

21.11 Means and Matrix Cauchy–Schwarz

Two inequalities about numbers refuse to die: $|\overline{a}b|^2 \leq |a|^2|b|^2$, and $2|\overline{a}b| \leq |a|^2 + |b|^2$. For matrices the products do not commute and the absolute values are matrices themselves, so it is not obvious what either statement should say. Sections 4 and 6 have given the right language: measure a matrix by a unitarily invariant norm, and compare two matrices by weak majorization of their singular values. In that language both inequalities survive, for **every** unitarily invariant norm at once. The second half of the section then builds a genuine geometric mean of two positive definite matrices, which is where two order facts from the earlier chapters pay off again: the operator monotonicity of $t \mapsto t^{1/2}$ from Chapter 13 §05 and that of $t \mapsto -1/t$ from Chapter 17 §11.

Throughout, $F = \mathbb{R}$ or $F = \mathbb{C}$ and all matrices are $n \times n$ over F , with $\|\cdot\|$ an arbitrary unitarily invariant norm on $M_n(F)$ (Definition 21.4.1). As always $\sigma_1(\mathbf{A}) \geq \cdots \geq \sigma_n(\mathbf{A}) \geq 0$ are the singular values in decreasing order and $\lambda_1(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{A})$ the eigenvalues of a Hermitian $\mathbf{A}$; for $\mathbf{A} \succeq \mathbf{O}$ the two lists agree. The order $\succeq$ is the Loewner order (Definition 13.5.1).

21.11.1 An off-diagonal block is small

Everything in the first half comes from one observation about a positive semidefinite matrix cut into four blocks: the corner blocks control the off-diagonal one, and they do so eigenvalue by eigenvalue.

Lemma 21.11.1. The Off-Diagonal Block of a Positive Semidefinite Matrix

Let $k, l \geq 1$, let $\mathbf{P} \in M_k(F)$, $\mathbf{Q} \in M_l(F)$, and $\mathbf{X} \in M_{k \times l}(F)$, and suppose that

$$\mathbf{M} = \begin{pmatrix} \mathbf{P} & \mathbf{X} \\ \mathbf{X}^* & \mathbf{Q} \end{pmatrix} \succeq \mathbf{O}.$$

Then $2\sigma_j(\mathbf{X}) \leq \lambda_j(\mathbf{M})$ for every $j = 1, \dots, \min(k, l)$.

Idea. Flipping the sign of the second block of coordinates is a unitary congruence, so it turns $\mathbf{M}$ into another positive semidefinite matrix — the same one with $\mathbf{X}$ replaced by $-\mathbf{X}$. Half the difference of the two is the Hermitian dilation of $\mathbf{X}$, whose eigenvalues are exactly $\pm\sigma_i(\mathbf{X})$ and zeros. Since what was subtracted is itself positive semidefinite, the dilation is dominated by $\frac{1}{2}\mathbf{M}$ in the Loewner order, and Chapter 17 §02 turns that into a comparison of eigenvalues index by index.

Proof. Put $\mathbf{S} = \mathbf{I}_k \oplus (-\mathbf{I}_l)$, which is Hermitian and unitary. By Theorem 8.1.1,

$$\mathbf{S}^*\mathbf{M}\mathbf{S} = \begin{pmatrix} \mathbf{P} & -\mathbf{X} \\ -\mathbf{X}^* & \mathbf{Q} \end{pmatrix},$$

and $\mathbf{S}^*\mathbf{M}\mathbf{S} \succeq \mathbf{O}$ because a congruence by an invertible matrix preserves positive semidefiniteness (Proposition 13.1.2 (a)). Subtracting,

$$\frac{1}{2}(\mathbf{M} - \mathbf{S}^*\mathbf{M}\mathbf{S}) = \begin{pmatrix} \mathbf{O} & \mathbf{X} \\ \mathbf{X}^* & \mathbf{O} \end{pmatrix} = \mathcal{H}(\mathbf{X}),$$

the Hermitian dilation of $\mathbf{X}$ (Proposition 17.9.3). Since $\frac{1}{2}\mathbf{S}^*\mathbf{M}\mathbf{S} \succeq \mathbf{O}$, this says

$$\mathcal{H}(\mathbf{X}) = \frac{1}{2}\mathbf{M} - \frac{1}{2}\mathbf{S}^*\mathbf{M}\mathbf{S} \preceq \frac{1}{2}\mathbf{M}.$$

By Corollary 17.2.4 the Loewner order compares every eigenvalue, so $\lambda_j(\mathcal{H}(\mathbf{X})) \leq \lambda_j(\frac{1}{2}\mathbf{M}) = \frac{1}{2}\lambda_j(\mathbf{M})$ for every j .

Finally, by Proposition 17.9.3, $\mathcal{H}(\mathbf{X})$ has eigenvalues $\pm\sigma_1(\mathbf{X}), \dots, \pm\sigma_p(\mathbf{X})$ and $k + l - 2p$ zeros, with multiplicity, where $p = \min(k, l)$. Sorted decreasingly, the first p of them are $\sigma_1(\mathbf{X}) \geq \dots \geq \sigma_p(\mathbf{X})$, since every remaining entry is $\leq 0 \leq \sigma_p(\mathbf{X})$. Hence $\lambda_j(\mathcal{H}(\mathbf{X})) = \sigma_j(\mathbf{X})$ for $j \leq p$, and the displayed inequality reads $2\sigma_j(\mathbf{X}) \leq \lambda_j(\mathbf{M})$. ■

21.11.2 A product against a sum of squares

The lemma becomes an inequality about products as soon as one remembers where positive semidefinite block matrices come from: $\mathbf{D}^*\mathbf{D}$, for $\mathbf{D}$ the block row built from the two matrices in question.

Theorem 21.11.2. Arithmetic–Geometric Mean Inequality for Matrices

Let $\mathbf{A}, \mathbf{B} \in M_n(F)$. Then

$$2\sigma_j(\mathbf{A}^*\mathbf{B}) \leq \lambda_j(\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*) \quad (j = 1, \dots, n),$$

and consequently, for every unitarily invariant norm on $M_n(F)$,

$$2\|\mathbf{A}^*\mathbf{B}\| \leq \|\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*\|.$$

Idea. Put the two matrices side by side in one $n \times 2n$ block row $\mathbf{D} = (\mathbf{A} \ \mathbf{B})$. Then $\mathbf{D}^*\mathbf{D}$ is the $2n \times 2n$ positive semidefinite matrix whose off-diagonal block is exactly $\mathbf{A}^*\mathbf{B}$, so Lemma 21.11.1 applies; and $\mathbf{D}\mathbf{D}^*$, which has the same nonzero eigenvalues, is the sum $\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*$. The norm statement is then the singular value statement read through §06's dominance theorem.

Proof. Let $\mathbf{D} = (\mathbf{A} \ \mathbf{B}) \in M_{n \times 2n}(F)$, the block row with $\mathbf{A}$ then $\mathbf{B}$. By Theorem 8.1.1,

$$\mathbf{D}^*\mathbf{D} = \begin{pmatrix} \mathbf{A}^*\mathbf{A} & \mathbf{A}^*\mathbf{B} \\ \mathbf{B}^*\mathbf{A} & \mathbf{B}^*\mathbf{B} \end{pmatrix}, \quad \mathbf{D}\mathbf{D}^* = \mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*.$$

The matrix $\mathbf{D}^*\mathbf{D}$ is Hermitian and satisfies $\mathbf{x}^*\mathbf{D}^*\mathbf{D}\mathbf{x} = \|\mathbf{D}\mathbf{x}\|^2 \geq 0$ for every $\mathbf{x} \in F^{2n}$, so $\mathbf{D}^*\mathbf{D} \succeq \mathbf{O}$ (Definition 13.1.1). Its off-diagonal block is $\mathbf{A}^*\mathbf{B}$, so Lemma 21.11.1 with $k = l = n$ gives

$$2\sigma_j(\mathbf{A}^*\mathbf{B}) \leq \lambda_j(\mathbf{D}^*\mathbf{D}) \quad (j = 1, \dots, n).$$

It remains to identify $\lambda_j(\mathbf{D}^*\mathbf{D})$ with $\lambda_j(\mathbf{D}\mathbf{D}^*)$ for $j \leq n$. Applying Theorem 9.10.4 to the pair $\mathbf{D}^* \in M_{2n \times n}$ and $\mathbf{D} \in M_{n \times 2n}$ gives $x^n p_{\mathbf{D}^*\mathbf{D}}(x) = x^{2n} p_{\mathbf{D}\mathbf{D}^*}(x)$, that is $p_{\mathbf{D}^*\mathbf{D}} = x^n p_{\mathbf{D}\mathbf{D}^*}$: the eigenvalue list of $\mathbf{D}^*\mathbf{D}$ is that of $\mathbf{D}\mathbf{D}^*$ with n extra zeros. Both matrices are positive

semidefinite, so all their eigenvalues are ≥ 0 (Theorem 13.1.4 (b)) and the first n entries of the two decreasing lists agree. This proves the first display.

