; = \; \sum_{a=1}^{m}\sum_{b=1}^{n} A_{ai}B_{bj}\,\mathbf{e}_a \otimes \mathbf{f}_b \; . \tag{10.205}$$

Consequentemente, podemos afirmar que os elementos de matriz de $A \otimes B$ na base $\mathcal{B}$ de $\mathbb{C}^m \otimes \mathbb{C}^n$ é

$$\big(A \otimes B\big)_{(a,\,b)(i,\,j)} \; = \; A_{ai}B_{bj} \; ,$$

pois assim, seguindo a mesma convenção, podemos reescrever (10.205), na forma

$$A \otimes B\left(\mathbf{e}_i \otimes \mathbf{f}_j\right) \; = \; \sum_{(a,\,b)} \big(A \otimes B\big)_{(a,\,b)(i,\,j)} \left(\mathbf{e}_a \otimes \mathbf{f}_b\right) \; ,$$

onde $\sum_{(a,\,b)}$ significa $\sum_{a=1}^{m}\sum_{b=1}^{n}$. Nessa representação os pares ordenados $(i,\, j) \in \{1,\, \ldots,\, m\} \times \{1,\, \ldots,\, n\}$ fazem o papel de índices das matrizes.

Se $A,\; A_1,\; A_2 \in \mathrm{Mat}\,(\mathbb{C},\; m)$ e $B,\; B_1,\; B_2 \in \mathrm{Mat}\,(\mathbb{C},\; n)$ é trivial demonstrar com uso de (10.204) que

$$A_1 \otimes B + A_2 \otimes B \; = \; (A_1 + A_2) \otimes B \; , \qquad A \otimes B_1 + A \otimes B_2 \; = \; A \otimes (B_1 + B_2) \tag{10.206}$$

e que (verifique!)

$$\big(A_1 \otimes B_1\big)\big(A_2 \otimes B_2\big) \; = \; \big(A_1A_2\big) \otimes \big(B_1B_2\big) \; . \tag{10.207}$$

E. 10.51 *Exercício*. Prove que $(A \otimes B)^T = A^T \otimes B^T$ e que $(A \otimes B)^* = A^* \otimes B^*$. ✠

Segue também de (10.204) que $\mathbb{1}_m \otimes \mathbb{1}_n$ é a matriz unidade em $\mathbb{C}^m \otimes \mathbb{C}^n$. Portanto, se A e B possuírem inversa, $A \otimes B$ também possuirá e valerá (verifique!)

$$\big(A \otimes B\big)^{-1} \; = \; \big(A^{-1}\big) \otimes \big(B^{-1}\big) \; . \tag{10.208}$$

Para a recíproca dessa afirmação precisamos aprender algo sobre determinantes de produtos tensoriais de matrizes.

• O determinante de um produto tensorial de matrizes

É imediato por (10.207) que $A \otimes B = \big(A \otimes \mathbb{1}_n\big)\big(\mathbb{1}_m \otimes B\big) = \big(\mathbb{1}_m \otimes B\big)\big(A \otimes \mathbb{1}_n\big)$. Segue que o determinante de $A \otimes B$ será dado por $\det\big(A \otimes B\big) = \det\big(A \otimes \mathbb{1}_n\big)\det\big(\mathbb{1}_m \otimes B\big)$. Vamos agora determinar os dois determinantes do lado direito.

Ordenemos os vetores da base $\mathcal{B}$ na forma $\big(\mathbf{e}_1 \otimes \mathbf{f}_1,\; \ldots,\; \mathbf{e}_m \otimes \mathbf{f}_1,\; \ldots,\; \mathbf{e}_1 \otimes \mathbf{f}_n,\; \ldots,\; \mathbf{e}_m \otimes \mathbf{f}_n\big)$. É claro que $\mathbb{C}^m \otimes \mathbb{C}^n$ quebra-se na soma direta de subespaços $V_1 \oplus \ldots \oplus V_n$, onde V_k é gerado pelos vetores $\mathbf{e}_1 \otimes \mathbf{f}_k,\; \ldots,\; \mathbf{e}_m \otimes \mathbf{f}_k$. Cada V_k é um subespaço invariante pela ação de $A \otimes \mathbb{1}_n$, que nele age como $(A\mathbf{e}_1) \otimes \mathbf{f}_k,\; \ldots,\; (A\mathbf{e}_n) \otimes \mathbf{f}_k$. Assim, ordenando os elementos da base $\mathcal{B}$ dessa maneira, $A \otimes \mathbb{1}_n$ assumirá a representação matricial na forma de n blocos diagonais, como apresentado à esquerda:

$$\begin{pmatrix} \boxed{A} & & \\ & \ddots & \\ & & \boxed{A} \end{pmatrix} .$$

Disso, fica evidente[43] que $\det\big(A \otimes \mathbb{1}_n\big) = \big(\det(A)\big)^n$ (Proposição 10.3, página 472).

Para o caso de $\det\big(\mathbb{1}_m \otimes B\big)$ a ideia é análoga. Ordenamos a base na forma $\big(\mathbf{e}_1 \otimes \mathbf{f}_1,\; \ldots,\; \mathbf{e}_1 \otimes \mathbf{f}_n,\; \ldots,\; \mathbf{e}_m \otimes \mathbf{f}_1,\; \ldots,\; \mathbf{e}_m \otimes \mathbf{f}_n\big)$ e fica claro que

[43] Lembrar que o reordenamento de uma base não altera o determinante de uma matriz, pois é realizado por uma transformação de similaridade envolvendo uma matriz de permutação.

