

9.6 The Cayley–Hamilton Theorem

An operator on an n -dimensional space now carries two monic polynomials. The characteristic polynomial p_T comes from determinants and has degree exactly n . The minimal polynomial m_T comes from the ideal of polynomials that kill T , and the previous section showed that it has the same roots as p_T . Chapter 7 stated, without proof, the fact that ties the two together: every square matrix satisfies its own characteristic polynomial. Chapter 7 also warned against a one-line fake proof of that statement. This section gives a real proof, built on the adjugate identity of Chapter 7 read over $F[x]$. It then draws the consequences: m_T divides p_T , inverses are polynomials, and high powers of a matrix reduce to low ones.

9.6.1 The statement, and a proof that is not one

Chapter 6 proved that every operator on an n -dimensional space is killed by some nonzero polynomial of degree at most n^2 (Theorem 6.7.2). The bound came from counting: $\mathcal{L}(V)$ has dimension n^2 . Counting cannot do better, yet in every example so far a polynomial of degree at most n sufficed. The characteristic polynomial always has degree n , and it is the natural candidate.

Theorem 9.6.1. Cayley–Hamilton Theorem

Let $n \geq 1$ and $\mathbf{A} \in M_n(F)$. Then

$$p_{\mathbf{A}}(\mathbf{A}) = \mathbf{O}.$$

Likewise, if V is a finite-dimensional vector space over F with $\dim V \geq 1$ and $T \in \mathcal{L}(V)$, then $p_T(T) = 0$.

Here $p_{\mathbf{A}}(\mathbf{A})$ means what it always means (Definition 0.9.14): if $p_{\mathbf{A}} = x^n + c_{n-1}x^{n-1} + \cdots + c_1x + c_0$, then

$$p_{\mathbf{A}}(\mathbf{A}) = \mathbf{A}^n + c_{n-1}\mathbf{A}^{n-1} + \cdots + c_1\mathbf{A} + c_0\mathbf{I}_n,$$

an $n \times n$ matrix, with the constant term becoming $c_0\mathbf{I}_n$.

For $n = 2$ the theorem can be checked by hand. With $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $p_{\mathbf{A}} = x^2 - (\operatorname{tr} \mathbf{A})x + \det \mathbf{A}$, and Exercise (C2: Nilpotent 2×2 matrices) (b) multiplied out

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

For $n = 3$, the same brute force already involves cubic expressions in nine entries, and for general n it is hopeless. We need an argument. There is a tempting one, and we name it before it can do any harm.

Warning. The fake proof of Cayley–Hamilton, again. “Since $p_{\mathbf{A}}(x) = \det(x\mathbf{I} - \mathbf{A})$, substitute $x = \mathbf{A}$ to get $p_{\mathbf{A}}(\mathbf{A}) = \det(\mathbf{A}\mathbf{I} - \mathbf{A}) = \det \mathbf{O} = 0$.” Chapter 7 (the warning at the end of its §10) explained why this is not a proof. The left side $p_{\mathbf{A}}(\mathbf{A})$ is a **matrix**, while $\det(\cdots)$ is a **scalar**, so the equation compares objects of different kinds. And substituting $\mathbf{A}$ for x **inside the entries** of $x\mathbf{I} - \mathbf{A}$ is not what substituting into a polynomial means: the entry $x - a_{11}$ would have to become the matrix $\mathbf{A} - a_{11}\mathbf{I}$, and the determinant of a matrix whose entries are matrices is undefined. The same “argument” applied to $q(x) = \operatorname{tr}(x\mathbf{I} - \mathbf{A}) = nx - \operatorname{tr} \mathbf{A}$ would prove $q(\mathbf{A}) = \mathbf{O}$, which is false for $\mathbf{A} = \operatorname{diag}(1, 0)$: there

$q(\mathbf{A}) = 2\mathbf{A} - \mathbf{I} = \text{diag}(1, -1)$. Any correct proof must keep x a polynomial variable until the very end, and only then put a matrix in its place.

9.6.2 A proof through the adjugate

The adjugate identity of Chapter 7 says $\mathbf{M} \text{adj } \mathbf{M} = (\det \mathbf{M}) \mathbf{I}$ for every square matrix $\mathbf{M}$, and the remark after Theorem 7.6.2 notes that its proof works over any commutative ring. Over $F[x]$, with $\mathbf{M} = x\mathbf{I} - \mathbf{A}$, it reads

$$(x\mathbf{I} - \mathbf{A}) \text{adj}(x\mathbf{I} - \mathbf{A}) = p_{\mathbf{A}}(x) \mathbf{I}. \quad (1)$$

This is an identity between matrices of **polynomials**, and it contains $p_{\mathbf{A}}$ on the right. The proof extracts $p_{\mathbf{A}}(\mathbf{A})$ from it legally.

Idea. Try $n = 2$ first. By Exercise (C3: An adjugate over $F[x]$) (b) (done there over $\mathbb{R}$, but the computation works over any field), $\text{adj}(x\mathbf{I} - \mathbf{A}) = x\mathbf{I} - \text{adj } \mathbf{A}$, a polynomial in x with **matrix coefficients** $\mathbf{B}_1 = \mathbf{I}$ and $\mathbf{B}_0 = -\text{adj } \mathbf{A}$. Expanding the left side of (1),

$$(x\mathbf{I} - \mathbf{A})(x\mathbf{I} - \text{adj } \mathbf{A}) = x^2 \mathbf{I} - x(\mathbf{A} + \text{adj } \mathbf{A}) + \mathbf{A} \text{adj } \mathbf{A}.$$

Comparing the coefficients of x^2 , x , and 1 with those of $(x^2 - (\text{tr } \mathbf{A})x + \det \mathbf{A})\mathbf{I}$ gives three matrix equations:

$$\mathbf{I} = \mathbf{I}, \quad -(\mathbf{A} + \text{adj } \mathbf{A}) = -(\text{tr } \mathbf{A}) \mathbf{I}, \quad \mathbf{A} \text{adj } \mathbf{A} = (\det \mathbf{A}) \mathbf{I}.$$

