

## 21.4 Unitarily Invariant Norms

Chapter 16 ended its survey of norms with a list of what it had left undone, and one entry read: “It has not said which norms are invariant under unitary multiplication, which is Chapter 21.” This section says it. The answer is that such a norm is nothing but a symmetric gauge function applied to the singular values, so §03’s inequality transfers wholesale from vectors to matrices — and the Ky Fan functionals that failed to be norms on Hermitian matrices become norms as soon as they are read on singular values instead.

**Throughout,**  $F = \mathbb{R}$  or  $F = \mathbb{C}$ ,  $\mathbf{A}, \mathbf{B} \in M_{m \times n}(F)$ ,  $p = \min(m, n)$ , and the decreasing list of singular values (Definition 13.9.1) is  $\sigma(\mathbf{A}) = (\sigma_1(\mathbf{A}), \dots, \sigma_p(\mathbf{A}))$ . Symmetric gauge functions are §03’s (Definition 21.3.1) and live on  $\mathbb{R}^p$ . For  $\mathbf{x} \in \mathbb{R}^p$  we write

$$\text{diag}_{m,n}(\mathbf{x}) \in M_{m \times n}(F)$$

for the matrix with  $x_1, \dots, x_p$  down the main diagonal and every other entry 0; for  $m = n$  this is the usual  $\text{diag}(x_1, \dots, x_n)$ . Finally, one consequence of Theorem 13.9.1 and Theorem 13.9.2 is used constantly: if  $\mathbf{A} = \mathbf{W}\mathbf{\Sigma}\mathbf{X}^*$  is a singular value decomposition and  $\mathbf{U} \in M_m(F)$ ,  $\mathbf{V} \in M_n(F)$  are unitary, then

$$\mathbf{U}\mathbf{A}\mathbf{V} = (\mathbf{U}\mathbf{W})\mathbf{\Sigma}(\mathbf{V}^*\mathbf{X})^*$$

is again of that shape, with  $\mathbf{U}\mathbf{W}$  and  $\mathbf{V}^*\mathbf{X}$  unitary, so

$$\sigma(\mathbf{U}\mathbf{A}\mathbf{V}) = \sigma(\mathbf{A}). \quad (21.4.1)$$

### 21.4.1 Norms that do not see the unitary factors

Two matrix norms in the book already ignore unitary factors. The Frobenius norm does, by Lemma 11.7.5; the spectral norm does, by Lemma 20.9.1 (c). Both facts were proved separately and both were used the same way: to replace  $\mathbf{A}$  by  $\mathbf{\Sigma}$  and reduce a question about a matrix to a question about a list of numbers. That is a property worth isolating, and it is the property Chapter 16 left unnamed.

*A norm is unitarily invariant when rotating the domain and the codomain does not change the measurement.*

#### Definition 21.4.1. Unitarily Invariant Norm

A **unitarily invariant norm** on  $M_{m \times n}(F)$  is a norm  $\|\cdot\|$  on  $M_{m \times n}(F)$ , regarded as a vector space over  $F$ , such that

$$\|\mathbf{U}\mathbf{A}\mathbf{V}\| = \|\mathbf{A}\|$$

for **every**  $\mathbf{A} \in M_{m \times n}(F)$  and **all** unitary  $\mathbf{U} \in M_m(F)$ ,  $\mathbf{V} \in M_n(F)$ .

The triple bars are standard and are used from here on for a norm known to be unitarily invariant; a single bar with a subscript, as in  $\|\mathbf{A}\|_2$  or  $\|\mathbf{A}\|_F$ , keeps naming a specific one. In words: the condition is not invariance under similarity, and not invariance under transposition; it is invariance under multiplying on the **left** by any unitary and on the **right** by any unitary, independently. The two unitaries are unrelated, which is what makes the condition strong. Asking only for invariance under  $\mathbf{A} \mapsto \mathbf{U}\mathbf{A}\mathbf{U}^*$ , with the same unitary on both sides, is a much weaker demand:  $|\text{tr } \mathbf{A}|$  has it, since  $\text{tr}(\mathbf{U}\mathbf{A}\mathbf{U}^*) = \text{tr } \mathbf{A}$  by Theorem 0.9.6 (3), yet it is not determined by  $\sigma(\mathbf{A})$ , because  $\mathbf{I}_2$  and  $\text{diag}(1, -1)$  both have singular values 1, 1 while their traces have absolute values 2 and 0.

**Example. The two familiar norms**

Check each against [Definition 21.4.1](#).

- (a) The Frobenius norm  $\|\mathbf{A}\|_F$ .
- (b) The spectral norm  $\|\mathbf{A}\|_2$ .

*Solution.*

- (a) It is a norm, being the norm of the Frobenius inner product of Chapter 11 §01, and  $\|\mathbf{UAV}\|_F = \|\mathbf{A}\|_F$  is Lemma 11.7.5.
- (b) It is a norm by Theorem 16.3.2 (b), and Lemma 20.9.1 (c) gives  $\|\mathbf{UAV}\|_2 = \|\mathbf{A}\|_2$  for unitary  $\mathbf{U}, \mathbf{V}$ . By Theorem 16.3.3 (c),  $\|\mathbf{A}\|_2 = \sigma_1(\mathbf{A})$ , so this norm too is a function of  $\sigma(\mathbf{A})$  alone — and [Eq. 21.4.1](#) gives a second proof of its invariance.

Both are built from  $\sigma(\mathbf{A})$ . The largest singular value and the square root of the sum of their squares are two entries in a whole family, and the family is the point of the section.

**Definition 21.4.2. Ky Fan k-Norms**

For  $1 \leq k \leq p$ , the **Ky Fan  $k$ -norm** of  $\mathbf{A} \in M_{m \times n}(F)$  is

$$\|\mathbf{A}\|_{(k)} = \sum_{i=1}^k \sigma_i(\mathbf{A}),$$

the sum of the  $k$  largest singular values. For  $k = 1$  it is the spectral norm  $\|\mathbf{A}\|_2$ ; for  $k = p$  it is called the **trace norm**.

Chapter 17 §06 measured a Hermitian matrix by  $s_k(\mathbf{A}) = \lambda_1(\mathbf{A}) + \cdots + \lambda_k(\mathbf{A})$  and found in Exercise (C2: Partial sums that behave like a norm) that this is no norm:  $s_1(\text{diag}(0, -1)) = 0$  although  $\text{diag}(0, -1) \neq \mathbf{O}$ . The exercise said that “the repair is to apply  $N_k$  [its name there for  $s_k$ ] to the singular values instead of the eigenvalues, which is what Chapter 21 does”. [Definition 21.4.2](#) is that repair, and the repair works. Positive definiteness now holds because singular values are nonnegative, so they can only sum to 0 by all vanishing, which by Theorem 13.9.1 forces  $\mathbf{A} = \mathbf{O}$ ; absolute homogeneity holds because  $(c\mathbf{A})^*(c\mathbf{A}) = |c|^2 \mathbf{A}^* \mathbf{A}$  gives  $\sigma(c\mathbf{A}) = |c| \sigma(\mathbf{A})$  through Definition 13.9.1; and unitary invariance is [Eq. 21.4.1](#). The triangle inequality is the one axiom still outstanding, and [Lemma 21.4.2](#) below supplies it.

