

21.6 Ky Fan Dominance

Three earlier chapters have deferred the same kind of statement to this one: that a matrix is closest to another in **every** unitarily invariant norm at once. Such a statement looks like infinitely many theorems, one for each norm. This section shows that it is one theorem about singular values, and that it can be tested on a short finite list of norms. Then it collects the debts.

21.6.1 One weak majorization, every norm

Section 3 proved that a symmetric gauge respects weak majorization, and Section 4 matched the gauges with the unitarily invariant norms. Put together, they say that a weak majorization of singular values forces an inequality of norms. What is new here is the **converse**: nothing else forces it, and the Ky Fan k -norms already see everything.

To compare two matrices in every unitarily invariant norm, compare the running totals of their singular values.

Theorem 21.6.1. Ky Fan Dominance Theorem

Let $\mathbf{A}, \mathbf{B} \in M_{m \times n}(F)$ with $F = \mathbb{R}$ or $\mathbb{C}$, and let $p = \min(m, n)$. The following are equivalent.

- (a) $\sigma(\mathbf{A}) \prec_w \sigma(\mathbf{B})$, that is, $\sum_{i \leq k} \sigma_i(\mathbf{A}) \leq \sum_{i \leq k} \sigma_i(\mathbf{B})$ for every $k = 1, \dots, p$.
- (b) $\|\mathbf{A}\| \leq \|\mathbf{B}\|$ for **every** unitarily invariant norm $\|\cdot\|$ on $M_{m \times n}(F)$.

Idea.

- (a) $\Rightarrow$ (b) is the dictionary of Sections 3 and 4 read in one direction: a unitarily invariant norm is a symmetric gauge evaluated at the singular values, and a symmetric gauge is monotone under weak majorization of nonnegative vectors. For (b) $\Rightarrow$ (a) we need, for each k , a unitarily invariant norm that computes the k -th running total. The Ky Fan k -norm is exactly that, so the p inequalities in (a) are p of the infinitely many inequalities in (b).

Proof.

- (a) $\Rightarrow$ (b). Let $\|\cdot\|$ be a unitarily invariant norm and let Φ be its symmetric gauge, so that $\|\mathbf{M}\| = \Phi(\sigma(\mathbf{M}))$ for every $\mathbf{M}$ (Theorem 21.4.3). The vectors $\sigma(\mathbf{A})$ and $\sigma(\mathbf{B})$ have nonnegative entries (Definition 13.9.1), so Theorem 21.3.4 (b) applies to the hypothesis $\sigma(\mathbf{A}) \prec_w \sigma(\mathbf{B})$ and gives $\Phi(\sigma(\mathbf{A})) \leq \Phi(\sigma(\mathbf{B}))$, which is $\|\mathbf{A}\| \leq \|\mathbf{B}\|$.
- (b) $\Rightarrow$ (a). Fix $k \in \{1, \dots, p\}$. Section 3's k -th Ky Fan gauge $\Phi_k(\mathbf{x}) = \sum_{i \leq k} |x|_i^\downarrow$ (Definition 21.3.2) is a symmetric gauge function, so by Theorem 21.4.3

$$\|\mathbf{M}\|_{(k)} := \Phi_k(\sigma(\mathbf{M})) = \sum_{i=1}^k \sigma_i(\mathbf{M})$$

is a unitarily invariant norm on $M_{m \times n}(F)$ — the Ky Fan k -norm of Definition 21.4.2 — the second equality holding because $\sigma(\mathbf{M})$ is already nonnegative and decreasing, so that its k largest absolute values are its first k entries. Applying (b) to this one norm

gives

$$\sum_{i=1}^k \sigma_i(\mathbf{A}) \leq \sum_{i=1}^k \sigma_i(\mathbf{B}).$$

As k was arbitrary and both lists are decreasing, this is (a), by Definition 17.8.1. This proves the theorem. ■

The theorem is a translation device, and it will be used in one direction almost every time: to prove an inequality for every unitarily invariant norm, produce the weak majorization. Note also what it says about testing: the condition “for every unitarily invariant norm” is equivalent to “for the p Ky Fan norms”, so a statement quantified over an infinite family of norms is decided by finitely many numbers.

Warning. One norm is never enough. For $\mathbf{A} = \text{diag}(1, 1)$ and $\mathbf{B} = \text{diag}(1, 0)$ the spectral norms agree, $\|\mathbf{A}\|_2 = \|\mathbf{B}\|_2 = 1$, yet $\sigma(\mathbf{A}) = (1, 1)$ is not weakly majorized by $\sigma(\mathbf{B}) = (1, 0)$, and indeed the trace norms are 2 and 1. Dominance needs **all** the running totals, and it compares matrices of the **same size**, since otherwise the two sides are norms on different spaces.

Quick check. Let $\mathbf{A}, \mathbf{B} \in M_{m \times n}(F)$ satisfy $\sigma_i(\mathbf{A}) \leq \sigma_i(\mathbf{B})$ for every i . Does $\|\mathbf{A}\| \leq \|\mathbf{B}\|$ hold for every unitarily invariant norm?

Solution. Yes. Entrywise domination of two decreasing nonnegative lists gives $\sum_{i \leq k} \sigma_i(\mathbf{A}) \leq \sum_{i \leq k} \sigma_i(\mathbf{B})$ for every k , which is $\sigma(\mathbf{A}) \prec_w \sigma(\mathbf{B})$; now apply [Theorem 21.6.1](#). The converse fails: $\sigma(\mathbf{A}) = (1, 1)$ and $\sigma(\mathbf{B}) = (2, 0)$ satisfy $\sigma(\mathbf{A}) \prec_w \sigma(\mathbf{B})$ with $\sigma_2(\mathbf{A}) > \sigma_2(\mathbf{B})$.

