O TEOREMA DE CAYLEY-HAMILTON

Oswaldo Rio Branco de Oliveira (IMEUSP) e Carlos Alexandre Gomes (UFRN)

http://www.ime.usp.br/~oliveira oliveira@ime.usp.br

Ano 2015/2017/2019/2021

OBJETIVO

Nesta notas apresentamos quatro demonstrações do importante Teorema de Cayley-Hamilton. A primeira demonstração é a mais trivial e curta que conhecemos mas não é a mais elementar pois faz uso do teorema fundamental da álgebra. A segunda demonstração é a mais elementar, não utiliza o teorema fundamental da álgebra e faz uso da decomposição de um operador linear. A terceira demonstração é frequente em livros textos e utiliza a noção de matriz adjunta. A quarta demonstração utiliza uma versão da Fórmula Integral de Cauchy no espaço das matrizes quadradas e complexas.

INTRODUÇÃO

Indiquemos por V um espaço vetorial de dimensão n sobre o corpo $\mathbb{K}$, fixado $\mathbb{K} = \mathbb{R}$ ou $\mathbb{K} = \mathbb{C}$. Consideremos uma aplicação linear

$$T : V \to V.$$

Fixemos uma base $\mathcal{B}$ do espaço V. Seja

$$[T]_{\mathcal{B}}$$

a matriz de representação de T em relação à base $\mathcal{B}$.

Seja $\mathcal{C}$ uma outra base de V. Escrevendo os elementos de $\mathcal{C}$ como combinação linear dos elementos de $\mathcal{B}$, encontramos a matriz de mudança de base

$$P = [I]_{\mathcal{B}}^{\mathcal{C}},$$

da base $\mathcal{C}$ para a base $\mathcal{B}$. Donde segue $P^{-1} = [I]_{\mathcal{C}}^{\mathcal{B}}$. Então, temos

$$[T]_{\mathcal{C}} = P^{-1}[T]_{\mathcal{B}}P.$$

Seja λ a variável em $\mathbb{K}$. Propriedades de matrizes e determinantes garantem

$$\det(\lambda I - [T]_{\mathcal{C}}) = \det\left(P^{-1}(\lambda I - [T]_{\mathcal{B}})P\right) = \det(\lambda I - [T]_{\mathcal{B}}).$$

Logo, podemos associar à aplicação T um polinômio que independe da representação de T. Formalizemos

Seja A uma matriz $n \times n$, com coeficientes em $\mathbb{K}$ [notação, $A \in M_n(\mathbb{K})$].

O polinômio característico de A é

$$p_A(\lambda) = \det(\lambda I - A), \text{ onde } \lambda \in \mathbb{K}.$$

Assim, duas **matrizes semelhantes** tem iguais polinômios característicos. Definimos *o polinômio característico de uma aplicação linear* $T : V \to V$, *como o polinômio característico de qualquer uma de suas representações matriciais.*

1. PRIMEIRA PROVA
(trivial e usa o teorema fudamental da ágebra)

Teorema (Cayley-Hamilton). *Sejam $A \in M_n(\mathbb{R})$ e $p_A(\lambda)$ seu polinômio característico. Então,*

$$p_A(A) = 0.$$

Primeira prova (elementar). Usa o *Teorema Fundamental da Álgebra (TFA)*.

O caso $n = 1$ é óbvio. Suponhamos o resultado válido para $n - 1$. Seja $T : \mathbb{C}^n \to \mathbb{C}^n$ o operador linear associado a A. O TFA (*teorema fundamental da álgebra*) garante um autovalor λ e um autovetor associado $v \neq 0$. Seja $\{v, \ldots, \}$ uma base de $\mathbb{C}^n$. Mudando de base, se preciso, podemos supor

$$A = \begin{pmatrix} \lambda & * \\ 0 & B \end{pmatrix}, \quad \text{onde } B \in M_{n-1}(\mathbb{C}).$$

Então temos, com I a identidade de ordem $n-1$ e z a variável complexa,

$$\begin{aligned} p_T(z) &= \det \begin{pmatrix} z - \lambda & * \\ 0 & zI - B \end{pmatrix} \\ &= (z - \lambda)\det(zI - B) \\ &= (z - \lambda)p_B(z). \end{aligned}$$

Donde segue (cheque, particularmente a terceira e a última igualdades)

$$\begin{aligned} p_T(T) &= (A - \lambda I)p_B(A) \\ &= \begin{pmatrix} 0 & * \\ 0 & B - \lambda I \end{pmatrix} p_B \begin{pmatrix} \lambda & * \\ 0 & B \end{pmatrix} \\ &= \begin{pmatrix} 0 & * \\ 0 & B - \lambda I \end{pmatrix} \begin{pmatrix} p_B(\lambda) & * \\ 0 & p_B(B) \end{pmatrix} \\ &= \begin{pmatrix} 0 & * \\ 0 & B - \lambda I \end{pmatrix} \begin{pmatrix} p_B(\lambda) & * \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \clubsuit \end{aligned}$$