$\mathbb{C}^m \otimes \mathbb{C}^n$ quebra-se na soma direta de subespaços $W_1 \oplus \ldots \oplus W_m$, onde W_k é gerado pelos vetores $\mathbf{e}_k \otimes \mathbf{f}_1, \ \ldots, \ \mathbf{e}_k \otimes \mathbf{f}_n$. Cada W_k é um subespaço invariante pela ação de $\mathbb{1}_m \otimes B$, que nele age como $\mathbf{e}_k \otimes (B\mathbf{f}_1), \ \ldots, \ \mathbf{e}_k \otimes (B\mathbf{f}_n)$. Com a base $\mathcal{B}$ assim ordenada, $\mathbb{1}_m \otimes B$ assumirá a representação matricial na forma de m blocos diagonais como apresentado à esquerda:

$$\begin{pmatrix} \boxed{B} & & \\ & \ddots & \\ & & \boxed{B} \end{pmatrix} .$$

Disso, fica evidente que $\det\big(\mathbb{1}_m \otimes B\big) = \big(\det(B)\big)^m$. Juntando as afirmações anteriores, estabelecemos que se $A \in \mathrm{Mat}\,(\mathbb{C}, \ m)$ e $B \in \mathrm{Mat}\,(\mathbb{C}, \ n)$, então

$$\det\big(A \otimes B\big) \ = \ \big(\det(A)\big)^n \big(\det(B)\big)^m \ . \tag{10.209}$$

Observe o leitor que o expoente de $\det(A)$ à direita é a ordem de B e *vice-versa*.

E. 10.52 *Exercício (fácil)*. Sejam $A_1 \in \mathrm{Mat}\,(\mathbb{C}, \ m_1)$, $A_2 \in \mathrm{Mat}\,(\mathbb{C}, \ m_2)$ e $A_3 \in \mathrm{Mat}\,(\mathbb{C}, \ m_3)$. Mostre que

$$\det\big(A_1 \otimes A_2 \otimes A_3\big) \ = \ \big(\det(A_1)\big)^{m_2 m_3} \big(\det(A_2)\big)^{m_1 m_3} \big(\det(A_3)\big)^{m_1 m_2} \ .$$

✠

A relação (10.209) diz-nos, entre outras coisas, que $A \otimes B$ é uma matriz inversível se e somente se A e B o forem. Em qualquer desses casos, valerá (10.208).

10.11 Propriedades Especiais de Determinantes

10.11.1 Expansão do Polinômio Característico

Seja $A \in \mathrm{Mat}\,(\mathbb{C}, \ n)$ e seja $p_A(\lambda) = \det(\lambda\mathbb{1} - A) = \sum_{m=0}^{n} c_m \lambda^m$, $\lambda \in \mathbb{C}$, seu polinômio característico. Desejamos obter uma fórmula explicita para os coeficientes c_m em termos de determinantes de submatrizes de A (vide abaixo). Vamos designar por $\mathbf{a}_k$ a k-ésima coluna de A, de sorte que, pela notação introduzida em (10.9), página 464, valha $A = [\![\mathbf{a}_1, \ \ldots, \ \mathbf{a}_n]\!]$. Recordando a definição de base canônica fornecida em (10.10) e (10.11), página 464, fica claro que $p_A(\lambda) = \det [\![\lambda\mathbf{e}_1 - \mathbf{a}_1, \ \ldots, \ \lambda\mathbf{e}_n - \mathbf{a}_n]\!]$. Usando a propriedade de multilinearidade do determinante (linearidade em relação a cada coluna), segue que

$$p_A(\lambda) \ = \ \sum_{m=1}^{n} (-1)^{n-m} \lambda^m \left(\sum_{1 \leq j_1 < \cdots < j_m \leq n} \det \Big[\!\!\Big[\mathbf{a}_1, \ \ldots, \ \mathbf{e}_{j_1} \ \ldots, \ \mathbf{e}_{j_m} \ \ldots, \ \mathbf{a}_n\Big]\!\!\Big] \right) + (-1)^n \det(A) \ ,$$

onde, para $1 \leq j_1 < \cdots < j_m \leq n$, $[\![\mathbf{a}_1, \ \ldots, \ \mathbf{e}_{j_1} \ \ldots, \ \mathbf{e}_{j_m} \ \ldots, \ \mathbf{a}_n]\!]$ é a matriz obtida a partir da matriz A substituindo sua j_l-ésima coluna por $\mathbf{e}_{j_l}$ para cada $l = 1, \ \ldots, \ m$. Note que no caso $m = n$, tem-se forçosamente $j_l = l$ para cada $l = 1, \ \ldots, \ n$ e $[\![\mathbf{a}_1, \ \ldots, \ \mathbf{e}_{j_1} \ \ldots, \ \mathbf{e}_{j_m} \ \ldots, \ \mathbf{a}_n]\!] = [\![\mathbf{e}_1, \ \ldots, \ \mathbf{e}_n]\!] = \mathbb{1}$. Com isso, escrevemos

$$p_A(\lambda) \ = \ \lambda^n + \sum_{m=1}^{n-1} (-1)^{n-m} \lambda^m \left(\sum_{1 \leq j_1 < \cdots < j_m \leq n} \det \Big[\!\!\Big[\mathbf{a}_1, \ \ldots, \ \mathbf{e}_{j_1} \ \ldots, \ \mathbf{e}_{j_m} \ \ldots, \ \mathbf{a}_n\Big]\!\!\Big] \right) + (-1)^n \det(A) \ .$$

