

13.4 Two Forms at Once

Chapter 12 asked when several operators can be diagonalized in one orthonormal basis, and answered: when they commute and each is normal (Theorem 12.7.1). That theorem is about **operators**, and its currency is the unitary similarity U^*TU . This section asks a question that looks similar and is not. Given two **forms** — two self-adjoint matrices, each thought of as the recipe $\mathbf{x} \mapsto \mathbf{x}^* \mathbf{A} \mathbf{x}$ — can one change of variable make both of them diagonal at once? No commuting is assumed, and in general the two matrices do not commute. The price is that the change of variable is no longer unitary, and the hypothesis is that one of the two forms is positive definite.

Throughout, $F = \mathbb{R}$ or $F = \mathbb{C}$, and “self-adjoint” means $\mathbf{A}^* = \mathbf{A}$ (over $\mathbb{R}$: symmetric).

13.4.1 Similarity is for operators, congruence is for forms

A matrix can play two roles, and the two roles come with different change-of-variable rules.

As an **operator**, $\mathbf{A}$ eats a vector and returns a vector. Rewriting $\mathbf{x} = \mathbf{S} \mathbf{y}$ in the equation $\mathbf{z} = \mathbf{A} \mathbf{x}$ turns it into $\mathbf{S}^{-1} \mathbf{z} = (\mathbf{S}^{-1} \mathbf{A} \mathbf{S}) \mathbf{y}$: the new matrix is $\mathbf{S}^{-1} \mathbf{A} \mathbf{S}$, and the relation is **similarity**. Eigenvalues, trace, determinant, and characteristic polynomial all survive it.

As a **form**, $\mathbf{A}$ eats a vector and returns the scalar $\mathbf{x}^* \mathbf{A} \mathbf{x}$. Rewriting $\mathbf{x} = \mathbf{S} \mathbf{y}$ turns that scalar into

$$(\mathbf{S} \mathbf{y})^* \mathbf{A} (\mathbf{S} \mathbf{y}) = \mathbf{y}^* (\mathbf{S}^* \mathbf{A} \mathbf{S}) \mathbf{y},$$

so the new matrix is $\mathbf{S}^* \mathbf{A} \mathbf{S}$. Two matrices $\mathbf{A}$ and $\mathbf{S}^* \mathbf{A} \mathbf{S}$ with $\mathbf{S}$ invertible are called **congruent**; this is the operation of Proposition 13.1.2, where it was already shown to preserve the sign of a form. Congruence is the right relation here for the plain reason that it is the one the question is phrased in: “make both forms diagonal” means “make both $\mathbf{x}^* \mathbf{A} \mathbf{x}$ and $\mathbf{x}^* \mathbf{B} \mathbf{x}$ into sums of squares in the same new coordinates”, and that is a statement about $\mathbf{S}^* \mathbf{A} \mathbf{S}$ and $\mathbf{S}^* \mathbf{B} \mathbf{S}$. Chapter 14 studies congruence in its own right, including what it does preserve; for now, note that it certainly does **not** preserve eigenvalues, since $\mathbf{S}^* \mathbf{I} \mathbf{S} = \mathbf{S}^* \mathbf{S}$ can be any positive definite matrix at all.

Congruence and similarity coincide in exactly one situation: when $\mathbf{S}$ is unitary, $\mathbf{S}^* = \mathbf{S}^{-1}$. That is the situation of Chapter 12, and it is what makes that chapter’s theorems so strong and its hypotheses so restrictive.

Warning. This is not Chapter 12’s simultaneous diagonalization. Theorem 12.7.1 needs the operators to **commute** and produces a **unitary** $\mathbf{U}$ with every $\mathbf{U}^* \mathbf{T} \mathbf{U}$ diagonal; eigenvalues are preserved. The theorem below needs **no commuting** and one of the two matrices to be **positive definite**, and produces a merely invertible $\mathbf{S}$; the eigenvalues of $\mathbf{B}$ are destroyed, and even $\mathbf{A}$ comes out as $\mathbf{I}$. The two theorems have no common generalization: neither hypothesis implies the other. When the matrices both commute and one is positive definite, both theorems apply and give different factorizations.

13.4.2 Diagonalizing two forms at once

Theorem 13.4.1. Diagonalizing Two Forms by Congruence

Let $\mathbf{A}, \mathbf{B} \in M_n(F)$ be self-adjoint with $\mathbf{A} \succ \mathbf{O}$. Then there is an invertible $\mathbf{S} \in M_n(F)$ with

$$\mathbf{S}^* \mathbf{A} \mathbf{S} = \mathbf{I}_n \quad \text{and} \quad \mathbf{S}^* \mathbf{B} \mathbf{S} = \mathbf{D}$$

for some **real** diagonal $\mathbf{D}$. Over $\mathbb{R}$, with $\mathbf{A}, \mathbf{B}$ symmetric, $\mathbf{S}$ may be taken real.

Idea. Two steps, and the first one is the whole idea. ① *Whiten.* The hypothesis $\mathbf{A} \succ \mathbf{O}$ gives an invertible positive definite square root $\mathbf{A}^{1/2}$, and the congruence by $\mathbf{A}^{-1/2}$ turns $\mathbf{A}$ into $\mathbf{I}$. That uses up the definiteness of $\mathbf{A}$ and nothing else. ② Now $\mathbf{I}$ is invariant under every **unitary** congruence, because a unitary congruence is a similarity; so we still have the whole spectral theorem available for the transformed $\mathbf{B}$, and it is self-adjoint. Diagonalize it and compose the two changes of variable.