21.6.2 Low-rank approximation in every unitarily invariant norm

Chapter 13 §11 proved that the truncated singular value decomposition $\mathbf{A}_k$ is the best rank- k approximation in the Frobenius norm, and Chapter 17 §09 proved the same for the spectral norm. Both promised more. Chapter 13 §11 said that the two statements are “instances of one theorem: $\mathbf{A}_k$ is a best rank- k approximation in **every** unitarily invariant norm”, and predicted the shape of the proof: “it needs inequalities between the singular values of $\mathbf{A}$ and of $\mathbf{A} - \mathbf{B}$, not a projection argument”. Chapter 16 §08 recorded the same debt — “The version covering every unitarily invariant norm at once is Chapter 21’s” — and Chapter 17 §09 said that the common minimizer “is optimal in every unitarily invariant norm at once, which is Chapter 21’s theorem”. Here it is, by the predicted route.

Corollary 21.6.2. Eckart–Young Theorem in Every Unitarily Invariant Norm

Let $\mathbf{A} \in M_{m \times n}(F)$ with singular values $\sigma_1 \geq \cdots \geq \sigma_p$, $p = \min(m, n)$, let $0 \leq k \leq p$, and let $\mathbf{A}_k$ be a truncation of Definition 13.11.1. Then for **every** $\mathbf{B} \in M_{m \times n}(F)$ with $\text{rank } \mathbf{B} \leq k$

and **every** unitarily invariant norm on $M_{m \times n}(F)$,

$$\|\mathbf{A} - \mathbf{A}_k\| \leq \|\mathbf{A} - \mathbf{B}\|,$$

and the minimum value is $\|\mathbf{A} - \mathbf{A}_k\| = \Phi(\sigma_{k+1}, \dots, \sigma_p, 0, \dots, 0)$, where Φ is the symmetric gauge of the norm.

Idea. The singular values of the error $\mathbf{A} - \mathbf{A}_k$ are the discarded $\sigma_{k+1}, \dots, \sigma_p$, padded with zeros. For an arbitrary competitor $\mathbf{B}$ of rank at most k , Weyl's inequality for singular values, applied with the splitting $\mathbf{A} = (\mathbf{A} - \mathbf{B}) + \mathbf{B}$, says that $\mathbf{B}$ can shift the singular value list of $\mathbf{A}$ by at most k places, because $\sigma_{k+1}(\mathbf{B}) = 0$. That is precisely $\sigma_i(\mathbf{A} - \mathbf{B}) \geq \sigma_{k+i}(\mathbf{A})$: the truncation wins index by index, which is more than dominance needs.

Proof. The error of the truncation. Fix the singular value decomposition $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^*$ (Theorem 13.9.1) used to form $\mathbf{A}_k = \sum_{i \leq k} \sigma_i \mathbf{u}_i \mathbf{v}_i^*$ (Definition 13.11.1). Then

$$\mathbf{A} - \mathbf{A}_k = \sum_{i > k} \sigma_i \mathbf{u}_i \mathbf{v}_i^* = \mathbf{U}\Sigma^{(k)}\mathbf{V}^*,$$

where $\Sigma^{(k)}$ is Σ with its first k diagonal entries replaced by 0. Since $\mathbf{U}$ and $\mathbf{V}$ are unitary, $(\mathbf{A} - \mathbf{A}_k)^*(\mathbf{A} - \mathbf{A}_k) = \mathbf{V}(\Sigma^{(k)})^*\Sigma^{(k)}\mathbf{V}^*$ has eigenvalues $\sigma_{k+1}^2, \dots, \sigma_p^2$ together with zeros, so by Definition 13.9.1

$$\sigma_i(\mathbf{A} - \mathbf{A}_k) = \sigma_{k+i}(\mathbf{A}) \quad (1 \leq i \leq p), \quad (21.6.1)$$

with the convention $\sigma_j(\mathbf{A}) = 0$ for **every** $j > p$. Chapter 17 §09 states it for $p < j \leq n$; we extend it here, because the indices $k+i$ above run as far as $2p$, which may exceed n .

The competitor. Let $\text{rank } \mathbf{B} \leq k$. Then $\sigma_{k+1}(\mathbf{B}) = 0$, since a matrix has exactly $\text{rank } \mathbf{B}$ nonzero singular values (Theorem 13.9.1). Fix $i \leq p$. If $i+k > n$, then $\sigma_{k+i}(\mathbf{A}) = 0 \leq \sigma_i(\mathbf{A} - \mathbf{B})$ by the convention. Otherwise $i + (k+1) - 1 = i+k \leq n$, so Theorem 17.9.4 applies to the sum $\mathbf{A} = (\mathbf{A} - \mathbf{B}) + \mathbf{B}$ with indices i and $k+1$ and gives

$$\sigma_{k+i}(\mathbf{A}) \leq \sigma_i(\mathbf{A} - \mathbf{B}) + \sigma_{k+1}(\mathbf{B}) = \sigma_i(\mathbf{A} - \mathbf{B}).$$

Either way, with Eq. 21.6.1,

$$\sigma_i(\mathbf{A} - \mathbf{A}_k) \leq \sigma_i(\mathbf{A} - \mathbf{B}) \quad (1 \leq i \leq p).$$

Summing the first k' of these for each $k' \leq p$ gives $\sigma(\mathbf{A} - \mathbf{A}_k) \prec_w \sigma(\mathbf{A} - \mathbf{B})$, so Theorem 21.6.1 gives $\|\mathbf{A} - \mathbf{A}_k\| \leq \|\mathbf{A} - \mathbf{B}\|$ for every unitarily invariant norm.

