

18.8 Extreme Points

Everything so far has gone from points to convex sets: a set of points has a convex hull, and Carathéodory's theorem says how few points each element of the hull needs. This section asks the reverse question: given a convex set, which of its points cannot be manufactured from the others? A triangle is the hull of its three corners, and no smaller set will do. The corners are the **extreme points**, and the main theorem, due to Minkowski, says that a compact convex set is always the hull of its extreme points. We then find the extreme points of three unit balls and of the density matrices, show that the permutation matrices are extreme among the doubly stochastic matrices, and end with a compact convex set whose extreme points do not form a closed set.

Throughout, the field is $\mathbb{R}$, and V is a finite-dimensional real inner product space: $\mathbb{R}^n$ with the dot product, or $M_n(\mathbb{R})$ with the Frobenius inner product $\langle \mathbf{A}, \mathbf{B} \rangle = \text{tr}(\mathbf{B}^\top \mathbf{A})$ of Chapter 11 §01. **Compact** is meant as in §03, and closed bounded sets are compact by Corollary 16.2.5. Conversely a compact set is closed, as a subsequence has the limit of the whole sequence, and bounded, as a sequence with $\|x_k\| > k$ has no convergent subsequence, convergent sequences being bounded; both facts belong to the elementary algebra of limits that Chapter 16's introduction imports.

18.8.1 Points that are not in between

A convex set contains the segment between any two of its points (Definition 18.1.1). In a filled triangle, a point in the middle of an edge sits between the two ends of that edge, and an interior point sits between many pairs. A corner sits between no pair: any segment of the triangle through a corner has the corner as an end. So the corners are the points you cannot do without: delete one from the list of points whose hull you take, and the hull shrinks.

An extreme point of a convex set is a point that is not strictly between two other points of the set.

Definition 18.8.1. Extreme Point

Let $C \subseteq V$ be a convex set. A point $x \in C$ is an **extreme point** of C if, **whenever** $x = ty + (1 - t)z$ with $y, z \in C$ and $0 < t < 1$ **strictly**, then $y = z = x$. The set of extreme points of C is written $\text{ext } C$.

In words: the only way to put x strictly inside a segment of C is the useless way, with both ends equal to x . The word **strictly** matters. With $t = 1$ allowed, every point would be $1 \cdot x + 0 \cdot z$ for any $z \in C$, and no point of a set with two elements or more would be extreme.

To test the definition, midpoints are enough.

Lemma 18.8.1. The Midpoint Test

Let $C \subseteq V$ be convex and $x \in C$. Then x is an extreme point of C if and only if, whenever $x = \frac{1}{2}(y + z)$ with $y, z \in C$, we have $y = z$.

Idea. If x sits at parameter $t \leq \frac{1}{2}$ on a segment $[z, y]$, shorten the segment to $[z, y']$ with y' at parameter $2t$; then x is its midpoint.

Proof. ($\Rightarrow$) Take $t = \frac{1}{2}$ in [Definition 18.8.1](#).

($\Leftarrow$) Let $x = ty + (1 - t)z$ with $y, z \in C$ and $0 < t < 1$. Swapping the names of y and z if necessary, we may assume $t \leq \frac{1}{2}$. Put $y' = 2ty + (1 - 2t)z$. Since $0 < 2t \leq 1$, y' lies on the segment from z to y , so $y' \in C$ by convexity. Moreover

$$\frac{1}{2}(y' + z) = ty + (1 - t)z = x,$$

so the hypothesis gives $y' = z$, that is, $2t(y - z) = 0$. As $t \neq 0$, $y = z$, and then $x = ty + (1 - t)y = y$. So $y = z = x$, as claimed. ■

Now some examples, simplest first.

1. **A single point.** $C = \{x\}$ is convex, and x is extreme: any $y, z \in C$ equal x .
2. **A closed segment** $C = \{ta + (1 - t)b : 0 \leq t \leq 1\}$ with $a \neq b$ has exactly two extreme points, a and b . A point x with parameter $0 < s < 1$ is not extreme, since $x = sa + (1 - s)b$ with $a \neq b$. The end a is extreme: if $a = \frac{1}{2}(y + z)$ with y, z at parameters $r, r' \in [0, 1]$, then comparing coefficients along the line gives $\frac{1}{2}(r + r') = 1$, which forces $r = r' = 1$ and $y = z = a$. (The coefficient comparison is legitimate because $t \mapsto ta + (1 - t)b = b + t(a - b)$ is injective, as $a - b \neq 0$.)
3. **The closed half-plane** $H = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 \geq 0\}$ has **no** extreme points. Every $\mathbf{x} \in H$ is the midpoint of $\mathbf{x} + \mathbf{e}_1$ and $\mathbf{x} - \mathbf{e}_1$, both in H , and they differ.
4. **The empty set** is convex and compact with no extreme points. It satisfies the main theorem below only because $\text{conv } \emptyset = \emptyset$, which is why the theorem's last sentence says "nonempty".

The closed disk $B_2 \subseteq \mathbb{R}^2$ is a richer example: its extreme points are the whole unit circle. The section on unit balls below proves this for every n .

A nonexample by minimal change. Take the closed triangle $T = \text{conv}\{(0, 0), (2, 0), (0, 2)\}$ and slide from the corner $(2, 0)$ along the edge to its midpoint $(1, 1)$. The point $(1, 1)$ is still in T , still on its boundary, and still on a line with all of T on one side ($x_1 + x_2 \leq 2$ on T). What fails is the one clause of [Definition 18.8.1](#): $(1, 1) = \frac{1}{2}((2, 0) + (0, 2))$ with $(2, 0) \neq (0, 2)$.

Warning. **Extreme is not the same as "on the boundary"**. The point $(1, 1)$ just used lies on the boundary of the triangle T , and every open disk around it contains points outside T ; yet it is not extreme. The other direction does hold, for every convex set: an extreme point is never an interior point, since an interior point $\mathbf{x}$ is the midpoint of $\mathbf{x} \pm \varepsilon \mathbf{e}_1$ for small $\varepsilon > 0$.

Quick check. Let $S = \{(0, 0), (2, 0), (0, 2), (1, 1), (1, 0)\}$. Which points of S are extreme points of $\text{conv } S$?

Solution. Only $(0, 0)$, $(2, 0)$, and $(0, 2)$. The hull is the triangle $T = \{x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \leq 2\}$ above, because $(1, 1)$ and $(1, 0) = \frac{1}{2}((0, 0) + (2, 0))$ already lie in T . Each of those two points is the midpoint of two distinct points of T , so neither is extreme. The

corners are extreme. For $(2, 0)$: if $(2, 0) = \frac{1}{2}(\mathbf{y} + \mathbf{z})$ with $\mathbf{y}, \mathbf{z} \in T$, then $y_1, z_1 \leq 2$ and $\frac{1}{2}(y_1 + z_1) = 2$ force $y_1 = z_1 = 2$, and the only point of T with first coordinate 2 is $(2, 0)$, because $x_2 \geq 0$ and $x_1 + x_2 \leq 2$. The other two corners are handled the same way, with x_2 and with $-x_1 - x_2$ in place of x_1 .

Why this definition. It does not ask x to be far from the rest of C , or on a corner; on the disk, extreme points have extreme neighbors arbitrarily close. It asks only that x not be an average of other points of C , which is what “cannot be manufactured from the others” means for a convex hull. Exercise C1 shows that x is extreme exactly when $C \setminus \{x\}$ is still convex.

The first result makes the indispensability precise: if a convex set is the hull of S , then S contains every extreme point.

Proposition 18.8.2. Extreme Points of a Hull Come From the Generators

Let $S \subseteq V$ be any set. Then $\text{ext}(\text{conv } S) \subseteq S$.

Idea. Split off one generator from a convex combination: what remains is again a point of the hull, and x lies strictly between the two.

Proof. Let x be an extreme point of $\text{conv } S$. By Theorem 18.1.3, $x = \sum_{i=1}^k t_i s_i$ with $s_i \in S$, every $t_i > 0$ (drop the zero terms) and $\sum_i t_i = 1$. If $k = 1$ then $x = s_1 \in S$. If $k \geq 2$, then $0 < t_1 < 1$, and

$$x = t_1 s_1 + (1 - t_1)r, \quad r = \sum_{i=2}^k \frac{t_i}{1 - t_1} s_i.$$

The coefficients of r are positive and add up to 1, so $r \in \text{conv } S$ by Theorem 18.1.3. Since x is extreme, $s_1 = r = x$. In either case $x \in S$. ■

In particular a **polytope**, the hull of finitely many points, has finitely many extreme points, all among the points one started from. Section 9 is built on this.

18.8.2 Faces

A single extreme point is too small an object for an induction on dimension. The right object is a piece of the set that behaves like an extreme point as a whole.

Definition 18.8.2. Face

Let $C \subseteq V$ be convex. A **face** of C is a convex subset $F \subseteq C$ such that, **whenever** $ty + (1 - t)z \in F$ with $y, z \in C$ and $0 < t < 1$, both y and z lie in F .