Proof. Since $\mathbf{A} \succ \mathbf{O}$, Theorem 13.2.1 gives a unique $\mathbf{A}^{1/2} \succeq \mathbf{O}$ with $(\mathbf{A}^{1/2})^2 = \mathbf{A}$. Its eigenvalues are the nonnegative square roots of the eigenvalues of $\mathbf{A}$, which are > 0 by Theorem 13.1.5; hence $\mathbf{A}^{1/2} \succ \mathbf{O}$ and in particular $\mathbf{A}^{1/2}$ is invertible. Write $\mathbf{A}^{-1/2} = (\mathbf{A}^{1/2})^{-1}$, which is again self-adjoint, since the inverse of an invertible self-adjoint matrix is self-adjoint.

Put $\mathbf{C} = \mathbf{A}^{-1/2} \mathbf{B} \mathbf{A}^{-1/2}$. Then $\mathbf{C}^* = \mathbf{A}^{-1/2} \mathbf{B}^* \mathbf{A}^{-1/2} = \mathbf{C}$, so $\mathbf{C}$ is self-adjoint. By the spectral theorem for matrices there is a unitary $\mathbf{U}$ and a real diagonal $\mathbf{D}$ with $\mathbf{C} = \mathbf{U} \mathbf{D} \mathbf{U}^*$ (Corollary 12.4.3 over $\mathbb{C}$; Corollary 12.5.2 over $\mathbb{R}$, with $\mathbf{U}$ orthogonal and real).

Set $\mathbf{S} = \mathbf{A}^{-1/2} \mathbf{U}$, a product of invertible matrices, hence invertible. Then

$$\mathbf{S}^* \mathbf{A} \mathbf{S} = \mathbf{U}^* \mathbf{A}^{-1/2} \mathbf{A} \mathbf{A}^{-1/2} \mathbf{U} = \mathbf{U}^* \mathbf{U} = \mathbf{I}_n,$$

where the middle equality uses $\mathbf{A}^{-1/2} \mathbf{A} \mathbf{A}^{-1/2} = \mathbf{A}^{-1/2} \mathbf{A}^{1/2} \mathbf{A}^{1/2} \mathbf{A}^{-1/2} = \mathbf{I}_n$. And

$$\mathbf{S}^* \mathbf{B} \mathbf{S} = \mathbf{U}^* \mathbf{A}^{-1/2} \mathbf{B} \mathbf{A}^{-1/2} \mathbf{U} = \mathbf{U}^* \mathbf{C} \mathbf{U} = \mathbf{D}.$$

Over $\mathbb{R}$ every matrix named here is real: $\mathbf{A}^{1/2}$ is real because it is a polynomial in $\mathbf{A}$ (Theorem 13.2.1), and $\mathbf{U}$ is real by Corollary 12.5.2. This proves the theorem. ■

The theorem says more than “both become diagonal”: the definite one becomes the **identity**, which is the strongest normalization available, and it costs nothing extra. Note also which hypothesis did which job. Definiteness of $\mathbf{A}$ was used exactly once, to invert $\mathbf{A}^{1/2}$. Self-adjointness of $\mathbf{B}$ was used exactly once, to apply the spectral theorem to $\mathbf{C}$. Nothing at all was assumed about $\mathbf{AB}$ versus $\mathbf{BA}$.

Quick check. Why can $\mathbf{S}$ not be taken unitary in Theorem 13.4.1?

Solution. A unitary $\mathbf{S}$ has $\mathbf{S}^* = \mathbf{S}^{-1}$, so $\mathbf{S}^* \mathbf{A} \mathbf{S} = \mathbf{I}$ would give $\mathbf{A} = \mathbf{S} \mathbf{I} \mathbf{S}^* = \mathbf{I}$. So a unitary $\mathbf{S}$ can only work when $\mathbf{A}$ is already the identity — and then the theorem is just the spectral

theorem for $\mathbf{B}$.

13.4.3 The generalized eigenvalue problem

The matrix $\mathbf{S}$ of the theorem is not an abstract object: its columns solve an eigenvalue problem in which the identity has been replaced by $\mathbf{A}$.

Definition 13.4.1. Generalized eigenvalue

Let $\mathbf{A}, \mathbf{B} \in M_n(F)$. A scalar $\lambda \in F$ is a **generalized eigenvalue** of the pair $(\mathbf{A}, \mathbf{B})$ if

$$\mathbf{B}\mathbf{x} = \lambda\mathbf{A}\mathbf{x} \quad \text{for some } \mathbf{x} \neq \mathbf{0},$$

and such an $\mathbf{x}$ is a **generalized eigenvector** for λ . The family of matrices $\mathbf{B} - \lambda\mathbf{A}$ is called the **pencil** of the pair.

Taking $\mathbf{A} = \mathbf{I}$ recovers Definition 9.2.1, so the word “generalized” is honest. The condition $\mathbf{B}\mathbf{x} = \lambda\mathbf{A}\mathbf{x}$ with $\mathbf{x} \neq \mathbf{0}$ says that $\mathbf{B} - \lambda\mathbf{A}$ is singular, that is $\det(\mathbf{B} - \lambda\mathbf{A}) = 0$; so the generalized eigenvalues are the roots of the polynomial $\det(\mathbf{B} - \lambda\mathbf{A})$ in λ , the **characteristic polynomial of the pencil**. Its degree is at most n , and exactly n when $\mathbf{A}$ is invertible, since the λ^n -coefficient is $(-1)^n \det \mathbf{A}$. When $\mathbf{A}$ is singular the degree can drop, and the polynomial can even be a nonzero constant, in which case the pair has no generalized eigenvalues at all; exercise C3 below is such a pair. From here on $\mathbf{A} \succ \mathbf{O}$, so this does not arise.

