

9.9 Commuting Operators

So far we have diagonalized or triangularized one operator at a time. Many problems hand us several at once: two matrices that describe the same system, or a matrix together with all its powers. Can a single basis make **all** of them diagonal, or all of them triangular? A necessary condition is easy to spot, because diagonal matrices commute. This section shows that commuting is also enough, as long as each operator can be handled on its own. The whole argument rests on one observation: an operator that commutes with T maps each eigenspace of T into itself, so we can search for common eigenvectors inside an eigenspace.

9.9.1 When can two operators share a basis?

Let V be finite-dimensional and $S, T \in \mathcal{L}(V)$. Suppose one basis $\mathcal{B}$ makes both $[S]_{\mathcal{B}}^{\mathcal{B}}$ and $[T]_{\mathcal{B}}^{\mathcal{B}}$ diagonal. Diagonal matrices multiply entry by entry on the diagonal, so $[S]_{\mathcal{B}}^{\mathcal{B}}[T]_{\mathcal{B}}^{\mathcal{B}} = [T]_{\mathcal{B}}^{\mathcal{B}}[S]_{\mathcal{B}}^{\mathcal{B}}$. By Corollary 4.1.3 this says $[ST]_{\mathcal{B}}^{\mathcal{B}} = [TS]_{\mathcal{B}}^{\mathcal{B}}$, and since $T \mapsto [T]_{\mathcal{B}}^{\mathcal{B}}$ is injective (Theorem 4.1.4), $ST = TS$. So **sharing a diagonalizing basis forces commuting**. The same is not true for triangular matrices: $\begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ are both upper triangular, and their products in the two orders are $\begin{pmatrix} 1 & 2 \\ 0 & 4 \end{pmatrix}$ and $\begin{pmatrix} 1 & 1 \\ 0 & 4 \end{pmatrix}$. We will see, though, that commuting is a convenient sufficient condition in the triangular case as well.

The question for this section is the converse: if $ST = TS$, can we find a common basis? Here are the commuting pairs to keep in mind.

- **Polynomials in one operator.** For every $p, q \in F[x]$, $p(T)q(T) = q(T)p(T)$ (Theorem 6.7.1 (c)). In particular T commutes with T^2 , with T^{-1} when it exists (both products are id_V), and with every scalar operator $c \text{id}_V$.
- **Diagonal matrices** commute with each other, and so do block diagonal matrices whose corresponding blocks commute (Theorem 8.1.4).
- **A noncommuting pair.** $\mathbf{E}_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $\mathbf{E}_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ give $\mathbf{E}_{12}\mathbf{E}_{21} = \mathbf{E}_{11} \neq \mathbf{E}_{22} = \mathbf{E}_{21}\mathbf{E}_{12}$.

The last pair already shows that something must be assumed. Over any field, the eigenvectors of $\mathbf{E}_{12}$ are the nonzero multiples of $\mathbf{e}_1$, and those of $\mathbf{E}_{21}$ are the nonzero multiples of $\mathbf{e}_2$. They share no eigenvector at all, let alone a basis of them.

9.9.2 Commuting operators preserve eigenspaces

How does the hypothesis $ST = TS$ interact with eigenvectors? Take $Tv = \lambda v$ and ask what T does to Sv . There is no S in the eigenvector equation yet, so we apply S to both sides.

Theorem 9.9.1. Commuting Operators Preserve Eigenspaces

Let V be a vector space over F , and let $S, T \in \mathcal{L}(V)$ with $ST = TS$. Then for every $\lambda \in F$, the eigenspace $E_{\lambda}(T)$ is S -invariant.

Proof. Let $v \in E_\lambda(T)$, so $Tv = \lambda v$. Since $ST = TS$ and S is linear,

$$T(Sv) = (TS)v = (ST)v = S(\lambda v) = \lambda Sv.$$

Hence $Sv \in E_\lambda(T)$ by Definition 9.2.2. This shows that S maps $E_\lambda(T)$ into itself. ■

No finite dimension is needed, and λ need not be an eigenvalue (then $E_\lambda(T) = \{0\}$ and there is nothing to prove). By symmetry, every eigenspace of S is T -invariant too.

In matrix language the theorem is a statement about block shapes. Take $T = \text{diag}(1, 1, 2)$ on F^3 , with $E_1(T) = \text{span}(\mathbf{e}_1, \mathbf{e}_2)$ and $E_2(T) = \text{span}(\mathbf{e}_3)$, and let S be any matrix with $ST = TS$. Both eigenspaces are S -invariant, and $F^3 = E_1(T) \oplus E_2(T)$, so by Theorem 9.1.5 the matrix S is block diagonal:

$$S = \begin{pmatrix} s_{11} & s_{12} & 0 \\ s_{21} & s_{22} & 0 \\ 0 & 0 & s_{33} \end{pmatrix}.$$

Conversely, every matrix of this shape commutes with T , since $T = \mathbf{I}_2 \oplus (2)$ and $\mathbf{I}_2$ commutes with every 2×2 block (Theorem 8.1.4). Inside the repeated eigenspace S can be anything; this freedom is what the rest of the section has to deal with.

Nonexample. Take $T = \text{diag}(1, 2)$ and the swap $S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Then $S\mathbf{e}_1 = \mathbf{e}_2 \notin E_1(T) = \text{span}(\mathbf{e}_1)$, so $E_1(T)$ is not S -invariant. The theorem is not contradicted: $ST = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}$ and $TS = \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}$ differ.

Quick check. Let $T = \mathbf{I}_2$ and $S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ over $\mathbb{R}$. Do S and T commute? Is every eigenvector of T an eigenvector of S ? Do S and T have a common eigenvector?

Solution. Yes, $\mathbf{I}_2$ commutes with every 2×2 matrix. No: every nonzero vector is an eigenvector of $\mathbf{I}_2$, but $S\mathbf{e}_2 = (1, 1)$ is not a multiple of $\mathbf{e}_2$. Yes: $\mathbf{e}_1$ is an eigenvector of both, for the eigenvalues 1 and 1. The theorem says S maps $E_1(T) = \mathbb{R}^2$ into itself, which is true but empty of information; the common eigenvector must be **searched for** inside that eigenspace, and here it is an eigenvector of S .

9.9.3 A common eigenvector

Here is the plan for finding a common eigenvector of commuting S and T . Pick an eigenvalue λ of T and put $U = E_\lambda(T) \neq \{0\}$. By Theorem 9.9.1, S restricts to an operator $S|_U$ on U . If $S|_U$ has an eigenvector u , then $Su = \mu u$ and, since $u \in U$, also $Tu = \lambda u$. So everything reduces to whether $S|_U$ has an eigenvalue. Over $\mathbb{C}$ it does (Theorem 9.2.4). Over a general field we need to know that the characteristic polynomial of a restriction splits when that of S

does. We record this, together with the same fact for the induced operator on the quotient, since triangularization will need both.

