

20.10 Singular Values and Polar Factors

Chapter 17 §09 proved that singular values are perfectly conditioned: $|\sigma_i(\mathbf{A} + \mathbf{E}) - \sigma_i(\mathbf{A})| \leq \|\mathbf{E}\|_2$, for any matrices at all. This section finishes that story and then starts a new one. First the Frobenius-norm version, Mirsky's theorem, which follows from Section 5's Hoffman–Wielandt inequality once the Hermitian dilation of Chapter 17 turns singular values into eigenvalues. Then the polar decomposition $\mathbf{A} = \mathbf{U}\mathbf{P}$: how far do the unitary factor and the positive factor move when $\mathbf{A}$ does? Both answers come from one tool used twice, the bound of Section 7 for a Sylvester equation with positive coefficients.

Throughout, the field is $\mathbb{C}$, $\|\cdot\|_2$ is the spectral norm and $\|\cdot\|_F$ the Frobenius norm, and Lemma 20.9.1 of Section 9 is used freely to move unitary factors and conjugate transposes in and out of either norm. For $\mathbf{A} \in M_{m \times n}(\mathbb{C})$ we put $p = \min(m, n)$, and $\sigma_1(\mathbf{A}) \geq \cdots \geq \sigma_p(\mathbf{A}) \geq 0$ are its singular values (Definition 13.9.1).

20.10.1 Singular values in the Frobenius norm

Here is where Chapter 17 left the question. Its Corollary 17.9.5 says that for **any** $\mathbf{A}, \mathbf{E} \in M_{m \times n}(\mathbb{C})$,

$$|\sigma_i(\mathbf{A} + \mathbf{E}) - \sigma_i(\mathbf{A})| \leq \|\mathbf{E}\|_2 \quad (1 \leq i \leq p),$$

with no symmetry, squareness, or smallness assumed. In the language of Section 5 this is a bound on the largest entry of the vector of differences, and the matching is simply by index. What it does not give is a good bound on **all** the differences together. Squaring and adding p copies of it gives

$$\sum_{i=1}^p (\sigma_i(\mathbf{A} + \mathbf{E}) - \sigma_i(\mathbf{A}))^2 \leq p \|\mathbf{E}\|_2^2,$$

and the factor p grows with the size of the matrices. For Hermitian eigenvalues, Section 5 removed that factor: Corollary 20.5.4 bounds the sum of squares by $\|\mathbf{E}\|_F^2$, with no dependence on n . Chapter 17 recorded, in Proposition 17.9.3, a device for carrying any theorem about Hermitian eigenvalues over to singular values, and promised that this chapter would lean on it. Here it is.

Theorem 20.10.1. Mirsky's Theorem

Let $\mathbf{A}, \mathbf{B} \in M_{m \times n}(\mathbb{C})$ and $p = \min(m, n)$. Then

$$\sum_{i=1}^p (\sigma_i(\mathbf{A}) - \sigma_i(\mathbf{B}))^2 \leq \|\mathbf{A} - \mathbf{B}\|_F^2.$$

Idea. Replace each matrix by its Hermitian dilation $\mathcal{H}(\cdot)$, whose eigenvalues are $\pm\sigma_i$ and zeros. Listed decreasingly, the eigenvalues of $\mathcal{H}(\mathbf{A})$ are $\sigma_1(\mathbf{A}), \dots, \sigma_p(\mathbf{A})$, then the zeros, then $-\sigma_p(\mathbf{A}), \dots, -\sigma_1(\mathbf{A})$, and the positions depend only on m and n , so $\mathcal{H}(\mathbf{B})$ has the same pattern. Apply the Hermitian Hoffman–Wielandt inequality to the two dilations. On the left each difference $\sigma_i(\mathbf{A}) - \sigma_i(\mathbf{B})$ appears twice, once from the top and once from the bottom, while the zeros cancel against zeros. On the right each entry of $\mathbf{A} - \mathbf{B}$ appears twice, once in each off-diagonal block. The two factors of 2 cancel.

Proof. For $\mathbf{M} \in M_{m \times n}(\mathbb{C})$, $\mathcal{H}(\mathbf{M}) = \begin{pmatrix} \mathbf{O} & \mathbf{M} \\ \mathbf{M}^* & \mathbf{O} \end{pmatrix}$ is Hermitian (Proposition 17.9.3). Since conjugate transposition is additive,

$$\mathcal{H}(\mathbf{A}) - \mathcal{H}(\mathbf{B}) = \begin{pmatrix} \mathbf{O} & \mathbf{A} - \mathbf{B} \\ (\mathbf{A} - \mathbf{B})^* & \mathbf{O} \end{pmatrix} = \mathcal{H}(\mathbf{A} - \mathbf{B}),$$

and since the entries of $\mathcal{H}(\mathbf{M})$ are those of $\mathbf{M}$, those of $\mathbf{M}^*$, and zeros,

$$\|\mathcal{H}(\mathbf{M})\|_F^2 = \|\mathbf{M}\|_F^2 + \|\mathbf{M}^*\|_F^2 = 2\|\mathbf{M}\|_F^2. \quad (20.10.1)$$

By Proposition 17.9.3, $\mathcal{H}(\mathbf{A}) \in M_{m+n}(\mathbb{C})$ has eigenvalues $\pm\sigma_1(\mathbf{A}), \dots, \pm\sigma_p(\mathbf{A})$ and $m+n-2p$ zeros, with multiplicity. Since every $\sigma_i(\mathbf{A}) \geq 0$, the list

$$\begin{aligned} \sigma_1(\mathbf{A}) \geq \dots \geq \sigma_p(\mathbf{A}) &\geq \underbrace{0 \geq \dots \geq 0}_{m+n-2p} \\ &\geq -\sigma_p(\mathbf{A}) \geq \dots \geq -\sigma_1(\mathbf{A}) \end{aligned}$$

is decreasing, and a multiset of real numbers has only one decreasing listing. So, with $N = m + n$,

$$\lambda_i(\mathcal{H}(\mathbf{A})) = \begin{cases} \sigma_i(\mathbf{A}) & 1 \leq i \leq p, \\ 0 & p < i \leq N - p, \\ -\sigma_{N+1-i}(\mathbf{A}) & N - p < i \leq N. \end{cases}$$

The same formula holds for $\mathbf{B}$, with the **same** three ranges of i , because m , n , and p are the same. Hence, summing over the three ranges, and writing $j = N + 1 - i$ in the third,

$$\begin{aligned} \sum_{i=1}^N (\lambda_i(\mathcal{H}(\mathbf{A})) - \lambda_i(\mathcal{H}(\mathbf{B})))^2 &= \sum_{i=1}^p (\sigma_i(\mathbf{A}) - \sigma_i(\mathbf{B}))^2 + 0 \\ &\quad + \sum_{j=1}^p (-\sigma_j(\mathbf{A}) + \sigma_j(\mathbf{B}))^2 \\ &= 2 \sum_{i=1}^p (\sigma_i(\mathbf{A}) - \sigma_i(\mathbf{B}))^2. \end{aligned}$$

By Corollary 20.5.4, applied to the Hermitian matrices $\mathcal{H}(\mathbf{A})$ and $\mathcal{H}(\mathbf{B})$, the left-hand side is at most $\|\mathcal{H}(\mathbf{A}) - \mathcal{H}(\mathbf{B})\|_F^2 = \|\mathcal{H}(\mathbf{A} - \mathbf{B})\|_F^2 = 2\|\mathbf{A} - \mathbf{B}\|_F^2$, by Eq. 20.10.1. Dividing by 2 proves the theorem. ■

This is the payment Chapter 17 §09 promised when it said that Chapter 20 leans on the dilation for perturbation bounds. Together with Corollary 17.9.5, it bounds the vector of singular-value differences in the two norms one reaches for first: its largest entry by $\|\mathbf{A} - \mathbf{B}\|_2$, its Euclidean length by $\|\mathbf{A} - \mathbf{B}\|_F$. Chapter 21 puts both inside one statement covering every unitarily invariant norm.

