

12.10 The Sylvester Equation

Chapter 8 wrote an operator with an invariant subspace as a block triangular matrix, with a corner block $\mathbf{C}$ recording how the subspace fails to have an invariant complement (Theorem 8.1.7). Chapter 10 spent its length on what to do when that corner cannot be cleared. The question of *when* it can be cleared has an exact answer, and the answer is a linear equation in the unknown matrix $\mathbf{X}$:

$$\mathbf{AX} - \mathbf{XB} = \mathbf{C}.$$

Chapter 8 already turned this equation into an ordinary linear system with the Kronecker product, and stopped one step short of the criterion because that step needs eigenvalues. We have eigenvalues now, and Schur's factorization of Section 1 to organize them. This section proves the criterion, uses it to clear corners, and reads off what it says about the matrices that commute with a given one.

Throughout, the field is $\mathbb{C}$ unless stated otherwise, and $\mathbf{A} \in M_m(\mathbb{C})$, $\mathbf{B} \in M_n(\mathbb{C})$, while the unknown $\mathbf{X}$ and the data $\mathbf{C}$ live in $M_{m \times n}(\mathbb{C})$.

12.10.1 The Sylvester operator

The left-hand side of the equation is linear in $\mathbf{X}$: replacing $\mathbf{X}$ by $\mathbf{X} + \mathbf{X}'$ or by $c\mathbf{X}$ does the same to $\mathbf{AX} - \mathbf{XB}$, because matrix multiplication is bilinear (Theorem 0.9.4). So the whole question is about one linear map, and it deserves a name.

Solving $\mathbf{AX} - \mathbf{XB} = \mathbf{C}$ means inverting a single linear operator on the space of $m \times n$ matrices.

Definition 12.10.1. Sylvester operator

Let $\mathbf{A} \in M_m(F)$ and $\mathbf{B} \in M_n(F)$. The **Sylvester operator** of the pair $(\mathbf{A}, \mathbf{B})$ is

$$\mathcal{S}_{\mathbf{A}, \mathbf{B}}: M_{m \times n}(F) \rightarrow M_{m \times n}(F), \quad \mathcal{S}_{\mathbf{A}, \mathbf{B}}(\mathbf{X}) = \mathbf{AX} - \mathbf{XB}.$$

The equation $\mathbf{AX} - \mathbf{XB} = \mathbf{C}$, for a given $\mathbf{C} \in M_{m \times n}(F)$, is the **Sylvester equation**.

In words: $\mathcal{S}_{\mathbf{A}, \mathbf{B}}$ multiplies by $\mathbf{A}$ on the left, multiplies by $\mathbf{B}$ on the right, and subtracts. It is an operator on a space of dimension mn , not on F^m or F^n , and that is the only thing that makes the equation look harder than $\mathbf{Ax} = \mathbf{b}$.

Three special cases are worth keeping in view.

1. $\mathbf{B} = \mathbf{O}$. Then $\mathcal{S}_{\mathbf{A}, \mathbf{O}}(\mathbf{X}) = \mathbf{AX}$, and the equation is n copies of $\mathbf{Ax} = \mathbf{b}$, one for each column of $\mathbf{X}$ (Exercise (B2: One-sided equations)). It is uniquely solvable for every $\mathbf{C}$ exactly when $\mathbf{A}$ is invertible, that is, when $0 \notin \text{spec}(\mathbf{A})$. Since $\text{spec}(\mathbf{O}) = \{0\}$, this already matches the criterion we are heading for.
2. $\mathbf{B} = \mathbf{A}$ and $\mathbf{C} = \mathbf{O}$. Then $\ker \mathcal{S}_{\mathbf{A}, \mathbf{A}}$ is the **commutant** of $\mathbf{A}$, the set of matrices commuting with $\mathbf{A}$. This is the question Chapter 8 opened with, and Chapter 10 §04 studied it for a single Jordan or Weyr block.
3. $m = n = 1$. Then $\mathcal{S}_{a, b}(x) = (a - b)x$, which is invertible exactly when $a \neq b$. The general criterion will be this one, run over all pairs of eigenvalues.

Nonexample by minimal change. The map $\mathbf{X} \mapsto \mathbf{AX} - \mathbf{XB} - \mathbf{X}^\top$ is still linear on $M_n(F)$, but it is not a Sylvester operator: the transpose is not of the form “multiply on one side”. Everything below uses the two-sided product shape and nothing else.

Chapter 8 already wrote the matrix of $\mathcal{S}_{\mathbf{A},\mathbf{B}}$. Stacking the columns of $\mathbf{X}$ with the vectorization vec of Definition 8.4.2 and applying the vec identity Theorem 8.4.4 twice, once with $\mathbf{B} = \mathbf{I}_n$ and once with $\mathbf{A} = \mathbf{I}_m$,

$$\begin{aligned}\text{vec}(\mathcal{S}_{\mathbf{A},\mathbf{B}}(\mathbf{X})) &= \mathbf{K} \text{vec} \mathbf{X}, \\ \mathbf{K} &:= \mathbf{I}_n \otimes \mathbf{A} - \mathbf{B}^\top \otimes \mathbf{I}_m.\end{aligned}\tag{12.10.1}$$

Here $\mathbf{K} \in M_{mn}(F)$, and $\text{vec}: M_{m \times n}(F) \rightarrow F^{mn}$ is an isomorphism. So Eq. 12.10.1 says that $\mathbf{K}$ is the matrix of $\mathcal{S}_{\mathbf{A},\mathbf{B}}$ with respect to the basis of matrix units ordered column by column. In particular $\mathcal{S}_{\mathbf{A},\mathbf{B}}$ is invertible if and only if $\mathbf{K}$ is.

That is exactly where Chapter 8 left the problem. What it could not do was compute the eigenvalues of $\mathbf{K}$.

12.10.2 The spectrum of a Kronecker sum

Chapter 8 proved how the Kronecker product treats rank, trace, and determinant (Theorem 8.4.3) and how it treats products, transposes, and inverses (Theorem 8.4.1). It proved nothing about eigenvalues, and Exercise (C2: Kronecker sums) got only as far as the case where both factors are already triangular. Schur closes the gap, because over $\mathbb{C}$ every matrix can be made triangular unitarily, and the mixed product rule carries a pair of unitary conjugations through the Kronecker product in one step.

Lemma 12.10.1. The Eigenvalues of a Kronecker Sum

Let $\mathbf{A} \in M_m(\mathbb{C})$ with eigenvalues $\lambda_1, \dots, \lambda_m$ and let $\mathbf{B} \in M_n(\mathbb{C})$ with eigenvalues $\mu_1, \dots, \mu_n$, each list repeated according to algebraic multiplicity. Then the mn eigenvalues of

$$\mathbf{K} = \mathbf{I}_n \otimes \mathbf{A} - \mathbf{B}^\top \otimes \mathbf{I}_m \in M_{mn}(\mathbb{C}),$$

listed with algebraic multiplicity, are the mn differences $\lambda_i - \mu_j$ ($1 \leq i \leq m$, $1 \leq j \leq n$).

