

17.8 The Diagonal and the Spectrum

A Hermitian matrix carries two lists of n real numbers: its diagonal entries, read off at a glance, and its eigenvalues, computed with effort. Since Chapter 9 we have known that the two lists have the same sum, the trace. This section asks what else the eigenvalues force on the diagonal, and answers the question completely. The answer needs a name, **majorization**. One half of it, Schur's theorem, is §06's bound on partial sums of the diagonal in the new language. The other half, Horn's theorem, says that nothing more is forced; it is proved by an induction driven by a single 2×2 rotation. The section ends with a short second proof of Hadamard's determinant inequality.

Throughout, the field is $\mathbb{R}$ or $\mathbb{C}$, and every matrix whose eigenvalues are indexed is Hermitian, with its eigenvalues real and listed **decreasingly** with multiplicity, $\lambda_1(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{A})$. Two vectors in $\mathbb{R}^n$ are attached to such an $\mathbf{A} \in M_n(F)$:

$$\begin{aligned}\mathbf{d}(\mathbf{A}) &= (a_{11}, a_{22}, \dots, a_{nn}), \\ \boldsymbol{\lambda}(\mathbf{A}) &= (\lambda_1(\mathbf{A}), \lambda_2(\mathbf{A}), \dots, \lambda_n(\mathbf{A})).\end{aligned}$$

The first is the **diagonal vector**, in the order the entries sit on the diagonal; the second is the **eigenvalue vector**, always in decreasing order. The diagonal entries are real because $\mathbf{A}^* = \mathbf{A}$ gives $a_{ii} = \overline{a_{ii}}$, and the eigenvalues are real by Theorem 12.2.1. The bold $\mathbf{d}(\mathbf{A})$ is a vector, not the diagonal matrix $\text{diag}(a_{11}, \dots, a_{nn})$.

Two facts about a Hermitian $\mathbf{A}$ are used below. The spectral theorem (Corollary 12.4.3 over $\mathbb{C}$, Corollary 12.5.2 over $\mathbb{R}$) makes $\mathbf{A}$ similar to the real diagonal matrix of its eigenvalues, so its characteristic polynomial splits as $\prod_i (x - \lambda_i(\mathbf{A}))$. Then Theorem 9.3.3 gives

$$\text{tr } \mathbf{A} = \sum_{i=1}^n \lambda_i(\mathbf{A}), \quad \det \mathbf{A} = \prod_{i=1}^n \lambda_i(\mathbf{A}).$$

17.8.1 What the diagonal already knows

Start with the smallest interesting case. The matrix

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

has $\mathbf{d}(\mathbf{A}) = (2, 2)$ and eigenvalues 3 and 1, with eigenvectors $(1, 1)$ and $(1, -1)$, so $\boldsymbol{\lambda}(\mathbf{A}) = (3, 1)$. The totals agree, $2 + 2 = 4 = 3 + 1$. But the two lists are not alike: the diagonal is perfectly even, while the spectrum is spread out.

A larger example shows the pattern better. The tridiagonal matrix

$$\mathbf{A} = \begin{pmatrix} 3 & 2 & 0 \\ 2 & 4 & 2 \\ 0 & 2 & 5 \end{pmatrix}$$

has eigenvectors $(1, 2, 2)$, $(-2, -1, 2)$ and $(2, -2, 1)$, with eigenvalues 7, 4 and 1. To check the first, $\mathbf{A}(1, 2, 2) = (3 + 4, 2 + 8 + 4, 4 + 10) = (7, 14, 14)$; the other two are checked the same way. So $\boldsymbol{\lambda}(\mathbf{A}) = (7, 4, 1)$, while $\mathbf{d}(\mathbf{A}) = (3, 4, 5)$. Sort the diagonal decreasingly to $(5, 4, 3)$ and compare the running totals of the two lists:

k	1	2	3
largest k diagonal entries	5	9	12
largest k eigenvalues	7	11	12

Each running total of the eigenvalues is at least the matching running total of the diagonal, and at $k = n$ the two are equal. None of this is new. Corollary 17.6.3 says that any k diagonal entries add up to at most $\lambda_1(\mathbf{A}) + \cdots + \lambda_k(\mathbf{A})$, with equality at $k = n$, and the k largest diagonal entries are one such choice. What is new is the point of view: to ask which diagonals are **possible**, the corollary's list of inequalities has to be read as a single relation between two vectors.

Two features of this comparison deserve notice. First, the diagonal had to be **sorted**, and this is forced. For a permutation matrix $\mathbf{P}_\sigma$ (Definition 3.4.2), $\mathbf{P}_\sigma^\top = \mathbf{P}_\sigma^{-1}$ by Lemma 3.4.7 (b), so $\mathbf{P}_\sigma^\top \mathbf{A} \mathbf{P}_\sigma$ is Hermitian with the same eigenvalues as $\mathbf{A}$, and its (j, j) entry is $\mathbf{e}_{\sigma(j)}^\top \mathbf{A} \mathbf{e}_{\sigma(j)} = a_{\sigma(j)\sigma(j)}$. Relabeling the basis shuffles the diagonal and fixes the spectrum, so any true relation between $\mathbf{d}(\mathbf{A})$ and $\boldsymbol{\lambda}(\mathbf{A})$ must ignore the order of the diagonal. Second, the comparison is between **running totals from the top**, and it measures how unevenly a fixed total is spread.

17.8.2 Majorization

One vector is majorized by another when the two have the same total but the second piles more of it onto its largest entries: every “top k ” sum of the second is at least that of the first.

The sorting step gets a notation first. For $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, the **decreasing rearrangement** $\mathbf{x}^\downarrow = (x_1^\downarrow, \dots, x_n^\downarrow)$ is the vector with the same entries, repeated as often as in $\mathbf{x}$, listed so that $x_1^\downarrow \geq x_2^\downarrow \geq \cdots \geq x_n^\downarrow$. For example, $(0, 3, 1, 3)^\downarrow = (3, 3, 1, 0)$. In this notation, $\boldsymbol{\lambda}(\mathbf{A}) = \boldsymbol{\lambda}(\mathbf{A})^\downarrow$ by convention, while $\mathbf{d}(\mathbf{A})$ usually needs sorting.

Definition 17.8.1. Majorization

Let $n \geq 1$ and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ be **real** vectors with the **same number** of entries. We say that $\mathbf{x}$ is **majorized by** $\mathbf{y}$, and write $\mathbf{x} \prec \mathbf{y}$, if

$$(M1) \quad \sum_{i=1}^k x_i^\downarrow \leq \sum_{i=1}^k y_i^\downarrow \text{ for every } k = 1, \dots, n-1; \text{ and}$$

$$(M2) \quad \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \text{ (equal totals).}$$

We say that $\mathbf{x}$ is **weakly majorized by** $\mathbf{y}$, and write $\mathbf{x} \prec_w \mathbf{y}$, if the inequality in (M1) holds for **every** $k = 1, \dots, n$, including $k = n$, and (M2) is dropped.

In words, clause by clause. (M1) sorts both vectors and compares their **running totals from the top**: the largest entry of $\mathbf{x}$ is at most the largest entry of $\mathbf{y}$, the two largest of $\mathbf{x}$ add up to at most the two largest of $\mathbf{y}$, and so on up to $k = n-1$. The word **every** matters: one failing k refutes $\mathbf{x} \prec \mathbf{y}$. (M2) says the totals are equal; a sum does not care about order, so no sorting is needed there. Weak majorization keeps (M1), adds its case $k = n$, which reads $\sum_i x_i \leq \sum_i y_i$, and asks nothing more. Comparing the definitions: $\mathbf{x} \prec \mathbf{y}$ holds **exactly when** $\mathbf{x} \prec_w \mathbf{y}$ and the totals are equal.

There is an equivalent way to read (M1) from the bottom. Subtracting the k -th running total from the equal totals of (M2) gives

$$\sum_{i=k+1}^n x_i^\downarrow \geq \sum_{i=k+1}^n y_i^\downarrow \quad (k = 1, \dots, n-1),$$

so under (M2), the $n-k$ **smallest** entries of $\mathbf{x}$ add up to at least the $n-k$ smallest entries of $\mathbf{y}$. The vector $\mathbf{y}$ is more extreme at both ends: its top is heavier and its bottom is lighter.

The definition makes sense because $\mathbf{x}^\downarrow$ is determined by $\mathbf{x}$: when two entries are equal, it does not matter which is written first. A consequence worth recording: **the relation sees only which numbers occur in each vector and how often, not where they sit.** If $\mathbf{x}'$ is a rearrangement of $\mathbf{x}$ and $\mathbf{y}'$ of $\mathbf{y}$, then $\mathbf{x} \prec \mathbf{y}$ if and only if $\mathbf{x}' \prec \mathbf{y}'$, because the rearrangements have the same decreasing rearrangements and the same totals.

Example. A chain, the two extremes, and size one

Check each claim against [Definition 17.8.1](#).

- (a) $(1, 1, 1) \prec (2, 1, 0) \prec (3, 0, 0)$.
- (b) For every $\mathbf{x} \in \mathbb{R}^n$ with mean $m = \frac{1}{n} \sum_i x_i$, the **flat vector** $m\mathbf{1} = (m, \dots, m)$ satisfies $m\mathbf{1} \prec \mathbf{x}$.
- (c) For every $\mathbf{x} \in \mathbb{R}^n$ with **nonnegative** entries and total $s = \sum_i x_i$, $\mathbf{x} \prec (s, 0, \dots, 0)$.
- (d) For $n = 1$, $\mathbf{x} \prec \mathbf{y}$ if and only if $\mathbf{x} = \mathbf{y}$.

Solution.

- (a) All three vectors are already decreasing, and all have total 3, so (M2) holds for both pairs. Their running totals are

$$(1, 1, 1): 1, 2, 3; \quad (2, 1, 0): 2, 3, 3; \quad (3, 0, 0): 3, 3, 3.$$

At $k = 1$ and $k = 2$ the first row is at most the second ($1 \leq 2, 2 \leq 3$), and the second is at most the third ($2 \leq 3, 3 \leq 3$). So (M1) holds for both pairs.

