

## 13.2 Square Roots and the Cholesky Factorization

A nonnegative real number has exactly one nonnegative square root, and it is the reason “non-negative” is a useful class of numbers at all: it lets us take lengths, define standard deviations, and undo a squaring. Section 1 produced the matrix analogue as one clause of a theorem. This section makes it an object with a name, adds the one property the clause did not mention — that the square root is a **polynomial** in the matrix — and then turns the same idea into the triangular factorization that elimination discovers, the one Chapter 3 promised and could not yet explain.

Throughout,  $F = \mathbb{R}$  or  $F = \mathbb{C}$ , and all matrices are over  $F$ .

### 13.2.1 The positive square root

Theorem 13.1.4 (d) already says that a positive semidefinite  $\mathbf{A}$  has exactly one positive semidefinite square root. That is an existence-and-uniqueness statement, so by the usual convention it deserves a notation.

*The square root of a positive matrix is what you get by taking square roots of its eigenvalues and leaving the eigenvectors alone.*

#### Theorem 13.2.1. The Positive Square Root

Let  $\mathbf{A} \in M_n(F)$  with  $\mathbf{A} \succeq \mathbf{O}$ . Then there is exactly one  $\mathbf{C} \in M_n(F)$  with  $\mathbf{C} \succeq \mathbf{O}$  and  $\mathbf{C}^2 = \mathbf{A}$ . We write  $\mathbf{C} = \mathbf{A}^{1/2}$ , also  $\sqrt{\mathbf{A}}$ . Moreover:

- (a)  $\mathbf{A}^{1/2} = p(\mathbf{A})$  for some polynomial  $p$  with real coefficients (the nodes  $\lambda_i$  and the values  $\sqrt{\lambda_i}$  are real, so the interpolation of Theorem 6.6.1 may be run over  $\mathbb{R}$ , and  $\mathbb{R}[x] \subseteq \mathbb{C}[x]$ );
- (b)  $\mathbf{A}^{1/2}\mathbf{B} = \mathbf{B}\mathbf{A}^{1/2}$  for **every**  $\mathbf{B} \in M_n(F)$  with  $\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A}$ ;
- (c)  $\text{nul}(\mathbf{A}^{1/2}) = \text{nul}(\mathbf{A})$  and  $\text{rank } \mathbf{A}^{1/2} = \text{rank } \mathbf{A}$ ; in particular  $\mathbf{A}^{1/2}$  is invertible if and only if  $\mathbf{A}$  is.

The eigenvalues of  $\mathbf{A}^{1/2}$  are the nonnegative square roots of those of  $\mathbf{A}$ , with the same eigenvectors.

*Idea.* Existence and uniqueness are Section 1. What is new is (a), and it is the reason (b) is free. On the spectrum, the square root is just a function  $\lambda \mapsto \sqrt{\lambda}$  defined at finitely many points, and a function defined at finitely many points is a polynomial — that is Lagrange interpolation from Chapter 6. The functional calculus of Chapter 12 says that applying a polynomial to  $\mathbf{A}$  and applying the corresponding function to the spectrum are the same operation, so the interpolating polynomial evaluated at  $\mathbf{A}$  is the square root. Once  $\mathbf{A}^{1/2}$  is a polynomial in  $\mathbf{A}$ , anything commuting with  $\mathbf{A}$  commutes with it, with no further work.

*Proof.* Existence and uniqueness are Theorem 13.1.4 (d).

- (a) Let  $\lambda_1, \dots, \lambda_k$  be the **distinct** eigenvalues of  $\mathbf{A}$  and let  $\mathbf{A} = \sum_i \lambda_i \mathbf{P}_i$  be its spectral resolution (Definition 12.8.1, Theorem 12.8.1), which exists because a Hermitian matrix

has an orthonormal basis of eigenvectors. Every  $\lambda_i$  is a nonnegative real number by Theorem 13.1.4 (b), so the numbers  $\sqrt{\lambda_i}$  are defined and real. Let  $f(\lambda) = \sqrt{\lambda}$  on  $\text{spec}(\mathbf{A})$ . By Theorem 12.8.3 (b) and (a),

$$f(\mathbf{A})^2 = (f^2)(\mathbf{A}) = \mathbf{A},$$

and by (c) and (d),  $f(\mathbf{A})$  is Hermitian with spectrum  $\{\sqrt{\lambda_1}, \dots, \sqrt{\lambda_k}\} \subseteq [0, \infty)$ , so  $f(\mathbf{A}) \succeq \mathbf{O}$  by Theorem 13.1.4 (b). Uniqueness gives  $f(\mathbf{A}) = \mathbf{A}^{1/2}$ .

The numbers  $\lambda_1, \dots, \lambda_k$  are distinct, so by Lagrange interpolation (Theorem 6.6.1) there is a polynomial  $p$  of degree at most  $k-1$ , with real coefficients, such that  $p(\lambda_i) = \sqrt{\lambda_i}$  for every  $i$ . By Theorem 12.8.3 (a),  $p(\mathbf{A}) = \sum_i p(\lambda_i) \mathbf{P}_i = f(\mathbf{A}) = \mathbf{A}^{1/2}$ .

(b) Suppose  $\mathbf{AB} = \mathbf{BA}$ . Then  $\mathbf{B}$  commutes with every power of  $\mathbf{A}$ , hence with every polynomial in  $\mathbf{A}$ , hence with  $p(\mathbf{A}) = \mathbf{A}^{1/2}$ .

(c) If  $\mathbf{A}^{1/2}\mathbf{x} = \mathbf{0}$  then  $\mathbf{Ax} = \mathbf{A}^{1/2}(\mathbf{A}^{1/2}\mathbf{x}) = \mathbf{0}$ . Conversely, if  $\mathbf{Ax} = \mathbf{0}$ , then since  $\mathbf{A}^{1/2}$  is Hermitian,

$$\|\mathbf{A}^{1/2}\mathbf{x}\|^2 = \langle \mathbf{A}^{1/2}\mathbf{x}, \mathbf{A}^{1/2}\mathbf{x} \rangle = \langle \mathbf{Ax}, \mathbf{x} \rangle = 0,$$

so  $\mathbf{A}^{1/2}\mathbf{x} = \mathbf{0}$ . Hence the two null spaces agree, and the ranks agree by Theorem 3.5.7. Invertibility is the case  $\text{nul} = \{\mathbf{0}\}$  (Theorem 3.4.4).

The last sentence is the spectral mapping statement of Theorem 12.8.3 (d) for  $f = \sqrt{\cdot}$ . ■

Two routes to the same matrix, then, and both are worth practicing: assemble  $\sum_i \sqrt{\lambda_i} \mathbf{P}_i$  from the projections, or find one interpolating polynomial and substitute.