Idea. Triangularize both ingredients at once. Schur (Corollary 12.1.3) writes $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^*$ and $\mathbf{B}^\top = \mathbf{V}\mathbf{S}\mathbf{V}^*$ with $\mathbf{T}, \mathbf{S}$ upper triangular. The mixed product rule then conjugates *both* summands of $\mathbf{K}$ by the single unitary matrix $\mathbf{V} \otimes \mathbf{U}$, because $\mathbf{I}_n \otimes \mathbf{A}$ and $\mathbf{B}^\top \otimes \mathbf{I}_m$ each have one trivial factor to absorb the other conjugation. What is left is $\mathbf{I}_n \otimes \mathbf{T} - \mathbf{S} \otimes \mathbf{I}_m$, and a Kronecker product with one triangular and one identity factor is triangular, so the eigenvalues are visible on the diagonal.

Proof. By Corollary 12.1.3 there are a unitary $\mathbf{U} \in M_m(\mathbb{C})$ and an upper triangular $\mathbf{T} \in M_m(\mathbb{C})$ with $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^*$ and $t_{ii} = \lambda_i$. By Proposition 9.10.5 (b), $p_{\mathbf{B}^\top} = p_{\mathbf{B}}$, so $\mathbf{B}^\top$ has eigenvalue list $\mu_1, \dots, \mu_n$; applying Corollary 12.1.3 to $\mathbf{B}^\top$ gives a unitary $\mathbf{V} \in M_n(\mathbb{C})$ and an upper triangular $\mathbf{S} \in M_n(\mathbb{C})$ with $\mathbf{B}^\top = \mathbf{V}\mathbf{S}\mathbf{V}^*$ and $s_{jj} = \mu_j$.

Put $\mathbf{W} = \mathbf{V} \otimes \mathbf{U} \in M_{mn}(\mathbb{C})$. Conjugating every entry of $\mathbf{A} \otimes \mathbf{B}$ conjugates every block $a_{ij}\mathbf{B}$, so $\overline{\mathbf{A} \otimes \mathbf{B}} = \overline{\mathbf{A}} \otimes \overline{\mathbf{B}}$ (Definition 8.4.1), and combining this with Theorem 8.4.1 (d) gives $(\mathbf{A} \otimes \mathbf{B})^* = \mathbf{A}^* \otimes \mathbf{B}^*$ whenever the product is defined. Hence $\mathbf{W}^* = \mathbf{V}^* \otimes \mathbf{U}^*$, and by the

mixed product rule Theorem 8.4.1 (c),

$$\mathbf{W}^* \mathbf{W} = (\mathbf{V}^* \mathbf{V}) \otimes (\mathbf{U}^* \mathbf{U}) = \mathbf{I}_n \otimes \mathbf{I}_m = \mathbf{I}_{mn},$$

so $\mathbf{W}$ is unitary. Using the mixed product rule twice more,

$$\mathbf{W}(\mathbf{I}_n \otimes \mathbf{T})\mathbf{W}^* = (\mathbf{V}\mathbf{I}_n\mathbf{V}^*) \otimes (\mathbf{U}\mathbf{T}\mathbf{U}^*) = \mathbf{I}_n \otimes \mathbf{A},$$

$$\mathbf{W}(\mathbf{S} \otimes \mathbf{I}_m)\mathbf{W}^* = (\mathbf{V}\mathbf{S}\mathbf{V}^*) \otimes (\mathbf{U}\mathbf{I}_m\mathbf{U}^*) = \mathbf{B}^\top \otimes \mathbf{I}_m.$$

Subtracting, $\mathbf{W}(\mathbf{I}_n \otimes \mathbf{T} - \mathbf{S} \otimes \mathbf{I}_m)\mathbf{W}^* = \mathbf{K}$, so $\mathbf{K}$ is similar to

$$\mathbf{R} := \mathbf{I}_n \otimes \mathbf{T} - \mathbf{S} \otimes \mathbf{I}_m.$$

R is upper triangular. Label the rows and columns of an $mn \times mn$ matrix by pairs as in Definition 8.4.1: row $(j-1)m+i$ with $1 \leq j \leq n$, $1 \leq i \leq m$, and likewise column $(l-1)m+k$. In those labels the entries of the two summands are

$$(\mathbf{I}_n \otimes \mathbf{T})_{(j-1)m+i, (l-1)m+k} = \delta_{jl} t_{ik},$$

$$(\mathbf{S} \otimes \mathbf{I}_m)_{(j-1)m+i, (l-1)m+k} = s_{jl} \delta_{ik}.$$

The first is nonzero only when $j = l$ and $i \leq k$, since $\mathbf{T}$ is upper triangular; then $(j-1)m+i \leq (l-1)m+k$. The second is nonzero only when $j \leq l$ and $i = k$; if $j = l$ the two labels are equal, and if $j < l$ then

$$(j-1)m+i \leq (j-1)m+m = jm \leq (l-1)m < (l-1)m+k,$$

using $i \leq m$, $j \leq l-1$, and $k \geq 1$. So each summand vanishes below the diagonal, and so does $\mathbf{R}$.

The diagonal. Taking $j = l$ and $i = k$ above, the diagonal entry of $\mathbf{R}$ in position $(j-1)m+i$ is $t_{ii} - s_{jj} = \lambda_i - \mu_j$. As (i, j) runs over all mn pairs, so does the position $(j-1)m+i$, by division with remainder.

By Theorem 9.7.4 (a), applied to the operator $T_{\mathbf{R}}$ on $\mathbb{C}^{mn}$ and the standard basis, $p_{\mathbf{R}} = \prod_{i,j} (x - (\lambda_i - \mu_j))$. Since $\mathbf{K} \sim \mathbf{R}$, Theorem 7.10.3 gives $p_{\mathbf{K}} = p_{\mathbf{R}}$. This proves the lemma. ■

Two remarks on what the proof used. The hypothesis $F = \mathbb{C}$ entered once, in Schur, and through Schur it is the Fundamental Theorem of Algebra again. And the unitarity of $\mathbf{W}$ was a convenience, not a necessity: any similarity triangularizing $\mathbf{A}$ and $\mathbf{B}^\top$ would do, but Schur hands us one for free and the Kronecker product of two unitary matrices is unitary, which is worth recording on its own.

12.10.3 When the equation is uniquely solvable

The criterion now costs one line.

Theorem 12.10.2. Sylvester's Theorem

Let $\mathbf{A} \in M_m(\mathbb{C})$ and $\mathbf{B} \in M_n(\mathbb{C})$. The following are equivalent.

- (a) $\text{spec}(\mathbf{A}) \cap \text{spec}(\mathbf{B}) = \emptyset$: the two matrices have **no common eigenvalue**.

- (b) The Sylvester operator $\mathcal{S}_{\mathbf{A},\mathbf{B}}: \mathbf{X} \mapsto \mathbf{A}\mathbf{X} - \mathbf{X}\mathbf{B}$ is invertible on $M_{m \times n}(\mathbb{C})$.
- (c) For every $\mathbf{C} \in M_{m \times n}(\mathbb{C})$, the equation $\mathbf{A}\mathbf{X} - \mathbf{X}\mathbf{B} = \mathbf{C}$ has **exactly one** solution $\mathbf{X}$.
- (d) The only $\mathbf{X} \in M_{m \times n}(\mathbb{C})$ with $\mathbf{A}\mathbf{X} = \mathbf{X}\mathbf{B}$ is $\mathbf{X} = \mathbf{O}$.