Como cada vetor-coluna $\mathbf{e}_{j_l}$ contém 1 na j_l-ésima linha, as demais linhas sendo nulas, as bem-conhecidas regras de cálculo de determinantes ensinam-nos que, para todo $m = 1, \ \ldots, \ n-1$,

$$\det \Big[\!\!\Big[\mathbf{a}_1, \ \ldots, \ \mathbf{e}_{j_1} \ \ldots, \ \mathbf{e}_{j_m} \ \ldots, \ \mathbf{a}_n\Big]\!\!\Big] \ = \ \det\Big(A_{j_1, \ldots, j_m}\Big) \ ,$$

$A_{j_1, \ldots, j_m}$ sendo a matriz de $\mathrm{Mat}\,(\mathbb{C}, \ n-m)$ (ou seja $(n-m) \times (n-m)$) obtida a partir de A eliminando-lhe as j_l-ésimas linhas e colunas para todo $l = 1, \ldots, m$. Assim, obtemos

$$p_A(\lambda) \ = \ \lambda^n + \sum_{m=1}^{n-1} (-1)^{n-m} \lambda^m \left(\sum_{1 \leq j_1 < \cdots < j_m \leq n} \det\Big(A_{j_1, \ldots, j_m}\Big) \right) + (-1)^n \det(A) \ , \tag{10.210}$$

onde é possível reconhecer os coeficientes de $p_A(\lambda)$.

Pelo Teorema de Hamilton-Cayley, Teorema 10.4, página 486, $p_A(A) = \mathbb{0}$ e, portanto,

$$A^n + \sum_{m=1}^{n-1} \left((-1)^{n-m} \sum_{1\leq j_1 < \cdots < j_m \leq n} \det\left(A_{j_1,\,\ldots,\,j_m}\right) \right) A^m + (-1)^n \det(A)\mathbb{1} \;=\; \mathbb{0} \;.$$

Como comentamos em (10.47), página 489, se A for inversível, obtém-se disso

$$A^{-1} \;=\; \frac{1}{(-1)^{n+1}\det(A)} \left[A^{n-1} + \sum_{m=1}^{n-1} \left((-1)^{n-m} \sum_{1\leq j_1 < \cdots < j_m \leq n} \det\left(A_{j_1,\,\ldots,\,j_m}\right) \right) A^{m-1} \right] . \tag{10.211}$$

10.11.2 A Desigualdade de Hadamard

Vamos nesta seção demonstrar uma desigualdade para determinantes de matrizes, a qual é muito útil, a chamada *desigualdade de Hadamard*[44].

Teorema 10.40 (Teorema do Determinante de Hadamard) *Seja $A \in \mathrm{Mat}(\mathbb{C},\, n)$. Então,*

$$|\det(A)|^2 \;\leq\; \prod_{j=1}^{n} \sum_{i=1}^{n} |A_{ij}|^2 \;, \tag{10.212}$$

sendo A_{ij} o elemento ij da matriz A. Segue disso que

$$|\det(A)| \;\leq\; n^{n/2} \left(\max_{ij} |A_{ij}| \right)^n \;, \tag{10.213}$$

para toda matriz $A \in \mathrm{Mat}(\mathbb{C},\, n)$. □

<u>*Comentários*</u>. **I.** Seja $A \in \mathrm{Mat}(\mathbb{C},\, n)$ e seja $\mathbf{a}_k \in \mathbb{C}^n$, $1 \leq k \leq n$, sua k-ésima coluna, de sorte que, pela notação introduzida em (10.9), página 464, valha $A = [\![\mathbf{a}_1,\, \ldots,\, \mathbf{a}_n]\!]$. A desigualdade de Hadamard (10.212) afirma que

$$|\det A| \;\leq\; \prod_{j=1}^{n} \|\mathbf{a}_j\|_2 \;, \tag{10.214}$$

onde, para $v \in \mathbb{C}^n$ com componentes $v_1,\, \ldots,\, v_n \in \mathbb{C}$, definimos $\|v\|_2 := \sqrt{\sum_{i=1}^n |v_i|^2}$, a chamada norma Euclidiana em $\mathbb{C}^n$.

II. Um ponto importante na estimativa (10.213) é o tipo de dependência em n que se tem do lado direito. Ela será usada, por exemplo, em estimativas de convergência da série de determinantes de Fredholm na Seção 20.2, página 1014.

III. A desiguldade (10.212), acima, será usada na Seção 11.7, página 623, para demonstrar a continuidade do determinante de matrizes. ♣

Prova do Teorema 10.40. A prova de (10.213) é elementar, por (10.212). Passemos à prova de (10.212).

Seja $A \in \mathrm{Mat}(\mathbb{C},\, n)$. Se A não tem inversa, então $\det(A) = 0$ e a desigualdade (10.212) é trivialmente satisfeita, não havendo o que se provar. Vamos então supor que A tenha inversa.