**Nonexamples by minimal change.** Both are honest norms on  $M_2(\mathbb{R})$  that fail only the invariance clause, and one witness refutes both. Let

$$\mathbf{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix},$$

a rotation through half a right angle, which is orthogonal since its columns are orthonormal.

- The entrywise maximum  $N(\mathbf{A}) = \max_{i,j} |a_{ij}|$  is a norm (it is the  $\infty$ -norm of the entry list). But  $N(\mathbf{I}_2) = 1$  while  $N(\mathbf{U}\mathbf{I}_2\mathbf{I}_2) = N(\mathbf{U}) = 1/\sqrt{2}$ .
- The operator 1-norm  $\|\mathbf{A}\|_1$ , the largest column sum of absolute values (Theorem 16.3.3 (a)), is a norm. But  $\|\mathbf{I}_2\|_1 = 1$  while  $\|\mathbf{U}\|_1 = 1/\sqrt{2} + 1/\sqrt{2} = \sqrt{2}$ .

Both failures are visible in the same picture:  $\mathbf{I}_2$  and  $\mathbf{U}$  have the same singular values, namely 1, 1, because both are orthogonal, and a unitarily invariant norm is therefore forced to give them the same value. A norm that reads individual entries or individual columns is reading a basis, and a unitary change of basis is exactly what Definition 21.4.1 forbids it to notice.

**Why this definition.** The clause is exactly what makes a norm a function of the singular values and nothing else, as the next lemma shows; that is its whole purpose. Weakening it to one-sided invariance  $\|\mathbf{U}\mathbf{A}\| = \|\mathbf{A}\|$  is not enough. The largest Euclidean length among the columns of  $\mathbf{A}$  is a norm with that property, since a unitary matrix preserves the length of each column; but it gives  $\text{diag}(2, 1)$  the value 2, while  $\text{diag}(2, 1)\mathbf{U}^\top = \frac{1}{\sqrt{2}}\begin{pmatrix} 2 & 2 \\ -1 & 1 \end{pmatrix}$  has both columns of length  $\sqrt{5/2}$ , and the two matrices have the same singular values. Strengthening it to include all invertible  $\mathbf{U}, \mathbf{V}$  is far too much, since every nonzero matrix of a given rank would then have the same norm.

*Quick check.* Is  $\mathbf{A} \mapsto \sigma_2(\mathbf{A})$  a unitarily invariant norm on  $M_2(\mathbb{R})$ ? Is  $\mathbf{A} \mapsto \sigma_1(\mathbf{A}) - \sigma_2(\mathbf{A})$ ?

*Solution.* Neither, and both fail (N1) of Definition 16.1.1 rather than the invariance clause: they are unitarily invariant by Eq. 21.4.1, but they are not norms. The matrix  $\mathbf{A} = \mathbf{e}_1\mathbf{e}_1^\top$  has  $\sigma(\mathbf{A}) = (1, 0)$ , so  $\sigma_2(\mathbf{A}) = 0$  with  $\mathbf{A} \neq \mathbf{O}$ . The identity has  $\sigma(\mathbf{I}_2) = (1, 1)$ , so  $\sigma_1(\mathbf{I}_2) - \sigma_2(\mathbf{I}_2) = 0$  with  $\mathbf{I}_2 \neq \mathbf{O}$ . This is the same failure Chapter 17 met with  $s_k$ , and it says that not every symmetric-looking formula in the singular values is a norm; the correspondence below says exactly which are.

## 21.4.2 A unitarily invariant norm is a function of the singular values

### Lemma 21.4.1. The Gauge of a Unitarily Invariant Norm

Let  $\|\cdot\|$  be a unitarily invariant norm on  $M_{m \times n}(F)$  and define

$$\Phi(\mathbf{x}) = \|\text{diag}_{m,n}(\mathbf{x})\| \quad (\mathbf{x} \in \mathbb{R}^p).$$

Then  $\Phi$  is a symmetric gauge function on  $\mathbb{R}^p$ , and

$$\|\mathbf{A}\| = \Phi(\sigma(\mathbf{A})) \quad \text{for every } \mathbf{A} \in M_{m \times n}(F).$$

In particular  $\|\mathbf{A}\|$  depends on  $\mathbf{A}$  only through  $\sigma(\mathbf{A})$ .

*Idea.* The map  $\mathbf{x} \mapsto \text{diag}_{m,n}(\mathbf{x})$  is linear and injective, so the norm axioms for  $\Phi$  are inherited entry for entry. The two invariance clauses are the two families of unitary matrices that act diagonally: a permutation matrix on each side rearranges the diagonal, a diagonal matrix of signs on one side flips its signs. The last formula is the singular value decomposition, read once.

*Proof.  $\Phi$  is a norm.* The map  $L(\mathbf{x}) = \text{diag}_{m,n}(\mathbf{x})$  is linear from  $\mathbb{R}^p$  to  $M_{m \times n}(F)$  and injective, since  $L(\mathbf{x}) = \mathbf{O}$  reads off as  $\mathbf{x} = \mathbf{0}$ . Hence  $\Phi = \|\cdot\| \circ L$  satisfies (N1) —  $\Phi(\mathbf{x}) = 0$  gives  $L(\mathbf{x}) = \mathbf{O}$  and  $\mathbf{x} = \mathbf{0}$  — and (N2) for real scalars, and (N3), all directly from the corresponding axioms for  $\|\cdot\|$ .

(G1). The letter  $\sigma$  is taken here, so write  $\pi$  for a permutation of  $\{1, \dots, p\}$ . Extend it to a permutation  $\hat{\pi}$  of  $\{1, \dots, m\}$  and to one  $\tilde{\pi}$  of  $\{1, \dots, n\}$  by fixing every index greater than  $p$ , and let  $\mathbf{P} \in M_m(F)$  and  $\mathbf{Q} \in M_n(F)$  be their permutation matrices (Definition 3.4.2), which are unitary by Lemma 3.4.7 (b). Writing  $\mathbf{D} = \text{diag}_{m,n}(\mathbf{x})$ , multiplying by  $\mathbf{P}$  on the left moves row  $k$  to row  $\hat{\pi}(k)$  and multiplying by  $\mathbf{Q}^\top$  on the right moves column  $l$  to column  $\tilde{\pi}(l)$ ; no entry is created or destroyed. The only nonzero entries of  $\mathbf{D}$  are  $d_{kk} = x_k$  for  $k \leq p$ , and they land in positions  $(\hat{\pi}(k), \tilde{\pi}(k)) = (\pi(k), \pi(k))$ , again on the main diagonal. Hence

$$\mathbf{P} \text{diag}_{m,n}(\mathbf{x}) \mathbf{Q}^\top = \text{diag}_{m,n}(\mathbf{P}_\pi \mathbf{x}).$$

Taking  $\|\cdot\|$  of both sides and using Definition 21.4.1 gives  $\Phi(\mathbf{P}_\pi \mathbf{x}) = \Phi(\mathbf{x})$ .

(G2). Let  $\varepsilon_1, \dots, \varepsilon_p$  be signs. The matrix  $\text{diag}(\varepsilon_1, \dots, \varepsilon_p, 1, \dots, 1) \in M_m(F)$  is real diagonal with entries of modulus 1, hence unitary, and multiplying on the left by it multiplies row  $k$  by  $\varepsilon_k$ . So it carries  $\text{diag}_{m,n}(\mathbf{x})$  to  $\text{diag}_{m,n}(\varepsilon_1 x_1, \dots, \varepsilon_p x_p)$ , and invariance gives (G2).

