Cayley–Hamilton theorem

Alexey Muranov

11th July 2026

1 Introduction

Theorem (Cayley–Hamilton theorem)**.** *Let R be a unital commutative ring (that is, with the unit 1_R). Let A be a square matrix with coefficients in R and χ_A be the characteristic polynomial of A. Then $\chi_A(A)$ is the zero matrix.*

There is a very short proof of this theorem based on the following two ingredients:

(1) the canonical *isomorphism* $R[X]^{n\times n} \simeq R^{n\times n}[X]$, that allows to "identify" the matrices $n \times n$ with coefficients from $R[X]$ and the polynomials in X with coefficients from $R^{n\times n}$ (the "identification" is between matrices with polynomial coefficients and polynomials with matrix coefficients),

(2) certain *left actions* of $R^{n\times n}[X]$ (and thus of $R[X]^{n\times n}$ as well) on $R^{n\times n}$, which are associated to elements of $R^{n\times n}$.

This proof essentially follows the one in *Algebra* by Serge Lang.[1]

2 Preliminaries

For each square matrix A with coefficients in a unital commutative ring,

$$(\operatorname{adj} A)A = I \det A,$$

where I is the identity matrix (of the same size as A).

Let R be a unital commutative ring. Then $R[X]$ is such as well. Let A be a square matrix with coefficients from R and I be the identity matrix of the same size. Let us consider the square matrix with coefficients from $R[X]$ given by the expression $A - IX$. We have:

$$\big(\operatorname{adj}(A - IX)\big)(A - IX) = I \det(A - IX) = I\chi_A.$$

[1] Serge Lang. *Algebra*. 3rd ed. Graduate Texts in Mathematics 211. Originally published by Addison-Wesley, 1993. New York, NY: Springer, 2002. xv+914. DOI: 10.1007/978-1-4613-0041-0.

3 Proof of Cayley–Hamilton theorem

In this section, R is a unital commutative ring, n is a natural number, A is an $n \times n$ matrix with coefficients from R, I is the $n \times n$ identity matrix, O is the $n \times n$ zero matrix ($A, I, O \in R^{n\times n}$, $n \in \mathbf{N}$).

Given arbitrary $\boldsymbol{B} \in R[X]^{n\times n}$ and $M \in R^{n\times n}$, with

$$\begin{aligned}\boldsymbol{B} &= B_0 + B_1 X + \cdots + B_p X^p \\ &= B_0 X^0 + B_1 X^1 + \cdots + B_p X^p, \qquad B_0, \ldots, B_p \in R^{n\times n},\end{aligned}$$

let

$$\begin{aligned}\boldsymbol{B} \lhd_A M &\overset{\text{def}}{=} B_0 M I + B_1 M A + \cdots + B_p M A^p \\ &= B_0 M A^0 + B_1 M A^1 + \cdots + B_p M A^p.\end{aligned}$$

Thus, in particular,

$$\begin{aligned}\boldsymbol{B} \lhd_A I &= B_0 I I + B_1 I A + \cdots + B_p I A^p \\ &= B_0 I + B_1 A + \cdots + B_p A^p = B_0 + B_1 A + \cdots + B_p A^p.\end{aligned}$$

With a view to prove the Cayley–Hamilton theorem, observe that

$$(A - IX) \lhd_A I = AI - IA = A - A = O,$$

and that for every $P = \alpha_0 + \alpha_1 X + \cdots + \alpha_p X^p \in R[X]$,

$$\begin{aligned}IP \lhd_A I &= (\alpha_0 I + \alpha_1 I X + \cdots + \alpha_p I X^p) \lhd_A I \\ &= \alpha_0 I I + \alpha_1 I A + \cdots + \alpha_p I A^p \\ &= \alpha_0 I + \alpha_1 A + \cdots + \alpha_p A^p \\ &= P(A).\end{aligned}$$

We will need a few properties of the operation $(\lhd_A)$, presented as the following lemmas.