Seja $\mathcal{A}$ o conjunto de todas as matrizes M de $\mathrm{Mat}(\mathbb{C},\, n)$ com a propriedade que

$$\sum_{i=1}^{n} |M_{ij}|^2 \;=\; \sum_{i=1}^{n} |A_{ij}|^2$$

para todo $j = 1,\, \ldots,\, n$. Claro está que $A \in \mathcal{A}$. É também claro que $\mathcal{A}$ é um subconjunto compacto de $\mathrm{Mat}(\mathbb{C},\, n)$ (visto aqui como $\mathbb{C}^{n^2}$). A função $|\det(M)|$ é contínua como função de M e, portanto, assume ao menos um máximo absoluto (não necessariamente único) em $\mathcal{A}$, por este ser compacto (teorema de Weierstrass). Seja $T \in \mathcal{A}$ um desses máximos. Note-se que $|\det(T)| \geq |\det(A)| > 0$ e, portanto, T tem inversa.

[44] Jacques Salomon Hadamard (1865–1963). A referência ao trabalho de Hadamard é: J. Hadamard, “Résolution d'une question relativ aux déterminants”, Bull. Sci. Math. **28**, 240-246 (1893).

Para todo $i = 1, \ldots, n$ vale por (10.21), página 468, que $\det(T) = \sum_{j=1}^{n} T_{ij}\mathrm{Cof}(T)_{ij}$, onde $\mathrm{Cof}(T)$, chamada de matriz dos cofatores de T, foi definida no enunciado do Teorema 10.1, página 467. Seja fixo esse i. Pela desigualdade de Cauchy-Schwarz, vale

$$|\det(T)|^2 \leq \left(\sum_{j=1}^{n} |T_{ij}|^2\right)\left(\sum_{j=1}^{n} |\mathrm{Cof}(T)_{ij}|^2\right) = \left(\sum_{j=1}^{n} |A_{ij}|^2\right)\left(\sum_{j=1}^{n} |\mathrm{Cof}(T)_{ij}|^2\right) . \tag{10.215}$$

A última igualdade sendo devida ao fato que $T \in \mathcal{A}$.

Como é bem sabido, para o produto escalar $\langle a,\ b\rangle := \sum_{k=1}^{n} \overline{a_k} b_k$, a desigualdade de Cauchy-Schwarz $|\langle a,\ b\rangle| \leq \|a\|\|b\|$ é uma igualdade se e somente se os vetores a e b forem proporcionais. Assim, tem-se a igualdade em (10.215) se e somente se existir $\lambda_i \in \mathbb{C}$ tal que $\overline{T_{ij}} = \lambda_i \mathrm{Cof}(T)_{ij}$ para todo j, ou seja, se a i-ésima linha de $\overline{T}$ for proporcional à i-ésima linha de $\mathrm{Cof}(T)$.

O ponto importante agora e notar que se tivermos a desigualdade estrita

$$|\det(T)|^2 < \left(\sum_{j=1}^{n} |A_{ij}|^2\right)\left(\sum_{j=1}^{n} |\mathrm{Cof}(T)_{ij}|^2\right) , \tag{10.216}$$

então T não pode maximizar o módulo de determinante entre as matrizes de $\mathcal{A}$. De fato, considere a matriz T' que é igual à matriz T, exceto sua i-ésima linha, que é dada por

$$T'_{ij} := \left(\frac{\sum_{j=1}^{n} |A_{ij}|^2}{\sum_{j=1}^{n} |\mathrm{Cof}(T)_{ij}|^2}\right)^{1/2} \overline{\mathrm{Cof}(T)_{ij}} ,$$

$j = 1, \ldots, n$. É claro que

$$\sum_{j=1}^{n} |T'_{ij}|^2 = \sum_{j=1}^{n} |A_{ij}|^2 ,$$

o que mostra que $T' \in \mathcal{A}$ (para as demais linhas T' coincide com T e não há o que provar, pois $T \in \mathcal{A}$). Fora isso, $\det(T') = \sum_{j=1}^{n} T'_{ij}\mathrm{Cof}(T)_{ij}$, pois $\mathrm{Cof}(T')_{ij} = \mathrm{Cof}(T)_{ij}$, já que T' e T só diferem na i-ésima linha. Assim,

$$\det(T') = \left(\frac{\sum_{j=1}^{n} |A_{ij}|^2}{\sum_{j=1}^{n} |\mathrm{Cof}(T)_{ij}|^2}\right)^{1/2} \sum_{j=1}^{n} |\mathrm{Cof}(T)_{ij}|^2 = \left(\sum_{j=1}^{n} |A_{ij}|^2\right)^{1/2}\left(\sum_{j=1}^{n} |\mathrm{Cof}(T)_{ij}|^2\right)^{1/2}$$

e concluímos por (10.216) que teríamos $|\det(T)| < \det(T')$, contrariando a hipótese que $|\det(T)|$ é máximo. Assim, devemos ter a igualdade em (10.215) e, pelos comentários acima, isso implica que existe $\lambda_i \in \mathbb{C}$ tal que $\overline{T_{ij}} = \lambda_i \mathrm{Cof}(T)_{ij}$ para todo j, ou seja, a i-ésima linha de $\overline{T}$ é proporcional à i-ésima linha de $\mathrm{Cof}(T)$. Como i é arbitrário, isso vale para todo i.