**The formula.** Let  $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^*$  be a singular value decomposition (Theorem 13.9.1). Then  $\Sigma = \text{diag}_{m,n}(\sigma(\mathbf{A}))$ , and  $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^*$  with  $\mathbf{U}$  and  $\mathbf{V}^*$  unitary, so  $\|\mathbf{A}\| = \|\Sigma\| = \Phi(\sigma(\mathbf{A}))$ . This proves the lemma. ■

So the map  $\|\cdot\| \mapsto \Phi$  loses nothing. The question is whether it is surjective: does every symmetric gauge come from a unitarily invariant norm? Evaluating  $\Phi$  at  $\sigma(\mathbf{A})$  certainly produces a unitarily invariant *function*, by Eq. 21.4.1, and (N1) and (N2) are easy. Everything hangs on the triangle inequality, and the triangle inequality hangs on one majorization.

### 21.4.3 Singular values are subadditive

#### Lemma 21.4.2. Weak Subadditivity of the Singular Values

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

$$\sigma(\mathbf{A} + \mathbf{B}) \prec_w \sigma(\mathbf{A}) + \sigma(\mathbf{B}),$$

that is,  $\sum_{i=1}^k \sigma_i(\mathbf{A} + \mathbf{B}) \leq \sum_{i=1}^k (\sigma_i(\mathbf{A}) + \sigma_i(\mathbf{B}))$  for every  $1 \leq k \leq p$ .

*Idea.* The statement is about singular values, and the tool — Corollary 17.6.5, which says that the top  $k$  eigenvalues of a Hermitian sum are subadditive — is about eigenvalues. The Hermitian dilation  $\mathcal{H}$  of Proposition 17.9.3 converts one into the other, and it does so **linearly**:  $\mathcal{H}(\mathbf{A} + \mathbf{B}) = \mathcal{H}(\mathbf{A}) + \mathcal{H}(\mathbf{B})$ , because the blocks add. Since the top  $p$  eigenvalues of  $\mathcal{H}(\mathbf{A})$  are exactly  $\sigma_1(\mathbf{A}), \dots, \sigma_p(\mathbf{A})$ , the corollary transfers verbatim.

*Proof.* Write  $\mathcal{H}(\mathbf{M}) = \begin{pmatrix} \mathbf{O} & \mathbf{M} \\ \mathbf{M}^* & \mathbf{O} \end{pmatrix} \in M_{m+n}(F)$  for the Hermitian dilation of Proposition 17.9.3. Adding block by block,  $\mathcal{H}(\mathbf{A} + \mathbf{B}) = \mathcal{H}(\mathbf{A}) + \mathcal{H}(\mathbf{B})$ .

By Proposition 17.9.3 the eigenvalues of  $\mathcal{H}(\mathbf{M})$ , with multiplicity, are  $\sigma_1(\mathbf{M}), \dots, \sigma_p(\mathbf{M})$ , then  $m + n - 2p$  zeros, then  $-\sigma_p(\mathbf{M}), \dots, -\sigma_1(\mathbf{M})$ . Every singular value is  $\geq 0$  and every listed negative number is  $\leq 0$ , so this list, read in the order written, is already decreasing. Hence

$$\lambda_i(\mathcal{H}(\mathbf{M})) = \sigma_i(\mathbf{M}) \quad (1 \leq i \leq p),$$

for every  $\mathbf{M} \in M_{m \times n}(F)$ .

Now fix  $1 \leq k \leq p$ . The matrices  $\mathcal{H}(\mathbf{A})$  and  $\mathcal{H}(\mathbf{B})$  are Hermitian of the same size, so Corollary 17.6.5 applies to them:

$$\sum_{i=1}^k \lambda_i(\mathcal{H}(\mathbf{A}) + \mathcal{H}(\mathbf{B})) \leq \sum_{i=1}^k \lambda_i(\mathcal{H}(\mathbf{A})) + \sum_{i=1}^k \lambda_i(\mathcal{H}(\mathbf{B})).$$

Replacing  $\mathcal{H}(\mathbf{A}) + \mathcal{H}(\mathbf{B})$  by  $\mathcal{H}(\mathbf{A} + \mathbf{B})$  on the left and using the displayed identity three times turns this into

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

Finally,  $\sigma(\mathbf{A}) + \sigma(\mathbf{B})$  is a sum of two decreasing vectors, hence decreasing, so it is its own decreasing rearrangement, and the displayed inequalities for  $k = 1, \dots, p$  are precisely (M1) for  $\prec_w$  in Definition 17.8.1. This proves the lemma. ■

Two remarks on the route. The dilation was chosen because it reduces the statement to a corollary already proved, in four lines and with no new variational argument; the alternative is to establish a Ky Fan maximum principle for singular values directly, maximizing  $\sum_{i \leq k} |\langle \mathbf{A} \mathbf{w}_i, \mathbf{q}_i \rangle|$  over pairs of orthonormal lists, which is a genuine theorem in its own right and is not needed here. And the conclusion really is only **weak** majorization: for  $\mathbf{A} = \mathbf{I}_2$  and  $\mathbf{B} = \text{diag}(-1, 1)$  we have  $\sigma(\mathbf{A}) = \sigma(\mathbf{B}) = (1, 1)$  and  $\mathbf{A} + \mathbf{B} = \text{diag}(0, 2)$ , so  $\sigma(\mathbf{A} + \mathbf{B}) = (2, 0)$  has total 2 while  $\sigma(\mathbf{A}) + \sigma(\mathbf{B}) = (2, 2)$  has total 4. Section 6 records this as a negative result: no strengthening to  $\prec$  is available.

The debts of the last two examples are now paid. By Lemma 21.4.2 with  $k$  fixed,

$$\|\mathbf{A} + \mathbf{B}\|_{(k)} \leq \|\mathbf{A}\|_{(k)} + \|\mathbf{B}\|_{(k)},$$

so the Ky Fan  $k$ -norms of Definition 21.4.2 really are norms, and so is the trace norm, the case  $k = p$ . Chapter 17 §06's Ky Fan functionals have been repaired exactly as that section promised.

## 21.4.4 The correspondence

### Theorem 21.4.3. Unitarily Invariant Norms Are Symmetric Gauges

Let  $m, n \geq 1$  and  $p = \min(m, n)$ . The assignments

$$\|\cdot\| \longmapsto \Phi_{\|\cdot\|}(\mathbf{x}) = \left\| \text{diag}_{m,n}(\mathbf{x}) \right\|, \quad \Phi \longmapsto \|\mathbf{A}\|_{\Phi} = \Phi(\sigma(\mathbf{A}))$$

are mutually inverse bijections between the unitarily invariant norms on  $M_{m \times n}(F)$  and the symmetric gauge functions on  $\mathbb{R}^p$ .

*Idea.* [Lemma 21.4.1](#) is one direction and already contains the fact that the round trip  $\|\cdot\| \mapsto \Phi \mapsto \|\cdot\|$  is the identity. For the other direction the only issue is the triangle inequality for  $\mathbf{A} \mapsto \Phi(\sigma(\mathbf{A}))$ , and it splits into two steps that have just been prepared:  $\sigma(\mathbf{A} + \mathbf{B}) \prec_w \sigma(\mathbf{A}) + \sigma(\mathbf{B})$ , then §03's theorem, then the triangle inequality for  $\Phi$  itself.

