

9.10 Spectral Mapping, Products in Both Orders, Left Eigenvectors

Once the eigenvalues of T are known, what can we say about the eigenvalues of T^2 , of $T^2 - 3T + I$, or of any polynomial in T ? And when two matrices are multiplied in the two possible orders, how are the eigenvalues of $\mathbf{AB}$ and $\mathbf{BA}$ related? This section answers both questions using the triangular forms of Section 7 and the block matrices of Chapter 8. Along the way it settles two debts: Chapter 7 showed $p_{\mathbf{AB}} = p_{\mathbf{BA}}$ only when $\mathbf{A}$ is invertible, and Chapter 8 proved that vanishing traces of all powers force a matrix to be nilpotent by a long computation, with no explanation of why. We end with left eigenvectors, the eigenvectors of the transpose, which pair with ordinary eigenvectors in a useful way and prepare the Markov chains of Section 11.

9.10.1 Eigenvalues of a polynomial in T

One half of the answer needs no theory. If $Tv = \lambda v$ with $v \neq 0$, then $T^k v = \lambda^k v$ for every k , and so $q(T)v = q(\lambda)v$ for every $q \in F[x]$ (Exercise (B3: Eigenvalues of powers and inverses)). So q **maps eigenvalues of T to eigenvalues of $q(T)$** , over every field and in every dimension.

The other half, that **every** eigenvalue of $q(T)$ arises this way, can fail. Part (c) of Exercise (B3: Eigenvalues of powers and inverses) already met the rotation $\mathbf{R} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ on $\mathbb{R}^2$: it has no real eigenvalue, while $\mathbf{R}^2 = -\mathbf{I}$ has the eigenvalue -1 . The eigenvalues of $\mathbf{R}$ that would map to -1 , namely $\pm i$, are missing from $\mathbb{R}$. So we expect the full statement to need a characteristic polynomial that splits. With that hypothesis we even get the multiplicities, because a triangular form shows all the eigenvalues at once on its diagonal.

The eigenvalues of $q(T)$ are the values of q at the eigenvalues of T , counted with multiplicity.

Theorem 9.10.1. Spectral Mapping Theorem

Let V be a finite-dimensional vector space over F with $\dim V = n \geq 1$, and let $T \in \mathcal{L}(V)$. Suppose p_T splits over F , say

$$p_T(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n),$$

where each eigenvalue appears as often as its algebraic multiplicity. Then for every $q \in F[x]$,

$$p_{q(T)}(x) = (x - q(\lambda_1))(x - q(\lambda_2)) \cdots (x - q(\lambda_n)),$$

and consequently

$$\text{spec}(q(T)) = q(\text{spec}(T)) = \{q(\lambda) : \lambda \in \text{spec}(T)\}.$$

In particular this holds for every operator on a nonzero finite-dimensional complex space, and for every $\mathbf{A} \in M_n(\mathbb{C})$.

Idea. Choose a basis $\mathcal{B}$ in which T is upper triangular (Section 7). Its diagonal is $\lambda_1, \dots, \lambda_n$. Products and sums of upper triangular matrices are upper triangular, and their diagonals are

computed position by position, so $q([T]_{\mathcal{B}}^{\mathcal{B}})$ is upper triangular with diagonal $q(\lambda_1), \dots, q(\lambda_n)$. The characteristic polynomial of a triangular matrix is read off its diagonal.

Proof. By Theorem 9.7.2 there is a basis $\mathcal{B}$ of V such that $\mathbf{U} = [T]_{\mathcal{B}}^{\mathcal{B}}$ is upper triangular. By Theorem 9.7.4, its diagonal entries are $\lambda_1, \dots, \lambda_n$ in some order; renumbering the λ_i , we may assume $u_{ii} = \lambda_i$.

Claim. If $\mathbf{X}, \mathbf{Y} \in M_n(F)$ are upper triangular, then so are $\mathbf{X} + \mathbf{Y}$, $c\mathbf{X}$ for $c \in F$, and $\mathbf{XY}$, with diagonal entries $x_{ii} + y_{ii}$, cx_{ii} , and $x_{ii}y_{ii}$.

Proof. The statements for $\mathbf{X} + \mathbf{Y}$ and $c\mathbf{X}$ are entrywise. For the product, $(\mathbf{XY})_{ij} = \sum_{k=1}^n x_{ik}y_{kj}$ (Definition 0.9.9). If $i > j$, every k has $k < i$ or $k \geq i > j$, so $x_{ik} = 0$ or $y_{kj} = 0$, and the sum is 0. If $i = j$, the only term that can be nonzero has $k \geq i$ and $k \leq i$, so $(\mathbf{XY})_{ii} = x_{ii}y_{ii}$. ■

By the claim and induction on k , $\mathbf{U}^k$ is upper triangular with diagonal entries $\lambda_1^k, \dots, \lambda_n^k$ (for $k = 0$, $\mathbf{U}^0 = \mathbf{I}$). Writing $q = \sum_k a_k x^k$, the claim again shows that $q(\mathbf{U}) = \sum_k a_k \mathbf{U}^k$ is upper triangular with diagonal entries $\sum_k a_k \lambda_i^k = q(\lambda_i)$. By Corollary 4.1.3, $[q(T)]_{\mathcal{B}}^{\mathcal{B}} = q(\mathbf{U})$, so by Definition 7.10.2 and the examples after Definition 7.10.1,

$$p_{q(T)} = p_{q(\mathbf{U})} = \prod_{i=1}^n (x - q(\lambda_i)).$$

For the spectra: by Theorem 9.2.1, $\mu \in F$ is an eigenvalue of $q(T)$ if and only if $p_{q(T)}(\mu) = \prod_i (\mu - q(\lambda_i)) = 0$, that is, if and only if $\mu = q(\lambda_i)$ for some i , since a product of nonzero field elements is nonzero. The λ_i are exactly the eigenvalues of T , each listed at least once (Theorem 9.2.1). Hence $\text{spec}(q(T)) = \{q(\lambda) : \lambda \in \text{spec}(T)\}$. Over $\mathbb{C}$, p_T always splits by Corollary 6.5.5. ■

The formula for $p_{q(T)}$ carries more than the spectrum. The algebraic multiplicity of μ as an eigenvalue of $q(T)$ is the number of indices i with $q(\lambda_i) = \mu$ (Lemma 6.4.1), so several eigenvalues of T can merge into one eigenvalue of $q(T)$, and their multiplicities add. Combined with Theorem 9.3.3, applied to $q(T)$, whose characteristic polynomial splits by the theorem, we get the most used consequence.

Corollary 9.10.2. Trace and Determinant of a Polynomial in T

Under the hypotheses of Theorem 9.10.1, for every $q \in F[x]$,

$$\text{tr } q(T) = q(\lambda_1) + \dots + q(\lambda_n), \quad \det q(T) = q(\lambda_1) \cdots q(\lambda_n).$$

In particular $\text{tr } T^k = \lambda_1^k + \dots + \lambda_n^k$ for every $k \in \mathbb{N}$.

Proof. By Theorem 9.10.1, $p_{q(T)} = \prod_i (x - q(\lambda_i))$ splits. Group equal values: if $\mu_1, \dots, \mu_s$ are the distinct values among the $q(\lambda_i)$, then μ_j occurs exactly $a_{q(T)}(\mu_j)$ times in the list (Lemma 6.4.1). By Theorem 9.3.3 applied to $q(T)$, $\text{tr } q(T) = \sum_j a_{q(T)}(\mu_j) \mu_j = \sum_i q(\lambda_i)$ and

$\det q(T) = \prod_j \mu_j^{a_{q(T)}(\mu_j)} = \prod_i q(\lambda_i)$. Take $q = x^k$ for the last statement. ■

Example. A Polynomial in a Non-Diagonalizable Matrix

Let $\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ -3 & 3 & 2 \end{pmatrix} \in M_3(\mathbb{R})$, for which Exercise (B1: Diagonalize or show impossible) found $p_{\mathbf{A}} = (x+1)(x-2)^2$ and showed that $\mathbf{A}$ is not diagonalizable. Without computing any eigenvectors, find $p_{\mathbf{A}^2}$, $\text{tr } \mathbf{A}^2$, and $\det \mathbf{A}^2$. Then find $p_{q(\mathbf{A})}$ for $q = x^2 - x - 2$, and decide whether $q(\mathbf{A}) = \mathbf{O}$.