Agora, como as linhas de $\overline{T}$ são proporcionais às de $\mathrm{Cof}(T)$, segue que

$$\det(T) = \sum_{j=1}^{n} T_{ij}\mathrm{Cof}(T)_{ij} = \frac{1}{\lambda_i}\sum_{j=1}^{n} |T_{ij}|^2 , = \frac{1}{\lambda_i}\sum_{j=1}^{n} |A_{ij}|^2$$

e pela multilinearidade do determinante, que

$$\overline{\det(T)} = \det(\overline{T}) = \lambda_1 \cdots \lambda_n \det(\mathrm{Cof}(T)) .$$

Dessas duas relações extraímos

$$\det(T)^n \;=\; \frac{1}{\lambda_1 \cdots \lambda_n} \prod_{i=1}^n \sum_{j=1}^n |A_{ij}|^2 \;=\; \frac{\det(\mathrm{Cof}(T))}{\overline{\det(T)}} \prod_{i=1}^n \sum_{j=1}^n |A_{ij}|^2 \;.$$

Como a relação (10.26) vale para qualquer matriz inversível, tem-se $\det(\mathrm{Cof}(T)) = \det(T)^{n-1}$ e, portanto, $|\det(T)|^2 = \prod_{i=1}^n \sum_{j=1}^n |A_{ij}|^2$. Por construção, T maximiza $|\det(T)|$ em $\mathcal{A}$. Como $A \in \mathcal{A}$, segue que

$$|\det(A)|^2 \;\leq\; \prod_{i=1}^n \sum_{j=1}^n |A_{ij}|^2 \;. \tag{10.217}$$

Isso prova o teorema. ∎

10.12 Exercícios Adicionais

E. 10.53 *Exercício*. **a)** Determine o polinômio característico da matriz

$$A = \begin{pmatrix} 5 & -2 & -7 \\ 0 & 2-3i & -5i \\ 0 & 0 & 1-4i \end{pmatrix} .$$

b) Verifique explicitamente a validade do Teorema de Hamilton-Cayley para a matriz A.

c) Usando o Teorema de Hamilton-Cayley calcule A^{-1}.

✠

E. 10.54 *Exercício*. Repita o exercício anterior para as matrizes

$$A_1 = \begin{pmatrix} 2 & -1 & 3 \\ 0 & -4+i & i \\ 0 & 0 & 2-7i \end{pmatrix} , \qquad A_2 = \begin{pmatrix} -5 & 0 & 0 \\ -8 & 8 & 0 \\ -3i & 1-9i & 4-5i \end{pmatrix} .$$

✠

E. 10.55 *Exercício*. Considere em $\mathbb{C}^n$ o seguinte produto escalar

$$\langle u,\ v\rangle_p = \sum_{a=1}^{n} \overline{u_a} v_a p_a ,$$

onde $p_a > 0$ para $a = 1, \ldots, n$. Seja uma matriz A, com elementos de matriz A_{ij}. Mostre que, com o produto escalar $\langle \cdot,\ \cdot\rangle_p$ o elemento de matriz $(A^{*p})_{ij}$ da adjunta A^{*p} da matriz A é dado por

$$(A^{*p})_{ij} = \frac{p_j}{p_i}\overline{A_{ji}} . \tag{10.218}$$

(Lembre-se que A^{*p} é definida de sorte que $\langle u,\ Av\rangle_p = \langle A^{*p}u,\ v\rangle_p$ para todos $u,\ v \in \mathbb{C}^n$).

Para a matriz adjunta definida em (10.218), verifique a validade das regras $(A^{*p})^{*p} = A$ e $(AB)^{*p} = B^{*p}A^{*p}$, para quaisquer matrizes $A,\ B \in \mathrm{Mat}\,(\mathbb{C},\ n)$. Calcule $\mathbb{1}^{*p}$.

Mostre que para quaisquer $u,\ v \in \mathbb{C}^n$ vale $\langle u,\ v\rangle_p = \langle u,\ Pv\rangle_{\mathbb{C}}$, onde $\langle u,\ v\rangle_{\mathbb{C}} = \sum_{a=1}^{n} \overline{u_a} v_a$ é o produto escalar usual em $\mathbb{C}^n$ e $P = \mathrm{diag}\,(p_1,\ \ldots,\ p_n)$. Conclua disso que $A^{*p} = P^{-1}A^*P$, onde A^* é a adjunta usual de A em relação ao produto escalar $\langle \cdot,\ \cdot\rangle_{\mathbb{C}}$: $(A^*)_{ij} = \overline{A_{ji}}$. ✠

E. 10.56 *Exercício*. Determine os autovalores da matriz $A = \begin{pmatrix} 4 & -i/2 \\ 2i & 5 \end{pmatrix}$. Essa matriz não é autoadjunta em relação ao produto escalar usual em $\mathbb{C}^2$, mas possui autovalores reais. Justifique esse fato mostrando, pelos exercícios anteriores, que A é autoadjunta em relação ao produto escalar $\langle u,\ v\rangle_p = 2\overline{u_1}v_1 + \overline{u_2}v_2/2$. Mostre a adjunta A^{*p} em relação a esse produto escalar é $A^{*p} = \begin{pmatrix} 4 & -i/2 \\ 2i & 5 \end{pmatrix} = A$ e constate explicitamente que $\langle u,\ Av\rangle_p = \langle Au,\ v\rangle_p$ para todos $u,\ v \in \mathbb{C}^2$. Determine os autovetores de A e constate que os mesmos são ortogonais em relação ao produto escalar $\langle \cdot,\ \cdot\rangle_p$. ✠

O exercício que segue generaliza o Exercício E. 10.55.