- (b) The totals are $nm = \sum_i x_i$, so (M2) holds. For (M1), fix k and let $a = \frac{1}{k} \sum_{i \leq k} x_i^\downarrow$ be the average of the k largest entries. Each of these is at least $x_k^\downarrow$, so $a \geq x_k^\downarrow$. Every remaining entry $x_j^\downarrow$ with $j > k$ is at most $x_k^\downarrow$, hence at most a . Therefore

$$nm = \sum_{i \leq k} x_i^\downarrow + \sum_{j > k} x_j^\downarrow \leq ka + (n - k)a = na,$$

so $m \leq a$, and multiplying by k gives $km \leq \sum_{i \leq k} x_i^\downarrow$. The left side is the k -th running total of $m\mathbf{1}$. This is (M1).

- (c) The totals are both s . Every running total of $(s, 0, \dots, 0)$ equals s , and $\sum_{i \leq k} x_i^\downarrow \leq s$ because the omitted entries $x_{k+1}^\downarrow, \dots, x_n^\downarrow$ are ≥ 0 . This is (M1). Nonnegativity is really needed: $(1, -1)$ has total 0, but $(1, -1) \prec (0, 0)$ fails at $k = 1$, since $1 > 0$.
- (d) For $n = 1$ there are no values of k in (M1), since k runs from 1 to 0. So only (M2) remains, and it says $x_1 = y_1$.

Parts (b) and (c) are the two ends of the order: among vectors with a given total, the flat vector is the most evenly spread, and among nonnegative ones, the vector that puts everything in one entry is the least. Part (d) is the degenerate case: with a single entry there is nothing to spread, and the relation collapses to equality.

The chain in (a) might suggest that any two vectors with the same total can be ranked this way. They cannot.

Example. Neither majorizes the other

Show that $\mathbf{x} = (4, 1, 1)$ and $\mathbf{y} = (3, 3, 0)$ have the same total, that neither $\mathbf{x} \prec \mathbf{y}$ nor $\mathbf{y} \prec \mathbf{x}$, and that both are majorized by $(4, 2, 0)$.

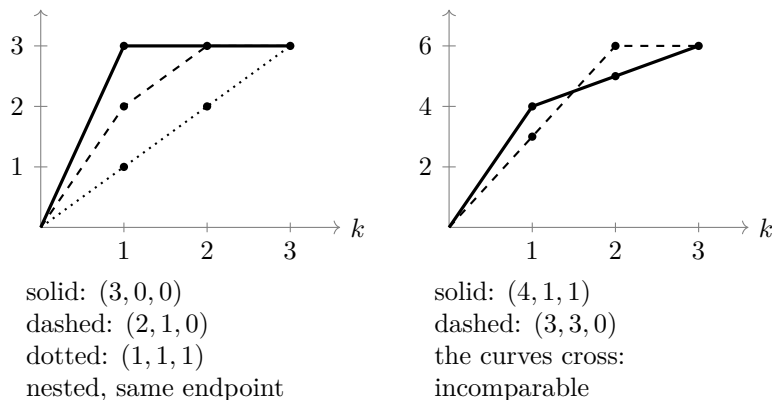
Solution. Both totals are 6, and both vectors are already decreasing. Their running totals are

$$\mathbf{x}: 4, 5, 6; \quad \mathbf{y}: 3, 6, 6.$$

At $k = 1$, $4 > 3$, so (M1) fails for $\mathbf{x} \prec \mathbf{y}$. At $k = 2$, $6 > 5$, so (M1) fails for $\mathbf{y} \prec \mathbf{x}$. The vector $(4, 2, 0)$ has total 6 and running totals 4, 6, 6, which are at least both 4, 5 and 3, 6 at $k = 1, 2$. So $\mathbf{x} \prec (4, 2, 0)$ and $\mathbf{y} \prec (4, 2, 0)$.

This is the nonexample by minimal change. Start from the comparable pair $(4, 1, 1) \prec (4, 2, 0)$, move one unit in the second vector from its top entry to its middle entry to get $(3, 3, 0)$, and the pair becomes incomparable. (M2) still holds; what fails is (M1), at a **different** k in each direction. Spreading is not a single number.

The picture makes this visible. Plot the running totals of a sorted vector against k , joined by segments, starting from $(0, 0)$. Then $\mathbf{x} \prec \mathbf{y}$ says that the curve of $\mathbf{y}$ lies on or above the curve of $\mathbf{x}$ and ends at the same height. Two curves can also cross.



Why this definition. Each clause earns its place, and changing any one of them changes the subject.

- *Sorting.* Relabeling a basis shuffles $\mathbf{d}(\mathbf{A})$ and fixes $\boldsymbol{\lambda}(\mathbf{A})$, so a relation comparing unsorted entries would answer a question the matrix does not pose.
- *Running totals from the top, with $\leq$.* These are the quantities that Theorem 17.6.2 and Corollary 17.6.3 control: the top k eigenvalues sum to a maximum. Totals from the bottom add nothing new, as shown above, because under (M2) they are the same information.
- *Equal totals.* This is the trace. Dropping it gives weak majorization, the right notion when there is no conservation law to enforce: $(1, 0) \prec_w (2, 0)$, but $(1, 0) \not\prec (2, 0)$, because the totals differ.

Warning. $\mathbf{x} \prec \mathbf{y}$ does not mean $x_i \leq y_i$, not even after sorting both vectors. In $(1, 1, 1) \prec (2, 1, 0)$, the third entries are 1 and 0, and $1 > 0$. This is no accident: subtracting (M1) at $k = n - 1$ from (M2) gives $x_n^\downarrow \geq y_n^\downarrow$ whenever $\mathbf{x} \prec \mathbf{y}$, so at the bottom the inequality always runs the other way. Majorization compares **sums of the largest entries**, not entries

one at a time.

Quick check. Arrange the five vectors $(2, 1, 1, 0)$, $(4, 0, 0, 0)$, $(1, 1, 1, 1)$, $(3, 1, 0, 0)$, and $(2, 2, 0, 0)$ in a chain $\mathbf{x}^{(1)} \prec \mathbf{x}^{(2)} \prec \dots \prec \mathbf{x}^{(5)}$.

Solution. All five are decreasing and all have total 4. Their running totals at $k = 1, 2, 3$ are

$$\begin{array}{ll} (1, 1, 1, 1): & 1, 2, 3; \\ (2, 1, 1, 0): & 2, 3, 4; \\ (2, 2, 0, 0): & 2, 4, 4; \\ (3, 1, 0, 0): & 3, 4, 4; \\ (4, 0, 0, 0): & 4, 4, 4. \end{array}$$

Read in this order, each triple is entrywise at most the next, so

$$(1, 1, 1, 1) \prec (2, 1, 1, 0) \prec (2, 2, 0, 0) \prec (3, 1, 0, 0) \prec (4, 0, 0, 0).$$

The ends are the flat vector and the concentrated vector of [Example \(A chain, the two extremes, and size one\)](#) (b) and (c), as they must be. These five happen to be comparable in pairs, but [Example \(Neither majorizes the other\)](#) shows that this is not always so.

17.8.3 Majorization is an order

The first small result is that $\prec$ behaves like an order, and it says precisely how it falls short of being a partial order on $\mathbb{R}^n$.

Proposition 17.8.1. Basic Properties of Majorization

Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$.

- (a) $\mathbf{x} \prec \mathbf{x}$.
- (b) If $\mathbf{x} \prec \mathbf{y}$ and $\mathbf{y} \prec \mathbf{z}$, then $\mathbf{x} \prec \mathbf{z}$.
- (c) If $\mathbf{x} \prec \mathbf{y}$ and $\mathbf{y} \prec \mathbf{x}$, then $\mathbf{x}^\downarrow = \mathbf{y}^\downarrow$; that is, $\mathbf{y}$ is a rearrangement of $\mathbf{x}$.

Statements (a) and (b) hold with $\prec_w$ in place of $\prec$.

Proof. For $k = 1, \dots, n$, write $X_k = \sum_{i \leq k} x_i^\downarrow$, and define Y_k and Z_k in the same way. Since $X_n = \sum_i x_i$, [Definition 17.8.1](#) says that $\mathbf{x} \prec \mathbf{y}$ means $X_k \leq Y_k$ for $k < n$ and $X_n = Y_n$.

- (a) $X_k \leq X_k$ for every k , and $X_n = X_n$.
- (b) For $k < n$, $X_k \leq Y_k \leq Z_k$. At $k = n$, $X_n = Y_n = Z_n$. Hence $\mathbf{x} \prec \mathbf{z}$.
- (c) The hypotheses give $X_k \leq Y_k$ and $Y_k \leq X_k$ for $k < n$, and $X_n = Y_n$, so $X_k = Y_k$ for every $k = 1, \dots, n$. Put $X_0 = Y_0 = 0$. Then for each k ,

$$x_k^\downarrow = X_k - X_{k-1} = Y_k - Y_{k-1} = y_k^\downarrow,$$

so $\mathbf{x}^\downarrow = \mathbf{y}^\downarrow$. Two vectors with the same decreasing rearrangement contain the same numbers, each the same number of times, so each is a rearrangement of the other.

For $\prec_w$, the arguments for (a) and (b) are the same with every $k \leq n$ treated like $k < n$. This proves the proposition. ■

So $\prec$ is reflexive and transitive. A relation with these two properties is called a **preorder**. Part (c) is antisymmetry up to rearrangement, and no more: $(1, 0) \prec (0, 1) \prec (1, 0)$, yet $(1, 0) \neq (0, 1)$. So $\prec$ is **not** a partial order on $\mathbb{R}^n$. Restricted to **decreasing** vectors, where a rearrangement is the vector itself, (c) is exact antisymmetry, and there $\prec$ is a partial order. By [Example \(Neither majorizes the other\)](#) it is **not total**, not even among nonnegative decreasing vectors with the same total. Like the Loewner order of Chapter 13 (Proposition 13.5.1), it leaves many pairs uncomparable.