Nothing so far forces those roots to be real: for $\mathbf{A} = \mathbf{I}$ they are just eigenvalues, and a general $\mathbf{B}$ has complex ones. Self-adjointness of $\mathbf{B}$ alone is not enough either, as the warning at the end of this section shows. It is definiteness of $\mathbf{A}$ that makes the problem behave.

Theorem 13.4.2. The Definite Generalized Eigenvalue Problem

Let $\mathbf{A}, \mathbf{B} \in M_n(F)$ be self-adjoint with $\mathbf{A} \succ \mathbf{O}$, and let $\mathbf{S}$ and $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$ be as in Theorem 13.4.1, with columns $\mathbf{s}_1, \dots, \mathbf{s}_n$ of $\mathbf{S}$. Then:

- (a) $\mathbf{B}\mathbf{s}_i = d_i\mathbf{A}\mathbf{s}_i$ for every i , and $\mathbf{s}_i^*\mathbf{A}\mathbf{s}_j = \delta_{ij}$: the columns of $\mathbf{S}$ are a basis of F^n of generalized eigenvectors, orthonormal for the inner product $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{A}} = \mathbf{y}^*\mathbf{A}\mathbf{x}$;
- (b) the generalized eigenvalues of $(\mathbf{A}, \mathbf{B})$ are exactly $d_1, \dots, d_n$, all **real**, and

$$\det(\mathbf{B} - \lambda\mathbf{A}) = \frac{1}{|\det \mathbf{S}|^2} \prod_{i=1}^n (d_i - \lambda);$$

- (c) $\mathbf{A}^{-1}\mathbf{B} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1}$; in particular $\mathbf{A}^{-1}\mathbf{B}$ is diagonalizable with real eigenvalues $d_1, \dots, d_n$, although it need not be self-adjoint.

Idea. Everything is already in the two equations $\mathbf{S}^*\mathbf{A}\mathbf{S} = \mathbf{I}$ and $\mathbf{S}^*\mathbf{B}\mathbf{S} = \mathbf{D}$. Substituting the first into the second turns $\mathbf{D}$ into $\mathbf{S}^*\mathbf{A}\mathbf{S}\mathbf{D}$, and canceling the invertible $\mathbf{S}^*$ leaves $\mathbf{B}\mathbf{S} = \mathbf{A}\mathbf{S}\mathbf{D}$, which read column by column is (a). Then (b) and (c) are two ways of undoing the substitution.

Proof.

- (a) From $\mathbf{S}^* \mathbf{A} \mathbf{S} = \mathbf{I}$ we get $\mathbf{D} = \mathbf{I} \mathbf{D} = \mathbf{S}^* \mathbf{A} \mathbf{S} \mathbf{D}$, and also $\mathbf{D} = \mathbf{S}^* \mathbf{B} \mathbf{S}$. Therefore $\mathbf{S}^* (\mathbf{B} \mathbf{S} - \mathbf{A} \mathbf{S} \mathbf{D}) = \mathbf{O}$, and $\mathbf{S}^*$ is invertible, so

$$\mathbf{B} \mathbf{S} = \mathbf{A} \mathbf{S} \mathbf{D}.$$

Column i of the left side is $\mathbf{B} \mathbf{s}_i$ and of the right side is $d_i \mathbf{A} \mathbf{s}_i$, since $\mathbf{D}$ is diagonal. Each $\mathbf{s}_i \neq \mathbf{0}$ because $\mathbf{S}$ is invertible, so each d_i is a generalized eigenvalue with eigenvector $\mathbf{s}_i$. The entry (i, j) of $\mathbf{S}^* \mathbf{A} \mathbf{S} = \mathbf{I}$ is $\mathbf{s}_i^* \mathbf{A} \mathbf{s}_j$, which is therefore δ_{ij} . Finally $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{A}} = \mathbf{y}^* \mathbf{A} \mathbf{x}$ is an inner product on F^n , exactly because $\mathbf{A} \succ \mathbf{O}$: it is sesquilinear, $\langle \mathbf{y}, \mathbf{x} \rangle_{\mathbf{A}} = \mathbf{x}^* \mathbf{A} \mathbf{y} = \mathbf{y}^* \mathbf{A}^* \mathbf{x} = \langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{A}}$, and $\langle \mathbf{x}, \mathbf{x} \rangle_{\mathbf{A}} = \mathbf{x}^* \mathbf{A} \mathbf{x} > 0$ for $\mathbf{x} \neq \mathbf{0}$.