Proof. By Eq. 12.10.1 the isomorphism vec turns $\mathcal{S}_{\mathbf{A},\mathbf{B}}$ into the matrix $\mathbf{K} = \mathbf{I}_n \otimes \mathbf{A} - \mathbf{B}^\top \otimes \mathbf{I}_m$: a matrix $\mathbf{X}$ solves $\mathbf{A}\mathbf{X} - \mathbf{X}\mathbf{B} = \mathbf{C}$ if and only if the vector $\text{vec } \mathbf{X}$ solves $\mathbf{K}\mathbf{y} = \text{vec } \mathbf{C}$, and every $\mathbf{y} \in \mathbb{C}^{mn}$ and every $\mathbf{z} \in \mathbb{C}^{mn}$ arise this way. Transported through vec , statements (b), (c), and (d) say respectively that $\mathbf{K}$ is invertible, that $\mathbf{K}\mathbf{y} = \mathbf{z}$ has exactly one solution for every $\mathbf{z}$, and that $\mathbf{K}\mathbf{y} = \mathbf{0}$ forces $\mathbf{y} = \mathbf{0}$. These three are equivalent by Theorem 3.4.4, parts (a), (e), and (b): (e) gives existence for every $\mathbf{z}$, and (b) upgrades it to uniqueness, since two solutions differ by a member of the null space.

It remains to match (a) with (b). By Theorem 9.2.6, $\mathbf{K}$ is invertible if and only if $0 \notin \text{spec}(\mathbf{K})$. By Lemma 12.10.1, $\text{spec}(\mathbf{K}) = \{\lambda_i - \mu_j\}$, where $\lambda_1, \dots, \lambda_m$ and $\mu_1, \dots, \mu_n$ list the eigenvalues of $\mathbf{A}$ and of $\mathbf{B}$. So $0 \in \text{spec}(\mathbf{K})$ if and only if $\lambda_i = \mu_j$ for some pair (i, j) , that is, if and only if $\mathbf{A}$ and $\mathbf{B}$ have a common eigenvalue. Negating both sides matches (a) with (b), and the proof is complete. ■

This is the statement Chapter 8 promised. At the end of Example (A Sylvester Equation as a Linear System) it was said that over $\mathbb{C}$ the equation $\mathbf{A}\mathbf{X} - \mathbf{X}\mathbf{B} = \mathbf{C}$ is uniquely solvable for every $\mathbf{C}$ exactly when $\mathbf{A}$ and $\mathbf{B}$ have no common eigenvalue, and that the criterion needs eigenvalues and would be proved in Chapter 12. Parts (a) and (c) of Theorem 12.10.2 are that sentence, and the eigenvalues it needed were supplied by Lemma 12.10.1. The quantitative version — *how badly* conditioned the equation is when the spectra are close but disjoint — is a separate matter, and belongs with perturbation theory in Chapter 20.

Part (d) is the version used most often in practice, because it is a statement with no data in it: to show that some matrix is zero, exhibit it as a solution of $\mathbf{A}\mathbf{X} = \mathbf{X}\mathbf{B}$ for a pair with disjoint spectra. The corollary on commutants below is that move and nothing else.

Example. Solving a Sylvester Equation

Over $\mathbb{C}$, solve $\mathbf{A}\mathbf{X} - \mathbf{X}\mathbf{B} = \mathbf{C}$ for

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 4 & -2 \\ 0 & 4 \end{pmatrix}.$$

Solution. The matrix $\mathbf{A}$ is upper triangular, so $\text{spec}(\mathbf{A}) = \{2, 3\}$ (Theorem 9.7.4); $\mathbf{B}$ is lower triangular, so $\text{spec}(\mathbf{B}) = \{1, -1\}$. The spectra are disjoint, and by Theorem 12.10.2 there is exactly one solution.

To find it, use [Eq. 12.10.1](#). Here $\mathbf{B}^\top = \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}$, so by Definition 8.4.1

$$\mathbf{I}_2 \otimes \mathbf{A} = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix}, \quad \mathbf{B}^\top \otimes \mathbf{I}_2 = \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix},$$

and their difference is

$$\mathbf{K} = \begin{pmatrix} 1 & 1 & -2 & 0 \\ 0 & 2 & 0 & -2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 4 \end{pmatrix}.$$

Its diagonal entries are 1, 2, 3, 4, which are the four differences $2-1$, $3-1$, $2-(-1)$, $3-(-1)$, as [Lemma 12.10.1](#) predicts; none is zero, confirming invertibility.

With $\text{vec } \mathbf{X} = (x_{11}, x_{21}, x_{12}, x_{22})$ and $\text{vec } \mathbf{C} = (4, 0, -2, 4)$, back substitution in $\mathbf{K} \text{vec } \mathbf{X} = \text{vec } \mathbf{C}$ gives $4x_{22} = 4$, so $x_{22} = 1$; then $3x_{12} + 1 = -2$, so $x_{12} = -1$; then $2x_{21} - 2 = 0$, so $x_{21} = 1$; then $x_{11} + 1 + 2 = 4$, so $x_{11} = 1$. Hence

$$\mathbf{X} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Check: $\mathbf{A}\mathbf{X} = \begin{pmatrix} 3 & -1 \\ 3 & 3 \end{pmatrix}$ and $\mathbf{X}\mathbf{B} = \begin{pmatrix} -1 & 1 \\ 3 & -1 \end{pmatrix}$, and their difference is $\mathbf{C}$.

Quick check. Let $\mathbf{A} \in M_3(\mathbb{C})$ have eigenvalues 1, 2, 2 and let $\mathbf{B} \in M_2(\mathbb{C})$ have eigenvalues 0, 3. List the eigenvalues of $\mathcal{S}_{\mathbf{A}, \mathbf{B}}$ with multiplicity, and decide whether $\mathbf{A}\mathbf{X} - \mathbf{X}\mathbf{B} = \mathbf{C}$ is uniquely solvable for every $\mathbf{C} \in M_{3 \times 2}(\mathbb{C})$.

Solution. By [Lemma 12.10.1](#) the six eigenvalues are the differences $\lambda_i - \mu_j$:

$$1 - 0 = 1, \quad 2 - 0 = 2, \quad 2 - 0 = 2, \quad 1 - 3 = -2, \quad 2 - 3 = -1, \quad 2 - 3 = -1.$$

None is zero — equivalently $\text{spec}(\mathbf{A}) = \{1, 2\}$ and $\text{spec}(\mathbf{B}) = \{0, 3\}$ are disjoint — so by [Theorem 12.10.2](#) the equation has exactly one solution for every $\mathbf{C}$.

Warning. The criterion depends on the sign. For the equation $\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{B} = \mathbf{C}$, which is the form Chapter 8 used in Example (A Sylvester Equation as a Linear System), the relevant operator is $\mathcal{S}_{\mathbf{A}, -\mathbf{B}}$, and $\text{spec}(-\mathbf{B}) = \{-\mu_j\}$. So the condition is $\lambda_i + \mu_j \neq 0$ for all i, j , **not** $\text{spec}(\mathbf{A}) \cap \text{spec}(\mathbf{B}) = \emptyset$. The two conditions are genuinely different: with $\mathbf{A} = \text{diag}(1, 2)$ and $\mathbf{B} = \begin{pmatrix} -1 & 5 \\ 0 & 3 \end{pmatrix}$ the spectra $\{1, 2\}$ and $\{-1, 3\}$ are disjoint, yet $1 + (-1) = 0$, so $\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{B} = \mathbf{C}$ is not solvable for every $\mathbf{C}$. Read the sign off the equation before quoting the criterion.

12.10.4 Clearing the corner

We can now answer the question the section opened with.