For the second, both $2\mathbf{A}^*\mathbf{B}$ and $\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*$ lie in $M_n(F)$, and the singular values of the positive semidefinite $\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*$ are its eigenvalues. So the first display says $\sigma_j(2\mathbf{A}^*\mathbf{B}) \leq \sigma_j(\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*)$ for every j . Two decreasing lists of nonnegative numbers with the first dominated entry by entry satisfy $\sigma(2\mathbf{A}^*\mathbf{B}) \prec_w \sigma(\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*)$, since each top- k partial sum is dominated term by term (Definition 17.8.1). Now Theorem 21.6.1 turns that weak majorization into $\|\mathbf{A}^*\mathbf{B}\| \leq \|\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*\|$, and $\|\mathbf{A}^*\mathbf{B}\| = 2\|\mathbf{A}^*\mathbf{B}\|$ by homogeneity of a norm. ■

Warning. The two sums are not interchangeable. It is tempting to write $\mathbf{A}^*\mathbf{A} + \mathbf{B}^*\mathbf{B}$ on the right, since that is what pairs with $\mathbf{A}^*\mathbf{B}$ in the block matrix $\mathbf{D}^*\mathbf{D}$. The resulting statement is **false**. Take

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

Then $\mathbf{A}^*\mathbf{A} = \text{diag}(1, 0)$ and $\mathbf{B}^*\mathbf{B} = \text{diag}(0, 1)$, so $\mathbf{A}^*\mathbf{A} + \mathbf{B}^*\mathbf{B} = \mathbf{I}$ and $\|\mathbf{I}\|_2 = 1$; while $\mathbf{A}^*\mathbf{B} = \mathbf{B}$ has $\|\mathbf{B}\|_2 = 1$, so $2\|\mathbf{A}^*\mathbf{B}\| = 2 > 1$ in the spectral norm. The correct right-hand side is $\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^* = \text{diag}(2, 0)$, of spectral norm 2 — and here the inequality is an equality.

21.11.3 Cauchy–Schwarz

Theorem 21.11.3. Cauchy–Schwarz for Unitarily Invariant Norms

Let $\mathbf{A}, \mathbf{B} \in M_n(F)$. Then for every unitarily invariant norm on $M_n(F)$,

$$\|\mathbf{A}^*\mathbf{B}\|^2 \leq \|\mathbf{A}^*\mathbf{A}\| \|\mathbf{B}^*\mathbf{B}\|.$$

Idea. The scalar inequality $2|\bar{a}b| \leq |a|^2 + |b|^2$ becomes $|\bar{a}b| \leq |a||b|$ by the classical trick of replacing a by ta and b by b/t and choosing t to balance the two terms. The product $\mathbf{A}^*\mathbf{B}$ is unchanged by that substitution, so the same trick upgrades Theorem 21.11.2. The one thing to check is that the two sides of the balance are the right quantities, and that is the identity $\|\mathbf{A}\mathbf{A}^*\| = \|\mathbf{A}^*\mathbf{A}\|$.

Proof. Put $a = \|\mathbf{A}^*\mathbf{A}\|$ and $b = \|\mathbf{B}^*\mathbf{B}\|$. If $a = 0$, then $\mathbf{A}^*\mathbf{A} = \mathbf{O}$ by the positivity clause of Definition 16.1.1, so $\|\mathbf{A}\mathbf{x}\|^2 = \mathbf{x}^*\mathbf{A}^*\mathbf{A}\mathbf{x} = 0$ for every $\mathbf{x}$ and $\mathbf{A} = \mathbf{O}$; then both sides of the asserted inequality are 0. Likewise if $b = 0$. So assume $a > 0$ and $b > 0$.

First, $\|\mathbf{A}\mathbf{A}^*\| = \|\mathbf{A}^*\mathbf{A}\|$. Indeed $\mathbf{A}\mathbf{A}^*$ and $\mathbf{A}^*\mathbf{A}$ are positive semidefinite with the same eigenvalues, namely $\sigma_i(\mathbf{A})^2$ (Definition 13.9.1 and Theorem 9.10.4 (b), the two matrices being of equal size n), hence the same singular values; and a unitarily invariant norm depends only on the singular values (Lemma 21.4.1).

Let $t > 0$ and apply Theorem 21.11.2 to the pair $t\mathbf{A}$ and $t^{-1}\mathbf{B}$. Their product is

$(t\mathbf{A})^*(t^{-1}\mathbf{B}) = \mathbf{A}^*\mathbf{B}$, unchanged, while

$$(t\mathbf{A})(t\mathbf{A})^* + (t^{-1}\mathbf{B})(t^{-1}\mathbf{B})^* = t^2\mathbf{A}\mathbf{A}^* + t^{-2}\mathbf{B}\mathbf{B}^*.$$

So, using the triangle inequality and homogeneity of $\|\cdot\|$ and then the identity just proved,

$$2\|\mathbf{A}^*\mathbf{B}\| \leq t^2\|\mathbf{A}\mathbf{A}^*\| + t^{-2}\|\mathbf{B}\mathbf{B}^*\| = t^2a + t^{-2}b.$$

By the arithmetic–geometric mean inequality for two positive numbers (Lemma 17.8.6), $t^2a + t^{-2}b \geq 2\sqrt{ab}$, with equality when $t^2a = t^{-2}b$; choosing $t = (b/a)^{1/4}$, which is legitimate since $a, b > 0$, makes the two terms equal and gives

$$2\|\mathbf{A}^*\mathbf{B}\| \leq 2\sqrt{ab}.$$

Dividing by 2 and squaring proves the theorem. ■

Remark. Three familiar inequalities are special cases. With $\|\cdot\| = \|\cdot\|_F$ it is $\|\mathbf{A}^*\mathbf{B}\|_F^2 \leq \|\mathbf{A}^*\mathbf{A}\|_F\|\mathbf{B}^*\mathbf{B}\|_F$; with the spectral norm it is $\|\mathbf{A}^*\mathbf{B}\|_2^2 \leq \|\mathbf{A}\|_2^2\|\mathbf{B}\|_2^2$, which also follows from submultiplicativity; and with the trace norm $\|\cdot\|_{(n)} = \sum_i \sigma_i(\cdot)$ it compares $\sum_i \sigma_i(\mathbf{A}^*\mathbf{B})$ with $(\sum_i \sigma_i(\mathbf{A})^2)^{1/2}(\sum_i \sigma_i(\mathbf{B})^2)^{1/2}$. At $n = 1$ all of them are $|\bar{a}b|^2 \leq |a|^2|b|^2$.

21.11.4 The geometric mean of two positive definite matrices

The inequalities above were about *norms* of products. The rest of the section is about an operation: given $\mathbf{A}, \mathbf{B} \succ \mathbf{O}$, produce a positive definite matrix that deserves the name $\sqrt{\mathbf{A}\mathbf{B}}$.

The arithmetic mean gives no trouble: $\frac{1}{2}(\mathbf{A} + \mathbf{B})$ is positive definite, symmetric in the two arguments, and monotone in each. The geometric mean is the problem. The naive candidate is $(\mathbf{A}\mathbf{B})^{1/2}$, and it does not exist: $\mathbf{A}\mathbf{B}$ need not be Hermitian, so Theorem 13.2.1 has nothing to say about it. The next candidate, $\mathbf{A}^{1/2}\mathbf{B}^{1/2}$, is a matrix, but usually not a Hermitian one. What works is to measure $\mathbf{B}$ in the coordinates that make $\mathbf{A}$ the identity, take the square root there, and translate back.