Solution. The square. $p_{\mathbf{A}}$ splits over $\mathbb{R}$ with the list $\lambda_1 = -1$, $\lambda_2 = \lambda_3 = 2$. By Theorem 9.10.1 with $q = x^2$, the list for $\mathbf{A}^2$ is 1, 4, 4, so

$$p_{\mathbf{A}^2} = (x-1)(x-4)^2, \quad \text{tr } \mathbf{A}^2 = 1 + 4 + 4 = 9, \quad \det \mathbf{A}^2 = 1 \cdot 4 \cdot 4 = 16$$

by Corollary 9.10.2. Check. Multiplying out,

$$\mathbf{A}^2 = \begin{pmatrix} 0 & 4 & 4 \\ -1 & 5 & 4 \\ -3 & 3 & 4 \end{pmatrix},$$

whose trace is $0 + 5 + 4 = 9$; and $\det \mathbf{A} = (-1) \cdot 2 \cdot 2 = -4$ by Theorem 9.3.3, so $\det \mathbf{A}^2 = (\det \mathbf{A})^2 = 16$ by Theorem 7.5.4.

The polynomial q. Here $q = (x-2)(x+1)$, so $q(-1) = 0$ and $q(2) = 0$. The list for $q(\mathbf{A})$ is 0, 0, 0, and $p_{q(\mathbf{A})} = x^3$: the only eigenvalue of $q(\mathbf{A})$ is 0. Yet $q(\mathbf{A}) \neq \mathbf{O}$. Indeed

$$q(\mathbf{A}) = \mathbf{A}^2 - \mathbf{A} - 2\mathbf{I} = \begin{pmatrix} -3 & 3 & 3 \\ -3 & 3 & 3 \\ 0 & 0 & 0 \end{pmatrix},$$

and one checks $q(\mathbf{A})^2 = \mathbf{O}$. The spectral mapping theorem sees only eigenvalues, and a nonzero matrix can have 0 as its only eigenvalue. That $q(\mathbf{A}) \neq \mathbf{O}$ here is no accident: if $q(\mathbf{A}) = \mathbf{O}$, the minimal polynomial of $\mathbf{A}$ would divide (Theorem 9.5.1) $(x-2)(x+1)$, which has distinct roots, and $\mathbf{A}$ would be diagonalizable by Theorem 9.8.2.

Warning. Without splitting, spectral mapping fails. For the real rotation $\mathbf{R}$ above, $\text{spec}(\mathbf{R}) = \emptyset$ in $\mathbb{R}$, so $q(\text{spec}(\mathbf{R})) = \emptyset$ for $q = x^2$, while $\text{spec}(\mathbf{R}^2) = \text{spec}(-\mathbf{I}) = \{-1\}$. Over $\mathbb{C}$ the theorem is restored: $\text{spec}(\mathbf{R}) = \{i, -i\}$, both square to -1 , and $p_{\mathbf{R}^2} = (x+1)^2$. Also, the theorem is about eigenvalues and algebraic multiplicities only: $\mathbf{A} = \text{diag}(1, -1)$ has two one-dimensional eigenspaces, while $\mathbf{A}^2 = \mathbf{I}$ has the single eigenspace $\mathbb{R}^2$.

Quick check. Let $\mathbf{A} \in M_2(\mathbb{C})$ have eigenvalues i and $-i$. Find $p_{\mathbf{A}^2}$ and $\det(\mathbf{A}^2 + \mathbf{I})$.

Solution. The list for $\mathbf{A}$ is $i, -i$ (two distinct roots of the degree-2 polynomial $p_{\mathbf{A}}$). By Theorem 9.10.1 with $q = x^2$, the list for $\mathbf{A}^2$ is $-1, -1$, so $p_{\mathbf{A}^2} = (x + 1)^2$. With $q = x^2 + 1$, Corollary 9.10.2 gives $\det(\mathbf{A}^2 + \mathbf{I}) = q(i)q(-i) = 0 \cdot 0 = 0$.

9.10.2 Traces of powers detect nilpotency

Chapter 8 proved, in Exercise (C2: When $\mathbf{A}$ commutes with $[\mathbf{A}, \mathbf{B}]$), that a matrix $\mathbf{C}$ over a field of characteristic 0 with $\text{tr } \mathbf{C}^k = 0$ for every $k \geq 1$ is nilpotent, by a careful argument with annihilating polynomials that gave no reason for the result. Here is the view through eigenvalues, for complex matrices. The traces of the powers are the **power sums** $\lambda_1^k + \cdots + \lambda_n^k$ of the eigenvalues (Corollary 9.10.2), and a nilpotent matrix is one whose eigenvalues are all 0. So the question becomes: if all power sums of n complex numbers vanish, must the numbers vanish? A Vandermonde determinant says yes.

Theorem 9.10.3. Nilpotent Matrices via Eigenvalues and Traces

Let $n \geq 1$ and $\mathbf{A} \in M_n(\mathbb{C})$. The following are equivalent:

- (a) $\mathbf{A}$ is nilpotent, that is, $\mathbf{A}^s = \mathbf{O}$ for some $s \geq 1$;
- (b) $\text{spec}(\mathbf{A}) = \{0\}$;
- (c) $p_{\mathbf{A}} = x^n$;
- (d) $\mathbf{A}^n = \mathbf{O}$;
- (e) $\text{tr } \mathbf{A}^k = 0$ for $k = 1, 2, \dots, n$.

The same holds for $\mathbf{A} \in M_n(F)$ when F is a subfield of $\mathbb{C}$, such as $\mathbb{Q}$ or $\mathbb{R}$, with $\text{spec}(\mathbf{A})$ in (b) taken over $\mathbb{C}$.

Idea. Plan: (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (a), and (c) $\Rightarrow$ (e) $\Rightarrow$ (b). All but one arrow are short. For (e) $\Rightarrow$ (b), suppose some eigenvalue is nonzero. Group the nonzero eigenvalues into distinct values $\mu_1, \dots, \mu_r$ with multiplicities $m_1, \dots, m_r \geq 1$. The vanishing traces say $\sum_j m_j \mu_j^k = 0$ for $k = 1, \dots, r$: a square linear system for the vector $(m_1, \dots, m_r)$, whose matrix is a Vandermonde matrix times the invertible $\text{diag}(\mu_j)$. So the m_j would be 0, but they are positive integers, and positive integers are nonzero in $\mathbb{C}$.

Proof.

- (a) $\Rightarrow$ (b). Let $\mathbf{A}^s = \mathbf{O}$ with $s \geq 1$. If $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ with $\mathbf{v} \neq \mathbf{0}$, then $\mathbf{A}^s\mathbf{v} = \lambda^s\mathbf{v}$ by induction on s , so $\lambda^s\mathbf{v} = \mathbf{0}$ and $\lambda^s = 0$ by Theorem 1.1.9; hence $\lambda = 0$, since a product of nonzero scalars is nonzero (Theorem 0.7.1 (f)). So $\text{spec}(\mathbf{A}) \subseteq \{0\}$. Conversely $\mathbf{A}$ is not injective: if it were, its s -fold composite $\mathbf{A}^s$ would be injective too (Theorem 0.4.3 (a)), whereas $\mathbf{A}^s = \mathbf{O}$ is not injective on $\mathbb{C}^n \neq \{\mathbf{0}\}$. Hence $0 \in \text{spec}(\mathbf{A})$ by Theorem 9.2.1 ((b) $\Rightarrow$ (a)), and $\text{spec}(\mathbf{A}) = \{0\}$.
- (b) $\Rightarrow$ (c). By Corollary 6.5.5, $p_{\mathbf{A}} = (x - z_1) \cdots (x - z_n)$ with $z_i \in \mathbb{C}$. Each z_i is an eigenvalue (Theorem 9.2.1), hence 0, so $p_{\mathbf{A}} = x^n$.
- (c) $\Rightarrow$ (d). By Corollary 9.7.3, $\mathbf{A} = \mathbf{P}\mathbf{U}\mathbf{P}^{-1}$ with $\mathbf{P}$ invertible and $\mathbf{U}$ upper triangular, and by Theorem 9.7.4 the diagonal entries of $\mathbf{U}$ are the roots of $p_{\mathbf{A}} = x^n$, all 0. Then

$\mathbf{U}\mathbf{e}_1 = \mathbf{0}$ and $\mathbf{U}\mathbf{e}_j \in \text{span}(\mathbf{e}_1, \dots, \mathbf{e}_{j-1})$ for $j \geq 2$, since column j of $\mathbf{U}$ has zeros from row j down. By induction on j , $\mathbf{U}^j \mathbf{e}_j = \mathbf{0}$, hence $\mathbf{U}^n \mathbf{e}_j = \mathbf{U}^{n-j} \mathbf{U}^j \mathbf{e}_j = \mathbf{0}$ for every j , and $\mathbf{U}^n = \mathbf{O}$. By Proposition 4.2.6 (c), $\mathbf{A}^n = \mathbf{P}\mathbf{U}^n\mathbf{P}^{-1} = \mathbf{O}$.