Note what was **not** assumed. Theorem 20.5.2 needs normal matrices, and Section 5 showed by a Jordan-block pair that it fails without normality. Mirsky's theorem holds for every pair of matrices, square or not, because the dilation manufactures the Hermitian hypothesis instead of assuming it.

Example. Two matrices that differ in one entry

Let $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$. Compute the singular values of both, and compare the two sides of [Theorem 20.10.1](#) with what Corollary 17.9.5 alone would give.

Solution. $\mathbf{A}$ is symmetric with $\mathbf{A}(1, 1) = 3(1, 1)$ and $\mathbf{A}(1, -1) = (1, -1)$, so $\mathbf{A}^\top \mathbf{A} = \mathbf{A}^2$ has eigenvalues 9 and 1, and $\sigma(\mathbf{A}) = (3, 1)$. For $\mathbf{B}$,

$$\mathbf{B}^\top \mathbf{B} = \begin{pmatrix} 4 & 2 \\ 2 & 5 \end{pmatrix},$$

with trace 9 and determinant 16, so its eigenvalues are $\frac{9 \pm \sqrt{17}}{2}$. These are perfect squares: $\left(\frac{\sqrt{17}+1}{2}\right)^2 = \frac{18 \pm 2\sqrt{17}}{4} = \frac{9 \pm \sqrt{17}}{2}$. Hence

$$\sigma_1(\mathbf{B}) = \frac{\sqrt{17}+1}{2} \approx 2.562, \quad \sigma_2(\mathbf{B}) = \frac{\sqrt{17}-1}{2} \approx 1.562.$$

The difference $\mathbf{A} - \mathbf{B} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ has $\|\mathbf{A} - \mathbf{B}\|_F = \|\mathbf{A} - \mathbf{B}\|_2 = 1$. The left side of Mirsky's inequality is

$$\left(\frac{5-\sqrt{17}}{2}\right)^2 + \left(\frac{3-\sqrt{17}}{2}\right)^2 = \frac{(42-10\sqrt{17})+(26-6\sqrt{17})}{4} = 17 - 4\sqrt{17} \approx 0.508,$$

comfortably below 1. The index-by-index bound gives $|\sigma_i(\mathbf{A}) - \sigma_i(\mathbf{B})| \leq 1$ for each i ; the actual values are about 0.438 and 0.562. On its own it bounds the sum of squares only by 2, and Mirsky halves that.

Quick check. Let $\mathbf{A} = \text{diag}(3, 1)$ and $\mathbf{B} = \text{diag}(1, 3)$. Compute both sides of [Theorem 20.10.1](#). Why is the left side not $(3-1)^2 + (1-3)^2$?

Solution. Both matrices have singular values $(3, 1)$ in decreasing order, so the left side is 0, while $\|\mathbf{A} - \mathbf{B}\|_F^2 = \|\text{diag}(2, -2)\|_F^2 = 8$. The theorem compares the **sorted** lists, the i -th largest with the i -th largest. Pairing the diagonal entries by position would give 8, which is also allowed, but it is not what the theorem measures.

Warning. Stable singular values do not mean stable singular vectors. Take $0 < \eta < 1$, $\mathbf{A} = \text{diag}(1 + \eta, 1 - \eta)$ and $\mathbf{B} = \begin{pmatrix} 1 & \eta \\ \eta & 1 \end{pmatrix}$, the pair of Example (A tiny perturbation that turns the eigenvectors through 45°). Both are positive definite with eigenvalues $1 \pm \eta$, so both have singular values $1 \pm \eta$, and the left side of Mirsky's inequality is 0. But the right singular vector for σ_1 is $\mathbf{e}_1$ for $\mathbf{A}$ and $\frac{1}{\sqrt{2}}(1, 1)$ for $\mathbf{B}$, a 45° turn, while $\|\mathbf{A} - \mathbf{B}\|_F = 2\eta$ is as small as we like. As with eigenvectors, singular vectors are controlled by gaps. Applying Section 9 to the dilations $\mathcal{H}(\mathbf{A})$ and $\mathcal{H}(\mathbf{B})$, whose eigenvectors are built from the singular vectors, is one way to bound them.

20.10.2 The square root is Lipschitz away from zero

The second half of the section is about the polar decomposition, and its positive factor is a square root, $|\mathbf{A}| = (\mathbf{A}^* \mathbf{A})^{1/2}$. So we start with square roots.

For positive numbers there is an exact formula for how far a square root moves:

$$\sqrt{p} - \sqrt{q} = \frac{p - q}{\sqrt{p} + \sqrt{q}},$$

the usual trick of multiplying by the conjugate. It shows that the square root is Lipschitz on $[c, \infty)$ with constant $1/(2\sqrt{c})$, and that no constant works near 0. For matrices the trick fails as written: with $\mathbf{S} = \mathbf{P}^{1/2}$ and $\mathbf{T} = \mathbf{Q}^{1/2}$, the product $(\mathbf{S} - \mathbf{T})(\mathbf{S} + \mathbf{T}) = \mathbf{S}^2 + \mathbf{ST} - \mathbf{TS} - \mathbf{T}^2$ is $\mathbf{P} - \mathbf{Q}$ only when $\mathbf{S}$ and $\mathbf{T}$ commute. But a different grouping of the same terms works with no commuting at all:

$$\mathbf{S}(\mathbf{S} - \mathbf{T}) + (\mathbf{S} - \mathbf{T})\mathbf{T} = \mathbf{S}^2 - \mathbf{ST} + \mathbf{ST} - \mathbf{T}^2 = \mathbf{P} - \mathbf{Q}. \quad (20.10.2)$$

This says that $\mathbf{X} = \mathbf{S} - \mathbf{T}$ solves the Sylvester equation $\mathbf{SX} + \mathbf{XT} = \mathbf{P} - \mathbf{Q}$, whose coefficients are positive definite. That is exactly the situation of Corollary 20.7.4.

Theorem 20.10.2. Perturbation of the Square Root

Let $\mathbf{P}, \mathbf{Q} \in M_n(\mathbb{C})$ be positive definite. Then

$$\|\mathbf{P}^{1/2} - \mathbf{Q}^{1/2}\| \leq \frac{\|\mathbf{P} - \mathbf{Q}\|}{\lambda_n(\mathbf{P})^{1/2} + \lambda_n(\mathbf{Q})^{1/2}},$$

where $\|\cdot\|$ is either $\|\cdot\|_2$ or $\|\cdot\|_F$.

Proof. Put $\mathbf{S} = \mathbf{P}^{1/2}$ and $\mathbf{T} = \mathbf{Q}^{1/2}$, the positive square roots of Theorem 13.2.1. By that theorem the eigenvalues of $\mathbf{S}$ are the nonnegative square roots of those of $\mathbf{P}$, and $t \mapsto \sqrt{t}$ is increasing, so the smallest eigenvalue of $\mathbf{S}$ is $\alpha := \lambda_n(\mathbf{P})^{1/2}$, which is positive because $\mathbf{P} \succ \mathbf{O}$. Every eigenvalue of the Hermitian matrix $\mathbf{S} - \alpha\mathbf{I}$ is therefore nonnegative, so $\mathbf{S} \succeq \alpha\mathbf{I}$ by Theorem 13.1.4 (b). In the same way $\mathbf{T} \succeq \beta\mathbf{I}$ with $\beta := \lambda_n(\mathbf{Q})^{1/2} > 0$.