The geometric mean of two positive definite matrices is the square root of the second one, computed after the first has been turned into the identity, and then translated back.

Definition 21.11.1. Geometric Mean of Positive Definite Matrices

Let $\mathbf{A}, \mathbf{B} \in M_n(F)$ with $\mathbf{A} \succ \mathbf{O}$ and $\mathbf{B} \succ \mathbf{O}$. Their **geometric mean** is

$$\mathbf{A} \# \mathbf{B} := \mathbf{A}^{1/2}(\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2})^{1/2}\mathbf{A}^{1/2},$$

where $\mathbf{A}^{1/2}$ is the positive square root of Theorem 13.2.1, $\mathbf{A}^{-1/2} := (\mathbf{A}^{1/2})^{-1}$, and the outer exponent $\frac{1}{2}$ is again the positive square root.

In words: conjugate $\mathbf{B}$ by $\mathbf{A}^{-1/2}$ — this is the **whitening** move of Chapter 13 §05, which

replaces the pair $(\mathbf{A}, \mathbf{B})$ by $(\mathbf{I}, \mathbf{C})$ with $\mathbf{C} = \mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}$ — take the positive square root of $\mathbf{C}$, and undo the conjugation.

Everything in the definition makes sense. $\mathbf{A} \succ \mathbf{O}$ gives $\mathbf{A}^{1/2} \succ \mathbf{O}$ (its eigenvalues are the positive square roots of those of $\mathbf{A}$, by Theorem 13.2.1), so $\mathbf{A}^{1/2}$ is invertible and $\mathbf{A}^{-1/2}$ is Hermitian. Hence $\mathbf{C} = \mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2} = (\mathbf{A}^{-1/2})^*\mathbf{B}\mathbf{A}^{-1/2}$ is a congruence of $\mathbf{B} \succ \mathbf{O}$ by an invertible matrix and is therefore $\succ \mathbf{O}$ (Proposition 13.1.2 (b)); so $\mathbf{C}^{1/2}$ exists, is unique, and is $\succ \mathbf{O}$. Finally $\mathbf{A}\#\mathbf{B} = (\mathbf{A}^{1/2})^*\mathbf{C}^{1/2}\mathbf{A}^{1/2}$ is another such congruence, so $\mathbf{A}\#\mathbf{B} \succ \mathbf{O}$. In particular $\mathbf{A}\#\mathbf{B}$ is Hermitian, which $\mathbf{A}^{1/2}\mathbf{B}^{1/2}$ is not.

Examples.

- **Numbers.** For $n = 1$ with $\mathbf{A} = (a)$, $\mathbf{B} = (b)$, and $a, b > 0$, the formula reads $\sqrt{a}(b/a)^{1/2}\sqrt{a} = \sqrt{ab}$. This is where the name comes from.
- **One argument the identity.** $\mathbf{A}\#\mathbf{I} = \mathbf{A}^{1/2}(\mathbf{A}^{-1})^{1/2}\mathbf{A}^{1/2} = \mathbf{A}^{1/2}\mathbf{A}^{-1/2}\mathbf{A}^{1/2} = \mathbf{A}^{1/2}$, using $(\mathbf{A}^{-1})^{1/2} = \mathbf{A}^{-1/2}$, which holds because $(\mathbf{A}^{-1/2})^2 = \mathbf{A}^{-1}$ and $\mathbf{A}^{-1/2} \succ \mathbf{O}$, so uniqueness in Theorem 13.2.1 applies. Likewise $\mathbf{A}\#\mathbf{A} = \mathbf{A}^{1/2}\mathbf{I}\mathbf{A}^{1/2} = \mathbf{A}$.
- **Commuting arguments.** If $\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A}$ then $\mathbf{A}\#\mathbf{B} = (\mathbf{A}\mathbf{B})^{1/2}$. Indeed $\mathbf{A}^{1/2}$ commutes with $\mathbf{B}$ (Theorem 13.2.1 (b)), hence so does $\mathbf{A}^{-1/2}$, so $\mathbf{C} = \mathbf{B}\mathbf{A}^{-1}$ commutes with $\mathbf{A}$, and $\mathbf{C}^{1/2}$, being a polynomial in $\mathbf{C}$ (Theorem 13.2.1 (a)), commutes with $\mathbf{A}$ too. Therefore

$$\begin{aligned} (\mathbf{A}\#\mathbf{B})^2 &= \mathbf{A}^{1/2}\mathbf{C}^{1/2}\mathbf{A}\mathbf{C}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}^{1/2}\mathbf{A}\mathbf{C}\mathbf{A}^{1/2} \\ &= \mathbf{A}^{1/2}\mathbf{A}\mathbf{B}\mathbf{A}^{-1}\mathbf{A}^{1/2} = \mathbf{A}^{1/2}\mathbf{B}\mathbf{A}^{1/2} = \mathbf{A}\mathbf{B}, \end{aligned}$$

using $\mathbf{C} = \mathbf{B}\mathbf{A}^{-1}$ and $\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A}$ once more; and $\mathbf{A}\#\mathbf{B} \succ \mathbf{O}$, so it is the positive square root of $\mathbf{A}\mathbf{B}$. For instance $\text{diag}(4, 1)\#\text{diag}(1, 9) = \text{diag}(2, 3)$.

- **A noncommuting pair, exactly.** Let

$$\mathbf{A} = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 8 & 10 \\ 10 & 17 \end{pmatrix}.$$

Then $\mathbf{A}^{1/2} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, and $\mathbf{A}^{-1/2} = \frac{1}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$, so $\mathbf{C} = \mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2} = \text{diag}(1, 4)$ and $\mathbf{C}^{1/2} = \text{diag}(1, 2)$. Hence

$$\mathbf{A}\#\mathbf{B} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 6 & 6 \\ 6 & 9 \end{pmatrix}.$$

Here $\mathbf{A}\mathbf{B} \neq \mathbf{B}\mathbf{A}$, as the (1, 2) entries 118 and 82 show. Note $\det(\mathbf{A}\#\mathbf{B}) = 18 = \sqrt{9 \cdot 36} = \sqrt{\det \mathbf{A} \det \mathbf{B}}$; Exercise C1 explains why.

A nonexample by minimal change. Drop the middle square root and take $\mathbf{A}^{1/2}\mathbf{B}^{1/2}$. With $\mathbf{A} = \text{diag}(4, 1)$ and $\mathbf{B} = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}$, the two square roots are $\text{diag}(2, 1)$ and $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, and

$$\mathbf{A}^{1/2}\mathbf{B}^{1/2} = \begin{pmatrix} 4 & 2 \\ 1 & 2 \end{pmatrix},$$

which is not Hermitian, so it is not $\succ \mathbf{O}$ and is not a mean of anything. Its square is $\begin{pmatrix} 18 & 12 \\ 6 & 6 \end{pmatrix}$, not $\mathbf{A}\mathbf{B} = \begin{pmatrix} 20 & 16 \\ 4 & 5 \end{pmatrix}$.

Warning. **Half of $\mathbf{A}$ times half of $\mathbf{B}$ is not a mean.** $\mathbf{A}^{1/2}\mathbf{B}^{1/2}$ is Hermitian only when $\mathbf{A}$ and $\mathbf{B}$ commute: $(\mathbf{A}^{1/2}\mathbf{B}^{1/2})^* = \mathbf{B}^{1/2}\mathbf{A}^{1/2}$, and if this equals $\mathbf{A}^{1/2}\mathbf{B}^{1/2}$, then squaring gives $\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A}$. The sandwich in Definition 21.11.1 exists precisely to repair this, and the price is that the formula no longer *looks* symmetric in $\mathbf{A}$ and $\mathbf{B}$. It is; that is the content of

Theorem 21.11.4 (b), and it is the one property here that needs an argument.

Quick check. Compute $\mathbf{A}\#\mathbf{B}$ for $\mathbf{A} = \text{diag}(4, 1)$ and $\mathbf{B} = \text{diag}(1, 9)$ straight from [Definition 21.11.1](#), and check the answer against $(\mathbf{A}\mathbf{B})^{1/2}$.