### Example. Two Square Roots

Compute  $\mathbf{A}^{1/2}$  for

$$(a) \quad \mathbf{A} = \begin{pmatrix} 13 & 12 \\ 12 & 13 \end{pmatrix};$$

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

in each case by the spectral resolution and by a polynomial.

*Solution.*

(a)  $\mathbf{A}$  is symmetric. Since  $\mathbf{A}(1, 1) = (25, 25)$  and  $\mathbf{A}(1, -1) = (1, -1)$ , the eigenvalues are 25 and 1, with eigenlines spanned by  $(1, 1)$  and  $(1, -1)$ . Both are positive, so  $\mathbf{A} \succ \mathbf{O}$ . The orthogonal projections onto the two eigenlines are

$$\mathbf{P}_1 = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{P}_2 = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$

so  $\mathbf{A} = 25\mathbf{P}_1 + \mathbf{P}_2$  and

$$\mathbf{A}^{1/2} = 5\mathbf{P}_1 + 1 \cdot \mathbf{P}_2 = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}.$$

By the polynomial route: we need  $p$  with  $p(25) = 5$  and  $p(1) = 1$ . The line through the two points is  $p(x) = \frac{1}{6}x + \frac{5}{6}$ , and

$$\frac{1}{6}\mathbf{A} + \frac{5}{6}\mathbf{I} = \frac{1}{6} \begin{pmatrix} 18 & 12 \\ 12 & 18 \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}.$$

Check:  $\begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}^2 = \begin{pmatrix} 13 & 12 \\ 12 & 13 \end{pmatrix}.$

- (b) Here  $\mathbf{B} = \mathbf{I} + \mathbf{J}$  with  $\mathbf{J}$  the all-ones matrix. Since  $\mathbf{J}\mathbf{1} = 3\mathbf{1}$  and  $\mathbf{J}\mathbf{x} = \mathbf{0}$  for every  $\mathbf{x}$  with entries summing to 0, the eigenvalues of  $\mathbf{J}$  are 3 and 0, so those of  $\mathbf{B}$  are 4 and 1; both are positive, so  $\mathbf{B} \succ \mathbf{O}$ . We need  $p$  with  $p(4) = 2$  and  $p(1) = 1$ , namely  $p(x) = \frac{1}{3}x + \frac{2}{3}$ , giving

$$\mathbf{B}^{1/2} = \frac{1}{3}\mathbf{B} + \frac{2}{3}\mathbf{I} = \frac{1}{3} \begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix}.$$

Check by squaring: the diagonal entry of the square is  $\frac{1}{9}(16 + 1 + 1) = 2$  and an off-diagonal entry is  $\frac{1}{9}(4 + 4 + 1) = 1$ , which reproduces  $\mathbf{B}$ . Only two distinct eigenvalues appear, so a polynomial of degree 1 suffices even though  $\mathbf{B}$  is  $3 \times 3$ .

**Warning.**  $\mathbf{A}^{1/2}$  is not the entrywise square root, and it is not the only square root. For  $\mathbf{A} = \begin{pmatrix} 13 & 12 \\ 12 & 13 \end{pmatrix}$  the entrywise roots  $\begin{pmatrix} \sqrt{13} & \sqrt{12} \\ \sqrt{12} & \sqrt{13} \end{pmatrix}$  form a matrix whose square is not  $\mathbf{A}$ .

And  $\mathbf{A}$  has other square roots:  $-\mathbf{A}^{1/2}$  is one, and  $5\mathbf{P}_1 - \mathbf{P}_2 = \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix}$  is another symmetric one. The word **positive** in “positive square root” is what makes the object unique; drop it and this  $\mathbf{A}$ , whose two eigenvalues are distinct and positive, has four symmetric square roots, one for each choice of sign on each eigenline. When an eigenvalue repeats there are infinitely many, symmetric ones included: every reflection  $\begin{pmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{pmatrix}$  is a symmetric square root of  $\mathbf{I}_2$ .

**Quick check.** Let  $\mathbf{A} \succeq \mathbf{O}$  and let  $\mathbf{U}$  be unitary. What is  $(\mathbf{U}^* \mathbf{A} \mathbf{U})^{1/2}$ ?

**Solution.**  $\mathbf{U}^* \mathbf{A}^{1/2} \mathbf{U}$ . It is positive semidefinite by Proposition 13.1.2 (a), since  $\mathbf{A}^{1/2} \succeq \mathbf{O}$ , and its square is  $\mathbf{U}^* \mathbf{A}^{1/2} \mathbf{U} \mathbf{U}^* \mathbf{A}^{1/2} \mathbf{U} = \mathbf{U}^* \mathbf{A} \mathbf{U}$  because  $\mathbf{U} \mathbf{U}^* = \mathbf{I}$ . Uniqueness in Theorem 13.2.1 does the rest. In words: the square root does not care which orthonormal coordinates we use, which is exactly what one wants of an operation defined on the spectrum.

## 13.2.2 The absolute value of a matrix

A complex number has a modulus  $|z| = \sqrt{\bar{z}z}$ , and the polar form  $z = e^{i\theta}|z|$  splits it into a rotation and a nonnegative stretch. Section 10 will split a matrix the same way, and the stretching factor it needs is built now. For **any** matrix, square or not,  $\mathbf{A}^* \mathbf{A}$  is positive semidefinite (Theorem 13.1.6), so it has a square root.

### Definition 13.2.1. Absolute Value of a Matrix

Let  $\mathbf{A} \in M_{m \times n}(F)$ . The **absolute value** of  $\mathbf{A}$  is

$$|\mathbf{A}| := (\mathbf{A}^* \mathbf{A})^{1/2} \in M_n(F),$$

the unique positive semidefinite square root of  $\mathbf{A}^* \mathbf{A}$ .

Note the size:  $|\mathbf{A}|$  is  $n \times n$ , not  $m \times n$ , because  $\mathbf{A}^* \mathbf{A}$  acts on the domain of  $\mathbf{A}$ . It is well defined because  $\mathbf{A}^* \mathbf{A} \succeq \mathbf{0}$  and Theorem 13.2.1 gives exactly one root.

### Proposition 13.2.2. What the absolute value preserves

Let  $\mathbf{A} \in M_{m \times n}(F)$ . Then for every  $\mathbf{x} \in F^n$ ,

$$\| |\mathbf{A}| \mathbf{x} \| = \| \mathbf{A} \mathbf{x} \|.$$

Consequently  $\text{nul} |\mathbf{A}| = \text{nul} \mathbf{A}$  and  $\text{rank} |\mathbf{A}| = \text{rank} \mathbf{A}$ .