So a face swallows every segment of C that passes through it at an inner point. The empty set and C itself are faces, and a one-point set $\{x\}$ is a face exactly when x is an extreme point: that is [Definition 18.8.1](#) word for word. The edges of a triangle are faces. The segment joining

the midpoints of two edges is convex but is not a face: if the triangle has corners a, b, c and the segment joins $\frac{1}{2}(a + b)$ to $\frac{1}{2}(a + c)$, its midpoint is $\frac{1}{2}a + \frac{1}{2} \cdot \frac{1}{2}(b + c)$, an inner point of a segment of the triangle whose ends a and $\frac{1}{2}(b + c)$ do not lie on it.

What makes faces useful for an induction is that they pass extremeness upwards.

Lemma 18.8.3. Extreme Points of a Face

Let $C \subseteq V$ be convex and F a face of C . Then $\text{ext } F = F \cap \text{ext } C$. In particular an extreme point of a face is an extreme point of C .

Proof. Let $x \in \text{ext } F$, and let $x = ty + (1 - t)z$ with $y, z \in C$ and $0 < t < 1$. Since F is a face, $y, z \in F$, and since x is extreme in F , $y = z = x$. So $x \in \text{ext } C$. Conversely a point of F that is extreme in C is extreme in the smaller set F , because the definition then quantifies over fewer pairs y, z . ■

The faces we can actually produce are the ones cut off by a hyperplane that has the whole set on one side. This is the situation of Theorem 18.3.6.

Proposition 18.8.4. Supporting Slices Are Faces

Let $C \subseteq V$ be convex, let $w \in V$ and $\alpha \in \mathbb{R}$, and suppose $\langle w, x \rangle \leq \alpha$ for every $x \in C$. Then:

- (a) $F = \{x \in C : \langle w, x \rangle = \alpha\}$ is a face of C ;
- (b) if F is a single point $\{c\}$, then c is an extreme point of C .

Idea. A weighted average of numbers that are all at most α equals α only if each of them does.

Proof.

- (a) F is convex, since if $\langle w, y \rangle = \langle w, z \rangle = \alpha$, the same holds for $ty + (1 - t)z$ by linearity of the inner product in its second slot. Let $y, z \in C$, $0 < t < 1$, and $ty + (1 - t)z \in F$. Then

$$\alpha = t\langle w, y \rangle + (1 - t)\langle w, z \rangle,$$

with $\langle w, y \rangle \leq \alpha$ and $\langle w, z \rangle \leq \alpha$. If either inequality were strict, the right side would be strictly less than $t\alpha + (1 - t)\alpha = \alpha$, because t and $1 - t$ are positive. So both are equalities, and $y, z \in F$.

- (b) By (a), $\{c\}$ is a face, and its only point is extreme in it. By [Lemma 18.8.3](#), $c \in \text{ext } C$. ■

A point c as in (b), the **unique** maximizer over C of some linear function $x \mapsto \langle w, x \rangle$, is called an **exposed point**. Part (b) says that exposed points are extreme; it is the most convenient test for extremeness.

18.8.3 Minkowski's theorem

The main theorem is proved by induction on dimension. Recall from §01 that a nonempty convex set C has dimension $\dim C = \dim \text{aff } C$; by Proposition 18.1.4, $\text{aff } C = x_0 + W_C$ with $W_C = \text{span}\{y - x_0 : y \in C\}$, independent of $x_0 \in C$. So $\dim C = \dim W_C$ (Definition 18.1.5 (b)). A convex set has dimension 0 exactly when it is a single point, and for every $y \in C$ the difference set $C - y = \{x - y : x \in C\}$ lies in W_C .

Theorem 18.8.5. Minkowski's Theorem

Let V be a finite-dimensional real inner product space and let $K \subseteq V$ be a compact convex set. Then

$$K = \text{conv}(\text{ext } K).$$

In particular, a nonempty compact convex set has at least one extreme point.

Idea. The inclusion $\supseteq$ is free, since K is convex and contains its extreme points. For $\subseteq$, induct on $\dim K$. Take $x \in K$ and a line through x in a direction of W_K . Compactness cuts the line down to a segment $[y, z]$ of K with x on it. An end y is not an interior point, and Theorem 18.3.6, applied **inside** W_K to the translate $K - y$, gives a hyperplane through y with K on one side. The slice of K on that hyperplane is a face of smaller dimension, so by induction y is a convex combination of its extreme points, and those are extreme in K by Lemma 18.8.3. The same goes for z , and x lies between y and z . The space matters: in all of V , a hyperplane containing all of K would qualify, and would give nothing.

Proof. ($\supseteq$) K is convex and contains $\text{ext } K$, so it contains $\text{conv}(\text{ext } K)$, the smallest convex set containing $\text{ext } K$ (Definition 18.1.4).

($\subseteq$) If $K = \emptyset$ there is nothing to prove. For nonempty K we prove, by induction on $d \geq 0$, the statement: *every nonempty compact convex set of dimension d , in any finite-dimensional real inner product space, is contained in the convex hull of its extreme points.*

If $d = 0$, then $K = \{x_0\}$, and x_0 is extreme, so $K = \text{conv}(\text{ext } K)$.

Let $d \geq 1$, assume the statement for all dimensions less than d , and let $\dim K = d$ and $x \in K$. Write $W = W_K$, and choose $u \in W$ with $u \neq 0$, which is possible because $\dim W = d \geq 1$.

Step 1: a segment through x . Let $T = \{t \in \mathbb{R} : x + tu \in K\}$, which contains 0. It is bounded: if $\|k\| \leq R$ for every $k \in K$, then for $t \in T$, $|t|\|u\| = \|(x + tu) - x\| \leq 2R$. It is closed: if $t_j \in T$ and $t_j \rightarrow t$, then $x + t_j u \rightarrow x + tu$, which lies in K because K is closed. So T is a nonempty closed bounded subset of $\mathbb{R}$, which is compact by fact (A3) of Chapter 16's introduction. The function $t \mapsto t$ is continuous, so by fact (A4) (the extreme value theorem) it attains a minimum a and a maximum b on T , and $a \leq 0 \leq b$. Put $y = x + au$ and $z = x + bu$, both in K . If $a < b$ then

$$x = \frac{b}{b-a} y + \frac{-a}{b-a} z,$$

a convex combination, since both coefficients are ≥ 0 and they add up to 1; if $a = b$ then $a = b = 0$ and $x = y = z$. Either way $x \in \text{conv}\{y, z\}$.

Step 2: a supporting hyperplane at y , inside W . With the inner product of V restricted

to it, $W \neq \{0\}$ is a finite-dimensional real inner product space. The set $K - y$ lies in W , is convex by Proposition 18.1.1 (c), and contains 0. But 0 is not an interior point of $K - y$ in W (Definition 18.3.2): for every $\varepsilon > 0$, the vector $-\varepsilon u \in W$, of norm $\varepsilon\|u\|$, is **not** in $K - y$, because $y - \varepsilon u = x + (a - \varepsilon)u \notin K$ by the choice of a . By Theorem 18.3.6, applied in W to $K - y$ and 0, there is a nonzero $w \in W$ with

$$\langle w, k - y \rangle \leq \langle w, 0 \rangle = 0 \quad \text{for every } k \in K.$$

Step 3: the face through y has smaller dimension. So $\langle w, k \rangle \leq \alpha := \langle w, y \rangle$ on K , and by Proposition 18.8.4 (a), $F = \{k \in K : \langle w, k \rangle = \alpha\}$ is a face of K . It is nonempty, as $y \in F$, and convex. It is bounded, being inside K , and closed: if $k_j \in F$ and $k_j \rightarrow k$, then $k \in K$, and $|\langle w, k_j \rangle - \langle w, k \rangle| \leq \|w\|\|k_j - k\| \rightarrow 0$ by Theorem 11.1.5, so $\langle w, k \rangle = \alpha$. So F is compact. Using $y \in F$ as base point, W_F is spanned by vectors $k - y$ with $k \in F$, each of which lies in

$$W' = \{v \in W : \langle w, v \rangle = 0\},$$

the kernel of the linear functional $v \mapsto \langle w, v \rangle$ on W . That functional is nonzero, since it sends w to $\|w\|^2 > 0$, so its image is $\mathbb{R}$ and Theorem 2.3.1 gives $\dim W' = d - 1$. Hence $\dim F = \dim W_F \leq d - 1$ by Theorem 1.6.7.

Step 4: induction. By the induction hypothesis, $y \in \text{conv}(\text{ext } F)$, and $\text{ext } F \subseteq \text{ext } K$ by Lemma 18.8.3. So $y \in \text{conv}(\text{ext } K)$. Running Steps 2 to 4 with $-u$ in place of u , which exchanges the roles of a and b , gives $z \in \text{conv}(\text{ext } K)$. Since $\text{conv}(\text{ext } K)$ is convex and contains y and z , Step 1 gives $x \in \text{conv}(\text{ext } K)$. As $x \in K$ was arbitrary, this proves the inclusion and the theorem. The last sentence of the statement follows, since $\text{conv } \emptyset = \emptyset$. ■

Warning. Compactness cannot be dropped. The quarter-plane $Q = \{x_1 \geq 0, x_2 \geq 0\}$ is closed and convex with exactly one extreme point, the origin, and $\text{conv}\{\mathbf{0}\} = \{\mathbf{0}\} \neq Q$. The open disk, bounded but not closed, has no extreme points at all, since each of its points $\mathbf{x}$ is the midpoint of $\mathbf{x} \pm \varepsilon \mathbf{e}_1$ for small $\varepsilon > 0$.