Solution. $\mathbf{A}^{1/2} = \text{diag}(2, 1)$ and $\mathbf{A}^{-1/2} = \text{diag}(\frac{1}{2}, 1)$, so $\mathbf{C} = \mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2} = \text{diag}(\frac{1}{4}, 9)$ and $\mathbf{C}^{1/2} = \text{diag}(\frac{1}{2}, 3)$. Hence

$$\mathbf{A}\#\mathbf{B} = \text{diag}(2, 1) \text{diag}(\frac{1}{2}, 3) \text{diag}(2, 1) = \text{diag}(2, 3).$$

The two matrices commute, and $\mathbf{A}\mathbf{B} = \text{diag}(4, 9)$, whose positive square root is $\text{diag}(2, 3)$. The two agree.

21.11.5 What the geometric mean does

The key is a description of $\mathbf{A}\#\mathbf{B}$ that never mentions $\mathbf{A}^{1/2}$. Chapter 13 §06's block test supplies it.

Theorem 21.11.4. Properties of the Geometric Mean

Let $\mathbf{A}, \mathbf{B}, \mathbf{A}', \mathbf{B}' \in M_n(F)$ be positive definite and let $\mathbf{S} \in M_n(F)$ be invertible.

(a) **Extremal characterization.** $\mathbf{A}\#\mathbf{B}$ is the largest Hermitian $\mathbf{X} \in M_n(F)$ for which

$$\begin{pmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X} & \mathbf{B} \end{pmatrix} \succeq \mathbf{O} :$$

this holds for $\mathbf{X} = \mathbf{A}\#\mathbf{B}$, and every Hermitian $\mathbf{X}$ for which it holds satisfies $\mathbf{X} \preceq \mathbf{A}\#\mathbf{B}$.

(b) **Symmetry.** $\mathbf{A}\#\mathbf{B} = \mathbf{B}\#\mathbf{A}$.

(c) **Congruence.** $(\mathbf{S}^*\mathbf{A}\mathbf{S})\#(\mathbf{S}^*\mathbf{B}\mathbf{S}) = \mathbf{S}^*(\mathbf{A}\#\mathbf{B})\mathbf{S}$.

(d) **Monotonicity.** If $\mathbf{A} \succeq \mathbf{A}'$ and $\mathbf{B} \succeq \mathbf{B}'$, then $\mathbf{A}\#\mathbf{B} \succeq \mathbf{A}'\#\mathbf{B}'$.

(e) **Arithmetic–geometric mean, and inversion.** $\mathbf{A}\#\mathbf{B} \preceq \frac{1}{2}(\mathbf{A} + \mathbf{B})$, and $(\mathbf{A}\#\mathbf{B})^{-1} = \mathbf{A}^{-1}\#\mathbf{B}^{-1}$.

Idea. Everything rests on (a), and (a) rests on whitening. With $\mathbf{A} \succ \mathbf{O}$ the block test Theorem 13.6.1 turns the block condition into $\mathbf{B} - \mathbf{X}\mathbf{A}^{-1}\mathbf{X} \succeq \mathbf{O}$; substituting $\mathbf{X} = \mathbf{A}^{1/2}\mathbf{Y}\mathbf{A}^{1/2}$ and conjugating by $\mathbf{A}^{-1/2}$ turns that into $\mathbf{Y}^2 \preceq \mathbf{C}$, with $\mathbf{C} = \mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}$. The largest Hermitian $\mathbf{Y}$ with $\mathbf{Y}^2 \preceq \mathbf{C}$ is $\mathbf{C}^{1/2}$, because $\mathbf{Y} \preceq |\mathbf{Y}| = (\mathbf{Y}^2)^{1/2}$ and the square root is operator monotone. Once (a) is available, (b), (c), and (d) are all statements about the set of admissible $\mathbf{X}$, and they read off the symmetry, the congruence-invariance, and the monotonicity of that set without any further computation.

Proof. Write $\mathbf{C} = \mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2} \succ \mathbf{O}$, so that $\mathbf{A}\#\mathbf{B} = \mathbf{A}^{1/2}\mathbf{C}^{1/2}\mathbf{A}^{1/2}$, and for Hermitian $\mathbf{X}$ put

$$\mathbf{M}(\mathbf{X}) = \begin{pmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X} & \mathbf{B} \end{pmatrix},$$

which is Hermitian because $\mathbf{X}^* = \mathbf{X}$. Call $\mathbf{X}$ **admissible** for $(\mathbf{A}, \mathbf{B})$ when $\mathbf{M}(\mathbf{X}) \succeq \mathbf{O}$.

(a). Since $\mathbf{A} \succ \mathbf{O}$, Theorem 13.6.1 (b) says that $\mathbf{X}$ is admissible exactly when the Schur complement $\mathbf{B} - \mathbf{X}\mathbf{A}^{-1}\mathbf{X}$ is $\succeq \mathbf{O}$. Every Hermitian $\mathbf{X}$ is uniquely $\mathbf{X} = \mathbf{A}^{1/2}\mathbf{Y}\mathbf{A}^{1/2}$ with $\mathbf{Y} = \mathbf{A}^{-1/2}\mathbf{X}\mathbf{A}^{-1/2}$ Hermitian, and then

$$\mathbf{X}\mathbf{A}^{-1}\mathbf{X} = \mathbf{A}^{1/2}\mathbf{Y}\mathbf{A}^{1/2}\mathbf{A}^{-1}\mathbf{A}^{1/2}\mathbf{Y}\mathbf{A}^{1/2} = \mathbf{A}^{1/2}\mathbf{Y}^2\mathbf{A}^{1/2}.$$

Conjugating by the invertible Hermitian $\mathbf{A}^{-1/2}$ and using Proposition 13.1.2 (a) in both directions,

$$\mathbf{B} - \mathbf{A}^{1/2}\mathbf{Y}^2\mathbf{A}^{1/2} \succeq \mathbf{O} \iff \mathbf{C} - \mathbf{Y}^2 \succeq \mathbf{O}.$$

Claim. If $\mathbf{Y}$ is Hermitian and $\mathbf{Y}^2 \preceq \mathbf{C}$, then $\mathbf{Y} \preceq \mathbf{C}^{1/2}$.

Proof. Let $\mathbf{Y} = \sum_j \mu_j \mathbf{P}_j$ be the spectral resolution (Theorem 12.8.1), so that $\mathbf{Y}^2 = \sum_j \mu_j^2 \mathbf{P}_j$ and the matrix $\sum_j |\mu_j| \mathbf{P}_j$ is $\succeq \mathbf{O}$ with square $\mathbf{Y}^2$; by uniqueness in Theorem 13.2.1 it is $(\mathbf{Y}^2)^{1/2}$. Then

$$(\mathbf{Y}^2)^{1/2} - \mathbf{Y} = \sum_j (|\mu_j| - \mu_j) \mathbf{P}_j \succeq \mathbf{O},$$

because $\mathbf{x}^* \mathbf{P}_j \mathbf{x} = \|\mathbf{P}_j \mathbf{x}\|^2 \geq 0$ and every coefficient is ≥ 0 . So $\mathbf{Y} \preceq (\mathbf{Y}^2)^{1/2}$. Also $\mathbf{Y}^2 \preceq \mathbf{C}$ with $\mathbf{Y}^2 \succeq \mathbf{O}$, so Proposition 13.5.4 gives $(\mathbf{Y}^2)^{1/2} \preceq \mathbf{C}^{1/2}$. Chaining the two (Proposition 13.5.1) gives $\mathbf{Y} \preceq \mathbf{C}^{1/2}$. ■

The choice $\mathbf{Y} = \mathbf{C}^{1/2}$ itself satisfies $\mathbf{C} - \mathbf{Y}^2 = \mathbf{O} \succeq \mathbf{O}$. So among the Hermitian $\mathbf{Y}$ with $\mathbf{C} - \mathbf{Y}^2 \succeq \mathbf{O}$ the largest is $\mathbf{C}^{1/2}$. Translating back by the order-preserving bijection $\mathbf{Y} \mapsto \mathbf{A}^{1/2}\mathbf{Y}\mathbf{A}^{1/2}$ (again Proposition 13.1.2 (a), in both directions), the largest admissible $\mathbf{X}$ is $\mathbf{A}^{1/2}\mathbf{C}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}\#\mathbf{B}$.