Now multiply the first by $\mathbf{A}^2$, the second by $\mathbf{A}$, the third by $\mathbf{I}$, and add. The right sides add up to $\mathbf{A}^2 - (\text{tr } \mathbf{A})\mathbf{A} + (\det \mathbf{A})\mathbf{I} = p_{\mathbf{A}}(\mathbf{A})$. The left sides add up to $\mathbf{A}^2 - \mathbf{A}^2 - \mathbf{A} \text{adj } \mathbf{A} + \mathbf{A} \text{adj } \mathbf{A}$, in which every term cancels against its neighbor. So $p_{\mathbf{A}}(\mathbf{A}) = \mathbf{O}$. This is the whole proof; for general n there are $n + 1$ equations and the cancellation is a telescoping sum. The variable x is never replaced by a matrix: we compare coefficients first, and only then multiply the resulting **matrix** equations by powers of $\mathbf{A}$.

Proof. Let $p_{\mathbf{A}} = c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0$ with $c_n = 1$ (Theorem 7.10.2).

Step 1: the adjugate as a polynomial with matrix coefficients. If $n = 1$, then $\text{adj}(x\mathbf{I} - \mathbf{A}) = (1)$ by Definition 7.6.2. If $n \geq 2$, each entry of $\text{adj}(x\mathbf{I} - \mathbf{A})$ is a cofactor $\pm \det \mathbf{N}$, where $\mathbf{N}$ is an $(n - 1) \times (n - 1)$ submatrix of $x\mathbf{I} - \mathbf{A}$ (Definition 7.6.2, Definition 7.6.1). Every entry of $\mathbf{N}$ lies in $F[x]_{\leq 1}$. A product of $n - 1$ such polynomials lies in $F[x]_{\leq n-1}$ (Theorem 0.8.3, the zero polynomial being harmless), and so does a sum of such products (Theorem 0.8.2); in particular the Leibniz sum $\det \mathbf{N}$ does (Definition 7.4.1). In both cases every entry of $\text{adj}(x\mathbf{I} - \mathbf{A})$ lies in $F[x]_{\leq n-1}$. For $0 \leq k \leq n - 1$, let $\mathbf{B}_k \in M_n(F)$ be the matrix whose (i, j) -entry is the coefficient of x^k in the (i, j) -entry of $\text{adj}(x\mathbf{I} - \mathbf{A})$. Then, entry by entry,

$$\text{adj}(x\mathbf{I} - \mathbf{A}) = \sum_{k=0}^{n-1} x^k \mathbf{B}_k.$$

Put $\mathbf{B}_{-1} = \mathbf{B}_n = \mathbf{O}$ for convenience.

Step 2: compare coefficients. By (1), which holds by Theorem 7.6.2 over the commutative ring $F[x]$ (the remark after it), the (i, j) -entries of both sides of (1) agree. On the left, by

Definition 0.9.9,

$$\sum_{l=1}^n (x\delta_{il} - a_{il}) \sum_{k=0}^{n-1} (\mathbf{B}_k)_{lj} x^k = \sum_{k=0}^{n-1} (\mathbf{B}_k)_{ij} x^{k+1} - \sum_{k=0}^{n-1} (\mathbf{A}\mathbf{B}_k)_{ij} x^k,$$

where $\delta_{il} = 1$ if $i = l$ and 0 otherwise. So for $0 \leq k \leq n$, the coefficient of x^k is $(\mathbf{B}_{k-1})_{ij} - (\mathbf{A}\mathbf{B}_k)_{ij}$. On the right the (i, j) -entry is $p_{\mathbf{A}} \delta_{ij}$, whose coefficient of x^k is $c_k \delta_{ij}$. Two polynomials are equal exactly when their coefficients are equal (Definition 0.8.1). Since this holds for every (i, j) ,

$$\mathbf{B}_{k-1} - \mathbf{A}\mathbf{B}_k = c_k \mathbf{I} \quad (0 \leq k \leq n). \quad (2)$$

Step 3: telescope. Multiply equation (2) for k on the left by $\mathbf{A}^k$, and add the $n+1$ equations. On the right we get $\sum_{k=0}^n c_k \mathbf{A}^k = p_{\mathbf{A}}(\mathbf{A})$ by Definition 0.9.14. On the left, writing $\mathbf{S}_k = \mathbf{A}^k \mathbf{B}_{k-1}$ for $0 \leq k \leq n+1$,

$$\sum_{k=0}^n (\mathbf{A}^k \mathbf{B}_{k-1} - \mathbf{A}^{k+1} \mathbf{B}_k) = \sum_{k=0}^n (\mathbf{S}_k - \mathbf{S}_{k+1}) = \mathbf{S}_0 - \mathbf{S}_{n+1} = \mathbf{B}_{-1} - \mathbf{A}^{n+1} \mathbf{B}_n = \mathbf{O},$$

since $\mathbf{B}_{-1} = \mathbf{B}_n = \mathbf{O}$. Hence $p_{\mathbf{A}}(\mathbf{A}) = \mathbf{O}$.

The operator version. Let $\mathcal{B}$ be a basis of V and $\mathbf{A} = [T]_{\mathcal{B}}^{\mathcal{B}}$. By Definition 7.10.2, $p_T = p_{\mathbf{A}}$. By Corollary 4.1.3, $[p_T(T)]_{\mathcal{B}}^{\mathcal{B}} = p_T(\mathbf{A}) = p_{\mathbf{A}}(\mathbf{A}) = \mathbf{O}$, so $p_T(T) = 0$ (Theorem 4.1.4). This proves the theorem. ■

Compare with the fake proof. There, $\mathbf{A}$ was pushed into the entries of $x\mathbf{I} - \mathbf{A}$. Here, (1) is first unpacked into the $n+1$ matrix equations (2), which contain no x at all, and $\mathbf{A}$ enters only when we multiply those equations by $\mathbf{A}^k$. Every multiplication is on the **left**, so no commutativity of matrices was used; the only commutativity used is inside $F[x]$, in the adjugate identity.