Lemma 9.9.2. Restrictions and Quotients Inherit Splitting

Let V be a finite-dimensional vector space over F , let $T \in \mathcal{L}(V)$, and let U be a T -invariant subspace with $\{0\} \neq U \neq V$. If p_T splits over F , then $p_{T|_U}$ and $p_{\bar{T}}$ split over F , where $\bar{T}$ is the induced operator on V/U .

Idea. Section 1 factors p_T as $p_{T|_U} p_{\bar{T}}$, so the two factors can only have roots that p_T has. A polynomial splits exactly when its root multiplicities add up to its degree (Theorem 6.4.3). Count: the multiplicities of the two factors add up to those of p_T , which use up the whole degree; so neither factor can be short of its own degree.

Proof. By Theorem 9.1.3, $p_T = fg$ with $f = p_{T|_U}$ and $g = p_{\bar{T}}$, both monic (Theorem 7.10.2). Let $c_1, \dots, c_k$ be the distinct roots of p_T in F . Every root of f or of g in F is a root of p_T (Theorem 0.8.5), so it is among the c_i , and a c_i that is not a root of f has $\text{mult}_{c_i}(f) = 0$. Hence, by Theorem 6.4.3 applied to f and to g ,

$$\sum_{i=1}^k \text{mult}_{c_i}(f) \leq \deg f, \quad \sum_{i=1}^k \text{mult}_{c_i}(g) \leq \deg g.$$

Adding, and using Theorem 6.4.2 and then Theorem 6.4.3 for the split polynomial p_T ,

$$\begin{aligned} \deg f + \deg g &= \deg p_T = \sum_{i=1}^k \text{mult}_{c_i}(p_T) \\ &= \sum_{i=1}^k \text{mult}_{c_i}(f) + \sum_{i=1}^k \text{mult}_{c_i}(g) \\ &\leq \deg f + \deg g. \end{aligned}$$

So both inequalities above are equalities, and f and g split over F by Theorem 6.4.3. ■

Now the common eigenvector. We prove it for a whole family of commuting operators at once, possibly infinite, since the proof costs no more and the next theorems need it. A family $\mathcal{F} \subseteq \mathcal{L}(V)$ is **commuting** if $ST = TS$ for all $S, T \in \mathcal{F}$.

Theorem 9.9.3. Commuting Operators Have a Common Eigenvector

Let $V \neq \{0\}$ be a finite-dimensional vector space over F , and let $\mathcal{F} \subseteq \mathcal{L}(V)$ be a nonempty commuting family such that p_T splits over F for every $T \in \mathcal{F}$. Then there is a nonzero $v \in V$ that is an eigenvector of **every** $T \in \mathcal{F}$. This applies to every commuting family when $F = \mathbb{C}$.

Idea. For two operators, the plan above works: an eigenvector of $S|_U$ with $U = E_\lambda(T)$. For many operators we would like to repeat it, shrinking U each time, and a family may be infinite. So we jump straight to the end of the shrinking: take a nonzero subspace W , invariant under the whole family, that is as small as possible. If some $T \in \mathcal{F}$ were not a scalar on W , an eigenspace of $T|_W$ would be a smaller such subspace. So every operator in the family is a scalar on W , and any nonzero vector of W is a common eigenvector.

Proof. Call a subspace $\mathcal{F}$ -**invariant** if it is T -invariant for every $T \in \mathcal{F}$. The space V is a nonzero $\mathcal{F}$ -invariant subspace, so among all nonzero $\mathcal{F}$ -invariant subspaces there is one, W , of smallest dimension.

Let $T \in \mathcal{F}$. Then $T|_W$ is an operator on $W \neq \{0\}$, and $p_{T|_W}$ splits: if $W = V$ this is the hypothesis, and otherwise it is [Lemma 9.9.2](#). Since $p_{T|_W}$ is monic of degree $\dim W \geq 1$, it has a root $\lambda \in F$, which is an eigenvalue of $T|_W$ by Theorem 9.2.1. Put

$$W' = \{w \in W : Tw = \lambda w\} = W \cap E_\lambda(T),$$

which is nonzero because it contains an eigenvector of $T|_W$. For every $S \in \mathcal{F}$, both W and $E_\lambda(T)$ are S -invariant, the first by the choice of W and the second by [Theorem 9.9.1](#); so W' is S -invariant by Proposition 9.1.4. Hence W' is a nonzero $\mathcal{F}$ -invariant subspace contained in W , and by the minimality of $\dim W$, $\dim W' = \dim W$. By Theorem 1.6.8, $W' = W$, that is, $Tw = \lambda w$ for every $w \in W$.

So every $T \in \mathcal{F}$ acts on W as multiplication by a scalar λ_T . Any nonzero $v \in W$ then satisfies $Tv = \lambda_T v$ for every $T \in \mathcal{F}$, as claimed. Over $\mathbb{C}$ every characteristic polynomial splits (Corollary 6.5.5), so the hypothesis holds automatically. ■

The proof never had to name the eigenvalues. It found a subspace on which every operator of the family is a scalar, and the scalars came for free.

Warning. The **splitting hypothesis cannot be dropped, and neither can commut-
ing**. Over $\mathbb{R}$, the rotation $\mathbf{R} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ commutes with itself, yet the family $\{\mathbf{R}\}$ has no real eigenvector at all, since $p_{\mathbf{R}} = x^2 + 1$ has no real root. Over $\mathbb{C}$, the noncommuting pair $\mathbf{E}_{12}, \mathbf{E}_{21}$ has no common eigenvector, as we saw above.

9.9.4 Simultaneous triangularization

Section 7 triangularized one operator by splitting off an eigenvector and passing to the quotient. For a commuting family the same recursion works, because the eigenvector we split off can be chosen common to all.

Theorem 9.9.4. Simultaneous Triangularization

Let V be a finite-dimensional vector space over F with $\dim V = n \geq 1$, and let $\mathcal{F} \subseteq \mathcal{L}(V)$ be a nonempty commuting family such that p_T splits over F for every $T \in \mathcal{F}$. Then there is

one basis $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}^{\mathcal{B}}$ is upper triangular for **every** $T \in \mathcal{F}$.

In particular, if $\mathbf{A}_1, \dots, \mathbf{A}_r \in M_n(\mathbb{C})$ commute pairwise, there is an invertible $\mathbf{P} \in M_n(\mathbb{C})$ such that $\mathbf{P}^{-1}\mathbf{A}_i\mathbf{P}$ is upper triangular for every i .

Idea. Take one common eigenvector, extend, recurse. ① [Theorem 9.9.3](#) gives v_1 , an eigenvector of every $T \in \mathcal{F}$, so $U = \text{span}(v_1)$ is invariant under the family. ② On V/U , the induced operators still commute, and their characteristic polynomials still split. ③ By induction on the dimension, one basis of V/U triangularizes all of them. ④ Lift it back and put v_1 in front; the block triangular form of Section 1 does the rest.

Proof. We use induction on n . If $n = 1$, every 1×1 matrix is upper triangular, so any basis works.