*Proof.* Since  $|\mathbf{A}|$  is Hermitian and  $|\mathbf{A}|^2 = \mathbf{A}^* \mathbf{A}$ ,

$$\| |\mathbf{A}| \mathbf{x} \|^2 = \langle |\mathbf{A}|^2 \mathbf{x}, \mathbf{x} \rangle = \langle \mathbf{A}^* \mathbf{A} \mathbf{x}, \mathbf{x} \rangle = \| \mathbf{A} \mathbf{x} \|^2,$$

using Lemma 11.4.1 twice. Both sides are nonnegative reals, so the norms agree. Hence  $|\mathbf{A}| \mathbf{x} = \mathbf{0}$  if and only if  $\mathbf{A} \mathbf{x} = \mathbf{0}$ , and the rank statement follows by Theorem 3.5.7. ■

So  $|\mathbf{A}|$  does exactly what  $\mathbf{A}$  does to lengths, and nothing else. That single identity is the whole engine of the polar decomposition: the map  $|\mathbf{A}| \mathbf{x} \mapsto \mathbf{A} \mathbf{x}$  preserves norms, so it extends to a unitary matrix, and  $\mathbf{A} = \mathbf{W} |\mathbf{A}|$  follows.

### Example. Two Absolute Values

Compute  $|\mathbf{A}|$  for  $\mathbf{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$  and for  $\mathbf{B} = \begin{pmatrix} 3 & 0 \\ 4 & 0 \end{pmatrix}$ , both over  $\mathbb{R}$ .

*Solution.* For the quarter turn,  $\mathbf{A}^\top \mathbf{A} = \mathbf{I}$ , so  $|\mathbf{A}| = \mathbf{I}^{1/2} = \mathbf{I}$ . This is right: a rotation changes no length, so the matrix that does the same thing to lengths and nothing else is the identity.

For  $\mathbf{B}$ ,  $\mathbf{B}^\top \mathbf{B} = \begin{pmatrix} 25 & 0 \\ 0 & 0 \end{pmatrix}$ , whose positive square root is  $|\mathbf{B}| = \begin{pmatrix} 5 & 0 \\ 0 & 0 \end{pmatrix}$ . Check [Proposition 13.2.2](#):  $\|\mathbf{B}\mathbf{x}\| = \|(3x_1, 4x_1)\| = 5|x_1| = \| |\mathbf{B}|\mathbf{x} \|$ . Both have rank 1 and null space  $\text{span}(\mathbf{e}_2)$ .

*Warning.*  $|\mathbf{A}|$  is not the entrywise absolute value, and  $|\mathbf{AB}| \neq |\mathbf{A}||\mathbf{B}|$ . For the quarter turn above, the entrywise absolute value is  $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ , while  $|\mathbf{A}| = \mathbf{I}$ . For the product rule, take

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

Then  $\mathbf{AB} = \mathbf{O}$ , so  $|\mathbf{AB}| = \mathbf{O}$ , while  $|\mathbf{A}| = \mathbf{A}$  and  $|\mathbf{B}| = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ , giving

$$|\mathbf{A}||\mathbf{B}| = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \neq \mathbf{O}.$$

That product is not even Hermitian, so it could not be the absolute value of anything. The modulus of a product of **numbers** is the product of the moduli because numbers commute; matrices do not.

### 13.2.3 Cholesky: elimination on a positive definite matrix

Chapter 3 ran elimination on a symmetric matrix and found that the two triangular factors are secretly the same one:  $\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{L}^\top$  (Theorem 3.6.8). It then observed that when every entry of  $\mathbf{D}$  is positive one can split  $\mathbf{D} = \mathbf{D}^{1/2}\mathbf{D}^{1/2}$  and write  $\mathbf{A} = (\mathbf{L}\mathbf{D}^{1/2})(\mathbf{L}\mathbf{D}^{1/2})^\top$ , called this the **Cholesky factorization**, and promised that Chapter 13 would connect it with positive definite matrices. Here is the connection, and it is the best kind: the hypothesis “every entry of  $\mathbf{D}$  is positive” is not an extra assumption at all. It holds exactly when  $\mathbf{A}$  is positive definite, and it can be checked while the elimination runs.

#### Theorem 13.2.3. Cholesky Factorization

Let  $\mathbf{A} \in M_n(F)$  be Hermitian. Then  $\mathbf{A} \succ \mathbf{O}$  if and only if

$$\mathbf{A} = \mathbf{L}\mathbf{L}^*$$

for some lower triangular  $\mathbf{L} \in M_n(F)$  whose diagonal entries are **real and strictly positive**. In that case  $\mathbf{L}$  is unique.

*Idea.*  $(\Leftarrow)$  is one line: such an  $\mathbf{L}$  is invertible, and  $\mathbf{L}\mathbf{L}^* = (\mathbf{L}^*)^*\mathbf{L}^*$  is Theorem 13.1.5 (c) with  $\mathbf{B} = \mathbf{L}^*$ .

$(\Rightarrow)$  is Chapter 3’s elimination with its hypothesis supplied, run symmetrically. One pass clears the first column and the first row at once and leaves a smaller matrix that is again

positive definite, so an induction produces  $\mathbf{A} = \mathbf{MDM}^*$  with  $\mathbf{M}$  unit lower triangular and every pivot  $d_k > 0$ . Absorbing  $\mathbf{D}^{1/2}$  into  $\mathbf{M}$  finishes it. For uniqueness, strip the positive diagonal off each candidate  $\mathbf{L}$ ; what is left is an LU factorization of  $\mathbf{A}$ , and Chapter 3 already proved those unique.

*Proof.* ( $\Leftarrow$ ) Suppose  $\mathbf{A} = \mathbf{L}\mathbf{L}^*$  with  $\mathbf{L}$  lower triangular and every  $\ell_{kk} > 0$ . Then  $\mathbf{L}$  is invertible by Lemma 3.6.2, hence so is  $\mathbf{L}^*$ , and  $\mathbf{A} = (\mathbf{L}^*)^*\mathbf{L}^*$  is positive definite by Theorem 13.1.5 ((c)  $\Rightarrow$  (a)).

( $\Rightarrow$ ) Suppose  $\mathbf{A} \succ \mathbf{O}$ .

*Claim.* There are a unit lower triangular  $\mathbf{M} \in M_n(F)$  and a diagonal  $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$  whose entries are **real and strictly positive**, with  $\mathbf{A} = \mathbf{MDM}^*$ .