E. 10.57 *Exercício*. Seja ω um produto escalar em $\mathbb{C}^n$. Pela Proposição 3.5, página 247, existe uma única matriz $M_\omega \in \mathrm{Mat}(\mathbb{C},\, n)$ autoadjunta e de autovalores positivos (e, portanto, inversível) tal que $\omega(x,\, y) = \langle x,\, M_\omega y\rangle_{\mathbb{C}}$ para todos $x,\, y \in \mathbb{C}^n$.

Seja $A \in \mathrm{Mat}(\mathbb{C},\, n)$ e seja $A^{*\omega} \in \mathrm{Mat}(\mathbb{C},\, n)$ sua adjunta em relação ao produto escalar ω: $\omega(x,\, Ay) = \omega(A^{*\omega}x,\, y)$ para todos $x,\, y \in \mathbb{C}^n$. Mostre que $A^{*\omega} = M_\omega^{-1}A^*M_\omega$, onde A^* é a adjunta usual de A em relação ao produto escalar $\langle\cdot,\, \cdot\rangle_{\mathbb{C}}$.

Mostre que para quaisquer matrizes $A,\, B \in \mathrm{Mat}(\mathbb{C},\, n)$ valem $(A^{*\omega})^{*\omega} = A$ e $(AB)^{*\omega} = B^{*\omega}A^{*\omega}$. Calcule $\mathbb{1}^{*\omega}$. ✠

E. 10.58 *Exercício*. [**Números de Fibonacci**]. A sequência de números conhecida como *sequência de Fibonacci*[45] foi introduzida na Seção 6.1.2, página 314, e foi lá estudada usando-se funções geratrizes. Neste exercício vamos estudá-la fazendo uso de matrizes e do Teorema Espectral.

A sequência de Fibonacci a_n, $n \in \mathbb{N}_0$, é a sequência definida recursivamente pela relação

$$a_{n+2} \;=\; a_{n+1} + a_n\,, \quad \forall\, n \geq 0\,. \tag{10.219}$$

Comummente adota-se $a_0 = 1$ e $a_1 = 1$, mas vamos deixar essa escolha de “condições iniciais” provisoriamente em aberto.

A relação (10.219) pode ser expressa de forma elegante com o uso de matrizes e vetores, da seguinte forma. Tem-se, trivialmente que

$$T\begin{pmatrix} x \\ y \end{pmatrix} \;=\; \begin{pmatrix} x+y \\ x \end{pmatrix}\,, \qquad \text{onde} \qquad T \;:=\; \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}\,.$$

Isso mostra que vale a seguinte relação para os elementos da sequência de Fibonacci:

$$\begin{pmatrix} a_{n+1} \\ a_n \end{pmatrix} \;=\; T\begin{pmatrix} a_n \\ a_{n-1} \end{pmatrix}\,, \qquad \forall\, n \in \mathbb{N}\,,$$

o que permite escrever

$$\begin{pmatrix} a_{n+1} \\ a_n \end{pmatrix} \;=\; T^n\begin{pmatrix} a_1 \\ a_0 \end{pmatrix}\,, \qquad \forall\, n \in \mathbb{N}_0\,. \tag{10.220}$$

Justifique! A matriz T é, por vezes, denominada *matriz de transferência*.

T é manifestamente autoadjunta e, portanto, diagonalizável (Teorema 10.15, página 510) e, portanto, satisfaz o Teorema Espectral (Teorema 10.7, página 494).

Mostre que os autovalores de T são $\lambda_\pm = \frac{1}{2}\big(1 \pm \sqrt{5}\big)$. Usando (10.60), mostre que a decomposição espectral de T é

$$T \;=\; \lambda_+ E_+ + \lambda_- E_-\,, \qquad \text{onde} \qquad E_\pm \;=\; \frac{\pm 1}{\sqrt{5}}\begin{pmatrix} \lambda_\pm & 1 \\ 1 & -\lambda_\mp \end{pmatrix}\,.$$

Conclua do Cálculo Funcional (Teorema 10.8, página 496) que, para $n \in \mathbb{N}$,

$$T^n \;=\; \big(\lambda_+\big)^n E_+ + \big(\lambda_-\big)^n E_- \;=\; \frac{1}{\sqrt{5}}\begin{pmatrix} \big(\lambda_+\big)^{n+1} - \big(\lambda_-\big)^{n+1} & \big(\lambda_+\big)^n - \big(\lambda_-\big)^n \\ \big(\lambda_+\big)^n - \big(\lambda_-\big)^n & \big(\lambda_+\big)^{n-1} - \big(\lambda_-\big)^{n-1} \end{pmatrix}\,,$$

(use que $\lambda_+\lambda_- = -1$).

Retornando com isso a (10.220), obtenha que

$$a_n \;=\; \frac{1}{\sqrt{5}}\Big[\Big(\big(\lambda_+\big)^{n-1} - \big(\lambda_-\big)^{n-1}\Big)a_0 + \Big(\big(\lambda_+\big)^n - \big(\lambda_-\big)^n\Big)a_1\Big]\,, \tag{10.221}$$

$n \in \mathbb{N}_0$. Essa é a expressão geral (em termos de n, a_0 e a_1) dos elementos a_n da sequência de Fibonacci.

