Chapter II: Convex Geometry

1 Convex, Affine, and Conic hulls

We will first introduce the concept of lines, segments, and rays, which play a central role for the definition of affine, convex, and conic sets. The construction of these three notions is similar to the construction of a vector space (which, as you should recall, relies on the possibility to take *linear combinations of points* $\sum_i \lambda_i \boldsymbol{x}_i$), except that here we will put some constraints on the coefficients λ_i, leading to what we call affine, convex, or conic combinations of points. We will present these constructions in parallel, since they only differ by the constraints imposed on the coefficients of the combinations.

Definition 1. (Lines, segments, rays). Let $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$.

- The *line* through $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is $\{\theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 : \ \theta \in \mathbb{R}\}$.
- The *segment* between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is $\{\theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 : \ \theta \in [0,1]\}$.
- The *ray* through $\boldsymbol{x}_1$ is $\{ \ \theta \boldsymbol{x}_1 : \theta \geq 0\}$.

Note that the ray through $\boldsymbol{x}$ is in fact the half-line with end-point $\mathbf{0}$ that goes through $\boldsymbol{x}$.

Definition 2. (Affine, Convex, and Conic sets). Let S be a subset of $\mathbb{R}^n$.

- We say that S is an *affine set* if it contains the line through any 2 points of S:
$$\boldsymbol{x}_1, \boldsymbol{x}_2 \in S, \theta \in \mathbb{R} \implies \theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 \in S.$$
- We say that S is *convex* if it contains the segment through any 2 points of S:
$$\boldsymbol{x}_1, \boldsymbol{x}_2 \in S, \theta \in [0,1] \implies \theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 \in S.$$
- We say that S is a *cone* if it contains the ray through any point of S:
$$\boldsymbol{x} \in S, \theta \geq 0 \implies \theta \boldsymbol{x} \in S.$$

By combining the last 2 points of the above definition, it is easy to see that a set C is a *convex cone* if and only if
$$\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in C, \forall \lambda_1, \lambda_2 \in \mathbb{R}_+, \quad \lambda_1 \boldsymbol{x}_1 + \lambda_2 \boldsymbol{x}_2 \in C.$$

In fact, we can also show that the above definitions can be extended by taking affine (convex, convex conic) combinations of more than just two points. This yields the following

Proposition 1. *A set $S \subseteq \mathbb{R}^n$ is affine (or convex, or a convex cone) iff it is stable by affine (or convex, or conic) combinations. More precisely,*

- *S is affine* $\iff$ $\forall k \in \mathbb{N}, \forall \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in S,\ \forall \boldsymbol{\lambda} \in \mathbb{R}^k$ *with* $\sum_{i=1}^k \lambda_i = 1, \qquad \sum_{i=1}^k \lambda_i \boldsymbol{x}_i \in S.$
- *S is convex* $\iff$ $\forall k \in \mathbb{N}, \forall \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in S,\ \forall \boldsymbol{\lambda} \in \mathbb{R}_+^k$ *with* $\sum_{i=1}^k \lambda_i = 1, \qquad \sum_{i=1}^k \lambda_i \boldsymbol{x}_i \in S.$
- *S is a convex cone* $\iff$ $\forall k \in \mathbb{N}, \forall \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in S,\ \forall \boldsymbol{\lambda} \in \mathbb{R}_+^k, \qquad \sum_{i=1}^k \lambda_i \boldsymbol{x}_i \in S.$

Proof. We only handle the "convex" case, the other two are similiar. The implication $\Longleftarrow$ is just the case $k = 2$. For $\Longrightarrow$, we proceed by induction on $k \in \mathbb{N}$. The case $k = 1$ is clear. Now, let S be a convex set, and assume that the induction hypothesis is true up to order k. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{k+1} \in \mathbb{R}^n$ and let $\boldsymbol{\lambda} \in \mathbb{R}_+^{k+1}$ satisfy $\sum_{i=1}^{k+1} \lambda_i = 1$. We can write

$$\sum_{i=1}^{k+1} \lambda_i \boldsymbol{x}_i = (1 - \lambda_{k+1}) \underbrace{\sum_{i=1}^{k} \frac{\lambda_i \boldsymbol{x}_i}{1 - \lambda_{k+1}}}_{\boldsymbol{x}_0} + \lambda_{k+1} \boldsymbol{x}_{k+1}$$