*Proof.* Induction on  $n$ . For  $n = 1$ , the single entry is  $a_{11} = \langle \mathbf{A}\mathbf{e}_1, \mathbf{e}_1 \rangle > 0$  by (P2) of Definition 13.1.1 in its strict form, so  $\mathbf{M} = (1)$  and  $\mathbf{D} = (a_{11})$  will do.

Let  $n \geq 2$ . Since  $\mathbf{A}^* = \mathbf{A}$ , we may write

$$\mathbf{A} = \begin{pmatrix} \alpha & \mathbf{c}^* \\ \mathbf{c} & \mathbf{A}' \end{pmatrix}, \quad \alpha = a_{11}, \quad \mathbf{c} \in F^{n-1}, \quad \mathbf{A}' \in M_{n-1}(F) \text{ Hermitian.}$$

Here  $\alpha = \langle \mathbf{A}\mathbf{e}_1, \mathbf{e}_1 \rangle > 0$ , so  $\alpha^{-1}$  exists and is a positive real number. Put

$$\mathbf{S} := \mathbf{A}' - \alpha^{-1}\mathbf{c}\mathbf{c}^*,$$

which is Hermitian, since  $(\mathbf{c}\mathbf{c}^*)^* = \mathbf{c}\mathbf{c}^*$  and  $\alpha$  is real.

*The smaller matrix is again positive definite.* Let  $\mathbf{y} \in F^{n-1}$  with  $\mathbf{y} \neq \mathbf{0}$ , set  $t = -\alpha^{-1}\mathbf{c}^*\mathbf{y}$  and  $\mathbf{x} = (t, \mathbf{y}) \in F^n$ , which is nonzero because  $\mathbf{y}$  is. Multiplying out in blocks (Theorem 8.1.1),

$$\langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle = \mathbf{x}^*\mathbf{A}\mathbf{x} = \alpha|t|^2 + \bar{t}\mathbf{c}^*\mathbf{y} + \mathbf{y}^*\mathbf{c}t + \mathbf{y}^*\mathbf{A}'\mathbf{y}.$$

With  $\mathbf{c}^*\mathbf{y} = -\alpha t$  and  $\mathbf{y}^*\mathbf{c} = \overline{\mathbf{c}^*\mathbf{y}} = -\alpha\bar{t}$ , the first three terms are

$$\alpha|t|^2 - \alpha|t|^2 - \alpha|t|^2 = -\alpha|t|^2 = -\alpha^{-1}|\mathbf{c}^*\mathbf{y}|^2,$$

so  $\langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle = \mathbf{y}^*(\mathbf{A}' - \alpha^{-1}\mathbf{c}\mathbf{c}^*)\mathbf{y} = \langle \mathbf{S}\mathbf{y}, \mathbf{y} \rangle$ . The left side is  $> 0$  because  $\mathbf{A} \succ \mathbf{O}$  and  $\mathbf{x} \neq \mathbf{0}$ . Hence  $\mathbf{S} \succ \mathbf{O}$ .

*The induction step.* By the induction hypothesis,  $\mathbf{S} = \mathbf{M}'\mathbf{D}'(\mathbf{M}')^*$  with  $\mathbf{M}' \in M_{n-1}(F)$  unit lower triangular and  $\mathbf{D}' = \text{diag}(d_2, \dots, d_n)$  real with positive entries. Put

$$\mathbf{M} = \begin{pmatrix} 1 & \mathbf{0}^\top \\ \alpha^{-1}\mathbf{c} & \mathbf{M}' \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} \alpha & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{D}' \end{pmatrix}.$$

Then  $\mathbf{M}$  is unit lower triangular and  $\mathbf{D}$  is diagonal with the positive real entries  $\alpha, d_2, \dots, d_n$ . Block multiplication (Theorem 8.1.1) gives

$$\mathbf{MDM}^* = \begin{pmatrix} \alpha & \mathbf{0}^\top \\ \mathbf{c} & \mathbf{M}'\mathbf{D}' \end{pmatrix} \begin{pmatrix} 1 & \alpha^{-1}\mathbf{c}^* \\ \mathbf{0} & (\mathbf{M}')^* \end{pmatrix} = \begin{pmatrix} \alpha & \mathbf{c}^* \\ \mathbf{c} & \alpha^{-1}\mathbf{c}\mathbf{c}^* + \mathbf{M}'\mathbf{D}'(\mathbf{M}')^* \end{pmatrix},$$

and  $\alpha^{-1}\mathbf{c}\mathbf{c}^* + \mathbf{M}'\mathbf{D}'(\mathbf{M}')^* = \alpha^{-1}\mathbf{c}\mathbf{c}^* + \mathbf{S} = \mathbf{A}'$ . So  $\mathbf{MDM}^* = \mathbf{A}$ , which proves the claim.

The claim is one pass of elimination performed symmetrically, and  $d_1 = \alpha, d_2, \dots, d_n$  are its pivots. Over  $\mathbb{R}$  it is the  $\mathbf{L}\mathbf{D}\mathbf{L}^\top$  factorization of Theorem 3.6.8 with the hypothesis that makes every pivot positive. ■

Write  $d_1, \dots, d_n$  for the diagonal entries of  $\mathbf{D}$ , all of them positive real numbers. Then  $\mathbf{D}^{1/2} := \text{diag}(\sqrt{d_1}, \dots, \sqrt{d_n})$  is a real matrix, and putting

$$\mathbf{L} = \mathbf{M}\mathbf{D}^{1/2}$$

gives a lower triangular matrix — its  $(i, j)$ -entry is  $m_{ij}\sqrt{d_j}$ , which vanishes for  $i < j$  — whose  $k$ -th diagonal entry is  $\sqrt{d_k} > 0$ , and

$$\mathbf{L}\mathbf{L}^* = \mathbf{M}\mathbf{D}^{1/2}(\mathbf{D}^{1/2})^*\mathbf{M}^* = \mathbf{M}\mathbf{D}\mathbf{M}^* = \mathbf{A}.$$

*Uniqueness.* Suppose  $\mathbf{A} = \mathbf{L}_1\mathbf{L}_1^* = \mathbf{L}_2\mathbf{L}_2^*$ , with each  $\mathbf{L}_i$  lower triangular with positive diagonal entries. Split off the diagonal: write  $\mathbf{\Delta}_i = \text{diag}((\mathbf{L}_i)_{11}, \dots, (\mathbf{L}_i)_{nn})$ , a real diagonal matrix with positive entries, and  $\mathbf{M}_i = \mathbf{L}_i\mathbf{\Delta}_i^{-1}$ , which is lower triangular with every diagonal entry 1. Then  $\mathbf{L}_i = \mathbf{M}_i\mathbf{\Delta}_i$  and

$$\mathbf{A} = \mathbf{M}_i\mathbf{\Delta}_i^2\mathbf{M}_i^* = \mathbf{M}_i(\mathbf{\Delta}_i^2\mathbf{M}_i^*).$$