Corollary 12.10.3. Block Diagonalization by a Sylvester Solution

Let $\mathbf{A} \in M_m(F)$, $\mathbf{B} \in M_n(F)$, and $\mathbf{C} \in M_{m \times n}(F)$, and put

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{O} & \mathbf{B} \end{pmatrix} \in M_{m+n}(F).$$

- (a) If $\mathbf{X} \in M_{m \times n}(F)$ satisfies $\mathbf{AX} - \mathbf{XB} = -\mathbf{C}$, then with $\mathbf{P} = \begin{pmatrix} \mathbf{I}_m & \mathbf{X} \\ \mathbf{O} & \mathbf{I}_n \end{pmatrix}$ we have $\mathbf{P}^{-1}\mathbf{MP} = \mathbf{A} \oplus \mathbf{B}$. In particular $\mathbf{M} \sim \mathbf{A} \oplus \mathbf{B}$.
- (b) Conversely, if $\mathbf{P}^{-1}\mathbf{MP} = \mathbf{A} \oplus \mathbf{B}$ for some $\mathbf{X} \in M_{m \times n}(F)$ and the matrix $\mathbf{P}$ built from it as in (a), then that $\mathbf{X}$ satisfies $\mathbf{AX} - \mathbf{XB} = -\mathbf{C}$.
- (c) If $F = \mathbb{C}$ and $\text{spec}(\mathbf{A}) \cap \text{spec}(\mathbf{B}) = \emptyset$, then a solution $\mathbf{X}$ exists, and it is unique; so $\mathbf{M} \sim \mathbf{A} \oplus \mathbf{B}$ for **every** corner $\mathbf{C}$.

Idea. The matrices $\begin{pmatrix} \mathbf{I}_m & \mathbf{X} \\ \mathbf{O} & \mathbf{I}_n \end{pmatrix}$ are exactly the changes of basis that add multiples of the first block of basis vectors to the second and leave the first alone. Conjugating $\mathbf{M}$ by one of them cannot touch the two diagonal blocks or create anything below the diagonal; all it can do is move the corner, and the block multiplication shows it moves the corner by $\mathbf{AX} - \mathbf{XB}$. So the corner can be driven to $\mathbf{O}$ precisely when $-\mathbf{C}$ lies in the image of the Sylvester operator.

Proof.

- (a) First, $\mathbf{P}$ is invertible with

$$\mathbf{P}^{-1} = \begin{pmatrix} \mathbf{I}_m & -\mathbf{X} \\ \mathbf{O} & \mathbf{I}_n \end{pmatrix},$$

since multiplying the two matrices in either order gives $\mathbf{I}_{m+n}$ by Theorem 8.1.1, the partitions being conformable. The same theorem gives

$$\mathbf{MP} = \begin{pmatrix} \mathbf{A} & \mathbf{AX} + \mathbf{C} \\ \mathbf{O} & \mathbf{B} \end{pmatrix},$$

and then

$$\mathbf{P}^{-1}\mathbf{MP} = \begin{pmatrix} \mathbf{A} & \mathbf{AX} + \mathbf{C} - \mathbf{XB} \\ \mathbf{O} & \mathbf{B} \end{pmatrix}.$$

By hypothesis $\mathbf{AX} - \mathbf{XB} = -\mathbf{C}$, so the corner is $\mathbf{O}$ and the result is $\mathbf{A} \oplus \mathbf{B}$ (Definition 8.1.2).

- (b) The computation in (a) used nothing about $\mathbf{X}$ until the last sentence. So if that $\mathbf{P}$ satisfies $\mathbf{P}^{-1}\mathbf{MP} = \mathbf{A} \oplus \mathbf{B}$, comparing the corners of the two sides gives $\mathbf{AX} + \mathbf{C} - \mathbf{XB} = \mathbf{O}$, which is the equation.
- (c) By Theorem 12.10.2 ((a) $\Rightarrow$ (c)), the equation $\mathbf{AX} - \mathbf{XB} = -\mathbf{C}$ has exactly one solution. Apply (a).


Example. Clearing a Corner

Block-diagonalize

$$\mathbf{M} = \begin{pmatrix} 2 & 1 & -4 & 2 \\ 0 & 3 & 0 & -4 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & -1 \end{pmatrix} \in M_4(\mathbb{Q}).$$

Solution. Partition $\mathbf{M}$ into 2×2 blocks. The diagonal blocks are the matrices $\mathbf{A}$ and $\mathbf{B}$ of [Example \(Solving a Sylvester Equation\)](#), and the corner is

$$\begin{pmatrix} -4 & 2 \\ 0 & -4 \end{pmatrix} = -\mathbf{C}$$

for the $\mathbf{C}$ of that example. So the equation to solve is $\mathbf{A}\mathbf{X} - \mathbf{X}\mathbf{B} = -(-\mathbf{C}) = \mathbf{C}$, which is the one already solved there, with the unique solution $\mathbf{X} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$. By [Corollary 12.10.3](#) (a), taking

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

gives $\mathbf{P}^{-1}\mathbf{M}\mathbf{P} = \mathbf{A} \oplus \mathbf{B}$, that is,

$$\mathbf{P}^{-1}\mathbf{M}\mathbf{P} = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & -1 \end{pmatrix}.$$

The eigenvalues 2, 3, 1, -1 are of course unchanged; what has changed is that the space has been split into two invariant pieces of dimension 2.

Remark. Part (b) restricts the shape of $\mathbf{P}$, and one may ask whether that is a real restriction: if $\mathbf{M}$ is similar to $\mathbf{A} \oplus \mathbf{B}$ by *any* invertible matrix, must the Sylvester equation be solvable? The answer is yes over every field. This is **Roth's removal rule**, and the proof is not a block computation: it goes through the characteristic matrix and the criterion $\mathbf{A} \sim \mathbf{B}$ if and only if $x\mathbf{I} - \mathbf{A} \approx x\mathbf{I} - \mathbf{B}$ over $F[x]$ (Theorem 10.7.5). We do not need it here and do not prove it; nothing later in the book depends on it.

Primary decomposition, seen again. Put [Corollary 12.10.3](#) next to Schur and a familiar theorem drops out with no extra work. Let $\mathbf{A} \in M_N(\mathbb{C})$ have distinct eigenvalues $\lambda_1, \dots, \lambda_k$ with algebraic multiplicities $a_1, \dots, a_k$. By Corollary 12.1.3 we may prescribe the order of the diagonal in a Schur factorization, so choose the order that lists all copies of λ_1 first, then all

copies of λ_2 , and so on. Partitioning the resulting triangular matrix along $N = a_1 + \cdots + a_k$ makes it block upper triangular with diagonal blocks $\mathbf{T}_1, \dots, \mathbf{T}_k$, where $\mathbf{T}_i$ is triangular with the single diagonal value λ_i , so $\text{spec}(\mathbf{T}_i) = \{\lambda_i\}$. Now peel the blocks off one at a time: $\text{spec}(\mathbf{T}_1)$ is disjoint from the spectrum of the rest, so [Corollary 12.10.3](#) (c) splits off $\mathbf{T}_1$, and induction on k finishes. The conclusion is that $\mathbf{A} \sim \mathbf{T}_1 \oplus \cdots \oplus \mathbf{T}_k$ with $\text{spec}(\mathbf{T}_i) = \{\lambda_i\}$ — the matrix form of the primary decomposition of Theorem 9.8.1, here obtained from a triangular form and a linear equation rather than from the kernel-splitting lemma.