Now, observe that the coefficients $\mu_i = \frac{\lambda_i}{1-\lambda_{k+1}}$ $(i = 1, \ldots, k)$ sum to 1, so by the induction hypothesis we have $\boldsymbol{x}_0 \in S$. It follows that $\sum_{i=1}^{k+1} \lambda_i \boldsymbol{x}_i = (1 - \lambda_{k+1})\boldsymbol{x}_0 + \lambda_{k+1}\boldsymbol{x}_{k+1} \in S$ by convexity of S. $\square$

Note: The vector $\sum_{i=1}^k \lambda_i \boldsymbol{x}_i$ is called an *affine combination* of the $\boldsymbol{x}_i$'s when the λ_i's sum to one, a *convex combination* of the $\boldsymbol{x}_i$'s when the λ_i's are nonnegative and sum to one, and a *conic combination* when the λ_i's are nonnegative (without restriction on their sum).

Now we can define the notions of affine, convex, and conic hull of a set S, which correspond to the smallest (in the sense of inclusion) affine set (convex set, convex cone) that contain S. In other words, the affine (convex, conic) hull of S is the intersection of all affine (convex, convex conic) supersets of S. This definition can be useful, but is not really practical, since we are talking about taking the intersection of an infinite family of supersets of S. Fortunately, there is an equivalent definition, which gives an explicit construction: the hulls are obtained by taking all possible affine (convex, conic) combinations of points in S.

Definition 3. (Affine, Convex, and Conic hull). Let S be a subset of $\mathbb{R}^n$.

- The *vector space* spanned by S is:

$$\mathbf{span}\, S := \{\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i :\ k \in \mathbb{N}, \quad \forall i \in [k], \boldsymbol{x}_i \in S, \quad \boldsymbol{\lambda} \in \mathbb{R}^k\}.$$

- The *affine hull* of S is:

$$\mathbf{aff}\, S := \{\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i :\ k \in \mathbb{N}, \quad \forall i \in [k], \boldsymbol{x}_i \in S, \quad \boldsymbol{\lambda} \in \mathbb{R}^k, \quad \mathbf{1}^T \boldsymbol{\lambda} = 1\}.$$

- The *convex hull* of S is:

$$\mathbf{conv}\, S := \{\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i :\ k \in \mathbb{N}, \quad \forall i \in [k], \boldsymbol{x}_i \in S, \quad \boldsymbol{\lambda} \in \mathbb{R}_+^k, \quad \mathbf{1}^T \boldsymbol{\lambda} = 1\}.$$

- The *conic hull* of S is:

$$\mathbf{cone}\, S := \{\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i :\ k \in \mathbb{N}, \quad \forall i \in [k], \boldsymbol{x}_i \in S, \quad \boldsymbol{\lambda} \in \mathbb{R}_+^k\}.$$

Proof. There is nothing to show for **span**, the definition is just given here for the sake of completeness, and to emphasize the parallel with the three hull-definitions.

We need to show that the above statements are indeed equivalent to the original definition of the different hulls (as intersection of all affine/convex/convex conic supersets). We'll do this for the convex case only, the other two are similar. Let R denote the set on the right hand side of the definition, and let S' denote an arbitrary convex superset of S. For any element $\boldsymbol{x} = \sum_i \lambda_i \boldsymbol{x}_i$ of R (with $\sum_i \lambda_i = 1$), we have that $\boldsymbol{x}_i \in S'$, so by convexity of S' it holds $\boldsymbol{x} \in S'$. This shows $R \subseteq S'$, and hence $R \subseteq \mathbf{conv}\, S$ because S' was arbitrary. For the converse inclusion, it is sufficient to show that R is a convex superset of S. For all $\boldsymbol{x} \in S$ we can write $\boldsymbol{x}$ as a convex combination of itself (i.e., with $k = 1$), so $S \subseteq R$. It remains to show that R is convex. To do so, let $\boldsymbol{x} = \sum_i \lambda_i \boldsymbol{x}_i \in R$ and $\boldsymbol{y} = \sum_j \mu_j \boldsymbol{y}_j \in R$ (with $\sum_i \lambda_i = \sum_j \mu_j = 1$). For some $\alpha \in [0, 1]$ we will show that convex combination $\boldsymbol{z} = \alpha\boldsymbol{x} + (1 - \alpha)\boldsymbol{y}$ is an element of R. Indeed,