The value. By Theorem 21.4.3 and Eq. 21.6.1,

$$\|\mathbf{A} - \mathbf{A}_k\| = \Phi(\sigma(\mathbf{A} - \mathbf{A}_k)) = \Phi(\sigma_{k+1}, \dots, \sigma_p, 0, \dots, 0).$$

This proves the corollary. ■

The proof gives more than was asked: the error of the truncation is smaller than that of any competitor **singular value by singular value**, not merely in every norm. Specializing Φ recovers the two known cases — $\Phi = \|\cdot\|_\infty$ gives σ_{k+1} , which is Theorem 17.9.6, and $\Phi = \|\cdot\|_2$ gives $(\sum_{i>k} \sigma_i^2)^{1/2}$, which is Theorem 13.11.3 — and now every other norm comes with them.

Example. One matrix, four norms

Let

$$\mathbf{A} = \begin{pmatrix} 6 & -4 & 0 \\ 8 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \in M_3(\mathbb{R}).$$

Find $\mathbf{A}_1$ and $\mathbf{A}_2$, and tabulate the approximation errors in the spectral, Frobenius, trace, and Ky Fan 2-norms. Compare $\mathbf{A}_1$ with the rank-one competitor $\mathbf{C} = \text{diag}(0, 0, 1)$.

Solution. Write $\mathbf{Q} = \frac{1}{5} \begin{pmatrix} 3 & -4 & 0 \\ 4 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}$, which is orthogonal, since its columns are orthonormal. Then $\mathbf{A} = \mathbf{Q} \text{diag}(10, 5, 1)$, so this is a singular value decomposition with $\mathbf{U} = \mathbf{Q}$, $\mathbf{V} = \mathbf{I}$, and $\sigma(\mathbf{A}) = (10, 5, 1)$. Hence

$$\mathbf{A}_1 = \begin{pmatrix} 6 & 0 & 0 \\ 8 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{A}_2 = \begin{pmatrix} 6 & -4 & 0 \\ 8 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

with errors $\mathbf{A} - \mathbf{A}_1 = \mathbf{Q} \text{diag}(0, 5, 1)$ and $\mathbf{A} - \mathbf{A}_2 = \mathbf{Q} \text{diag}(0, 0, 1)$, whose singular value lists are $(5, 1, 0)$ and $(1, 0, 0)$, as [Eq. 21.6.1](#) predicts.

error measured in	$\mathbf{A} - \mathbf{A}_1$	$\mathbf{A} - \mathbf{A}_2$	$\mathbf{A} - \mathbf{C}$
spectral $\ \cdot\ _2$	5	1	10
Frobenius $\ \cdot\ _F$	$\sqrt{26}$	1	$\sqrt{125}$
trace norm	6	1	15
Ky Fan 2-norm (Definition 21.4.2)	6	1	15

The competitor $\mathbf{C} = \text{diag}(0, 0, 1)$ has rank 1, and $\mathbf{A} - \mathbf{C} = \mathbf{A}_2 = \mathbf{Q} \text{diag}(10, 5, 0)$ has singular values $(10, 5, 0)$. It loses to $\mathbf{A}_1$ in all four norms at once, as it must: $(5, 1, 0)$ is dominated entry by entry by $(10, 5, 0)$. Note also that the Ky Fan 2-norm and the trace norm agree on these three errors, because each has at most two nonzero singular values.

21.6.3 The nearest unitary matrix

Chapter 13 §10 proved that the unitary polar factor $\mathbf{W}$ of $\mathbf{A} = \mathbf{W}|\mathbf{A}|$ is the nearest unitary matrix in the Frobenius norm (Corollary 13.10.2) and warned: “It is true, though not proved here, that $\mathbf{W}$ also minimizes the distance in every unitarily invariant norm; Chapter 21 takes up that family of norms.” That is this corollary.

The proof passes through a symmetric list of the form $(z_1, \dots, z_q, 0, \dots, 0, -z_q, \dots, -z_1)$, and has to read off its top- k sums. Section 7 meets the same list again, so we settle the counting once.

Lemma 21.6.3. Top Sums of a Symmetric List

Let $z_1, \dots, z_q$ be real numbers, let $r \geq 0$ be an integer, and let $\mathbf{c} \in \mathbb{R}^{2q+r}$ be the list

$$\mathbf{c} = (z_1, \dots, z_q, \underbrace{0, \dots, 0}_r, -z_q, \dots, -z_1).$$

Then for every $k \leq q$, the largest sum of k entries of $\mathbf{c}$, which is the running total $\sum_{j \leq k} c_j^\downarrow$, equals the sum of the k largest of the numbers $|z_1|, \dots, |z_q|$.

Proof. Fix $k \leq q$. Write S for the sum of the k largest of the $|z_i|$. Since the largest sum of k entries of $\mathbf{c}$ is obtained by taking the k largest entries, it is $\sum_{j \leq k} c_j^\downarrow$, and the two descriptions agree.