Combined with Carathéodory's theorem (Theorem 18.2.2), Minkowski's theorem says more: every point of a compact convex K of dimension d is a convex combination of at most $d + 1$ extreme points. To see this, pick $x_0 \in K$ and apply Theorem 18.2.2 to the translate $\text{ext } K - x_0$, which lies in the d -dimensional space W_K ; translating back preserves convex combinations, because their coefficients add up to 1.

The consequence used most often concerns linear functions. By the Riesz representation theorem (Theorem 11.5.1), every linear functional on V is $x \mapsto \langle x, a \rangle$ for some $a \in V$, and it is continuous by Theorem 11.1.5, since $|\langle x, a \rangle - \langle x', a \rangle| \leq \|a\|\|x - x'\|$.

Corollary 18.8.6. Linear Functions Peak at Extreme Points

Let $K \subseteq V$ be a nonempty compact convex set and $a \in V$. Then the maximum of $x \mapsto \langle x, a \rangle$ over K exists and is attained at an extreme point of K . The same holds for the minimum.

Idea. Minkowski's theorem writes a maximizer as an average of extreme points, and an average of values at most M equals M only if every value does.

Proof. By fact (A4) of Chapter 16's introduction, the continuous function $x \mapsto \langle x, a \rangle$ attains a maximum M on K , at some $x^* \in K$. By Theorem 18.8.5 and Theorem 18.1.3, $x^* = \sum_{i=1}^k t_i v_i$ with $v_i \in \text{ext } K$, $t_i > 0$, and $\sum_i t_i = 1$. Then

$$M = \langle x^*, a \rangle = \sum_{i=1}^k t_i \langle v_i, a \rangle,$$

while each $\langle v_i, a \rangle \leq M$. If some $\langle v_i, a \rangle < M$, the weighted average would be strictly less than M , because every $t_i > 0$. So $\langle v_i, a \rangle = M$ for every i , and v_1 is an extreme point at which the maximum is attained. For the minimum, apply this to $-a$. ■

This is why linear programs are solved at corners: when one of Section 6's feasible regions is bounded and nonempty, the corollary puts an optimum at an extreme point, and Section 9 shows that there are finitely many. The same corollary gives a new view of Chapter 17's variational formulas, as we will see after the density matrices.

18.8.4 Three unit balls

The unit balls of the three norms of Chapter 16 (Definition 16.1.2, Example (The 1-, 2-, and ∞ -norms on F^n)) are compact and convex. They are convex by the remark after Definition 16.1.2. They are bounded, and closed, because $\|\mathbf{x}_k\| \leq 1$ and $\mathbf{x}_k \rightarrow \mathbf{x}$ give $\|\mathbf{x}\| \leq \|\mathbf{x}_k\| + \|\mathbf{x} - \mathbf{x}_k\| \leq 1 + \|\mathbf{x} - \mathbf{x}_k\|$ for every k , hence $\|\mathbf{x}\| \leq 1$; so they are compact by Corollary 16.2.5.

Example. The diamond, the cube, and the sphere

In $\mathbb{R}^n$ with $n \geq 1$, show that:

- (a) $\text{ext } B_1 = \{\pm \mathbf{e}_1, \dots, \pm \mathbf{e}_n\}$, $2n$ points;
- (b) $\text{ext } B_\infty = \{\mathbf{s} : \text{every } s_i = \pm 1\}$, the 2^n sign vectors;
- (c) $\text{ext } B_2 = \{\mathbf{x} : \|\mathbf{x}\|_2 = 1\}$, the whole unit sphere.

Solution. In each case we show that the listed points are extreme and that no other point is. The second half uses the same device each time: a point $\mathbf{x}$ of the ball is not extreme as soon as we find $\mathbf{v} \neq \mathbf{0}$ with $\mathbf{x} \pm \mathbf{v}$ both in the ball, because $\mathbf{x}$ is their midpoint.

- (a) *The points $\pm \mathbf{e}_i$ are extreme.* Let $\mathbf{e}_i = \frac{1}{2}(\mathbf{y} + \mathbf{z})$ with $\|\mathbf{y}\|_1, \|\mathbf{z}\|_1 \leq 1$. Then

$$1 = \frac{1}{2}(y_i + z_i) \leq \frac{1}{2}(|y_i| + |z_i|) \leq \frac{1}{2}(\|\mathbf{y}\|_1 + \|\mathbf{z}\|_1) \leq 1,$$

so every inequality is an equality. The first forces $y_i = |y_i|$ and $z_i = |z_i|$. The second, since $|y_i| \leq \|\mathbf{y}\|_1$ and $|z_i| \leq \|\mathbf{z}\|_1$ separately, forces $|y_i| = \|\mathbf{y}\|_1$, so all other entries of $\mathbf{y}$ vanish, and the third forces $\|\mathbf{y}\|_1 = 1$. Hence $\mathbf{y} = \mathbf{e}_i$, and likewise $\mathbf{z} = \mathbf{e}_i$. By Lemma 18.8.1, $\mathbf{e}_i$ is extreme; for $-\mathbf{e}_i$ apply this to $-\mathbf{y}, -\mathbf{z}$.

No other point is extreme. Let $\mathbf{x} \in B_1$. If $\|\mathbf{x}\|_1 < 1$, put $\varepsilon = 1 - \|\mathbf{x}\|_1 > 0$ and $\mathbf{v} = \varepsilon \mathbf{e}_1$; then $\|\mathbf{x} \pm \mathbf{v}\|_1 \leq \|\mathbf{x}\|_1 + \varepsilon = 1$. If $\|\mathbf{x}\|_1 = 1$ and $\mathbf{x}$ is not some $\pm \mathbf{e}_i$, then $\mathbf{x}$ has two nonzero entries x_i, x_j with $i \neq j$. Let $\sigma_i, \sigma_j \in \{\pm 1\}$ be their signs, $\varepsilon = \min(|x_i|, |x_j|) > 0$, and $\mathbf{v} = \varepsilon(\sigma_i \mathbf{e}_i - \sigma_j \mathbf{e}_j)$. Adding $\mathbf{v}$ moves $|x_i|$ up by ε and $|x_j|$ down by ε , and subtracting $\mathbf{v}$ does the opposite; neither entry changes sign, because $\varepsilon \leq |x_i|, |x_j|$. So $\|\mathbf{x} \pm \mathbf{v}\|_1 = \|\mathbf{x}\|_1 = 1$.

- (b) *Sign vectors are extreme.* Let $\mathbf{s} = \frac{1}{2}(\mathbf{y} + \mathbf{z})$ with $\|\mathbf{y}\|_\infty, \|\mathbf{z}\|_\infty \leq 1$. For each i , $s_i = \pm 1$ is the average of y_i and z_i , both in $[-1, 1]$. If $s_i = 1$, then $y_i + z_i = 2$ with $y_i, z_i \leq 1$ forces $y_i = z_i = 1$; if $s_i = -1$, then $y_i + z_i = -2$ with $y_i, z_i \geq -1$ forces $y_i = z_i = -1$. So $\mathbf{y} = \mathbf{z} = \mathbf{s}$.

No other point is extreme. If $\mathbf{x} \in B_\infty$ is not a sign vector, some $|x_i| < 1$. Put $\mathbf{v} = (1 - |x_i|)\mathbf{e}_i \neq \mathbf{0}$. Then the i -th entries of $\mathbf{x} \pm \mathbf{v}$ are $x_i \pm (1 - |x_i|)$, both in $[-1, 1]$, and the other entries are unchanged.

- (c) *Unit vectors are extreme.* Let $\|\mathbf{x}\|_2 = 1$ and $\mathbf{x} = \frac{1}{2}(\mathbf{y} + \mathbf{z})$ with $\|\mathbf{y}\|_2, \|\mathbf{z}\|_2 \leq 1$. The parallelogram law of Chapter 11 §01, which comes from expanding with Theorem 11.1.2 (c), gives

$$\|\mathbf{y} - \mathbf{z}\|_2^2 = 2\|\mathbf{y}\|_2^2 + 2\|\mathbf{z}\|_2^2 - \|\mathbf{y} + \mathbf{z}\|_2^2 \leq 2 + 2 - \|2\mathbf{x}\|_2^2 = 0,$$

so $\mathbf{y} = \mathbf{z}$.

No other point is extreme. If $\|\mathbf{x}\|_2 < 1$, put $\mathbf{v} = (1 - \|\mathbf{x}\|_2)\mathbf{e}_1$; then $\|\mathbf{x} \pm \mathbf{v}\|_2 \leq \|\mathbf{x}\|_2 + (1 - \|\mathbf{x}\|_2) = 1$.

In (a) and (b), Minkowski's theorem (Theorem 18.8.5) now makes B_1 the hull of the $2n$ points $\pm \mathbf{e}_i$, and B_∞ the hull of the 2^n sign vectors. In (c) it says only that the ball is the hull of the sphere, and the infinitely many extreme points are why B_2 is not a polytope.

18.8.5 Density matrices

Sums and nonnegative multiples of positive semidefinite matrices are positive semidefinite, because $\mathbf{x}^\top (s\mathbf{A} + t\mathbf{B})\mathbf{x} = s\mathbf{x}^\top \mathbf{A}\mathbf{x} + t\mathbf{x}^\top \mathbf{B}\mathbf{x}$; Chapter 13 §01 recorded this in Exercise (C2: The cone of positive matrices) (a). Cutting that cone with a single linear condition gives a compact convex set whose extreme points we can find exactly.