Nothing in the proof divides. The determinant, the adjugate identity, coefficient comparison, and matrix multiplication all make sense over a commutative ring, and [Exercise \(C1: Cayley–Hamilton over a commutative ring\)](#) asks you to check that the theorem holds, with the same proof, for matrices with entries in a commutative ring such as $\mathbb{Z}$ or $\mathbb{R}[t]$.

The proof also produced more than it was asked for. Equation (2) with $k = n$ says $\mathbf{B}_{n-1} = \mathbf{I}$, and the others say $\mathbf{B}_{k-1} = \mathbf{A}\mathbf{B}_k + c_k \mathbf{I}$. Working down from $\mathbf{B}_{n-1}$, every $\mathbf{B}_k$ is a polynomial in $\mathbf{A}$; [Exercise \(C2: The adjugate is a polynomial in the matrix\)](#) turns this into a formula for $\text{adj } \mathbf{A}$.

9.6.3 Consequences

The first consequence is the promised comparison of the two polynomials.

Corollary 9.6.2. The Minimal Polynomial Divides the Characteristic Polynomial

Let V be finite-dimensional with $\dim V = n \geq 1$, and let $T \in \mathcal{L}(V)$.

- (a) $m_T \mid p_T$; in particular $1 \leq \deg m_T \leq n$.

(b) If p_T splits over F , say $p_T = (x - \lambda_1)^{a_1} \cdots (x - \lambda_k)^{a_k}$ with distinct λ_i , then

$$m_T = (x - \lambda_1)^{s_1} \cdots (x - \lambda_k)^{s_k} \quad \text{with } 1 \leq s_i \leq a_i = a_T(\lambda_i).$$

The same holds for $\mathbf{A} \in M_n(F)$ with $m_{\mathbf{A}}$ and $p_{\mathbf{A}}$. In particular, over $\mathbb{C}$, (b) applies to every $\mathbf{A} \in M_n(\mathbb{C})$.

Proof.

- (a) By Theorem 9.6.1, $p_T(T) = 0$, so $m_T \mid p_T$ by Theorem 9.5.1. Since p_T is monic of degree n (Theorem 7.10.2), $\deg m_T \leq n$ by Proposition 6.1.2 (c), and $\deg m_T \geq 1$ by Theorem 9.5.1, as $V \neq \{0\}$.
- (b) A polynomial $x - \lambda$ of degree 1 is irreducible: if $x - \lambda = gh$, then $\deg g + \deg h = 1$ (Theorem 0.8.3), so one factor is constant. The $x - \lambda_i$ are distinct and monic. By (a) and Lemma 9.5.2, $m_T = \prod_i (x - \lambda_i)^{s_i}$ with $0 \leq s_i \leq a_i$. Each λ_i is a root of p_T , hence an eigenvalue (Theorem 9.2.1), hence a root of m_T (Theorem 9.5.3), so $s_i \geq 1$. Finally $a_i = \text{mult}_{\lambda_i}(p_T) = a_T(\lambda_i)$ by Lemma 6.4.1, since the other factors do not vanish at λ_i . The matrix case is the operator $T_{\mathbf{A}}$, and over $\mathbb{C}$ every $p_{\mathbf{A}}$ splits (Corollary 6.5.5). ■

This pays off a promise from Chapter 6: the bound n^2 of Theorem 6.7.2 improves to n , and p_T is an explicit annihilating polynomial of degree exactly n . It also makes Method 2 of the previous section a finite search that always succeeds. For a matrix with $p_{\mathbf{A}} = (x - 2)^2(x - 3)$, the corollary leaves exactly two candidates, $(x - 2)(x - 3)$ and $(x - 2)^2(x - 3)$, and one matrix product decides between them.

Warning. Cayley–Hamilton gives an annihilating polynomial, not the minimal one. From $p_{\mathbf{A}}(\mathbf{A}) = \mathbf{O}$ we may **not** conclude $m_{\mathbf{A}} = p_{\mathbf{A}}$; only $m_{\mathbf{A}} \mid p_{\mathbf{A}}$. For $\mathbf{A} = \text{diag}(1, 1, 2)$, $p_{\mathbf{A}} = (x - 1)^2(x - 2)$ kills $\mathbf{A}$, and so does the proper divisor $(x - 1)(x - 2) = m_{\mathbf{A}}$. Nor may we conclude that every divisor of $p_{\mathbf{A}}$ with the right roots kills $\mathbf{A}$: for $\mathbf{J} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $x - 1$ has the right root and does not.

The second consequence expresses inverses through powers.

Corollary 9.6.3. The Inverse Is a Polynomial in the Matrix

Let $n \geq 1$, $\mathbf{A} \in M_n(F)$, and $p_{\mathbf{A}} = x^n + c_{n-1}x^{n-1} + \cdots + c_1x + c_0$. Then $c_0 = (-1)^n \det \mathbf{A}$. If $\mathbf{A}$ is invertible, then $c_0 \neq 0$ and

$$\mathbf{A}^{-1} = -\frac{1}{c_0}(\mathbf{A}^{n-1} + c_{n-1}\mathbf{A}^{n-2} + \cdots + c_2\mathbf{A} + c_1\mathbf{I}).$$

(For $n = 1$ the bracket is $\mathbf{I}$ alone, and the formula reads $\mathbf{A}^{-1} = -c_0^{-1}\mathbf{I}$.) The same holds for an invertible operator T on a space of dimension $n \geq 1$, with id_V in place of $\mathbf{I}$.