Let $n \geq 2$, and assume the theorem for spaces of dimension $n - 1$. By [Theorem 9.9.3](#) there is $v_1 \neq 0$ with $Tv_1 = \lambda_T v_1$ for every $T \in \mathcal{F}$. Let $U = \text{span}(v_1)$; it is T -invariant for every $T \in \mathcal{F}$ (Proposition 9.1.2), and $\{0\} \neq U \neq V$ since $\dim U = 1 < n$. For $T \in \mathcal{F}$ let $\bar{T}$ be the induced operator on V/U (Definition 9.1.2), and put $\bar{\mathcal{F}} = \{\bar{T} : T \in \mathcal{F}\}$.

The family $\bar{\mathcal{F}}$ is commuting: for $S, T \in \mathcal{F}$ and $v \in V$,

$$\bar{S}\bar{T}(v + U) = \bar{S}(Tv + U) = STv + U = TSv + U = \bar{T}\bar{S}(v + U).$$

Each $p_{\bar{T}}$ splits by [Lemma 9.9.2](#), and $\dim V/U = n - 1$ by Theorem 4.4.5. By the induction hypothesis there are $w_2, \dots, w_n \in V$ such that $\bar{\mathcal{B}} = (w_2 + U, \dots, w_n + U)$ is a basis of V/U in which every $[\bar{T}]_{\bar{\mathcal{B}}}^{\bar{\mathcal{B}}}$ is upper triangular.

Let $\mathcal{B} = (v_1, w_2, \dots, w_n)$. It is linearly independent: if $a_1v_1 + a_2w_2 + \dots + a_nw_n = 0$, then passing to V/U , where $v_1 + U$ is the zero coset, gives $a_2(w_2 + U) + \dots + a_n(w_n + U) = 0$, so $a_2 = \dots = a_n = 0$ by the independence of $\bar{\mathcal{B}}$; then $a_1v_1 = 0$ with $v_1 \neq 0$ gives $a_1 = 0$. A linearly independent list of length $n = \dim V$ is a basis (Theorem 1.6.4). Now $\mathcal{B}$ is a basis of U extended to a basis of V , and $\bar{\mathcal{B}}$ is exactly the list of cosets in Theorem 9.1.3. Hence for every $T \in \mathcal{F}$,

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} \lambda_T & * \\ \mathbf{0} & [\bar{T}]_{\bar{\mathcal{B}}}^{\bar{\mathcal{B}}} \end{pmatrix},$$

with a 1×1 upper-left block, a zero column below it, and an upper triangular lower-right block. Such a matrix is upper triangular. This completes the induction.

For the matrix statement, apply this to the family $\{T_{\mathbf{A}_1}, \dots, T_{\mathbf{A}_r}\}$ on $\mathbb{C}^n$, whose characteristic polynomials split by Corollary 6.5.5, and let $\mathbf{P}$ have the vectors of $\mathcal{B}$ as columns; then $\mathbf{P}^{-1}\mathbf{A}_i\mathbf{P} = [T_{\mathbf{A}_i}]_{\mathcal{B}}^{\mathcal{B}}$ (Theorem 4.2.3). ■

A first use: if S and T commute and both characteristic polynomials split, the diagonal entries of $[S + T]_{\mathcal{B}}^{\mathcal{B}}$ and $[ST]_{\mathcal{B}}^{\mathcal{B}}$ are the sums and products of the diagonal entries of $[S]_{\mathcal{B}}^{\mathcal{B}}$ and $[T]_{\mathcal{B}}^{\mathcal{B}}$, position by position. For the sum this is entrywise; for the product, the (i, i) -entry of $\mathbf{XY}$ is $\sum_k x_{ik}y_{ki}$ (Definition 0.9.9), and when $\mathbf{X}$ and $\mathbf{Y}$ are upper triangular only $k = i$ can contribute, since $x_{ik} = 0$ for $k < i$ and $y_{ki} = 0$ for $k > i$. By Theorem 9.7.4, those diagonal entries are the eigenvalues. So the eigenvalues of $S + T$ are of the form $\lambda + \mu$, with λ an eigenvalue of S

and μ one of T , **paired by position in a common triangular form**. Section 2 warned that nothing like this holds without commuting.

9.9.5 Simultaneous diagonalization

Diagonalization is the stronger wish, and it needs a stronger hypothesis: each operator must be diagonalizable on its own. The name for the goal:

Definition 9.9.1. Simultaneously Diagonalizable

Let V be finite-dimensional. A family $\mathcal{F} \subseteq \mathcal{L}(V)$ is **simultaneously diagonalizable** if there is **one** basis $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}^{\mathcal{B}}$ is diagonal for **every** $T \in \mathcal{F}$. Matrices $\mathbf{A}_1, \dots, \mathbf{A}_r \in M_n(F)$ are simultaneously diagonalizable over F if there is **one** invertible $\mathbf{P} \in M_n(F)$ with every $\mathbf{P}^{-1}\mathbf{A}_i\mathbf{P}$ diagonal.

Equivalently, by the unwinding in Section 4, V has a basis whose vectors are eigenvectors of every $T \in \mathcal{F}$. The opening of this section showed that a simultaneously diagonalizable family is commuting. For diagonalizable operators the converse holds.

Theorem 9.9.5. Simultaneous Diagonalization

Let V be a finite-dimensional vector space over F with $\dim V \geq 1$, and let $S, T \in \mathcal{L}(V)$ be diagonalizable. Then S and T are simultaneously diagonalizable if and only if $ST = TS$.

For matrices: if $\mathbf{A}, \mathbf{B} \in M_n(F)$ are diagonalizable over F , there is an invertible $\mathbf{P} \in M_n(F)$ with $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ and $\mathbf{P}^{-1}\mathbf{B}\mathbf{P}$ both diagonal if and only if $\mathbf{AB} = \mathbf{BA}$.

Idea. Split V into the eigenspaces of T . On each piece T is a scalar, so **any** basis of the piece consists of eigenvectors of T ; we are free to choose it to suit S . The piece is S -invariant because S commutes with T , and S restricted to it is diagonalizable because S is (Section 8). So each piece has a basis of eigenvectors of S , and gluing these bases together gives the common basis.

Proof. ($\Rightarrow$) Let $\mathcal{B}$ be a basis with $[S]_{\mathcal{B}}^{\mathcal{B}}$ and $[T]_{\mathcal{B}}^{\mathcal{B}}$ diagonal. Diagonal matrices commute, so by Corollary 4.1.3, $[ST]_{\mathcal{B}}^{\mathcal{B}} = [S]_{\mathcal{B}}^{\mathcal{B}}[T]_{\mathcal{B}}^{\mathcal{B}} = [T]_{\mathcal{B}}^{\mathcal{B}}[S]_{\mathcal{B}}^{\mathcal{B}} = [TS]_{\mathcal{B}}^{\mathcal{B}}$. The map $R \mapsto [R]_{\mathcal{B}}^{\mathcal{B}}$ is injective (Theorem 4.1.4), so $ST = TS$.