(d) $\Rightarrow$ (a). Take $s = n$.

(e) $\Rightarrow$ (e). The list of eigenvalues of $\mathbf{A}$ with multiplicity is $0, \dots, 0$, so by Corollary 9.10.2, $\text{tr } \mathbf{A}^k = 0^k + \dots + 0^k = 0$ for every $k \geq 1$.

(f) $\Rightarrow$ (b). By Corollary 6.5.5, $p_{\mathbf{A}}$ splits; let $\lambda_1, \dots, \lambda_n$ be the list of its roots with multiplicity. Since $n \geq 1$, $\text{spec}(\mathbf{A})$ is nonempty, so it suffices to show that every $\lambda_i = 0$. Suppose not, let $\mu_1, \dots, \mu_r$ ($1 \leq r \leq n$) be the distinct nonzero values among the λ_i , and let $m_j \geq 1$ be the number of indices i with $\lambda_i = \mu_j$. For $1 \leq k \leq r$, the zero eigenvalues contribute $0^k = 0$, so by Corollary 9.10.2 and (e),

$$\sum_{j=1}^r m_j \mu_j^k = \text{tr } \mathbf{A}^k = 0.$$

In matrix form, $\mathbf{W}\mathbf{m} = \mathbf{0}$, where $\mathbf{m} = (m_1, \dots, m_r) \in \mathbb{C}^r$ and $\mathbf{W} \in M_r(\mathbb{C})$ has (k, j) -entry $\mu_j^k = \mu_j^{k-1} \mu_j$. Hence $\mathbf{W} = V(\mu_1, \dots, \mu_r)^\top \text{diag}(\mu_1, \dots, \mu_r)$, where $V(\mu_1, \dots, \mu_r)$ is the Vandermonde matrix with (j, k) -entry μ_j^{k-1} (Theorem 7.8.1). The μ_j are distinct, so V is invertible (Corollary 7.8.2), and so is $V^\top$, with inverse $(V^{-1})^\top$ (Theorem 0.9.5). The μ_j are nonzero, so $\text{diag}(\mu_1, \dots, \mu_r)$ is invertible. Therefore $\mathbf{W}$ is invertible, and $\mathbf{m} = \mathbf{W}^{-1}\mathbf{0} = \mathbf{0}$. But m_1 is a positive integer, and $m_1 \neq 0$ in $\mathbb{C}$. This contradiction shows that every $\lambda_i = 0$, so $\text{spec}(\mathbf{A}) = \{0\}$.

For a subfield $F \subseteq \mathbb{C}$, regard $\mathbf{A}$ as a complex matrix. Its powers, their traces, its characteristic polynomial, and the condition $\mathbf{A}^s = \mathbf{O}$ are computed by the same arithmetic in F and in $\mathbb{C}$, so (a), (c), (d), and (e) mean the same thing over F and over $\mathbb{C}$, and the equivalences carry over. ■

In the setting of Exercise (C2: When $\mathbf{A}$ commutes with $[\mathbf{A}, \mathbf{B}]$), where $\mathbf{C} = [\mathbf{A}, \mathbf{B}]$ commutes with $\mathbf{A}$, part (a) of that exercise gives $\text{tr } \mathbf{C}^k = 0$ for all $k \geq 1$, and for complex matrices the theorem now shows at once that $\mathbf{C}^n = \mathbf{O}$. That exercise remains the route for a general field of characteristic 0.

Warning. The characteristic matters. Over $\mathbb{F}_p$, the identity $\mathbf{I}_p \in M_p(\mathbb{F}_p)$ has $\text{tr } \mathbf{I}_p^k = p \cdot 1 = 0$ for every k , yet $\mathbf{I}_p^k = \mathbf{I}_p \neq \mathbf{O}$. The proof above breaks exactly where it used that the positive integer m_1 is nonzero in the field: here the multiplicity $m_1 = p$ of the eigenvalue 1 is 0 in $\mathbb{F}_p$.

9.10.3 AB and BA

Chapter 7, in Exercise (B3: $\mathbf{AB}$ and $\mathbf{BA}$), showed that $p_{\mathbf{AB}} = p_{\mathbf{BA}}$ for square $\mathbf{A}$ and $\mathbf{B}$ when $\mathbf{A}$ is invertible, because then $\mathbf{BA} = \mathbf{A}^{-1}(\mathbf{AB})\mathbf{A}$ is similar to $\mathbf{AB}$. Without invertibility the two products need not be similar, and for rectangular $\mathbf{A}$ and $\mathbf{B}$ they do not even have the same size: if $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times m$, then $\mathbf{AB}$ is $m \times m$ and $\mathbf{BA}$ is $n \times n$. Their characteristic polynomials have degrees m and n . Multiplying by powers of x repairs the degrees, and the identity then holds in general.

Theorem 9.10.4. $\mathbf{AB}$ and $\mathbf{BA}$ Have the Same Nonzero Eigenvalues

Let $m, n \geq 1$, $\mathbf{A} \in M_{m \times n}(F)$, and $\mathbf{B} \in M_{n \times m}(F)$. Then

$$x^n p_{\mathbf{AB}}(x) = x^m p_{\mathbf{BA}}(x).$$

Consequently:

- (a) for every $\lambda \neq 0$ in F , $a_{\mathbf{AB}}(\lambda) = a_{\mathbf{BA}}(\lambda)$; in particular $\mathbf{AB}$ and $\mathbf{BA}$ have the same nonzero eigenvalues;
- (b) if $m = n$, then $p_{\mathbf{AB}} = p_{\mathbf{BA}}$; and if $m \geq n$, then $p_{\mathbf{AB}} = x^{m-n} p_{\mathbf{BA}}$.

Idea. We want one big matrix that “contains” $\mathbf{AB}$ and another that “contains” $\mathbf{BA}$, and that are similar to each other. Put them in $(m+n) \times (m+n)$ block form:

$$\mathbf{M} = \begin{pmatrix} \mathbf{AB} & \mathbf{O} \\ \mathbf{B} & \mathbf{O} \end{pmatrix}, \quad \mathbf{N} = \begin{pmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{B} & \mathbf{BA} \end{pmatrix}.$$

Both are block triangular, so their characteristic polynomials are $p_{\mathbf{AB}} \cdot x^n$ and $x^m \cdot p_{\mathbf{BA}}$. The invertible block matrix $\mathbf{S} = \begin{pmatrix} \mathbf{I}_m & \mathbf{A} \\ \mathbf{O} & \mathbf{I}_n \end{pmatrix}$ turns one into the other: $\mathbf{MS}$ and $\mathbf{SN}$ are both $\begin{pmatrix} \mathbf{AB} & \mathbf{ABA} \\ \mathbf{B} & \mathbf{BA} \end{pmatrix}$.