*Proof. Step 1:  $\|\cdot\|_\Phi$  is a unitarily invariant norm.* Let  $\Phi$  be a symmetric gauge function on  $\mathbb{R}^p$  and put  $\|\mathbf{A}\|_\Phi = \Phi(\sigma(\mathbf{A}))$ .

(N1).  $\Phi(\sigma(\mathbf{A})) \geq 0$ , and it vanishes only if  $\sigma(\mathbf{A}) = \mathbf{0}$ , which by Theorem 13.9.1 forces  $\mathbf{A} = \mathbf{U}\mathbf{O}\mathbf{V}^* = \mathbf{O}$ .

(N2). For  $c \in F$ ,  $(c\mathbf{A})^*(c\mathbf{A}) = |c|^2 \mathbf{A}^* \mathbf{A}$ , so each eigenvalue of  $(c\mathbf{A})^*(c\mathbf{A})$  is  $|c|^2$  times the matching eigenvalue of  $\mathbf{A}^* \mathbf{A}$  and Definition 13.9.1 gives  $\sigma(c\mathbf{A}) = |c|\sigma(\mathbf{A})$ . Hence  $\|c\mathbf{A}\|_\Phi = \Phi(|c|\sigma(\mathbf{A})) = |c| \|\mathbf{A}\|_\Phi$ .

(N3). The vectors  $\sigma(\mathbf{A} + \mathbf{B})$  and  $\sigma(\mathbf{A}) + \sigma(\mathbf{B})$  have nonnegative entries, and [Lemma 21.4.2](#) gives  $\sigma(\mathbf{A} + \mathbf{B}) \prec_w \sigma(\mathbf{A}) + \sigma(\mathbf{B})$ . So Theorem 21.3.4 (b) applies, and then the triangle inequality for the norm  $\Phi$ :

$$\begin{aligned} \|\mathbf{A} + \mathbf{B}\|_\Phi &= \Phi(\sigma(\mathbf{A} + \mathbf{B})) \leq \Phi(\sigma(\mathbf{A}) + \sigma(\mathbf{B})) \\ &\leq \Phi(\sigma(\mathbf{A})) + \Phi(\sigma(\mathbf{B})) = \|\mathbf{A}\|_\Phi + \|\mathbf{B}\|_\Phi. \end{aligned}$$

Invariance. By [Eq. 21.4.1](#),  $\sigma(\mathbf{U}\mathbf{A}\mathbf{V}) = \sigma(\mathbf{A})$  for unitary  $\mathbf{U}, \mathbf{V}$ , so  $\|\mathbf{U}\mathbf{A}\mathbf{V}\|_\Phi = \|\mathbf{A}\|_\Phi$ .

**Step 2: the round trips.** Start from a unitarily invariant norm  $\|\cdot\|$ . By [Lemma 21.4.1](#) its  $\Phi = \Phi_{\|\cdot\|}$  is a symmetric gauge and  $\|\mathbf{A}\| = \Phi(\sigma(\mathbf{A})) = \|\mathbf{A}\|_\Phi$  for every  $\mathbf{A}$ , so the second assignment undoes the first.

Start instead from a symmetric gauge  $\Phi$  on  $\mathbb{R}^p$ , and let  $\Psi$  be the gauge of the norm  $\|\cdot\|_\Phi$ . For  $\mathbf{x} \in \mathbb{R}^p$ , the matrix  $\text{diag}_{m,n}(\mathbf{x})$  has  $\text{diag}_{m,n}(\mathbf{x})^* \text{diag}_{m,n}(\mathbf{x}) = \text{diag}(x_1^2, \dots, x_p^2, 0, \dots, 0)$ , whose eigenvalue list in decreasing order is  $(|\mathbf{x}|^\downarrow)^2$  padded with zeros; so Definition 13.9.1 gives  $\sigma(\text{diag}_{m,n}(\mathbf{x})) = |\mathbf{x}|^\downarrow$ . Hence

$$\Psi(\mathbf{x}) = \left\| \text{diag}_{m,n}(\mathbf{x}) \right\|_\Phi = \Phi(|\mathbf{x}|^\downarrow) = \Phi(\mathbf{x}),$$

the last equality by [Eq. 21.3.1](#). So the first assignment undoes the second, and the two maps are mutually inverse bijections. This proves the theorem. ■

The theorem is a machine for producing matrix norms: choose any symmetric gauge and it hands back a unitarily invariant norm, with the triangle inequality free of charge. Feeding it the gauges of Example (The p-norms, the Ky Fan gauges, and size one) (a) gives the most-used family.

**Example. The Schatten norms**

The letter  $p$  is reserved in this section for  $\min(m, n)$ , so write the exponent as  $r$ . For  $1 \leq r \leq \infty$ , the **Schatten  $r$ -norm** is

$$\|\mathbf{A}\|_{S_r} = \|\sigma(\mathbf{A})\|_r = \left( \sum_{i=1}^p \sigma_i(\mathbf{A})^r \right)^{1/r},$$

read as  $\|\mathbf{A}\|_{S_\infty} = \sigma_1(\mathbf{A})$  at  $r = \infty$ . Identify  $S_1$ ,  $S_2$ , and  $S_\infty$ , and explain why each  $\|\cdot\|_{S_r}$  is a unitarily invariant norm.

*Solution.* By Example (The  $p$ -norms, the Ky Fan gauges, and size one) (a),  $\|\cdot\|_r$  is a symmetric gauge function on  $\mathbb{R}^p$  for every  $1 \leq r \leq \infty$ . So [Theorem 21.4.3](#) makes  $\mathbf{A} \mapsto \|\sigma(\mathbf{A})\|_r$  a unitarily invariant norm, with no further checking. At the three familiar exponents:  $\|\cdot\|_{S_1}$  is the trace norm, since  $\|\sigma(\mathbf{A})\|_1 = \sum_i \sigma_i(\mathbf{A})$ ;  $\|\cdot\|_{S_2}$  is the Frobenius norm, since  $\|\mathbf{A}\|_F^2 = \|\Sigma\|_F^2 = \sum_i \sigma_i(\mathbf{A})^2$  by Lemma 11.7.5 applied to  $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^*$ ; and  $\|\cdot\|_{S_\infty}$  is the spectral norm, by Theorem 16.3.3 (c). The three norms Chapter 16 measured matrices with are three points of one family. A concrete reading: the symmetric  $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$  has eigenvalues 3 and 1, both positive, so  $\mathbf{A}^*\mathbf{A} = \mathbf{A}^2$  has eigenvalues 9 and 1 and  $\sigma(\mathbf{A}) = (3, 1)$  by Definition 13.9.1. Hence  $\|\mathbf{A}\|_{S_1} = 4$ ,  $\|\mathbf{A}\|_{S_2} = \sqrt{10}$  — which matches  $\|\mathbf{A}\|_F = \sqrt{4 + 1 + 1 + 4}$  — and  $\|\mathbf{A}\|_{S_\infty} = 3$ .

## 21.4.5 What every unitarily invariant norm satisfies

Chapter 20 §09 collected the facts it needed about moving factors in and out of the spectral and Frobenius norms (Lemma 20.9.1). Part (b) of that lemma —  $\|\mathbf{X}\mathbf{M}\mathbf{Y}\| \leq \|\mathbf{X}\|_2 \|\mathbf{M}\| \|\mathbf{Y}\|_2$  — was proved separately for the two norms. It holds for all of them at once.