(b). Let $\mathbf{W} = \begin{pmatrix} \mathbf{O} & \mathbf{I} \\ \mathbf{I} & \mathbf{O} \end{pmatrix}$, which is Hermitian and unitary, hence invertible. By Theorem 8.1.1,

$$\mathbf{W}^* \begin{pmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X} & \mathbf{B} \end{pmatrix} \mathbf{W} = \begin{pmatrix} \mathbf{B} & \mathbf{X} \\ \mathbf{X} & \mathbf{A} \end{pmatrix},$$

so by Proposition 13.1.2 (a), applied to $\mathbf{W}$ and to $\mathbf{W}^{-1}$, a Hermitian $\mathbf{X}$ is admissible for $(\mathbf{A}, \mathbf{B})$ if and only if it is admissible for $(\mathbf{B}, \mathbf{A})$. The two admissible sets coincide, so they have the same largest element; by (a) that element is $\mathbf{A}\#\mathbf{B}$ on one side and $\mathbf{B}\#\mathbf{A}$ on the other, and a partial order has at most one largest element (Proposition 13.5.1).

(c). Put $\mathbf{T} = \mathbf{S} \oplus \mathbf{S}$. By Theorem 8.1.1,

$$\mathbf{T}^* \begin{pmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X} & \mathbf{B} \end{pmatrix} \mathbf{T} = \begin{pmatrix} \mathbf{S}^* \mathbf{A} \mathbf{S} & \mathbf{S}^* \mathbf{X} \mathbf{S} \\ \mathbf{S}^* \mathbf{X} \mathbf{S} & \mathbf{S}^* \mathbf{B} \mathbf{S} \end{pmatrix},$$

and $\mathbf{T}$ is invertible, so $\mathbf{X}$ is admissible for $(\mathbf{A}, \mathbf{B})$ exactly when $\mathbf{S}^* \mathbf{X} \mathbf{S}$ is admissible for $(\mathbf{S}^* \mathbf{A} \mathbf{S}, \mathbf{S}^* \mathbf{B} \mathbf{S})$ (Proposition 13.1.2 (a), both ways). The map $\mathbf{X} \mapsto \mathbf{S}^* \mathbf{X} \mathbf{S}$ is a bijection of the Hermitian matrices that preserves $\preceq$ in both directions, so it carries the largest element

of one admissible set to the largest element of the other. Both $\mathbf{S}^*\mathbf{A}\mathbf{S}$ and $\mathbf{S}^*\mathbf{B}\mathbf{S}$ are $\succ \mathbf{O}$ by Proposition 13.1.2 (b), so (a) applies to them, and the claim follows.

(d). Let $\mathbf{X}$ be admissible for $(\mathbf{A}', \mathbf{B}')$. Then

$$\begin{pmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X} & \mathbf{B} \end{pmatrix} = \begin{pmatrix} \mathbf{A}' & \mathbf{X} \\ \mathbf{X} & \mathbf{B}' \end{pmatrix} + \begin{pmatrix} \mathbf{A} - \mathbf{A}' & \mathbf{O} \\ \mathbf{O} & \mathbf{B} - \mathbf{B}' \end{pmatrix},$$

a sum of two positive semidefinite matrices — the second because its quadratic form is $\mathbf{y}_1^*(\mathbf{A} - \mathbf{A}')\mathbf{y}_1 + \mathbf{y}_2^*(\mathbf{B} - \mathbf{B}')\mathbf{y}_2 \geq 0$. So $\mathbf{X}$ is admissible for $(\mathbf{A}, \mathbf{B})$. Taking $\mathbf{X} = \mathbf{A}'\#\mathbf{B}'$, which is admissible for $(\mathbf{A}', \mathbf{B}')$ by (a), gives $\mathbf{A}'\#\mathbf{B}' \preceq \mathbf{A}\#\mathbf{B}$ by the maximality in (a).

(e). Since $\mathbf{C}^{1/2}$ is Hermitian, $(\mathbf{I} - \mathbf{C}^{1/2})^2 \succeq \mathbf{O}$ by Theorem 13.1.4 (c), and $(\mathbf{C}^{1/2})^2 = \mathbf{C}$ gives

$$\frac{1}{2}(\mathbf{I} + \mathbf{C}) - \mathbf{C}^{1/2} = \frac{1}{2}(\mathbf{I} - \mathbf{C}^{1/2})^2 \succeq \mathbf{O}.$$

Conjugating by $\mathbf{A}^{1/2}$ and using $\mathbf{A}^{1/2}\mathbf{C}\mathbf{A}^{1/2} = \mathbf{B}$,

$$\frac{1}{2}(\mathbf{A} + \mathbf{B}) - \mathbf{A}\#\mathbf{B} = \frac{1}{2}\mathbf{A}^{1/2}(\mathbf{I} - \mathbf{C}^{1/2})^2\mathbf{A}^{1/2} \succeq \mathbf{O},$$

by Proposition 13.1.2 (a). For the inversion formula, $\mathbf{A}\#\mathbf{B} = \mathbf{A}^{1/2}\mathbf{C}^{1/2}\mathbf{A}^{1/2}$ is a product of invertible matrices, with inverse $\mathbf{A}^{-1/2}\mathbf{C}^{-1/2}\mathbf{A}^{-1/2}$, where $\mathbf{C}^{-1/2} = (\mathbf{C}^{1/2})^{-1}$. On the other side, using $(\mathbf{A}^{-1})^{1/2} = \mathbf{A}^{-1/2}$ and $(\mathbf{A}^{-1})^{-1/2} = \mathbf{A}^{1/2}$,

$$\mathbf{A}^{-1}\#\mathbf{B}^{-1} = \mathbf{A}^{-1/2}(\mathbf{A}^{1/2}\mathbf{B}^{-1}\mathbf{A}^{1/2})^{1/2}\mathbf{A}^{-1/2},$$

and $\mathbf{A}^{1/2}\mathbf{B}^{-1}\mathbf{A}^{1/2} = (\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2})^{-1} = \mathbf{C}^{-1}$, whose positive square root is $\mathbf{C}^{-1/2}$, since $(\mathbf{C}^{-1/2})^2 = \mathbf{C}^{-1}$ and $\mathbf{C}^{-1/2} \succ \mathbf{O}$. So $\mathbf{A}^{-1}\#\mathbf{B}^{-1} = \mathbf{A}^{-1/2}\mathbf{C}^{-1/2}\mathbf{A}^{-1/2} = (\mathbf{A}\#\mathbf{B})^{-1}$. ■

Part (a) is the statement to remember, because it is the one with no square roots in it: $\mathbf{A}\#\mathbf{B}$ is the best Hermitian matrix that can sit in the off-diagonal corner above $\mathbf{A}$ and $\mathbf{B}$. Read beside [Lemma 21.11.1](#), which says that such a corner cannot be too large in singular value, part (a) says exactly how large it can be in the Loewner order.

Quick check. Take $\mathbf{B} = \mathbf{A}$. Part (a) then says that $\mathbf{A}$ is the largest Hermitian $\mathbf{X}$ with $\begin{pmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X} & \mathbf{A} \end{pmatrix} \succeq \mathbf{O}$. Check by hand that $\mathbf{X} = \mathbf{A}$ is admissible and that $\mathbf{X} = 2\mathbf{A}$ is not.