- (b) From $\mathbf{S}^* \mathbf{A} \mathbf{S} = \mathbf{I}$ and $\mathbf{S}^* \mathbf{B} \mathbf{S} = \mathbf{D}$ we get $\mathbf{A} = (\mathbf{S}^*)^{-1} \mathbf{S}^{-1}$ and $\mathbf{B} = (\mathbf{S}^*)^{-1} \mathbf{D} \mathbf{S}^{-1}$, hence

$$\mathbf{B} - \lambda \mathbf{A} = (\mathbf{S}^*)^{-1} (\mathbf{D} - \lambda \mathbf{I}) \mathbf{S}^{-1}.$$

Taking determinants with Theorem 7.5.4, and using $\det(\mathbf{S}^*) = \overline{\det \mathbf{S}}$,

$$\det(\mathbf{B} - \lambda \mathbf{A}) = \frac{\det(\mathbf{D} - \lambda \mathbf{I})}{\det \mathbf{S}^* \det \mathbf{S}} = \frac{1}{|\det \mathbf{S}|^2} \prod_i (d_i - \lambda),$$

the last step by Theorem 7.4.6. Since $|\det \mathbf{S}|^2 \neq 0$, the roots of $\det(\mathbf{B} - \lambda \mathbf{A})$ are exactly $d_1, \dots, d_n$; and λ is a generalized eigenvalue if and only if $\mathbf{B} - \lambda \mathbf{A}$ is singular, that is if and only if $\det(\mathbf{B} - \lambda \mathbf{A}) = 0$ (Theorem 3.4.4). The d_i are real because $\mathbf{D} = \mathbf{S}^* \mathbf{B} \mathbf{S}$ is self-adjoint, so its diagonal entries equal their own conjugates.

- (c) Inverting $\mathbf{A} = (\mathbf{S}^*)^{-1} \mathbf{S}^{-1}$ gives $\mathbf{A}^{-1} = \mathbf{S} \mathbf{S}^*$, so

$$\mathbf{A}^{-1} \mathbf{B} = \mathbf{S} \mathbf{S}^* (\mathbf{S}^*)^{-1} \mathbf{D} \mathbf{S}^{-1} = \mathbf{S} \mathbf{D} \mathbf{S}^{-1}.$$

This proves the theorem. ■

Part (c) explains the phenomenon. The matrix $\mathbf{A}^{-1} \mathbf{B}$ is what one would naively form to turn $\mathbf{B} \mathbf{x} = \lambda \mathbf{A} \mathbf{x}$ into an ordinary eigenvalue problem, and it is usually not self-adjoint, so there is no reason from Chapter 12 for its eigenvalues to be real. But it **is** self-adjoint with respect to the inner product $\langle \cdot, \cdot \rangle_{\mathbf{A}}$, and reality and the orthogonal eigenbasis come from that inner product rather than the standard one. The congruence $\mathbf{S}$ is exactly the change of variable that turns $\langle \cdot, \cdot \rangle_{\mathbf{A}}$ into the standard inner product.

Example. A pencil that does not commute

Let

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 & 3 \\ 3 & 6 \end{pmatrix}.$$

Check that $\mathbf{A} \succ \mathbf{O}$, that $\mathbf{A} \mathbf{B} \neq \mathbf{B} \mathbf{A}$, and that $\mathbf{B}$ is indefinite. Find the generalized eigenvalues and an $\mathbf{A}$ -orthonormal basis of generalized eigenvectors, and write down $\mathbf{S}$.

Solution. Setting up. $\mathbf{A}$ is symmetric with leading principal minors $2 > 0$ and $4 - 1 = 3 > 0$,

so $\mathbf{A} \succ \mathbf{O}$ by Theorem 13.1.5. Its own eigenvalues are 1 and 3. The products are

$$\mathbf{AB} = \begin{pmatrix} 3 & 12 \\ 6 & 15 \end{pmatrix}, \quad \mathbf{BA} = \begin{pmatrix} 3 & 6 \\ 12 & 15 \end{pmatrix},$$

which differ, so Chapter 12's theorem does not apply. Also $\det \mathbf{B} = -9 < 0$, so $\mathbf{B}$ has one positive and one negative eigenvalue and is neither positive nor negative semidefinite.

The pencil.

$$\mathbf{B} - \lambda \mathbf{A} = \begin{pmatrix} -2\lambda & 3 - \lambda \\ 3 - \lambda & 6 - 2\lambda \end{pmatrix},$$

so

$$\begin{aligned} \det(\mathbf{B} - \lambda \mathbf{A}) &= -2\lambda(6 - 2\lambda) - (3 - \lambda)^2 \\ &= 4\lambda^2 - 12\lambda - 9 + 6\lambda - \lambda^2 \\ &= 3\lambda^2 - 6\lambda - 9 = 3(\lambda - 3)(\lambda + 1). \end{aligned}$$

The generalized eigenvalues are $\lambda = -1$ and $\lambda = 3$, both real, as Theorem 13.4.2 (b) promises.

Eigenvectors. For $\lambda = -1$, $\mathbf{B} + \mathbf{A} = \begin{pmatrix} 2 & 4 \\ 4 & 8 \end{pmatrix}$, whose kernel is spanned by $\mathbf{x}_1 = (2, -1)$. For

$\lambda = 3$, $\mathbf{B} - 3\mathbf{A} = \begin{pmatrix} -6 & 0 \\ 0 & 0 \end{pmatrix}$, whose kernel is spanned by $\mathbf{x}_2 = (0, 1)$. Checking directly: $\mathbf{B}\mathbf{x}_1 = (-3, 0)$ and $\mathbf{A}\mathbf{x}_1 = (3, 0)$, so $\mathbf{B}\mathbf{x}_1 = -\mathbf{A}\mathbf{x}_1$; and $\mathbf{B}\mathbf{x}_2 = (3, 6) = 3(1, 2) = 3\mathbf{A}\mathbf{x}_2$.