[45] Leonardo Pisano, cognominado “Fibonacci” (1170–1250).

Para o caso particular em que $a_0 = 1$ e $a_1 = 1$, obtenha disso que

$$a_n \;=\; \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right] , \tag{10.222}$$

para todo $n \geq 0$. Para isso, mostre que $(\lambda_\pm)^n + (\lambda_\pm)^{n-1} = (\lambda_\pm)^n\Big(1 + (\lambda_\pm)^{-1}\Big) = (\lambda_\pm)^n(1-\lambda_\mp) = (\lambda_\pm)^{n+1}$.

A expressão (10.222) coincide com o resultado apresentado em (6.6), página 314, e lá obtido por outros meios. ✠

E. 10.59 *Exercício*. [**Números de Fibonacci Generalizados**]. Este exercício generaliza o Exercício E. 10.58.

Considere a *sequência de Fibonacci generalizada*:

$$a_{n+2} \;=\; \alpha a_{n+1} + \beta a_n \;, \quad \forall\, n \geq 0 \;, \tag{10.223}$$

onde α e β são constantes (reais ou complexas). A matriz de transferência T associada a essa sequência é

$$T \;:=\; \begin{pmatrix} \alpha & \beta \\ 1 & 0 \end{pmatrix} .$$

Mostre que os seus autovalores são $\lambda_\pm = \frac{1}{2}\Big(\alpha \pm \sqrt{\alpha^2 + 4\beta}\Big)$.

Considere primeiramente o caso em que $\alpha^2 + 4\beta \neq 0$. Nessa situação, os autovalores λ_+ e λ_- são distintos e, portanto, T é diagonalizável (pela Proposição 10.22, página 503) e aplicam-se novamente o Teorema Espectral e o Cálculo Funcional.

Repita o procedimento do Exercício E. 10.58 para obter a expressão geral (em termos de n, a_0 e a_1) dos elementos a_n da sequência de Fibonacci generalizada. O resultado é que

$$T^n \;=\; \frac{1}{\sqrt{\alpha^2+4\beta}} \begin{pmatrix} (\lambda_+)^{n+1} - (\lambda_-)^{n+1} & \beta\Big((\lambda_+)^n - (\lambda_-)^n\Big) \\ (\lambda_+)^n - (\lambda_-)^n & \beta\Big((\lambda_+)^{n-1} - (\lambda_-)^{n-1}\Big) \end{pmatrix} ,$$

donde obtém-se que

$$a_n \;=\; \frac{1}{\sqrt{\alpha^2+4\beta}}\left[\beta\Big((\lambda_+)^{n-1} - (\lambda_-)^{n-1}\Big)a_0 + \Big((\lambda_+)^n - (\lambda_-)^n\Big)a_1\right] . \tag{10.224}$$

Esta é a expressão geral (em termos de n, a_0, a_1 α e β) da sequência de Fibonacci generalizada para o caso $\beta \neq -\alpha^2/4$.

No caso em que $\alpha^2 + 4\beta = 0$, mostre que T não é diagonalizável. Para isso, mostre, por exemplo, que seus autovetores são todos múltiplos do vetor $\binom{\alpha/2}{1}$ e, portanto, compõem um subespaço unidimensional.

O que se pode fazer nessa situação para determinar T^n? Proceda da seguinte forma: escreva

$$T \;=\; \begin{pmatrix} \alpha & -\alpha^2/4 \\ 1 & 0 \end{pmatrix} \;=\; \frac{\alpha}{2}\mathbb{1} + N \;, \qquad \text{onde} \qquad N \;=\; \begin{pmatrix} \alpha/2 & -\alpha^2/4 \\ 1 & -\alpha/2 \end{pmatrix} .$$

Constate que $N^2 = 0$ e conclua que a representação $T = \frac{\alpha}{2}\mathbb{1} + N$ é a *forma de Jordan* de T. Pelo binômio de Newton, teremos, para $n \geq 1$,

$$T^n \;=\; \left(\frac{\alpha}{2}\mathbb{1} + N\right)^n \;=\; \sum_{p=0}^{n}\binom{n}{p}\left(\frac{\alpha}{2}\right)^{n-p} N^p \;\stackrel{N^2=0}{=}\; \sum_{p=0}^{1}\binom{n}{p}\left(\frac{\alpha}{2}\right)^{n-p} N^p \;=\; \left(\frac{\alpha}{2}\right)^n\mathbb{1} + n\left(\frac{\alpha}{2}\right)^{n-1} N \;.$$

Portanto,

$$T^n \;=\; \begin{pmatrix} (1+n)\left(\frac{\alpha}{2}\right)^n & -n\left(\frac{\alpha}{2}\right)^{n+1} \\ n\left(\frac{\alpha}{2}\right)^{n-1} & (1-n)\left(\frac{\alpha}{2}\right)^n \end{pmatrix} ,$$

e, portanto,

$$a_n \;=\; (1-n)\left(\frac{\alpha}{2}\right)^n a_0 + n\left(\frac{\alpha}{2}\right)^{n-1} a_1 \;. \tag{10.225}$$

Esta é a expressão geral (em termos de n, a_0, a_1 e α) da sequência de Fibonacci generalizada para o caso $\beta = -\alpha^2/4$.