Solution. For $\mathbf{X} = \mathbf{A}$ the quadratic form at $(\mathbf{y}_1, \mathbf{y}_2)$ is

$$\mathbf{y}_1^*\mathbf{A}\mathbf{y}_1 + \mathbf{y}_1^*\mathbf{A}\mathbf{y}_2 + \mathbf{y}_2^*\mathbf{A}\mathbf{y}_1 + \mathbf{y}_2^*\mathbf{A}\mathbf{y}_2 = (\mathbf{y}_1 + \mathbf{y}_2)^*\mathbf{A}(\mathbf{y}_1 + \mathbf{y}_2) \geq 0,$$

so $\mathbf{X} = \mathbf{A}$ is admissible — as it must be, since $\mathbf{A}\#\mathbf{A} = \mathbf{A}$ by the second example after [Definition 21.11.1](#). For $\mathbf{X} = 2\mathbf{A}$ take $\mathbf{y}_2 = -\mathbf{y}_1$: the four terms give $\mathbf{y}_1^*\mathbf{A}\mathbf{y}_1 - 4\mathbf{y}_1^*\mathbf{A}\mathbf{y}_1 + \mathbf{y}_1^*\mathbf{A}\mathbf{y}_1 = -2\mathbf{y}_1^*\mathbf{A}\mathbf{y}_1$, which is < 0 for $\mathbf{y}_1 \neq \mathbf{0}$ because $\mathbf{A} \succ \mathbf{O}$. So $2\mathbf{A}$ is not admissible.

21.11.6 The chain of three means

Beside the arithmetic mean $\frac{1}{2}(\mathbf{A} + \mathbf{B})$ and the geometric mean $\mathbf{A} \# \mathbf{B}$ sits a third. The **harmonic mean** of two positive definite matrices is the inverse of the arithmetic mean of their inverses,

$$\left(\frac{\mathbf{A}^{-1} + \mathbf{B}^{-1}}{2} \right)^{-1},$$

which at $n = 1$ is the familiar $2ab/(a + b)$. For two positive numbers the three are ordered harmonic $\leq$ geometric $\leq$ arithmetic, and the same order holds for matrices.

Corollary 21.11.5. Harmonic, Geometric, and Arithmetic

Let $\mathbf{A}, \mathbf{B} \in M_n(F)$ be positive definite. Then

$$\left(\frac{\mathbf{A}^{-1} + \mathbf{B}^{-1}}{2} \right)^{-1} \preceq \mathbf{A} \# \mathbf{B} \preceq \frac{\mathbf{A} + \mathbf{B}}{2}.$$

Proof. The right inequality is [Theorem 21.11.4](#) (e). For the left, apply that same part to the positive definite matrices $\mathbf{A}^{-1}$ and $\mathbf{B}^{-1}$:

$$\mathbf{A}^{-1} \# \mathbf{B}^{-1} \preceq \frac{1}{2}(\mathbf{A}^{-1} + \mathbf{B}^{-1}).$$

Both sides are positive definite — the left by the remark after [Definition 21.11.1](#), the right as a positive combination of positive definite matrices. Since $t \mapsto -1/t$ is operator monotone on $(0, \infty)$ ([Proposition 17.11.3](#)), applying it reverses the inequality between the inverses:

$$\left(\frac{1}{2}(\mathbf{A}^{-1} + \mathbf{B}^{-1}) \right)^{-1} \preceq (\mathbf{A}^{-1} \# \mathbf{B}^{-1})^{-1}.$$

By [Theorem 21.11.4](#) (e) the right side is $\mathbf{A} \# \mathbf{B}$. ■

The three means agree when $\mathbf{A} = \mathbf{B}$, and for the pair of the worked example they are

$$\frac{1}{5} \begin{pmatrix} 28 & 26 \\ 26 & 37 \end{pmatrix} \preceq \begin{pmatrix} 6 & 6 \\ 6 & 9 \end{pmatrix} \preceq \begin{pmatrix} \frac{13}{2} & 7 \\ 7 & 11 \end{pmatrix},$$

the two differences being $\frac{2}{5} \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ and $\frac{1}{2} \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$, each a nonnegative multiple of a rank-one positive semidefinite matrix.

Remark. The geometric mean is the first genuinely noncommutative *average* in the book, and it is built out of two order facts proved long before this chapter: the operator monotonicity of $t \mapsto t^{1/2}$, which is Chapter 13 §05's [Proposition 13.5.4](#), used in the claim inside [Theorem 21.11.4](#), and the operator monotonicity of $t \mapsto -1/t$, which is Chapter 17 §11's [Proposition 17.11.3](#), used in [Corollary 21.11.5](#). Without the first, the extremal characterization would fail; without the second, the chain would have no left-hand end.

21.11.7 Exercises

21.11.7.1 A. Check your understanding

Exercise. A1

- (a) Define $\mathbf{A}\#\mathbf{B}$ for positive definite $\mathbf{A}, \mathbf{B}$, and say why each square root in the formula exists.
- (b) State the extremal characterization of $\mathbf{A}\#\mathbf{B}$, and explain in one sentence how it makes the symmetry $\mathbf{A}\#\mathbf{B} = \mathbf{B}\#\mathbf{A}$ obvious.
- (c) Decide whether the following is correct, and justify your answer: “ $\mathbf{A}^{1/2}\mathbf{B}^{1/2}$ is positive definite whenever $\mathbf{A}$ and $\mathbf{B}$ are”.
- (d) State the matrix Cauchy–Schwarz inequality and the arithmetic–geometric mean inequality for singular values, [Theorem 21.11.2](#) — not the one for the geometric mean $\mathbf{A}\#\mathbf{B}$ — and say which one implies the other and how.
- (e) Which of $\mathbf{A}^*\mathbf{A} + \mathbf{B}^*\mathbf{B}$ and $\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*$ bounds $2\|\mathbf{A}^*\mathbf{B}\|$, and what goes wrong with the other?

Solution.

- (a) $\mathbf{A}\#\mathbf{B} = \mathbf{A}^{1/2}(\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2})^{1/2}\mathbf{A}^{1/2}$ ([Definition 21.11.1](#)). The outer square root exists because $\mathbf{A} \succ \mathbf{O}$ ([Theorem 13.2.1](#)), and $\mathbf{A}^{1/2} \succ \mathbf{O}$ is invertible; the inner one because $\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}$ is a congruence of $\mathbf{B} \succ \mathbf{O}$ by an invertible matrix, hence $\succ \mathbf{O}$ ([Proposition 13.1.2 \(b\)](#)).
- (b) It is the largest Hermitian $\mathbf{X}$ with $\begin{pmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X}^* & \mathbf{B} \end{pmatrix} \succeq \mathbf{O}$ ([Theorem 21.11.4 \(a\)](#)). Swapping the two diagonal blocks is a congruence by the unitary $\begin{pmatrix} \mathbf{O} & \mathbf{I} \\ \mathbf{I} & \mathbf{O} \end{pmatrix}$, which changes neither the condition nor $\mathbf{X}$, so $(\mathbf{A}, \mathbf{B})$ and $(\mathbf{B}, \mathbf{A})$ have the same admissible set and therefore the same largest element.
- (c) Incorrect. It need not even be Hermitian: with $\mathbf{A} = \text{diag}(4, 1)$ and $\mathbf{B} = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}$, $\mathbf{A}^{1/2}\mathbf{B}^{1/2} = \begin{pmatrix} 4 & 2 \\ 1 & 2 \end{pmatrix}$. It is Hermitian exactly when $\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A}$.
- (d) Cauchy–Schwarz: $\|\mathbf{A}^*\mathbf{B}\|^2 \leq \|\mathbf{A}^*\mathbf{A}\| \|\mathbf{B}^*\mathbf{B}\|$ ([Theorem 21.11.3](#)). Arithmetic–geometric mean for singular values: $2\|\mathbf{A}^*\mathbf{B}\| \leq \|\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*\|$ ([Theorem 21.11.2](#)). The second implies the first: replace $\mathbf{A}$ by $t\mathbf{A}$ and $\mathbf{B}$ by $t^{-1}\mathbf{B}$, which leaves $\mathbf{A}^*\mathbf{B}$ alone, use the triangle inequality, and pick t to balance the two terms.
- (e) $\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*$ does ([Theorem 21.11.2](#)). With $\mathbf{A}^*\mathbf{A} + \mathbf{B}^*\mathbf{B}$ the statement is false: the warning after that theorem gives $\mathbf{A}, \mathbf{B}$ with $\mathbf{A}^*\mathbf{A} + \mathbf{B}^*\mathbf{B} = \mathbf{I}$ of spectral norm 1 and $2\|\mathbf{A}^*\mathbf{B}\|_2 = 2$.