17.8.4 Schur's theorem

The running totals in the 3×3 example were not a coincidence. Restated in the language of majorization, §06's Corollary 17.6.3 says that for every Hermitian matrix the diagonal is more evenly spread than the spectrum:

Theorem 17.8.2. Schur's Majorization Theorem

Let $F = \mathbb{R}$ or $F = \mathbb{C}$, and let $\mathbf{A} \in M_n(F)$ be Hermitian. Then

$$\mathbf{d}(\mathbf{A}) \prec \lambda(\mathbf{A}).$$

Explicitly, if $\mathbf{d}(\mathbf{A})^\downarrow = (a_1^\downarrow, \dots, a_n^\downarrow)$ lists the diagonal entries of $\mathbf{A}$ decreasingly, then

$$\begin{aligned} a_1^\downarrow + \dots + a_k^\downarrow &\leq \lambda_1(\mathbf{A}) + \dots + \lambda_k(\mathbf{A}) \quad (1 \leq k \leq n-1), \\ a_1^\downarrow + \dots + a_n^\downarrow &= \lambda_1(\mathbf{A}) + \dots + \lambda_n(\mathbf{A}). \end{aligned}$$

Proof. Both vectors lie in $\mathbb{R}^n$, as noted at the start of the section. Fix k with $1 \leq k \leq n$. List the positions $1, \dots, n$ in an order in which the diagonal entries decrease, breaking ties arbitrarily, and let $i_1 < \dots < i_k$ be the first k positions of that list, written in increasing order. Then $a_{i_1 i_1}, \dots, a_{i_k i_k}$ are $a_1^\downarrow, \dots, a_k^\downarrow$ in some order, so Corollary 17.6.3 gives

$$a_1^\downarrow + \dots + a_k^\downarrow = \sum_{j=1}^k a_{i_j i_j} \leq \lambda_1(\mathbf{A}) + \dots + \lambda_k(\mathbf{A}),$$

with equality when $k = n$. For $k \leq n-1$ this is (M1). For $k = n$ it is (M2), since a sum does not depend on the order of its terms. This proves the theorem. ■

So Schur's theorem is Corollary 17.6.3 restated. What the restatement adds is a relation between two vectors, and a relation has a converse that can be asked about. Read from the bottom, (M1) says that the $n-k$ smallest diagonal entries add up to **at least** the $n-k$ smallest eigenvalues. At the two ends of [Example \(A chain, the two extremes, and size one\)](#), a **diagonal** matrix has $\mathbf{d}(\mathbf{A})$ a rearrangement of $\lambda(\mathbf{A})$, the most spread a diagonal can be; a **constant** diagonal t is the flat vector $t\mathbf{1}$, where Schur's theorem says only that $nt = \lambda_1(\mathbf{A}) + \dots + \lambda_n(\mathbf{A})$.

Example. Coordinate vectors against eigenvectors

For the matrix

$$\mathbf{A} = \begin{pmatrix} 3 & 2 & 0 \\ 2 & 4 & 2 \\ 0 & 2 & 5 \end{pmatrix}$$

from the start of the section, whose running totals were tabulated there, exhibit for $k = 2$ the coordinate pair that the proof of Corollary 17.6.3 feeds to Theorem 17.6.2 and an orthonormal pair that attains the maximum in Theorem 17.6.2. Account for the difference between the two sums.

Solution. We found $\lambda(\mathbf{A}) = (7, 4, 1)$, with unit eigenvectors $\mathbf{u}_1 = \frac{1}{3}(1, 2, 2)$, $\mathbf{u}_2 = \frac{1}{3}(-2, -1, 2)$, and $\mathbf{u}_3 = \frac{1}{3}(2, -2, 1)$. They are pairwise orthogonal (for instance $\langle (1, 2, 2), (-2, -1, 2) \rangle = -2 - 2 + 4 = 0$), and each vector in parentheses has length 3.

For $k = 2$ the proof uses the two largest diagonal entries, $a_{33} = 5$ and $a_{22} = 4$, so the coordinate pair is $(\mathbf{e}_2, \mathbf{e}_3)$, and $\langle \mathbf{A}\mathbf{e}_2, \mathbf{e}_2 \rangle + \langle \mathbf{A}\mathbf{e}_3, \mathbf{e}_3 \rangle = 9$. The pair $(\mathbf{u}_1, \mathbf{u}_2)$ gives $\langle \mathbf{A}\mathbf{u}_1, \mathbf{u}_1 \rangle + \langle \mathbf{A}\mathbf{u}_2, \mathbf{u}_2 \rangle = 7 + 4 = 11$, since $\langle \mathbf{A}\mathbf{u}_i, \mathbf{u}_i \rangle = \lambda_i \|\mathbf{u}_i\|^2 = \lambda_i$. The coordinate pair falls short of the eigenvector pair by 2, and that shortfall is exactly the slack in Schur's theorem at $k = 2$. Ky Fan's maximum is attained by eigenvectors, and the diagonal only ever sees coordinate vectors.

Quick check. A Hermitian 3×3 matrix $\mathbf{A}$ has diagonal entries 4, -1 , 2. Without knowing anything else about $\mathbf{A}$, what does Theorem 17.8.2 say about $\lambda_1(\mathbf{A})$, about $\lambda_1(\mathbf{A}) + \lambda_2(\mathbf{A})$, and about $\lambda_3(\mathbf{A})$?

Solution. Sorted, $\mathbf{d}(\mathbf{A})^\downarrow = (4, 2, -1)$, with running totals 4, 6, 5. So $\lambda_1(\mathbf{A}) \geq 4$, $\lambda_1(\mathbf{A}) + \lambda_2(\mathbf{A}) \geq 6$, and $\lambda_1(\mathbf{A}) + \lambda_2(\mathbf{A}) + \lambda_3(\mathbf{A}) = 5$. Subtracting the middle inequality from the last equality gives $\lambda_3(\mathbf{A}) \leq 5 - 6 = -1$. In particular $\mathbf{A}$ is indefinite, since $\lambda_1(\mathbf{A}) > 0 > \lambda_3(\mathbf{A})$, and that was visible without computing any eigenvalue.

17.8.5 Horn's converse

Schur's theorem gives a **necessary** condition on a diagonal. Is it also sufficient? Given $\mathbf{d} \prec \lambda$, is there a Hermitian matrix with eigenvalues $\lambda_1, \dots, \lambda_n$ and diagonal entries $d_1, \dots, d_n$, in the given order? For $n = 2$ a rotation answers the question. Turning the eigenbasis of $\text{diag}(\lambda_1, \lambda_2)$ moves its first diagonal entry through every value between λ_2 and λ_1 , and the trace fixes the second entry.

Lemma 17.8.3. The Two-by-Two Rotation

Let $\lambda_1 \geq \lambda_2$ and t be real numbers with $\lambda_2 \leq t \leq \lambda_1$. Put $b = \sqrt{(\lambda_1 - t)(t - \lambda_2)}$. Then there

is an orthogonal $\mathbf{R} \in \mathrm{O}(2)$ with

$$\mathbf{R} \operatorname{diag}(\lambda_1, \lambda_2) \mathbf{R}^\top = \begin{pmatrix} t & b \\ b & \lambda_1 + \lambda_2 - t \end{pmatrix}.$$

Idea. If the first row of $\mathbf{R}$ is $(c, -s)$ with $c^2 + s^2 = 1$, the $(1, 1)$ entry of $\mathbf{R} \operatorname{diag}(\lambda_1, \lambda_2) \mathbf{R}^\top$ is $c^2 \lambda_1 + s^2 \lambda_2$, an average of λ_1 and λ_2 with weights c^2 and s^2 . Every t between them is such an average, and solving for the weights gives c and s .

Proof. If $\lambda_1 = \lambda_2$, then $t = \lambda_1 = \lambda_2$ and $b = 0$, and $\mathbf{R} = \mathbf{I}$ works. Otherwise $\lambda_1 - \lambda_2 > 0$, and

$$c = \sqrt{\frac{t - \lambda_2}{\lambda_1 - \lambda_2}}, \quad s = \sqrt{\frac{\lambda_1 - t}{\lambda_1 - \lambda_2}}$$

are real numbers, because both radicands are ≥ 0 by the hypothesis on t . They satisfy $c^2 + s^2 = 1$. Let $\mathbf{R} = \begin{pmatrix} c & -s \\ s & c \end{pmatrix}$. Then $\mathbf{R}^\top \mathbf{R} = (c^2 + s^2) \mathbf{I} = \mathbf{I}$, so $\mathbf{R} \in \mathrm{O}(2)$. Multiplying out,

$$\mathbf{R} \operatorname{diag}(\lambda_1, \lambda_2) \mathbf{R}^\top = \begin{pmatrix} c^2 \lambda_1 + s^2 \lambda_2 & cs(\lambda_1 - \lambda_2) \\ cs(\lambda_1 - \lambda_2) & s^2 \lambda_1 + c^2 \lambda_2 \end{pmatrix}.$$

Since $s^2 = 1 - c^2$, the $(1, 1)$ entry is $\lambda_2 + c^2(\lambda_1 - \lambda_2) = \lambda_2 + (t - \lambda_2) = t$. The two diagonal entries add up to $(c^2 + s^2)(\lambda_1 + \lambda_2) = \lambda_1 + \lambda_2$, so the $(2, 2)$ entry is $\lambda_1 + \lambda_2 - t$. Finally $cs(\lambda_1 - \lambda_2) = \sqrt{(t - \lambda_2)(\lambda_1 - t)} = b$. This proves the lemma. ■

This is the converse for $n = 2$: if $\lambda_1 \geq \lambda_2$ and $(d_1, d_2) \prec (\lambda_1, \lambda_2)$, then $d_1, d_2 \leq d_1^\downarrow \leq \lambda_1$ by (M1), so by (M2) $d_1 = \lambda_1 + \lambda_2 - d_2 \geq \lambda_2$, so $t = d_1$ is allowed. For larger n , the lemma is used once, on a well-chosen pair of eigenvalues, to put d_1 in the corner, and induction does the rest.