Definition 18.8.3. Density Matrix

A **density matrix** of size n is a real symmetric matrix $\mathbf{X} \in M_n(\mathbb{R})$ with $\mathbf{X} \succeq \mathbf{0}$ and $\text{tr } \mathbf{X} = 1$. We write $\mathcal{D}_n$ for the set of them.

The name comes from quantum mechanics, which uses the complex Hermitian version; the proofs below carry over, with $\mathbf{x}^*$ for $\mathbf{x}^\top$, $|\mathbf{x}^* \mathbf{y}|^2$ for $(\mathbf{x}^\top \mathbf{y})^2$, and Corollary 12.4.3 for Corollary 12.5.2. The set $\mathcal{D}_n$ is convex: if $\mathbf{X}, \mathbf{Y} \in \mathcal{D}_n$ and $0 \leq t \leq 1$, then $t\mathbf{X} + (1-t)\mathbf{Y}$ is symmetric, $\mathbf{x}^\top (t\mathbf{X} + (1-t)\mathbf{Y})\mathbf{x} = t\mathbf{x}^\top \mathbf{X}\mathbf{x} + (1-t)\mathbf{x}^\top \mathbf{Y}\mathbf{x} \geq 0$, and the trace is $t + (1-t) = 1$. It is compact in $M_n(\mathbb{R})$: the three conditions are preserved by entrywise limits, since symmetry and trace are linear and each inequality $\mathbf{x}^\top \mathbf{X}_j \mathbf{x} \geq 0$ survives the limit; and it is bounded, since by Corollary 12.5.2

$$\|\mathbf{X}\|_F^2 = \text{tr}(\mathbf{X}^2) = \sum_{i=1}^n \lambda_i(\mathbf{X})^2 \leq \left(\sum_{i=1}^n \lambda_i(\mathbf{X}) \right)^2 = 1,$$

where the inequality uses $\lambda_i(\mathbf{X}) \geq 0$ (Theorem 13.1.4).

For a unit vector $\mathbf{x} \in \mathbb{R}^n$, the matrix $\mathbf{x}\mathbf{x}^\top$ is in $\mathcal{D}_n$: it is symmetric, $\mathbf{y}^\top \mathbf{x}\mathbf{x}^\top \mathbf{y} = (\mathbf{x}^\top \mathbf{y})^2 \geq 0$, and $\text{tr}(\mathbf{x}\mathbf{x}^\top) = \mathbf{x}^\top \mathbf{x} = 1$. It is the orthogonal projection onto the line $\text{span}\{\mathbf{x}\}$. One more fact about positive semidefinite matrices is needed: a zero value of the quadratic form pins down a null vector.

Lemma 18.8.7. A Zero of the Form Is a Null Vector

Let $\mathbf{B} \in M_n(\mathbb{R})$ be symmetric with $\mathbf{B} \succeq \mathbf{O}$, and let $\mathbf{y} \in \mathbb{R}^n$ satisfy $\mathbf{y}^\top \mathbf{B} \mathbf{y} = 0$. Then $\mathbf{B} \mathbf{y} = \mathbf{0}$.

Proof. By Corollary 12.5.2, $\mathbf{B} = \sum_{i=1}^n \mu_i \mathbf{q}_i \mathbf{q}_i^\top$ for an orthonormal basis $(\mathbf{q}_1, \dots, \mathbf{q}_n)$ of eigenvectors, with eigenvalues $\mu_i \geq 0$ by Theorem 13.1.4. Then $0 = \mathbf{y}^\top \mathbf{B} \mathbf{y} = \sum_i \mu_i (\mathbf{q}_i^\top \mathbf{y})^2$, a sum of nonnegative terms, so $\mu_i (\mathbf{q}_i^\top \mathbf{y})^2 = 0$ for every i . Hence $\mu_i (\mathbf{q}_i^\top \mathbf{y}) = 0$ for every i , and $\mathbf{B} \mathbf{y} = \sum_i \mu_i (\mathbf{q}_i^\top \mathbf{y}) \mathbf{q}_i = \mathbf{0}$. ■

Theorem 18.8.8. Extreme Points of the Density Matrices

For $n \geq 1$,

$$\text{ext } \mathcal{D}_n = \{\mathbf{x}\mathbf{x}^\top : \mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\| = 1\},$$

the orthogonal projections of rank one. Every $\mathbf{X} \in \mathcal{D}_n$ is a convex combination of at most n of them.

Idea. The spectral theorem writes $\mathbf{X}$ as $\sum_i \lambda_i \mathbf{q}_i \mathbf{q}_i^\top$, with weights $\lambda_i \geq 0$ adding up to $\text{tr } \mathbf{X} = 1$. That is already a convex combination of rank-one projections, which gives the last sentence and shows at once that a matrix with two positive eigenvalues is not extreme. For the converse, suppose $\mathbf{x}\mathbf{x}^\top$ is an average of two density matrices. On every vector orthogonal to $\mathbf{x}$, the form of $\mathbf{x}\mathbf{x}^\top$ is zero, so the forms of both pieces are zero there too, and Lemma 18.8.7 turns that into “both pieces kill $\mathbf{x}^\perp$ ”. A symmetric matrix that kills $\mathbf{x}^\perp$ is a multiple of $\mathbf{x}\mathbf{x}^\top$, and the trace fixes the multiple.

Proof. Let $\mathbf{X} \in \mathcal{D}_n$. By Corollary 12.5.2, $\mathbf{X} = \sum_{i=1}^n \lambda_i \mathbf{q}_i \mathbf{q}_i^\top$ with $(\mathbf{q}_1, \dots, \mathbf{q}_n)$ orthonormal and $\lambda_i = \lambda_i(\mathbf{X})$. Each $\lambda_i \geq 0$ by Theorem 13.1.4, and $\sum_i \lambda_i = \text{tr } \mathbf{X} = 1$, the trace being the sum of the eigenvalues. Dropping the zero terms, this writes $\mathbf{X}$ as a convex combination of at most n matrices $\mathbf{q}_i \mathbf{q}_i^\top$, which proves the last sentence.

Only rank-one projections are extreme. At least one eigenvalue of $\mathbf{X}$ is positive, since the trace is 1. Suppose first that $\mathbf{X}$ has at least two positive eigenvalues. As they are listed decreasingly, $\lambda_1 \geq \lambda_2 > 0$, and then $0 < \lambda_1 < 1$, because the total is 1 and λ_2 is part of it. Put

$$\mathbf{Y} = \frac{1}{1 - \lambda_1} \sum_{i=2}^n \lambda_i \mathbf{q}_i \mathbf{q}_i^\top.$$

Its coefficients are nonnegative and add up to 1, so $\mathbf{Y} \in \mathcal{D}_n$ by the convexity of $\mathcal{D}_n$, and $\mathbf{X} = \lambda_1 \mathbf{q}_1 \mathbf{q}_1^\top + (1 - \lambda_1) \mathbf{Y}$. The two ends differ, since $\mathbf{q}_1 \mathbf{q}_1^\top \mathbf{q}_1 = \mathbf{q}_1$ while $\mathbf{Y} \mathbf{q}_1 = \mathbf{0}$ by orthonormality. So $\mathbf{X}$ is not extreme. If instead $\mathbf{X}$ has exactly one positive eigenvalue, that eigenvalue is the whole trace, 1, and $\mathbf{X} = \mathbf{q}_1 \mathbf{q}_1^\top$. So every extreme point has the form $\mathbf{x}\mathbf{x}^\top$

with $\|\mathbf{x}\| = 1$.

Every rank-one projection is extreme. Let $\|\mathbf{x}\| = 1$ and $\mathbf{xx}^\top = \frac{1}{2}(\mathbf{B} + \mathbf{C})$ with $\mathbf{B}, \mathbf{C} \in \mathcal{D}_n$. For every $\mathbf{y}$ with $\mathbf{x}^\top \mathbf{y} = 0$,

$$0 = (\mathbf{x}^\top \mathbf{y})^2 = \mathbf{y}^\top \mathbf{xx}^\top \mathbf{y} = \frac{1}{2} \mathbf{y}^\top \mathbf{B} \mathbf{y} + \frac{1}{2} \mathbf{y}^\top \mathbf{C} \mathbf{y},$$

a sum of two nonnegative numbers, so $\mathbf{y}^\top \mathbf{B} \mathbf{y} = 0$, and $\mathbf{B} \mathbf{y} = \mathbf{0}$ by Lemma 18.8.7. Now let $\mathbf{v} \in \mathbb{R}^n$ be arbitrary, and write $\mathbf{v} = (\mathbf{x}^\top \mathbf{v})\mathbf{x} + \mathbf{y}$ with $\mathbf{y} = \mathbf{v} - (\mathbf{x}^\top \mathbf{v})\mathbf{x}$, so that $\mathbf{x}^\top \mathbf{y} = \mathbf{x}^\top \mathbf{v} - \mathbf{x}^\top \mathbf{v} = 0$ because $\mathbf{x}^\top \mathbf{x} = 1$. Then $\mathbf{B} \mathbf{v} = (\mathbf{x}^\top \mathbf{v})\mathbf{B} \mathbf{x}$ for every $\mathbf{v}$, that is, $\mathbf{B} = (\mathbf{B} \mathbf{x})\mathbf{x}^\top$. Since $\mathbf{B}$ is symmetric, also $\mathbf{B} = \mathbf{B}^\top = \mathbf{x}(\mathbf{B} \mathbf{x})^\top$, and applying both expressions to $\mathbf{x}$ gives $\mathbf{B} \mathbf{x} = (\mathbf{x}^\top \mathbf{B} \mathbf{x})\mathbf{x}$. Hence $\mathbf{B} = c\mathbf{xx}^\top$ with $c = \mathbf{x}^\top \mathbf{B} \mathbf{x}$, and $1 = \text{tr} \mathbf{B} = c \text{tr}(\mathbf{xx}^\top) = c$. So $\mathbf{B} = \mathbf{xx}^\top$, and by symmetry of the argument $\mathbf{C} = \mathbf{xx}^\top$. By Lemma 18.8.1, $\mathbf{xx}^\top$ is extreme. This proves the theorem. ■