21.11.7.2 B. Practice

Exercise. B1: A geometric mean, exactly

Let

$$\mathbf{A} = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 13 & 20 \\ 20 & 37 \end{pmatrix}.$$

- (a) Verify that $\mathbf{A}^{1/2} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ and compute $\mathbf{C} = \mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}$ and $\mathbf{A}\#\mathbf{B}$.
- (b) Check the arithmetic–geometric mean inequality $\mathbf{A}\#\mathbf{B} \preceq \frac{1}{2}(\mathbf{A} + \mathbf{B})$ directly, by com-

putting the difference and its eigenvalues.

- (c) Verify the extremal characterization for your answer by computing the Schur complement $\mathbf{B} - \mathbf{X}\mathbf{A}^{-1}\mathbf{X}$ with $\mathbf{X} = \mathbf{A}\#\mathbf{B}$.

Solution.

- (a) $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}^2 = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix} = \mathbf{A}$, and $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ has eigenvalues 3 and 1, both positive, so it is $\succ \mathbf{O}$ and is the positive square root (Theorem 13.2.1). Its determinant is 3, so $\mathbf{A}^{-1/2} = \frac{1}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$. Then

$$\mathbf{A}^{-1/2}\mathbf{B} = \frac{1}{3} \begin{pmatrix} 6 & 3 \\ 27 & 54 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 9 & 18 \end{pmatrix},$$

and multiplying on the right by $\mathbf{A}^{-1/2}$ gives $\mathbf{C} = \frac{1}{3} \begin{pmatrix} 3 & 0 \\ 0 & 27 \end{pmatrix} = \text{diag}(1, 9)$. Hence $\mathbf{C}^{1/2} = \text{diag}(1, 3)$ and

$$\mathbf{A}\#\mathbf{B} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 7 & 8 \\ 8 & 13 \end{pmatrix}.$$

- (b) $\frac{1}{2}(\mathbf{A} + \mathbf{B}) = \begin{pmatrix} 9 & 12 \\ 12 & 21 \end{pmatrix}$, so the difference is

$$\frac{1}{2}(\mathbf{A} + \mathbf{B}) - \mathbf{A}\#\mathbf{B} = \begin{pmatrix} 2 & 4 \\ 4 & 8 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \end{pmatrix},$$

of trace 10 and determinant $16 - 16 = 0$, so its eigenvalues are 10 and 0 and it is $\succeq \mathbf{O}$ (Theorem 13.1.4 (b)). The inequality holds, with the difference singular.

- (c) $\det \mathbf{A} = 9$, so $\mathbf{A}^{-1} = \frac{1}{9} \begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix}$. With $\mathbf{X} = \begin{pmatrix} 7 & 8 \\ 8 & 13 \end{pmatrix}$,

$$\mathbf{X}\mathbf{A}^{-1} = \frac{1}{9} \begin{pmatrix} 3 & 12 \\ -12 & 33 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 1 & 4 \\ -4 & 11 \end{pmatrix},$$

and $\mathbf{X}\mathbf{A}^{-1}\mathbf{X} = \frac{1}{3} \begin{pmatrix} 39 & 60 \\ 60 & 111 \end{pmatrix} = \begin{pmatrix} 13 & 20 \\ 20 & 37 \end{pmatrix} = \mathbf{B}$. So the Schur complement is $\mathbf{O}$, which is $\succeq \mathbf{O}$: $\mathbf{X}$ is admissible, and the block matrix is singular, as it must be at the largest admissible $\mathbf{X}$.

Exercise. B2: The two inequalities on one pair

Let $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ in $M_2(\mathbb{R})$.

- (a) Compute $\mathbf{A}^*\mathbf{B}$ and $\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*$, and verify $2\sigma_j(\mathbf{A}^*\mathbf{B}) \leq \lambda_j(\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*)$ for $j = 1, 2$ using traces and determinants.
 (b) Verify Theorem 21.11.3 for this pair in the Frobenius norm.

Solution.

- (a) $\mathbf{A}^* = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, so $\mathbf{A}^*\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$, and $\mathbf{A}\mathbf{A}^* = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, $\mathbf{B}\mathbf{B}^* = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$, so $\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^* = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}$, with eigenvalues 5 and 1.

The singular values of $\mathbf{M} = \mathbf{A}^*\mathbf{B}$ come from $\mathbf{M}^*\mathbf{M} = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$, of trace 6 and determinant 1, so $\sigma_1^2, \sigma_2^2 = 3 \pm 2\sqrt{2} = (\sqrt{2} \pm 1)^2$ and $\sigma_1 = \sqrt{2} + 1$, $\sigma_2 = \sqrt{2} - 1$. Then $2\sigma_1 = 2\sqrt{2} + 2 \approx 4.83 \leq 5$ and $2\sigma_2 = 2\sqrt{2} - 2 \approx 0.83 \leq 1$. Both hold; the first is $2\sqrt{2} \leq 3$, that is $8 \leq 9$, and the second is $2\sqrt{2} \leq 3$ again.

- (b) $\mathbf{A}^*\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ and $\mathbf{B}^*\mathbf{B} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, each of Frobenius norm $\sqrt{1+1+1+4} = \sqrt{7}$. And $\|\mathbf{A}^*\mathbf{B}\|_F^2 = 1+4+1 = 6$. So the inequality reads $6 \leq \sqrt{7} \cdot \sqrt{7} = 7$, which holds.

Exercise. B3: Means of commuting matrices

Let $\mathbf{A} = \text{diag}(a_1, \dots, a_n)$ and $\mathbf{B} = \text{diag}(b_1, \dots, b_n)$ with all $a_i, b_i > 0$. Compute the three means of [Corollary 21.11.5](#) and deduce the classical chain of inequalities between the harmonic, geometric, and arithmetic means of two positive numbers.

Solution. The matrices commute, so $\mathbf{A}\#\mathbf{B} = (\mathbf{A}\mathbf{B})^{1/2} = \text{diag}(\sqrt{a_i b_i})$ by the third example after [Definition 21.11.1](#). Also $\frac{1}{2}(\mathbf{A} + \mathbf{B}) = \text{diag}(\frac{a_i + b_i}{2})$ and

$$\left(\frac{\mathbf{A}^{-1} + \mathbf{B}^{-1}}{2}\right)^{-1} = \text{diag}\left(\frac{2a_i b_i}{a_i + b_i}\right).$$

A diagonal matrix is $\succeq \mathbf{O}$ exactly when its entries are ≥ 0 , so [Corollary 21.11.5](#) says precisely

$$\frac{2a_i b_i}{a_i + b_i} \leq \sqrt{a_i b_i} \leq \frac{a_i + b_i}{2}$$

for every i , the classical chain for the two positive numbers a_i, b_i . The right inequality is Lemma 17.8.6 with two terms, and the left is the right one applied to $1/a_i$ and $1/b_i$ and inverted.

21.11.7.3 C. Going deeper

Exercise. C1: The determinant of a geometric mean

Let $\mathbf{A}, \mathbf{B} \in M_n(F)$ be positive definite.

- Prove that $\det(\mathbf{A}\#\mathbf{B}) = \sqrt{\det \mathbf{A} \det \mathbf{B}}$.
- Deduce from [Corollary 21.11.5](#) that $\sqrt{\det \mathbf{A} \det \mathbf{B}} \leq \det(\frac{1}{2}(\mathbf{A} + \mathbf{B}))$.

Hint for (b): the Loewner order compares determinants of positive definite matrices.

Solution.