($\geq$) For each index i , the number $|z_i|$ is an entry of $\mathbf{c}$: it is z_i in position i if $z_i \geq 0$, and $-z_i$ in position $2q + r + 1 - i$ otherwise. Distinct indices i give distinct positions, so the k largest of the $|z_i|$ occur at k distinct positions of $\mathbf{c}$, and their sum S is one of the sums being maximized.

($\leq$) Let k positions be given. The entries in the middle range contribute 0, and the remaining positions belong to at most k of the indices i . For one such index, the chosen positions contribute z_i , or $-z_i$, or (if both are chosen) 0, hence at most $|z_i|$ in every case. So the sum of the chosen entries is at most $\sum_{i \in J} |z_i|$ for a set J of at most k indices, and since every $|z_i| \geq 0$ that is at most S . This proves the lemma. ■

Corollary 21.6.4. The Polar Factor Is Nearest in Every Unitarily Invariant Norm

Let $\mathbf{A} \in M_n(\mathbb{C})$ with polar decomposition $\mathbf{A} = \mathbf{W}|\mathbf{A}|$ (Theorem 13.10.1). Then for every unitary $\mathbf{Q} \in M_n(\mathbb{C})$ and every unitarily invariant norm on $M_n(\mathbb{C})$,

$$\|\mathbf{A} - \mathbf{W}\| \leq \|\mathbf{A} - \mathbf{Q}\|.$$

Idea. Peel off $\mathbf{W}$. Unitary invariance turns the comparison of $\mathbf{A} - \mathbf{Q}$ with $\mathbf{A} - \mathbf{W}$ into a comparison of $|\mathbf{A}| - \mathbf{U}$ with $|\mathbf{A}| - \mathbf{I}$, where $\mathbf{U} = \mathbf{W}^*\mathbf{Q}$ runs over all unitary matrices: the question is why the identity is the unitary matrix closest to a positive semidefinite one. By dominance it is enough to weakly majorize singular values, and singular values of a difference of two matrices are eigenvalues of a difference of two Hermitian ones, through the dilation of Chapter 17 §09. There, Lidskii's theorem is waiting: it already says that the vector of eigenvalue shifts is majorized by the eigenvalues of the difference, and choosing the right k entries of that majorization is the whole computation.

Proof. Step 1: reduce to $|\mathbf{A}|$ and $\mathbf{I}$. Let $\mathbf{Q}$ be unitary and put $\mathbf{U} := \mathbf{W}^*\mathbf{Q}$, which is unitary. Then

$$\mathbf{A} - \mathbf{Q} = \mathbf{W}|\mathbf{A}| - \mathbf{Q} = \mathbf{W}(|\mathbf{A}| - \mathbf{W}^*\mathbf{Q}) = \mathbf{W}(|\mathbf{A}| - \mathbf{U}),$$

so $\|\mathbf{A} - \mathbf{Q}\| = \|\mathbf{A} - \mathbf{U}\|$ by unitary invariance (Definition 21.4.1). Taking $\mathbf{Q} = \mathbf{W}$ gives $\mathbf{U} = \mathbf{I}$ and $\|\mathbf{A} - \mathbf{W}\| = \|\mathbf{A} - \mathbf{I}\|$. Writing $\mathbf{S} = |\mathbf{A}| \succeq \mathbf{O}$, it suffices to prove

$$\|\mathbf{S} - \mathbf{I}\| \leq \|\mathbf{S} - \mathbf{U}\| \quad \text{for every unitary } \mathbf{U}.$$

By Theorem 21.6.1 this is the weak majorization $\sigma(\mathbf{S} - \mathbf{I}) \prec_w \sigma(\mathbf{S} - \mathbf{U})$.

Step 2: the singular values of $\mathbf{S} - \mathbf{I}$. Let $t_1 \geq \dots \geq t_n \geq 0$ be the eigenvalues of $\mathbf{S}$, which are nonnegative because $\mathbf{S} \succeq \mathbf{O}$ (Theorem 13.1.4). The spectral theorem writes $\mathbf{S} = \mathbf{Y} \text{diag}(t_1, \dots, t_n) \mathbf{Y}^*$ with $\mathbf{Y}$ unitary, and this is a singular value decomposition, so $\sigma_i(\mathbf{S}) = t_i$. The matrix $\mathbf{S} - \mathbf{I}$ is Hermitian with eigenvalues $t_i - 1$, and $(\mathbf{S} - \mathbf{I})^*(\mathbf{S} - \mathbf{I}) = (\mathbf{S} - \mathbf{I})^2$ has eigenvalues $(t_i - 1)^2$, so by Definition 13.9.1

$$\sigma(\mathbf{S} - \mathbf{I}) = (|t_1 - 1|, \dots, |t_n - 1|)^\downarrow.$$

Write $s_j = \sigma_j(\mathbf{S} - \mathbf{U})$ for the numbers we must bound from below.