By Eq. 20.10.2, $\mathbf{X} = \mathbf{S} - \mathbf{T}$ satisfies $\mathbf{SX} + \mathbf{XT} = \mathbf{P} - \mathbf{Q}$. Since $\alpha + \beta > 0$, Corollary 20.7.4 applies to this equation, whose solution is therefore unique and equal to $\mathbf{X}$, and gives

$$\|\mathbf{S} - \mathbf{T}\| \leq \frac{\|\mathbf{P} - \mathbf{Q}\|}{\alpha + \beta}$$

in both the spectral and the Frobenius norm. This proves the theorem. ■

For $n = 1$ this is the scalar identity, as an inequality that happens to be an equality. It is sharp for every n : with $\mathbf{P} = \text{diag}(4, 100)$ and $\mathbf{Q} = \text{diag}(1, 100)$, $\mathbf{P}^{1/2} - \mathbf{Q}^{1/2} = \text{diag}(1, 0)$ has both norms equal to 1, and the bound is $3/(2 + 1) = 1$; padding both matrices with a further block $100\mathbf{I}_{n-2}$ changes neither side of the inequality, so the equality holds for every $n \geq 2$. The

difference sits exactly in the direction where both matrices are smallest, which is the worst place for it.

Remark. The proof used only $\alpha + \beta > 0$, so the conclusion holds as soon as one of $\mathbf{P}$ and $\mathbf{Q}$ is positive definite and the other positive semidefinite.

Warning. **The square root is not Lipschitz on all positive semidefinite matrices.** Take $\mathbf{P} = \varepsilon \mathbf{I}_n$ with $0 < \varepsilon < 1$ and $\mathbf{Q} = \mathbf{O}$. Then $\|\mathbf{P} - \mathbf{Q}\|_2 = \varepsilon$, while $\mathbf{P}^{1/2} - \mathbf{Q}^{1/2} = \sqrt{\varepsilon} \mathbf{I}_n$ has spectral norm $\sqrt{\varepsilon}$, which is larger by the factor $1/\sqrt{\varepsilon}$. That factor is unbounded as $\varepsilon \rightarrow 0^+$, so no inequality $\|\mathbf{P}^{1/2} - \mathbf{Q}^{1/2}\|_2 \leq C\|\mathbf{P} - \mathbf{Q}\|_2$ can hold for all $\mathbf{P}, \mathbf{Q} \succeq \mathbf{O}$. The denominator of the theorem is what records this. By the remark the theorem still applies, with denominator $\sqrt{\varepsilon} + 0$, and the bound $\varepsilon/\sqrt{\varepsilon} = \sqrt{\varepsilon}$ is attained. [Exercise \(C1: The square root is continuous everywhere\)](#) shows that the square root is nevertheless continuous everywhere, with a square-root modulus.

20.10.3 The positive polar factor

Recall Theorem 13.10.1: every $\mathbf{A} \in M_n(\mathbb{C})$ is a unitary matrix times the positive semidefinite $|\mathbf{A}| = (\mathbf{A}^* \mathbf{A})^{1/2}$, and when $\mathbf{A}$ is invertible both factors are unique. In this section we write the decomposition as

$$\mathbf{A} = \mathbf{U}\mathbf{P}, \quad \mathbf{U} \text{ unitary}, \quad \mathbf{P} = |\mathbf{A}|,$$

keeping the letter $\mathbf{W}$ for another use below. For invertible $\mathbf{A}$ the matrix $\mathbf{A}^* \mathbf{A}$ is positive definite, since $\mathbf{x}^* \mathbf{A}^* \mathbf{A} \mathbf{x} = \|\mathbf{A}\mathbf{x}\|^2 > 0$ for $\mathbf{x} \neq \mathbf{0}$. Its eigenvalues are the $\sigma_i(\mathbf{A})^2$ (Definition 13.9.1), so

$$\lambda_n(\mathbf{A}^* \mathbf{A}) = \sigma_n(\mathbf{A})^2 > 0, \quad \mathbf{P} = |\mathbf{A}| \succeq \sigma_n(\mathbf{A}) \mathbf{I},$$

the second because the eigenvalues of $|\mathbf{A}|$ are the square roots $\sigma_i(\mathbf{A})$ (Theorem 13.2.1) and Theorem 13.1.4 (b) applies to $|\mathbf{A}| - \sigma_n(\mathbf{A}) \mathbf{I}$ as before.

Theorem 20.10.3. Perturbation of the Positive Polar Factor

Let $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{C})$ be invertible. Then

$$\| |\mathbf{A}| - |\mathbf{B}| \| \leq \frac{(\|\mathbf{A}\|_2 + \|\mathbf{B}\|_2) \|\mathbf{A} - \mathbf{B}\|}{\sigma_n(\mathbf{A}) + \sigma_n(\mathbf{B})},$$

where $\|\cdot\|$ is either $\|\cdot\|_2$ or $\|\cdot\|_F$.

Proof. Both $\mathbf{A}^* \mathbf{A}$ and $\mathbf{B}^* \mathbf{B}$ are positive definite, with $\lambda_n(\mathbf{A}^* \mathbf{A})^{1/2} = \sigma_n(\mathbf{A})$ and $\lambda_n(\mathbf{B}^* \mathbf{B})^{1/2} = \sigma_n(\mathbf{B})$, as just shown. So [Theorem 20.10.2](#) gives

$$\| |\mathbf{A}| - |\mathbf{B}| \| \leq \frac{\|\mathbf{A}^* \mathbf{A} - \mathbf{B}^* \mathbf{B}\|}{\sigma_n(\mathbf{A}) + \sigma_n(\mathbf{B})}.$$

Now $\mathbf{A}^* \mathbf{A} - \mathbf{B}^* \mathbf{B} = \mathbf{A}^*(\mathbf{A} - \mathbf{B}) + (\mathbf{A} - \mathbf{B})^* \mathbf{B}$, as expanding the right side shows. By Lemma 20.9.1 (b) and (a),

$$\|\mathbf{A}^*(\mathbf{A} - \mathbf{B})\| \leq \|\mathbf{A}^*\|_2 \|\mathbf{A} - \mathbf{B}\| = \|\mathbf{A}\|_2 \|\mathbf{A} - \mathbf{B}\|,$$

and likewise $\|(\mathbf{A} - \mathbf{B})^* \mathbf{B}\| \leq \|(\mathbf{A} - \mathbf{B})^*\| \|\mathbf{B}\|_2 = \|\mathbf{A} - \mathbf{B}\| \|\mathbf{B}\|_2$. The triangle inequality for the norm (Theorem 16.3.2 (b) for $\|\cdot\|_2$, Corollary 11.1.6 for the Frobenius norm, which is induced by an inner product) gives $\|\mathbf{A}^* \mathbf{A} - \mathbf{B}^* \mathbf{B}\| \leq (\|\mathbf{A}\|_2 + \|\mathbf{B}\|_2) \|\mathbf{A} - \mathbf{B}\|$, and substituting proves the theorem. ■

When $\mathbf{B}$ is close to $\mathbf{A}$, the constant is about $2\|\mathbf{A}\|_2/(2\sigma_n(\mathbf{A})) = \kappa_2(\mathbf{A})$, the condition number of Chapter 16. So by this estimate $|\mathbf{A}|$ is as sensitive as $\mathbf{A}$ is ill-conditioned. It is a crude estimate, as the example below shows. Sharper ones are known, but this is the one the Sylvester bound gives with no extra work.

20.10.4 The unitary polar factor

For a nonzero complex number, the unitary polar factor is the phase $z/|z|$. If two numbers z and w have the same modulus r , their phases differ by exactly $|z - w|/r$. So the phase moves at the rate of the perturbation divided by the modulus, and it is undefined at $z = 0$. The theorem below is the matrix form of this, with σ_n as the modulus.