($\Leftarrow$) Suppose $ST = TS$. Let $\lambda_1, \dots, \lambda_k$ be the distinct eigenvalues of T , and put $U_i = E_{\lambda_i}(T)$. Since T is diagonalizable, Theorem 9.4.1 (c) gives

$$V = U_1 \oplus \cdots \oplus U_k.$$

Each U_i is S -invariant (Theorem 9.9.1), and S is diagonalizable, so $S|_{U_i}$ is diagonalizable by Corollary 9.8.3. Hence U_i has a basis $\mathcal{B}_i$ of eigenvectors of $S|_{U_i}$ (Theorem 9.4.1 (b)), which are eigenvectors of S . They are also eigenvectors of T , being nonzero vectors of $U_i = E_{\lambda_i}(T)$. Let $\mathcal{B}$ be the list $\mathcal{B}_1, \dots, \mathcal{B}_k$. Since the sum is direct, $\mathcal{B}$ is a basis of V by Theorem 1.7.5 ((a) $\Rightarrow$ (d)). Every vector of $\mathcal{B}$ is an eigenvector of both S and T , so $[S]_{\mathcal{B}}^{\mathcal{B}}$ and $[T]_{\mathcal{B}}^{\mathcal{B}}$ are diagonal (Theorem 9.4.1, (a) $\Leftrightarrow$ (b)).

For matrices, apply this to $T_{\mathbf{A}}$ and $T_{\mathbf{B}}$ on F^n , which are diagonalizable because $\mathbf{A}$ and $\mathbf{B}$ are

(Definition 9.4.1), and commute because $\mathbf{AB} = \mathbf{BA}$; let $\mathbf{P}$ have the common eigenvectors as columns, as in Section 4. ■

The proof is also an algorithm. Diagonalize $\mathbf{A}$; in each eigenspace of $\mathbf{A}$ of dimension at least 2, write down the matrix of $\mathbf{B}$ restricted to it and diagonalize that small matrix; eigenspaces of dimension 1 need no work, since their basis vector is automatically an eigenvector of $\mathbf{B}$.

Example. Diagonalizing Two Commuting Matrices Together

Let

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 2 \\ -2 & 0 & 3 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & -2 & 2 \\ 0 & 1 & 0 \\ 0 & -2 & 3 \end{pmatrix} \in M_3(\mathbb{R}).$$

Check that $\mathbf{AB} = \mathbf{BA}$, and find an invertible $\mathbf{P}$ such that $\mathbf{P}^{-1}\mathbf{AP}$ and $\mathbf{P}^{-1}\mathbf{BP}$ are both diagonal.

Solution. They commute. Multiplying out,

$$\mathbf{AB} = \begin{pmatrix} 1 & -2 & 2 \\ -2 & 1 & 2 \\ -2 & -2 & 5 \end{pmatrix} = \mathbf{BA}.$$

Diagonalize A. Expanding $\det(x\mathbf{I} - \mathbf{A})$ along the first row, whose only nonzero entry is $x - 1$,

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

For $\lambda = 1$, $\mathbf{A} - \mathbf{I} = \begin{pmatrix} 0 & 0 & 0 \\ -2 & 0 & 2 \\ -2 & 0 & 2 \end{pmatrix}$ has rank 1 and reduces to $x = z$, so $E_1(\mathbf{A})$ is the plane with

basis $\mathbf{u}_1 = (0, 1, 0)$, $\mathbf{u}_2 = (1, 0, 1)$. For $\lambda = 3$, $\mathbf{A} - 3\mathbf{I} = \begin{pmatrix} -2 & 0 & 0 \\ -2 & -2 & 2 \\ -2 & 0 & 0 \end{pmatrix}$ gives $x = 0$ and $y = z$, so $E_3(\mathbf{A}) = \text{span}((0, 1, 1))$. Since $g(1) = 2 = a(1)$ and $g(3) = 1 = a(3)$, $\mathbf{A}$ is diagonalizable.

The same computation for $\mathbf{B}$ gives $p_{\mathbf{B}} = (x - 1)^2(x - 3)$, with $\mathbf{B} - \mathbf{I} = \begin{pmatrix} 0 & -2 & 2 \\ 0 & 0 & 0 \\ 0 & -2 & 2 \end{pmatrix}$ of rank 1 and $\mathbf{B} - 3\mathbf{I}$ of rank 2, so $\mathbf{B}$ is diagonalizable as well.

A basis of eigenvectors of $\mathbf{A}$ need not work for $\mathbf{B}$: $\mathbf{B}\mathbf{u}_1 = (-2, 1, -2)$ is not a multiple of $\mathbf{u}_1$. This is where the search inside the eigenspace comes in.

The one-dimensional eigenspace. $E_3(\mathbf{A})$ is $\mathbf{B}$ -invariant (Theorem 9.9.1) and one-dimensional, so $\mathbf{v}_3 = (0, 1, 1)$ is an eigenvector of $\mathbf{B}$: indeed $\mathbf{B}\mathbf{v}_3 = (0, 1, 1) = 1 \cdot \mathbf{v}_3$.

The plane $E_1(\mathbf{A})$. It is $\mathbf{B}$ -invariant, and

$$\mathbf{B}\mathbf{u}_1 = (-2, 1, -2) = \mathbf{u}_1 - 2\mathbf{u}_2, \quad \mathbf{B}\mathbf{u}_2 = (3, 0, 3) = 3\mathbf{u}_2.$$

So the matrix of $\mathbf{B}|_{E_1(\mathbf{A})}$ in the basis $(\mathbf{u}_1, \mathbf{u}_2)$ is $\mathbf{M} = \begin{pmatrix} 1 & 0 \\ -2 & 3 \end{pmatrix}$, whose columns are the coordinates just found. It is lower triangular with eigenvalues 1 and 3. From $\mathbf{M} - \mathbf{I} = \begin{pmatrix} 0 & 0 \\ -2 & 2 \end{pmatrix}$ the eigenvector is $(1, 1)$, which stands for $\mathbf{u}_1 + \mathbf{u}_2 = (1, 1, 1)$; from $\mathbf{M} - 3\mathbf{I} = \begin{pmatrix} -2 & 0 \\ -2 & 0 \end{pmatrix}$ it is $(0, 1)$, which stands for $\mathbf{u}_2 = (1, 0, 1)$.