Normalizing. $\mathbf{x}_1^\top \mathbf{A}\mathbf{x}_1 = (2, -1) \cdot (3, 0) = 6$ and $\mathbf{x}_2^\top \mathbf{A}\mathbf{x}_2 = (0, 1) \cdot (1, 2) = 2$, while $\mathbf{x}_1^\top \mathbf{A}\mathbf{x}_2 = (2, -1) \cdot (1, 2) = 0$: the two are already $\mathbf{A}$ -orthogonal, as they must be, having different generalized eigenvalues. Scaling to $\mathbf{A}$ -length 1,

$$\mathbf{S} = \begin{pmatrix} 2/\sqrt{6} & 0 \\ -1/\sqrt{6} & 1/\sqrt{2} \end{pmatrix}, \quad \mathbf{S}^\top \mathbf{A} \mathbf{S} = \mathbf{I}_2, \quad \mathbf{S}^\top \mathbf{B} \mathbf{S} = \begin{pmatrix} -1 & 0 \\ 0 & 3 \end{pmatrix}.$$

So in the coordinates $\mathbf{x} = \mathbf{S}\mathbf{y}$ the two forms read $\mathbf{x}^\top \mathbf{A} \mathbf{x} = y_1^2 + y_2^2$ and $\mathbf{x}^\top \mathbf{B} \mathbf{x} = -y_1^2 + 3y_2^2$. The signs $(-, +)$ match the signs of the eigenvalues of $\mathbf{B}$ itself, which are $3 \pm 3\sqrt{2}$; Chapter 14 shows that this is no coincidence.

The problem $\mathbf{B}\mathbf{x} = \lambda \mathbf{A}\mathbf{x}$ with $\mathbf{A} \succ \mathbf{O}$ is the one an engineer meets first: for a vibrating structure, $\mathbf{A}$ is the mass matrix and $\mathbf{B}$ the stiffness matrix, the generalized eigenvalues are the squared natural frequencies, and the $\mathbf{A}$ -orthonormal eigenvectors are the normal modes, in which the motion decouples into independent one-dimensional oscillations.

13.4.4 When neither form is definite

Definiteness of one of the two matrices is not decoration.

Warning. Two self-adjoint matrices, neither definite, need not be simultaneously diagonalizable by congruence. Take the real symmetric pair

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

Suppose $\mathbf{S}^\top \mathbf{A} \mathbf{S}$ and $\mathbf{S}^\top \mathbf{B} \mathbf{S}$ were both diagonal for some invertible real $\mathbf{S}$, say $\mathbf{D}_\mathbf{A}$ and $\mathbf{D}_\mathbf{B}$.

Since $\mathbf{A}$ is invertible, so is $\mathbf{D}_\mathbf{A}$, and from $\mathbf{A} = (\mathbf{S}^\top)^{-1}\mathbf{D}_\mathbf{A}\mathbf{S}^{-1}$ and $\mathbf{B} = (\mathbf{S}^\top)^{-1}\mathbf{D}_\mathbf{B}\mathbf{S}^{-1}$ we get

$$\mathbf{A}^{-1}\mathbf{B} = \mathbf{S}\mathbf{D}_\mathbf{A}^{-1}\mathbf{S}^\top(\mathbf{S}^\top)^{-1}\mathbf{D}_\mathbf{B}\mathbf{S}^{-1} = \mathbf{S}(\mathbf{D}_\mathbf{A}^{-1}\mathbf{D}_\mathbf{B})\mathbf{S}^{-1},$$

so $\mathbf{A}^{-1}\mathbf{B}$ would be similar to a real diagonal matrix and would have real eigenvalues. But $\mathbf{A}^{-1} = \mathbf{A}$ here, and

$$\mathbf{A}^{-1}\mathbf{B} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

has characteristic polynomial $\lambda^2 + 1$, with roots $\pm i$. So no such $\mathbf{S}$ exists — and the pencil has **no** generalized eigenvalues in $\mathbb{R}$ at all.

The obstruction in the warning is exactly the failure of part (c) of [Theorem 13.4.2](#). That part is in fact the right way to test a pair: if $\mathbf{A}$ is invertible, a necessary condition for simultaneous diagonalization by congruence is that $\mathbf{A}^{-1}\mathbf{B}$ be diagonalizable with real eigenvalues, and the theorem says that $\mathbf{A} \succ \mathbf{O}$ is enough to guarantee it.

13.4.5 Exercises

13.4.5.1 A. Check your understanding

Exercise. A1

- Define what it means for two matrices to be **congruent**, and say which of similarity and congruence is the natural relation for a form.
- State [Theorem 13.4.1](#), with all hypotheses.
- Name two differences between [Theorem 13.4.1](#) and Theorem 12.7.1.
- Determine whether the following statement is correct, and justify your answer: the generalized eigenvalues of $(\mathbf{A}, \mathbf{B})$ are eigenvalues of $\mathbf{B}$.
- For $\mathbf{A} \succ \mathbf{O}$ and $\mathbf{B}$ self-adjoint, what inner product makes $\mathbf{A}^{-1}\mathbf{B}$ self-adjoint?

Solution.