For instance, $\mathbf{X} = \frac{1}{15} \begin{pmatrix} 9 & 2 \\ 2 & 6 \end{pmatrix}$ is symmetric with trace 1, and $\mathbf{X}(2, 1) = \frac{2}{3}(2, 1)$, $\mathbf{X}(1, -2) = \frac{1}{3}(1, -2)$. With $\mathbf{q}_1 = \frac{1}{\sqrt{5}}(2, 1)$ and $\mathbf{q}_2 = \frac{1}{\sqrt{5}}(1, -2)$, the proof's splitting is

$$\begin{aligned} \mathbf{X} &= \frac{2}{3} \mathbf{q}_1 \mathbf{q}_1^\top + \frac{1}{3} \mathbf{q}_2 \mathbf{q}_2^\top \\ &= \frac{2}{3} \cdot \frac{1}{5} \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix} + \frac{1}{3} \cdot \frac{1}{5} \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}, \end{aligned}$$

and the entries check: $\frac{8+1}{15} = \frac{9}{15}$, $\frac{4-2}{15} = \frac{2}{15}$, $\frac{2+4}{15} = \frac{6}{15}$.

Here the spectral theorem handed us the decomposition promised by Minkowski's theorem, with at most n terms. Carathéodory's theorem, in the coset of symmetric matrices of trace 1, of dimension $n(n+1)/2 - 1$, would only have promised $n(n+1)/2$. Structure beats counting.

Now the promised instance of Corollary 18.8.6. For a real symmetric $\mathbf{A}$, the function $\mathbf{X} \mapsto \text{tr}(\mathbf{AX}) = \langle \mathbf{X}, \mathbf{A} \rangle$ is linear on $M_n(\mathbb{R})$, with $\langle \cdot, \cdot \rangle$ the Frobenius inner product.

Example. The largest eigenvalue as a linear maximum

Let $\mathbf{A} \in M_n(\mathbb{R})$ be symmetric. Show that $\lambda_1(\mathbf{A}) = \max\{\text{tr}(\mathbf{AX}) : \mathbf{X} \in \mathcal{D}_n\}$, and find the maximum for $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$.

Solution. The set $\mathcal{D}_n$ is nonempty, compact, and convex, so by Corollary 18.8.6 the maximum exists and is attained at an extreme point, which by Theorem 18.8.8 is $\mathbf{xx}^\top$ with $\|\mathbf{x}\| = 1$. There,

$$\text{tr}(\mathbf{Axx}^\top) = \text{tr}(\mathbf{x}^\top \mathbf{A} \mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x} \leq \lambda_1(\mathbf{A})$$

by the cyclic property of the trace and Lemma 13.5.2. The bound is attained when $\mathbf{x}$ is a unit eigenvector for $\lambda_1(\mathbf{A})$, by the same lemma, so the maximum is $\lambda_1(\mathbf{A})$.

For the given $\mathbf{A}$, $\mathbf{A}(1, 1) = (3, 3)$ and $\mathbf{A}(1, -1) = (1, -1)$, so $\lambda_1(\mathbf{A}) = 3$, and the maximizer

is $\mathbf{x}\mathbf{x}^\top$ with $\mathbf{x} = \frac{1}{\sqrt{2}}(1, 1)$:

$$\mathbf{X} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{A}\mathbf{X} = \frac{1}{2} \begin{pmatrix} 3 & 3 \\ 3 & 3 \end{pmatrix}, \quad \text{tr}(\mathbf{A}\mathbf{X}) = 3.$$

This is the case $k = 1$ of Ky Fan's maximum principle (Theorem 17.6.2); for $k > 1$, $\text{tr}(\mathbf{Q}^\top \mathbf{A}\mathbf{Q}) = \text{tr}(\mathbf{A}\mathbf{Q}\mathbf{Q}^\top)$ is again linear in the projection $\mathbf{Q}\mathbf{Q}^\top$, but this section does not show that the rank- k projections are the extreme points of any convex set.

18.8.6 Doubly stochastic matrices

Chapter 9 met **stochastic** matrices, whose entries are nonnegative and whose columns add up to 1 (Definition 9.11.1). Asking the rows to add up to 1 as well gives a set with remarkable corners.

Definition 18.8.4. Doubly Stochastic Matrix

A matrix $\mathbf{M} \in M_n(\mathbb{R})$ is **doubly stochastic** if all its entries are nonnegative and **every** row and **every** column adds up to 1. We write Ω_n for the set of doubly stochastic $n \times n$ matrices.

Every permutation matrix $\mathbf{P}_\sigma$ is doubly stochastic, since it has exactly one 1 in each row and each column and zeros elsewhere (the remark after Definition 3.4.2). So is $\frac{1}{n}\mathbf{J}$, the matrix with every entry $\frac{1}{n}$. The set Ω_n is convex, since nonnegativity and the $2n$ linear conditions on sums survive convex combinations. It is compact in $M_n(\mathbb{R})$: its entries lie in $[0, 1]$, so it is bounded, and each defining condition, being a nonstrict inequality or a linear equation in the entries, survives limits. By Minkowski's theorem, Ω_n is the hull of its extreme points, and the natural candidates are the permutation matrices.

Proposition 18.8.9. Permutation Matrices Are Extreme

Every permutation matrix is an extreme point of Ω_n .

Idea. Nonnegative entries cannot cancel, so a zero of $\mathbf{P}$ forces zeros in both pieces, and the row sums then leave no freedom.

Proof. Let $\mathbf{P} = \mathbf{P}_\sigma$ and $\mathbf{P} = t\mathbf{A} + (1-t)\mathbf{B}$ with $\mathbf{A}, \mathbf{B} \in \Omega_n$ and $0 < t < 1$. Wherever $\mathbf{P}$ has a zero entry, $0 = ta_{ij} + (1-t)b_{ij}$ with both terms nonnegative, so $a_{ij} = b_{ij} = 0$. Each row of $\mathbf{A}$ therefore has at most one nonzero entry, in the position where $\mathbf{P}$ has its 1, and since the row adds up to 1, that entry is 1. So $\mathbf{A} = \mathbf{P}$, and likewise $\mathbf{B} = \mathbf{P}$. ■

The proof never used the column sums. The converse is much deeper.

Birkhoff's theorem. *The extreme points of Ω_n are exactly the $n!$ permutation matrices. Equivalently, every doubly stochastic matrix is a convex combination of permutation matrices.*

The two forms are equivalent by what we have. If the permutation matrices are all the extreme points, Minkowski's theorem (Theorem 18.8.5) makes Ω_n their hull. Conversely, if Ω_n is the hull of the permutation matrices, Proposition 18.8.2 puts every extreme point among them, and Proposition 18.8.9 shows that each of them is one. We do not prove Birkhoff's theorem here; Chapter 19 does. For $n = 2$ it can be checked by hand, as the next Quick check shows.

Quick check. Show that every $\mathbf{M} \in \Omega_2$ has the form $\begin{pmatrix} t & 1-t \\ 1-t & t \end{pmatrix}$ with $0 \leq t \leq 1$, and deduce Birkhoff's theorem for $n = 2$.

Solution. Let $t = m_{11}$. The first row sum gives $m_{12} = 1 - t$, the first column sum gives $m_{21} = 1 - t$, and the second row sum gives $m_{22} = 1 - (1 - t) = t$. Nonnegativity gives $0 \leq t \leq 1$. So

$$\mathbf{M} = t \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + (1 - t) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

a convex combination of the two 2×2 permutation matrices. Hence Ω_2 is the segment between them, and its extreme points are its two ends.

Birkhoff's theorem would also give a second route to Schur's theorem, Theorem 17.8.2, which Chapter 17 §08 proved without it. If $\mathbf{A} = \mathbf{QDQ}^\top$ with $\mathbf{Q}$ orthogonal and $\mathbf{D} = \text{diag}(\lambda_1(\mathbf{A}), \dots, \lambda_n(\mathbf{A}))$ (Corollary 12.5.2, after reordering), then $\mathbf{d}(\mathbf{A}) = \mathbf{S}\boldsymbol{\lambda}(\mathbf{A})$ with $\mathbf{S} = (q_{ij}^2)$, doubly stochastic because the rows and columns of $\mathbf{Q}$ are unit vectors, so $\mathbf{d}(\mathbf{A})$ would be an average of rearrangements of $\boldsymbol{\lambda}(\mathbf{A})$, each majorized by $\boldsymbol{\lambda}(\mathbf{A})$, hence majorized by $\boldsymbol{\lambda}(\mathbf{A})$ itself by Example (The vectors majorized by a fixed vector).