Proof. By Theorem 7.10.2, $c_0 = (-1)^n \det \mathbf{A}$, which is nonzero when $\mathbf{A}$ is invertible (Theorem 7.5.3). By Theorem 9.6.1,

$$\mathbf{A}(\mathbf{A}^{n-1} + c_{n-1}\mathbf{A}^{n-2} + \cdots + c_1\mathbf{I}) = \mathbf{A}^n + c_{n-1}\mathbf{A}^{n-1} + \cdots + c_1\mathbf{A} = -c_0\mathbf{I}.$$

Multiplying by $-c_0^{-1}$, the matrix in parentheses times $-c_0^{-1}$ is a right inverse of $\mathbf{A}$, hence the inverse by Theorem 3.4.5. For an operator, apply this to $[T]_{\mathcal{B}}^{\mathcal{B}}$ and use Corollary 4.1.3. ■

So inverting an $n \times n$ matrix never requires more than its first $n - 1$ powers and the coefficients of $p_{\mathbf{A}}$. The theorem also reduces **every** power of $\mathbf{A}$ to the first $n - 1$. Divide x^N by $p_{\mathbf{A}}$ (Theorem 6.1.1): $x^N = qp_{\mathbf{A}} + r$ with $\deg r < n$. Then by Theorem 6.7.1 and Theorem 9.6.1,

$$\mathbf{A}^N = q(\mathbf{A})p_{\mathbf{A}}(\mathbf{A}) + r(\mathbf{A}) = r(\mathbf{A}).$$

When $p_{\mathbf{A}}$ has n distinct roots $\lambda_1, \dots, \lambda_n$, the remainder can be found without long division. Substituting $x = \lambda_i$ in $x^N = qp_{\mathbf{A}} + r$ gives $r(\lambda_i) = \lambda_i^N$ for each i , and these n conditions determine $r \in F[x]_{\leq n-1}$ by Lagrange interpolation (Theorem 6.6.1). This substitution is legal: λ_i is a **scalar** plugged into an identity of polynomials.

Example. Cayley–Hamilton for a 3×3 Matrix

Let

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \in M_3(\mathbb{Q}).$$

Compute $p_{\mathbf{A}}$ and verify $p_{\mathbf{A}}(\mathbf{A}) = \mathbf{O}$. Then use it to find $\mathbf{A}^{-1}$ and $\mathbf{A}^5$, and write down the matrices $\mathbf{B}_k$ of the proof.

Solution. Characteristic polynomial. Expand $\det(x\mathbf{I} - \mathbf{A})$ along the first row (Theorem 7.6.1 over $\mathbb{Q}[x]$):

$$\begin{aligned} p_{\mathbf{A}} &= \det \begin{pmatrix} x-1 & -1 & 0 \\ 0 & x & -1 \\ -1 & 0 & x-1 \end{pmatrix} \\ &= (x-1) \det \begin{pmatrix} x & -1 \\ 0 & x-1 \end{pmatrix} + (-1)(-1)^{1+2} \det \begin{pmatrix} 0 & -1 \\ -1 & x-1 \end{pmatrix}. \end{aligned}$$

The two minors are $x(x-1)$ and -1 , so $p_{\mathbf{A}} = x(x-1)^2 - 1 = x^3 - 2x^2 + x - 1$. As a check, $\text{tr } \mathbf{A} = 2$ matches the coefficient -2 of x^2 , and $c_0 = -1 = (-1)^3 \det \mathbf{A}$ says $\det \mathbf{A} = 1$.

Verification. Multiplying out,

$$\mathbf{A}^2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 2 & 1 & 1 \end{pmatrix}, \quad \mathbf{A}^3 = \mathbf{A}\mathbf{A}^2 = \begin{pmatrix} 2 & 1 & 2 \\ 2 & 1 & 1 \\ 3 & 2 & 2 \end{pmatrix}.$$

Then $\mathbf{A}^3 - 2\mathbf{A}^2 + \mathbf{A} - \mathbf{I}$ has first row $(2 - 2 + 1 - 1, 1 - 2 + 1, 2 - 2 + 0) = (0, 0, 0)$, second row

$(2-2+0, 1-0+0-1, 1-2+1) = (0, 0, 0)$, and third row $(3-4+1, 2-2+0, 2-2+1-1) = (0, 0, 0)$. So $p_{\mathbf{A}}(\mathbf{A}) = \mathbf{O}$, as the theorem predicts.

Inverse. Here $c_0 = -1$, $c_1 = 1$, $c_2 = -2$. By [Corollary 9.6.3](#),

$$\mathbf{A}^{-1} = \mathbf{A}^2 - 2\mathbf{A} + \mathbf{I} = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 1 & -1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Check: the first row of $\mathbf{A} \mathbf{A}^{-1}$ is $(0+1, -1+1, 1-1) = (1, 0, 0)$, and the other rows work the same way. The entries are integers, as they must be for an integer matrix of determinant 1 ([Corollary 7.6.5](#)).

Fifth power. The relation $\mathbf{A}^3 = 2\mathbf{A}^2 - \mathbf{A} + \mathbf{I}$ lets us lower degrees step by step:

$$\mathbf{A}^4 = 2\mathbf{A}^3 - \mathbf{A}^2 + \mathbf{A} = 2(2\mathbf{A}^2 - \mathbf{A} + \mathbf{I}) - \mathbf{A}^2 + \mathbf{A} = 3\mathbf{A}^2 - \mathbf{A} + 2\mathbf{I},$$

$$\mathbf{A}^5 = 3\mathbf{A}^3 - \mathbf{A}^2 + 2\mathbf{A} = 3(2\mathbf{A}^2 - \mathbf{A} + \mathbf{I}) - \mathbf{A}^2 + 2\mathbf{A} = 5\mathbf{A}^2 - \mathbf{A} + 3\mathbf{I}.$$