The second factor is upper triangular, its  $(k, l)$ -entry being  $(\mathbf{\Delta}_i^2)_{kk}(\overline{\mathbf{M}_i})_{lk}$ , which vanishes for  $k > l$ . So for  $i = 1$  and  $i = 2$  this displays an LU factorization of  $\mathbf{A}$ , which is invertible:  $\mathbf{L}_1$  is triangular with nonzero diagonal, hence invertible by Lemma 3.6.2, and  $\mathbf{A} = \mathbf{L}_1\mathbf{L}_1^*$  is a product of two invertible matrices. By Theorem 3.6.5 the two agree:

$$\mathbf{M}_1 = \mathbf{M}_2 \quad \text{and} \quad \mathbf{\Delta}_1^2\mathbf{M}_1^* = \mathbf{\Delta}_2^2\mathbf{M}_2^*.$$

Canceling the invertible  $\mathbf{M}_1^* = \mathbf{M}_2^*$  gives  $\mathbf{\Delta}_1^2 = \mathbf{\Delta}_2^2$ , and since both have positive diagonal entries,  $\mathbf{\Delta}_1 = \mathbf{\Delta}_2$ . Hence  $\mathbf{L}_1 = \mathbf{M}_1\mathbf{\Delta}_1 = \mathbf{M}_2\mathbf{\Delta}_2 = \mathbf{L}_2$ . This proves the theorem. ■

*Remark.* The pivots are determinant ratios. For  $i, j \leq k$ , a term of  $(\mathbf{M}\mathbf{D}\mathbf{M}^*)_{ij} = \sum_l m_{il}(\mathbf{D}\mathbf{M}^*)_{lj}$  with  $l > k$  has  $l > i$  and so  $m_{il} = 0$ ; the sum therefore stops at  $l = k$ , which says that the leading principal submatrices multiply:  $\mathbf{A}_k = \mathbf{M}_k\mathbf{D}_k(\mathbf{M}^*)_k = \mathbf{M}_k\mathbf{D}_k\mathbf{M}_k^*$ . Taking determinants with Theorem 7.5.4 and Theorem 7.4.6 gives  $\det \mathbf{A}_k = d_1 \cdots d_k$ , so  $d_1 = \det \mathbf{A}_1$  and  $d_k = \det \mathbf{A}_k / \det \mathbf{A}_{k-1}$  for  $k \geq 2$ . Read the other way, this is Sylvester's criterion (Theorem 13.1.5 (d)) appearing inside the elimination: the leading minors are positive exactly when the pivots are.

Reading the equation  $\mathbf{A} = \mathbf{L}\mathbf{L}^*$  entry by entry turns it into a recipe, and the recipe is the standard way to factor by hand. Comparing the  $(j, j)$  entry of both sides gives  $a_{jj} = \sum_{k \leq j} |\ell_{jk}|^2$ , and comparing the  $(i, j)$  entry for  $i > j$  gives  $a_{ij} = \sum_{k \leq j} \ell_{ik}\overline{\ell_{jk}}$ . Solving each for its one new unknown, column by column:

$$\begin{aligned} \ell_{jj} &= \left( a_{jj} - \sum_{k < j} |\ell_{jk}|^2 \right)^{1/2}, \\ \ell_{ij} &= \left( a_{ij} - \sum_{k < j} \ell_{ik}\overline{\ell_{jk}} \right) / \ell_{jj} \quad (i > j). \end{aligned}$$

**Example. A Cholesky Factorization, and One That Fails**

Factor

$$\mathbf{A} = \begin{pmatrix} 4 & -2 & 6 \\ -2 & 5 & -1 \\ 6 & -1 & 26 \end{pmatrix}$$

as  $\mathbf{L}\mathbf{L}^\top$ , and run the same algorithm on the matrix  $\mathbf{A}'$  obtained from  $\mathbf{A}$  by changing the (3,3) entry to 9.

*Solution. Column 1.*  $\ell_{11} = \sqrt{4} = 2$ ; then  $\ell_{21} = -2/2 = -1$  and  $\ell_{31} = 6/2 = 3$ .

*Column 2.*  $\ell_{22} = \sqrt{5 - (-1)^2} = \sqrt{4} = 2$ ; then

$$\ell_{32} = \frac{a_{32} - \ell_{31}\ell_{21}}{\ell_{22}} = \frac{-1 - 3(-1)}{2} = 1.$$

*Column 3.*  $\ell_{33} = \sqrt{26 - 3^2 - 1^2} = \sqrt{16} = 4$ . So

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

and multiplying back confirms  $\mathbf{L}\mathbf{L}^\top = \mathbf{A}$ . The diagonal entries 2, 2, 4 are positive, so  $\mathbf{A} \succ \mathbf{O}$  by [Theorem 13.2.3](#); their squares 4, 4, 16 are the pivots  $d_k = \det \mathbf{A}_k / \det \mathbf{A}_{k-1}$ , consistent with  $\det \mathbf{A}_1 = 4$ ,  $\det \mathbf{A}_2 = 16$ ,  $\det \mathbf{A} = 256$ .

*The failure.* For  $\mathbf{A}'$ , columns 1 and 2 are unchanged, and the last step asks for

$$\ell_{33} = \sqrt{9 - 3^2 - 1^2} = \sqrt{-1},$$

which is not a positive real number. The algorithm stops, and by [Theorem 13.2.3](#) its stopping is a proof:  $\mathbf{A}'$  is not positive definite. Indeed  $\det \mathbf{A}' = -16 < 0$ , so Sylvester's criterion fails at  $k = 3$ .

The example shows the practical point. **The algorithm is the test.** To decide whether a numerical Hermitian matrix is positive definite, one does not compute eigenvalues and one does not compute  $n$  determinants; one starts the Cholesky recursion and watches whether every quantity under a square root stays positive. Sylvester's criterion guarantees that this verdict is correct, because  $d_1 \cdots d_k = \det \mathbf{A}_k$ .

*Remark.* Two remarks on cost, both of which belong properly to Chapter 24. First, Cholesky does about half the arithmetic of a general LU factorization, because it computes only  $\mathbf{L}$  and never the second triangular factor. Second, and less obvious, it needs no pivoting: the recursion above never divides by something that positive definiteness allows to be zero, so the row swaps that Chapter 3 introduced for stability are simply not required here. Positive definiteness is the one structural hypothesis under which elimination is unconditionally safe.