Theorem 17.8.4. Horn's Theorem

Let $n \geq 1$, and let $\mathbf{d} = (d_1, \dots, d_n)$ and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)$ be vectors in $\mathbb{R}^n$ with $\mathbf{d} \prec \boldsymbol{\lambda}$. Then there is an orthogonal $\mathbf{Q} \in \mathrm{O}(n)$ such that the matrix

$$\mathbf{A} = \mathbf{Q} \operatorname{diag}(\lambda_1, \dots, \lambda_n) \mathbf{Q}^\top$$

has diagonal entries $d_1, d_2, \dots, d_n$, **in this order**. So $\mathbf{A}$ is a real symmetric matrix with $\mathbf{d}(\mathbf{A}) = \mathbf{d}$ and $\boldsymbol{\lambda}(\mathbf{A}) = \boldsymbol{\lambda}^\downarrow$.

Idea. Induction on n , one diagonal entry at a time.

① Permutation matrices are orthogonal and shuffle diagonal entries, so it is enough to treat $\mathbf{d}$ and $\boldsymbol{\lambda}$ both decreasing. ② The top and bottom clauses of majorization put d_1 between λ_n and λ_1 , hence between two **consecutive** eigenvalues, $\lambda_k \geq d_1 \geq \lambda_{k+1}$. The lemma rotates that pair so that d_1 sits in the corner, with $\eta = \lambda_k + \lambda_{k+1} - d_1$ next to it. ③ The Claim: what is left over, the diagonal $d_2, \dots, d_n$ and the spectrum with λ_k, λ_{k+1} replaced by η , satisfies the hypothesis again, in size $n - 1$. ④ Induction rotates the lower $(n - 1) \times (n - 1)$ block

into place without touching the corner.

Step ③ is where the choice of k matters.

Proof. For every $\mathbf{Q} \in \mathbf{O}(n)$, the matrix $\mathbf{A} = \mathbf{Q} \operatorname{diag}(\lambda_1, \dots, \lambda_n) \mathbf{Q}^\top$ is real and symmetric, and by Corollary 12.5.2 its eigenvalues, with multiplicity, are $\lambda_1, \dots, \lambda_n$; listed decreasingly they form $\lambda^\downarrow$. So only the diagonal needs work. We use two facts repeatedly. A product of orthogonal matrices is orthogonal, since $(\mathbf{XY})^\top \mathbf{XY} = \mathbf{Y}^\top \mathbf{X}^\top \mathbf{XY} = \mathbf{I}$. And for a permutation $\pi \in S_n$ and any $\mathbf{M} \in M_n(\mathbb{R})$, Definition 3.4.2 gives

$$(\mathbf{P}_\pi^\top \mathbf{M} \mathbf{P}_\pi)_{ij} = \mathbf{e}_{\pi(i)}^\top \mathbf{M} \mathbf{e}_{\pi(j)} = m_{\pi(i)\pi(j)}, \quad (\dagger)$$

where $\mathbf{P}_\pi$ is orthogonal by Lemma 3.4.7 (b).

Step 1: reduction to decreasing vectors. Suppose the theorem holds whenever $\mathbf{d}$ and λ are both decreasing, and let $\mathbf{d} \prec \lambda$ be arbitrary. Majorization sees only rearrangements (as noted after Definition 17.8.1), so $\mathbf{d}^\downarrow \prec \lambda^\downarrow$, and there is $\mathbf{Q}_0 \in \mathbf{O}(n)$ such that $\mathbf{B} = \mathbf{Q}_0 \operatorname{diag}(\lambda^\downarrow) \mathbf{Q}_0^\top$ has diagonal $\mathbf{d}^\downarrow$. Here $\operatorname{diag}(\lambda^\downarrow)$ is the diagonal matrix with diagonal $\lambda_1^\downarrow, \dots, \lambda_n^\downarrow$. Since $\lambda^\downarrow$ is a rearrangement of λ , and $\mathbf{d}^\downarrow$ of $\mathbf{d}$, there are permutations $\sigma, \tau \in S_n$ with $\lambda_{\sigma(i)} = \lambda_i^\downarrow$ and $d_{\tau(j)}^\downarrow = d_j$ for all i, j . By $(\dagger)$, $\mathbf{P}_\sigma^\top \operatorname{diag}(\lambda_1, \dots, \lambda_n) \mathbf{P}_\sigma = \operatorname{diag}(\lambda^\downarrow)$, and the (j, j) entry of $\mathbf{P}_\tau^\top \mathbf{B} \mathbf{P}_\tau$ is $b_{\tau(j)\tau(j)} = d_{\tau(j)}^\downarrow = d_j$. So $\mathbf{Q} = \mathbf{P}_\tau^\top \mathbf{Q}_0 \mathbf{P}_\sigma^\top$ is orthogonal and

$$\begin{aligned} \mathbf{Q} \operatorname{diag}(\lambda_1, \dots, \lambda_n) \mathbf{Q}^\top &= \mathbf{P}_\tau^\top \mathbf{Q}_0 \operatorname{diag}(\lambda^\downarrow) \mathbf{Q}_0^\top \mathbf{P}_\tau \\ &= \mathbf{P}_\tau^\top \mathbf{B} \mathbf{P}_\tau \end{aligned}$$

has diagonal $\mathbf{d}$. It therefore suffices to prove the theorem for decreasing $\mathbf{d}$ and λ , and we do so by induction on n .

Step 2: the corner. For $n = 1$, $\mathbf{d} \prec \lambda$ means $d_1 = \lambda_1$ (Example (A chain, the two extremes, and size one) (d)), and $\mathbf{Q} = (1)$ works. Let $n \geq 2$, assume the theorem for $n - 1$ (for vectors in any order, by Step 1), and let $\mathbf{d}$ and λ be decreasing with $\mathbf{d} \prec \lambda$. By (M1) at $k = 1$, $d_1 \leq \lambda_1$. Subtracting (M1) at $k = n - 1$ from (M2) gives $d_n \geq \lambda_n$, so $d_1 \geq d_n \geq \lambda_n$. Let k be the smallest index in $\{1, \dots, n - 1\}$ with $\lambda_{k+1} \leq d_1$; one exists, since $\lambda_n \leq d_1$. Then $\lambda_k \geq d_1$: for $k = 1$ this is $d_1 \leq \lambda_1$, and for $k \geq 2$ the minimality of k gives $\lambda_k > d_1$. So

$$\lambda_k \geq d_1 \geq \lambda_{k+1}.$$

By Lemma 17.8.3 with $t = d_1$, there are $\mathbf{R} \in \mathbf{O}(2)$ and a real b with

$$\mathbf{R} \operatorname{diag}(\lambda_k, \lambda_{k+1}) \mathbf{R}^\top = \begin{pmatrix} d_1 & b \\ b & \eta \end{pmatrix}, \quad \eta = \lambda_k + \lambda_{k+1} - d_1.$$

Moreover $\lambda_{k+1} \leq \eta \leq \lambda_k$, since $\eta - \lambda_{k+1} = \lambda_k - d_1 \geq 0$ and $\lambda_k - \eta = d_1 - \lambda_{k+1} \geq 0$.

Step 3: the leftover data. Put

$$\begin{aligned} \mathbf{d}' &= (d_2, \dots, d_n), \\ \lambda' &= (\eta, \lambda_1, \dots, \lambda_{k-1}, \lambda_{k+2}, \dots, \lambda_n), \end{aligned}$$

two vectors in $\mathbb{R}^{n-1}$. So λ' is λ with the pair λ_k, λ_{k+1} replaced by the single number η , written first.

Claim. $\mathbf{d}' \prec \lambda'$.

Proof. By Step 2, $\lambda_{k-1} \geq \lambda_k \geq \eta \geq \lambda_{k+1} \geq \lambda_{k+2}$, for those of these indices that exist. So

$$(\mu_1, \dots, \mu_{n-1}) = (\lambda_1, \dots, \lambda_{k-1}, \eta, \lambda_{k+2}, \dots, \lambda_n)$$

is $(\lambda')^\downarrow$, and $\mathbf{d}'$ is already decreasing. For (M2), $\sum_i d'_i = \sum_i d_i - d_1 = \sum_i \lambda_i - d_1 = \sum_i \lambda'_i$, by (M2) for $\mathbf{d} \prec \lambda$ and the definition of η . For (M1), fix m with $1 \leq m \leq n-2$; when $n=2$ there is nothing to check.

Case $m \leq k-1$. Then $\mu_i = \lambda_i$ for $i \leq m$. For each such i , $d_{i+1} \leq d_1 \leq \lambda_k \leq \lambda_i$, because $\mathbf{d}$ and λ are decreasing and $i < k$. Summing over $i \leq m$ gives $d_2 + \dots + d_{m+1} \leq \mu_1 + \dots + \mu_m$.