18.8.7 When the extreme points are not closed

In every example so far, the extreme points have formed a closed set: finitely many points, a sphere, or the rank-one projections. That is not always so.

Example. A double cone over a circle

In $\mathbb{R}^3$, let $\Gamma = \{(1, 0, 0) + (\mathbf{u}, 0) : \mathbf{u} \in \mathbb{R}^2, \|\mathbf{u}\| = 1\}$ be the circle of radius 1 centered at $(1, 0, 0)$ in the plane $x_3 = 0$, where $(\mathbf{u}, 0) = (u_1, u_2, 0)$; it passes through the origin, at $\mathbf{u} = (-1, 0)$. Let $\mathbf{p} = (0, 0, 1)$, $\mathbf{q} = (0, 0, -1)$, $S = \Gamma \cup \{\mathbf{p}, \mathbf{q}\}$, and $K = \text{conv } S$. Show that K is compact and convex, that

$$\text{ext } K = (\Gamma \setminus \{\mathbf{0}\}) \cup \{\mathbf{p}, \mathbf{q}\},$$

and that this set is not closed.

Solution. Compact and convex. K is convex as a convex hull. The set S is bounded and closed: a limit of points of Γ satisfies the closed conditions $(x_1 - 1)^2 + x_2^2 = 1$ and $x_3 = 0$, which describe Γ , and adding two points keeps it closed. So S is compact, and K is compact by Corollary 18.2.3.

The candidates. By Proposition 18.8.2, $\text{ext } K \subseteq S$.

The origin is not extreme. $\mathbf{0} \in \Gamma \subseteq K$, and $\mathbf{0} = \frac{1}{2}(\mathbf{p} + \mathbf{q})$ with $\mathbf{p} \neq \mathbf{q}$.

The points $\mathbf{p}$ and $\mathbf{q}$ are extreme. Let $\varphi(\mathbf{x}) = x_3$. On S , $\varphi(\mathbf{p}) = 1$, $\varphi(\mathbf{q}) = -1$, and $\varphi = 0$ on Γ . A point $\mathbf{x} = \sum_i t_i \mathbf{s}_i$ of K , written as a convex combination with every $t_i > 0$ (Theorem 18.1.3), has $\varphi(\mathbf{x}) = \sum_i t_i \varphi(\mathbf{s}_i) \leq 1$, with equality only if every $\mathbf{s}_i = \mathbf{p}$, that is, only if $\mathbf{x} = \mathbf{p}$. So $\varphi \leq 1$ on K with equality exactly at $\mathbf{p}$, and $\mathbf{p}$ is extreme by Proposition 18.8.4 (b). The same argument with $-\varphi$ handles $\mathbf{q}$.

The other points of Γ are extreme. Fix $\mathbf{c} = (1, 0, 0) + (\mathbf{u}, 0)$ with $\mathbf{c} \neq \mathbf{0}$, that is, $\mathbf{u} \neq (-1, 0)$; then $u_1 > -1$, because a unit vector with $u_1 = -1$ is $(-1, 0)$. Let $\varphi(\mathbf{x}) = \langle \mathbf{u}, (x_1, x_2) \rangle = u_1 x_1 + u_2 x_2$, so that $\varphi(\mathbf{c}) = u_1 + \|\mathbf{u}\|^2 = u_1 + 1$. For a point $\mathbf{s} = (1, 0, 0) + (\mathbf{u}', 0)$ of Γ ,

$$\varphi(\mathbf{s}) = u_1 + \langle \mathbf{u}, \mathbf{u}' \rangle \leq u_1 + \|\mathbf{u}\| \|\mathbf{u}'\| = u_1 + 1$$

by Theorem 11.1.5. Equality forces $(\mathbf{u}, \mathbf{u}')$ to be linearly dependent, so $\mathbf{u}' = \pm \mathbf{u}$, and $\mathbf{u}' = -\mathbf{u}$ would give $\langle \mathbf{u}, \mathbf{u}' \rangle = -1$; so equality holds only at $\mathbf{u}' = \mathbf{u}$, that is, at $\mathbf{s} = \mathbf{c}$. Also $\varphi(\mathbf{p}) = \varphi(\mathbf{q}) = 0 < 1 + u_1 = \varphi(\mathbf{c})$. So $\varphi(\mathbf{s}) < \varphi(\mathbf{c})$ for every $\mathbf{s} \in S$ other than $\mathbf{c}$. As before, a convex combination $\mathbf{x} = \sum_i t_i \mathbf{s}_i$ with all $t_i > 0$ has $\varphi(\mathbf{x}) \leq \varphi(\mathbf{c})$, with equality only if every $\mathbf{s}_i = \mathbf{c}$. By Proposition 18.8.4 (b), $\mathbf{c}$ is extreme.

Not closed. For $k \geq 1$ let $\mathbf{c}_k = (1 - \sqrt{1 - 1/k^2}, 1/k, 0)$. It is $(1, 0, 0) + (\mathbf{u}_k, 0)$ with $\mathbf{u}_k = (-\sqrt{1 - 1/k^2}, 1/k)$, a unit vector, so $\mathbf{c}_k \in \Gamma$; and $\mathbf{c}_k \neq \mathbf{0}$, as its second entry is $1/k$. So every $\mathbf{c}_k$ is extreme. For $0 \leq h \leq 1$ we have $(1 - h)^2 \leq 1 - h$, hence $1 - h \leq \sqrt{1 - h} \leq 1$ and $0 \leq 1 - \sqrt{1 - h} \leq h$. With $h = 1/k^2$, the first entry of $\mathbf{c}_k$ lies in $[0, 1/k^2]$, so $\mathbf{c}_k \rightarrow \mathbf{0}$, which is not extreme. So $\text{ext } K$ contains a sequence converging to a point outside it, and it is not closed (Definition 18.1.6).

Geometrically, K is a double cone over the disk bounded by Γ , with apexes $\mathbf{p}$ and $\mathbf{q}$. The segment from $\mathbf{p}$ to $\mathbf{q}$ touches the rim of the disk at the origin, which robs the origin of its extremeness while leaving its neighbors on the circle intact. So the set of extreme points of a compact convex set need not be closed, and hence need not be compact.

18.8.8 Exercises

18.8.8.1 A. Check your understanding

Exercise. A1

- State the definition of an extreme point of a convex set $C \subseteq V$.
- True or false: every boundary point of a compact convex set in $\mathbb{R}^2$ is an extreme point. Justify your answer.
- True or false: there is a nonempty closed convex subset of $\mathbb{R}^2$ with no extreme points. Justify your answer.
- State Theorem 18.8.5, and name the two facts from Chapter 16's introduction that its proof uses.
- True or false: if $K = \text{conv } S$ with S finite, then every point of S is an extreme point of K . Justify your answer.

Solution.

- (a) A point $x \in C$ is extreme if $x = ty + (1 - t)z$ with $y, z \in C$ and $0 < t < 1$ forces $y = z = x$.
- (b) False. In the square $[-1, 1]^2$, the point $(1, 0)$ is on the boundary, since $(1 + \varepsilon, 0)$ lies outside for every $\varepsilon > 0$, but it is the midpoint of $(1, 1)$ and $(1, -1)$, both in the square.
- (c) True. The closed half-plane $\{x_2 \geq 0\}$ is one: each of its points $\mathbf{x}$ is the midpoint of $\mathbf{x} \pm \mathbf{e}_1$, which lie in it. A line is another.
- (d) A compact convex subset K of a finite-dimensional real inner product space is the convex hull of its extreme points. The proof uses (A3), to see that the set of t with $x + tu \in K$ is compact, and (A4), the extreme value theorem, to find its smallest and largest elements.
- (e) False. Take $S = \{0, 1, 2\} \subseteq \mathbb{R}$. Then $K = [0, 2]$, whose extreme points are 0 and 2; the point $1 = \frac{1}{2}(0 + 2)$ of S is not extreme. What is true is the reverse inclusion $\text{ext } K \subseteq S$ ([Proposition 18.8.2](#)).

18.8.8.2 B. Practice

Exercise. B1: Finding extreme points

Determine the extreme points of each of the following convex subsets of $\mathbb{R}^2$. Justify your answer.

- (a) $\text{conv}\{(0, 0), (4, 0), (0, 2), (2, 1), (1, 1)\}$.
- (b) $P = \{(x, y) : x^2 \leq y \leq 1\}$.
- (c) The strip $\{(x, y) : 0 \leq y \leq 1\}$.

Hint: for (b), use [Proposition 18.8.4](#) (b) with a tangent line.

Solution.

- (a) By [Proposition 18.8.2](#), every extreme point is one of the five listed points. The point $(2, 1) = \frac{1}{2}((4, 0) + (0, 2))$ is not extreme. Neither is $(1, 1) = \frac{1}{2}((2, 1) + (0, 1))$, since $(0, 1) = \frac{1}{2}((0, 0) + (0, 2))$ lies in the hull, and $(2, 1) \neq (0, 1)$. The three corners are extreme: each is the unique maximizer over the five points of a linear function, namely $-x - y$, x , and y respectively, hence the unique maximizer over the hull by the convex-combination argument of [Example \(A double cone over a circle\)](#), and [Proposition 18.8.4](#) (b) applies. So the extreme points are $(0, 0)$, $(4, 0)$, and $(0, 2)$.