- $\mathbf{A}$ and $\mathbf{C}$ are congruent if $\mathbf{C} = \mathbf{S}^*\mathbf{A}\mathbf{S}$ for some invertible $\mathbf{S}$. Congruence is the natural relation for a form, since the substitution $\mathbf{x} = \mathbf{S}\mathbf{y}$ turns $\mathbf{x}^*\mathbf{A}\mathbf{x}$ into $\mathbf{y}^*(\mathbf{S}^*\mathbf{A}\mathbf{S})\mathbf{y}$; similarity $\mathbf{S}^{-1}\mathbf{A}\mathbf{S}$ is the natural relation for an operator.
- If $\mathbf{A}, \mathbf{B} \in M_n(F)$ are self-adjoint and $\mathbf{A} \succ \mathbf{O}$, there is an invertible $\mathbf{S}$ with $\mathbf{S}^*\mathbf{A}\mathbf{S} = \mathbf{I}_n$ and $\mathbf{S}^*\mathbf{B}\mathbf{S}$ real diagonal.
- Chapter 12's theorem assumes the operators commute and are normal, and delivers a **unitary** change of basis preserving eigenvalues; this one assumes nothing about commuting but assumes one matrix is positive definite, and delivers a merely invertible $\mathbf{S}$ under which eigenvalues are not preserved.
- Incorrect. In [Example \(A pencil that does not commute\)](#) the generalized eigenvalues are -1 and 3 , while the eigenvalues of $\mathbf{B}$ are $3 \pm 3\sqrt{2}$. They agree when $\mathbf{A} = \mathbf{I}$, and not in general.
- $\langle \mathbf{x}, \mathbf{y} \rangle_\mathbf{A} = \mathbf{y}^*\mathbf{A}\mathbf{x}$ ([Theorem 13.4.2](#) (a)).

13.4.5.2 B. Practice

Exercise. B1: A generalized eigenvalue problem

Let

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

- (a) Verify that $\mathbf{A} \succ \mathbf{O}$ and find the generalized eigenvalues of $(\mathbf{A}, \mathbf{B})$.
 (b) Find generalized eigenvectors and an invertible $\mathbf{S}$ with $\mathbf{S}^\top \mathbf{A} \mathbf{S} = \mathbf{I}_2$ and $\mathbf{S}^\top \mathbf{B} \mathbf{S}$ diagonal.

Solution.

- (a) The leading principal minors of $\mathbf{A}$ are $1 > 0$ and $2 - 1 = 1 > 0$, so $\mathbf{A} \succ \mathbf{O}$ by Theorem 13.1.5. Then

$$\mathbf{B} - \lambda \mathbf{A} = \begin{pmatrix} 2 - \lambda & 2 - \lambda \\ 2 - \lambda & 1 - 2\lambda \end{pmatrix},$$

so

$$\begin{aligned} \det(\mathbf{B} - \lambda \mathbf{A}) &= (2 - \lambda)(1 - 2\lambda) - (2 - \lambda)^2 \\ &= (2 - \lambda)[(1 - 2\lambda) - (2 - \lambda)] \\ &= (2 - \lambda)(-1 - \lambda) = (\lambda - 2)(\lambda + 1). \end{aligned}$$

 The generalized eigenvalues are 2 and -1 .

- (b) For $\lambda = 2$: $\mathbf{B} - 2\mathbf{A} = \begin{pmatrix} 0 & 0 \\ 0 & -3 \end{pmatrix}$, with kernel spanned by $\mathbf{x}_1 = (1, 0)$. For $\lambda = -1$:

$\mathbf{B} + \mathbf{A} = \begin{pmatrix} 3 & 3 \\ 3 & 3 \end{pmatrix}$, with kernel spanned by $\mathbf{x}_2 = (1, -1)$. Check: $\mathbf{B}\mathbf{x}_1 = (2, 2) = 2\mathbf{A}\mathbf{x}_1$ since $\mathbf{A}\mathbf{x}_1 = (1, 1)$; and $\mathbf{B}\mathbf{x}_2 = (0, 1)$ while $\mathbf{A}\mathbf{x}_2 = (0, -1)$, so $\mathbf{B}\mathbf{x}_2 = -\mathbf{A}\mathbf{x}_2$.

Now $\mathbf{x}_1^\top \mathbf{A} \mathbf{x}_1 = (1, 0) \cdot (1, 1) = 1$ and $\mathbf{x}_2^\top \mathbf{A} \mathbf{x}_2 = (1, -1) \cdot (0, -1) = 1$, and $\mathbf{x}_1^\top \mathbf{A} \mathbf{x}_2 = (1, 0) \cdot (0, -1) = 0$. Both are already $\mathbf{A}$ -unit vectors, so

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

Exercise. B2: Whitening by hand

Let $\mathbf{A} = \text{diag}(4, 9)$ and $\mathbf{B} = \begin{pmatrix} 4 & 6 \\ 6 & 9 \end{pmatrix}$. Follow the two steps of the proof of Theorem 13.4.1: compute $\mathbf{A}^{-1/2}$, then $\mathbf{C} = \mathbf{A}^{-1/2} \mathbf{B} \mathbf{A}^{-1/2}$, diagonalize $\mathbf{C}$, and write down $\mathbf{S}$ and $\mathbf{S}^\top \mathbf{B} \mathbf{S}$.

Solution. $\mathbf{A}$ is diagonal with positive entries, so $\mathbf{A}^{1/2} = \text{diag}(2, 3)$ and $\mathbf{A}^{-1/2} = \text{diag}(\frac{1}{2}, \frac{1}{3})$. Then

$$\mathbf{C} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 4 & 6 \\ 6 & 9 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Its eigenvalues are 2, with unit eigenvector $\frac{1}{\sqrt{2}}(1, 1)$, and 0, with unit eigenvector $\frac{1}{\sqrt{2}}(1, -1)$.