Step 3: dilate. Let $\mathcal{H}(\cdot)$ be the Hermitian dilation of Proposition 17.9.3. Conjugate transposition is additive, so $\mathcal{H}(\mathbf{S}) - \mathcal{H}(\mathbf{U}) = \mathcal{H}(\mathbf{S} - \mathbf{U})$, and all three are Hermitian matrices in $M_{2n}(\mathbb{C})$. By Proposition 17.9.3, the eigenvalues of $\mathcal{H}(\mathbf{M})$ are $\pm\sigma_i(\mathbf{M})$, and since the σ_i are nonnegative and decreasing, the decreasing list is

$$\lambda(\mathcal{H}(\mathbf{M})) = (\sigma_1(\mathbf{M}), \dots, \sigma_n(\mathbf{M}), -\sigma_n(\mathbf{M}), \dots, -\sigma_1(\mathbf{M})).$$

Applied to $\mathbf{S}$, to $\mathbf{U}$ (all of whose singular values are 1) and to $\mathbf{S} - \mathbf{U}$, this gives the three lists

$$\begin{aligned} \lambda(\mathcal{H}(\mathbf{S})) &= (t_1, \dots, t_n, -t_n, \dots, -t_1), \\ \lambda(\mathcal{H}(\mathbf{U})) &= (1, \dots, 1, -1, \dots, -1), \\ \lambda(\mathcal{H}(\mathbf{S} - \mathbf{U})) &= (s_1, \dots, s_n, -s_n, \dots, -s_1). \end{aligned}$$

Step 4: apply Lidskii's theorem. Chapter 17 §07 proved, in the remark following Theorem 17.7.2, that $\lambda(\mathbf{X}) - \lambda(\mathbf{Y}) \prec \lambda(\mathbf{X} - \mathbf{Y})$ for Hermitian $\mathbf{X}, \mathbf{Y}$. Applied to $\mathbf{X} = \mathcal{H}(\mathbf{S})$ and $\mathbf{Y} = \mathcal{H}(\mathbf{U})$,

$$\lambda(\mathcal{H}(\mathbf{S})) - \lambda(\mathcal{H}(\mathbf{U})) \prec \lambda(\mathcal{H}(\mathbf{S} - \mathbf{U})).$$

By Step 3 the vector on the left is

$$\mathbf{d} = (t_1 - 1, \dots, t_n - 1, -(t_n - 1), \dots, -(t_1 - 1)) \in \mathbb{R}^{2n},$$

which is the symmetric list of Lemma 21.6.3 with $q = n$, $r = 0$, and $z_i = t_i - 1$. Fix $k \leq n$. By that lemma the k -th running total of $\mathbf{d}^\downarrow$ is the sum of the k largest of the numbers $|t_i - 1|$, which by Step 2 is $\sum_{j \leq k} \sigma_j(\mathbf{S} - \mathbf{I})$. The right-hand list is decreasing with nonnegative first half, so its k -th running total is $\sum_{j \leq k} s_j$. So (M1) of Definition 17.8.1, at the index k , reads

$$\sum_{j=1}^k \sigma_j(\mathbf{S} - \mathbf{I}) \leq \sum_{j=1}^k s_j.$$

As this holds for every $k \leq n$, it is the weak majorization required in Step 1. This proves the corollary. ■

The mechanism deserves a name, because Section 7 runs it again in general: **a Hermitian**

dilation turns a statement about singular values into one about eigenvalues, where Lidskii's theorem applies. What was special here is that both matrices were **square**, so that their dilations had no zeros in the middle of the spectrum and no bookkeeping was needed; that one of the two was unitary is what made its dilated eigenvalue list the constant ± 1 .

Remark. The inequality of Corollary 21.6.4 cannot be improved to $\sigma_j(\mathbf{A} - \mathbf{W}) \leq \sigma_j(\mathbf{A} - \mathbf{Q})$ index by index. Take $\mathbf{A} = \mathbf{S} = \text{diag}(2, 0)$, which is positive semidefinite, so that $|\mathbf{A}| = \mathbf{A}$ and $\mathbf{W} = \mathbf{I}$ is a polar factor, and take $\mathbf{Q} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Then the matrix $\mathbf{S} - \mathbf{Q} = \begin{pmatrix} 2 & -1 \\ -1 & 0 \end{pmatrix}$ is symmetric with eigenvalues $1 \pm \sqrt{2}$, so $\sigma(\mathbf{S} - \mathbf{Q}) = (1 + \sqrt{2}, \sqrt{2} - 1) \approx (2.414, 0.414)$, while $\sigma(\mathbf{S} - \mathbf{I}) = (1, 1)$. The second singular value is smaller for $\mathbf{Q}$, and only the running totals, $1 \leq 2.414$ and $2 \leq 2.828$, go the right way. So the weak majorization, and not something cruder, is what the corollary needs.

21.6.4 Two smaller consequences

A compression never increases a unitarily invariant norm.

Corollary 21.6.5. Compressions Are Contractions

Let $\mathbf{A} \in M_n(F)$ and let $\mathbf{P} \in M_n(F)$ be an orthogonal projection. Then $\|\mathbf{PAP}\| \leq \|\mathbf{A}\|$ for every unitarily invariant norm.

Proof. If $\mathbf{P} = \mathbf{O}$ the left side is $\|\mathbf{O}\| = 0 \leq \|\mathbf{A}\|$. Otherwise $\mathbf{P}$ is Hermitian with eigenvalues in $\{0, 1\}$, at least one of them equal to 1, so $\|\mathbf{P}\|_2 = 1$ by Lemma 17.3.5. Applying the two bounds of Proposition 21.4.4 (b) in turn,

$$\|\mathbf{PAP}\| \leq \|\mathbf{P}\|_2 \|\mathbf{AP}\| \leq \|\mathbf{P}\|_2 \|\mathbf{A}\| \|\mathbf{P}\|_2 = \|\mathbf{A}\|,$$

as claimed. ■

And a negative result, about the subadditivity of singular values. Section 4 proved $\sigma(\mathbf{A} + \mathbf{B}) \prec_w \sigma(\mathbf{A}) + \sigma(\mathbf{B})$ (Lemma 21.4.2), where the right-hand side is the entrywise sum of the two decreasing lists. The weak majorization cannot be upgraded to majorization.