Proof. Let $\mathbf{M}, \mathbf{N}, \mathbf{S} \in M_{m+n}(F)$ be the block matrices displayed in the Idea, partitioned by m, n in rows and columns. By Theorem 8.1.1,

$$\begin{aligned} \mathbf{MS} &= \begin{pmatrix} \mathbf{AB} \cdot \mathbf{I}_m + \mathbf{O} & \mathbf{AB} \cdot \mathbf{A} + \mathbf{O} \\ \mathbf{B} \cdot \mathbf{I}_m + \mathbf{O} & \mathbf{B} \cdot \mathbf{A} + \mathbf{O} \end{pmatrix} = \begin{pmatrix} \mathbf{AB} & \mathbf{ABA} \\ \mathbf{B} & \mathbf{BA} \end{pmatrix}, \\ \mathbf{SN} &= \begin{pmatrix} \mathbf{O} + \mathbf{AB} & \mathbf{O} + \mathbf{A} \cdot \mathbf{BA} \\ \mathbf{O} + \mathbf{B} & \mathbf{O} + \mathbf{BA} \end{pmatrix} = \begin{pmatrix} \mathbf{AB} & \mathbf{ABA} \\ \mathbf{B} & \mathbf{BA} \end{pmatrix}, \end{aligned}$$

using associativity (Theorem 0.9.4). So $\mathbf{MS} = \mathbf{SN}$. The matrix $\mathbf{S}$ is invertible: by Theorem 8.1.1, $\begin{pmatrix} \mathbf{I}_m & \mathbf{A} \\ \mathbf{O} & \mathbf{I}_n \end{pmatrix} \begin{pmatrix} \mathbf{I}_m & -\mathbf{A} \\ \mathbf{O} & \mathbf{I}_n \end{pmatrix} = \begin{pmatrix} \mathbf{I}_m & -\mathbf{A} + \mathbf{A} \\ \mathbf{O} & \mathbf{I}_n \end{pmatrix} = \mathbf{I}_{m+n}$, and Theorem 3.4.5 applies. Hence $\mathbf{S}^{-1}\mathbf{MS} = \mathbf{N}$, and $p_{\mathbf{M}} = p_{\mathbf{N}}$ by Theorem 7.10.3.

Now compute both sides. The matrix $x\mathbf{I}_{m+n} - \mathbf{M} = \begin{pmatrix} x\mathbf{I}_m - \mathbf{AB} & \mathbf{O} \\ -\mathbf{B} & x\mathbf{I}_n \end{pmatrix}$ is lower block triangular with square diagonal blocks, so by Theorem 7.8.4, which holds over the commutative ring $F[x]$ (the remarks before Example (A 3×3 Characteristic Polynomial)), and by Theorem 7.4.6 for $x\mathbf{I}_n$,

$$p_{\mathbf{M}}(x) = \det(x\mathbf{I}_m - \mathbf{AB}) \det(x\mathbf{I}_n) = p_{\mathbf{AB}}(x) x^n.$$

Likewise $x\mathbf{I}_{m+n} - \mathbf{N} = \begin{pmatrix} x\mathbf{I}_m & \mathbf{O} \\ -\mathbf{B} & x\mathbf{I}_n - \mathbf{BA} \end{pmatrix}$ gives $p_{\mathbf{N}}(x) = x^m p_{\mathbf{BA}}(x)$. Since $p_{\mathbf{M}} = p_{\mathbf{N}}$, this proves $x^n p_{\mathbf{AB}} = x^m p_{\mathbf{BA}}$.

- (a) Let $\lambda \neq 0$. Then λ is not a root of x^n , so $\text{mult}_\lambda(x^n) = 0$, and similarly for x^m . By Theorem 6.4.2,

$$a_{\mathbf{AB}}(\lambda) = \text{mult}_\lambda(x^n p_{\mathbf{AB}}) = \text{mult}_\lambda(x^m p_{\mathbf{BA}}) = a_{\mathbf{BA}}(\lambda).$$

A scalar is an eigenvalue exactly when its algebraic multiplicity is positive (Theorem 9.2.1, Definition 6.4.1), so the nonzero eigenvalues agree.

- (b) If $m \geq n$, then $x^n p_{\mathbf{AB}} = x^n (x^{m-n} p_{\mathbf{BA}})$, and canceling the nonzero polynomial x^n (Corollary 0.8.4) gives $p_{\mathbf{AB}} = x^{m-n} p_{\mathbf{BA}}$. For $m = n$ this is $p_{\mathbf{AB}} = p_{\mathbf{BA}}$. ■

This pays off the promise of Exercise (B3: $\mathbf{AB}$ and $\mathbf{BA}$): $p_{\mathbf{AB}} = p_{\mathbf{BA}}$ for **all** square $\mathbf{A}$ and $\mathbf{B}$, invertible or not. Comparing coefficients (Theorem 7.10.2) recovers $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$ and $\det(\mathbf{AB}) = \det(\mathbf{BA})$ once more.

The eigenvectors travel between the two products as well, and by hand. If $\mathbf{AB}\mathbf{w} = \lambda\mathbf{w}$ with $\lambda \neq 0$, then

$$\mathbf{BA}(\mathbf{B}\mathbf{w}) = \mathbf{B}(\mathbf{AB}\mathbf{w}) = \lambda\mathbf{B}\mathbf{w},$$

and $\mathbf{B}\mathbf{w} \neq \mathbf{0}$, since $\mathbf{A}(\mathbf{B}\mathbf{w}) = \lambda\mathbf{w} \neq \mathbf{0}$. So $\mathbf{B}\mathbf{w}$ is an eigenvector of $\mathbf{BA}$ for the same λ : multiplying an eigenvector of $\mathbf{AB}$ by $\mathbf{B}$ produces one of $\mathbf{BA}$. For $\lambda = 0$ the recipe can fail, because $\mathbf{B}\mathbf{w}$ may be $\mathbf{0}$ — which is the warning below in another guise.

A rank-one example. Let $\mathbf{u}, \mathbf{v} \in F^n$ with $n \geq 2$. Take $\mathbf{A} = \mathbf{u} \in M_{n \times 1}(F)$ and $\mathbf{B} = \mathbf{v}^\top \in M_{1 \times n}(F)$. Then $\mathbf{AB} = \mathbf{uv}^\top$ is $n \times n$, and $\mathbf{BA} = \mathbf{v}^\top \mathbf{u}$ is the 1×1 matrix whose entry is the number $c = \mathbf{v}^\top \mathbf{u}$, so $p_{\mathbf{BA}} = x - c$. Applying the theorem with its m equal to our n , and its n equal to 1, gives $x p_{\mathbf{uv}^\top} = x^n (x - c)$, and canceling x ,

$$p_{\mathbf{uv}^\top}(x) = x^{n-1} (x - \mathbf{v}^\top \mathbf{u}).$$

For instance, the all-ones matrix $\mathbf{1}\mathbf{1}^\top \in M_n(\mathbb{R})$, with $\mathbf{1} = (1, \dots, 1)$, has $p = x^{n-1}(x - n)$: eigenvalue 0 with multiplicity $n - 1$, and eigenvalue n . Evaluating at -1 recovers Theorem 7.8.5:

$$\begin{aligned} \det(\mathbf{I} + \mathbf{uv}^\top) &= (-1)^n \det(-\mathbf{I} - \mathbf{uv}^\top) = (-1)^n p_{\mathbf{uv}^\top}(-1) \\ &= (-1)^n (-1)^{n-1} (-1 - \mathbf{v}^\top \mathbf{u}) = 1 + \mathbf{v}^\top \mathbf{u}. \end{aligned}$$

Warning. The zero eigenvalue behaves differently, and $\mathbf{AB}$ need not be similar to $\mathbf{BA}$. Take $\mathbf{A} = \mathbf{E}_{12}$ and $\mathbf{B} = \mathbf{E}_{11}$ in $M_2(F)$. Then $\mathbf{AB} = \mathbf{E}_{12}\mathbf{E}_{11} = \mathbf{O}$ and $\mathbf{BA} = \mathbf{E}_{11}\mathbf{E}_{12} = \mathbf{E}_{12}$. Both have $p = x^2$, as the theorem demands. But $g_{\mathbf{AB}}(0) = 2$ while $g_{\mathbf{BA}}(0) = 1$, and $\mathbf{AB} = \mathbf{O}$ is similar only to $\mathbf{O}$, so $\mathbf{AB} \not\sim \mathbf{BA}$. For rectangular matrices even the algebraic multiplicities of 0 differ, by $|m - n|$.