Case $m \geq k$. The first m entries of $(\lambda')^\downarrow$ are $\lambda_1, \dots, \lambda_{k-1}$, then η , then $\lambda_{k+2}, \dots, \lambda_{m+1}$. So

$$\begin{aligned} \mu_1 + \dots + \mu_m &= (\lambda_1 + \dots + \lambda_{m+1}) - \lambda_k - \lambda_{k+1} + \eta \\ &= (\lambda_1 + \dots + \lambda_{m+1}) - d_1. \end{aligned}$$

Since $m+1 \leq n-1$, (M1) for $\mathbf{d} \prec \lambda$ applies at $m+1$ and gives

$$\begin{aligned} d_2 + \dots + d_{m+1} &= (d_1 + \dots + d_{m+1}) - d_1 \\ &\leq (\lambda_1 + \dots + \lambda_{m+1}) - d_1 = \mu_1 + \dots + \mu_m. \end{aligned}$$

In both cases (M1) holds, which proves the claim. ■

Step 4: assembly. Let $\sigma \in S_n$ list the indices $k, k+1$ first and the others after them in increasing order: $\sigma(1) = k$, $\sigma(2) = k+1$, and $\sigma(3), \dots, \sigma(n)$ are $1, \dots, k-1, k+2, \dots, n$. By (†),

$$\mathbf{P}_\sigma^\top \text{diag}(\lambda_1, \dots, \lambda_n) \mathbf{P}_\sigma = \begin{pmatrix} \text{diag}(\lambda_k, \lambda_{k+1}) & \mathbf{O} \\ \mathbf{O} & \mathbf{D}_1 \end{pmatrix},$$

where $\mathbf{D}_1 = \text{diag}(\lambda_1, \dots, \lambda_{k-1}, \lambda_{k+2}, \dots, \lambda_n) \in M_{n-2}(\mathbb{R})$; when $n=2$ the blocks involving $\mathbf{D}_1$ are absent. Let $\mathbf{R}_1 = \begin{pmatrix} \mathbf{R} & \mathbf{O} \\ \mathbf{O} & \mathbf{I}_{n-2} \end{pmatrix}$. Multiplying blocks and using Step 2,

$$\mathbf{M} = \mathbf{R}_1 \mathbf{P}_\sigma^\top \text{diag}(\lambda_1, \dots, \lambda_n) \mathbf{P}_\sigma \mathbf{R}_1^\top = \begin{pmatrix} d_1 & b & \mathbf{0}^\top \\ b & \eta & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_1 \end{pmatrix} = \begin{pmatrix} d_1 & b \mathbf{e}_1^\top \\ b \mathbf{e}_1 & \mathbf{E} \end{pmatrix},$$

where $\mathbf{e}_1$ is the first standard basis vector of $\mathbb{R}^{n-1}$ and $\mathbf{E} = \text{diag}(\eta, \lambda_1, \dots, \lambda_{k-1}, \lambda_{k+2}, \dots, \lambda_n)$ is the diagonal matrix of λ' . By the Claim and the inductive hypothesis, there is $\mathbf{W} \in \text{O}(n-1)$ such that $\mathbf{W} \mathbf{E} \mathbf{W}^\top$ has diagonal entries $d_2, \dots, d_n$. Let $\mathbf{W}_1 = \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{W} \end{pmatrix}$. Then

$$\mathbf{W}_1 \mathbf{M} \mathbf{W}_1^\top = \begin{pmatrix} d_1 & b (\mathbf{W} \mathbf{e}_1)^\top \\ b \mathbf{W} \mathbf{e}_1 & \mathbf{W} \mathbf{E} \mathbf{W}^\top \end{pmatrix},$$

whose diagonal entries are $d_1, d_2, \dots, d_n$. The block matrices $\mathbf{R}_1$ and $\mathbf{W}_1$ are orthogonal, since $\mathbf{R}_1^\top \mathbf{R}_1$ and $\mathbf{W}_1^\top \mathbf{W}_1$ are block diagonal with blocks $\mathbf{R}^\top \mathbf{R} = \mathbf{I}$, $\mathbf{I}$, and 1 , $\mathbf{W}^\top \mathbf{W} = \mathbf{I}$. So $\mathbf{Q} = \mathbf{W}_1 \mathbf{R}_1 \mathbf{P}_\sigma^\top$ is orthogonal, and $\mathbf{Q} \text{diag}(\lambda_1, \dots, \lambda_n) \mathbf{Q}^\top = \mathbf{W}_1 \mathbf{M} \mathbf{W}_1^\top$. This completes the induction and proves the theorem. ■

Over $\mathbb{C}$ nothing more is needed, because an orthogonal matrix is unitary and a real symmetric matrix is Hermitian. Together with Schur's theorem, this settles the question the section began with.

Theorem 17.8.5. Schur–Horn Theorem

Let $F = \mathbb{R}$ or $F = \mathbb{C}$, and let $\mathbf{d}, \boldsymbol{\lambda} \in \mathbb{R}^n$. There is a Hermitian $\mathbf{A} \in M_n(F)$ with $\boldsymbol{\lambda}(\mathbf{A}) = \boldsymbol{\lambda}^\downarrow$ and $\mathbf{d}(\mathbf{A}) = \mathbf{d}$ if and only if $\mathbf{d} \prec \boldsymbol{\lambda}$. In words: the diagonals of the Hermitian matrices with spectrum $\boldsymbol{\lambda}$ are **exactly** the vectors majorized by $\boldsymbol{\lambda}$, and real symmetric matrices already give all of them.

Proof. $(\Rightarrow)$ By Theorem 17.8.2, $\mathbf{d} = \mathbf{d}(\mathbf{A}) \prec \boldsymbol{\lambda}(\mathbf{A}) = \boldsymbol{\lambda}^\downarrow$. Since majorization sees only rearrangements, $\mathbf{d} \prec \boldsymbol{\lambda}$.

$(\Leftarrow)$ Theorem 17.8.4 gives a real symmetric $\mathbf{A} \in M_n(\mathbb{R}) \subseteq M_n(F)$ with $\mathbf{d}(\mathbf{A}) = \mathbf{d}$ and $\boldsymbol{\lambda}(\mathbf{A}) = \boldsymbol{\lambda}^\downarrow$, and a real symmetric matrix is Hermitian. ■

Example. Building a matrix from its diagonal and spectrum

Find a real symmetric matrix with eigenvalues 6, 1, 0 and diagonal (4, 2, 1).

Solution. Both vectors are decreasing, and $(4, 2, 1) \prec (6, 1, 0)$: the running totals are 4, 6 against 6, 7, and both totals are 7. Follow the proof of Theorem 17.8.4.

The corner. $d_1 = 4$ lies between $\lambda_1 = 6$ and $\lambda_2 = 1$, so $k = 1$, $\eta = 6 + 1 - 4 = 3$, and Lemma 17.8.3 gives $b = \sqrt{(6-4)(4-1)} = \sqrt{6}$. After the first rotation the matrix is

$$\mathbf{M} = \begin{pmatrix} 4 & \sqrt{6} & 0 \\ \sqrt{6} & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The leftover. $\mathbf{d}' = (2, 1)$ and $\boldsymbol{\lambda}' = (3, 0)$, with $2 \leq 3$ and $2 + 1 = 3 + 0$, so $\mathbf{d}' \prec \boldsymbol{\lambda}'$, as the Claim promises. The lemma with $t = 2$ gives $c = \sqrt{2/3}$, $s = \sqrt{1/3}$, and

$$\mathbf{W} \operatorname{diag}(3, 0) \mathbf{W}^\top = \begin{pmatrix} 2 & \sqrt{2} \\ \sqrt{2} & 1 \end{pmatrix}, \quad \mathbf{W} \mathbf{e}_1 = (c, s).$$

Assembly. The first row becomes $(4, \sqrt{6}c, \sqrt{6}s) = (4, 2, \sqrt{2})$, so

$$\mathbf{A} = \begin{pmatrix} 4 & 2 & \sqrt{2} \\ 2 & 2 & \sqrt{2} \\ \sqrt{2} & \sqrt{2} & 1 \end{pmatrix}.$$

Check directly: $\mathbf{A}(3, 2, \sqrt{2}) = (18, 12, 6\sqrt{2}) = 6(3, 2, \sqrt{2})$, $\mathbf{A}(-\sqrt{2}, \sqrt{2}, 1) = (-\sqrt{2}, \sqrt{2}, 1)$, and $\mathbf{A}(0, -1, \sqrt{2}) = (0, 0, 0)$. Three eigenvectors with distinct eigenvalues 6, 1, 0 give the whole spectrum, and the diagonal is (4, 2, 1).

17.8.6 Hadamard's inequality from the diagonal

Chapter 13 §07 proved Theorem 13.7.3: for a positive semidefinite $\mathbf{A}$, $\det \mathbf{A} \leq a_{11}a_{22} \cdots a_{nn}$, with equality if and only if $\mathbf{A}$ is diagonal or some $a_{ii} = 0$. That proof inducts on the size through Fischer's inequality (Corollary 13.7.2). A much shorter route scales $\mathbf{A}$ so that its diagonal is flat, and then compares the product of the eigenvalues with their sum. That comparison is the inequality between the arithmetic and geometric means, which the book has not yet proved; it is pure algebra, and it comes first.

Lemma 17.8.6. The Arithmetic–Geometric Mean Inequality

Let $n \geq 1$ and let $t_1, \dots, t_n$ be **nonnegative** real numbers. Then

$$t_1 t_2 \cdots t_n \leq \left(\frac{t_1 + t_2 + \cdots + t_n}{n} \right)^n,$$

with equality if and only if $t_1 = t_2 = \cdots = t_n$.

Idea. Scale so that the mean is 1. If the numbers are not all 1, one of them lies above 1 and another below. Merge that pair into the single number $t_j + t_k - 1$. This leaves $n - 1$ nonnegative numbers with mean 1, so induction applies to them. The merge can only increase the product, because

$$(t_j + t_k - 1) - t_j t_k = (t_j - 1)(1 - t_k),$$

and both factors are positive. So the original product is at most, and when it is nonzero strictly less than, a product that induction bounds by 1.

Proof. Step 1: mean 1. We prove by induction on $n \geq 1$: if $t_1, \dots, t_n \geq 0$ and $t_1 + \cdots + t_n = n$, then $t_1 \cdots t_n \leq 1$, with equality only when every $t_i = 1$. For $n = 1$ the hypothesis says $t_1 = 1$, and the product is 1.

Let $n \geq 2$ and assume the claim for $n - 1$ numbers. If every $t_i = 1$, the product is 1. Otherwise some $t_j > 1$: if every $t_i \leq 1$ with at least one strictly smaller, the sum would be less than n . Similarly some $t_k < 1$, since if every $t_i \geq 1$ with at least one strictly larger, the sum would exceed n . Then $j \neq k$. Neither the sum nor the product depends on the order of the numbers, so we may relabel so that $j = n - 1$ and $k = n$. Put

$$s = t_{n-1} + t_n - 1.$$

Then $s > 0$, since $t_{n-1} > 1$ and $t_n \geq 0$. The $n - 1$ numbers $t_1, \dots, t_{n-2}, s$ are nonnegative and add up to $n - 1$. So by the inductive hypothesis $p s \leq 1$, where $p = t_1 \cdots t_{n-2} \geq 0$ (with $p = 1$ when $n = 2$). Moreover

$$s - t_{n-1} t_n = (t_{n-1} - 1)(1 - t_n) > 0.$$

If $p = 0$, then $t_1 \cdots t_n = 0 < 1$. If $p > 0$, then $t_1 \cdots t_n = p t_{n-1} t_n < p s \leq 1$. In either case

the product is strictly less than 1 when the t_i are not all 1. This completes the induction.