12.10.5 Commuting with a matrix of disjoint spectra

Here is the promised use of [Theorem 12.10.2](#) (d). It says that disjoint spectra force a block structure on anything that commutes.

Corollary 12.10.4. Disjoint Spectra Split the Commutant

Let $\mathbf{A}_1 \in M_{n_1}(\mathbb{C}), \dots, \mathbf{A}_r \in M_{n_r}(\mathbb{C})$ have **pairwise disjoint** spectra, and put $\mathbf{D} = \mathbf{A}_1 \oplus \cdots \oplus \mathbf{A}_r \in M_N(\mathbb{C})$, $N = n_1 + \cdots + n_r$. Let $\mathbf{Z} \in M_N(\mathbb{C})$ satisfy $\mathbf{D}\mathbf{Z} = \mathbf{Z}\mathbf{D}$, and partition $\mathbf{Z}$ into blocks $\mathbf{Z}_{pq} \in M_{n_p \times n_q}(\mathbb{C})$ for the partition $N = n_1 + \cdots + n_r$. Then

$$\mathbf{Z}_{pq} = \mathbf{O} \quad \text{for all } p \neq q,$$

so $\mathbf{Z} = \mathbf{Z}_{11} \oplus \cdots \oplus \mathbf{Z}_{rr}$ is block **diagonal**, and each diagonal block satisfies $\mathbf{A}_p \mathbf{Z}_{pp} = \mathbf{Z}_{pp} \mathbf{A}_p$.

Proof. By Theorem 8.1.1, the (p, q) block of $\mathbf{D}\mathbf{Z}$ is $\mathbf{A}_p \mathbf{Z}_{pq}$ and the (p, q) block of $\mathbf{Z}\mathbf{D}$ is $\mathbf{Z}_{pq} \mathbf{A}_q$, because $\mathbf{D}$ has only one nonzero block in each block row and in each block column. So $\mathbf{D}\mathbf{Z} = \mathbf{Z}\mathbf{D}$ is equivalent to

$$\mathbf{A}_p \mathbf{Z}_{pq} = \mathbf{Z}_{pq} \mathbf{A}_q \quad (1 \leq p, q \leq r).$$

Let $p \neq q$. Then $\text{spec}(\mathbf{A}_p) \cap \text{spec}(\mathbf{A}_q) = \emptyset$ by hypothesis, so [Theorem 12.10.2](#) ((a) $\Rightarrow$ (d)) forces $\mathbf{Z}_{pq} = \mathbf{O}$. The diagonal equations $\mathbf{A}_p \mathbf{Z}_{pp} = \mathbf{Z}_{pp} \mathbf{A}_p$ are the remaining cases $p = q$, and they carry no information beyond commuting. ■

Read the statement twice, because a similar-sounding claim is false.

Warning. It is the disjointness of the spectra that does the work, not the block diagonal shape. Nothing here says that a matrix commuting with a block diagonal matrix is block diagonal, and nothing here says that a commutant consists of triangular matrices. Both fail as soon as two diagonal blocks share an eigenvalue. Chapter 10 §04 computes the commutant of $\mathbf{J} = \mathbf{J}_2(0) \oplus \mathbf{J}_1(0)$ explicitly and finds it to be

$$\left\{ \begin{pmatrix} a & b & c \\ 0 & a & 0 \\ 0 & h & i \end{pmatrix} : a, b, c, h, i \in F \right\},$$

which contains matrices with a nonzero entry **below** the diagonal, in position (3, 2). That

does not contradict [Corollary 12.10.4](#): the two blocks $\mathbf{J}_2(0)$ and $\mathbf{J}_1(0)$ have the same spectrum $\{0\}$, so the corollary has nothing to say about them. What is true for a single repeated eigenvalue is weaker and needs a different normal form: the commutant of a basic **Weyr** matrix is block upper triangular for the Weyr partition (Proposition 10.4.6), and the Jordan form has no such property. Note too that even when [Corollary 12.10.4](#) applies, it says nothing about the diagonal blocks $\mathbf{Z}_{pp}$ themselves, which range over the whole commutant of $\mathbf{A}_p$.

The corollary is the reason spectral separation is so useful. Combined with the primary decomposition above, it says that over $\mathbb{C}$ the commutant of any matrix splits along its distinct eigenvalues: after the change of basis that block-diagonalizes $\mathbf{A}$ by spectrum, everything commuting with $\mathbf{A}$ becomes block diagonal too, and the hard part of describing a commutant is always the single-eigenvalue part. The pleasant case of the same phenomenon appeared in Theorem 12.7.1, where the pieces are honest eigenspaces and the splitting is orthogonal.

12.10.6 Lyapunov's equation

One specialization is important enough to have its own name.

Definition 12.10.2. Lyapunov equation

Let $\mathbf{A} \in M_n(\mathbb{C})$ and $\mathbf{Q} \in M_n(\mathbb{C})$. The **Lyapunov equation** for the pair $(\mathbf{A}, \mathbf{Q})$ is

$$\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{A}^* = -\mathbf{Q},$$

with unknown $\mathbf{X} \in M_n(\mathbb{C})$. Call $\mathbf{A}$ **stable** when every $\lambda \in \text{spec}(\mathbf{A})$ has $\text{Re } \lambda < 0$.

This is the Sylvester equation with $\mathbf{B} = -\mathbf{A}^*$, since $\mathbf{A}\mathbf{X} - \mathbf{X}(-\mathbf{A}^*) = \mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{A}^*$. To apply [Theorem 12.10.2](#) we need $\text{spec}(-\mathbf{A}^*)$.

The eigenvalues of $\mathbf{A}^*$ are the conjugates of those of $\mathbf{A}$. Indeed $\det(\mathbf{M}^*) = \overline{\det \mathbf{M}}$ for every $\mathbf{M} \in M_n(\mathbb{C})$, because $\det(\mathbf{M}^\top) = \det \mathbf{M}$ (Theorem 7.4.5) and the determinant is a sum of products of entries with coefficients ± 1 (Definition 7.4.1), so conjugating every entry conjugates the value. Applying this to $\mathbf{M} = \lambda \mathbf{I} - \mathbf{A}$, whose conjugate transpose is $\bar{\lambda} \mathbf{I} - \mathbf{A}^*$, gives $\det(\bar{\lambda} \mathbf{I} - \mathbf{A}^*) = \overline{\det(\lambda \mathbf{I} - \mathbf{A})}$. So one determinant vanishes exactly when the other does, and $\lambda \in \text{spec}(\mathbf{A})$ if and only if $\bar{\lambda} \in \text{spec}(\mathbf{A}^*)$ (Theorem 9.2.1).

Therefore $\text{spec}(-\mathbf{A}^*) = \{-\bar{\lambda} : \lambda \in \text{spec}(\mathbf{A})\}$, and [Theorem 12.10.2](#) says the Lyapunov equation has exactly one solution for every $\mathbf{Q}$ precisely when

$$\lambda + \bar{\mu} \neq 0 \quad \text{for all } \lambda, \mu \in \text{spec}(\mathbf{A}).$$

If $\mathbf{A}$ is stable this holds, because then $\text{Re}(\lambda + \bar{\mu}) = \text{Re } \lambda + \text{Re } \mu < 0$. So a stable matrix gives a uniquely solvable Lyapunov equation, whatever $\mathbf{Q}$ is.