**Proposition 21.4.4. Basic Properties**

Let  $\|\cdot\|$  be a unitarily invariant norm on  $M_{m \times n}(F)$ , with gauge  $\Phi$  as in [Lemma 21.4.1](#), and let  $\mathbf{A} \in M_{m \times n}(F)$ .

- (a)  $\sigma(\mathbf{A}^*) = \sigma(\mathbf{A})$ ; consequently, if  $\|\cdot\|'$  is the unitarily invariant norm on  $M_{n \times m}(F)$  attached to the same gauge  $\Phi$  by [Theorem 21.4.3](#), then  $\|\mathbf{A}^*\|' = \|\mathbf{A}\|$ . For  $m = n$  this reads  $\|\mathbf{A}^*\| = \|\mathbf{A}\|$ .
- (b)  $\|\mathbf{X}\mathbf{A}\| \leq \|\mathbf{X}\|_2 \|\mathbf{A}\|$  for every  $\mathbf{X} \in M_m(F)$ , and  $\|\mathbf{A}\mathbf{Y}\| \leq \|\mathbf{A}\| \|\mathbf{Y}\|_2$  for every  $\mathbf{Y} \in M_n(F)$ .
- (c) If  $\|\cdot\|$  is **normalized**, meaning  $\|\mathbf{e}_1\mathbf{e}_1^*\| = 1$  for the  $m \times n$  matrix  $\mathbf{e}_1\mathbf{e}_1^*$  with a single 1 in position  $(1, 1)$ , then

$$\|\mathbf{A}\|_2 \leq \|\mathbf{A}\| \leq \|\mathbf{A}\|_{(p)}.$$

- (d) A normalized unitarily invariant norm on  $M_n(F)$  is submultiplicative:  $\|\mathbf{A}\mathbf{B}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|$ .

*Idea.* Everything reduces to one inequality between singular values,  $\sigma_i(\mathbf{XA}) \leq \|\mathbf{X}\|_2 \sigma_i(\mathbf{A})$ , and then to §03's monotonicity Proposition 21.3.1. That inequality is one line of the min-max theorem: the subspace on which  $\mathbf{A}$  stretches by at most  $\sigma_i(\mathbf{A})$  also works for  $\mathbf{XA}$ , with the factor  $\|\mathbf{X}\|_2$  paid once. Part (c) is Corollary 21.3.2, read through the correspondence, and (d) is (b) and (c) combined.

*Proof.*

- (a) By Theorem 13.9.1 write  $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^*$ ; then  $\mathbf{A}^* = \mathbf{V}\Sigma^*\mathbf{U}^*$ , and  $\Sigma^* = \text{diag}_{n,m}(\sigma(\mathbf{A}))$  has the same diagonal. By Theorem 13.9.2 applied to  $\mathbf{A}^*$ ,  $\sigma(\mathbf{A}^*) = \sigma(\mathbf{A})$ . The consequence is Lemma 21.4.1 used twice:  $\|\mathbf{A}^*\|' = \Phi(\sigma(\mathbf{A}^*)) = \Phi(\sigma(\mathbf{A})) = \|\mathbf{A}\|$ .
- (b) Everything rests on one inequality between singular values.

*Claim.*  $\sigma_k(\mathbf{XA}) \leq \|\mathbf{X}\|_2 \sigma_k(\mathbf{A})$  for every  $1 \leq k \leq n$ , with the convention  $\sigma_k = 0$  for  $k > p$ .

*Proof.* Fix a singular value decomposition of  $\mathbf{A}$  with right singular vectors  $\mathbf{v}_1, \dots, \mathbf{v}_n$ , and let  $T_k = \text{span}(\mathbf{v}_k, \dots, \mathbf{v}_n)$ , a subspace of  $F^n$  of dimension  $n - k + 1$ . By Lemma 17.9.1 (a),  $\|\mathbf{Ax}\| \leq \sigma_k(\mathbf{A})\|\mathbf{x}\|$  for every  $\mathbf{x} \in T_k$ . By Theorem 16.3.2 (a),  $\|\mathbf{X}(\mathbf{Ax})\| \leq \|\mathbf{X}\|_2 \|\mathbf{Ax}\|$ , so

$$\|(\mathbf{XA})\mathbf{x}\| \leq \|\mathbf{X}\|_2 \sigma_k(\mathbf{A}) \|\mathbf{x}\| \quad (\mathbf{x} \in T_k).$$

The matrices  $\mathbf{XA}$  and  $\mathbf{A}$  both have  $n$  columns, so Theorem 17.9.2 applies to  $\mathbf{XA}$  with subspaces of  $F^n$ :  $\sigma_k(\mathbf{XA})$  is the **minimum** over subspaces of dimension  $n - k + 1$  of the largest stretch on them, and  $T_k$  is one competitor. Hence  $\sigma_k(\mathbf{XA}) \leq \|\mathbf{X}\|_2 \sigma_k(\mathbf{A})$ . ■

Now  $\mathbf{XA}$  and  $\mathbf{A}$  lie in the same space  $M_{m \times n}(F)$ , and the claim says  $|\sigma_i(\mathbf{XA})| \leq \|\mathbf{X}\|_2 \sigma_i(\mathbf{A})$  for  $1 \leq i \leq p$ . By Proposition 21.3.1 and then (N2) for  $\Phi$ ,

$$\|\mathbf{XA}\| = \Phi(\sigma(\mathbf{XA})) \leq \Phi(\|\mathbf{X}\|_2 \sigma(\mathbf{A})) = \|\mathbf{X}\|_2 \Phi(\sigma(\mathbf{A})) = \|\mathbf{X}\|_2 \|\mathbf{A}\|.$$

For the second inequality, apply the first to  $\mathbf{A}^*$  and  $\mathbf{Y}^*$  inside  $M_{n \times m}(F)$ , with the norm  $\|\cdot\|'$  of part (a):  $\|(\mathbf{AY})^*\|' = \|\mathbf{Y}^*\mathbf{A}^*\|' \leq \|\mathbf{Y}^*\|_2 \|\mathbf{A}^*\|'$ . By (a) the left side is  $\|\mathbf{AY}\|$  and  $\|\mathbf{A}^*\|' = \|\mathbf{A}\|$ , while  $\|\mathbf{Y}^*\|_2 = \|\mathbf{Y}\|_2$  by Lemma 20.9.1 (a).

- (c) The matrix  $\mathbf{e}_1 \mathbf{e}_1^*$  is  $\text{diag}_{m,n}(\mathbf{e}_1)$ , so the normalization says  $\Phi(\mathbf{e}_1) = 1$ . By Corollary 21.3.2,  $\Phi(\mathbf{e}_1)\|\mathbf{x}\|_\infty \leq \Phi(\mathbf{x}) \leq \Phi(\mathbf{e}_1)\|\mathbf{x}\|_1$  for every  $\mathbf{x} \in \mathbb{R}^p$ . At  $\mathbf{x} = \sigma(\mathbf{A})$  this reads

$$\sigma_1(\mathbf{A}) \leq \Phi(\sigma(\mathbf{A})) \leq \sum_{i=1}^p \sigma_i(\mathbf{A}),$$

which is the claim, since  $\sigma_1(\mathbf{A}) = \|\mathbf{A}\|_2$  by Theorem 16.3.3 (c) and the right-hand side is  $\|\mathbf{A}\|_{(p)}$  by Definition 21.4.2.