Note-se que no caso $\alpha = 2$ (e $\beta = -1$), obtém-se disso $a_n = a_0 + n(a_1 - a_0)$, que exibe um comportamento dominante linear em relação a n, e não exponencial, como em todos os casos anteriores. Em particular, se $a_0 = a_1$, a sequência é constante. ✠

E. 10.60 *Exercício*. Demonstre a *identidade de polarização para matrizes*: para $A, B \in \mathrm{Mat}(\mathbb{C}, n)$ tem-se:

$$B^*A = \frac{1}{4}\sum_{k=0}^{3} i^k \big(A + i^k B\big)^* \big(A + i^k B\big) . \tag{10.226}$$

Sugestão: simplesmente expanda o lado direito e constate a igualdade. ✠

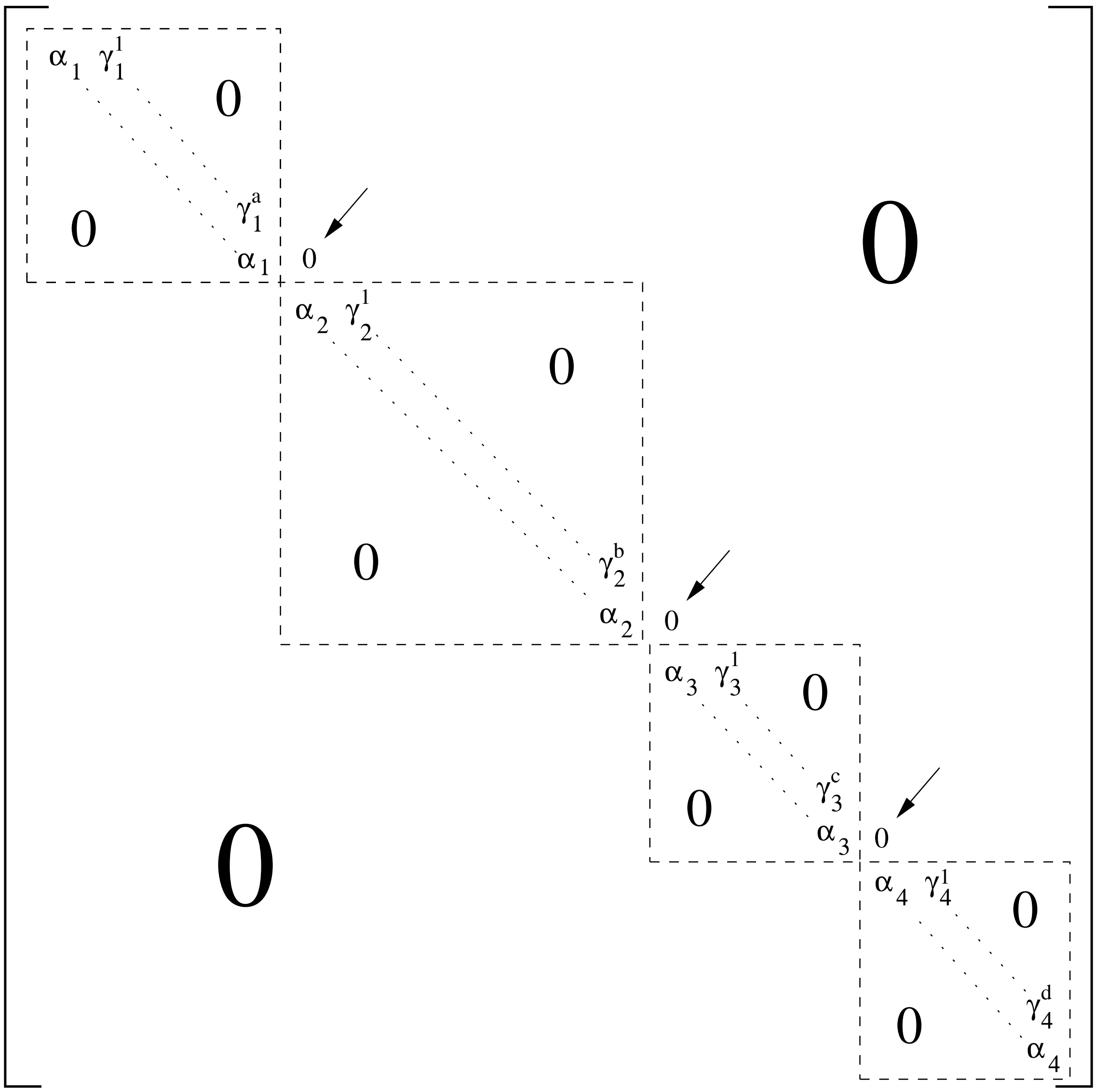

Figura 10.6: Forma canônica de uma matriz com 4 autovalores distintos α_1, α_2, α_3 e α_4. Os γ's assumem apenas os valores 0 ou 1, de acordo com as regras explicadas acima. Todos os elementos fora da diagonal principal e da primeira supradiagonal são nulos. As setas indicam zeros que ocorrem na primeira supradiagonal nos pontos onde ocorre transição entre os blocos, consequência do fato de esses elementos estarem fora dos blocos.