(Equivalently, $x^5 = (x^2 + 2x + 3)p_{\mathbf{A}} + (5x^2 - x + 3)$.) Hence

$$\mathbf{A}^5 = 5 \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 2 & 1 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} + 3\mathbf{I} = \begin{pmatrix} 7 & 4 & 5 \\ 5 & 3 & 4 \\ 9 & 5 & 7 \end{pmatrix}.$$

The matrices of the proof. By (2), $\mathbf{B}_2 = \mathbf{I}$, $\mathbf{B}_1 = \mathbf{A}\mathbf{B}_2 + c_2\mathbf{I} = \mathbf{A} - 2\mathbf{I}$, and $\mathbf{B}_0 = \mathbf{A}\mathbf{B}_1 + c_1\mathbf{I} = \mathbf{A}^2 - 2\mathbf{A} + \mathbf{I}$. So

$$\begin{aligned} \text{adj}(x\mathbf{I} - \mathbf{A}) &= x^2\mathbf{I} + x(\mathbf{A} - 2\mathbf{I}) + (\mathbf{A}^2 - 2\mathbf{A} + \mathbf{I}) \\ &= \begin{pmatrix} x^2 - x & x - 1 & 1 \\ 1 & x^2 - 2x + 1 & x - 1 \\ x & 1 & x^2 - x \end{pmatrix}, \end{aligned}$$

which one can confirm entry by entry from cofactors of $x\mathbf{I} - \mathbf{A}$; for instance the $(1, 3)$ -entry is the cofactor $C_{31} = \det \begin{pmatrix} -1 & 0 \\ x & -1 \end{pmatrix} = 1$. The last equation of (2), $-\mathbf{A}\mathbf{B}_0 = c_0\mathbf{I} = -\mathbf{I}$, is the statement $\mathbf{A}^{-1} = \mathbf{B}_0$ found above.

Quick check. Let $\mathbf{A} \in M_2(\mathbb{R})$ with $\text{tr } \mathbf{A} = 5$ and $\det \mathbf{A} = 6$. Write $\mathbf{A}^2$ and $\mathbf{A}^3$ as combinations of $\mathbf{A}$ and $\mathbf{I}$, and write $\mathbf{A}^{-1}$ the same way.

Solution. $p_{\mathbf{A}} = x^2 - 5x + 6$, so by [Theorem 9.6.1](#) $\mathbf{A}^2 = 5\mathbf{A} - 6\mathbf{I}$. Then $\mathbf{A}^3 = 5\mathbf{A}^2 - 6\mathbf{A} = 5(5\mathbf{A} - 6\mathbf{I}) - 6\mathbf{A} = 19\mathbf{A} - 30\mathbf{I}$. By [Corollary 9.6.3](#) with $c_0 = 6$ and $c_1 = -5$, $\mathbf{A}^{-1} = -\frac{1}{6}(\mathbf{A} - 5\mathbf{I}) = \frac{1}{6}(5\mathbf{I} - \mathbf{A})$.

Alternatively. When p_T splits, there is a second proof that avoids determinants of polynomial matrices entirely. The next section shows that such an operator has an upper triangular matrix, and a triangular matrix reveals directly why the product $(T - \lambda_1 \text{id}_V) \cdots (T - \lambda_n \text{id}_V)$ is zero. That proof is set out as an exercise there. The adjugate proof has the advantage of working

over every field at once, with no splitting hypothesis.

9.6.4 Exercises

9.6.4.1 A. Check your understanding

Exercise. A1

- State the Cayley–Hamilton Theorem for $\mathbf{A} \in M_n(F)$.
- True or false: $\deg m_{\mathbf{A}} \leq n$ for every $\mathbf{A} \in M_n(F)$, $n \geq 1$. Justify your answer.
- In one sentence, why is “ $p_{\mathbf{A}}(\mathbf{A}) = \det(\mathbf{A}\mathbf{I} - \mathbf{A}) = 0$ ” not a proof?
- Suppose $p_{\mathbf{A}} = (x - 1)^2(x + 2)$. Which polynomials can $m_{\mathbf{A}}$ be?
- True or false: if $p_{\mathbf{A}}(\mathbf{A}) = \mathbf{O}$, then $p_{\mathbf{A}}$ is the minimal polynomial of $\mathbf{A}$. Justify your answer.
- How do you write $\mathbf{A}^{-1}$ as a polynomial in $\mathbf{A}$ when $\mathbf{A}$ is invertible?

Solution.

- If $n \geq 1$ and $\mathbf{A} \in M_n(F)$, then $p_{\mathbf{A}}(\mathbf{A}) = \mathbf{O}$ (Theorem 9.6.1).
- True. By Corollary 9.6.2, $m_{\mathbf{A}} \mid p_{\mathbf{A}}$, and $\deg p_{\mathbf{A}} = n$.
- It substitutes a matrix into the entries of $x\mathbf{I} - \mathbf{A}$ and then equates the matrix $p_{\mathbf{A}}(\mathbf{A})$ with the scalar $\det \mathbf{O}$, neither of which is legitimate.
- By Corollary 9.6.2 (b), $m_{\mathbf{A}} = (x - 1)^s(x + 2)^t$ with $1 \leq s \leq 2$ and $t = 1$: so $(x - 1)(x + 2)$ or $(x - 1)^2(x + 2)$. Both occur, for $\text{diag}(1, 1, -2)$ and for $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \oplus (-2)$ (Proposition 9.5.6).
- False. $p_{\mathbf{I}_2} = (x - 1)^2$ kills $\mathbf{I}_2$, but $m_{\mathbf{I}_2} = x - 1$.
- If $p_{\mathbf{A}} = x^n + c_{n-1}x^{n-1} + \cdots + c_0$, then $c_0 = (-1)^n \det \mathbf{A} \neq 0$ and $\mathbf{A}^{-1} = -c_0^{-1}(\mathbf{A}^{n-1} + c_{n-1}\mathbf{A}^{n-2} + \cdots + c_1\mathbf{I})$ (Corollary 9.6.3).