- (d) By (b) and then (c),  $\|\mathbf{AB}\| \leq \|\mathbf{A}\|_2 \|\mathbf{B}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|$ . This proves the proposition. ■

Applying (b) twice, once on each side, gives  $\|\mathbf{XAY}\| \leq \|\mathbf{X}\|_2 \|\mathbf{A}\| \|\mathbf{Y}\|_2$ . That is Lemma 20.9.1 (b) with the two named norms replaced by any unitarily invariant one, so the debt announced at the head of this subsection is paid. Taking  $\mathbf{X} = \mathbf{U}$  and  $\mathbf{Y} = \mathbf{V}$  unitary recovers Definition 21.4.1 itself: a unitary matrix has orthonormal columns, so  $\|\mathbf{U}\|_2 = \|\mathbf{V}\|_2 = 1$  by Lemma 20.9.1 (c), and

the bound reads  $\|\mathbf{UAV}\| \leq \|\mathbf{A}\|$ ; applying it again to  $\mathbf{U}^*(\mathbf{UAV})\mathbf{V}^* = \mathbf{A}$  gives  $\|\mathbf{A}\| \leq \|\mathbf{UAV}\|$ , and the two together give equality.

*Warning.* A unitarily invariant norm need not be submultiplicative, and the normalization in (d) is what supplies the lower bound that rules it out. Half the spectral norm,  $\|\mathbf{A}\| = \frac{1}{2}\|\mathbf{A}\|_2$ , is a unitarily invariant norm — it is  $\Phi(\sigma(\mathbf{A}))$  for the symmetric gauge  $\Phi = \frac{1}{2}\|\cdot\|_\infty$ . At  $\mathbf{A} = \mathbf{B} = \mathbf{I}_n$  it gives  $\|\mathbf{AB}\| = \frac{1}{2}$  while  $\|\mathbf{A}\|\|\mathbf{B}\| = \frac{1}{4}$ , so submultiplicativity fails. Doubling instead of halving is harmless:  $2\|\cdot\|_2$  is submultiplicative. What Proposition 21.4.4 (d) really needs is  $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|$ , and Corollary 21.3.2 delivers that as soon as  $\|\mathbf{e}_1\mathbf{e}_1^*\| \geq 1$ , normalized or not; here  $\|\mathbf{e}_1\mathbf{e}_1^*\| = \frac{1}{2}$ , and the bound runs the wrong way. Submultiplicativity is a property of the **scaling** of a norm, not of its shape.

Theorem 21.4.3 is the section's payment of Chapter 16 §08's debt: the norms invariant under unitary multiplication are exactly the symmetric gauges of the singular values, and there are as many of them as there are norms on  $\mathbb{R}^p$  that ignore order and sign. What the correspondence buys is that any inequality between two singular value lists in the weak majorization order becomes an inequality between the matrices in **every** one of these norms at once. Section 6 makes that the definition of dominance and uses it to settle low-rank approximation and the nearest unitary matrix; Section 7 does the same for Lidskii's and Mirsky's theorems.

## 21.4.6 Exercises

### 21.4.6.1 A. Check your understanding

#### Exercise. A1

- (a) State Definition 21.4.1, saying exactly which matrices  $\mathbf{U}$  and  $\mathbf{V}$  range over.
- (b) Explain in one sentence why a unitarily invariant norm can be computed from  $\sigma(\mathbf{A})$  alone.
- (c) True or false, with a reason:  $\mathbf{A} \mapsto \max_{i,j} |a_{ij}|$  is a unitarily invariant norm on  $M_2(\mathbb{R})$ .
- (d) Which axiom of Definition 16.1.1 is the only hard one for  $\mathbf{A} \mapsto \Phi(\sigma(\mathbf{A}))$ , and which two earlier results prove it?

*Solution.*

- (a) A norm  $\|\cdot\|$  on  $M_{m \times n}(F)$  with  $\|\mathbf{UAV}\| = \|\mathbf{A}\|$  for every  $\mathbf{A}$  and all unitary  $\mathbf{U} \in M_m(F)$  and  $\mathbf{V} \in M_n(F)$ , the two chosen independently.
- (b) Because  $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$  with  $\mathbf{\Sigma} = \text{diag}_{m,n}(\sigma(\mathbf{A}))$ , so invariance gives  $\|\mathbf{A}\| = \|\mathbf{\Sigma}\|$ , a quantity built from  $\sigma(\mathbf{A})$  alone (Lemma 21.4.1).
- (c) False. It is a norm but not unitarily invariant:  $\mathbf{I}_2$  and the rotation  $\mathbf{U} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$  have entrywise maxima 1 and  $1/\sqrt{2}$ , while  $\mathbf{U} = \mathbf{U}\mathbf{I}_2\mathbf{U}$ .
- (d) The triangle inequality (N3). It follows from Lemma 21.4.2, which gives  $\sigma(\mathbf{A} + \mathbf{B}) \prec_w \sigma(\mathbf{A}) + \sigma(\mathbf{B})$ , together with Theorem 21.3.4 (b), which turns a weak majorization of nonnegative vectors into an inequality of gauges.

### 21.4.6.2 B. Practice

**Exercise. B1: Which of these are unitarily invariant norms**

Determine which of the following are unitarily invariant norms on  $M_2(\mathbb{R})$ . Justify your answer, naming the clause that fails when one does.

- (a)  $\|\mathbf{A}\| = \sigma_1(\mathbf{A}) + \sigma_2(\mathbf{A})$ .
- (b)  $\|\mathbf{A}\| = |a_{11}| + |a_{12}| + |a_{21}| + |a_{22}|$ .
- (c)  $\|\mathbf{A}\| = (\sigma_1(\mathbf{A})^3 + \sigma_2(\mathbf{A})^3)^{1/3}$ .
- (d)  $\|\mathbf{A}\| = |\operatorname{tr} \mathbf{A}|$ .
- (e)  $\|\mathbf{A}\| = \max(\sigma_1(\mathbf{A}), 2\sigma_2(\mathbf{A}))$ .