- (b) *Points (a, a^2) with $-1 \leq a \leq 1$ are extreme.* Let $\varphi(x, y) = 2ax - y$. For $(x, y) \in P$,

$$\varphi(x, y) \leq 2ax - x^2 = a^2 - (x - a)^2 \leq a^2,$$

with equality only if $x = a$ and $y = x^2 = a^2$. So (a, a^2) is the unique maximizer of φ over P , and it lies in P because $a^2 \leq 1$. By [Proposition 18.8.4](#) (b) it is extreme.

No other point is extreme. Let $(x, y) \in P$ with $y \neq x^2$, so $x^2 < y \leq 1$. If $y < 1$, let $\varepsilon = \min(y - x^2, 1 - y) > 0$; then $(x, y \pm \varepsilon) \in P$ and (x, y) is their midpoint. If $y = 1$, then $|x| < 1$; let $\varepsilon = 1 - |x| > 0$; then $(x \pm \varepsilon)^2 \leq (|x| + \varepsilon)^2 = 1$, so $(x \pm \varepsilon, 1) \in P$, and $(x, 1)$ is their midpoint. So $\text{ext } P = \{(a, a^2) : -1 \leq a \leq 1\}$.

- (c) None: every point $\mathbf{x}$ of the strip is the midpoint of $\mathbf{x} \pm \mathbf{e}_1$, which lie in the strip. The strip is closed and convex but not bounded, so Minkowski's theorem ([Theorem 18.8.5](#)) does not apply.

Exercise. B2: A density matrix split into pure pieces

Let

$$\mathbf{X} = \begin{pmatrix} 1/2 & 1/4 \\ 1/4 & 1/2 \end{pmatrix}, \quad \mathbf{Y} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} 1/2 & 1 \\ 1 & 1/2 \end{pmatrix}.$$

- (a) For each of the three matrices, decide whether it lies in $\mathcal{D}_2$, and if so whether it is an extreme point.
- (b) Write $\mathbf{X}$ as a convex combination of two extreme points of $\mathcal{D}_2$.

Solution.

- (a) All three are symmetric with trace 1. For $\mathbf{X}$: $\mathbf{X}(1,1) = \frac{3}{4}(1,1)$ and $\mathbf{X}(1,-1) = \frac{1}{4}(1,-1)$, so its eigenvalues are $\frac{3}{4}$ and $\frac{1}{4}$, both nonnegative. Hence $\mathbf{X} \in \mathcal{D}_2$ (Theorem 13.1.4), and it has two positive eigenvalues, so it is **not** extreme by Theorem 18.8.8. For $\mathbf{Y}$: $\mathbf{Y} = \mathbf{xx}^\top$ with $\mathbf{x} = \frac{1}{\sqrt{2}}(1,1)$, a unit vector, so $\mathbf{Y} \in \mathcal{D}_2$ and it **is** extreme. For $\mathbf{Z}$: $\det \mathbf{Z} = \frac{1}{4} - 1 = -\frac{3}{4} < 0$, so the product of its eigenvalues is negative and one of them is negative. Hence $\mathbf{Z}$ is not positive semidefinite, and $\mathbf{Z} \notin \mathcal{D}_2$.
- (b) With the unit eigenvectors $\mathbf{q}_1 = \frac{1}{\sqrt{2}}(1,1)$ and $\mathbf{q}_2 = \frac{1}{\sqrt{2}}(1,-1)$,

$$\begin{aligned} \mathbf{X} &= \frac{3}{4} \mathbf{q}_1 \mathbf{q}_1^\top + \frac{1}{4} \mathbf{q}_2 \mathbf{q}_2^\top \\ &= \frac{3}{4} \cdot \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \frac{1}{4} \cdot \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \end{aligned}$$

The entries check: the diagonal is $\frac{3}{8} + \frac{1}{8} = \frac{1}{2}$ and the off-diagonal is $\frac{3}{8} - \frac{1}{8} = \frac{1}{4}$. The weights $\frac{3}{4}, \frac{1}{4}$ are nonnegative and add up to 1, and both matrices are extreme points by Theorem 18.8.8.

Exercise. B3: A doubly stochastic matrix from permutations

Let

$$\mathbf{M} = \frac{1}{6} \begin{pmatrix} 3 & 2 & 1 \\ 1 & 3 & 2 \\ 2 & 1 & 3 \end{pmatrix}.$$

- (a) Check that $\mathbf{M} \in \Omega_3$.
- (b) Write $\mathbf{M}$ as a convex combination of three permutation matrices. Hence explain why $\mathbf{M}$ is not an extreme point of Ω_3 .

Solution.

- (a) The entries are nonnegative. Each row of $6\mathbf{M}$ is a rearrangement of 3, 2, 1, and so is each column (the columns are (3, 1, 2), (2, 3, 1), (1, 2, 3)), so every row and column of $\mathbf{M}$ adds up to $6/6 = 1$.
- (b) The entries equal to 3 sit on the diagonal, the entries equal to 2 in positions (1, 2), (2, 3), (3, 1), and the entries equal to 1 in positions (1, 3), (2, 1), (3, 2). Each of these three patterns has exactly one position in each row and each column, so each

is a permutation matrix:

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad \mathbf{P}^2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

and $\mathbf{I}$. Hence

$$\mathbf{M} = \frac{1}{2}\mathbf{I} + \frac{1}{3}\mathbf{P} + \frac{1}{6}\mathbf{P}^2,$$

with weights nonnegative and adding up to 1. Since $\mathbf{M} = \frac{1}{2}\mathbf{I} + \frac{1}{2}(\frac{2}{3}\mathbf{P} + \frac{1}{3}\mathbf{P}^2)$, with $\frac{2}{3}\mathbf{P} + \frac{1}{3}\mathbf{P}^2 \in \Omega_3$ by convexity and different from $\mathbf{I}$ (its diagonal is zero), $\mathbf{M}$ is the midpoint of two distinct points of Ω_3 and is not extreme.

18.8.8.3 C. Going deeper

Exercise. C1: Deleting an extreme point

Let $C \subseteq V$ be convex and $x \in C$.

- (a) Prove that x is an extreme point of C if and only if $C \setminus \{x\}$ is convex.
- (b) Deduce that x is an extreme point of C if and only if $x \notin \text{conv}(C \setminus \{x\})$.

Solution.

- (a) ($\Rightarrow$) Let $y, z \in C \setminus \{x\}$ and $0 \leq t \leq 1$. The point $w = ty + (1-t)z$ lies in C by convexity. If $t = 0$ or $t = 1$, w is z or y , which is not x . If $0 < t < 1$ and $w = x$, extremality would give $y = x$, which is false. So $w \neq x$, and $w \in C \setminus \{x\}$. ($\Leftarrow$) Suppose $x = ty + (1-t)z$ with $y, z \in C$, $0 < t < 1$, and not both equal to x . If exactly one of them, say z , equals x , then $x = ty + (1-t)x$, so $t(y-x) = 0$ and $y = x$, a contradiction. So neither equals x , both lie in $C \setminus \{x\}$, and their combination x does not: $C \setminus \{x\}$ is not convex.
- (b) If x is extreme, $C \setminus \{x\}$ is convex by (a), so it is its own hull (Definition 18.1.4), which does not contain x . If x is not extreme, (a) gives $y, z \in C \setminus \{x\}$ and $t \in [0, 1]$ with $ty + (1-t)z \notin C \setminus \{x\}$; this point lies in C , so it is x , and $x \in \text{conv}(C \setminus \{x\})$.

Exercise. C2: Every contraction is an average of orthogonal matrices

Let $\mathcal{B} = \{\mathbf{A} \in M_n(\mathbb{R}) : \|\mathbf{A}\|_2 \leq 1\}$, the unit ball of the operator 2-norm (Definition 16.3.2), so that $\mathbf{A} \in \mathcal{B}$ exactly when $\|\mathbf{Ax}\| \leq \|\mathbf{x}\|$ for every $\mathbf{x} \in \mathbb{R}^n$.

- (a) Prove that every orthogonal matrix is an extreme point of $\mathcal{B}$.
- (b) Prove that every $\mathbf{A} \in \mathcal{B}$ that is not orthogonal is **not** an extreme point of $\mathcal{B}$.
- (c) Deduce that every $\mathbf{A} \in \mathcal{B}$ is a convex combination of orthogonal matrices.

Hint for (b): use the singular value decomposition (Theorem 13.9.1) and change only the smallest singular value.

Solution.