9.6.4.2 B. Practice

Exercise. B1: Inverse and a hundredth power

Let $\mathbf{A} = \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix} \in M_2(\mathbb{R})$.

- Find $p_{\mathbf{A}}$, and use the Cayley–Hamilton Theorem to compute $\mathbf{A}^{-1}$.
- Hence find $\mathbf{A}^{100}$ as a combination $a\mathbf{A} + b\mathbf{I}$, and write it as a matrix.

Solution.

- $\text{tr } \mathbf{A} = 3$ and $\det \mathbf{A} = 2$, so $p_{\mathbf{A}} = x^2 - 3x + 2 = (x - 1)(x - 2)$. By Theorem 9.6.1,

$\mathbf{A}^2 - 3\mathbf{A} + 2\mathbf{I} = \mathbf{O}$, that is, $\mathbf{A}(3\mathbf{I} - \mathbf{A}) = 2\mathbf{I}$. By [Corollary 9.6.3](#),

$$\mathbf{A}^{-1} = \frac{1}{2}(3\mathbf{I} - \mathbf{A}) = \begin{pmatrix} 0 & 1 \\ -\frac{1}{2} & \frac{3}{2} \end{pmatrix}.$$

Check: $\mathbf{A}\mathbf{A}^{-1} = \begin{pmatrix} 0+1 & 3-3 \\ 0 & 1 \end{pmatrix} = \mathbf{I}$.

- (b) By Theorem 6.1.1, $x^{100} = qp_{\mathbf{A}} + (ax + b)$ for some q and scalars a, b . Substituting the roots $x = 1$ and $x = 2$ of $p_{\mathbf{A}}$ gives $1 = a + b$ and $2^{100} = 2a + b$. So $a = 2^{100} - 1$ and $b = 2 - 2^{100}$. By [Theorem 9.6.1](#), $\mathbf{A}^{100} = q(\mathbf{A})p_{\mathbf{A}}(\mathbf{A}) + a\mathbf{A} + b\mathbf{I} = a\mathbf{A} + b\mathbf{I}$, that is,

$$\mathbf{A}^{100} = (2^{100} - 1)\mathbf{A} + (2 - 2^{100})\mathbf{I} = \begin{pmatrix} 2^{101} - 1 & 2 - 2^{101} \\ 2^{100} - 1 & 2 - 2^{100} \end{pmatrix}.$$

For $k = 1$ in place of 100, the same recipe gives $a = 1$, $b = 0$, and returns $\mathbf{A}$, as it should.

Exercise. B2: Trace zero in size two

- (a) Let $\mathbf{A} \in M_2(F)$ with $\text{tr } \mathbf{A} = 0$. Prove that $\mathbf{A}^2 = -(\det \mathbf{A})\mathbf{I}$.
 (b) Deduce that for all $\mathbf{X}, \mathbf{Y} \in M_2(F)$, the matrix $(\mathbf{XY} - \mathbf{YX})^2$ is a scalar multiple of $\mathbf{I}$.
 (c) Show that the converse of (a) fails: give $\mathbf{A} \in M_2(\mathbb{R})$ with $\mathbf{A}^2$ scalar and $\text{tr } \mathbf{A} \neq 0$.

Solution.

- (a) By [Theorem 9.6.1](#), $\mathbf{A}^2 - (\text{tr } \mathbf{A})\mathbf{A} + (\det \mathbf{A})\mathbf{I} = \mathbf{O}$. With $\text{tr } \mathbf{A} = 0$, this is $\mathbf{A}^2 = -(\det \mathbf{A})\mathbf{I}$.
 (b) By Theorem 4.5.3 applied to $T_{\mathbf{X}}, T_{\mathbf{Y}}$, or directly by $\text{tr}(\mathbf{XY}) = \text{tr}(\mathbf{YX})$, we have $\text{tr}(\mathbf{XY} - \mathbf{YX}) = 0$. By (a), $(\mathbf{XY} - \mathbf{YX})^2 = -\det(\mathbf{XY} - \mathbf{YX})\mathbf{I}$.
 (c) $\mathbf{A} = \mathbf{I}_2$ has $\mathbf{A}^2 = \mathbf{I}_2$ and $\text{tr } \mathbf{A} = 2 \neq 0$.

Exercise. B3: Minimal polynomial from one product

Each of the real matrices

$$\mathbf{M}_1 = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ -1 & 1 & 3 \end{pmatrix}, \quad \mathbf{M}_2 = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ -1 & 0 & 3 \end{pmatrix}$$

has characteristic polynomial $(x-2)^2(x-3)$. Find $m_{\mathbf{M}_1}$ and $m_{\mathbf{M}_2}$. Are $\mathbf{M}_1$ and $\mathbf{M}_2$ similar?

Solution. By [Corollary 9.6.2](#) (b), each minimal polynomial is $(x-2)(x-3)$ or $(x-2)^2(x-3)$, and only the first needs testing.

For $\mathbf{M}_1$: $\mathbf{M}_1 - 2\mathbf{I} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ -1 & 1 & 1 \end{pmatrix}$ and $\mathbf{M}_1 - 3\mathbf{I} = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 0 \\ -1 & 1 & 0 \end{pmatrix}$. The first row of the product $(\mathbf{M}_1 - 2\mathbf{I})(\mathbf{M}_1 - 3\mathbf{I})$ is row 2 of $\mathbf{M}_1 - 3\mathbf{I}$, namely $(0, -1, 0) \neq \mathbf{0}$. So $m_{\mathbf{M}_1} = (x - 2)^2(x - 3)$.