So $\mathbf{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ and $\mathbf{D} = \text{diag}(2, 0)$, and

$$\mathbf{S} = \mathbf{A}^{-1/2} \mathbf{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & -\frac{1}{3} \end{pmatrix}, \quad \mathbf{S}^\top \mathbf{A} \mathbf{S} = \mathbf{I}_2, \quad \mathbf{S}^\top \mathbf{B} \mathbf{S} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}.$$

As a check, $\det(\mathbf{B} - \lambda \mathbf{A}) = (4 - 4\lambda)(9 - 9\lambda) - 36 = 36\lambda^2 - 72\lambda = 36\lambda(\lambda - 2)$, whose roots 0 and 2 are the diagonal entries of $\mathbf{S}^\top \mathbf{B} \mathbf{S}$.

Exercise. B3: A pair that resists

Let $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. Determine whether there is an invertible real $\mathbf{S}$ making $\mathbf{S}^\top \mathbf{A} \mathbf{S}$ and $\mathbf{S}^\top \mathbf{B} \mathbf{S}$ both diagonal. Justify your answer.

Solution. There is not. Both matrices are symmetric and $\mathbf{A}$ is invertible, with $\mathbf{A}^{-1} = \mathbf{A}$. If such an $\mathbf{S}$ existed, the computation in the warning above would give $\mathbf{A}^{-1} \mathbf{B} = \mathbf{S}(\mathbf{D}_{\mathbf{A}}^{-1} \mathbf{D}_{\mathbf{B}}) \mathbf{S}^{-1}$, so $\mathbf{A}^{-1} \mathbf{B}$ would be diagonalizable. But

$$\mathbf{A}^{-1} \mathbf{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

which is nonzero with $(\mathbf{A}^{-1} \mathbf{B})^2 = \mathbf{O}$, hence has 0 as its only eigenvalue and is not diagonalizable (a diagonalizable matrix with only the eigenvalue 0 is $\mathbf{O}$). Here the generalized eigenvalues are both 0 and real, so reality of the spectrum is not the obstruction; diagonalizability is.

13.4.5.3 C. Going deeper

Exercise. C1: The $\mathbf{A}$ -inner product

Let $\mathbf{A} \succ \mathbf{O}$ and $\mathbf{B}$ be self-adjoint in $M_n(F)$, and write $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{A}} = \mathbf{y}^* \mathbf{A} \mathbf{x}$.

- Prove that $\langle \cdot, \cdot \rangle_{\mathbf{A}}$ is an inner product on F^n .
- Prove that $T = T_{\mathbf{A}^{-1} \mathbf{B}}$ is self-adjoint on $(F^n, \langle \cdot, \cdot \rangle_{\mathbf{A}})$.
- Deduce Theorem 13.4.2 (a) and (b) from the spectral theorem applied to T , without using Theorem 13.4.1.

Solution.

- Sesquilinearity in the first slot is clear from the formula. Conjugate symmetry: $\overline{\langle \mathbf{y}, \mathbf{x} \rangle_{\mathbf{A}}} = \mathbf{x}^* \mathbf{A} \mathbf{y} = \mathbf{y}^* \mathbf{A}^* \mathbf{x} = \mathbf{y}^* \mathbf{A} \mathbf{x} = \langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{A}}$, using $\mathbf{A}^* = \mathbf{A}$. Positive definiteness: $\langle \mathbf{x}, \mathbf{x} \rangle_{\mathbf{A}} = \mathbf{x}^* \mathbf{A} \mathbf{x} > 0$ for $\mathbf{x} \neq \mathbf{0}$, which is Definition 13.1.1 for $\mathbf{A} \succ \mathbf{O}$.
- For all $\mathbf{x}, \mathbf{y}$,

$$\langle T \mathbf{x}, \mathbf{y} \rangle_{\mathbf{A}} = \mathbf{y}^* \mathbf{A} \mathbf{A}^{-1} \mathbf{B} \mathbf{x} = \mathbf{y}^* \mathbf{B} \mathbf{x},$$

and

$$\langle \mathbf{x}, T\mathbf{y} \rangle_{\mathbf{A}} = (\mathbf{A}^{-1}\mathbf{B}\mathbf{y})^* \mathbf{A}\mathbf{x} = \mathbf{y}^* \mathbf{B}^* \mathbf{A}^{-1} \mathbf{A}\mathbf{x} = \mathbf{y}^* \mathbf{B}\mathbf{x},$$

using $(\mathbf{A}^{-1})^* = \mathbf{A}^{-1}$ and $\mathbf{B}^* = \mathbf{B}$. The two agree, so T is self-adjoint for this inner product (Definition 11.6.1).

- (c) By Theorem 12.5.1 (over $\mathbb{R}$) or Theorem 12.4.2 (over $\mathbb{C}$) applied to the self-adjoint T on the inner product space $(F^n, \langle \cdot, \cdot \rangle_{\mathbf{A}})$, there is a basis $\mathbf{s}_1, \dots, \mathbf{s}_n$ of F^n , orthonormal for $\langle \cdot, \cdot \rangle_{\mathbf{A}}$, with $T\mathbf{s}_i = d_i\mathbf{s}_i$ and $d_i \in \mathbb{R}$ (Theorem 12.2.1). The equation $\mathbf{A}^{-1}\mathbf{B}\mathbf{s}_i = d_i\mathbf{s}_i$ is $\mathbf{B}\mathbf{s}_i = d_i\mathbf{A}\mathbf{s}_i$, and $\mathbf{s}_i^* \mathbf{A}\mathbf{s}_j = \langle \mathbf{s}_j, \mathbf{s}_i \rangle_{\mathbf{A}} = \delta_{ij}$; this is (a). For (b), λ is a generalized eigenvalue exactly when $\mathbf{A}^{-1}\mathbf{B} - \lambda\mathbf{I}$ is singular, that is when λ is an eigenvalue of $\mathbf{A}^{-1}\mathbf{B}$, and those are $d_1, \dots, d_n$.