2. SEGUNDA PROVA
(elementar e não usa o teorema fundamental da álgebra)

Teorema (Cayley-Hamilton). *Sejam V um $\mathbb{K}$-espaço vetorial de dimensão finita n e $T : V \to V$ linear e $p(\lambda)$ o polinômio característico de T. Então,*

$$p(T) = 0.$$

Prova. *Extraída da referência [1], apostila de H. P. Bueno.*

Consideremos um vetor não nulo e arbitrário $v \in V$. Mostremos $p(T)v = 0$. Seja m o maior natural tal que o conjunto

$$\mathcal{A} = \{v, Tv, \dots, T^{m-1}v\}$$

é linearmente independente. Existem coeficientes $\alpha_0, \dots, \alpha_{m-1}$ tais que

$$T^m v = \alpha_0 v + \dots + \alpha_{m-1} T^{m-1} v. \tag{1.1}$$

Seja W o subespaço gerado pelo conjunto linearmente independente $\mathcal{A}$.

◇ Afirmação. $T(W) \subset W$. De fato, dado $w \in W$ segue que existem escalares $\beta_0, \dots, \beta_{m-1}$ tais que $w = \beta_0 v + \beta_1 Tv + \dots + \beta_{m-1} T^{m-1} v$. Logo,

$$Tw = \beta_0 Tv + \beta_1 T^2 v + \dots + \beta_{m-1} T^m v.$$

A identidade (1.1) garante que $Tw \in W$. Logo, $T(W) \subset W$.

Consideremos a restrição $T|_W : W \to W$. A representação de $T|_W$ na base ordenada $\mathcal{A} = \{v, Tv, \dots, T^{m-1}v\}$ é

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & \alpha_0 \\ 1 & 0 & \cdots & 0 & \alpha_1 \\ 0 & 1 & \cdots & 0 & \alpha_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \alpha_{m-1} \end{pmatrix}.$$

Logo,

$$\det(\lambda I - A) = \begin{vmatrix} \lambda & 0 & \cdots & 0 & -\alpha_0 \\ -1 & \lambda & \cdots & 0 & -\alpha_1 \\ 0 & -1 & \cdots & 0 & -\alpha_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & \lambda - \alpha_{m-1} \end{vmatrix}$$

$$= \lambda \begin{vmatrix} \lambda & 0 & \cdots & -\alpha_1 \\ -1 & \lambda & \cdots & -\alpha_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -1 & \lambda - \alpha_{m-1} \end{vmatrix} +$$

$$-\alpha_0(-1)^{m+1} \begin{vmatrix} -1 & \lambda & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 \end{vmatrix}.$$

O último determinante acima é $(-1)^{m-1}$. Logo, o coeficiente independente do polinômio característico $\det(\lambda I - A)$ é $-\alpha_0$. É claro que o coeficiente dominante de $\det(\lambda I - A)$ é $+1$. Então, por iteração segue

$$\det(\lambda I - A) = -\alpha_0 - \alpha_1\lambda - \alpha_2\lambda^2 - \cdots - \alpha_{m-1}\lambda^{m-1} + \lambda^m = p_W(\lambda),$$

onde $p_W(\lambda)$ é o polinômio característico da restrição $T|_W$. Desta forma, pela equação (1.1) segue

$$(1.2) \qquad p_W(T)v = T^m v - \alpha_0 v - \alpha_1 T v - \alpha_2 T^2 v - \cdots - \alpha_{m-1}T^{m-1}v = 0.$$

- **Afirmação.** Temos $p(\lambda) = q(\lambda)p_W(\lambda)$ para algum polinômio $q(\lambda)$. De fato, completando $\mathcal{A}$ a uma base de V obtemos a representação

  $$[T] = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} \quad [A \text{ de ordem } m \text{ e } C \text{ de ordem } n{-}m].$$

  Então, por propriedades de determinantes encontramos

  $$\det(\lambda I - T) = \det(\lambda I - A)\det(\lambda I - C) = p_W(\lambda)q(\lambda) = q(\lambda)p_W(\lambda).$$

Devido a (1.2) concluímos então

$$p(T)v = q(T)p_W(T)v = 0 \clubsuit$$

3. TERCEIRA PROVA
(usa matriz adjunta e não usa o teorema fundamental da álgebra)

Dada uma matriz $A = (a_{ij})$ em $M_n(\mathbb{K})$, seja A_{ij} o determinante da matriz quadrada de ordem $n-1$ que surge ao eliminarmos a i-ésima linha e a j-ésima coluna da matriz A. A **matriz dos cofatores de** A é de tamanho $n \times n$ e definida por $C = (c_{ij})$ com $c_{ij} = (-1)^{i+j}A_{ij}$. A **matriz adjunta de** A é a matriz transposta de C. Notação,

$$\mathrm{adj}(A) = C^t.$$

Nestas notas, utilizaremos o seguinte resultado (sem prová-lo).