*Warning.* **Cholesky needs definiteness, not just semidefiniteness, for uniqueness.** The matrix  $\mathbf{A} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$  is positive semidefinite and does factor as  $\mathbf{L}\mathbf{L}^\top$ , for instance with  $\mathbf{L} = \mathbf{A}$ . But the diagonal entry  $\ell_{11} = 0$  is not positive, so [Theorem 13.2.3](#) does not apply — and indeed uniqueness fails, since  $\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$  and  $\begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}$  are two different lower triangular matrices  $\mathbf{L}$  with  $\mathbf{L}\mathbf{L}^\top = \mathbf{A}$ . Also note that the  $\mathbf{L}$  of Cholesky is **not** the  $\mathbf{L}$  of Chapter 3's LU factorization: the latter is unit lower triangular, and the two differ by the factor  $\mathbf{D}^{1/2}$ .

*Quick check.* Let  $\mathbf{A} \succ \mathbf{O}$  with Cholesky factor  $\mathbf{L}$ . What is  $\det \mathbf{A}$  in terms of  $\mathbf{L}$ , and what is the Cholesky factor of  $\mathbf{A}^{-1}$ ?

*Solution.*  $\det \mathbf{A} = \det \mathbf{L} \det \mathbf{L}^* = |\ell_{11} \cdots \ell_{nn}|^2 = (\ell_{11} \cdots \ell_{nn})^2$ , by Theorem 7.5.4 and Theorem 7.4.6, the last step because the  $\ell_{kk}$  are positive reals. For the inverse,  $\mathbf{A}^{-1} = (\mathbf{L}\mathbf{L}^*)^{-1} = (\mathbf{L}^*)^{-1}\mathbf{L}^{-1} = \mathbf{M}\mathbf{M}^*$  with  $\mathbf{M} = (\mathbf{L}^*)^{-1}$ , which is **upper** triangular; so this is not itself a Cholesky factorization. The Cholesky factor of  $\mathbf{A}^{-1}$  is a different matrix, which is why solving  $\mathbf{A}\mathbf{x} = \mathbf{b}$  is done by two triangular substitutions with  $\mathbf{L}$  rather than by inverting anything.

## 13.2.4 Exercises

### 13.2.4.1 A. Check your understanding

#### Exercise. A1

- State what  $\mathbf{A}^{1/2}$  is, including the two conditions that pin it down.
- Define  $|\mathbf{A}|$  for  $\mathbf{A} \in M_{m \times n}(F)$  and say what size it is.
- State the Cholesky factorization theorem.
- True or false: every Hermitian matrix has exactly one Hermitian square root. Justify your answer.
- Explain in one sentence why the Cholesky algorithm is a test for positive definiteness.

*Solution.*

- For  $\mathbf{A} \succeq \mathbf{O}$ ,  $\mathbf{A}^{1/2}$  is the unique matrix  $\mathbf{C}$  with  $\mathbf{C} \succeq \mathbf{O}$  and  $\mathbf{C}^2 = \mathbf{A}$  ([Theorem 13.2.1](#)). Both conditions are needed: positivity, and squaring to  $\mathbf{A}$ .
- $|\mathbf{A}| = (\mathbf{A}^* \mathbf{A})^{1/2}$ , which is  $n \times n$ , the size of the domain of  $\mathbf{A}$  ([Definition 13.2.1](#)).
- A Hermitian  $\mathbf{A}$  is positive definite if and only if  $\mathbf{A} = \mathbf{L}\mathbf{L}^*$  with  $\mathbf{L}$  lower triangular with positive real diagonal entries, and then  $\mathbf{L}$  is unique ([Theorem 13.2.3](#)).
- False, twice over. A Hermitian matrix with a negative eigenvalue has no Hermitian square root at all, since  $\mathbf{C}^2 \succeq \mathbf{O}$  always. And a positive semidefinite  $\mathbf{A}$  with two distinct eigenvalues has at least two Hermitian square roots,  $\pm \mathbf{A}^{1/2}$ . Uniqueness needs the root itself to be positive semidefinite.

- (e) Because the algorithm computes  $\ell_{jj}^2 = \det \mathbf{A}_j / \det \mathbf{A}_{j-1}$  one step at a time, so it produces a positive real  $\ell_{jj}$  at every step exactly when Sylvester's criterion holds.

### 13.2.4.2 B. Practice

#### Exercise. B1: Square roots

Compute  $\mathbf{A}^{1/2}$  for each of the following, and in each case give a polynomial  $p$  with  $p(\mathbf{A}) = \mathbf{A}^{1/2}$ .

- (a)  $\mathbf{A} = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}$ .
- (b)  $\mathbf{A} = \begin{pmatrix} 10 & -6 \\ -6 & 10 \end{pmatrix}$ .

*Solution.*

- (a)  $\mathbf{A}(1, 1) = (9, 9)$  and  $\mathbf{A}(1, -1) = (1, -1)$ , so the eigenvalues are 9 and 1 with eigenlines  $\text{span}((1, 1))$  and  $\text{span}((1, -1))$ ; both positive, so  $\mathbf{A} \succ \mathbf{O}$ . With  $\mathbf{P}_1, \mathbf{P}_2$  the projections onto those lines,

$$\mathbf{A}^{1/2} = 3\mathbf{P}_1 + \mathbf{P}_2 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

For the polynomial, solve  $p(9) = 3$ ,  $p(1) = 1$ :  $p(x) = \frac{1}{4}x + \frac{3}{4}$ , and  $\frac{1}{4}\mathbf{A} + \frac{3}{4}\mathbf{I} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ .

Check:  $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}^2 = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}$ .

- (b)  $\mathbf{A}(1, -1) = (16, -16)$  and  $\mathbf{A}(1, 1) = (4, 4)$ , so the eigenvalues are 16 and 4. Then

$$\mathbf{A}^{1/2} = 4 \cdot \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + 2 \cdot \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}.$$

For the polynomial,  $p(16) = 4$  and  $p(4) = 2$  give  $p(x) = \frac{1}{6}x + \frac{4}{3}$ , and  $\frac{1}{6}\mathbf{A} + \frac{4}{3}\mathbf{I} = \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}$ . Check:  $\begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}^2 = \begin{pmatrix} 10 & -6 \\ -6 & 10 \end{pmatrix}$ .

#### Exercise. B2: Cholesky, and a planted failure

- (a) Find the Cholesky factor of  $\mathbf{A} = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 5 & 1 \\ -1 & 1 & 14 \end{pmatrix}$ , and hence write down  $\det \mathbf{A}$ .
- (b) Run the same algorithm on  $\mathbf{A}' = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 5 & 1 \\ -1 & 1 & 9 \end{pmatrix}$  and say what the outcome proves.