For $\mathbf{M}_2$: $\mathbf{M}_2 - 2\mathbf{I} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix}$ and $\mathbf{M}_2 - 3\mathbf{I} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ -1 & 0 & 0 \end{pmatrix}$. The only nonzero row of the product is $-(\text{row } 1) + (\text{row } 3)$ of $\mathbf{M}_2 - 3\mathbf{I}$, which is $(1, 0, 0) + (-1, 0, 0) = \mathbf{0}$. So the product is $\mathbf{0}$, and $m_{\mathbf{M}_2} = (x - 2)(x - 3)$.

The minimal polynomials differ, so $\mathbf{M}_1 \not\sim \mathbf{M}_2$ by Theorem 9.5.5.

9.6.4.3 C. Going deeper

Exercise. C1: Cayley–Hamilton over a commutative ring

Let R be a commutative ring (the remark in Chapter 7, §4), and suppose, as for $F[x]$, that the polynomials in x with coefficients in R form a commutative ring $R[x]$ in which two polynomials are equal exactly when all their coefficients are equal. For $\mathbf{A} \in M_n(R)$, define $p_{\mathbf{A}} = \det(x\mathbf{I} - \mathbf{A}) \in R[x]$.

- Go through the proof of Theorem 9.6.1 and name, for each step, the fact it uses. Explain why each fact holds over R and $R[x]$, and conclude that $p_{\mathbf{A}}(\mathbf{A}) = \mathbf{0}$ for every $\mathbf{A} \in M_n(R)$.
- Let $R = \mathbb{R}[t]$ and $\mathbf{A} = \begin{pmatrix} t & 1 \\ t^2 & 0 \end{pmatrix}$. Compute $p_{\mathbf{A}}$ and verify $p_{\mathbf{A}}(\mathbf{A}) = \mathbf{0}$ directly.
- Let $\mathbf{A} \in M_n(\mathbb{Z})$ with $\det \mathbf{A} = \pm 1$. Using (a) with $R = \mathbb{Z}$, show that $\mathbf{A}^{-1}$ has integer entries.

Solution.

- The shape of $p_{\mathbf{A}}$.* In the Leibniz sum for $\det(x\mathbf{I} - \mathbf{A})$, the identity permutation contributes $(x - a_{11}) \cdots (x - a_{nn})$, of degree n with leading coefficient 1, and every other term omits at least two diagonal entries, so it has degree at most $n - 2$. Hence $p_{\mathbf{A}} = x^n + c_{n-1}x^{n-1} + \cdots + c_0$ with $c_k \in R$, which is what the proof uses in place of Theorem 7.10.2; only sums and products of entries appear. *Step 1* uses the Leibniz formula over $R[x]$ (which only multiplies and adds entries) and the degree bounds for sums and products, which hold because the degree of a product of polynomials is at most the sum of the degrees; no inverse is needed for the upper bound. *Step 2* uses the adjugate identity over the commutative ring $R[x]$, which holds by the remark after Theorem 7.6.2, and coefficient comparison in $R[x]$, which is the assumption on $R[x]$. *Step 3* uses only associativity and distributivity of matrix multiplication over R and a telescoping sum. No step divides, so each works verbatim over R , and $p_{\mathbf{A}}(\mathbf{A}) = \mathbf{0}$.

- $x\mathbf{I} - \mathbf{A} = \begin{pmatrix} x - t & -1 \\ -t^2 & x \end{pmatrix}$, so $p_{\mathbf{A}} = x(x - t) - t^2 = x^2 - tx - t^2$. Then

$$\mathbf{A}^2 = \begin{pmatrix} t^2 + t^2 & t \\ t^3 & t^2 \end{pmatrix}, \quad \mathbf{A}^2 - t\mathbf{A} - t^2\mathbf{I} = \begin{pmatrix} 2t^2 - t^2 - t^2 & t - t \\ t^3 - t^3 & t^2 - 0 - t^2 \end{pmatrix} = \mathbf{0}.$$

- (c) The Leibniz formula for $\det(x\mathbf{I} - \mathbf{A})$ only adds and multiplies integers and x , so the coefficients c_k of $p_{\mathbf{A}}$ are integers. Regarding $\mathbf{A}$ as a rational matrix, Theorem 7.10.2 gives $c_0 = (-1)^n \det \mathbf{A} = \pm 1$. By (a), $\mathbf{A}(\mathbf{A}^{n-1} + c_{n-1}\mathbf{A}^{n-2} + \cdots + c_1\mathbf{I}) = -c_0\mathbf{I}$. Since $c_0^2 = 1$, multiplying by $-c_0$ gives $\mathbf{A}\mathbf{B} = \mathbf{I}$ with $\mathbf{B} = -c_0(\mathbf{A}^{n-1} + \cdots + c_1\mathbf{I}) \in M_n(\mathbb{Z})$. Regarding $\mathbf{A}$ and $\mathbf{B}$ as rational matrices, Theorem 3.4.5 gives $\mathbf{B} = \mathbf{A}^{-1}$, which therefore has integer entries. (This reproves part of Corollary 7.6.5.)

Exercise. C2: The adjugate is a polynomial in the matrix

Let $n \geq 2$, $\mathbf{A} \in M_n(F)$, $p_{\mathbf{A}} = x^n + c_{n-1}x^{n-1} + \cdots + c_0$, and let $\mathbf{B}_0, \dots, \mathbf{B}_{n-1}$ be as in the proof of Theorem 9.6.1.