Step 2: the general case. Let $m = \frac{1}{n}(t_1 + \cdots + t_n) \geq 0$. If $m = 0$, then every $t_i = 0$, since nonnegative numbers with sum 0 are all 0. Both sides are then 0, and the t_i are all equal. If $m > 0$, the numbers t_i/m are nonnegative with sum n . By Step 1,

$$\frac{t_1}{m} \cdots \frac{t_n}{m} \leq 1, \quad \text{that is,} \quad t_1 \cdots t_n \leq m^n,$$

with equality if and only if every $t_i/m = 1$, that is, every $t_i = m$. Numbers that are all equal to some c have mean $m = c$, so “every $t_i = m$ ” is the same as “all the t_i are equal”. This proves the lemma. ■

In the language of this section, the lemma says that among nonnegative vectors with a fixed total, the **flat** vector, which is the bottom of the majorization order (Example (A chain, the two extremes, and size one) (b)), has the largest product.

Corollary 17.8.7. Hadamard's Inequality, Second Proof

Let $F = \mathbb{R}$ or $F = \mathbb{C}$, and let $\mathbf{A} \in M_n(F)$ with $\mathbf{A} \succeq \mathbf{O}$. Then

$$\det \mathbf{A} \leq a_{11}a_{22} \cdots a_{nn},$$

with equality if and only if $\mathbf{A}$ is diagonal or some $a_{ii} = 0$.

This is Theorem 13.7.3 (a) again, with the same hypothesis and the same equality condition. What is new is the proof.

Idea. Suppose first that every diagonal entry equals 1. Then $\text{tr } \mathbf{A} = n$, so the eigenvalues, which are nonnegative, add up to n , and Lemma 17.8.6 gives $\det \mathbf{A} = \prod_i \lambda_i(\mathbf{A}) \leq 1 = \prod_i a_{ii}$. A general diagonal is made flat by a diagonal congruence $\mathbf{A} \mapsto \mathbf{SAS}$ with $\mathbf{S} = \text{diag}(a_{11}^{-1/2}, \dots, a_{nn}^{-1/2})$. This congruence preserves positivity and divides both sides of the inequality by the same number $a_{11} \cdots a_{nn}$. Zero diagonal entries, where $\mathbf{S}$ does not exist, are handled first and separately.

Proof. Each $a_{ii} = \langle \mathbf{A}\mathbf{e}_i, \mathbf{e}_i \rangle \geq 0$ by Definition 13.1.1, and each $\lambda_i(\mathbf{A}) \geq 0$ by Theorem 13.1.4 (b).

Case 1: some $a_{ii} = 0$. The right-hand side is 0. Suppose $\det \mathbf{A} \neq 0$. By Theorem 9.3.3, $\det \mathbf{A} = \prod_j \lambda_j(\mathbf{A})$, so every $\lambda_j(\mathbf{A})$ is nonzero, hence positive. By Theorem 13.1.5 (b), $\mathbf{A} \succ \mathbf{O}$, so $a_{ii} = \langle \mathbf{A}\mathbf{e}_i, \mathbf{e}_i \rangle > 0$, a contradiction. Hence $\det \mathbf{A} = 0$, and the inequality holds with equality, as the statement says it should when some $a_{ii} = 0$.

Case 2: every $a_{ii} > 0$. Let $\mathbf{S} = \text{diag}(a_{11}^{-1/2}, \dots, a_{nn}^{-1/2})$, a real diagonal matrix with nonzero diagonal, so $\mathbf{S}^* = \mathbf{S}$ and $\mathbf{S}$ is invertible. Put $\mathbf{B} = \mathbf{S}^* \mathbf{A} \mathbf{S} = \mathbf{SAS}$. By Proposition 13.1.2 (a), $\mathbf{B} \succeq \mathbf{O}$. Its entries are $b_{ij} = a_{ij} / \sqrt{a_{ii}a_{jj}}$, so $b_{ii} = 1$ for every i , and $\text{tr } \mathbf{B} = n$. By Theorem 9.3.3,

$$\lambda_1(\mathbf{B}) + \cdots + \lambda_n(\mathbf{B}) = n.$$

The $\lambda_i(\mathbf{B})$ are nonnegative by Theorem 13.1.4 (b). So Lemma 17.8.6 and Theorem 9.3.3 give

$$\det \mathbf{B} = \prod_{i=1}^n \lambda_i(\mathbf{B}) \leq \left(\frac{n}{n}\right)^n = 1,$$

with equality if and only if all $\lambda_i(\mathbf{B})$ are equal, hence all equal to 1. By Theorem 7.5.4, and since $\det \mathbf{S} = (a_{11} \cdots a_{nn})^{-1/2}$,

$$\det \mathbf{B} = (\det \mathbf{S})^2 \det \mathbf{A} = \frac{\det \mathbf{A}}{a_{11} a_{22} \cdots a_{nn}}.$$

Multiplying $\det \mathbf{B} \leq 1$ by the positive number $a_{11} \cdots a_{nn}$ gives $\det \mathbf{A} \leq a_{11} \cdots a_{nn}$.

For equality in Case 2: equality holds exactly when every $\lambda_i(\mathbf{B}) = 1$. If so, the spectral theorem (Corollary 12.4.3, or Corollary 12.5.2 over $\mathbb{R}$) gives $\mathbf{B} = \mathbf{U}\mathbf{I}\mathbf{U}^* = \mathbf{I}$ for some unitary (or orthogonal) $\mathbf{U}$. Then $\mathbf{A} = \mathbf{S}^{-1}\mathbf{B}\mathbf{S}^{-1} = \mathbf{S}^{-2} = \text{diag}(a_{11}, \dots, a_{nn})$ is diagonal. Conversely, if $\mathbf{A}$ is diagonal, then $\mathbf{B} = \mathbf{S}\mathbf{A}\mathbf{S} = \mathbf{I}$, and every $\lambda_i(\mathbf{B}) = 1$. So in Case 2, equality holds if and only if $\mathbf{A}$ is diagonal. Together with Case 1, this is the stated equality condition. This proves the corollary. ■

For the matrix of Example (Coordinate vectors against eigenvectors), which is positive definite since its eigenvalues are 7, 4, 1, the inequality reads $\det \mathbf{A} = 7 \cdot 4 \cdot 1 = 28 \leq 60 = 3 \cdot 4 \cdot 5$.

The two proofs show different things. The first proves more along the way: Fischer's inequality for blocks of any sizes. The second shows that once the diagonal is scaled flat, the trace and the arithmetic–geometric mean inequality suffice. Of $\mathbf{d}(\mathbf{B}) \prec \boldsymbol{\lambda}(\mathbf{B})$ it uses only clause (M2), which for a matrix is just the trace.

Remark. A proof by majorization in earnest would use all of $\mathbf{d}(\mathbf{A}) \prec \boldsymbol{\lambda}(\mathbf{A})$, without scaling, together with the fact that for vectors with **nonnegative** entries, $\mathbf{x} \prec \mathbf{y}$ implies $x_1 \cdots x_n \geq y_1 \cdots y_n$ (the product is **Schur-concave**). That fact is true but **not proved in this book yet**; Chapter 21 proves it, for the whole class of functions that reverse majorization.

Majorization is the language this chapter's eigenvalue inequalities were already using, and the Schur–Horn theorem shows that it fits the diagonal problem exactly. The same language compresses the inequality of Section 7 for arbitrary index sets, Theorem 17.7.2, into one line. The k largest entries of the vector $\boldsymbol{\lambda}(\mathbf{A}) - \boldsymbol{\lambda}(\mathbf{B})$ add up to the largest value of $\sum_j (\lambda_{i_j}(\mathbf{A}) - \lambda_{i_j}(\mathbf{B}))$ over index sets of size k , and the totals are $\text{tr } \mathbf{A} - \text{tr } \mathbf{B} = \text{tr}(\mathbf{A} - \mathbf{B})$. So a bound of every such sum by $\sum_{j \leq k} \lambda_j(\mathbf{A} - \mathbf{B})$ says exactly that

$$\boldsymbol{\lambda}(\mathbf{A}) - \boldsymbol{\lambda}(\mathbf{B}) \prec \boldsymbol{\lambda}(\mathbf{A} - \mathbf{B}).$$

The diagonal of a Hermitian matrix and the eigenvalues of a difference are both controlled by one ordering of vectors in $\mathbb{R}^n$.

17.8.7 Exercises

17.8.7.1 A. Check your understanding

Exercise. A1

- (a) State the definitions of $\mathbf{x} \prec \mathbf{y}$ and $\mathbf{x} \prec_w \mathbf{y}$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
- (b) Decide whether $(2, 1, 1) \prec (2, 2, 0)$, and whether $(2, 2, 0) \prec (2, 1, 1)$.
- (c) True or false: if $\mathbf{x} \prec \mathbf{y}$, then $x_1 \leq y_1$. Justify your answer.
- (d) State [Theorem 17.8.2](#). Which result of §06 is it a restatement of?
- (e) State [Theorem 17.8.5](#). Which half is proved by induction on n , and which lemma does each step of that induction use?

Solution.