Theorem 20.10.4. Perturbation of the Unitary Polar Factor

Let $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{C})$ be invertible, with polar decompositions (Theorem 13.10.1) $\mathbf{A} = \mathbf{U}\mathbf{P}$ and $\mathbf{B} = \mathbf{V}\mathbf{Q}$, where $\mathbf{U}, \mathbf{V}$ are unitary, $\mathbf{P} = |\mathbf{A}|$, and $\mathbf{Q} = |\mathbf{B}|$. Then

$$\|\mathbf{U} - \mathbf{V}\| \leq \frac{2\|\mathbf{A} - \mathbf{B}\|}{\sigma_n(\mathbf{A}) + \sigma_n(\mathbf{B})},$$

where $\|\cdot\|$ is either $\|\cdot\|_2$ or $\|\cdot\|_F$.

Idea. Measure $\mathbf{U}$ against $\mathbf{V}$ by the single unitary matrix $\mathbf{W} = \mathbf{V}^* \mathbf{U}$, which equals $\mathbf{I}$ exactly when $\mathbf{U} = \mathbf{V}$; then $\|\mathbf{U} - \mathbf{V}\| = \|\mathbf{W} - \mathbf{I}\|$. Now write $\mathbf{A} - \mathbf{B}$ in the two polar frames. Seen through $\mathbf{V}^*$ on the left it is $\mathbf{W}\mathbf{P} - \mathbf{Q}$. Its conjugate transpose, seen through $\mathbf{U}$ on the right, is $\mathbf{P} - \mathbf{Q}\mathbf{W}$. The difference of these two expressions is $(\mathbf{W} - \mathbf{I})\mathbf{P} + \mathbf{Q}(\mathbf{W} - \mathbf{I})$, a Sylvester equation for $\mathbf{W} - \mathbf{I}$ with positive definite coefficients $\mathbf{Q}$ and $\mathbf{P}$, and its right-hand side has norm at most $2\|\mathbf{A} - \mathbf{B}\|$.

Proof. Put $\mathbf{W} = \mathbf{V}^* \mathbf{U}$, which is unitary. Since $\mathbf{U}^* \mathbf{U} = \mathbf{V}^* \mathbf{V} = \mathbf{I}$, $\mathbf{P}^* = \mathbf{P}$, and $\mathbf{Q}^* = \mathbf{Q}$,

$$\begin{aligned} \mathbf{V}^*(\mathbf{A} - \mathbf{B}) &= \mathbf{V}^* \mathbf{U} \mathbf{P} - \mathbf{V}^* \mathbf{V} \mathbf{Q} = \mathbf{W} \mathbf{P} - \mathbf{Q}, \\ (\mathbf{A} - \mathbf{B})^* \mathbf{U} &= \mathbf{P} \mathbf{U}^* \mathbf{U} - \mathbf{Q} \mathbf{V}^* \mathbf{U} = \mathbf{P} - \mathbf{Q} \mathbf{W}. \end{aligned}$$

Subtracting the second identity from the first,

$$\begin{aligned} \mathbf{V}^*(\mathbf{A} - \mathbf{B}) - (\mathbf{A} - \mathbf{B})^* \mathbf{U} &= \mathbf{W} \mathbf{P} - \mathbf{P} + \mathbf{Q} \mathbf{W} - \mathbf{Q} \\ &= \mathbf{Q}(\mathbf{W} - \mathbf{I}) + (\mathbf{W} - \mathbf{I})\mathbf{P}. \end{aligned} \tag{20.10.3}$$

So $\mathbf{X} = \mathbf{W} - \mathbf{I}$ solves $\mathbf{Q}\mathbf{X} + \mathbf{X}\mathbf{P} = \mathbf{Y}$, with $\mathbf{Y}$ the left side of Eq. 20.10.3. As shown before Theorem 20.10.3, $\mathbf{Q} \succeq \sigma_n(\mathbf{B})\mathbf{I}$ and $\mathbf{P} \succeq \sigma_n(\mathbf{A})\mathbf{I}$, and $\sigma_n(\mathbf{A}) + \sigma_n(\mathbf{B}) > 0$. So Corollary 20.7.4,

applied with $\mathbf{Q}$ and $\mathbf{P}$ as its two coefficients, gives

$$\|\mathbf{W} - \mathbf{I}\| \leq \frac{\|\mathbf{Y}\|}{\sigma_n(\mathbf{A}) + \sigma_n(\mathbf{B})}$$

in both norms. By Lemma 20.9.1 (c) and (a), $\|\mathbf{V}^*(\mathbf{A} - \mathbf{B})\| = \|\mathbf{A} - \mathbf{B}\|$ and $\|(\mathbf{A} - \mathbf{B})^*\mathbf{U}\| = \|(\mathbf{A} - \mathbf{B})^*\| = \|\mathbf{A} - \mathbf{B}\|$, so $\|\mathbf{Y}\| \leq 2\|\mathbf{A} - \mathbf{B}\|$ by the triangle inequality in either norm (Theorem 16.3.2 (b), Corollary 11.1.6). Finally $\mathbf{U} - \mathbf{V} = \mathbf{V}(\mathbf{V}^*\mathbf{U} - \mathbf{I}) = \mathbf{V}(\mathbf{W} - \mathbf{I})$, so $\|\mathbf{U} - \mathbf{V}\| = \|\mathbf{W} - \mathbf{I}\|$ by Lemma 20.9.1 (c). Combining the three displays proves the theorem. ■

Notice the difference from [Theorem 20.10.3](#). There the constant was a condition number, but here only σ_n appears. The unitary factor feels how close $\mathbf{A}$ is to the singular matrices, which is $\sigma_n(\mathbf{A})$ by Corollary 17.9.7, and nothing else.

Quick check. For $n = 1$, what does the theorem say about two nonzero complex numbers z and w ? Test it on $z = 1$, $w = -1$.

Solution. A nonzero $z \in \mathbb{C} = M_1(\mathbb{C})$ has polar decomposition $z = \frac{z}{|z|} \cdot |z|$, with $\sigma_1(z) = |z|$, and both norms are the modulus. So the theorem says

$$\left| \frac{z}{|z|} - \frac{w}{|w|} \right| \leq \frac{2|z - w|}{|z| + |w|}.$$

For $z = 1$, $w = -1$ both sides equal 2. So the constant 2 in the theorem cannot be lowered, even for $n = 1$.

Example. The polar factors of two nearby matrices

Let $\mathbf{A} = \begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix}$, whose polar decomposition Example (A polar decomposition) found to be $\mathbf{U} = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}$, $|\mathbf{A}| = \sqrt{5} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, with $\sigma(\mathbf{A}) = (3\sqrt{5}, \sqrt{5})$. Let $\mathbf{B} = \begin{pmatrix} 3 & 0 \\ 6 & 5 \end{pmatrix}$. Find the polar decomposition of $\mathbf{B}$, compute $\|\mathbf{U} - \mathbf{V}\|$, and compare it with [Theorem 20.10.4](#).