Example. Weak majorization is the most that holds

Let $\mathbf{A} = \text{diag}(1, 0)$ and $\mathbf{B} = \text{diag}(-1, 0)$. Then $\sigma(\mathbf{A}) = \sigma(\mathbf{B}) = (1, 0)$, so $\sigma(\mathbf{A}) + \sigma(\mathbf{B}) = (2, 0)$, while $\mathbf{A} + \mathbf{B} = \mathbf{O}$ has $\sigma(\mathbf{A} + \mathbf{B}) = (0, 0)$. The totals are 0 and 2, so $\sigma(\mathbf{A} + \mathbf{B}) \prec \sigma(\mathbf{A}) + \sigma(\mathbf{B})$ is false, although the weak majorization holds with room to spare.

The failure is not an accident of the example. Take $\mathbf{B} = -\mathbf{A}$ for any $\mathbf{A} \neq \mathbf{O}$: then $\sigma(\mathbf{B}) = \sigma(\mathbf{A})$, so the right-hand list has total $2 \sum_i \sigma_i(\mathbf{A}) > 0$, while $\sigma(\mathbf{A} + \mathbf{B}) = \mathbf{0}$ has total 0. Cancellation destroys singular value mass, and clause (M2) of Definition 17.8.1 — equal totals — therefore fails for every nonzero $\mathbf{A}$. Weak majorization is the most Lemma 21.4.2 can assert.

21.6.5 Exercises

21.6.5.1 A. Check your understanding

Exercise. A1

- (a) State [Theorem 21.6.1](#), and say which family of norms proves the direction that is not a direct consequence of Sections 3 and 4.
- (b) Define the Ky Fan k -norm and say why it is a unitarily invariant norm.
- (c) Determine whether the following is true, and justify your answer: if $\|\mathbf{A}\|_F \leq \|\mathbf{B}\|_F$ and $\|\mathbf{A}\|_2 \leq \|\mathbf{B}\|_2$, then $\|\mathbf{A}\| \leq \|\mathbf{B}\|$ for every unitarily invariant norm.
- (d) In the proof of [Corollary 21.6.2](#), where is the hypothesis $\text{rank } \mathbf{B} \leq k$ used?
- (e) Determine whether the following is true, and justify your answer: for $\mathbf{A} \in M_n(\mathbb{C})$ invertible, the nearest unitary matrix in the trace norm is the same as the nearest one in the Frobenius norm.

Solution.

- (a) For $\mathbf{A}, \mathbf{B}$ of the same size, $\sigma(\mathbf{A}) \prec_w \sigma(\mathbf{B})$ if and only if $\|\mathbf{A}\| \leq \|\mathbf{B}\|$ for every unitarily invariant norm. The direction (b) $\Rightarrow$ (a) is the one needing new input, and it is proved by applying the hypothesis to the Ky Fan k -norms.
- (b) $\|\mathbf{M}\|_{(k)} = \sum_{i \leq k} \sigma_i(\mathbf{M})$ (Definition 21.4.2). It is the unitarily invariant norm attached by Theorem 21.4.3 to the k -th Ky Fan gauge Definition 21.3.2, evaluated at $\sigma(\mathbf{M})$.
- (c) False. Two norms do not decide the family. Take $\mathbf{A} = \text{diag}(1, 1, 0)$ and $\mathbf{B} = \text{diag}(\sqrt{2}, 0, 0)$: then $\|\mathbf{A}\|_F = \sqrt{2} = \|\mathbf{B}\|_F$ and $\|\mathbf{A}\|_2 = 1 \leq \sqrt{2} = \|\mathbf{B}\|_2$, but the trace norms are 2 and $\sqrt{2}$, so $\|\mathbf{A}\| \leq \|\mathbf{B}\|$ fails for the trace norm.
- (d) Only to give $\sigma_{k+1}(\mathbf{B}) = 0$, which is what makes the Weyl inequality with second index $k+1$ say something about $\mathbf{A} - \mathbf{B}$ alone.
- (e) False: in the trace norm the minimizer need not be unique, so there is no *the* nearest unitary matrix. The polar factor $\mathbf{W}$ is nearest in both norms, by [Corollary 21.6.4](#), and for invertible $\mathbf{A}$ it is the **only** Frobenius minimizer (Corollary 13.10.2); in the trace norm it can share the minimum. Take $\mathbf{A} = \text{diag}(2, \frac{1}{2})$, which is positive definite, so $|\mathbf{A}| = \mathbf{A}$ and $\mathbf{W} = \mathbf{I}$. Then $\mathbf{A} - \mathbf{I} = \text{diag}(1, -\frac{1}{2})$ has trace norm $\frac{3}{2}$. The rotation $\mathbf{Q} = \frac{1}{5} \begin{pmatrix} 4 & -3 \\ 3 & 4 \end{pmatrix}$ is unitary, and

$$\mathbf{A} - \mathbf{Q} = \frac{1}{10} \begin{pmatrix} 12 & 6 \\ -6 & -3 \end{pmatrix}$$

has determinant 0 and is nonzero, so it has rank 1: its second singular value is 0, and the first is $\|\mathbf{A} - \mathbf{Q}\|_F = \frac{1}{10} \sqrt{144 + 36 + 36 + 9} = \frac{3}{2}$. Its trace norm is therefore $\frac{3}{2}$ as well, so $\mathbf{I}$ and $\mathbf{Q}$ are both nearest in the trace norm. In the Frobenius norm they are not tied: $\|\mathbf{A} - \mathbf{I}\|_F = \frac{\sqrt{5}}{2} < \frac{3}{2} = \|\mathbf{A} - \mathbf{Q}\|_F$.