What makes the equation important is a statement we are not yet equipped to prove. Call a Hermitian matrix $\mathbf{X}$ **positive definite** when $\mathbf{x}^* \mathbf{X} \mathbf{x} > 0$ for every nonzero $\mathbf{x} \in \mathbb{C}^n$. Then: *if $\mathbf{A}$ is stable and $\mathbf{Q}$ is positive definite, the unique solution $\mathbf{X}$ of $\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{A}^* = -\mathbf{Q}$ is positive definite; and conversely, if some positive definite $\mathbf{X}$ solves the equation for some positive definite $\mathbf{Q}$, then $\mathbf{A}$ is stable.* Positivity is the subject of Chapter 13, where the tools to handle $\mathbf{x}^* \mathbf{X} \mathbf{x}$ are built, and the consequence for differential equations — that a stable $\mathbf{A}$ makes every solution of $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ decay, with $\mathbf{x}^* \mathbf{X} \mathbf{x}$ as the quantity that measures the decay — belongs to Chapter 16. We record the equation here because its *solvability*, which is the part Sylvester's theorem settles, is exactly what makes the rest possible.

12.10.7 Exercises

12.10.7.1 A. Check your understanding

Exercise. A1

- (a) Define the Sylvester operator of a pair $(\mathbf{A}, \mathbf{B})$ with $\mathbf{A} \in M_m(F)$, $\mathbf{B} \in M_n(F)$, and state the dimension of the space it acts on.
- (b) Write down the matrix of the Sylvester operator with respect to the vectorization vec , and say which Chapter 8 result produces it.
- (c) State the criterion for $\mathbf{AX} - \mathbf{XB} = \mathbf{C}$ to have exactly one solution for every $\mathbf{C}$, over $\mathbb{C}$.
- (d) True or false: if $\mathbf{AX} - \mathbf{XB} = \mathbf{C}$ has at least one solution for every $\mathbf{C}$, then that solution is unique. Justify your answer.
- (e) True or false: if $\text{spec}(\mathbf{A}) \cap \text{spec}(\mathbf{B}) = \emptyset$, then $\mathbf{AX} + \mathbf{XB} = \mathbf{C}$ has exactly one solution for every $\mathbf{C}$. Justify your answer.

Solution.

- (a) $\mathcal{S}_{\mathbf{A}, \mathbf{B}}(\mathbf{X}) = \mathbf{AX} - \mathbf{XB}$ for $\mathbf{X} \in M_{m \times n}(F)$ (Definition 12.10.1). It acts on $M_{m \times n}(F)$, of dimension mn .
- (b) $\mathbf{K} = \mathbf{I}_n \otimes \mathbf{A} - \mathbf{B}^\top \otimes \mathbf{I}_m$, by Eq. 12.10.1; it comes from the vec identity Theorem 8.4.4 applied twice, with $\text{vec}(\mathbf{AX}) = (\mathbf{I}_n \otimes \mathbf{A}) \text{vec } \mathbf{X}$ and $\text{vec}(\mathbf{XB}) = (\mathbf{B}^\top \otimes \mathbf{I}_m) \text{vec } \mathbf{X}$.
- (c) Exactly when $\mathbf{A}$ and $\mathbf{B}$ have no common eigenvalue, that is $\text{spec}(\mathbf{A}) \cap \text{spec}(\mathbf{B}) = \emptyset$ (Theorem 12.10.2).
- (d) True. Existence for every $\mathbf{C}$ says $\mathcal{S}_{\mathbf{A}, \mathbf{B}}$ is surjective; an operator on the finite-dimensional space $M_{m \times n}(F)$ is surjective if and only if it is injective (Theorem 3.4.4 through vec , or Rank–Nullity), and injectivity is uniqueness.
- (e) False. The criterion for the $+$ form is $\lambda + \mu \neq 0$ for all $\lambda \in \text{spec}(\mathbf{A})$, $\mu \in \text{spec}(\mathbf{B})$, since $\mathbf{AX} + \mathbf{XB} = \mathcal{S}_{\mathbf{A}, -\mathbf{B}}(\mathbf{X})$. With $\mathbf{A} = \text{diag}(1, 2)$ and $\mathbf{B} = \text{diag}(-1, 3)$ the spectra are disjoint but $1 + (-1) = 0$, so the equation is not uniquely solvable.

12.10.7.2 B. Practice

Exercise. B1: A Sylvester equation by vectorization

Over $\mathbb{Q}$, let

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

- (a) Compute $\text{spec}(\mathbf{A})$ and $\text{spec}(\mathbf{B})$, and decide whether $\mathbf{AX} - \mathbf{XB} = \mathbf{C}$ has exactly one solution.
- (b) Write down $\mathbf{K} = \mathbf{I}_2 \otimes \mathbf{A} - \mathbf{B}^\top \otimes \mathbf{I}_2$, and check that its diagonal entries are the four differences $\lambda_i - \mu_j$.
- (c) Solve the equation.

Solution.

- (a) $\mathbf{A}$ is upper triangular with diagonal 3, 1, so $\text{spec}(\mathbf{A}) = \{3, 1\}$ (Theorem 9.7.4). For $\mathbf{B}$, $p_{\mathbf{B}} = \det \begin{pmatrix} x+1 & 0 \\ -1 & x \end{pmatrix} = x(x+1)$, so $\text{spec}(\mathbf{B}) = \{0, -1\}$. The spectra are disjoint, so by Theorem 12.10.2 there is exactly one solution for this (or any) $\mathbf{C}$.

- (b) $\mathbf{B}^{\top} = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$, so

$$\mathbf{I}_2 \otimes \mathbf{A} = \begin{pmatrix} 3 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{B}^{\top} \otimes \mathbf{I}_2 = \begin{pmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

and

$$\mathbf{K} = \begin{pmatrix} 4 & 1 & -1 & 0 \\ 0 & 2 & 0 & -1 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Its diagonal is 4, 2, 3, 1, and the differences $\lambda_i - \mu_j$ are $3 - (-1) = 4$, $1 - (-1) = 2$, $3 - 0 = 3$, $1 - 0 = 1$, as Lemma 12.10.1 asserts.

- (c) With $\text{vec } \mathbf{X} = (x_{11}, x_{21}, x_{12}, x_{22})$ and $\text{vec } \mathbf{C} = (2, 0, 1, -2)$, back substitution gives $x_{22} = -2$; then $3x_{12} + x_{22} = 1$, so $x_{12} = 1$; then $2x_{21} - x_{22} = 0$, so $x_{21} = -1$; then $4x_{11} + x_{21} - x_{12} = 2$, so $x_{11} = 1$. Hence

$$\mathbf{X} = \begin{pmatrix} 1 & 1 \\ -1 & -2 \end{pmatrix}.$$

Check: $\mathbf{A}\mathbf{X} = \begin{pmatrix} 2 & 1 \\ -1 & -2 \end{pmatrix}$ and $\mathbf{X}\mathbf{B} = \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}$, and their difference is $\mathbf{C}$.

Exercise. B2: Block-diagonalizing a 3×3 matrix

Let

$$\mathbf{M} = \begin{pmatrix} 2 & 1 & 2 \\ 0 & 2 & 3 \\ 0 & 0 & 5 \end{pmatrix} \in M_3(\mathbb{Q}).$$

- (a) Partition $\mathbf{M}$ as $3 = 2 + 1$, name the blocks $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and explain why Corollary 12.10.3 applies.
 (b) Solve the corresponding Sylvester equation.
 (c) Write down $\mathbf{P}$ and verify that $\mathbf{P}^{-1}\mathbf{M}\mathbf{P} = \mathbf{A} \oplus \mathbf{B}$.