Solution. Finding $\mathbf{V}$. For a real 2×2 matrix of positive determinant, look for $\mathbf{V}$ among the rotations $\mathbf{R}_\varphi = \begin{pmatrix} c & -s \\ s & c \end{pmatrix}$, with $c = \cos \varphi$, $s = \sin \varphi$. For $\mathbf{B} = (b_{ij})$, the off-diagonal entries of $\mathbf{R}_\varphi^\top \mathbf{B}$ are $cb_{12} + sb_{22}$ and $-sb_{11} + cb_{21}$, and they agree exactly when $(b_{11} + b_{22})s = (b_{21} - b_{12})c$. For our $\mathbf{B}$ this reads $8s = 6c$, so $(c, s) = (\frac{4}{5}, \frac{3}{5})$, and

$$\mathbf{V} = \frac{1}{5} \begin{pmatrix} 4 & -3 \\ 3 & 4 \end{pmatrix},$$

$$\mathbf{V}^\top \mathbf{B} = \frac{1}{5} \begin{pmatrix} 4 & 3 \\ -3 & 4 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 6 & 5 \end{pmatrix} = \begin{pmatrix} 6 & 3 \\ 3 & 4 \end{pmatrix}.$$

This symmetric matrix has trace 10 and determinant 15, so both its eigenvalues, $5 \pm \sqrt{10}$, are positive, and it is positive definite. Thus $\mathbf{B} = \mathbf{V}(\mathbf{V}^\top \mathbf{B})$ is a unitary matrix times a positive definite one. Squaring gives $(\mathbf{V}^\top \mathbf{B})^2 = \mathbf{B}^\top \mathbf{V} \mathbf{V}^\top \mathbf{B} = \mathbf{B}^\top \mathbf{B}$, so $\mathbf{V}^\top \mathbf{B} = |\mathbf{B}|$ by uniqueness

of the positive square root (Theorem 13.2.1). Since $\mathbf{B}$ is invertible, Theorem 13.10.1 now gives $\mathbf{V}$ as its unitary polar factor. Its singular values are the eigenvalues of $|\mathbf{B}|$, $\sigma(\mathbf{B}) = (5 + \sqrt{10}, 5 - \sqrt{10}) \approx (8.162, 1.838)$. The same recipe applied to $\mathbf{A}$ gives $8s = 4c$, which recovers $\mathbf{U}$.

The distance. $\mathbf{U}$ and $\mathbf{V}$ are the rotations through angles $\varphi_{\mathbf{A}}$ and $\varphi_{\mathbf{B}}$ with

$$\cos(\varphi_{\mathbf{B}} - \varphi_{\mathbf{A}}) = \frac{2}{\sqrt{5}} \cdot \frac{4}{5} + \frac{1}{\sqrt{5}} \cdot \frac{3}{5} = \frac{11}{5\sqrt{5}}.$$

For two rotations, $(\mathbf{U} - \mathbf{V})^\top(\mathbf{U} - \mathbf{V}) = 2\mathbf{I} - (\mathbf{U}^\top\mathbf{V} + \mathbf{V}^\top\mathbf{U}) = (2 - 2\cos(\varphi_{\mathbf{B}} - \varphi_{\mathbf{A}}))\mathbf{I}$, so both singular values of $\mathbf{U} - \mathbf{V}$ equal $\sqrt{2 - 22/(5\sqrt{5})}$ and

$$\|\mathbf{U} - \mathbf{V}\|_2 \approx 0.180, \quad \|\mathbf{U} - \mathbf{V}\|_F \approx 0.254.$$

The bound. $\mathbf{A} - \mathbf{B} = \begin{pmatrix} 0 & 0 \\ -2 & 0 \end{pmatrix}$, so $\|\mathbf{A} - \mathbf{B}\|_2 = \|\mathbf{A} - \mathbf{B}\|_F = 2$, and

$$\frac{2 \cdot 2}{\sigma_2(\mathbf{A}) + \sigma_2(\mathbf{B})} = \frac{4}{\sqrt{5} + 5 - \sqrt{10}} \approx 0.982$$

in both norms. The theorem holds, with a factor of about 5 to spare in the spectral norm and 4 in the Frobenius norm. For the positive factors, Theorem 20.10.3 gives

$$\| |\mathbf{A}| - |\mathbf{B}| \| \leq \frac{(3\sqrt{5} + 5 + \sqrt{10}) \cdot 2}{\sqrt{5} + 5 - \sqrt{10}} \approx 7.30,$$

while $|\mathbf{A}| - |\mathbf{B}| = \begin{pmatrix} 2\sqrt{5}-6 & \sqrt{5}-3 \\ \sqrt{5}-3 & 2\sqrt{5}-4 \end{pmatrix}$ has spectral norm about 1.79 and Frobenius norm about 1.93. That bound is correct but crude, as promised.

Warning. At a singular matrix the unitary factor jumps, and no bound of this kind survives. For $0 < \varepsilon \leq 1$ let

$$\begin{aligned} \mathbf{A}_\varepsilon &= \text{diag}(1, \varepsilon) = \mathbf{I} \text{diag}(1, \varepsilon), \\ \mathbf{B}_\varepsilon &= \text{diag}(1, -\varepsilon) = \text{diag}(1, -1) \text{diag}(1, \varepsilon). \end{aligned}$$

These are polar decompositions: the unitary factors are $\mathbf{U} = \mathbf{I}$ and $\mathbf{V} = \text{diag}(1, -1)$, and both positive factors are $\text{diag}(1, \varepsilon)$. As $\varepsilon \rightarrow 0^+$, both matrices tend to the singular $\text{diag}(1, 0)$, and $\|\mathbf{A}_\varepsilon - \mathbf{B}_\varepsilon\| = 2\varepsilon \rightarrow 0$ in both norms. Yet $\mathbf{U} - \mathbf{V} = \text{diag}(0, 2)$ has spectral and Frobenius norm 2 for every ε . So the unitary factor has no limit at $\text{diag}(1, 0)$, and it cannot be chosen continuously there. This matches the nonuniqueness in Theorem 13.10.1: both $\mathbf{I}$ and $\text{diag}(1, -1)$ are unitary polar factors of $\text{diag}(1, 0)$. The theorem, with $\sigma_2(\mathbf{A}_\varepsilon) = \sigma_2(\mathbf{B}_\varepsilon) = \varepsilon$, gives the bound $(2 \cdot 2\varepsilon)/(2\varepsilon) = 2$, which is attained. Padding both matrices with an identity block, $\mathbf{A}_\varepsilon \oplus \mathbf{I}_{n-2}$ and $\mathbf{B}_\varepsilon \oplus \mathbf{I}_{n-2}$, gives the same equality for every $n \geq 2$: the constant 2 cannot be improved, and the denominator cannot be dropped.

Corollary 13.10.2 identified the unitary polar factor as the unitary matrix nearest to $\mathbf{A}$ in the Frobenius norm. Theorem 20.10.4 adds that for invertible $\mathbf{A}$ this nearest unitary matrix moves continuously with $\mathbf{A}$, at a rate no worse than $2/(\sigma_n(\mathbf{A}) + \sigma_n(\mathbf{B}))$. A matrix that has drifted from unitarity can therefore be put back reliably, as long as it stays well away from the singular matrices.

20.10.5 Exercises

20.10.5.1 A. Check your understanding

Exercise. A1

- State Mirsky's theorem. In its proof by the Hermitian dilation of Proposition 17.9.3, where does each of the two factors of 2 come from?
- True or false: $|\sigma_i(\mathbf{A}) - \sigma_i(\mathbf{B})| \leq \|\mathbf{A} - \mathbf{B}\|_F$ for every i . Justify.
- True or false: the unitary polar factor is a continuous function of $\mathbf{A}$ on the invertible matrices, and it extends continuously to all of $M_n(\mathbb{C})$. Justify.
- Which identity turns the perturbation of a square root into a Sylvester equation, and which result then bounds it?
- Why does Theorem 20.10.4 hold in both the spectral and the Frobenius norm?

Solution.