21.6.5.2 B. Practice

Exercise. B1: Dominance by hand

Let $\mathbf{A}, \mathbf{B} \in M_3(\mathbb{R})$ have singular values $\sigma(\mathbf{A}) = (3, 2, 2)$ and $\sigma(\mathbf{B}) = (4, 3, 1)$.

- (a) Determine whether $\|\mathbf{A}\| \leq \|\mathbf{B}\|$ for every unitarily invariant norm.
 (b) Answer the same question for $\sigma(\mathbf{A}') = (3, 3, 2)$ in place of $\sigma(\mathbf{A})$.

Solution.

- (a) The running totals are 3, 5, 7 for $\mathbf{A}$ and 4, 7, 8 for $\mathbf{B}$, and $3 \leq 4$, $5 \leq 7$, $7 \leq 8$. So $\sigma(\mathbf{A}) \prec_w \sigma(\mathbf{B})$ and [Theorem 21.6.1](#) gives the inequality for every unitarily invariant norm.
 (b) The running totals of $\sigma(\mathbf{A}')$ are 3, 6, 8, all at most 4, 7, 8, so again yes. This one is worth noticing, because the index-by-index comparison **fails**: $\sigma_3(\mathbf{A}') = 2 > 1 = \sigma_3(\mathbf{B})$. Dominance is a statement about running totals, and it is strictly weaker than entrywise domination.

Exercise. B2: Best approximations of a 2 by 2 matrix

Let $\mathbf{A} = \begin{pmatrix} 3 & 4 \\ 4 & -3 \end{pmatrix}$. Compute $\sigma(\mathbf{A})$, find the best rank-one approximation, and compute the error in the spectral, Frobenius, and trace norms. What is unusual about this matrix?

Solution. $\mathbf{A}^\top \mathbf{A} = \begin{pmatrix} 25 & 0 \\ 0 & 25 \end{pmatrix}$, so $\sigma(\mathbf{A}) = (5, 5)$. By [Corollary 21.6.2](#) the best rank-one approximation is $\mathbf{A}_1 = 5\mathbf{u}_1\mathbf{v}_1^*$ for a singular value decomposition of $\mathbf{A}$. One such decomposition is $\mathbf{A} = \mathbf{U} \operatorname{diag}(5, 5) \mathbf{V}^\top$ with $\mathbf{V} = \mathbf{I}$ and $\mathbf{U} = \mathbf{A}/5$, which is orthogonal because $\mathbf{A}^\top \mathbf{A} = 25\mathbf{I}$; it gives $\mathbf{u}_1 = \frac{1}{5}(3, 4)$, $\mathbf{v}_1 = \mathbf{e}_1$ and

$$\mathbf{A}_1 = \begin{pmatrix} 3 & 0 \\ 4 & 0 \end{pmatrix}, \quad \mathbf{A} - \mathbf{A}_1 = \begin{pmatrix} 0 & 4 \\ 0 & -3 \end{pmatrix}.$$

The error has singular values $(\sigma_2, 0) = (5, 0)$, as [Corollary 21.6.2](#) predicts: spectral error 5, Frobenius error 5, trace-norm error 5. What is unusual is that $\sigma_1 = \sigma_2$, so the truncation is not unique: $\mathbf{A}/5$ is symmetric and orthogonal, and **every** unit vector is a right singular vector. Every rank-one matrix of the form $5\mathbf{u}\mathbf{v}^*$ arising from a singular value decomposition is a best approximation, and they are genuinely different matrices.

Exercise. B3: A compression

Let $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$ and let $\mathbf{P} = \mathbf{e}_1\mathbf{e}_1^\top$. Compute $\|\mathbf{PAP}\|$ and $\|\mathbf{A}\|$ for the trace norm, and check [Corollary 21.6.5](#).

Solution. $\mathbf{PAP} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$, whose singular values are $(2, 0)$, so its trace norm is 2. The matrix $\mathbf{A}$ is symmetric with $\operatorname{tr} \mathbf{A} = 5$ and $\det \mathbf{A} = 5$, so its eigenvalues are $(5 \pm \sqrt{5})/2$, both positive; hence $\mathbf{A} \succ \mathbf{O}$, its singular values are its eigenvalues, and its trace norm is $\operatorname{tr} \mathbf{A} = 5$. Indeed $2 \leq 5$.

21.6.5.3 C. Going deeper

Exercise. C1: Dominance for sums

Let $\mathbf{A}, \mathbf{B} \in M_{m \times n}(\mathbb{C})$.

- (a) Prove that $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$ follows from Lemma 21.4.2 together with Theorem 21.3.4 (b), without using that $\|\cdot\|$ is a norm.
- (b) Deduce that $\sum_{i \leq k} \sigma_i(\mathbf{A} + \mathbf{B}) \leq \sum_{i \leq k} \sigma_i(\mathbf{A}) + \sum_{i \leq k} \sigma_i(\mathbf{B})$ for every k .

Solution.