Solution.

- (a) The blocks are $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$, $\mathbf{B} = (5)$, and $\mathbf{C} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$. Both diagonal blocks are

triangular, so $\text{spec}(\mathbf{A}) = \{2\}$ and $\text{spec}(\mathbf{B}) = \{5\}$ (Theorem 9.7.4); these are disjoint, so [Corollary 12.10.3](#) (c) applies and a unique $\mathbf{X} \in M_{2 \times 1}(\mathbb{Q})$ exists.

- (b) With $\mathbf{X} = \mathbf{x} = (x_1, x_2)$, the equation $\mathbf{A}\mathbf{x} - 5\mathbf{x} = -\mathbf{C}$ reads $(\mathbf{A} - 5\mathbf{I}_2)\mathbf{x} = -\mathbf{C}$, that is

$$\begin{pmatrix} -3 & 1 \\ 0 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -2 \\ -3 \end{pmatrix}.$$

The second row gives $x_2 = 1$, and the first then gives $-3x_1 + 1 = -2$, so $x_1 = 1$. Hence $\mathbf{x} = (1, 1)$.

- (c) Taking

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{P}^{-1} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix},$$

we get $\mathbf{MP} = \begin{pmatrix} 2 & 1 & 5 \\ 0 & 2 & 5 \\ 0 & 0 & 5 \end{pmatrix}$ and then

$$\mathbf{P}^{-1}\mathbf{MP} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 5 \end{pmatrix} = \mathbf{A} \oplus \mathbf{B},$$

as [Corollary 12.10.3](#) (a) predicts.

Exercise. B3: Which equations are uniquely solvable?

Determine, for each of the following, whether the stated equation has exactly one solution $\mathbf{X} \in M_2(\mathbb{C})$ for **every** right-hand side $\mathbf{C} \in M_2(\mathbb{C})$. Justify your answer.

- (a) $\mathbf{AX} - \mathbf{XB} = \mathbf{C}$ with $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$.
- (b) $\mathbf{AX} - \mathbf{XB} = \mathbf{C}$ with $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.
- (c) $\mathbf{AX} + \mathbf{XB} = \mathbf{C}$ with $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} -1 & 5 \\ 0 & 3 \end{pmatrix}$.
- (d) $\mathbf{AX} - \mathbf{XB} = \mathbf{C}$ with $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$.
- (e) $\mathbf{AX} - \mathbf{XB} = \mathbf{C}$ with $\mathbf{A} \in M_2(\mathbb{C})$ nilpotent and $\mathbf{B} \in M_2(\mathbb{C})$ invertible.

Solution.

- (a) Yes. $\text{spec}(\mathbf{A}) = \{1\}$ and $\text{spec}(\mathbf{B}) = \{2, 3\}$ are disjoint (Theorem 9.7.4), so [Theorem 12.10.2](#) applies.
- (b) No. Both characteristic polynomials are $x^2 + 1$, so $\text{spec}(\mathbf{A}) = \text{spec}(\mathbf{B}) = \{i, -i\}$, which is as far from disjoint as possible. A witness is easy to produce here: $\mathbf{X} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

gives $\mathbf{A}\mathbf{X} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} = \mathbf{X}\mathbf{B}$, so $\mathbf{X}$ is a nonzero member of $\ker \mathcal{S}_{\mathbf{A},\mathbf{B}}$ and condition (d) of [Theorem 12.10.2](#) fails.

- (c) No. This is the $+$ form, so the operator is $\mathcal{S}_{\mathbf{A},-\mathbf{B}}$ and the condition is $\lambda + \mu \neq 0$. Here $\text{spec}(\mathbf{A}) = \{1, 2\}$ and $\text{spec}(\mathbf{B}) = \{-1, 3\}$, and $1 + (-1) = 0$.
- (d) No. Both are triangular with diagonal 1, 1, so $\text{spec}(\mathbf{A}) = \text{spec}(\mathbf{B}) = \{1\}$.
- (e) Yes. A nilpotent matrix has $\text{spec}(\mathbf{A}) = \{0\}$, since $\mathbf{A}^k = \mathbf{O}$ forces $\lambda^k = 0$ for every eigenvalue λ ; and $0 \notin \text{spec}(\mathbf{B})$ because $\mathbf{B}$ is invertible ([Theorem 9.2.6](#)). So the spectra are disjoint.

12.10.7.3 C. Going deeper

Exercise. C1: Eigenvectors of the Sylvester operator

Let $\mathbf{A} \in M_m(\mathbb{C})$ and $\mathbf{B} \in M_n(\mathbb{C})$, and let $\mathcal{S} = \mathcal{S}_{\mathbf{A},\mathbf{B}}$.

- (a) Suppose $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ with $\mathbf{x} \in \mathbb{C}^m$ nonzero, and suppose $\mathbf{y} \in \mathbb{C}^n$ is a left eigenvector of $\mathbf{B}$ for μ , so that $\mathbf{y}^\top \mathbf{B} = \mu \mathbf{y}^\top$ ([Definition 9.10.1](#)). Prove that $\mathbf{x}\mathbf{y}^\top \neq \mathbf{O}$ and that $\mathcal{S}(\mathbf{x}\mathbf{y}^\top) = (\lambda - \mu)\mathbf{x}\mathbf{y}^\top$.
- (b) Deduce that every difference $\lambda - \mu$ with $\lambda \in \text{spec}(\mathbf{A})$ and $\mu \in \text{spec}(\mathbf{B})$ is an eigenvalue of $\mathcal{S}$.
- (c) Hence give a second proof that $\mathcal{S}$ is not invertible when $\mathbf{A}$ and $\mathbf{B}$ have a common eigenvalue, one that uses no Kronecker products.

Solution.

- (a) Choose i with $x_i \neq 0$ and j with $y_j \neq 0$; these exist because $\mathbf{x}$ and $\mathbf{y}$ are nonzero. The (i, j) entry of $\mathbf{x}\mathbf{y}^\top$ is $x_i y_j \neq 0$, so $\mathbf{x}\mathbf{y}^\top \neq \mathbf{O}$. For the eigenvalue equation, associativity of matrix multiplication ([Theorem 0.9.4](#)) gives

$$\mathcal{S}(\mathbf{x}\mathbf{y}^\top) = (\mathbf{A}\mathbf{x})\mathbf{y}^\top - \mathbf{x}(\mathbf{y}^\top \mathbf{B}) = \lambda\mathbf{x}\mathbf{y}^\top - \mu\mathbf{x}\mathbf{y}^\top,$$

which is $(\lambda - \mu)\mathbf{x}\mathbf{y}^\top$.