Exercise. C2: Signs of the generalized eigenvalues

Let $\mathbf{A} \succ \mathbf{O}$ and $\mathbf{B}$ be self-adjoint in $M_n(F)$, with generalized eigenvalues $d_1, \dots, d_n$.

- (a) Prove that $\mathbf{B} \succeq \mathbf{O}$ if and only if every $d_i \geq 0$, and $\mathbf{B} \succ \mathbf{O}$ if and only if every $d_i > 0$.
 (b) Deduce that if $\mathbf{A} \succ \mathbf{O}$ and $\mathbf{B} \succ \mathbf{O}$, then $\det \mathbf{B} / \det \mathbf{A} = \prod_i d_i > 0$.

Solution.

- (a) Let $\mathbf{S}$ be as in Theorem 13.4.1, so $\mathbf{S}^* \mathbf{B} \mathbf{S} = \mathbf{D} = \text{diag}(d_1, \dots, d_n)$. For $\mathbf{x} \in F^n$ put $\mathbf{y} = \mathbf{S}^{-1}\mathbf{x}$; as $\mathbf{x}$ runs over F^n so does $\mathbf{y}$, and $\mathbf{x} \neq \mathbf{0}$ exactly when $\mathbf{y} \neq \mathbf{0}$. Then

$$\mathbf{x}^* \mathbf{B} \mathbf{x} = \mathbf{y}^* \mathbf{S}^* \mathbf{B} \mathbf{S} \mathbf{y} = \mathbf{y}^* \mathbf{D} \mathbf{y} = \sum_i d_i |y_i|^2.$$

If every $d_i \geq 0$, this is ≥ 0 for all $\mathbf{y}$, so $\mathbf{B} \succeq \mathbf{O}$. Conversely, taking $\mathbf{y} = \mathbf{e}_i$ gives $d_i \geq 0$. The same argument with strict inequalities gives the second claim, using that $\sum_i d_i |y_i|^2 > 0$ for $\mathbf{y} \neq \mathbf{0}$ when every $d_i > 0$.

- (b) Taking determinants in $\mathbf{S}^* \mathbf{B} \mathbf{S} = \mathbf{D}$ and $\mathbf{S}^* \mathbf{A} \mathbf{S} = \mathbf{I}$ gives $|\det \mathbf{S}|^2 \det \mathbf{B} = \prod_i d_i$ and $|\det \mathbf{S}|^2 \det \mathbf{A} = 1$. Dividing, $\det \mathbf{B} / \det \mathbf{A} = \prod_i d_i$, which is > 0 by (a).

Exercise. C3: Dropping definiteness of $\mathbf{A}$

Let $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, so that $\mathbf{A} \succeq \mathbf{O}$ but $\mathbf{A} \not\succ \mathbf{O}$.

- (a) Compute $\det(\mathbf{B} - \lambda\mathbf{A})$ and deduce that the pair has no generalized eigenvalues at all.
 (b) Prove that no invertible $\mathbf{S} \in M_2(\mathbb{R})$ makes $\mathbf{S}^\top \mathbf{A} \mathbf{S}$ and $\mathbf{S}^\top \mathbf{B} \mathbf{S}$ both diagonal.

Hint for (b): compare the degree in λ of $\det(\mathbf{B} - \lambda\mathbf{A})$ on both sides.

Solution.

- (a) $\mathbf{B} - \lambda\mathbf{A} = \begin{pmatrix} -\lambda & 1 \\ 1 & 0 \end{pmatrix}$, with determinant -1 . This is never 0, so $\mathbf{B} - \lambda\mathbf{A}$ is invertible for every λ and there is no $\mathbf{x} \neq \mathbf{0}$ with $\mathbf{B}\mathbf{x} = \lambda\mathbf{A}\mathbf{x}$.

- (b) Suppose $\mathbf{S}^\top \mathbf{A} \mathbf{S} = \text{diag}(a_1, a_2)$ and $\mathbf{S}^\top \mathbf{B} \mathbf{S} = \text{diag}(b_1, b_2)$. As in the proof of [Theorem 13.4.2](#) (b),

$$\det(\mathbf{B} - \lambda \mathbf{A}) = \frac{1}{(\det \mathbf{S})^2} (b_1 - \lambda a_1)(b_2 - \lambda a_2).$$

By (a) the left side is the constant -1 , so the right side is a nonzero constant. It is nonzero, so $b_1 b_2 \neq 0$ and hence $\det(\mathbf{S}^\top \mathbf{B} \mathbf{S}) \neq 0$; it is constant in λ , so $a_1 = a_2 = 0$, which forces $\mathbf{S}^\top \mathbf{A} \mathbf{S} = \mathbf{O}$ and therefore $\mathbf{A} = \mathbf{O}$. That is false, so no such $\mathbf{S}$ exists. (Alternatively: $\text{rank}(\mathbf{S}^\top \mathbf{A} \mathbf{S}) = \text{rank} \mathbf{A} = 1$, so exactly one a_i is nonzero, and the right side then has degree 1 in λ , not 0.)