The common basis. Put $\mathbf{v}_1 = (1, 0, 1)$, $\mathbf{v}_2 = (1, 1, 1)$, $\mathbf{v}_3 = (0, 1, 1)$ as columns:

$$\mathbf{P} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \text{diag}(1, 1, 3), \quad \mathbf{P}^{-1}\mathbf{B}\mathbf{P} = \text{diag}(3, 1, 1).$$

Check. $\mathbf{A}\mathbf{v}_1 = \mathbf{v}_1$, $\mathbf{A}\mathbf{v}_2 = \mathbf{v}_2$, $\mathbf{A}\mathbf{v}_3 = (0, 3, 3) = 3\mathbf{v}_3$; $\mathbf{B}\mathbf{v}_1 = (3, 0, 3) = 3\mathbf{v}_1$, $\mathbf{B}\mathbf{v}_2 = \mathbf{v}_2$, $\mathbf{B}\mathbf{v}_3 = \mathbf{v}_3$. Expanding along the first row, $\det \mathbf{P} = 1(1 - 1) - 1(0 - 1) + 0 = 1 \neq 0$, so $\mathbf{P}$ is invertible, with $\mathbf{P}^{-1} = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 1 & -1 \\ -1 & 0 & 1 \end{pmatrix}$.

Notice that neither matrix alone determines $\mathbf{P}$: $\mathbf{A}$ cannot tell $\mathbf{v}_1$ from $\mathbf{v}_2$, and $\mathbf{B}$ cannot tell $\mathbf{v}_2$ from $\mathbf{v}_3$. Together they pin down all three lines.

Warning. Commuting does not make an operator diagonalizable. The theorem assumes that both S and T are diagonalizable. $\mathbf{J} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ commutes with $\mathbf{I}_2$, and $\mathbf{I}_2$ is diagonal, but $\mathbf{J}$ is not diagonalizable (Section 4), so no basis makes both diagonal. What survives is [Theorem 9.9.4](#): $\mathbf{J}$ and $\mathbf{I}_2$ are already upper triangular together.

The theorem extends from two operators to any number, and even to infinitely many. An infinite family looks hopeless for an induction, but it lives inside the finite-dimensional space $\mathcal{L}(V)$, so finitely many of its members already span everything the family spans.

Theorem 9.9.6. Simultaneous Diagonalization of a Commuting Family

Let V be a finite-dimensional vector space over F with $\dim V \geq 1$, and let $\mathcal{F} \subseteq \mathcal{L}(V)$ be a family of diagonalizable operators. Then $\mathcal{F}$ is simultaneously diagonalizable if and only if it is commuting.

Idea. ① **Finite first.** For finitely many commuting diagonalizable $T_1, \dots, T_r$, induct on r : split V into eigenspaces of T_1 , and on each piece apply the case $r - 1$ to the restrictions of $T_2, \dots, T_r$. ② **Reduce to finite.** A longest linearly independent list $T_1, \dots, T_r$ taken from

$\mathcal{F}$ spans every member of $\mathcal{F}$, and has length at most $\dim \mathcal{L}(V)$. ③ A basis that diagonalizes $T_1, \dots, T_r$ diagonalizes every linear combination of them.

Proof. ($\Rightarrow$) If one basis makes every $[T]_{\mathcal{B}}^{\mathcal{B}}$ diagonal, any two of these matrices commute, and the argument of Theorem 9.9.5 ($\Rightarrow$) gives $ST = TS$ for all $S, T \in \mathcal{F}$.

($\Leftarrow$) Suppose $\mathcal{F}$ is commuting. If $\mathcal{F}$ is empty, any basis works; assume not.

Step 1: finitely many operators. We prove by induction on $r \geq 1$: for every nonzero finite-dimensional space W and all pairwise commuting diagonalizable $T_1, \dots, T_r \in \mathcal{L}(W)$, some basis of W consists of eigenvectors of every T_i . For $r = 1$ this is Theorem 9.4.1. Let $r \geq 2$, and let $\lambda_1, \dots, \lambda_k$ be the distinct eigenvalues of T_1 , with $U_j = E_{\lambda_j}(T_1) \neq \{0\}$. By Theorem 9.4.1 (c), $W = U_1 \oplus \dots \oplus U_k$. Each U_j is T_i -invariant for every i (Theorem 9.9.1), each $T_i|_{U_j}$ is diagonalizable (Corollary 9.8.3), and the restrictions $T_2|_{U_j}, \dots, T_r|_{U_j}$ commute pairwise, since they agree with $T_2, \dots, T_r$ on U_j . By the induction hypothesis, U_j has a basis $\mathcal{B}_j$ of common eigenvectors of $T_2, \dots, T_r$; they are eigenvectors of T_1 too, as nonzero vectors of U_j . The list $\mathcal{B}_1, \dots, \mathcal{B}_k$ is a basis of W by Theorem 1.7.5 ((a) $\Rightarrow$ (d)). This completes the induction.

Step 2: a finite spanning subfamily. Let $N = \dim \mathcal{L}(V) = (\dim V)^2$ (Theorem 4.1.4). Every linearly independent list in $\mathcal{L}(V)$ has length at most N (Theorem 1.6.5), so among the linearly independent lists of members of $\mathcal{F}$ there is one of greatest length, say $(T_1, \dots, T_r)$, with $r \geq 1$ because a single nonzero member is independent (if every member of $\mathcal{F}$ is 0, any basis works). For every $T \in \mathcal{F}$, the list $(T_1, \dots, T_r, T)$ is longer, hence dependent, so $T \in \text{span}(T_1, \dots, T_r)$ by Lemma 1.5.3.

Step 3. By Step 1 there is a basis $\mathcal{B}$ of V with every $[T_i]_{\mathcal{B}}^{\mathcal{B}}$ diagonal. For $T \in \mathcal{F}$, write $T = c_1 T_1 + \dots + c_r T_r$; then $[T]_{\mathcal{B}}^{\mathcal{B}} = c_1 [T_1]_{\mathcal{B}}^{\mathcal{B}} + \dots + c_r [T_r]_{\mathcal{B}}^{\mathcal{B}}$ (Corollary 4.1.3), a linear combination of diagonal matrices, hence diagonal. So $\mathcal{B}$ diagonalizes every member of $\mathcal{F}$. ■

A typical infinite family is $\{p(T) : p \in F[x]\}$ for a diagonalizable T . It is commuting, each $p(T)$ is diagonalizable (Section 8), and here one basis works for the whole family without any of the theorem's machinery, namely an eigenbasis of T : an eigenvector of T for λ is an eigenvector of $p(T)$ for $p(\lambda)$ (Exercise (B3: Eigenvalues of powers and inverses)). The theorem says the same for families that are not generated by one operator in so visible a way.

9.9.6 Exercises

9.9.6.1 A. Check your understanding

Exercise. A1

- (a) State the theorem relating commuting operators and eigenspaces.
- (b) True or false: if $ST = TS$, then every eigenvector of T is an eigenvector of S . Justify your answer.
- (c) True or false: any two commuting matrices in $M_2(\mathbb{R})$ have a common eigenvector in $\mathbb{R}^2$. Justify your answer.

- (d) True or false: if $\mathbf{A}, \mathbf{B} \in M_n(F)$ are simultaneously diagonalizable, then $\mathbf{AB} = \mathbf{BA}$. Justify your answer.
- (e) True or false: if $\mathbf{AB} = \mathbf{BA}$ and $\mathbf{A}$ is diagonalizable, then $\mathbf{B}$ is diagonalizable. Justify your answer.
- (f) In the proof of [Theorem 9.9.5](#), which hypothesis makes each eigenspace of T invariant under S , and which makes the restriction of S diagonalizable?