Quick check. Let $\mathbf{A} \in M_{2 \times 5}(\mathbb{R})$ and $\mathbf{B} \in M_{5 \times 2}(\mathbb{R})$ with $p_{\mathbf{AB}} = x^2 - 3x + 2$. What is $p_{\mathbf{BA}}$, and what are $\text{tr}(\mathbf{BA})$ and $\det(\mathbf{BA})$?

Solution. By Theorem 9.10.4 with $m = 2$, $n = 5$, $x^5 p_{\mathbf{AB}} = x^2 p_{\mathbf{BA}}$, so $p_{\mathbf{BA}} = x^3(x^2 - 3x + 2) =$

$x^5 - 3x^4 + 2x^3$. By Theorem 7.10.2, $\text{tr}(\mathbf{BA}) = 3$ (equal to $\text{tr}(\mathbf{AB})$), and $\det(\mathbf{BA}) = (-1)^5 \cdot 0 = 0$, as it must be, since $\mathbf{BA}$ has rank at most $2 < 5$.

9.10.4 Left eigenvectors

An eigenvector multiplies $\mathbf{A}$ from the right: $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Row vectors can multiply from the left, and in applications, especially Markov chains, the row version is often the one with a meaning.

A left eigenvector is a row that $\mathbf{A}$ only stretches when it acts from the right of the row.

Definition 9.10.1. Left Eigenvector

Let $\mathbf{A} \in M_n(F)$ and $\lambda \in F$. A **left eigenvector** of $\mathbf{A}$ for λ is a **nonzero** $\mathbf{y} \in F^n$ with

$$\mathbf{y}^\top \mathbf{A} = \lambda \mathbf{y}^\top.$$

To distinguish them, the eigenvectors of Definition 9.2.1 are also called **right eigenvectors**.

In words: $\mathbf{y}^\top$ is a $1 \times n$ row, $\mathbf{y}^\top \mathbf{A}$ is again a row, and we ask that it be a multiple of $\mathbf{y}^\top$. For example, $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}$ and $\mathbf{y} = (1, 1)$ give $\mathbf{y}^\top \mathbf{A} = \begin{pmatrix} 4 & 4 \end{pmatrix} = 4\mathbf{y}^\top$, so $(1, 1)$ is a left eigenvector for 4, whereas the right eigenvectors for 4 are the multiples of $(2, 3)$ (Example (Eigenvalues and Eigenspaces of a 2×2 Matrix)). For a diagonal matrix, $\mathbf{e}_i$ is both a left and a right eigenvector for the i -th diagonal entry. And $\mathbf{y} = \mathbf{0}$ always satisfies the equation, which is why it is excluded, exactly as for right eigenvectors.

Nonexample by minimal change. For the same $\mathbf{A}$, change $\mathbf{y} = (1, 1)$ to $\mathbf{y} = (1, -1)$. Then $\mathbf{y}^\top \mathbf{A} = \begin{pmatrix} 1-3 & 2-2 \end{pmatrix} = \begin{pmatrix} -2 & 0 \end{pmatrix}$, which is not a multiple of $\begin{pmatrix} 1 & -1 \end{pmatrix}$: the row is nonzero, but the defining equation $\mathbf{y}^\top \mathbf{A} = \lambda \mathbf{y}^\top$ has no solution λ . Note that $(1, -1)$ is a right eigenvector of $\mathbf{A}$, for -1 ; the two notions do not agree.

Transposing turns the definition into one we know.

Proposition 9.10.5. Left Eigenvectors Are Eigenvectors of the Transpose

Let $\mathbf{A} \in M_n(F)$ and $\lambda \in F$.

- (a) $\mathbf{y}$ is a left eigenvector of $\mathbf{A}$ for λ if and only if $\mathbf{y}$ is an eigenvector of $\mathbf{A}^\top$ for λ .
- (b) $\mathbf{A}$ has a left eigenvector for λ if and only if $\lambda \in \text{spec}(\mathbf{A})$. Moreover $p_{\mathbf{A}^\top} = p_{\mathbf{A}}$, and $a_{\mathbf{A}^\top}(\lambda) = a_{\mathbf{A}}(\lambda)$, $g_{\mathbf{A}^\top}(\lambda) = g_{\mathbf{A}}(\lambda)$.

Proof.

- (a) By Theorem 0.9.5, $(\mathbf{y}^\top \mathbf{A})^\top = \mathbf{A}^\top \mathbf{y}$ and $(\lambda \mathbf{y}^\top)^\top = \lambda \mathbf{y}$. Since transposing twice gives back the original, $\mathbf{y}^\top \mathbf{A} = \lambda \mathbf{y}^\top$ if and only if $\mathbf{A}^\top \mathbf{y} = \lambda \mathbf{y}$.
- (b) First, $p_{\mathbf{A}^\top} = p_{\mathbf{A}}$: by Theorem 0.9.5, $(x\mathbf{I} - \mathbf{A})^\top = x\mathbf{I} - \mathbf{A}^\top$, and transposing does not change a determinant (Theorem 7.4.5, which holds over the commutative ring $F[x]$ by the remark after it), so

$$p_{\mathbf{A}^\top} = \det(x\mathbf{I} - \mathbf{A}^\top) = \det((x\mathbf{I} - \mathbf{A})^\top) = \det(x\mathbf{I} - \mathbf{A}) = p_{\mathbf{A}}.$$

Hence $a_{\mathbf{A}^\top}(\lambda) = a_{\mathbf{A}}(\lambda)$, and $\text{spec}(\mathbf{A}^\top) = \text{spec}(\mathbf{A})$ by Theorem 9.2.1. With (a), $\mathbf{A}$ has a left eigenvector for λ if and only if $\lambda \in \text{spec}(\mathbf{A}^\top) = \text{spec}(\mathbf{A})$. Finally, $\mathbf{A}^\top - \lambda \mathbf{I} = (\mathbf{A} - \lambda \mathbf{I})^\top$, which has the same rank as $\mathbf{A} - \lambda \mathbf{I}$ (Theorem 3.5.5), so by Theorem 3.5.7, $g_{\mathbf{A}^\top}(\lambda) = n - \text{rank}(\mathbf{A} - \lambda \mathbf{I})^\top = n - \text{rank}(\mathbf{A} - \lambda \mathbf{I}) = g_{\mathbf{A}}(\lambda)$. ■

Remark. For an operator T on V , left eigenvectors correspond to eigenvectors of the dual map T' on V^* (Chapter 5), whose matrix in the dual bases is the transpose (Theorem 5.4.2): an eigenvector of T' is a functional $\varphi \neq 0$ with $\varphi \circ T = \lambda \varphi$.

Left and right eigenvectors for the **same** eigenvalue are usually unrelated, as $(1, 1)$ and $(2, 3)$ show. For **different** eigenvalues, though, they are always orthogonal in the sense of the plain product $\mathbf{y}^\top \mathbf{v}$.

Theorem 9.10.6. Left and Right Eigenvectors Are Biorthogonal

Let $\mathbf{A} \in M_n(F)$, let $\mathbf{y}$ be a left eigenvector of $\mathbf{A}$ for μ , and let $\mathbf{v}$ be a right eigenvector of $\mathbf{A}$ for λ . If $\lambda \neq \mu$, then $\mathbf{y}^\top \mathbf{v} = 0$.

Proof. Compute the 1×1 matrix $\mathbf{y}^\top \mathbf{A} \mathbf{v}$ in two ways, using associativity (Theorem 0.9.4):

$$\mu \mathbf{y}^\top \mathbf{v} = (\mathbf{y}^\top \mathbf{A}) \mathbf{v} = \mathbf{y}^\top \mathbf{A} \mathbf{v} = \mathbf{y}^\top (\mathbf{A} \mathbf{v}) = \lambda \mathbf{y}^\top \mathbf{v}.$$

Hence $(\lambda - \mu) \mathbf{y}^\top \mathbf{v} = 0$. Since $\lambda - \mu \neq 0$, this forces $\mathbf{y}^\top \mathbf{v} = 0$ (Theorem 0.7.1). ■

For the example, $(1, 1)$ is a left eigenvector for 4 and $(1, -1)$ a right eigenvector for -1 , and indeed $1 - 1 = 0$.