- Theorem 20.10.1: $\sum_{i=1}^p (\sigma_i(\mathbf{A}) - \sigma_i(\mathbf{B}))^2 \leq \|\mathbf{A} - \mathbf{B}\|_F^2$. On the eigenvalue side, each difference $\sigma_i(\mathbf{A}) - \sigma_i(\mathbf{B})$ occurs twice in the sorted spectra of the dilations, once as σ_i near the top and once as $-\sigma_i$ near the bottom. On the norm side, each entry of $\mathbf{A} - \mathbf{B}$ occurs twice in $\mathcal{H}(\mathbf{A} - \mathbf{B})$, once in $\mathbf{A} - \mathbf{B}$ and once, conjugated, in $(\mathbf{A} - \mathbf{B})^*$ (Eq. 20.10.1).
- True. Every term of the sum in Mirsky's theorem is nonnegative, so each one is at most the sum, which is at most $\|\mathbf{A} - \mathbf{B}\|_F^2$. (It also follows from Corollary 17.9.5 and $\|\cdot\|_2 \leq \|\cdot\|_F$, Proposition 16.3.4.)
- False as a whole. The first half is true: by Theorem 20.10.4, $\|\mathbf{U} - \mathbf{V}\|_2 \leq 2\|\mathbf{A} - \mathbf{B}\|_2 / (\sigma_n(\mathbf{A}) + \sigma_n(\mathbf{B}))$, which tends to 0 as $\mathbf{B} \rightarrow \mathbf{A}$ for fixed invertible $\mathbf{A}$, because $\sigma_n(\mathbf{B}) \geq 0$. The second half is false: in the last warning of the section, $\mathbf{A}_\varepsilon$ and $\mathbf{B}_\varepsilon$ both tend to $\text{diag}(1, 0)$ while their unitary factors stay at distance 2, so no value at $\text{diag}(1, 0)$ makes the factor continuous there.
- $\mathbf{S}(\mathbf{S} - \mathbf{T}) + (\mathbf{S} - \mathbf{T})\mathbf{T} = \mathbf{S}^2 - \mathbf{T}^2$ (Eq. 20.10.2), which exhibits $\mathbf{S} - \mathbf{T}$ as the solution of $\mathbf{S}\mathbf{X} + \mathbf{X}\mathbf{T} = \mathbf{P} - \mathbf{Q}$. Section 7's Corollary 20.7.4 bounds that solution.
- Each step holds in both norms: Corollary 20.7.4 is stated for both, and so are the facts of Lemma 20.9.1 used to remove the unitary factors and conjugate transposes.

20.10.5.2 B. Practice

Exercise. B1: Mirsky by hand

Let $\mathbf{A} = \begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix}$, with $\sigma(\mathbf{A}) = (3\sqrt{5}, \sqrt{5})$ from Example (A polar decomposition), and $\mathbf{B} = \begin{pmatrix} 3 & 0 \\ 4 & 3 \end{pmatrix}$.

- Show that $\sigma(\mathbf{B}) = (\sqrt{13} + 2, \sqrt{13} - 2)$.
- Verify Theorem 20.10.1 for this pair.
- What bound on $\sum_i (\sigma_i(\mathbf{A}) - \sigma_i(\mathbf{B}))^2$ does Corollary 17.9.5 give? Hence say what Mirsky's theorem gains here.

Solution.

- (a) $\mathbf{B}^\top \mathbf{B} = \begin{pmatrix} 25 & 12 \\ 12 & 9 \end{pmatrix}$ has trace 34 and determinant $225 - 144 = 81$, so its eigenvalues are $17 \pm \sqrt{289 - 81} = 17 \pm 4\sqrt{13}$. Since $(\sqrt{13} \pm 2)^2 = 17 \pm 4\sqrt{13}$ and $\sqrt{13} - 2 > 0$, the singular values are $\sqrt{13} \pm 2 \approx 5.606, 1.606$.

- (b) $\mathbf{A} - \mathbf{B} = \text{diag}(0, 2)$, so $\|\mathbf{A} - \mathbf{B}\|_F^2 = 4$. The left side is

$$\begin{aligned} & (3\sqrt{5} - \sqrt{13} - 2)^2 + (\sqrt{5} - \sqrt{13} + 2)^2 \\ &= 84 - 8\sqrt{65} - 8\sqrt{5} \approx 1.613, \end{aligned}$$

expanding each square (the $\sqrt{13}$ terms cancel). Indeed $1.613 \leq 4$.

- (c) $\|\mathbf{A} - \mathbf{B}\|_2 = 2$, so each $|\sigma_i(\mathbf{A}) - \sigma_i(\mathbf{B})| \leq 2$, and the sum of the two squares is at most 8. Mirsky's theorem gives 4, half as much. The true value is about 1.613: the differences are about 1.103 and 0.631.

Exercise. B2: A square root that does not commute

Let $\mathbf{S} = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}$, $\mathbf{T} = \begin{pmatrix} 3 & 2 \\ 2 & 4 \end{pmatrix}$, $\mathbf{P} = \mathbf{S}^2$, and $\mathbf{Q} = \mathbf{T}^2$.

- (a) Compute $\mathbf{P}$ and $\mathbf{Q}$, and explain why $\mathbf{S} = \mathbf{P}^{1/2}$ and $\mathbf{T} = \mathbf{Q}^{1/2}$.
 (b) Find $\lambda_2(\mathbf{P})$ and $\lambda_2(\mathbf{Q})^{1/2}$.
 (c) Compute $\|\mathbf{P} - \mathbf{Q}\|_2$ and $\|\mathbf{P} - \mathbf{Q}\|_F$, and hence the two bounds of Theorem 20.10.2. Compare them with $\|\mathbf{S} - \mathbf{T}\|$.

Solution.

- (a) Squaring, $\mathbf{P} = \begin{pmatrix} 13 & 12 \\ 12 & 13 \end{pmatrix}$ and $\mathbf{Q} = \begin{pmatrix} 13 & 14 \\ 14 & 20 \end{pmatrix}$. $\mathbf{S}$ is symmetric with eigenvalues 5 and 1 (eigenvectors $(1, \pm 1)$), and $\mathbf{T}$ is symmetric with trace 7 and determinant 8, so both eigenvalues of each are positive, and $\mathbf{S}, \mathbf{T} \succ \mathbf{O}$. By the uniqueness in Theorem 13.2.1, $\mathbf{S} = \mathbf{P}^{1/2}$ and $\mathbf{T} = \mathbf{Q}^{1/2}$.
- (b) The eigenvalues of $\mathbf{P}$ are 25 and 1, so $\lambda_2(\mathbf{P}) = 1$. The eigenvalues of $\mathbf{T}$ are $\frac{7 \pm \sqrt{17}}{2}$, and by Theorem 13.2.1 those of $\mathbf{Q}$ are their squares, so $\lambda_2(\mathbf{Q})^{1/2} = \lambda_2(\mathbf{T}) = \frac{7 - \sqrt{17}}{2} \approx 1.438$.
- (c) $\mathbf{P} - \mathbf{Q} = \begin{pmatrix} 0 & -2 \\ -2 & -7 \end{pmatrix}$ is symmetric with characteristic polynomial $t^2 + 7t - 4$, so its eigenvalues are $\frac{-7 \pm \sqrt{65}}{2}$, and by Lemma 17.3.5 $\|\mathbf{P} - \mathbf{Q}\|_2 = \frac{7 + \sqrt{65}}{2} \approx 7.531$. Also $\|\mathbf{P} - \mathbf{Q}\|_F = \sqrt{4 + 4 + 49} = \sqrt{57} \approx 7.550$. The denominator is $1 + 1.438 = 2.438$, so the bounds are about 3.089 (spectral) and 3.096 (Frobenius). Since $\mathbf{S} - \mathbf{T} = \text{diag}(0, -1)$, both of its norms equal 1. Hence the theorem holds here with a factor of about 3 to spare. It is not an equality because $\mathbf{S} - \mathbf{T}$ does not sit along the eigenvector of $\mathbf{P}$ and $\mathbf{Q}$ where both are smallest.

Exercise. B3: A nearly sharp instance

Let $\mathbf{A} = 2\mathbf{I}_2$ and $\mathbf{B} = \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}$.