Solution.

- (a) If $S, T \in \mathcal{L}(V)$ and $ST = TS$, then $E_\lambda(T)$ is S -invariant for every $\lambda \in F$ ([Theorem 9.9.1](#)).
- (b) False. $T = \mathbf{I}_2$ commutes with $S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$; the vector $\mathbf{e}_2$ is an eigenvector of T , but $S\mathbf{e}_2 = (1, 1)$ is not a multiple of $\mathbf{e}_2$.
- (c) False. The rotation $\mathbf{R} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ commutes with itself and has no real eigenvector, since $p_{\mathbf{R}} = x^2 + 1$ has no real root.
- (d) True. If $\mathbf{P}^{-1}\mathbf{AP} = \mathbf{D}$ and $\mathbf{P}^{-1}\mathbf{BP} = \mathbf{D}'$ are diagonal, then $\mathbf{AB} = \mathbf{PDP}^{-1}\mathbf{PD}'\mathbf{P}^{-1} = \mathbf{PDD}'\mathbf{P}^{-1} = \mathbf{PD}'\mathbf{D}\mathbf{P}^{-1} = \mathbf{BA}$, since diagonal matrices commute.
- (e) False. $\mathbf{A} = \mathbf{I}_2$ is diagonalizable and commutes with $\mathbf{B} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, which is not diagonalizable.
- (f) The hypothesis $ST = TS$ gives the invariance ([Theorem 9.9.1](#)). The hypothesis that S is diagonalizable gives, through Corollary 9.8.3, that $S|_{E_\lambda(T)}$ is diagonalizable.

9.9.6.2 B. Practice

Exercise. B1: Common eigenvectors

Let

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & 2 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 2 & 1 & 0 \\ -1 & 4 & 0 \\ 2 & -2 & 5 \end{pmatrix} \in M_3(\mathbb{R}).$$

- (a) Verify that $\mathbf{AB} = \mathbf{BA}$.
- (b) Find every common eigenvector of $\mathbf{A}$ and $\mathbf{B}$.
- (c) Hence show that $\mathbf{A}$ and $\mathbf{B}$ are not simultaneously diagonalizable, and explain which hypothesis of [Theorem 9.9.5](#) fails.

Solution.

- (a) Multiplying out,

$$\mathbf{AB} = \begin{pmatrix} 2 & 1 & 0 \\ -1 & 4 & 0 \\ 7 & -7 & 10 \end{pmatrix} = \mathbf{BA}.$$

For instance, the $(3, 1)$ -entry of $\mathbf{AB}$ is row $(1, -1, 2)$ of $\mathbf{A}$ times column $(2, -1, 2)$ of

$\mathbf{B}$, namely $2 + 1 + 4 = 7$, and that of $\mathbf{BA}$ is row $(2, -2, 5)$ of $\mathbf{B}$ times column $(1, 0, 1)$ of $\mathbf{A}$, namely $2 + 0 + 5 = 7$.

(b) $\mathbf{A}$ is lower triangular, so $p_{\mathbf{A}} = (x - 1)^2(x - 2)$. $\mathbf{A} - \mathbf{I} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & -1 & 1 \end{pmatrix}$ has rank 1,

so $E_1(\mathbf{A}) = \{(x, y, z) : x - y + z = 0\}$ with basis $\mathbf{u}_1 = (1, 1, 0)$, $\mathbf{u}_2 = (-1, 0, 1)$.

$\mathbf{A} - 2\mathbf{I} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & -1 & 0 \end{pmatrix}$ gives $E_2(\mathbf{A}) = \text{span}(\mathbf{e}_3)$.

A common eigenvector is an eigenvector of $\mathbf{A}$, so it lies in $E_1(\mathbf{A})$ or $E_2(\mathbf{A})$, and it must be an eigenvector of $\mathbf{B}$ there. On the line $E_2(\mathbf{A})$: $\mathbf{B}\mathbf{e}_3 = (0, 0, 5) = 5\mathbf{e}_3$, so every nonzero multiple of $\mathbf{e}_3$ is a common eigenvector. On the plane $E_1(\mathbf{A})$, which is $\mathbf{B}$ -invariant by [Theorem 9.9.1](#):

$$\mathbf{B}\mathbf{u}_1 = (3, 3, 0) = 3\mathbf{u}_1, \quad \mathbf{B}\mathbf{u}_2 = (-2, 1, 3) = \mathbf{u}_1 + 3\mathbf{u}_2.$$

So in the basis $\mathcal{B} = (\mathbf{u}_1, \mathbf{u}_2)$, $[\mathbf{B}|_{E_1(\mathbf{A})}]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}$. Its only eigenvalue is 3, and its eigenvectors are the nonzero multiples of $(1, 0)$, that is, of $\mathbf{u}_1$. Hence the common eigenvectors are exactly the nonzero multiples of $(1, 1, 0)$ and the nonzero multiples of $(0, 0, 1)$.

(c) By (b), any list of common eigenvectors lies in $\text{span}((1, 1, 0), \mathbf{e}_3)$, a plane, so no three of them form a basis of $\mathbb{R}^3$. Hence no basis diagonalizes both. The failing hypothesis is that $\mathbf{B}$ is diagonalizable: $p_{\mathbf{B}} = (x - 3)^2(x - 5)$ (from the block computation, $p_{\mathbf{B}} =$

$p_{\mathbf{B}|_{E_1(\mathbf{A})}} p_{\mathbf{B}}$, or directly), and $\mathbf{B} - 3\mathbf{I} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 1 & 0 \\ 2 & -2 & 2 \end{pmatrix}$ has rank 2, so $g_{\mathbf{B}}(3) = 1 < 2 = a_{\mathbf{B}}(3)$.

Exercise. B2: Decide simultaneous diagonalizability

For each pair of real matrices, determine whether they are simultaneously diagonalizable over $\mathbb{R}$. If they are, find a suitable $\mathbf{P}$. Justify your answer.

(a) $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$

(b) $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$

(c) $\mathbf{A} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}.$

Solution.

(a) Yes. $\mathbf{AB} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = \mathbf{BA}$. Both are diagonalizable: $\mathbf{A}$ has $p_{\mathbf{A}} = x^2 - 2x - 3 = (x - 3)(x + 1)$ and $\mathbf{B}$ has $p_{\mathbf{B}} = x^2 - 1 = (x - 1)(x + 1)$, each with distinct eigenvalues (Corollary 9.4.2). By [Theorem 9.9.5](#) a common $\mathbf{P}$ exists. The eigenspaces of $\mathbf{A}$ are

lines, so their spanning vectors $(1, 1)$ and $(1, -1)$ work: $\mathbf{P} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \text{diag}(3, -1)$, $\mathbf{P}^{-1}\mathbf{B}\mathbf{P} = \text{diag}(1, -1)$, since $\mathbf{B}(1, 1) = (1, 1)$ and $\mathbf{B}(1, -1) = (-1, 1)$.