- (a) Using equation (2) of the proof, show that $\mathbf{B}_0 = \mathbf{A}^{n-1} + c_{n-1}\mathbf{A}^{n-2} + \cdots + c_2\mathbf{A} + c_1\mathbf{I}$.
 (b) Show that the constant matrix $\mathbf{B}_0$ equals $\text{adj}(-\mathbf{A})$, and that $\text{adj}(-\mathbf{A}) = (-1)^{n-1} \text{adj} \mathbf{A}$.
 (c) Deduce that for **every** $\mathbf{A} \in M_n(F)$, invertible or not,

$$\text{adj} \mathbf{A} = (-1)^{n-1}(\mathbf{A}^{n-1} + c_{n-1}\mathbf{A}^{n-2} + \cdots + c_2\mathbf{A} + c_1\mathbf{I}).$$

- (d) Check (c) for $\mathbf{S} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 0 & 1 \end{pmatrix}$, whose adjugate Chapter 7 computed (the warning after Corollary 7.6.3), given that $p_{\mathbf{S}} = x^3 - 6x^2 + 2x$.

Solution.

- (a) Equation (2) with $k = n$ gives $\mathbf{B}_{n-1} = c_n\mathbf{I} = \mathbf{I}$, and for $1 \leq k \leq n-1$ it gives $\mathbf{B}_{k-1} = \mathbf{A}\mathbf{B}_k + c_k\mathbf{I}$. By downward induction on k , $\mathbf{B}_{k-1} = \mathbf{A}^{n-k} + c_{n-1}\mathbf{A}^{n-k-1} + \cdots + c_k\mathbf{I}$: this holds for $k = n$, and if it holds for k , then $\mathbf{B}_{k-2} = \mathbf{A}\mathbf{B}_{k-1} + c_{k-1}\mathbf{I} = \mathbf{A}^{n-k+1} + c_{n-1}\mathbf{A}^{n-k} + \cdots + c_k\mathbf{A} + c_{k-1}\mathbf{I}$. The case $k = 1$ is the claim.
 (b) Each entry of $\text{adj}(x\mathbf{I} - \mathbf{A})$ is a cofactor $\pm \det \mathbf{N}$ of $x\mathbf{I} - \mathbf{A}$, and its constant coefficient is its value at 0. By Lemma 8.3.7, $(\det \mathbf{N})(0) = \det \mathbf{N}(0)$, and $\mathbf{N}(0)$ is the corresponding submatrix of $-\mathbf{A}$. So the matrix of constant coefficients, $\mathbf{B}_0$, is $\text{adj}(-\mathbf{A})$. Each cofactor of $-\mathbf{A}$ is $\pm$ the determinant of an $(n-1) \times (n-1)$ submatrix of $-\mathbf{A}$, which is -1 times the corresponding submatrix of $\mathbf{A}$. Multiplying each of the $n-1$ rows by -1 multiplies the determinant by $(-1)^{n-1}$ (Theorem 7.5.1). Hence $\text{adj}(-\mathbf{A}) = (-1)^{n-1} \text{adj} \mathbf{A}$.
 (c) By (a) and (b), $(-1)^{n-1} \text{adj} \mathbf{A} = \mathbf{B}_0 = \mathbf{A}^{n-1} + c_{n-1}\mathbf{A}^{n-2} + \cdots + c_1\mathbf{I}$; multiply by $(-1)^{n-1}$. Nothing assumed $\det \mathbf{A} \neq 0$.

(d) Here $n = 3$, $c_2 = -6$, $c_1 = 2$, so the formula predicts $\text{adj } \mathbf{S} = \mathbf{S}^2 - 6\mathbf{S} + 2\mathbf{I}$. Now

$$\begin{aligned}\mathbf{S}^2 &= \begin{pmatrix} 8 & 10 & 18 \\ 16 & 20 & 36 \\ 2 & 2 & 4 \end{pmatrix}, \\ \mathbf{S}^2 - 6\mathbf{S} + 2\mathbf{I} &= \begin{pmatrix} 8 - 6 + 2 & 10 - 12 & 18 - 18 \\ 16 - 12 & 20 - 24 + 2 & 36 - 36 \\ 2 - 6 & 2 & 4 - 6 + 2 \end{pmatrix} \\ &= \begin{pmatrix} 4 & -2 & 0 \\ 4 & -2 & 0 \\ -4 & 2 & 0 \end{pmatrix},\end{aligned}$$

which is the adjugate found in Chapter 7.

Exercise. C3: Nilpotent matrices of size n

Let F be any field, $n \geq 1$, and $\mathbf{A} \in M_n(F)$ with $\mathbf{A}^k = \mathbf{O}$ for some $k \geq 1$.

- (a) Prove that $m_{\mathbf{A}} = x^s$ for some s with $1 \leq s \leq n$.
- (b) Deduce that $\mathbf{A}^n = \mathbf{O}$.
- (c) Give a 3×3 example with $\mathbf{A}^2 \neq \mathbf{O}$, and explain why no 3×3 matrix has $\mathbf{A}^3 \neq \mathbf{O} = \mathbf{A}^4$.

Solution.

- (a) x^k annihilates $\mathbf{A}$, so $m_{\mathbf{A}} \mid x^k$ (Theorem 9.5.1). The polynomial x is monic and irreducible (degree 1), so by Lemma 9.5.2, $m_{\mathbf{A}} = x^s$ with $0 \leq s \leq k$. By Corollary 9.6.2 (a), $1 \leq s = \deg m_{\mathbf{A}} \leq n$.
- (b) $x^n = x^{n-s} x^s$, so $m_{\mathbf{A}} \mid x^n$ and $\mathbf{A}^n = \mathbf{O}$ by Theorem 9.5.1.
- (c) The matrix $\mathbf{N}$ of the Quick check in Section 5 has $\mathbf{N}^2 \neq \mathbf{O} = \mathbf{N}^3$. If $\mathbf{A} \in M_3(F)$ satisfies $\mathbf{A}^4 = \mathbf{O}$, then $\mathbf{A}$ is nilpotent and $\mathbf{A}^3 = \mathbf{O}$ by (b), so $\mathbf{A}^3 \neq \mathbf{O} = \mathbf{A}^4$ is impossible. Note that no splitting or complex numbers were needed; compare Exercise (C2: Nilpotent operators), which worked over $\mathbb{C}$.