- (a) Find the polar decompositions of $\mathbf{A}$ and $\mathbf{B}$.
 (b) Compute $\|\mathbf{U} - \mathbf{V}\|_2$ and $\|\mathbf{U} - \mathbf{V}\|_F$, and compare them with the bounds of Theo-

rem 20.10.4.

- (c) Compute $\| |\mathbf{A}| - |\mathbf{B}| \|_2$ and the bound of [Theorem 20.10.3](#).

Solution.

- (a) $\mathbf{A} = \mathbf{I} \cdot 2\mathbf{I}$, with $\mathbf{U} = \mathbf{I}$ and $|\mathbf{A}| = 2\mathbf{I} \succ \mathbf{O}$. The columns of $\mathbf{B}$ are orthogonal, each of length $\sqrt{5}$, so $\mathbf{V} = \frac{1}{\sqrt{5}}\mathbf{B}$ is orthogonal and $\mathbf{B} = \mathbf{V} \cdot \sqrt{5}\mathbf{I}$, with $\sqrt{5}\mathbf{I} \succ \mathbf{O}$. Both matrices are invertible, so these are their polar decompositions (Theorem 13.10.1), and $\sigma_2(\mathbf{A}) = 2$, $\sigma_2(\mathbf{B}) = \sqrt{5}$.
- (b) $\mathbf{V}$ is the rotation with $\cos \varphi = \frac{2}{\sqrt{5}}$, so, as in [Example \(The polar factors of two nearby matrices\)](#), $(\mathbf{U} - \mathbf{V})^\top (\mathbf{U} - \mathbf{V}) = (2 - 2\cos \varphi)\mathbf{I} = (2 - \frac{4}{\sqrt{5}})\mathbf{I}$. Hence $\|\mathbf{U} - \mathbf{V}\|_2 = \sqrt{2 - 4/\sqrt{5}} \approx 0.4595$ and $\|\mathbf{U} - \mathbf{V}\|_F = \sqrt{2} \|\mathbf{U} - \mathbf{V}\|_2 \approx 0.6498$. Next, $\mathbf{A} - \mathbf{B} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ is orthogonal, so $\|\mathbf{A} - \mathbf{B}\|_2 = 1$ and $\|\mathbf{A} - \mathbf{B}\|_F = \sqrt{2}$. The bounds are

$$\frac{2 \cdot 1}{2 + \sqrt{5}} \approx 0.4721, \quad \frac{2\sqrt{2}}{2 + \sqrt{5}} \approx 0.6677.$$

Both hold, and each is within 3% of the true value.

- (c) $|\mathbf{A}| - |\mathbf{B}| = (2 - \sqrt{5})\mathbf{I}$, of spectral norm $\sqrt{5} - 2 \approx 0.236$. The bound is $(\|\mathbf{A}\|_2 + \|\mathbf{B}\|_2)\|\mathbf{A} - \mathbf{B}\|_2 / (2 + \sqrt{5}) = (2 + \sqrt{5}) \cdot 1 / (2 + \sqrt{5}) = 1$, a factor of about 4 larger.

20.10.5.3 C. Going deeper

Exercise. C1: The square root is continuous everywhere

Let $\mathbf{S}, \mathbf{T} \in M_n(\mathbb{C})$ be positive semidefinite.

- (a) Let λ be an eigenvalue of the Hermitian matrix $\mathbf{S} - \mathbf{T}$, with unit eigenvector $\mathbf{x}$. Prove that $\mathbf{S}^2 - \mathbf{T}^2 = (\mathbf{S} - \mathbf{T})^2 + (\mathbf{S} - \mathbf{T})\mathbf{T} + \mathbf{T}(\mathbf{S} - \mathbf{T})$, and that $\mathbf{x}^*(\mathbf{S}^2 - \mathbf{T}^2)\mathbf{x} = \lambda^2 + 2\lambda \mathbf{x}^*\mathbf{T}\mathbf{x}$.
- (b) Deduce that if $\lambda \geq 0$ then $\lambda^2 \leq \|\mathbf{S}^2 - \mathbf{T}^2\|_2$.
- (c) Prove that $\|\mathbf{P}^{1/2} - \mathbf{Q}^{1/2}\|_2 \leq \|\mathbf{P} - \mathbf{Q}\|_2^{1/2}$ for all positive semidefinite $\mathbf{P}, \mathbf{Q} \in M_n(\mathbb{C})$, and show that the exponent $\frac{1}{2}$ cannot be replaced by anything larger.

Hint: for (c), apply (b) to the pair $(\mathbf{S}, \mathbf{T})$ or to the pair $(\mathbf{T}, \mathbf{S})$, according to the sign of the eigenvalue of $\mathbf{S} - \mathbf{T}$ of largest modulus.

Solution.

- (a) Expanding,

$$\begin{aligned} & (\mathbf{S} - \mathbf{T})^2 + (\mathbf{S} - \mathbf{T})\mathbf{T} + \mathbf{T}(\mathbf{S} - \mathbf{T}) \\ &= \mathbf{S}^2 - \mathbf{S}\mathbf{T} - \mathbf{T}\mathbf{S} + \mathbf{T}^2 + \mathbf{S}\mathbf{T} - \mathbf{T}^2 + \mathbf{T}\mathbf{S} - \mathbf{T}^2 = \mathbf{S}^2 - \mathbf{T}^2. \end{aligned}$$

Since $(\mathbf{S} - \mathbf{T})\mathbf{x} = \lambda\mathbf{x}$ with λ real (Theorem 12.2.1) and $\mathbf{S} - \mathbf{T}$ Hermitian, also $\mathbf{x}^*(\mathbf{S} - \mathbf{T}) = \lambda\mathbf{x}^*$. Hence $\mathbf{x}^*(\mathbf{S} - \mathbf{T})^2\mathbf{x} = \lambda^2\mathbf{x}^*\mathbf{x} = \lambda^2$, $\mathbf{x}^*(\mathbf{S} - \mathbf{T})\mathbf{T}\mathbf{x} = \lambda\mathbf{x}^*\mathbf{T}\mathbf{x}$, and $\mathbf{x}^*\mathbf{T}(\mathbf{S} - \mathbf{T})\mathbf{x} = \lambda\mathbf{x}^*\mathbf{T}\mathbf{x}$. Adding gives the claim.

- (b) $\mathbf{x}^*\mathbf{T}\mathbf{x} \geq 0$ because $\mathbf{T} \succeq \mathbf{O}$, so if $\lambda \geq 0$, (a) gives $\lambda^2 \leq \mathbf{x}^*(\mathbf{S}^2 - \mathbf{T}^2)\mathbf{x}$. By Theorem 11.1.5 and Theorem 16.3.2 (a), $\mathbf{x}^*\mathbf{M}\mathbf{x} \leq \|\mathbf{M}\mathbf{x}\| \|\mathbf{x}\| \leq \|\mathbf{M}\|_2$ for any $\mathbf{M}$ and unit $\mathbf{x}$. So

$$\lambda^2 \leq \|\mathbf{S}^2 - \mathbf{T}^2\|_2.$$

- (c) Put $\mathbf{S} = \mathbf{P}^{1/2}$ and $\mathbf{T} = \mathbf{Q}^{1/2}$, so $\mathbf{S}^2 - \mathbf{T}^2 = \mathbf{P} - \mathbf{Q}$. By Lemma 17.3.5, $\|\mathbf{S} - \mathbf{T}\|_2 = |\lambda|$ for some eigenvalue λ of $\mathbf{S} - \mathbf{T}$. If $\lambda \geq 0$, (b) gives $\lambda^2 \leq \|\mathbf{P} - \mathbf{Q}\|_2$. If $\lambda < 0$, then $-\lambda > 0$ is an eigenvalue of $\mathbf{T} - \mathbf{S}$, and (b) applied to the pair $(\mathbf{T}, \mathbf{S})$ gives $\lambda^2 \leq \|\mathbf{T}^2 - \mathbf{S}^2\|_2 = \|\mathbf{P} - \mathbf{Q}\|_2$. In both cases $\|\mathbf{P}^{1/2} - \mathbf{Q}^{1/2}\|_2 = |\lambda| \leq \|\mathbf{P} - \mathbf{Q}\|_2^{1/2}$.