Lemma. *The operation $(\lhd_A)$ is linear in second argument:*

$$\boldsymbol{B} \lhd_A (M + N) = \boldsymbol{B} \lhd_A M + \boldsymbol{B} \lhd_A N \qquad \text{and} \qquad \boldsymbol{B} \lhd_A \alpha M = \alpha(\boldsymbol{B} \lhd_A M)$$

for all $\boldsymbol{B} \in R[X]^{n\times n}$, $M, N \in R^{n\times n}$, and $\alpha \in R$.

Exercise. Prove this lemma.

Lemma. *The operation $(\lhd_A)$ is linear in first argument:*

$$(\boldsymbol{B} + \boldsymbol{C}) \lhd_A M = \boldsymbol{B} \lhd_A M + \boldsymbol{C} \lhd_A M \qquad \text{and} \qquad \alpha\boldsymbol{B} \lhd_A M = \alpha(\boldsymbol{B} \lhd_A M)$$

for all $M \in R^{n\times n}$, $\boldsymbol{B}, \boldsymbol{C} \in R[X]^{n\times n}$, and $\alpha \in R$.

Exercise. Prove this lemma.

Lemma. *For all* $M \in R^{n\times n}$ *and* $\boldsymbol{B}, \boldsymbol{C} \in R[X]^{n\times n}$,

$$\boldsymbol{B}\boldsymbol{C} \triangleleft_A M = \boldsymbol{B} \triangleleft_A (\boldsymbol{C} \triangleleft_A M).$$

Outline of a proof. It is enough to deal with the case where $\boldsymbol{B} = BX^p$ and $\boldsymbol{C} = CX^q$, with $B, C \in R^{n\times n}$ and $p, q \in \mathbf{N}$. The general case will follow by bilinearity. And we have:

$$BX^pCX^q \triangleleft_A M = BCX^pX^q \triangleleft_A M = BCX^{p+q} \triangleleft_A M = BCMA^{p+q},$$

and

$$BX^p \triangleleft_A (CX^q \triangleleft_A M) = BX^p \triangleleft_A CMA^q = BCMA^qA^p = BCMA^{q+p}. \qquad \square$$

Proof of Cayley–Hamilton theorem.

$$\begin{aligned}
\chi_A(A) &= I\chi_A \triangleleft_A I \\
&= I\det(A - IX) \triangleleft_A I \\
&= \big(\operatorname{adj}(A - IX)\big)(A - IX) \triangleleft_A I \\
&= \big(\operatorname{adj}(A - IX)\big) \triangleleft_A \big((A - IX) \triangleleft_A I\big) \\
&= \big(\operatorname{adj}(A - IX)\big) \triangleleft_A O \\
&= O.
\end{aligned} \qquad \square$$

4 Generalisation

The method used above to prove the Cayley–Hamilton theorem allows to prove seemingly more general statements, like the following proposition:

Proposition. *Let* $A_1, \ldots, A_m, B_1, \ldots, B_m \in R^{n\times n}$ *be such that*:

(1) $B_1A_1 + \cdots + B_mA_m = O$,

(2) $A_1, \ldots, A_m$ *commute pairwise.*

Let

$$\begin{aligned}
\boldsymbol{B} &= B_1X_1 + \cdots + B_mX_m \in R[X_1, \ldots, X_m]^{n\times n}, \\
P &= \det \boldsymbol{B} \in R[X_1, \ldots, X_m].
\end{aligned}$$

Then

$$P(A_1, \ldots, A_m) = O.$$

Example. Let $A, B \in R^{n\times n}$ be such that $BA = AB$. Let

$$P = \det(BX - AY) \in R[X, Y].$$

Then $P(A, B) = O$.

Exercise. Let $A = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 3 & -4 \\ 4 & 3 \end{pmatrix}$ ($A, B \in \mathbf{Z}^{2\times 2}$). Let

$$P = \det(BX - AY) \in \mathbf{Z}[X, Y].$$

(1) Compute AB and BA.

(2) Compute $P(A, B)$.