- (b) No. Both are diagonalizable (distinct eigenvalues 1, 2), but $\mathbf{A}\mathbf{B} = \begin{pmatrix} 1 & 2 \\ 0 & 4 \end{pmatrix} \neq \begin{pmatrix} 1 & 1 \\ 0 & 4 \end{pmatrix} = \mathbf{B}\mathbf{A}$, so by [Theorem 9.9.5](#) ($\Rightarrow$) no common $\mathbf{P}$ exists.
- (c) No. They commute, since $\mathbf{A} = 2\mathbf{I}$. But $\mathbf{B}$ is not diagonalizable: $p_{\mathbf{B}} = (x - 2)^2$ and $\mathbf{B} - 2\mathbf{I}$ has rank 1, so $g(2) = 1 < 2$. If $\mathbf{P}^{-1}\mathbf{B}\mathbf{P}$ were diagonal, $\mathbf{B}$ would be diagonalizable (Definition 9.4.1).

Exercise. B3: Matrices commuting with a diagonal matrix

- (a) Let $\mathbf{D} = \text{diag}(1, 2, 3) \in M_3(\mathbb{R})$. Prove that a matrix $\mathbf{B} \in M_3(\mathbb{R})$ commutes with $\mathbf{D}$ if and only if $\mathbf{B}$ is diagonal.
- (b) Let $\mathbf{D}' = \text{diag}(1, 1, 2) \in M_3(\mathbb{R})$. Describe all $\mathbf{B} \in M_3(\mathbb{R})$ with $\mathbf{B}\mathbf{D}' = \mathbf{D}'\mathbf{B}$, and find one that is not diagonalizable.

Solution.

- (a) ($\Leftarrow$) Diagonal matrices commute. ($\Rightarrow$) Suppose $\mathbf{B}\mathbf{D} = \mathbf{D}\mathbf{B}$. The eigenspaces of $\mathbf{D}$ are $E_1(\mathbf{D}) = \text{span}(\mathbf{e}_1)$, $E_2(\mathbf{D}) = \text{span}(\mathbf{e}_2)$, and $E_3(\mathbf{D}) = \text{span}(\mathbf{e}_3)$. By [Theorem 9.9.1](#) each is $\mathbf{B}$ -invariant, so $\mathbf{B}\mathbf{e}_j \in \text{span}(\mathbf{e}_j)$ for each j . The j -th column of $\mathbf{B}$ is $\mathbf{B}\mathbf{e}_j$, so it has zeros off the diagonal, and $\mathbf{B}$ is diagonal. (Directly: $(\mathbf{D}\mathbf{B})_{ij} = i b_{ij}$ and $(\mathbf{B}\mathbf{D})_{ij} = b_{ij} j$, so $(i - j)b_{ij} = 0$, and $b_{ij} = 0$ for $i \neq j$.)
- (b) As computed in the text for $\text{diag}(1, 1, 2)$, $\mathbf{B}\mathbf{D}' = \mathbf{D}'\mathbf{B}$ exactly when $\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & 0 \\ b_{21} & b_{22} & 0 \\ 0 & 0 & b_{33} \end{pmatrix}$ with arbitrary entries: $E_1(\mathbf{D}') = \text{span}(\mathbf{e}_1, \mathbf{e}_2)$ and $E_2(\mathbf{D}') = \text{span}(\mathbf{e}_3)$ are $\mathbf{B}$ -invariant, which by [Theorem 9.1.5](#) forces this shape, and conversely $\mathbf{D}' = \mathbf{I}_2 \oplus (2)$ commutes with every such block matrix ([Theorem 8.1.4](#)). For example $\mathbf{B} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ commutes with $\mathbf{D}'$, and it is not diagonalizable: $p_{\mathbf{B}} = x(x - 1)^2$ and $\mathbf{B} - \mathbf{I}$ has rank 2, so $g(1) = 1 < 2 = a(1)$.

9.9.6.3 C. Going deeper

Exercise. C1: The commutant of a matrix with distinct eigenvalues

Let $\mathbf{A} \in M_n(F)$, $n \geq 1$, have n distinct eigenvalues $\lambda_1, \dots, \lambda_n \in F$.

- (a) Prove that if $\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A}$, then $\mathbf{B}$ is diagonalizable and every eigenvector of $\mathbf{A}$ is an eigenvector of $\mathbf{B}$.
- (b) Prove that if $\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A}$, then $\mathbf{B} = q(\mathbf{A})$ for some $q \in F[x]$ of degree at most $n - 1$.
- (c) Deduce that $\{\mathbf{B} \in M_n(F) : \mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A}\}$ is a subspace of $M_n(F)$ of dimension n , with basis $(\mathbf{I}, \mathbf{A}, \dots, \mathbf{A}^{n-1})$.

Hint: for (b), use Lagrange interpolation.

Solution.

- (a) By Corollary 9.4.2, $\mathbf{A}$ is diagonalizable, say $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$ with $\mathbf{D}$ diagonal. A diagonal matrix is upper triangular, so its diagonal entries are the eigenvalues of $\mathbf{A}$, each occurring as often as its algebraic multiplicity (Theorem 9.7.4 (b)); here there are n distinct eigenvalues and n diagonal places, so each occurs once, and after permuting the columns of $\mathbf{P}$ and the diagonal entries of $\mathbf{D}$ together we may assume $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_n)$. Let $\mathbf{v}_i$ be column i of $\mathbf{P}$. Each $E_{\lambda_i}(\mathbf{A})$ has dimension at least 1, and the dimensions add up to at most n (Corollary 9.2.3), so each has dimension exactly 1 and $E_{\lambda_i}(\mathbf{A}) = \text{span}(\mathbf{v}_i)$. Every eigenvector of $\mathbf{A}$ for λ_i is a nonzero multiple of $\mathbf{v}_i$. By Theorem 9.9.1, $\mathbf{B}\mathbf{v}_i \in \text{span}(\mathbf{v}_i)$, so $\mathbf{B}\mathbf{v}_i = \mu_i\mathbf{v}_i$ for some $\mu_i \in F$, and $\mathbf{v}_i$, with every nonzero multiple of it, is an eigenvector of $\mathbf{B}$. The columns of $\mathbf{P}$ form a basis of eigenvectors of $\mathbf{B}$, so $\mathbf{P}^{-1}\mathbf{B}\mathbf{P} = \text{diag}(\mu_1, \dots, \mu_n)$ and $\mathbf{B}$ is diagonalizable.
- (b) The nodes $\lambda_1, \dots, \lambda_n$ are distinct, so by Theorem 6.6.1 there is $q \in F[x]$ with $\deg q \leq n-1$ and $q(\lambda_i) = \mu_i$ for every i . By Theorem 9.4.3,

$$q(\mathbf{A}) = \mathbf{P} \text{diag}(q(\lambda_1), \dots, q(\lambda_n)) \mathbf{P}^{-1} = \mathbf{P} \text{diag}(\mu_1, \dots, \mu_n) \mathbf{P}^{-1} = \mathbf{B}.$$