- (a) $\mathbf{x} \prec \mathbf{y}$ means that $\sum_{i \leq k} x_i^\downarrow \leq \sum_{i \leq k} y_i^\downarrow$ for every $k = 1, \dots, n-1$, and $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$ ([Definition 17.8.1](#)). $\mathbf{x} \prec_w \mathbf{y}$ means that the same inequality holds for every $k = 1, \dots, n$, with no condition on the totals.
- (b) Both vectors are decreasing with total 4. Their running totals at $k = 1, 2$ are 2, 3 and 2, 4. Since $2 \leq 2$ and $3 \leq 4$, $(2, 1, 1) \prec (2, 2, 0)$. The reverse fails at $k = 2$, since $4 > 3$.
- (c) False. The definition compares **sorted** vectors, so the unsorted first entries can go either way. Take $\mathbf{x} = (1, 0)$ and $\mathbf{y} = (0, 1)$. Both have decreasing rearrangement $(1, 0)$, so (M1) at $k = 1$ reads $1 \leq 1$, and both totals are 1; hence $\mathbf{x} \prec \mathbf{y}$. But $x_1 = 1 > 0 = y_1$. What is true is $x_1^\downarrow \leq y_1^\downarrow$, which is (M1) at $k = 1$.
- (d) For a Hermitian $\mathbf{A} \in M_n(F)$, $F = \mathbb{R}$ or $\mathbb{C}$, $\mathbf{d}(\mathbf{A}) \prec \boldsymbol{\lambda}(\mathbf{A})$. It restates Corollary 17.6.3: applied to the positions of the k largest diagonal entries, the corollary gives (M1), and its equality at $k = n$ gives (M2).
- (e) For $\mathbf{d}, \boldsymbol{\lambda} \in \mathbb{R}^n$, there is a Hermitian matrix over $\mathbb{R}$ or $\mathbb{C}$ with diagonal $\mathbf{d}$, in the given order, and eigenvalues $\lambda_1, \dots, \lambda_n$ if and only if $\mathbf{d} \prec \boldsymbol{\lambda}$. The half “ $\mathbf{d} \prec \boldsymbol{\lambda}$ implies such a matrix exists” is Horn’s theorem, [Theorem 17.8.4](#), proved by induction on n . Each step uses [Lemma 17.8.3](#), applied to a pair of consecutive eigenvalues $\lambda_k \geq d_1 \geq \lambda_{k+1}$, to put d_1 in the corner. The other half is Schur’s theorem.

17.8.7.2 B. Practice**Exercise. B1: Determine which pairs are comparable**

For each pair, determine whether $\mathbf{x} \prec \mathbf{y}$ and whether $\mathbf{x} \prec_w \mathbf{y}$. Where a relation fails, name the exact clause and the value of k at which it fails. Justify your answer.

- (a) $\mathbf{x} = (2, 2, 1, 1)$, $\mathbf{y} = (3, 2, 1, 0)$.
- (b) $\mathbf{x} = (0, 3, 1)$, $\mathbf{y} = (4, 0, 0)$.
- (c) $\mathbf{x} = (1, 1, 0)$, $\mathbf{y} = (2, 1, 0)$.
- (d) $\mathbf{x} = (-1, 2, 2)$, $\mathbf{y} = (3, 0, 0)$.
- (e) $\mathbf{x} = (3, 1, 1, 1)$, $\mathbf{y} = (2, 2, 2, 0)$; decide also whether $\mathbf{y} \prec \mathbf{x}$.

Solution. In each case, sort first, then compare running totals.

- (a) Both are decreasing with total 6. Running totals: 2, 4, 5, 6 and 3, 5, 6, 6. Every entry of the first list is at most the matching entry of the second, so $\mathbf{x} \prec \mathbf{y}$ and hence $\mathbf{x} \prec_w \mathbf{y}$.

- (b) $\mathbf{x}^\downarrow = (3, 1, 0)$ with running totals 3, 4, 4, and $\mathbf{y}$ has running totals 4, 4, 4. The totals agree and $3 \leq 4$, $4 \leq 4$, so $\mathbf{x} \prec \mathbf{y}$ and $\mathbf{x} \prec_w \mathbf{y}$. This is [Example \(A chain, the two extremes, and size one\)](#) (c), applied after sorting.
- (c) Running totals 1, 2, 2 and 2, 3, 3. Every inequality holds, including at $k = 3$, so $\mathbf{x} \prec_w \mathbf{y}$. But the totals are $2 \neq 3$, so **(M2) fails** and $\mathbf{x} \not\prec \mathbf{y}$.
- (d) $\mathbf{x}^\downarrow = (2, 2, -1)$ with running totals 2, 4, 3, and $\mathbf{y}$ has running totals 3, 3, 3. The totals agree, but at $k = 2$, $4 > 3$, so **(M1) fails at $k = 2$** . Neither $\mathbf{x} \prec \mathbf{y}$ nor $\mathbf{x} \prec_w \mathbf{y}$ holds.
- (e) Both are decreasing with total 6. Running totals: 3, 4, 5, 6 and 2, 4, 6, 6. At $k = 1$, $3 > 2$, so **(M1) fails at $k = 1$** and $\mathbf{x} \not\prec \mathbf{y}$, and also $\mathbf{x} \not\prec_w \mathbf{y}$. For $\mathbf{y} \prec \mathbf{x}$, the inequalities $2 \leq 3$ and $4 \leq 4$ hold, but $6 > 5$ at $k = 3$, so **(M1) fails at $k = 3$** . The pair is incomparable, like [Example \(Neither majorizes the other\)](#).

Exercise. B2: Schur's theorem for one matrix

Let

$$\mathbf{A} = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 4 & 0 \\ 2 & 0 & 7 \end{pmatrix}.$$

- (a) Verify that $(2, 1, 4)$, $(-1, -2, 1)$, and $(-3, 2, 1)$ are eigenvectors of $\mathbf{A}$, and find $\lambda(\mathbf{A})$.
- (b) Write down $\mathbf{d}(\mathbf{A})$ and verify every clause of $\mathbf{d}(\mathbf{A}) \prec \lambda(\mathbf{A})$.
- (c) Hence verify Hadamard's inequality (Theorem 13.7.3) for $\mathbf{A}$.

Solution.

- (a) Multiply out:

$$\begin{aligned} \mathbf{A}(2, 1, 4) &= (6 + 2 + 8, 4 + 4, 4 + 28) = (16, 8, 32) = 8(2, 1, 4), \\ \mathbf{A}(-1, -2, 1) &= (-3 - 4 + 2, -2 - 8, -2 + 7) = (-5, -10, 5) = 5(-1, -2, 1), \\ \mathbf{A}(-3, 2, 1) &= (-9 + 4 + 2, -6 + 8, -6 + 7) = (-3, 2, 1) = 1(-3, 2, 1). \end{aligned}$$

Three distinct eigenvalues of a 3×3 matrix account for the whole spectrum, so $\lambda(\mathbf{A}) = (8, 5, 1)$.

- (b) $\mathbf{d}(\mathbf{A}) = (3, 4, 7)$, and $\mathbf{d}(\mathbf{A})^\downarrow = (7, 4, 3)$. (M1): at $k = 1$, $7 \leq 8$; at $k = 2$, $7 + 4 = 11 \leq 13 = 8 + 5$. (M2): $3 + 4 + 7 = 14 = 8 + 5 + 1$. Hence $\mathbf{d}(\mathbf{A}) \prec \lambda(\mathbf{A})$, as [Theorem 17.8.2](#) requires.
- (c) All eigenvalues are positive, so $\mathbf{A} \succ \mathbf{O}$ by Theorem 13.1.5 (b), and [Corollary 17.8.7](#) applies. By Theorem 9.3.3, $\det \mathbf{A} = 8 \cdot 5 \cdot 1 = 40$, while $a_{11}a_{22}a_{33} = 3 \cdot 4 \cdot 7 = 84$. Hence $40 \leq 84$. The inequality is strict, as it must be, since $\mathbf{A}$ is not diagonal and no diagonal entry is 0.

Exercise. B3: Which diagonals are ruled out?

A Hermitian 3×3 matrix has eigenvalues 5, 1, 0. For each proposed list of diagonal entries, determine whether [Theorem 17.8.2](#) rules it out. Justify your answer.

- (a) 4, 2, 0.

- (b) 6, 0, 0.
- (c) 3, 3, 1.
- (d) 2, 2, 2.
- (e) -1, 4, 3.

For the lists that are not ruled out, what does [Theorem 17.8.5](#) add?

Solution. Here $\lambda = (5, 1, 0)$, with running totals 5, 6 and total 6.

- (a) Running totals 4, 6, total 6. Since $4 \leq 5$ and $6 \leq 6$, the list is **not** ruled out.
- (b) At $k = 1$, $6 > 5$. **Ruled out.** No diagonal entry can exceed λ_1 .
- (c) The total is $7 \neq 6$. **Ruled out** by (M2): the trace would be wrong.
- (d) This is the flat vector with total 6, which is majorized by every vector with total 6 ([Example \(A chain, the two extremes, and size one\)](#) (b)). **Not** ruled out.
- (e) Sorted, it is $(4, 3, -1)$, with running totals 4, 7 and total 6. At $k = 2$, $7 > 6$. **Ruled out.** Independently, $-1 < 0 = \lambda_3$ contradicts the bottom-end inequality of the warning.

Schur's theorem is only a necessary condition, so for (a) and (d) it shows only that nothing forbids these diagonals. [Theorem 17.8.5](#) adds that they occur: for each of them there is a real symmetric matrix with eigenvalues 5, 1, 0 and that diagonal.

17.8.7.3 C. Going deeper

Exercise. C1: Majorization squeezes the range

- (a) Prove that if $\mathbf{x} \prec \mathbf{y}$, then $y_n^\downarrow \leq x_n^\downarrow \leq x_1^\downarrow \leq y_1^\downarrow$.
- (b) Deduce that every diagonal entry of a Hermitian $\mathbf{A}$ lies in $[\lambda_n(\mathbf{A}), \lambda_1(\mathbf{A})]$.
- (c) Show that the conclusion of (b), together with equal totals, does **not** imply majorization. Prove that no Hermitian matrix has eigenvalues 4, 1, 1, 0 and diagonal entries 3, 3, 0, 0, although every one of these entries lies in $[0, 4]$ and the totals agree.