Idea of a proof of the proposition. We may define the operation $(\triangleleft_{A_1,\dots,A_m})$ so that for all $\boldsymbol{B} \in R[X_1, \dots, X_m]^{n\times n}$ and $M \in R^{n\times n}$, we have $\boldsymbol{B} \triangleleft_{A_1,\dots,A_m} M \in R^{n\times n}$, and so that

$$\begin{aligned} &P(A_1, \dots, A_m) \\ &\quad = IP \triangleleft_{A_1,\dots,A_m} I \\ &\quad = I \det(B_1X_1 + \cdots + B_mX_m) \triangleleft_{A_1,\dots,A_m} I \\ &\quad = \big(\operatorname{adj}(B_1X_1 + \cdots + B_mX_m)\big)(B_1X_1 + \cdots + B_mX_m) \triangleleft_{A_1,\dots,A_m} I \\ &\quad = \big(\operatorname{adj}(B_1X_1 + \cdots + B_mX_m)\big) \triangleleft_{A_1,\dots,A_m} \big((B_1X_1 + \cdots + B_mX_m) \triangleleft_{A_1,\dots,A_m} I\big) \\ &\quad = \big(\operatorname{adj}(B_1X_1 + \cdots + B_mX_m)\big) \triangleleft_{A_1,\dots,A_m} O \\ &\quad = O. \end{aligned}$$

□

Proposition. *Let $A_1, \dots, A_m, B_1, \dots, B_m \in R^{n\times n}$ be such that:*

(1) $A_1B_1 + \cdots + A_mB_m = O$,

(2) *$A_1, \dots, A_m$ commute pairwise.*

Let

$$\begin{aligned} &\boldsymbol{B} = B_1X_1 + \cdots + B_mX_m = X_1B_1 + \cdots + X_mB_m \in R[X_1, \dots, X_m]^{n\times n}, \\ &P = \det \boldsymbol{B} \in R[X_1, \dots, X_m]. \end{aligned}$$

Then

$$P(A_1, \dots, A_m) = O.$$

Idea of a proof. We may define the operation $(\triangleright_{A_1,\dots,A_m})$ so that for all $\boldsymbol{B} \in R[X_1, \dots, X_m]^{n\times n}$ and $M \in R^{n\times n}$, we have $M \triangleright_{A_1,\dots,A_m} \boldsymbol{B} \in R^{n\times n}$, and so that

$$\begin{aligned} &P(A_1, \dots, A_m) \\ &\quad = I \triangleright_{A_1,\dots,A_m} PI \\ &\quad = I \triangleright_{A_1,\dots,A_m} \det(X_1B_1 + \cdots + X_mB_m)I \\ &\quad = I \triangleright_{A_1,\dots,A_m} (X_1B_1 + \cdots + X_mB_m) \operatorname{adj}(X_1B_1 + \cdots + X_mB_m) \\ &\quad = \big(I \triangleright_{A_1,\dots,A_m} (X_1B_1 + \cdots + X_mB_m)\big) \triangleright_{A_1,\dots,A_m} \operatorname{adj}(X_1B_1 + \cdots + X_mB_m) \\ &\quad = O \triangleright_{A_1,\dots,A_m} \operatorname{adj}(X_1B_1 + \cdots + X_mB_m) \\ &\quad = O. \end{aligned}$$

□

How about other generalisations? If one looks for other generalisations of the Cayley–Hamilton theorem, the one may ask himself, for example, the following question:

> Is it true that if $A, B, C \in R^{n\times n}$, $BA - AC = O$, and we set
>
> $$P = \det(BX - XC) = \det(BX - CX) = \big(\det(B - C)\big)X^n \in R[X],$$
>
> then $P(A) = O$?

This is not the case for $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $C = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, for example.

References

Lang, Serge. *Algebra*. 3rd ed. Graduate Texts in Mathematics 211. Originally published by Addison-Wesley, 1993. New York, NY: Springer, 2002. xv+914. DOI: 10.1007/978-1-4613-0041-0.