*Solution.*

(a)  $\ell_{11} = \sqrt{1} = 1$ ,  $\ell_{21} = 2$ ,  $\ell_{31} = -1$ . Then  $\ell_{22} = \sqrt{5-4} = 1$  and

$$\ell_{32} = \frac{1 - (-1)(2)}{1} = 3.$$

Finally  $\ell_{33} = \sqrt{14 - 1 - 9} = \sqrt{4} = 2$ . So

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 3 & 2 \end{pmatrix},$$

and  $\mathbf{L}\mathbf{L}^\top = \mathbf{A}$ . All diagonal entries are positive, so  $\mathbf{A} \succ \mathbf{O}$ . By the Quick check above,  $\det \mathbf{A} = (1 \cdot 1 \cdot 2)^2 = 4$ .

(b) Columns 1 and 2 are identical, and the last step asks for  $\ell_{33} = \sqrt{9 - 1 - 9} = \sqrt{-1}$ . The algorithm fails, so  $\mathbf{A}'$  is not positive definite ([Theorem 13.2.3](#)). Confirming by Sylvester:  $\det \mathbf{A}'_1 = 1$ ,  $\det \mathbf{A}'_2 = 1$ , and  $\det \mathbf{A}' = 1 \cdot 44 - 2 \cdot 19 - 7 = -1 < 0$ .

### Exercise. B3: An absolute value

Let  $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \in M_2(\mathbb{R})$ . Compute  $|\mathbf{A}|$ , verify  $\| |\mathbf{A}| \mathbf{x} \| = \| \mathbf{A} \mathbf{x} \|$  directly, and determine whether  $|\mathbf{A}| = \mathbf{A}$ .

*Solution.*  $\mathbf{A}^\top \mathbf{A} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$ , whose eigenvalues are 4 and 0 with eigenlines  $\text{span}((1, 1))$  and  $\text{span}((1, -1))$ . Hence

$$|\mathbf{A}| = 2 \cdot \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \mathbf{O} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \mathbf{A}.$$

Directly:  $\mathbf{A}\mathbf{x} = (x_1 + x_2)(1, 1)$ , so  $\| \mathbf{A}\mathbf{x} \| = \sqrt{2} |x_1 + x_2|$ , and the same computation applies to  $|\mathbf{A}| = \mathbf{A}$ . The equality  $|\mathbf{A}| = \mathbf{A}$  is no accident:  $\mathbf{A} \succeq \mathbf{O}$  here, and for a positive semidefinite  $\mathbf{A}$  one has  $\mathbf{A}^* \mathbf{A} = \mathbf{A}^2$ , so  $|\mathbf{A}| = (\mathbf{A}^2)^{1/2} = \mathbf{A}$  by uniqueness in [Theorem 13.2.1](#).

### 13.2.4.3 C. Going deeper

#### Exercise. C1: The trace of a product of positive matrices

Let  $\mathbf{A}, \mathbf{B} \in M_n(F)$  with  $\mathbf{A} \succeq \mathbf{O}$  and  $\mathbf{B} \succeq \mathbf{O}$ .

- (a) Prove that  $\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{A}^{1/2}\mathbf{B}\mathbf{A}^{1/2})$ , and deduce that  $\text{tr}(\mathbf{A}\mathbf{B}) \geq 0$ .
- (b) Prove that  $\text{tr}(\mathbf{A}\mathbf{B}) = 0$  if and only if  $\mathbf{A}\mathbf{B} = \mathbf{O}$ .

*Hint for (b): a positive semidefinite matrix of trace 0 is  $\mathbf{O}$ .*

*Solution.*

- (a) By Theorem 0.9.6,  $\text{tr}(\mathbf{XY}) = \text{tr}(\mathbf{YX})$ . With  $\mathbf{X} = \mathbf{A}^{1/2}$  and  $\mathbf{Y} = \mathbf{A}^{1/2}\mathbf{B}$ ,

$$\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{A}^{1/2}\mathbf{A}^{1/2}\mathbf{B}) = \text{tr}(\mathbf{A}^{1/2}\mathbf{BA}^{1/2}).$$

Now  $\mathbf{A}^{1/2}\mathbf{BA}^{1/2} = (\mathbf{A}^{1/2})^*\mathbf{BA}^{1/2} \succeq \mathbf{O}$  by Proposition 13.1.2 (a), since  $\mathbf{A}^{1/2}$  is Hermitian. The trace of a positive semidefinite matrix is  $\geq 0$  by Exercise (C1: Zero trace forces zero) (a). Hence  $\text{tr}(\mathbf{AB}) \geq 0$ .

- (b)  $(\Leftarrow)$  is clear.  $(\Rightarrow)$  Suppose  $\text{tr}(\mathbf{AB}) = 0$ . By (a),  $\mathbf{C} := \mathbf{A}^{1/2}\mathbf{BA}^{1/2} \succeq \mathbf{O}$  and  $\text{tr} \mathbf{C} = 0$ , so  $\mathbf{C} = \mathbf{O}$  by Exercise (C1: Zero trace forces zero) (b). Put  $\mathbf{M} = \mathbf{B}^{1/2}\mathbf{A}^{1/2}$ . Then

$$\mathbf{M}^*\mathbf{M} = \mathbf{A}^{1/2}\mathbf{B}^{1/2}\mathbf{B}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}^{1/2}\mathbf{BA}^{1/2} = \mathbf{C} = \mathbf{O},$$

so  $\mathbf{M} = \mathbf{O}$  by Lemma 11.4.3 (its null space is all of  $F^n$ , hence so is  $\text{nul } \mathbf{M}$ ). Therefore

$$\mathbf{AB} = \mathbf{A}^{1/2}(\mathbf{A}^{1/2}\mathbf{B}^{1/2})\mathbf{B}^{1/2} = \mathbf{A}^{1/2}\mathbf{M}^*\mathbf{B}^{1/2} = \mathbf{O}.$$

### Exercise. C2: Commuting positive matrices

Let  $\mathbf{A} \in M_n(F)$  with  $\mathbf{A} \succeq \mathbf{O}$ .

- (a) Prove **without** using Theorem 13.2.1 (a) that if  $\mathbf{BA} = \mathbf{AB}$ , then  $\mathbf{BA}^{1/2} = \mathbf{A}^{1/2}\mathbf{B}$ .  
*Hint: show that  $\mathbf{B}$  maps each eigenspace of  $\mathbf{A}$  into itself.*
- (b) Let  $\mathbf{B} \succeq \mathbf{O}$  with  $\mathbf{AB} = \mathbf{BA}$ . Prove that  $\mathbf{AB} \succeq \mathbf{O}$  and that  $(\mathbf{AB})^{1/2} = \mathbf{A}^{1/2}\mathbf{B}^{1/2}$ .