Lema. *Seja $A \in M_n(\mathbb{K})$. Então temos*

$$A.adj(A) = (\det A)I,$$

com $I = I_n$ a matriz identidade de ordem n. Em particular, se $\det A \neq 0$, segue

$$A^{-1} = \frac{1}{\det A} adj(A).$$

Teorema (Cayley-Hamilton). *Seja V um $\mathbb{K}$-espaço vetorial de dimensão finita n e $T : V \to V$ linear e $p(\lambda)$ o polinômio característico de T. Então,*

$$p(T) = 0.$$

Prova.

Seja $\mathcal{B}$ uma base de V e $A = [T]_{\mathcal{B}}$ a matriz de T nesta base. Consideremos $A_\lambda = \lambda I - A$, com I a matriz identidade de ordem n. Assim, $p(\lambda) = \det A_\lambda$. Seja

$$B = \mathrm{adj}(A_\lambda) = (b_{ij}).$$

Logo, cada b_{ij} é um polinômio de grau no máximo $n-1$ na variável λ. Escrevamos, para cada par i, j, tal polinômio como

$$b_{ij} = b_{ij}^{(0)} + b_{ij}^{(1)}\lambda + \cdots + b_{ij}^{(n-1)}\lambda^{n-1}.$$

A matriz dos coeficientes do termo de ordem k (fixa) destes polinômios é

$$B^{(k)} = \begin{pmatrix} b_{11}^{(k)} & b_{12}^{(k)} & \cdots & b_{1n}^{(k)} \\ \vdots & \vdots & & \vdots \\ b_{n1}^{(k)} & b_{n2}^{(k)} & \cdots & b_{nn}^{(k)} \end{pmatrix}, \quad \text{onde } k = 0, 1, \ldots, n-1.$$

Então, segue

$$B = B^{(0)} + B^{(1)}\lambda + \cdots + B^{(n-1)}\lambda^{n-1}.$$

Como $p(\lambda) = \det(\lambda I - A)$ é mônico e de grau n, podemos escrever

$$p(\lambda) = a_0 + a_1\lambda + \cdots + \lambda^n.$$

Escrevemos também

$$BA_\lambda = \operatorname{adj}(A_\lambda)A_\lambda = (\det A_\lambda)I = \det(\lambda I - A).I$$

Substituindo as expressões para B, A_λ e $\det(\lambda I - A) = p(\lambda)$ encontramos

$$\left(B^{(0)} + B^{(1)}\lambda + \cdots + B^{(n-1)}\lambda^{n-1}\right)(\lambda I - A) = (a_0 + a_1\lambda + \cdots + \lambda^n)I.$$

Donde segue

$$-B_0A + (B^{(0)} - B^{(1)}A)\lambda + \cdots + B^{(n-1)}\lambda^n = a_0 I + a_1 I\lambda + \cdots + I\lambda^n.$$

Igualando-se os coeficientes dos termos de mesmo grau obtemos

$$\begin{cases} a_0 I & = & -B^{(0)}A \\ a_1 I & = & B^{(0)} - B^{(1)}A \\ & \vdots & \\ a_{n-1}I & = & B^{(n-2)} - B^{(n-1)}A \\ I & = & B^{(n-1)}. \end{cases}$$

Multiplicando-se essas equações por $I, A, A^2, \ldots, A^n$, segue que

$$\begin{cases} a_0 I & = & -B^{(0)}A \\ a_1 A & = & B^{(0)}A - B^{(1)}A^2 \\ & \vdots & \\ a_{n-1}A^{n-1} & = & B^{(n-2)}A^{n-1} - B^{(n-1)}A^n \\ A^n & = & B^{(n-1)}A^n. \end{cases}$$

Adicionando-se membro a membro as igualdades acima encontramos

$$a_0 I + a_1 A + \cdots + a_{n-1}A^{n-1} + A^n = 0 \Rightarrow \ p_A(A) = 0 \clubsuit$$

4. QUARTA PROVA
(usa fórmula integral de Cauchy e matriz adjunta)

Dada uma matriz $A = (a_{ij})$ em $M_n(\mathbb{K})$, seja A_{ij} o determinante da matriz quadrada de ordem $n-1$ que surge ao eliminarmos a i-ésima linha e a j-ésima coluna da matriz A. A **matriz dos cofatores de** A é de tamanho $n \times n$ e definida por $C = (c_{ij})$ com $c_{ij} = (-1)^{i+j} A_{ij}$. A **matriz adjunta de** A é a matriz transposta de C. Notação,

$$\mathrm{adj}(A) = C^t.$$

Nestas notas, utilizaremos o seguinte resultado (sem prová-lo).