For a diagonalizable matrix the left eigenvectors come for free with the diagonalization, and they give a formula for all powers that we use in Section 11.

Proposition 9.10.7. Expansion in Left and Right Eigenvectors

Let $\mathbf{A} = \mathbf{P} \mathbf{D} \mathbf{P}^{-1}$ with $\mathbf{P} \in M_n(F)$ invertible and $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_n)$. Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be the columns of $\mathbf{P}$, and let $\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top$ be the rows of $\mathbf{P}^{-1}$. Then:

- (a) $\mathbf{v}_i$ is a right eigenvector and $\mathbf{y}_i$ a left eigenvector of $\mathbf{A}$ for λ_i ;
- (b) $\mathbf{y}_i^\top \mathbf{v}_j = 1$ if $i = j$ and 0 if $i \neq j$;
- (c) for every $q \in F[x]$,

$$q(\mathbf{A}) = \sum_{i=1}^n q(\lambda_i) \mathbf{v}_i \mathbf{y}_i^\top, \quad \text{in particular} \quad \mathbf{A}^k = \sum_{i=1}^n \lambda_i^k \mathbf{v}_i \mathbf{y}_i^\top \quad (k \in \mathbb{N}).$$

Proof.

- (a) From $\mathbf{AP} = \mathbf{PD}$, column i reads $\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i$ (Section 4), and $\mathbf{v}_i \neq \mathbf{0}$ because $\mathbf{P}$ is invertible. From $\mathbf{P}^{-1}\mathbf{A} = \mathbf{D}\mathbf{P}^{-1}$, row i reads $\mathbf{y}_i^\top \mathbf{A} = \lambda_i\mathbf{y}_i^\top$ (Theorem 0.9.2), and $\mathbf{y}_i \neq \mathbf{0}$ because $\mathbf{P}^{-1}$ is invertible.
- (b) The (i, j) -entry of $\mathbf{P}^{-1}\mathbf{P} = \mathbf{I}$ is row i of $\mathbf{P}^{-1}$ times column j of $\mathbf{P}$, that is, $\mathbf{y}_i^\top \mathbf{v}_j$.
- (c) By Theorem 9.4.3, $q(\mathbf{A}) = \mathbf{P}q(\mathbf{D})\mathbf{P}^{-1}$. The columns of $\mathbf{P}q(\mathbf{D})$ are $q(\lambda_i)\mathbf{v}_i$, and the rows of $\mathbf{P}^{-1}$ are $\mathbf{y}_i^\top$, so Corollary 8.1.3 gives $q(\mathbf{A}) = \sum_i q(\lambda_i)\mathbf{v}_i\mathbf{y}_i^\top$. Take $q = x^k$.

■

Part (b) is biorthogonality with a normalization: $\mathbf{y}_i^\top \mathbf{v}_i = 1$. Part (c) splits $\mathbf{A}$ into rank-one pieces $\mathbf{v}_i\mathbf{y}_i^\top$, one per eigenvalue, and each piece is simply multiplied by λ_i^k when $\mathbf{A}$ is raised to the power k . In the running example, $\mathbf{P} = \begin{pmatrix} 2 & 1 \\ 3 & -1 \end{pmatrix}$ and $\mathbf{D} = \text{diag}(4, -1)$ give $\mathbf{P}^{-1} = \frac{1}{5} \begin{pmatrix} 1 & 1 \\ 3 & -2 \end{pmatrix}$, so $\mathbf{y}_1 = \frac{1}{5}(1, 1)$ and $\mathbf{y}_2 = \frac{1}{5}(3, -2)$, and

$$\mathbf{A}^k = \frac{4^k}{5} \begin{pmatrix} 2 & 2 \\ 3 & 3 \end{pmatrix} + \frac{(-1)^k}{5} \begin{pmatrix} 3 & -2 \\ -3 & 2 \end{pmatrix}.$$

For $k = 1$ this is $\frac{1}{5} \begin{pmatrix} 8-3 & 8+2 \\ 12+3 & 12-2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}$, as it should be.

9.10.5 Exercises

9.10.5.1 A. Check your understanding

Exercise. A1

- (a) State the Spectral Mapping Theorem, including its hypothesis on p_T .
- (b) True or false: for every $\mathbf{A} \in M_n(\mathbb{R})$, every real eigenvalue of $\mathbf{A}^2$ is the square of a real eigenvalue of $\mathbf{A}$. Justify your answer.
- (c) True or false: if $\mathbf{A} \in M_{2 \times 3}(F)$ and $\mathbf{B} \in M_{3 \times 2}(F)$, then $p_{\mathbf{AB}} = p_{\mathbf{BA}}$. Justify your answer.
- (d) True or false: for all $\mathbf{A}, \mathbf{B} \in M_n(F)$, $\mathbf{AB}$ is similar to $\mathbf{BA}$. Justify your answer.
- (e) Define a left eigenvector of $\mathbf{A} \in M_n(F)$, and say how left eigenvectors of $\mathbf{A}$ are related to $\mathbf{A}^\top$.
- (f) True or false: a matrix $\mathbf{A} \in M_3(\mathbb{C})$ with $\text{tr } \mathbf{A} = \text{tr } \mathbf{A}^2 = \text{tr } \mathbf{A}^3 = 0$ satisfies $\mathbf{A}^3 = \mathbf{O}$. Justify your answer.

Solution.

- (a) If V is finite-dimensional with $\dim V = n \geq 1$, $T \in \mathcal{L}(V)$, and $p_T = \prod_{i=1}^n (x - \lambda_i)$ splits over F , then for every $q \in F[x]$, $p_{q(T)} = \prod_i (x - q(\lambda_i))$ and $\text{spec}(q(T)) = q(\text{spec}(T))$ (Theorem 9.10.1).
- (b) False. The rotation $\mathbf{R} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ has no real eigenvalue, while $\mathbf{R}^2 = -\mathbf{I}$ has the real eigenvalue -1 .

- (c) False. They have different degrees, 2 and 3. The correct identity is $x^3 p_{\mathbf{AB}} = x^2 p_{\mathbf{BA}}$, that is, $p_{\mathbf{BA}} = x p_{\mathbf{AB}}$ (Theorem 9.10.4).
- (d) False. $\mathbf{A} = \mathbf{E}_{12}$, $\mathbf{B} = \mathbf{E}_{11}$ give $\mathbf{AB} = \mathbf{O}$ and $\mathbf{BA} = \mathbf{E}_{12} \neq \mathbf{O}$, and the only matrix similar to $\mathbf{O}$ is $\mathbf{O}$.
- (e) A nonzero $\mathbf{y} \in F^n$ with $\mathbf{y}^\top \mathbf{A} = \lambda \mathbf{y}^\top$ for some $\lambda \in F$. Equivalently, $\mathbf{y}$ is an eigenvector of $\mathbf{A}^\top$ for λ (Proposition 9.10.5).
- (f) True. By Theorem 9.10.3 with $n = 3$, (e) $\Rightarrow$ (d).

9.10.5.2 B. Practice

Exercise. B1: Eigenvalues of a polynomial in $\mathbf{A}$

Let $\mathbf{A} \in M_3(\mathbb{C})$ have $p_{\mathbf{A}} = (x - 1)^2(x + 2)$.

- (a) Find $p_{\mathbf{A}^2 + \mathbf{A} + \mathbf{I}}$, $\text{tr}(\mathbf{A}^2 + \mathbf{A} + \mathbf{I})$, and $\det(\mathbf{A}^2 + \mathbf{A} + \mathbf{I})$. Is $\mathbf{A}^2 + \mathbf{A} + \mathbf{I}$ invertible?
- (b) Find $p_{\mathbf{A}^2 + \mathbf{A} - 2\mathbf{I}}$. Hence show that $\mathbf{A}^2 + \mathbf{A} - 2\mathbf{I}$ is nilpotent.
- (c) Find $\text{tr} \mathbf{A}^5$ and $\det(\mathbf{A}^{-1})$.

Solution. The list of eigenvalues of $\mathbf{A}$ with multiplicity is 1, 1, -2 , since $p_{\mathbf{A}}$ splits.