*Solution.*

- (a) Let  $\lambda \in \text{spec}(\mathbf{A})$  and  $\mathbf{Av} = \lambda\mathbf{v}$ . Then  $\mathbf{A}(\mathbf{Bv}) = \mathbf{B}(\mathbf{Av}) = \lambda\mathbf{Bv}$ , so  $\mathbf{B}$  maps  $E_\lambda(\mathbf{A})$  into itself. Since  $\mathbf{A} \succeq \mathbf{O}$ , its distinct eigenvalues  $\lambda_1, \dots, \lambda_k$  are nonnegative reals, and  $\mathbf{A}^{1/2} = \sum_i \sqrt{\lambda_i} \mathbf{P}_i$  with  $\mathbf{P}_i$  the orthogonal projection onto  $E_{\lambda_i}(\mathbf{A})$  (proof of Theorem 13.2.1). For  $\mathbf{v} \in E_{\lambda_j}(\mathbf{A})$  we have  $\mathbf{P}_i\mathbf{v} = \mathbf{0}$  for  $i \neq j$  and  $\mathbf{P}_j\mathbf{v} = \mathbf{v}$ , so  $\mathbf{A}^{1/2}\mathbf{v} = \sqrt{\lambda_j}\mathbf{v}$ . Now  $F^n = E_{\lambda_1}(\mathbf{A}) \oplus \dots \oplus E_{\lambda_k}(\mathbf{A})$  (Theorem 12.8.1), so it suffices to check  $\mathbf{BA}^{1/2}\mathbf{v} = \mathbf{A}^{1/2}\mathbf{Bv}$  for  $\mathbf{v} \in E_{\lambda_j}(\mathbf{A})$ . The left side is  $\sqrt{\lambda_j}\mathbf{Bv}$ ; the right side is  $\mathbf{A}^{1/2}(\mathbf{Bv}) = \sqrt{\lambda_j}\mathbf{Bv}$  as well, because  $\mathbf{Bv} \in E_{\lambda_j}(\mathbf{A})$  by the first sentence. Hence  $\mathbf{BA}^{1/2} = \mathbf{A}^{1/2}\mathbf{B}$ .
- (b) By (a),  $\mathbf{A}^{1/2}$  commutes with  $\mathbf{B}$ . Applying (a) once more, this time to  $\mathbf{B} \succeq \mathbf{O}$  with the commuting matrix  $\mathbf{A}^{1/2}$ , gives that  $\mathbf{A}^{1/2}$  commutes with  $\mathbf{B}^{1/2}$ . Put  $\mathbf{C} = \mathbf{A}^{1/2}\mathbf{B}^{1/2}$ . Then

$$\mathbf{C}^* = (\mathbf{B}^{1/2})^*(\mathbf{A}^{1/2})^* = \mathbf{B}^{1/2}\mathbf{A}^{1/2} = \mathbf{C},$$

so  $\mathbf{C}$  is Hermitian, and  $\mathbf{C}^2 = \mathbf{A}^{1/2}\mathbf{B}^{1/2}\mathbf{A}^{1/2}\mathbf{B}^{1/2} = \mathbf{AB}$  after one commutation. It remains to see  $\mathbf{C} \succeq \mathbf{O}$ . The matrices  $\mathbf{A}^{1/2}$  and  $\mathbf{B}^{1/2}$  are commuting Hermitian matrices, so by Corollary 12.7.2 (over  $\mathbb{C}$ ) or Corollary 12.7.3 (over  $\mathbb{R}$ ) there is a unitary  $\mathbf{U}$  with  $\mathbf{U}^*\mathbf{A}^{1/2}\mathbf{U} = \text{diag}(\alpha_i)$  and  $\mathbf{U}^*\mathbf{B}^{1/2}\mathbf{U} = \text{diag}(\beta_i)$ , all  $\alpha_i, \beta_i \geq 0$ . Then  $\mathbf{U}^*\mathbf{CU} = \text{diag}(\alpha_i\beta_i) \succeq \mathbf{O}$ , so  $\mathbf{C} \succeq \mathbf{O}$  by Proposition 13.1.2 (a). Hence  $\mathbf{AB} = \mathbf{C}^2 \succeq \mathbf{O}$  and, by uniqueness in Theorem 13.2.1,  $(\mathbf{AB})^{1/2} = \mathbf{C} = \mathbf{A}^{1/2}\mathbf{B}^{1/2}$ .

**Exercise. C3: Why definiteness is needed**

- (a) Give a positive semidefinite  $\mathbf{A} \in M_2(\mathbb{R})$ , not positive definite, together with two different lower triangular  $\mathbf{L}$  satisfying  $\mathbf{A} = \mathbf{L}\mathbf{L}^\top$ . Which hypothesis of [Theorem 13.2.3](#) fails?
- (b) Prove that if  $\mathbf{A} \succeq \mathbf{O}$  then  $\mathbf{A} + t\mathbf{I} \succ \mathbf{O}$  for every real  $t > 0$ , and hence that every positive semidefinite matrix is a limit of positive definite ones in the entrywise sense.

*Solution.*

- (a) Take  $\mathbf{A} = \text{diag}(0, 1)$ . Then  $\mathbf{L}_1 = \text{diag}(0, 1)$  and  $\mathbf{L}_2 = \text{diag}(0, -1)$  are both lower triangular with  $\mathbf{L}_i\mathbf{L}_i^\top = \mathbf{A}$ . The hypothesis that fails is that the diagonal entries of  $\mathbf{L}$  be strictly positive:  $\ell_{11} = 0$  here, which is exactly the first pivot vanishing. Uniqueness in [Theorem 13.2.3](#) is not available without it.
- (b)  $\mathbf{A} + t\mathbf{I}$  is Hermitian, and for  $\mathbf{x} \neq \mathbf{0}$ ,

$$\langle (\mathbf{A} + t\mathbf{I})\mathbf{x}, \mathbf{x} \rangle = \langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle + t\|\mathbf{x}\|^2 \geq t\|\mathbf{x}\|^2 > 0,$$

so  $\mathbf{A} + t\mathbf{I} \succ \mathbf{O}$ . Each entry of  $\mathbf{A} + t\mathbf{I}$  differs from the corresponding entry of  $\mathbf{A}$  by  $t$  or by 0, so letting  $t$  decrease to 0 recovers  $\mathbf{A}$  entry by entry. This is the standard device for extending a statement from positive definite to positive semidefinite matrices whenever the statement survives taking limits; Theorem 13.1.4 already used it in the form  $\det(\mathbf{A} + t\mathbf{I}) > 0$ .