Lema. *Seja $A \in M_n(\mathbb{K})$. Então temos*

$$A.adj(A) = (\det A)I,$$

com $I = I_n$ a matriz identidade de ordem n. Em particular, se $\det A \neq 0$, segue

$$A^{-1} = \frac{1}{\det A} adj(A).$$

Façamos algumas observações e introduzamos algumas notações e definições.

Notações. Fixemos $A \in M_n(\mathbb{K})$ e seu polinômio característico

$$p(\lambda) = \det(\lambda I - A), \text{ onde } \lambda \in \mathbb{C}.$$

Observação 1. Cada entrada da matriz adjunta de A é dada por um polinômio, em n^2 variáveis e de grau $n-1$, nas n^2 entradas da matriz A.

Definição. Consideremos uma função matricial $F = F(z)$, onde z é a variável complexa, que assume valores no espaço de matrizes $M_n(\mathbb{K})$. Isto é,

$$F(z) = \begin{pmatrix} F_{11}(z) & \cdots & F_{1n}(z) \\ \vdots & & \vdots \\ F_{n1}(z) & \cdots & F_{nn}(z) \end{pmatrix}.$$

A integral de F ao longo de uma curva γ no plano complexo, é feita coordenada a coordenada e resulta em uma matriz de integrais. Isto é,

$$\int_\gamma F(z)dz = \begin{pmatrix} \int_\gamma F_{11}(z)dz & \cdots & \int_\gamma F_{1n}(z)dz \\ \vdots & & \vdots \\ \int_\gamma F_{n1}(z)dz & \cdots & \int_\gamma F_{nn}(z) \end{pmatrix}.$$

Observação 2. Seja M uma matriz em $M_n(\mathbb{K})$. Dada $g = g(z)$ a valores complexos e definida numa curva γ no plano complexo, é trivial ver que temos

$$\left(\int_\gamma g(z)dz\right)M = \int_\gamma [g(z)M]dz.$$

Observação 3. Dados dois números complexos, z e λ, com $|z| = r$ e $|\lambda| < r$, por progressão geométrica temos

$$\frac{1}{z-\lambda} = \frac{1}{z\left(1-\frac{\lambda}{z}\right)} = \frac{1}{z}\sum_{n=0}^{+\infty}\left(\frac{z}{\lambda}\right)^n = \sum_{n=0}^{+\infty}\frac{\lambda^n}{z^{n+1}}.$$

Ainda mais, fixado o ponto z, tal série converge absolutamente em todo ponto λ da bola aberta $B(0,r) = \{w \in \mathbb{C} : |w| < r\}$ e converge uniformemente, na variável λ, em todo disco fechado $D(0,\rho) = \{w \in \mathbb{C} : |w| \leq \rho\}$ contido na bola $B(0,r)$.

Observação 4. Seja $\{e_1,\ldots,e_n\}$ a base canônica de $\mathbb{K}^n$. Identifiquemos $A \in M_n(\mathbb{K})$ com o operador linear $A : \mathbb{K}^n \to \mathbb{K}^n$ definido por $A(v) = Av$, com $v \in M_{n\times 1}(\mathbb{K})$ um vetor coluna. Identifiquemos $M_{n\times 1}(\mathbb{K}) \equiv \mathbb{K}^n$ e $A(v) \equiv Av$. Introduzamos a norma

$$\|A\| = \sup\{|Av| : v \in \mathbb{K}^n \text{ e } |v| = 1\}.$$

Fixemos R, com $R > \|A\|$. Consideremos um número λ com $|\lambda| = R$. Então, argumentando analogamente ao que foi acima para o número $1/(z-\lambda)$ é fácil ver que no espaço de matrizes $M_n(\mathbb{K})$ vale a seguinte fórmula para série de matrizes,

$$(\lambda I - A)^{-1} = \sum_{n=0}^{+\infty}\frac{A^n}{\lambda^{n+1}}.$$

Notemos que a convergência é absoluta. Ainda mais, a convergência é uniforme na variável complexa λ, sobre a circunferência $\Gamma = \{w \in \mathbb{C} : |w| = R\}$.