For the exponent, take $\mathbf{P} = \varepsilon \mathbf{I}$ and $\mathbf{Q} = \mathbf{O}$ with $0 < \varepsilon < 1$, as in the warning after Theorem 20.10.2. Then $\|\mathbf{P}^{1/2} - \mathbf{Q}^{1/2}\|_2 = \varepsilon^{1/2}$ and $\|\mathbf{P} - \mathbf{Q}\|_2 = \varepsilon$. If $\varepsilon^{1/2} \leq C\varepsilon^r$ held for all small ε with some $r > \frac{1}{2}$, then $C \geq \varepsilon^{1/2-r} \rightarrow \infty$ as $\varepsilon \rightarrow 0^+$, which is impossible. So the square root, which is not Lipschitz near singular matrices, is still continuous on all positive semidefinite matrices, with modulus $t \mapsto t^{1/2}$.

Exercise. C2: Both Eckart–Young lower bounds from perturbation theory

Let $\mathbf{A} \in M_{m \times n}(\mathbb{C})$, let $0 \leq k < p = \min(m, n)$, and let $\mathbf{B} \in M_{m \times n}(\mathbb{C})$ have rank at most k .

- (a) Explain why $\sigma_i(\mathbf{B}) = 0$ for $i > k$.
- (b) Prove, using Theorem 20.10.1, that $\|\mathbf{A} - \mathbf{B}\|_F^2 \geq \sum_{i=k+1}^p \sigma_i(\mathbf{A})^2$.
- (c) Prove, using Corollary 17.9.5, that $\|\mathbf{A} - \mathbf{B}\|_2 \geq \sigma_{k+1}(\mathbf{A})$.
- (d) Compare with Theorem 13.11.3 and Theorem 17.9.6.

Solution.

- (a) By Theorem 13.9.3 the number of nonzero singular values of $\mathbf{B}$ is $\text{rank } \mathbf{B} \leq k$, and they come first in the decreasing list, so $\sigma_i(\mathbf{B}) = 0$ for $i > k$.
- (b) By Theorem 20.10.1, and dropping the nonnegative terms with $i \leq k$,

$$\begin{aligned} \|\mathbf{A} - \mathbf{B}\|_F^2 &\geq \sum_{i=1}^p (\sigma_i(\mathbf{A}) - \sigma_i(\mathbf{B}))^2 \\ &\geq \sum_{i=k+1}^p (\sigma_i(\mathbf{A}) - 0)^2, \end{aligned}$$

using (a) in the last step.

- (c) By Corollary 17.9.5 with $i = k+1 \leq p$ and $\mathbf{E} = \mathbf{B} - \mathbf{A}$,

$$\sigma_{k+1}(\mathbf{A}) = |\sigma_{k+1}(\mathbf{A}) - \sigma_{k+1}(\mathbf{B})| \leq \|\mathbf{B} - \mathbf{A}\|_2 = \|\mathbf{A} - \mathbf{B}\|_2,$$

using (a) and $\sigma_{k+1}(\mathbf{A}) \geq 0$.

- (d) These are exactly the lower-bound halves of the Frobenius and the spectral Eckart–Young theorems. Both theorems also show that the bounds are attained, by the truncation $\mathbf{A}_k$, which has singular values $\sigma_1(\mathbf{A}), \dots, \sigma_k(\mathbf{A}), 0, \dots, 0$, so that it makes every term with $i \leq k$ vanish. So “the best rank- k approximation” is a perturbation theorem read backwards: no matrix of rank k can be closer to $\mathbf{A}$ than the singular values of $\mathbf{A}$ allow, because singular values cannot move further than the perturbation.

Exercise. C3: The nearest unitary to an almost-unitary matrix

Let $\mathbf{E} \in M_n(\mathbb{C})$ with $\|\mathbf{E}\|_2 < 1$, let $\mathbf{A} = \mathbf{I} + \mathbf{E}$, and let $\mathbf{U}$ be the unitary polar factor of $\mathbf{A}$.

- (a) Prove that $\sigma_n(\mathbf{A}) \geq 1 - \|\mathbf{E}\|_2 > 0$, so that $\mathbf{A}$ is invertible.
- (b) Prove that $\|\mathbf{U} - \mathbf{I}\| \leq 2\|\mathbf{E}\|/(2 - \|\mathbf{E}\|_2)$ in both the spectral and the Frobenius norm.
- (c) Deduce that if $\|\mathbf{E}\|_2 \leq \frac{1}{2}$ then $\|\mathbf{U} - \mathbf{I}\|_F \leq \frac{4}{3}\|\mathbf{E}\|_F$. Compare with the bound $\|\mathbf{U} - \mathbf{I}\|_F \leq 2\|\mathbf{E}\|_F$ that Corollary 13.10.2 and the triangle inequality give with no computation.

Solution.

- (a) The identity has all singular values equal to 1. By Corollary 17.9.5 with $i = n$, $|\sigma_n(\mathbf{I} + \mathbf{E}) - 1| \leq \|\mathbf{E}\|_2$, so $\sigma_n(\mathbf{A}) \geq 1 - \|\mathbf{E}\|_2 > 0$. An $n \times n$ matrix with n nonzero singular values has rank n (Theorem 13.9.3), so $\mathbf{A}$ is invertible.
- (b) Apply Theorem 20.10.4 to $\mathbf{A}$ and $\mathbf{B} = \mathbf{I}$, whose polar decomposition is $\mathbf{I} = \mathbf{I} \cdot \mathbf{I}$, with $\mathbf{V} = \mathbf{I}$ and $\sigma_n(\mathbf{I}) = 1$. Since $\mathbf{A} - \mathbf{I} = \mathbf{E}$,

$$\|\mathbf{U} - \mathbf{I}\| \leq \frac{2\|\mathbf{E}\|}{\sigma_n(\mathbf{A}) + 1} \leq \frac{2\|\mathbf{E}\|}{2 - \|\mathbf{E}\|_2},$$

using (a) to bound the denominator from below. This holds in both norms because the theorem does.

- (c) If $\|\mathbf{E}\|_2 \leq \frac{1}{2}$ then $2 - \|\mathbf{E}\|_2 \geq \frac{3}{2}$, and (b) in the Frobenius norm gives $\|\mathbf{U} - \mathbf{I}\|_F \leq \frac{4}{3}\|\mathbf{E}\|_F$.

For comparison, Corollary 13.10.2 says that $\mathbf{U}$ is at least as close to $\mathbf{A}$ as any unitary matrix is, and $\mathbf{I}$ is one, so $\|\mathbf{A} - \mathbf{U}\|_F \leq \|\mathbf{A} - \mathbf{I}\|_F = \|\mathbf{E}\|_F$. The triangle inequality (Corollary 11.1.6) then gives

$$\|\mathbf{U} - \mathbf{I}\|_F \leq \|\mathbf{U} - \mathbf{A}\|_F + \|\mathbf{A} - \mathbf{I}\|_F \leq 2\|\mathbf{E}\|_F.$$

So the soft argument already locates $\mathbf{U}$ within $2\|\mathbf{E}\|_F$ of $\mathbf{I}$. Part (b) improves the constant to $\frac{4}{3}$ when $\|\mathbf{E}\|_2 \leq \frac{1}{2}$, and to $2/(2 - \|\mathbf{E}\|_2)$, which tends to 1, as $\mathbf{E} \rightarrow \mathbf{O}$. It also gives the spectral-norm statement, about which Corollary 13.10.2 says nothing.