- (a) With $q = x^2 + x + 1$, $q(1) = 3$ and $q(-2) = 4 - 2 + 1 = 3$. By Theorem 9.10.1, $p_q(\mathbf{A}) = (x - 3)^3$, and by Corollary 9.10.2, $\text{tr} q(\mathbf{A}) = 9$ and $\det q(\mathbf{A}) = 27 \neq 0$. So $\mathbf{A}^2 + \mathbf{A} + \mathbf{I}$ is invertible (Theorem 7.5.3).
- (b) With $r = x^2 + x - 2 = (x - 1)(x + 2)$, $r(1) = r(-2) = 0$, so $p_r(\mathbf{A}) = x^3$ by Theorem 9.10.1. By Theorem 9.10.3 ((c) $\Rightarrow$ (d)), $r(\mathbf{A})^3 = \mathbf{O}$, so $\mathbf{A}^2 + \mathbf{A} - 2\mathbf{I}$ is nilpotent.
- (c) By Corollary 9.10.2, $\text{tr} \mathbf{A}^5 = 1 + 1 + (-2)^5 = -30$. Also $\det \mathbf{A} = 1 \cdot 1 \cdot (-2) = -2 \neq 0$ (Theorem 9.3.3), so $\mathbf{A}$ is invertible and $\det(\mathbf{A}^{-1}) = (\det \mathbf{A})^{-1} = -\frac{1}{2}$ (Corollary 7.5.6).

Exercise. B2: $\mathbf{AB}$ and $\mathbf{BA}$ for a 2×3 example

Let $\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 1 & 1 \end{pmatrix}$.

- (a) Compute $\mathbf{AB}$, $\mathbf{BA}$, $p_{\mathbf{AB}}$, and $p_{\mathbf{BA}}$, and verify that $x^3 p_{\mathbf{AB}} = x^2 p_{\mathbf{BA}}$.
- (b) Hence list the eigenvalues of $\mathbf{BA}$ without further computation, and find an eigenvector of $\mathbf{BA}$ for each nonzero eigenvalue by applying $\mathbf{B}$ to eigenvectors of $\mathbf{AB}$.

Solution.

- (a) By row-times-column,

$$\mathbf{AB} = \begin{pmatrix} 2 & 1 \\ 0 & -2 \end{pmatrix}, \quad \mathbf{BA} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & -2 & 0 \\ 1 & 2 & 1 \end{pmatrix}.$$

$\mathbf{AB}$ is upper triangular, so $p_{\mathbf{AB}} = (x-2)(x+2) = x^2 - 4$. Expanding $\det(x\mathbf{I} - \mathbf{BA})$ along the second row, whose only nonzero entry is $x+2$ in position $(2,2)$,

$$\begin{aligned} p_{\mathbf{BA}} &= (x+2) \det \begin{pmatrix} x-1 & -1 \\ -1 & x-1 \end{pmatrix} \\ &= (x+2)((x-1)^2 - 1) = (x+2)x(x-2) = x^3 - 4x. \end{aligned}$$

Then $x^3 p_{\mathbf{AB}} = x^5 - 4x^3 = x^2(x^3 - 4x) = x^2 p_{\mathbf{BA}}$.

- (b) The eigenvalues of $\mathbf{BA}$ are the roots of $x(x-2)(x+2)$: $0, 2, -2$, matching [Theorem 9.10.4](#) (the nonzero ones are those of $\mathbf{AB}$). If $\mathbf{AB}\mathbf{w} = \lambda\mathbf{w}$, then $\mathbf{BA}(\mathbf{B}\mathbf{w}) = \mathbf{B}(\mathbf{AB}\mathbf{w}) = \lambda\mathbf{B}\mathbf{w}$, so $\mathbf{B}\mathbf{w}$ is an eigenvector of $\mathbf{BA}$ as long as $\mathbf{B}\mathbf{w} \neq \mathbf{0}$. For $\lambda = 2$: $\mathbf{w} = (1, 0)$, $\mathbf{B}\mathbf{w} = (1, 0, 1)$, and $\mathbf{BA}(1, 0, 1) = (2, 0, 2)$. For $\lambda = -2$: $\mathbf{AB} + 2\mathbf{I} = \begin{pmatrix} 4 & 1 \\ 0 & 0 \end{pmatrix}$ gives $\mathbf{w} = (1, -4)$, $\mathbf{B}\mathbf{w} = (1, 4, -3)$, and $\mathbf{BA}(1, 4, -3) = (1-3, -8, 1+8-3) = (-2, -8, 6) = -2(1, 4, -3)$.

Exercise. B3: Left eigenvectors of a 2×2 matrix

Let $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -6 & 5 \end{pmatrix} \in M_2(\mathbb{R})$, whose right eigenvectors $(1, 2)$ for 2 and $(1, 3)$ for 3 were found in Exercise (B1: Diagonalize or show impossible).

- Find a left eigenvector of $\mathbf{A}$ for each eigenvalue.
- Verify [Theorem 9.10.6](#) for these vectors, and rescale the left eigenvectors so that $\mathbf{y}_i^\top \mathbf{v}_i = 1$.
- Hence write $\mathbf{A}^k$ as a combination of two rank-one matrices, and check the formula for $k = 0$ and $k = 1$.

Solution.

- By [Proposition 9.10.5](#), we need eigenvectors of $\mathbf{A}^\top = \begin{pmatrix} 0 & -6 \\ 1 & 5 \end{pmatrix}$. For 2: $\mathbf{A}^\top - 2\mathbf{I} = \begin{pmatrix} -2 & -6 \\ 1 & 3 \end{pmatrix}$ gives $y_1 = -3y_2$, so $(3, -1)$. For 3: $\mathbf{A}^\top - 3\mathbf{I} = \begin{pmatrix} -3 & -6 \\ 1 & 2 \end{pmatrix}$ gives $y_1 = -2y_2$, so $(2, -1)$. Check: $\begin{pmatrix} 3 & -1 \end{pmatrix} \mathbf{A} = \begin{pmatrix} 6 & -2 \end{pmatrix}$ and $\begin{pmatrix} 2 & -1 \end{pmatrix} \mathbf{A} = \begin{pmatrix} 6 & -3 \end{pmatrix}$.
- $(3, -1) \cdot (1, 3) = 0$ and $(2, -1) \cdot (1, 2) = 0$, as the theorem predicts for different eigenvalues. Also $(3, -1) \cdot (1, 2) = 1$ and $(2, -1) \cdot (1, 3) = -1$, so take $\mathbf{y}_1 = (3, -1)$ and $\mathbf{y}_2 = (-2, 1)$.
- With $\mathbf{P} = \begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix}$, the rows of $\mathbf{P}^{-1} = \begin{pmatrix} 3 & -1 \\ -2 & 1 \end{pmatrix}$ are exactly $\mathbf{y}_1^\top, \mathbf{y}_2^\top$, and by [Proposition 9.10.7](#),

$$\mathbf{A}^k = 2^k \begin{pmatrix} 1 \\ 2 \end{pmatrix} \begin{pmatrix} 3 & -1 \end{pmatrix} + 3^k \begin{pmatrix} 1 \\ 3 \end{pmatrix} \begin{pmatrix} -2 & 1 \end{pmatrix} = 2^k \begin{pmatrix} 3 & -1 \\ 6 & -2 \end{pmatrix} + 3^k \begin{pmatrix} -2 & 1 \\ -6 & 3 \end{pmatrix}.$$

$$\text{For } k = 0: \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \text{ For } k = 1: \begin{pmatrix} 6-6 & -2+3 \\ 12-18 & -4+9 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -6 & 5 \end{pmatrix} = \mathbf{A}.$$

9.10.5.3 C. Going deeper

Exercise. C1: Traces of powers of $\mathbf{AB}$ and $\mathbf{BA}$

Let $\mathbf{A} \in M_{m \times n}(F)$ and $\mathbf{B} \in M_{n \times m}(F)$.

- Prove that $\mathbf{B}(\mathbf{AB})^{k-1}\mathbf{A} = (\mathbf{BA})^k$ for every $k \geq 1$, and deduce that $\text{tr}((\mathbf{AB})^k) = \text{tr}((\mathbf{BA})^k)$.
- Now let $F = \mathbb{C}$. Give a second proof of the trace identity from [Theorem 9.10.4](#) and [Corollary 9.10.2](#).