- Write $\mathbf{C} = \mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}$. Taking determinants in [Definition 21.11.1](#) and using multiplicativity (Theorem 7.5.4),

$$\det(\mathbf{A}\#\mathbf{B}) = \det(\mathbf{A}^{1/2})^2 \det(\mathbf{C}^{1/2}) = \det \mathbf{A} \cdot \det(\mathbf{C}^{1/2}).$$

Now $(\mathbf{C}^{1/2})^2 = \mathbf{C}$ gives $\det(\mathbf{C}^{1/2})^2 = \det \mathbf{C}$, and $\det(\mathbf{C}^{1/2}) > 0$ because $\mathbf{C}^{1/2} \succ \mathbf{O}$ has positive eigenvalues, so $\det(\mathbf{C}^{1/2}) = \sqrt{\det \mathbf{C}}$. Finally $\det \mathbf{C} = \det(\mathbf{A}^{-1/2})^2 \det \mathbf{B} = \det \mathbf{B} / \det \mathbf{A}$. Hence

$$\det(\mathbf{A}\#\mathbf{B}) = \det \mathbf{A} \cdot \sqrt{\det \mathbf{B} / \det \mathbf{A}} = \sqrt{\det \mathbf{A} \det \mathbf{B}}.$$

- By [Corollary 21.11.5](#), $\mathbf{A}\#\mathbf{B} \preceq \frac{1}{2}(\mathbf{A} + \mathbf{B})$, and both are positive definite. For positive

definite $\mathbf{X} \preceq \mathbf{Y}$ we have $\det \mathbf{X} \leq \det \mathbf{Y}$: by Corollary 17.2.4, $\lambda_i(\mathbf{Y}) \geq \lambda_i(\mathbf{X}) > 0$ for every i , and the determinant of a Hermitian matrix is the product of its eigenvalues with multiplicity (Theorem 9.3.3, whose hypothesis holds because the characteristic polynomial of a Hermitian matrix splits over $\mathbb{R}$ by Theorem 12.2.1), a product of positive numbers increasing in each factor. Applying this and part (a) gives the inequality. At $n = 1$ it is again the classical arithmetic–geometric mean inequality.

Exercise. C2: The geometric mean solves an equation

Let $\mathbf{A}, \mathbf{B} \in M_n(F)$ be positive definite.

- (a) Prove that $\mathbf{X} = \mathbf{A} \# \mathbf{B}$ is the **only** positive definite solution of $\mathbf{X} \mathbf{A}^{-1} \mathbf{X} = \mathbf{B}$.
- (b) Use (a) to give a second proof that $\mathbf{A} \# \mathbf{B} = \mathbf{B} \# \mathbf{A}$.

Solution.

- (a) Write $\mathbf{C} = \mathbf{A}^{-1/2} \mathbf{B} \mathbf{A}^{-1/2}$. A positive definite $\mathbf{X}$ may be written uniquely as $\mathbf{X} = \mathbf{A}^{1/2} \mathbf{Y} \mathbf{A}^{1/2}$ with $\mathbf{Y} = \mathbf{A}^{-1/2} \mathbf{X} \mathbf{A}^{-1/2}$, and $\mathbf{Y} \succ \mathbf{O}$ by Proposition 13.1.2 (b). As computed in the proof of Theorem 21.11.4 (a), $\mathbf{X} \mathbf{A}^{-1} \mathbf{X} = \mathbf{A}^{1/2} \mathbf{Y}^2 \mathbf{A}^{1/2}$. So $\mathbf{X} \mathbf{A}^{-1} \mathbf{X} = \mathbf{B}$ is equivalent to $\mathbf{A}^{1/2} \mathbf{Y}^2 \mathbf{A}^{1/2} = \mathbf{A}^{1/2} \mathbf{C} \mathbf{A}^{1/2}$, that is to $\mathbf{Y}^2 = \mathbf{C}$, since $\mathbf{A}^{1/2}$ is invertible. By the uniqueness of the positive square root (Theorem 13.2.1), the only positive definite solution is $\mathbf{Y} = \mathbf{C}^{1/2}$, that is $\mathbf{X} = \mathbf{A} \# \mathbf{B}$.
- (b) Let $\mathbf{X} = \mathbf{A} \# \mathbf{B}$, so $\mathbf{X} \mathbf{A}^{-1} \mathbf{X} = \mathbf{B}$ by (a) and $\mathbf{X} \succ \mathbf{O}$ is invertible. Multiplying that identity on the left and on the right by $\mathbf{X}^{-1}$ gives $\mathbf{A}^{-1} = \mathbf{X}^{-1} \mathbf{B} \mathbf{X}^{-1}$, and inverting both sides gives $\mathbf{A} = \mathbf{X} \mathbf{B}^{-1} \mathbf{X}$. So $\mathbf{X}$ is a positive definite solution of $\mathbf{X} \mathbf{B}^{-1} \mathbf{X} = \mathbf{A}$, which by (a), with the roles of $\mathbf{A}$ and $\mathbf{B}$ exchanged, has $\mathbf{B} \# \mathbf{A}$ as its only positive definite solution. Hence $\mathbf{A} \# \mathbf{B} = \mathbf{B} \# \mathbf{A}$.

Exercise. C3: Cauchy–Schwarz for the Ky Fan norms, directly

Let $\mathbf{A}, \mathbf{B} \in M_n(F)$ and $1 \leq k \leq n$.

- (a) Let $\mathbf{A}^* \mathbf{B} = \mathbf{W} \mathbf{\Sigma} \mathbf{Z}^*$ be a singular value decomposition (Theorem 13.9.1), and let $\mathbf{U}_k$ and $\mathbf{V}_k$ be the matrices of the first k columns of $\mathbf{W}$ and of $\mathbf{Z}$. Show that $\text{tr}(\mathbf{U}_k^* \mathbf{A}^* \mathbf{B} \mathbf{V}_k) = \sum_{j \leq k} \sigma_j(\mathbf{A}^* \mathbf{B})$.
- (b) Deduce, using the Cauchy–Schwarz inequality for the Frobenius inner product (Theorem 11.1.5) and the Ky Fan maximum principle (Theorem 17.6.2), that

$$\sum_{j \leq k} \sigma_j(\mathbf{A}^* \mathbf{B}) \leq \left(\sum_{j \leq k} \lambda_j(\mathbf{A}^* \mathbf{A}) \right)^{1/2} \left(\sum_{j \leq k} \lambda_j(\mathbf{B}^* \mathbf{B}) \right)^{1/2}.$$

Solution.

- (a) The columns of $\mathbf{W}$ and $\mathbf{Z}$ are orthonormal, so $\mathbf{U}_k^* \mathbf{W} = (\mathbf{I}_k \ \mathbf{O})$ and $\mathbf{Z}^* \mathbf{V}_k = \begin{pmatrix} \mathbf{I}_k \\ \mathbf{O} \end{pmatrix}$. Hence

$$\mathbf{U}_k^* \mathbf{A}^* \mathbf{B} \mathbf{V}_k = \mathbf{U}_k^* \mathbf{W} \mathbf{\Sigma} \mathbf{Z}^* \mathbf{V}_k = \text{diag}(\sigma_1, \dots, \sigma_k),$$

the top-left $k \times k$ corner of Σ , whose trace is $\sum_{j \leq k} \sigma_j(\mathbf{A}^* \mathbf{B})$.

(b) Writing the trace as the Frobenius inner product $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{tr}(\mathbf{Y}^* \mathbf{X})$ of Chapter 11 §01,

$$\text{tr}(\mathbf{U}_k^* \mathbf{A}^* \mathbf{B} \mathbf{V}_k) = \text{tr}((\mathbf{A} \mathbf{U}_k)^* (\mathbf{B} \mathbf{V}_k)) = \langle \mathbf{B} \mathbf{V}_k, \mathbf{A} \mathbf{U}_k \rangle,$$

so by Theorem 11.1.5 it is at most $\|\mathbf{A} \mathbf{U}_k\|_F \|\mathbf{B} \mathbf{V}_k\|_F$. Now $\mathbf{U}_k^* \mathbf{U}_k = \mathbf{I}_k$, so $\mathbf{U}_k$ is an isometry and

$$\|\mathbf{A} \mathbf{U}_k\|_F^2 = \text{tr}(\mathbf{U}_k^* \mathbf{A}^* \mathbf{A} \mathbf{U}_k) \leq \sum_{j \leq k} \lambda_j(\mathbf{A}^* \mathbf{A})$$

by Theorem 17.6.2 applied to the Hermitian matrix $\mathbf{A}^* \mathbf{A}$. The same bound holds for $\|\mathbf{B} \mathbf{V}_k\|_F^2$ with $\mathbf{B}^* \mathbf{B}$. Combining with (a) gives the display. (Taking $k = n$ recovers [Theorem 21.11.3](#) for the trace norm.)