*Solution.*

- (a) Yes: this is the trace norm, the Ky Fan 2-norm of [Definition 21.4.2](#), equivalently  $\|\cdot\|_{S_1}$  of [Example \(The Schatten norms\)](#). Alternatively, it is  $\Phi(\sigma(\mathbf{A}))$  for  $\Phi = \|\cdot\|_1$ , so [Theorem 21.4.3](#) applies.
- (b) No. It is a norm — the 1-norm of the entry list — but it is not unitarily invariant: it gives  $\mathbf{I}_2$  the value 2 and the rotation  $\mathbf{U}$  above the value  $4/\sqrt{2} = 2\sqrt{2}$ , although  $\sigma(\mathbf{I}_2) = \sigma(\mathbf{U}) = (1, 1)$ .
- (c) Yes: it is  $\Phi(\sigma(\mathbf{A}))$  for  $\Phi = \|\cdot\|_3$ , a symmetric gauge by [Example \(The p-norms, the Ky Fan gauges, and size one\)](#) (a), so [Theorem 21.4.3](#) applies. This is the Schatten 3-norm.
- (d) No. It is unitarily *similarity* invariant but not a norm:  $\operatorname{tr} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = 0$  with the matrix nonzero, so (N1) fails. It is not unitarily invariant either, since  $\operatorname{tr}(\mathbf{U}\mathbf{A}\mathbf{V})$  generally differs from  $\operatorname{tr} \mathbf{A}$ .
- (e) No. By [Theorem 21.4.3](#) the question is whether  $\Phi(\mathbf{x}) = \max(|x|_1^\downarrow, 2|x|_2^\downarrow)$  is a symmetric gauge on  $\mathbb{R}^2$ , and it is not a norm: subadditivity fails. Take  $\mathbf{x} = (1, \frac{1}{2})$  and  $\mathbf{y} = (\frac{1}{2}, 1)$ . Then  $\Phi(\mathbf{x}) = \max(1, 1) = 1$  and  $\Phi(\mathbf{y}) = 1$ , while  $\mathbf{x} + \mathbf{y} = (\frac{3}{2}, \frac{3}{2})$  gives  $\Phi(\mathbf{x} + \mathbf{y}) = \max(\frac{3}{2}, 3) = 3 > 2$ . The failure is easy to see in the matrices as well:  $\operatorname{diag}(1, \frac{1}{2})$  and  $\operatorname{diag}(\frac{1}{2}, 1)$  each get the value 1, and their sum  $\frac{3}{2}\mathbf{I}_2$  gets 3. The trap is that  $\mathbf{x} \mapsto 2|x|_2^\downarrow$  is not a norm — it vanishes at  $\mathbf{e}_1$  — so the maximum of it with  $\Phi_1$  inherits nothing. Contrast exercise C1 below, where both members of the maximum are norms.

**Exercise. B2: Measuring one matrix in every way**

Let

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

- (a) Compute  $\sigma(\mathbf{A})$  and  $\sigma(\mathbf{B})$ , and then  $\|\mathbf{A}\|_2$ ,  $\|\mathbf{A}\|_F$ ,  $\|\mathbf{A}\|_{(1)}$ , and  $\|\mathbf{A}\|_{(2)}$ .
- (b) Compute  $\sigma(\mathbf{A} + \mathbf{B})$  and verify  $\sigma(\mathbf{A} + \mathbf{B}) \prec_w \sigma(\mathbf{A}) + \sigma(\mathbf{B})$  directly, saying for which  $k$  the inequality is strict.
- (c) Verify [Proposition 21.4.4](#) (c) for  $\mathbf{A}$  and the Frobenius norm, and check that the Frobenius norm is normalized.

*Solution.*

- (a)  $\mathbf{A}$  is real symmetric with eigenvectors  $(1, 1)$  and  $(1, -1)$  for the eigenvalues 3 and 1.

Hence  $\mathbf{A}^* \mathbf{A} = \mathbf{A}^2$  has eigenvalues 9 and 1, and Definition 13.9.1 gives  $\sigma(\mathbf{A}) = (3, 1)$ , as in [Example \(The Schatten norms\)](#). Likewise  $\mathbf{B}$  has eigenvectors  $(1, -1)$  and  $(1, 1)$  for 3 and 1, so  $\sigma(\mathbf{B}) = (3, 1)$ . Hence  $\|\mathbf{A}\|_2 = 3$ ,  $\|\mathbf{A}\|_F = \sqrt{9 + 1} = \sqrt{10}$  (which agrees with  $\sqrt{4 + 1 + 1 + 4}$ ),  $\|\mathbf{A}\|_{(1)} = 3$  and  $\|\mathbf{A}\|_{(2)} = 4$ .

- (b)  $\mathbf{A} + \mathbf{B} = \text{diag}(4, 4)$ , so  $\sigma(\mathbf{A} + \mathbf{B}) = (4, 4)$ , while  $\sigma(\mathbf{A}) + \sigma(\mathbf{B}) = (6, 2)$ . The partial sums are 4, 8 against 6, 8: strict at  $k = 1$ , equality at  $k = 2$ . So  $\sigma(\mathbf{A} + \mathbf{B}) \prec_w \sigma(\mathbf{A}) + \sigma(\mathbf{B})$ , and here the totals happen to agree, so the relation is even a majorization.
- (c) The Frobenius norm has  $\|\mathbf{e}_1 \mathbf{e}_1^\top\|_F = 1$ , so it is normalized. The chain reads  $3 = \|\mathbf{A}\|_2 \leq \sqrt{10} \leq 4 = \|\mathbf{A}\|_{(2)}$ , and  $9 \leq 10 \leq 16$  confirms it.

### Exercise. B3: Moving a factor across

Let  $\mathbf{A} \in M_{m \times n}(F)$ , let  $\|\cdot\|$  be a unitarily invariant norm on  $M_{m \times n}(F)$ , and let  $P$  be an orthogonal projection of  $F^m$  onto a subspace, with matrix  $\mathbf{P}$ .

- (a) Prove that  $\|\mathbf{P}\|_2 \leq 1$ .
- (b) Deduce that  $\|\mathbf{P}\mathbf{A}\| \leq \|\mathbf{A}\|$ .
- (c) Give an example with  $m = n = 2$  in which the inequality in (b) is strict for the Frobenius norm.

*Solution.*

- (a) For  $\mathbf{x} \in F^m$ , write  $\mathbf{x} = \mathbf{P}\mathbf{x} + (\mathbf{x} - \mathbf{P}\mathbf{x})$ , an orthogonal decomposition, so  $\|\mathbf{x}\|^2 = \|\mathbf{P}\mathbf{x}\|^2 + \|\mathbf{x} - \mathbf{P}\mathbf{x}\|^2 \geq \|\mathbf{P}\mathbf{x}\|^2$  by Theorem 11.1.3. Hence  $\|\mathbf{P}\mathbf{x}\| \leq \|\mathbf{x}\|$  for every  $\mathbf{x}$ , and  $\|\mathbf{P}\|_2 \leq 1$  by Definition 16.3.2.
- (b) By [Proposition 21.4.4](#) (b) with  $\mathbf{X} = \mathbf{P}$ ,  $\|\mathbf{P}\mathbf{A}\| \leq \|\mathbf{P}\|_2 \|\mathbf{A}\| \leq \|\mathbf{A}\|$ .
- (c) Take  $\mathbf{P} = \mathbf{e}_1 \mathbf{e}_1^\top$  and  $\mathbf{A} = \mathbf{I}_2$ . Then  $\mathbf{P}\mathbf{A} = \mathbf{e}_1 \mathbf{e}_1^\top$  has  $\|\mathbf{P}\mathbf{A}\|_F = 1$ , while  $\|\mathbf{I}_2\|_F = \sqrt{2}$ .

### 21.4.6.3 C. Going deeper

#### Exercise. C1: A gauge built from two Ky Fan gauges

On  $\mathbb{R}^p$  let  $\Phi(\mathbf{x}) = \max(\Phi_1(\mathbf{x}), \frac{2}{k}\Phi_k(\mathbf{x}))$  for a fixed  $2 \leq k \leq p$ , with  $\Phi_1$  and  $\Phi_k$  the Ky Fan gauges of Definition 21.3.2.