Solution.

- By induction on k . For $k = 1$ both sides are $\mathbf{BA}$. If $\mathbf{B}(\mathbf{AB})^{k-1}\mathbf{A} = (\mathbf{BA})^k$, then by associativity

$$\begin{aligned} \mathbf{B}(\mathbf{AB})^k\mathbf{A} &= \mathbf{B}(\mathbf{AB})^{k-1}(\mathbf{AB})\mathbf{A} = (\mathbf{B}(\mathbf{AB})^{k-1}\mathbf{A})(\mathbf{BA}) \\ &= (\mathbf{BA})^k(\mathbf{BA}) = (\mathbf{BA})^{k+1}. \end{aligned}$$

Now $(\mathbf{AB})^k = \mathbf{A}(\mathbf{B}(\mathbf{AB})^{k-1})$, with $\mathbf{A}$ of size $m \times n$ and $\mathbf{B}(\mathbf{AB})^{k-1}$ of size $n \times m$. By Theorem 0.9.6 (3),

$$\text{tr}((\mathbf{AB})^k) = \text{tr}(\mathbf{A} \cdot \mathbf{B}(\mathbf{AB})^{k-1}) = \text{tr}(\mathbf{B}(\mathbf{AB})^{k-1} \cdot \mathbf{A}) = \text{tr}((\mathbf{BA})^k).$$

- Over $\mathbb{C}$, $p_{\mathbf{AB}}$ and $p_{\mathbf{BA}}$ split. Suppose $m \geq n$ (otherwise swap the roles of $\mathbf{A}$ and $\mathbf{B}$). By [Theorem 9.10.4](#), $p_{\mathbf{AB}} = x^{m-n}p_{\mathbf{BA}}$, so if $\lambda_1, \dots, \lambda_n$ is the list of eigenvalues of $\mathbf{BA}$ with multiplicity, then the list for $\mathbf{AB}$ is $\lambda_1, \dots, \lambda_n$ together with $m - n$ zeros. For $k \geq 1$ the zeros contribute $0^k = 0$, so by [Corollary 9.10.2](#), $\text{tr}((\mathbf{AB})^k) = \sum_i \lambda_i^k = \text{tr}((\mathbf{BA})^k)$.

Exercise. C2: Adding a commuting nilpotent matrix

Let $\mathbf{A}, \mathbf{N} \in M_n(\mathbb{C})$, $n \geq 1$, with $\mathbf{AN} = \mathbf{NA}$ and $\mathbf{N}$ nilpotent.

- Prove that $p_{\mathbf{A}+\mathbf{N}} = p_{\mathbf{A}}$; in particular $\text{spec}(\mathbf{A} + \mathbf{N}) = \text{spec}(\mathbf{A})$ and $\det(\mathbf{A} + \mathbf{N}) = \det \mathbf{A}$.
- Deduce that if $\mathbf{A}$ is invertible, so is $\mathbf{A} + \mathbf{N}$.
- Show that (a) fails without $\mathbf{AN} = \mathbf{NA}$.

Hint: triangularize $\mathbf{A}$ and $\mathbf{N}$ together.

Solution.

- By Theorem 9.9.4 there is an invertible $\mathbf{P}$ with $\mathbf{U} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ and $\mathbf{U}' = \mathbf{P}^{-1}\mathbf{N}\mathbf{P}$ upper triangular. Since $\mathbf{N}$ is nilpotent, $\text{spec}(\mathbf{N}) = \{0\}$ (Exercise (C2: Nilpotent operators)), so every diagonal entry of $\mathbf{U}'$ is 0 by Theorem 9.7.4. Then $\mathbf{P}^{-1}(\mathbf{A} + \mathbf{N})\mathbf{P} = \mathbf{U} + \mathbf{U}'$ is upper triangular with the same diagonal as $\mathbf{U}$. By Theorem 7.10.3 and the examples

after Definition 7.10.1,

$$p_{\mathbf{A}+\mathbf{N}} = p_{\mathbf{U}+\mathbf{U}'} = \prod_i (x - u_{ii}) = p_{\mathbf{U}} = p_{\mathbf{A}}.$$

Equal characteristic polynomials have the same roots (Theorem 9.2.1) and the same constant term $(-1)^n \det$ (Theorem 7.10.2).

(b) By (a), $\det(\mathbf{A} + \mathbf{N}) = \det \mathbf{A} \neq 0$, so $\mathbf{A} + \mathbf{N}$ is invertible (Theorem 7.5.3).

(c) Let $\mathbf{A} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ and $\mathbf{N} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Both are nilpotent, so $p_{\mathbf{A}} = x^2$, but $\mathbf{A} + \mathbf{N} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ has $p = x^2 - 1$ and eigenvalues ± 1 . Here $\mathbf{AN} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \neq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \mathbf{NA}$.

Exercise. C3: How many traces, and which field?

- (a) Let $\omega = e^{2\pi i/3} \in \mathbb{C}$ and $\mathbf{C} = \text{diag}(1, \omega, \omega^2)$. Show that $\text{tr } \mathbf{C} = \text{tr } \mathbf{C}^2 = 0$ but $\mathbf{C}$ is not nilpotent. Explain why this does not contradict Theorem 9.10.3.
- (b) Let $n \geq 2$. Give a matrix $\mathbf{C} \in M_n(\mathbb{C})$ with $\text{tr } \mathbf{C}^k = 0$ for $k = 1, \dots, n-1$ that is not nilpotent.
- (c) Let $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{C})$ and suppose $\mathbf{C} = \mathbf{AB} - \mathbf{BA}$ commutes with $\mathbf{A}$. Prove that $\mathbf{C}^n = \mathbf{O}$.
- (d) Over $\mathbb{F}_2$, find a nonnilpotent $\mathbf{C} \in M_2(\mathbb{F}_2)$ with $\text{tr } \mathbf{C}^k = 0$ for every $k \geq 1$.

Solution.

- (a) $\omega^3 = 1$ and $\omega \neq 1$, so $0 = \omega^3 - 1 = (\omega - 1)(\omega^2 + \omega + 1)$ gives $1 + \omega + \omega^2 = 0$. Hence $\text{tr } \mathbf{C} = 1 + \omega + \omega^2 = 0$, and $\text{tr } \mathbf{C}^2 = 1 + \omega^2 + \omega^4 = 1 + \omega^2 + \omega = 0$. But $\mathbf{C}^k = \text{diag}(1, \omega^k, \omega^{2k}) \neq \mathbf{O}$ for every k . No contradiction: the theorem, for $n = 3$, requires $\text{tr } \mathbf{C}^k = 0$ for $k = 1, 2, 3$, and $\text{tr } \mathbf{C}^3 = 1 + 1 + 1 = 3 \neq 0$.
- (b) Let $\zeta = e^{2\pi i/n}$ and $\mathbf{C} = \text{diag}(1, \zeta, \dots, \zeta^{n-1})$. For $1 \leq k \leq n-1$, $\zeta^k \neq 1$ and $(\zeta^k)^n = 1$, so $\text{tr } \mathbf{C}^k = \sum_{j=0}^{n-1} (\zeta^k)^j = \frac{(\zeta^k)^n - 1}{\zeta^k - 1} = 0$ by the geometric sum. But $\mathbf{C}^k$ has the entry 1 in position $(1, 1)$ for all k , so $\mathbf{C}$ is not nilpotent.
- (c) By Exercise (C2: When $\mathbf{A}$ commutes with $[\mathbf{A}, \mathbf{B}]$) (a), $\text{tr } \mathbf{C}^k = 0$ for every $k \geq 1$, in particular for $k = 1, \dots, n$. By Theorem 9.10.3 ((e) $\Rightarrow$ (d)), $\mathbf{C}^n = \mathbf{O}$.
- (d) $\mathbf{C} = \mathbf{I}_2$: $\text{tr } \mathbf{C}^k = 1 + 1 = 0$ in $\mathbb{F}_2$ for every k , and $\mathbf{C}^k = \mathbf{I}_2 \neq \mathbf{O}$.