- (a) Let $\mathbf{Q}$ be orthogonal and $\mathbf{Q} = \frac{1}{2}(\mathbf{B} + \mathbf{C})$ with $\mathbf{B}, \mathbf{C} \in \mathcal{B}$. Let $\mathbf{x}$ be a unit vector. Then $\mathbf{Q}\mathbf{x}$ is a unit vector, $\mathbf{B}\mathbf{x}$ and $\mathbf{C}\mathbf{x}$ lie in the Euclidean unit ball B_2 , and $\mathbf{Q}\mathbf{x} = \frac{1}{2}(\mathbf{B}\mathbf{x} + \mathbf{C}\mathbf{x})$. By [Example \(The diamond, the cube, and the sphere\)](#) (c), $\mathbf{Q}\mathbf{x}$ is an extreme point of B_2 , so $\mathbf{B}\mathbf{x} = \mathbf{C}\mathbf{x}$ by [Lemma 18.8.1](#). This holds for every unit $\mathbf{x}$, hence for every $\mathbf{x}$ by scaling, so $\mathbf{B} = \mathbf{C}$. By [Lemma 18.8.1](#) again, $\mathbf{Q}$ is extreme.
- (b) By Theorem 13.9.1, $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ with $\mathbf{U}, \mathbf{V}$ orthogonal and $\mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_n)$, $\sigma_1 \geq \dots \geq \sigma_n \geq 0$, and $\sigma_1 = \|\mathbf{A}\|_2 \leq 1$ by Theorem 16.3.3 (c). If $\sigma_n = 1$, then every $\sigma_i = 1$ and $\mathbf{A} = \mathbf{U}\mathbf{V}^\top$ is orthogonal. So $\sigma_n < 1$. Put $\varepsilon = 1 - \sigma_n > 0$, let $\mathbf{u}_n, \mathbf{v}_n$ be the last columns of $\mathbf{U}, \mathbf{V}$, and $\mathbf{A}_\pm = \mathbf{A} \pm \varepsilon \mathbf{u}_n \mathbf{v}_n^\top = \mathbf{U}\mathbf{D}_\pm \mathbf{V}^\top$ with $\mathbf{D}_\pm = \text{diag}(\sigma_1, \dots, \sigma_{n-1}, \sigma_n \pm \varepsilon)$. The last diagonal entries are 1 and $2\sigma_n - 1 \in [-1, 1]$, so every diagonal entry d_i of $\mathbf{D}_\pm$ has $|d_i| \leq 1$. Since $\mathbf{U}$ and $\mathbf{V}^\top$ preserve Euclidean length,
- $$\|\mathbf{A}_\pm \mathbf{x}\| = \|\mathbf{D}_\pm \mathbf{V}^\top \mathbf{x}\| \leq \|\mathbf{V}^\top \mathbf{x}\| = \|\mathbf{x}\|,$$
- the inequality because $\mathbf{D}_\pm$ multiplies each coordinate by a number of absolute value at most 1. So $\mathbf{A}_\pm \in \mathcal{B}$, $\mathbf{A} = \frac{1}{2}(\mathbf{A}_+ + \mathbf{A}_-)$, and $\mathbf{A}_+ \neq \mathbf{A}_-$ because $\mathbf{u}_n \mathbf{v}_n^\top \neq \mathbf{O}$. Hence $\mathbf{A}$ is not extreme.
- (c) $\mathcal{B}$ is convex, being the unit ball of a norm (the remark after Definition 16.1.2). It is closed and bounded for $\|\cdot\|_2$: if $\|\mathbf{A}_j - \mathbf{A}\|_2 \rightarrow 0$ with $\|\mathbf{A}_j\|_2 \leq 1$, then $\|\mathbf{A}\|_2 \leq \|\mathbf{A}_j\|_2 + \|\mathbf{A} - \mathbf{A}_j\|_2 \leq 1 + \|\mathbf{A} - \mathbf{A}_j\|_2$ for every j , and letting $j \rightarrow \infty$ gives $\|\mathbf{A}\|_2 \leq 1$. By Corollary 16.2.5 it is compact for the Frobenius norm as well. By (a) and (b), ext $\mathcal{B}$ is the set of orthogonal matrices, and [Theorem 18.8.5](#) gives $\mathcal{B} = \text{conv}(\text{O}(n))$.

Exercise. C3: The corners of the Loewner interval

Let $\mathcal{E} = \{\mathbf{X} \in M_n(\mathbb{R}) : \mathbf{X}^\top = \mathbf{X}, \mathbf{O} \preceq \mathbf{X} \preceq \mathbf{I}\}$, where $\mathbf{X} \preceq \mathbf{I}$ means $\mathbf{I} - \mathbf{X} \succeq \mathbf{O}$ (Definition 13.5.1).

- (a) Prove that every symmetric $\mathbf{P}$ with $\mathbf{P}^2 = \mathbf{P}$, that is, every orthogonal projection, is an extreme point of $\mathcal{E}$.
- (b) Prove that no other point of $\mathcal{E}$ is extreme.
- (c) Deduce that every $\mathbf{X} \in \mathcal{E}$ is a convex combination of orthogonal projections, and give such a combination for $\mathbf{X} = \text{diag}(\frac{1}{2}, \frac{1}{4})$.

Hint for (a): use [Lemma 18.8.7](#) twice, once on $\ker \mathbf{P}$ and once on $\text{im } \mathbf{P}$.

Solution.

- (a) First, $\mathbf{P} \in \mathcal{E}$: its eigenvalues satisfy $\lambda^2 = \lambda$, so each is 0 or 1, and $\mathbf{P}$ and $\mathbf{I} - \mathbf{P}$ have eigenvalues in $\{0, 1\}$, hence are positive semidefinite (Theorem 13.1.4). Let $\mathbf{P} = \frac{1}{2}(\mathbf{A} + \mathbf{B})$ with $\mathbf{A}, \mathbf{B} \in \mathcal{E}$. If $\mathbf{P}\mathbf{x} = \mathbf{0}$, then $0 = \mathbf{x}^\top \mathbf{P}\mathbf{x} = \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} + \frac{1}{2}\mathbf{x}^\top \mathbf{B}\mathbf{x}$ with both terms ≥ 0 , so $\mathbf{x}^\top \mathbf{A}\mathbf{x} = 0$ and $\mathbf{A}\mathbf{x} = \mathbf{0}$ by [Lemma 18.8.7](#). If $\mathbf{P}\mathbf{x} = \mathbf{x}$, then $0 = \mathbf{x}^\top (\mathbf{I} - \mathbf{P})\mathbf{x} = \frac{1}{2}\mathbf{x}^\top (\mathbf{I} - \mathbf{A})\mathbf{x} + \frac{1}{2}\mathbf{x}^\top (\mathbf{I} - \mathbf{B})\mathbf{x}$, so $\mathbf{x}^\top (\mathbf{I} - \mathbf{A})\mathbf{x} = 0$ and, since $\mathbf{I} - \mathbf{A} \succeq \mathbf{O}$, $(\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$, that is, $\mathbf{A}\mathbf{x} = \mathbf{x}$. Every $\mathbf{v} \in \mathbb{R}^n$ is $(\mathbf{v} - \mathbf{P}\mathbf{v}) + \mathbf{P}\mathbf{v}$, where $\mathbf{P}(\mathbf{v} - \mathbf{P}\mathbf{v}) = \mathbf{0}$ and $\mathbf{P}(\mathbf{P}\mathbf{v}) = \mathbf{P}\mathbf{v}$. So $\mathbf{A}\mathbf{v} = \mathbf{0} + \mathbf{P}\mathbf{v} = \mathbf{P}\mathbf{v}$ for every $\mathbf{v}$, and $\mathbf{A} = \mathbf{P}$; then $\mathbf{B} = 2\mathbf{P} - \mathbf{A} = \mathbf{P}$. By [Lemma 18.8.1](#), $\mathbf{P}$ is extreme.
- (b) Let $\mathbf{X} \in \mathcal{E}$ not be idempotent. By Corollary 12.5.2, $\mathbf{X} = \sum_i \lambda_i \mathbf{q}_i \mathbf{q}_i^\top$ with $\mathbf{q}_i$ orthonor-

mal, and every $\lambda_i \in [0, 1]$ because $\mathbf{X} \succeq \mathbf{O}$ and $\mathbf{I} - \mathbf{X} \succeq \mathbf{O}$ have eigenvalues λ_i and $1 - \lambda_i$. If every λ_i were 0 or 1, $\mathbf{X}^2 = \sum_i \lambda_i^2 \mathbf{q}_i \mathbf{q}_i^\top = \mathbf{X}$; so some λ_1 , say, lies strictly between 0 and 1. Put $\varepsilon = \min(\lambda_1, 1 - \lambda_1) > 0$ and $\mathbf{X}_\pm = \mathbf{X} \pm \varepsilon \mathbf{q}_1 \mathbf{q}_1^\top$. These have the same eigenvectors, with λ_1 replaced by $\lambda_1 \pm \varepsilon \in [0, 1]$, so $\mathbf{X}_\pm \in \mathcal{E}$; they differ, and $\mathbf{X}$ is their midpoint.

- (c) $\mathcal{E}$ is convex (both conditions survive convex combinations, as for $\mathcal{D}_n$), closed (each condition survives limits), and bounded (its eigenvalues lie in $[0, 1]$, so $\|\mathbf{X}\|_F^2 = \sum_i \lambda_i^2 \leq n$). By (a), (b) and [Theorem 18.8.5](#), $\mathcal{E}$ is the convex hull of the orthogonal projections. For $\mathbf{X} = \text{diag}(\frac{1}{2}, \frac{1}{4})$, reading off the eigenvalues in layers,

$$\mathbf{X} = \frac{1}{4} \mathbf{I} + \frac{1}{4} \text{diag}(1, 0) + \frac{1}{2} \mathbf{O},$$

a convex combination of the projections $\mathbf{I}$, $\text{diag}(1, 0)$, and $\mathbf{O}$ with weights $\frac{1}{4}, \frac{1}{4}, \frac{1}{2}$. The diagonal entries check: $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ and $\frac{1}{4}$.