Solution.

- (a) The inequality $x_1^\downarrow \leq y_1^\downarrow$ is (M1) at $k = 1$ (for $n = 1$, $\mathbf{x} = \mathbf{y}$ by [Example \(A chain, the two extremes, and size one\)](#) (d), and there is nothing to prove). The middle inequality holds because $\mathbf{x}^\downarrow$ is decreasing. For the left inequality, subtract (M1) at $k = n - 1$ from (M2):

$$\begin{aligned} x_n^\downarrow &= \sum_{i=1}^n x_i - \sum_{i=1}^{n-1} x_i^\downarrow \\ &\geq \sum_{i=1}^n y_i - \sum_{i=1}^{n-1} y_i^\downarrow = y_n^\downarrow. \end{aligned}$$

- (b) By [Theorem 17.8.2](#), $\mathbf{d}(\mathbf{A}) \prec \lambda(\mathbf{A})$. Apply (a) with $\mathbf{x} = \mathbf{d}(\mathbf{A})$ and $\mathbf{y} = \lambda(\mathbf{A})$, using $\lambda(\mathbf{A})^\downarrow = \lambda(\mathbf{A})$. Every diagonal entry lies between the smallest and the largest diagonal entry, which (a) places between $\lambda_n(\mathbf{A})$ and $\lambda_1(\mathbf{A})$. (This also follows from Lemma 13.5.2 at $\mathbf{x} = \mathbf{e}_i$. The point of (a) is that it is a property of majorization itself.)
- (c) The vector $(3, 3, 0, 0)$ has running totals 3, 6, 6, and $(4, 1, 1, 0)$ has 4, 5, 6. At $k = 2$,

$6 > 5$, so $(3, 3, 0, 0) \not\prec (4, 1, 1, 0)$, and by [Theorem 17.8.2](#) no Hermitian matrix has this diagonal and this spectrum. Yet $0 \leq 3, 3, 0, 0 \leq 4$, and both totals are 6. The range condition controls only the running totals at $k = 1$ and $k = n - 1$, while majorization constrains the middle ones too, which is why the example needs $n \geq 4$.

Exercise. C2: A weaker determinant bound

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

- (a) Prove that $\det \mathbf{A} \leq \left(\frac{1}{n} \operatorname{tr} \mathbf{A}\right)^n$, with equality if and only if $\mathbf{A} = c\mathbf{I}$ for some real $c \geq 0$.
- (b) Prove that $a_{11} \cdots a_{nn} \leq \left(\frac{1}{n} \operatorname{tr} \mathbf{A}\right)^n$. Deduce that Hadamard's inequality ([Theorem 13.7.3](#)) always implies the bound in (a).
- (c) For

$$\mathbf{A} = \begin{pmatrix} 2 & 0 & 2 \\ 0 & 5 & 2 \\ 2 & 2 & 6 \end{pmatrix},$$

whose eigenvectors are $(1, 2, 3)$, $(1, -2, 1)$ and $(-4, -1, 2)$, compute the three numbers $\det \mathbf{A}$, $a_{11}a_{22}a_{33}$ and $\left(\frac{1}{3} \operatorname{tr} \mathbf{A}\right)^3$.

Solution.

- (a) The eigenvalues $\lambda_i(\mathbf{A})$ are nonnegative by [Theorem 13.1.4 \(b\)](#). By [Theorem 9.3.3](#) and [Lemma 17.8.6](#),

$$\det \mathbf{A} = \prod_i \lambda_i(\mathbf{A}) \leq \left(\frac{1}{n} \sum_i \lambda_i(\mathbf{A})\right)^n = \left(\frac{1}{n} \operatorname{tr} \mathbf{A}\right)^n.$$

Equality holds if and only if all $\lambda_i(\mathbf{A})$ equal a common value $c \geq 0$. By the spectral theorem this happens if and only if $\mathbf{A} = \mathbf{U}(c\mathbf{I})\mathbf{U}^* = c\mathbf{I}$. Conversely, $\mathbf{A} = c\mathbf{I}$ gives equality directly.

- (b) The a_{ii} are nonnegative (as in the proof of [Corollary 17.8.7](#)) and add up to $\operatorname{tr} \mathbf{A}$. So [Lemma 17.8.6](#) gives $\prod_i a_{ii} \leq \left(\frac{1}{n} \operatorname{tr} \mathbf{A}\right)^n$. Combined with [Corollary 17.8.7](#), this gives $\det \mathbf{A} \leq \prod_i a_{ii} \leq \left(\frac{1}{n} \operatorname{tr} \mathbf{A}\right)^n$, so the Hadamard bound always sits between $\det \mathbf{A}$ and the bound of (a). The bound in (a) sees only the total of the diagonal, while Hadamard's bound sees each entry, and so it is never worse.
- (c) Check the eigenvectors: $\mathbf{A}(1, 2, 3) = (2+6, 10+6, 2+4+18) = (8, 16, 24)$; $\mathbf{A}(1, -2, 1) = (2+2, -10+2, 2-4+6) = (4, -8, 4)$; $\mathbf{A}(-4, -1, 2) = (-8+4, -5+4, -8-2+12) = (-4, -1, 2)$. So $\lambda(\mathbf{A}) = (8, 4, 1)$, and $\mathbf{A} \succ \mathbf{O}$. Then

$$\begin{aligned} \det \mathbf{A} &= 8 \cdot 4 \cdot 1 = 32, & a_{11}a_{22}a_{33} &= 2 \cdot 5 \cdot 6 = 60, \\ \left(\frac{\operatorname{tr} \mathbf{A}}{3}\right)^3 &= \left(\frac{13}{3}\right)^3 = \frac{2197}{27} \approx 81.4, \end{aligned}$$

and $32 \leq 60 \leq 81.4$, with both inequalities strict.

Exercise. C3: The possible diagonals form a convex set

For $\mathbf{x} \in \mathbb{R}^n$ and $1 \leq k \leq n$, write $s_k(\mathbf{x}) = x_1^\downarrow + \cdots + x_k^\downarrow$.

- (a) Prove that $s_k(\mathbf{x}) = \max\{\sum_{i \in I} x_i : I \subseteq \{1, \dots, n\}, |I| = k\}$.
- (b) Deduce that $s_k(\mathbf{x} + \mathbf{y}) \leq s_k(\mathbf{x}) + s_k(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
- (c) Let $\boldsymbol{\lambda} \in \mathbb{R}^n$. Prove that if $\mathbf{d} \prec \boldsymbol{\lambda}$ and $\mathbf{d}' \prec \boldsymbol{\lambda}$, then $t\mathbf{d} + (1-t)\mathbf{d}' \prec \boldsymbol{\lambda}$ for every $0 \leq t \leq 1$.
- (d) Hence show that if $\mathbf{A}$ and $\mathbf{B}$ are Hermitian $n \times n$ matrices with the same eigenvalues, then for every $0 \leq t \leq 1$ there is a real symmetric matrix with those eigenvalues and diagonal $t\mathbf{d}(\mathbf{A}) + (1-t)\mathbf{d}(\mathbf{B})$.

Hint for (a): if $v_1 \geq \cdots \geq v_k$ are the entries x_i , $i \in I$, in decreasing order, count how many entries of $\mathbf{x}$ are at least v_j .

Solution.

- (a) Let $|I| = k$, and list the entries x_i , $i \in I$, decreasingly as $v_1 \geq \cdots \geq v_k$. For each $j \leq k$, the entries $v_1, \dots, v_j$ sit at j different positions of $\mathbf{x}$ and are all $\geq v_j$. If $x_j^\downarrow < v_j$, then, since $\mathbf{x}^\downarrow$ is decreasing, only $x_1^\downarrow, \dots, x_{j-1}^\downarrow$ could be $\geq v_j$, which is fewer than j entries. Hence $v_j \leq x_j^\downarrow$ for every j , and summing gives $\sum_{i \in I} x_i \leq s_k(\mathbf{x})$. The bound is attained by the positions of the k largest entries, chosen as in the proof of [Theorem 17.8.2](#). This proves (a).
- (b) Let I be a set of k positions at which $\mathbf{x} + \mathbf{y}$ attains the maximum in (a). Applying (a) three times,

$$s_k(\mathbf{x} + \mathbf{y}) = \sum_{i \in I} x_i + \sum_{i \in I} y_i \leq s_k(\mathbf{x}) + s_k(\mathbf{y}).$$
- (c) Put $\mathbf{z} = t\mathbf{d} + (1-t)\mathbf{d}'$. Its total is $t \sum_i \lambda_i + (1-t) \sum_i \lambda_i = \sum_i \lambda_i$, by (M2) for both hypotheses. For a scalar $c \geq 0$, multiplying by c does not change the order of the entries, so $s_k(c\mathbf{x}) = c s_k(\mathbf{x})$. For $1 \leq k \leq n-1$, (b) and (M1) for both hypotheses give

$$s_k(\mathbf{z}) \leq t s_k(\mathbf{d}) + (1-t) s_k(\mathbf{d}') \leq t s_k(\boldsymbol{\lambda}) + (1-t) s_k(\boldsymbol{\lambda}) = s_k(\boldsymbol{\lambda}).$$

So $\mathbf{z} \prec \boldsymbol{\lambda}$.

- (d) Let $\boldsymbol{\lambda} = \lambda(\mathbf{A}) = \lambda(\mathbf{B})$. By [Theorem 17.8.2](#), $\mathbf{d}(\mathbf{A}) \prec \boldsymbol{\lambda}$ and $\mathbf{d}(\mathbf{B}) \prec \boldsymbol{\lambda}$. By (c), $t\mathbf{d}(\mathbf{A}) + (1-t)\mathbf{d}(\mathbf{B}) \prec \boldsymbol{\lambda}$. By [Theorem 17.8.4](#) there is a real symmetric matrix with eigenvalues $\boldsymbol{\lambda}$ and exactly this diagonal. So among Hermitian matrices with a fixed spectrum, the set of diagonals contains the segment between any two of its points: it is **convex**.