- (a) By Lemma 21.4.2, $\sigma(\mathbf{A} + \mathbf{B}) \prec_w \sigma(\mathbf{A}) + \sigma(\mathbf{B})$, and all three vectors have nonnegative entries. Let Φ be the gauge of the norm. By Theorem 21.3.4 (b), $\Phi(\sigma(\mathbf{A} + \mathbf{B})) \leq \Phi(\sigma(\mathbf{A}) + \sigma(\mathbf{B}))$, and by the triangle inequality for the norm Φ on $\mathbb{R}^p$ the right side is at most $\Phi(\sigma(\mathbf{A})) + \Phi(\sigma(\mathbf{B}))$. By Theorem 21.4.3 these are $\|\mathbf{A} + \mathbf{B}\|$, and $\|\mathbf{A}\| + \|\mathbf{B}\|$. (The triangle inequality used is that of Φ , a norm on $\mathbb{R}^p$, not that of $\|\cdot\|$.)
- (b) Apply (a) to the Ky Fan k -norm, whose value at $\mathbf{M}$ is $\sum_{i \leq k} \sigma_i(\mathbf{M})$.

Exercise. C2: Distance to the singular matrices, in every norm

Let $\mathbf{A} \in M_n(\mathbb{C})$ be invertible with singular values $\sigma_1 \geq \cdots \geq \sigma_n > 0$, and let $\|\cdot\|$ be a unitarily invariant norm with gauge Φ .

- (a) Prove that $\min\{\|\mathbf{A} - \mathbf{B}\| : \mathbf{B} \in M_n(\mathbb{C}) \text{ singular}\} = \Phi(\sigma_n, 0, \dots, 0)$.
- (b) Deduce that this minimum equals $\sigma_n \Phi(\mathbf{e}_1)$, and identify it for the spectral, Frobenius, and trace norms.

Hint: a matrix is singular exactly when its rank is at most $n - 1$.

Solution.

- (a) By Theorem 7.5.3 (d), $\mathbf{B}$ is singular if and only if $\text{rank } \mathbf{B} \leq n - 1$. So the minimum is over the set appearing in Corollary 21.6.2 with $k = n - 1$, and that corollary gives the value $\Phi(\sigma_n, 0, \dots, 0)$, attained at the truncation $\mathbf{A}_{n-1}$, which is singular because its rank is $n - 1$.
- (b) The vector $(\sigma_n, 0, \dots, 0)$ is $\sigma_n \mathbf{e}_1$, and Φ is a norm, so $\Phi(\sigma_n \mathbf{e}_1) = \sigma_n \Phi(\mathbf{e}_1)$ as $\sigma_n > 0$. For the spectral norm $\Phi = \|\cdot\|_\infty$ and $\Phi(\mathbf{e}_1) = 1$; for the Frobenius norm $\Phi = \|\cdot\|_2$ and $\Phi(\mathbf{e}_1) = 1$; for the trace norm $\Phi = \|\cdot\|_1$ and $\Phi(\mathbf{e}_1) = 1$. All three give σ_n . This is not an accident: the normalization $\Phi(\mathbf{e}_1) = 1$ of Section 4 is exactly the condition that makes every unitarily invariant norm agree on rank-one matrices with one nonzero singular value.

Exercise. C3: Unitary matrices are far apart

Let $\mathbf{U}, \mathbf{Q} \in M_n(\mathbb{C})$ be unitary with $\mathbf{U} \neq \mathbf{Q}$.

- (a) Prove that $\sigma(\mathbf{U} - \mathbf{Q}) = (|\mu_1 - 1|, \dots, |\mu_n - 1|)^\downarrow$, where $\mu_1, \dots, \mu_n$ are the eigenvalues of $\mathbf{Q}^* \mathbf{U}$.

- (b) Deduce that $\|\mathbf{U} - \mathbf{Q}\| = \|\text{diag}(\mu_1 - 1, \dots, \mu_n - 1)\|$ for every unitarily invariant norm, and that the trace-norm distance between two distinct unitary matrices can be as large as $2n$.

Solution.

- (a) $\mathbf{U} - \mathbf{Q} = \mathbf{Q}(\mathbf{Q}^*\mathbf{U} - \mathbf{I})$, and multiplying on the left by a unitary matrix does not change singular values, since $(\mathbf{QM})^*(\mathbf{QM}) = \mathbf{M}^*\mathbf{M}$ (Definition 13.9.1). The matrix $\mathbf{Z} := \mathbf{Q}^*\mathbf{U}$ is unitary, hence normal, so by the spectral theorem $\mathbf{Z} = \mathbf{Y} \text{diag}(\mu_1, \dots, \mu_n) \mathbf{Y}^*$ with $\mathbf{Y}$ unitary, and $\mathbf{Z} - \mathbf{I} = \mathbf{Y} \text{diag}(\mu_i - 1) \mathbf{Y}^*$. Then $(\mathbf{Z} - \mathbf{I})^*(\mathbf{Z} - \mathbf{I}) = \mathbf{Y} \text{diag}(|\mu_i - 1|^2) \mathbf{Y}^*$, so the singular values are the numbers $|\mu_i - 1|$ sorted decreasingly (Definition 13.9.1).
- (b) Both matrices in the claimed equality have the same singular value list, so Theorem 21.4.3 gives equal norms. Each μ_i has modulus 1, so $|\mu_i - 1| \leq 2$, with equality when $\mu_i = -1$. Taking $\mathbf{Q} = \mathbf{I}$ and $\mathbf{U} = -\mathbf{I}$ gives every $\mu_i = -1$ and trace-norm distance $\sum_i 2 = 2n$.