- (a) Prove that  $\Phi$  is a symmetric gauge function.
- (b) Write down the unitarily invariant norm it induces on  $M_{m \times n}(F)$ , and decide whether it is normalized.
- (c) For  $p = k = 2$ , show that  $\Phi(\mathbf{x}) = \max(|x_1|_1^\downarrow, |x_1| + |x_2|) = |x_1| + |x_2|$ , and conclude that the norm of (b) is the trace norm in that case.

*Solution.*

- (a) A maximum of finitely many norms is a norm: it is nonnegative, absolutely homogeneous, and subadditive because each member is, and it vanishes only where all members do, hence only at  $\mathbf{0}$ . Here both members are symmetric gauges by Example (The p-

norms, the Ky Fan gauges, and size one) (b), and a maximum of permutation invariant and sign invariant functions has the same invariances. So  $\Phi$  satisfies Definition 21.3.1.

(b) By Theorem 21.4.3 the induced norm is

$$\|\mathbf{A}\| = \max\left(\sigma_1(\mathbf{A}), \frac{2}{k} \sum_{i \leq k} \sigma_i(\mathbf{A})\right).$$

It is normalized: at  $\mathbf{A} = \mathbf{e}_1 \mathbf{e}_1^*$  the singular values are  $1, 0, \dots, 0$ , so the two candidates are 1 and  $2/k \leq 1$ , and the maximum is 1.

(c) With  $p = k = 2$ ,  $\frac{2}{2}\Phi_2(\mathbf{x}) = |x_1| + |x_2|$ , which is at least  $\Phi_1(\mathbf{x}) = |x|_1$ . So  $\Phi = \|\cdot\|_1$ , and the induced norm is  $\sigma_1(\mathbf{A}) + \sigma_2(\mathbf{A})$ , the trace norm.

### Exercise. C2: How much scaling submultiplicativity can stand

Let  $c > 0$  and consider  $N_c(\mathbf{A}) = c \|\mathbf{A}\|_2$  on  $M_n(F)$ ,  $n \geq 1$ .

- (a) Prove that  $N_c$  is a unitarily invariant norm for every  $c > 0$ , and compute its gauge.
- (b) Determine exactly which  $c$  make  $N_c$  submultiplicative. *Hint: test at  $\mathbf{A} = \mathbf{B} = \mathbf{I}_n$ .*
- (c) Do the same for  $M_c(\mathbf{A}) = c \|\mathbf{A}\|_{(n)}$ , the scaled trace norm. Show that  $\mathbf{A} = \mathbf{B} = \mathbf{I}_n$  is now the **wrong** test matrix, and find one that gives the sharp answer.

*Solution.*

- (a) A positive multiple of a norm is a norm, and multiplying by  $c$  does not disturb  $N_c(\mathbf{UAV}) = N_c(\mathbf{A})$ . Its gauge is  $\mathbf{x} \mapsto N_c(\text{diag}(\mathbf{x})) = c\|\mathbf{x}\|_\infty$ , since the largest singular value of  $\text{diag}(\mathbf{x})$  is  $\|\mathbf{x}\|_\infty$ .
- (b) Exactly  $c \geq 1$ . If  $c \geq 1$ , then  $N_c(\mathbf{AB}) = c\|\mathbf{AB}\|_2 \leq c\|\mathbf{A}\|_2\|\mathbf{B}\|_2 \leq c^2\|\mathbf{A}\|_2\|\mathbf{B}\|_2 = N_c(\mathbf{A})N_c(\mathbf{B})$ , using submultiplicativity of the spectral norm (Theorem 16.3.2 (d)) and  $c \leq c^2$ . If  $c < 1$ , take  $\mathbf{A} = \mathbf{B} = \mathbf{I}_n$ : then  $N_c(\mathbf{AB}) = c$  while  $N_c(\mathbf{A})N_c(\mathbf{B}) = c^2 < c$ .
- (c) The answer is again exactly  $c \geq 1$ . Sufficiency: by Proposition 21.4.4 (b) and (c),  $\|\mathbf{AB}\|_{(n)} \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_{(n)} \leq \|\mathbf{A}\|_{(n)} \|\mathbf{B}\|_{(n)}$ , so for  $c \geq 1$ ,

$$M_c(\mathbf{AB}) = c \|\mathbf{AB}\|_{(n)} \leq c \|\mathbf{A}\|_{(n)} \|\mathbf{B}\|_{(n)} \leq c^2 \|\mathbf{A}\|_{(n)} \|\mathbf{B}\|_{(n)} = M_c(\mathbf{A})M_c(\mathbf{B}).$$

Necessity: at  $\mathbf{A} = \mathbf{B} = \mathbf{I}_n$  one gets  $M_c(\mathbf{I}_n) = cn$  and the requirement  $cn \leq c^2 n^2$ , that is  $c \geq 1/n$  — a weaker condition, so for  $n \geq 2$  the identity is the wrong test matrix. Take instead  $\mathbf{A} = \mathbf{B} = \mathbf{e}_1 \mathbf{e}_1^*$ , which is idempotent with singular values  $1, 0, \dots, 0$ . Then  $M_c(\mathbf{AB}) = c$  while  $M_c(\mathbf{A})M_c(\mathbf{B}) = c^2$ , and  $c \leq c^2$  forces  $c \geq 1$ . The moral: submultiplicativity is tested by matrices of the **smallest** rank, where the trace norm and the spectral norm agree, not by the identity, where they differ by a factor of  $n$ .

### Exercise. C3: The correspondence in the smallest case

- (a) Describe all unitarily invariant norms on  $M_{1 \times n}(F)$ , the row vectors. *Hint: what is  $p$ ?*
- (b) Deduce that on  $M_{1 \times n}(F)$  every unitarily invariant norm is a positive multiple of the Euclidean norm of the row.

- (c) Explain why the answer to (a) does not contradict the existence of many different unitarily invariant norms on  $M_2(F)$ .

*Solution.*

- (a) Here  $p = \min(1, n) = 1$ , so [Theorem 21.4.3](#) matches unitarily invariant norms on  $M_{1 \times n}(F)$  with symmetric gauge functions on  $\mathbb{R}^1$ . By Example (The  $p$ -norms, the Ky Fan gauges, and size one) (c), every norm on  $\mathbb{R}^1$  is such a gauge, and each is  $\Phi(t) = c|t|$  with  $c > 0$ . So the unitarily invariant norms are exactly  $\mathbf{A} \mapsto c \sigma_1(\mathbf{A})$ ,  $c > 0$ .
- (b) A row vector  $\mathbf{a}^\top$  has  $\mathbf{A}^* \mathbf{A} = \bar{\mathbf{a}} \mathbf{a}^\top$  of rank at most 1, with the single nonzero eigenvalue  $\mathbf{a}^* \mathbf{a} = \|\mathbf{a}\|_2^2$ ; so  $\sigma_1(\mathbf{A}) = \|\mathbf{a}\|_2$ . Hence every unitarily invariant norm on  $M_{1 \times n}(F)$  is  $c \|\mathbf{a}\|_2$ .
- (c) Because the correspondence is with norms on  $\mathbb{R}^p$ , and  $p = 1$  leaves only one shape. For  $M_2(F)$ ,  $p = 2$ , and there are many norms on  $\mathbb{R}^2$  that are permutation and sign invariant — the whole family  $\|\cdot\|_q$  for  $1 \leq q \leq \infty$ , for instance, giving the Schatten norms of [Example \(The Schatten norms\)](#). The rigidity in (a) is a statement about  $\mathbb{R}^1$ , not about matrices.