Consideremos em $\mathbb{K}^n$ a norma euclidiana (i.e., a norma usual). Escrevendo $A = (a_{jk})_{1\leq j,k\leq n}$ e e_k para o k-ésimo vetor canônico em $\mathbb{K}^n$, valem as desigualdades

$$|a_{jk}| \leq |A(e_k)| \leq \|A\| \leq \sqrt{\sum_{l,m}|a_{lm}|^2}.$$

[Se $v = v_1e_1 + \cdots + v_ne_n$, pela desigualdade de Cauchy-Schwartz em $\mathbb{R}^n$ temos $|Av| \leq |v_1||Ae_1| + \cdots + |v_n||Ae_n| \leq \sqrt{v_1^2 + \cdots + v_n^2}\sqrt{|Ae_1|^2 + \cdots + |Ae_n|^2} = |v|^2\sqrt{\sum_l a_{l1}^2 + \cdots + \sum_l a_{ln}^2}$ $= |v|^2\sqrt{\sum_{l,m} a_{lm}^2}$. Isto prova a terceira desigualdade acima. As demais são triviais.]

Logo, uma sequência (e uma série) de matrizes converge uniformemente em $M_n(\mathbb{K})$ se e só se ocorre a convergência uniforme coordenada a coordenada.

CONCLUSÃO. (Extraída do artigo de McCarthy [3]).

Seja $k = 0, 1, 2, \ldots$. Temos então

$$(zI - A)^{-1}A^k = \sum_{n=0}^{+\infty} \frac{A^{n+k}}{z^{n+1}}$$

e

$$\begin{aligned}
\int_\Gamma (zI - A)^{-1}A^k dz &= \int_\Gamma \left(\sum_{n=0}^{+\infty} \frac{A^{n+k}}{z^{n+1}} \right) dz \\
&= \sum_{n=0}^{+\infty} \int_\Gamma \frac{A^{n+k}}{z^{n+1}} dz \\
&= \sum_{n=0}^{+\infty} \left(\int_\Gamma \frac{1}{z^{n+1}} dz \right) A^{n+k} \\
&= 2\pi i A^k.
\end{aligned}$$

Donde segue

$$A^k = \frac{1}{2\pi i} \int_\Gamma (zI - A)A^k dz.$$

Portanto, se $p = p(\lambda)$ é o polinômio característo de A, obtemos o seguinte caso particular da fórmula integral de Cauchy,

$$\boxed{p(A) = \frac{1}{2\pi i} \int_\Gamma p(z)(zI - A)^{-1} dz}$$

Reintroduzindo a definição do polinômio característico e utilizando o lema sobre matrizes adjuntas enunciado no início desta seção, encontramos

$$\begin{aligned}
p(A) &= \frac{1}{2\pi i} \int_\Gamma [\det(zI - A)](zI - A)^{-1} dz \\
&= \frac{1}{2\pi i} \int_\Gamma \operatorname{adj}(zI - A) dz
\end{aligned}$$

Por fim, notemos que cada entrada da matriz adjunta $\operatorname{adj}(zI - A)$ é um polinômio na variável z e assim sua integral ao longo da curva fechada Γ é zero. Segue então

$$p(A) = 0 \clubsuit$$

Oswaldo Rio Branco de Oliveira

REFERÊNCIAS

1. Bueno, H. P., *Funções de Matrizes (Versão Preliminar, 2002)* - Universidade Federal de Minas Gerais - Departamento de Matemática. Disponível em `http://www.researchgate.net/publication/268338571_Funes_de_Matrizes`

2. Fraleigh, J. B., *A First Course In Abstract Algebra*, 7th ed., Addison-Wesley, 2003.

3. McCarthy, Charles A., *The Cayley-Hamilton Theorem*, The American Mathematical Monthly, Vol. **82**, No. 4 (apr., 1975), pp. 390–391.

4. Oliveira, Oswaldo R. B. de, *The Fundamental Theorem of Algebra: An elementary and direct proof*, Math. Intelligencer **33** (2), 1-2, (2011).

5. Oliveira, Oswaldo R. B. de, *The Fundamental Theorem of Algebra: from the four basic operations*, The American Mathematical Monthly, Vol. **119**, pp. 753-758, (2012).

6. Ulhoa, F. C. e Lourenço, M. L., *Um Curso de Álgebra Linear*, Edusp, segunda edição (2013).

Departamento de Matemática
Universidade de São Paulo
e-mail: oliveira@ime.usp.br
`http://www.ime.usp.br/~oliveira`