- (b) Let $\lambda \in \text{spec}(\mathbf{A})$ and $\mu \in \text{spec}(\mathbf{B})$. A right eigenvector $\mathbf{x}$ for λ exists by [Definition 9.2.1](#), and a left eigenvector $\mathbf{y}$ for μ exists by [Proposition 9.10.5 \(b\)](#). By (a), $\mathbf{x}\mathbf{y}^\top$ is a nonzero matrix with $\mathcal{S}(\mathbf{x}\mathbf{y}^\top) = (\lambda - \mu)\mathbf{x}\mathbf{y}^\top$, so $\lambda - \mu$ is an eigenvalue of $\mathcal{S}$ with eigenvector $\mathbf{x}\mathbf{y}^\top$.
- (c) If $\lambda \in \text{spec}(\mathbf{A}) \cap \text{spec}(\mathbf{B})$, apply (b) with $\mu = \lambda$: then 0 is an eigenvalue of $\mathcal{S}$, so $\ker \mathcal{S} \neq \{\mathbf{O}\}$ and $\mathcal{S}$ is not invertible. This shows one implication of [Theorem 12.10.2](#) directly. (It does not give the converse: knowing every $\lambda - \mu$ is an eigenvalue does not by itself show there are no others.)

Exercise. C2: The kernel when the spectra meet

Let $\mathbf{A} \in M_m(\mathbb{C})$ and $\mathbf{B} \in M_n(\mathbb{C})$ have eigenvalue lists $\lambda_1, \dots, \lambda_m$ and $\mu_1, \dots, \mu_n$ with algebraic multiplicity, and let $d = \#\{(i, j) : \lambda_i = \mu_j\}$.

- (a) Prove that $\dim \ker \mathcal{S}_{\mathbf{A},\mathbf{B}} \geq 1$ if and only if $d \geq 1$.

- (b) Prove that $\dim \ker \mathcal{S}_{\mathbf{A}, \mathbf{B}} \leq d$.
 (c) Show that the inequality in (b) can be strict, by computing $\ker \mathcal{S}_{\mathbf{A}, \mathbf{B}}$ for $\mathbf{A} = \mathbf{B} = \mathbf{J}_2(0)$.

Solution.

- (a) $d \geq 1$ says exactly that $\mathbf{A}$ and $\mathbf{B}$ have a common eigenvalue, and by [Theorem 12.10.2](#) ((a) $\Leftrightarrow$ (d)) that is exactly the failure of $\ker \mathcal{S}_{\mathbf{A}, \mathbf{B}} = \{\mathbf{O}\}$.
 (b) The operator $\mathcal{S}_{\mathbf{A}, \mathbf{B}}$ has matrix $\mathbf{K}$ ([Eq. 12.10.1](#)), so $\dim \ker \mathcal{S}_{\mathbf{A}, \mathbf{B}} = g_{\mathbf{K}}(0)$, the geometric multiplicity of 0 as an eigenvalue of $\mathbf{K}$ (when $0 \notin \text{spec}(\mathbf{K})$ both sides are 0 and there is nothing to prove). By [Lemma 12.10.1](#) the eigenvalue list of $\mathbf{K}$ consists of the mn differences $\lambda_i - \mu_j$, so the algebraic multiplicity is $a_{\mathbf{K}}(0) = d$. By [Theorem 9.3.1](#), $g_{\mathbf{K}}(0) \leq a_{\mathbf{K}}(0) = d$.
 (c) Here $m = n = 2$ and $\lambda_1 = \lambda_2 = \mu_1 = \mu_2 = 0$, so $d = 4$. The kernel is the commutant of $\mathbf{J} = \mathbf{J}_2(0)$: writing $\mathbf{X} = \begin{pmatrix} p & q \\ r & s \end{pmatrix}$,

$$\mathbf{J}\mathbf{X} = \begin{pmatrix} r & s \\ 0 & 0 \end{pmatrix}, \quad \mathbf{X}\mathbf{J} = \begin{pmatrix} 0 & p \\ 0 & r \end{pmatrix},$$

so $\mathbf{J}\mathbf{X} = \mathbf{X}\mathbf{J}$ forces $r = 0$ and $s = p$, and leaves q free. Hence $\ker \mathcal{S}_{\mathbf{J}, \mathbf{J}} = \text{span}(\mathbf{I}_2, \mathbf{J})$, of dimension $2 < 4 = d$.

Exercise. C3: Only the scalars commute with everything

Let F be a field, $n \geq 2$, and let $\mathbf{E}_{ij} \in M_n(F)$ be the matrix unit with 1 in position (i, j) and 0 elsewhere.

- (a) Suppose $\mathbf{A} \in M_n(F)$ satisfies $\mathbf{A}\mathbf{E}_{ij} = \mathbf{E}_{ij}\mathbf{A}$ for all $i \neq j$. Prove that $\mathbf{A} = c\mathbf{I}_n$ for some $c \in F$.
 (b) Deduce that the only matrices commuting with every $\mathbf{X} \in M_n(F)$ are the scalar matrices, that is, $\ker \mathcal{S}_{\mathbf{A}, \mathbf{A}} = M_n(F)$ forces $\mathbf{A}$ scalar.
 (c) Deduce that the only matrices commuting with every **invertible** matrix are again the scalar matrices.

Hint for (c): $\mathbf{I}_n + \mathbf{E}_{ij}$ is invertible when $i \neq j$.

Solution.

- (a) Fix $i \neq j$. By [Theorem 0.9.2](#), $\mathbf{A}\mathbf{E}_{ij}$ has column j equal to $\mathbf{A}\mathbf{e}_i$, that is column i of $\mathbf{A}$, and all other columns zero; while $\mathbf{E}_{ij}\mathbf{A}$ has row i equal to row j of $\mathbf{A}$, and all other rows zero. Compare entries.
 Take $k \neq i$. The (k, j) entry of $\mathbf{A}\mathbf{E}_{ij}$ is a_{ki} , and the (k, j) entry of $\mathbf{E}_{ij}\mathbf{A}$ is 0, because row $k \neq i$ of $\mathbf{E}_{ij}\mathbf{A}$ is zero. Hence $a_{ki} = 0$ whenever $k \neq i$. Since $n \geq 2$, every index i has some $j \neq i$ available, so this holds for every i , and $\mathbf{A}$ is diagonal.
 Now the (i, j) entry: on the left it is a_{ii} (the i -th entry of column i of $\mathbf{A}$), and on the right it is a_{jj} (the j -th entry of row j of $\mathbf{A}$). So $a_{ii} = a_{jj}$ for all $i \neq j$. Writing c for this common value, $\mathbf{A} = c\mathbf{I}_n$.
 (b) If $\mathbf{A}\mathbf{X} = \mathbf{X}\mathbf{A}$ for every $\mathbf{X} \in M_n(F)$, then in particular for $\mathbf{X} = \mathbf{E}_{ij}$ with $i \neq j$, so

- (a) gives $\mathbf{A} = c\mathbf{I}_n$. Conversely every scalar matrix commutes with everything. Hence $\ker \mathcal{S}_{\mathbf{A}, \mathbf{A}} = M_n(F)$ if and only if $\mathbf{A}$ is scalar.
- (c) For $i \neq j$ we have $\mathbf{E}_{ij}^2 = \mathbf{O}$, so $(\mathbf{I}_n + \mathbf{E}_{ij})(\mathbf{I}_n - \mathbf{E}_{ij}) = \mathbf{I}_n$ and $\mathbf{I}_n + \mathbf{E}_{ij}$ is invertible. If $\mathbf{A}$ commutes with every invertible matrix, it commutes with $\mathbf{I}_n + \mathbf{E}_{ij}$, hence with $\mathbf{E}_{ij}$ (subtract $\mathbf{A}\mathbf{I}_n = \mathbf{I}_n\mathbf{A}$), for all $i \neq j$. By (a), $\mathbf{A} = c\mathbf{I}_n$.