- (c) If $\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A}$ and $\mathbf{A}\mathbf{B}' = \mathbf{B}'\mathbf{A}$, then $\mathbf{A}(c\mathbf{B} + \mathbf{B}') = c\mathbf{B}\mathbf{A} + \mathbf{B}'\mathbf{A} = (c\mathbf{B} + \mathbf{B}')\mathbf{A}$, and $\mathbf{A}\mathbf{O} = \mathbf{O}\mathbf{A}$; so the set $\mathcal{C}$ is a subspace. Each $\mathbf{A}^k$ commutes with $\mathbf{A}$, so $\text{span}(\mathbf{I}, \mathbf{A}, \dots, \mathbf{A}^{n-1}) \subseteq \mathcal{C}$, and (b) gives the reverse inclusion. For independence, let $c_0\mathbf{I} + c_1\mathbf{A} + \dots + c_{n-1}\mathbf{A}^{n-1} = \mathbf{O}$ and put $q = \sum_k c_k x^k$. By Theorem 9.4.3, $\mathbf{O} = q(\mathbf{A}) = \mathbf{P} \text{diag}(q(\lambda_1), \dots, q(\lambda_n)) \mathbf{P}^{-1}$, so $q(\lambda_i) = 0$ for all i . A polynomial of degree at most $n-1$ with n distinct roots is zero (Corollary 6.1.4), so every $c_k = 0$. Hence $(\mathbf{I}, \mathbf{A}, \dots, \mathbf{A}^{n-1})$ is a basis of $\mathcal{C}$, and $\dim \mathcal{C} = n$.

Exercise. C2: Commuting nilpotent matrices

Recall that $\mathbf{N} \in M_n(F)$ is **nilpotent** if $\mathbf{N}^a = \mathbf{O}$ for some $a \geq 1$ (Exercise (C2: Nilpotent operators)). Let $\mathbf{N}, \mathbf{M} \in M_n(F)$ with $\mathbf{NM} = \mathbf{MN}$, $\mathbf{N}^a = \mathbf{O}$, and $\mathbf{M}^b = \mathbf{O}$ for some $a, b \geq 1$.

- (a) Prove that $\mathbf{NM}$ is nilpotent.
- (b) Prove that $(\mathbf{N} + \mathbf{M})^{a+b-1} = \mathbf{O}$, so that $\mathbf{N} + \mathbf{M}$ is nilpotent.
- (c) Show that both conclusions can fail without $\mathbf{NM} = \mathbf{MN}$.
- (d) Now let $F = \mathbb{C}$. Give a second proof of (b) using Theorem 9.9.4.

Hint: for (b), expand the power as a sum of products of $a+b-1$ factors, each equal to $\mathbf{N}$ or $\mathbf{M}$.

Solution.

- (a) Since $\mathbf{NM} = \mathbf{MN}$, induction on k gives $(\mathbf{NM})^k = \mathbf{N}^k \mathbf{M}^k$: if it holds for k , then $(\mathbf{NM})^{k+1} = \mathbf{N}^k \mathbf{M}^k \mathbf{NM} = \mathbf{N}^k \mathbf{NM}^k \mathbf{M} = \mathbf{N}^{k+1} \mathbf{M}^{k+1}$, where $\mathbf{M}^k \mathbf{N} = \mathbf{NM}^k$ follows from $\mathbf{MN} = \mathbf{NM}$ by induction as well. With $k = a$, $(\mathbf{NM})^a = \mathbf{N}^a \mathbf{M}^a = \mathbf{O} \cdot \mathbf{M}^a = \mathbf{O}$.
- (b) Let $s = a+b-1$. By distributivity, $(\mathbf{N} + \mathbf{M})^s$ is the sum of all 2^s products $\mathbf{X}_1 \mathbf{X}_2 \dots \mathbf{X}_s$ with each $\mathbf{X}_t \in \{\mathbf{N}, \mathbf{M}\}$. Since $\mathbf{N}$ and $\mathbf{M}$ commute, each such product can be rearranged, one adjacent swap at a time, into $\mathbf{N}^i \mathbf{M}^{s-i}$, where i is the number of factors

equal to $\mathbf{N}$. If $i \geq a$, then $\mathbf{N}^i = \mathbf{N}^a \mathbf{N}^{i-a} = \mathbf{O}$. If $i \leq a-1$, then $s-i \geq s-a+1 = b$, so $\mathbf{M}^{s-i} = \mathbf{O}$. Either way the product is $\mathbf{O}$, so $(\mathbf{N} + \mathbf{M})^s = \mathbf{O}$.

- (c) Let $\mathbf{N} = \mathbf{E}_{12}$ and $\mathbf{M} = \mathbf{E}_{21}$ in $M_2(F)$. Both square to $\mathbf{O}$. But $\mathbf{NM} = \mathbf{E}_{11}$ satisfies $\mathbf{E}_{11}^k = \mathbf{E}_{11} \neq \mathbf{O}$ for all $k \geq 1$, and $\mathbf{N} + \mathbf{M} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ satisfies $(\mathbf{N} + \mathbf{M})^2 = \mathbf{I}$, so $(\mathbf{N} + \mathbf{M})^k$ is $\mathbf{I}$ or $\mathbf{N} + \mathbf{M}$, never $\mathbf{O}$. Neither is nilpotent, and indeed $\mathbf{NM} = \mathbf{E}_{11} \neq \mathbf{E}_{22} = \mathbf{MN}$.
- (d) By [Theorem 9.9.4](#) there is an invertible $\mathbf{P} \in M_n(\mathbb{C})$ with $\mathbf{U} = \mathbf{P}^{-1}\mathbf{N}\mathbf{P}$ and $\mathbf{U}' = \mathbf{P}^{-1}\mathbf{M}\mathbf{P}$ 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; likewise for $\mathbf{U}'$. So $\mathbf{U} + \mathbf{U}' = \mathbf{P}^{-1}(\mathbf{N} + \mathbf{M})\mathbf{P}$ is upper triangular with zero diagonal. A matrix $\mathbf{X}$ of this kind is nilpotent: $\mathbf{X}\mathbf{e}_1 = \mathbf{0}$ and $\mathbf{X}\mathbf{e}_j \in \text{span}(\mathbf{e}_1, \dots, \mathbf{e}_{j-1})$, so by induction $\mathbf{X}^j\mathbf{e}_j = \mathbf{0}$ and $\mathbf{X}^n = \mathbf{O}$. Hence $(\mathbf{N} + \mathbf{M})^n = \mathbf{P}(\mathbf{U} + \mathbf{U}')^n\mathbf{P}^{-1} = \mathbf{O}$.