$$\boldsymbol{z} = \sum_i \alpha\lambda_i \boldsymbol{x}_i + \sum_j (1-\alpha)\mu_j \boldsymbol{y}_j, \quad \text{with } \sum_i \alpha\lambda_i + \sum_j (1-\alpha)\mu_j = \alpha + (1-\alpha) = 1,$$

so $\boldsymbol{z}$ is a convex combination of the $\boldsymbol{x}_i$'s and the $\boldsymbol{y}_j$'s, and $\boldsymbol{z} \in R$. This shows that R is convex, and we have shown $\mathbf{conv}\, S \subseteq R$. □

As intuition suggests, an affine set is just a vector space *shifted* by some constant vector. Indeed, let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathbb{R}^n$ and consider the affine set $L = \{\sum_{i=1}^k \lambda_i \boldsymbol{x}_i : \ \boldsymbol{\lambda} \in \mathbb{R}^k, \quad \mathbf{1}^T \boldsymbol{\lambda} = 1\}$. By using the relation $\lambda_1 = 1 - \sum_{i=2}^k \lambda_i$, we can rewrite L as

$$L = \{\boldsymbol{x}_k + \sum_{i=1}^{k-1} \lambda_i (\boldsymbol{x}_i - \boldsymbol{x}_k) : \ \boldsymbol{\lambda} \in \mathbb{R}^{k-1}\} = \boldsymbol{x}_k + \mathbf{span}\, \{\boldsymbol{x}_1 - \boldsymbol{x}_k, \boldsymbol{x}_2 - \boldsymbol{x}_k, \ldots, \boldsymbol{x}_{k-1} - \boldsymbol{x}_k\},$$

where $\mathbf{span}\, S$ denote the vector space spanned by the vectors of S. More generally, the following proposition shows that every affine space can be written as the sum of a subspace and a constant vector (also called *offset*), or equivalently as the set of solutions of a linear equation:

Proposition 2. *Let L be an affine space. Then, we have $L = \boldsymbol{x}_0 + V$, where $\boldsymbol{x}_0$ is* any *element of L, and V is a vector space. The subspace $V = L - \boldsymbol{x}_0$ does not depend on the choice of $\boldsymbol{x}_0 \in L$. Thus, we can define the dimension of L as the dimension of its associated subspace V.*

If L is an affine space of $\mathbb{R}^n$ of dimension $m \in \mathbb{N}$, then there exists a full-column-rank matrix $A \in \mathbb{R}^{n \times m}$ and a vector $\boldsymbol{b} \in \mathbb{R}^n$ such that

$$L = \{A\boldsymbol{y} + \boldsymbol{b} : \ \boldsymbol{y} \in \mathbb{R}^m\}.$$

Alternatively, there exists a full-row-rank matrix $F \in \mathbb{R}^{m \times n}$ and a vector $\boldsymbol{g} \in \mathbb{R}^m$ such that

$$L = \{\boldsymbol{x} \in \mathbb{R}^n : \ F\boldsymbol{x} = \boldsymbol{g}\}.$$

Proof. Let $\boldsymbol{x}_0 \in L$. We first show that $V = L - \boldsymbol{x}_0 := \{\boldsymbol{x} - \boldsymbol{x}_0 : \boldsymbol{x} \in L\}$ is a vector space. To do this, we select $\boldsymbol{x}_1, \boldsymbol{x}_2 \in L$ and $\lambda \in \mathbb{R}$, we let $\boldsymbol{y}_1 = \boldsymbol{x}_1 - \boldsymbol{x}_0 \in V$, $\boldsymbol{y}_2 = \boldsymbol{x}_2 - \boldsymbol{x}_0 \in V$, and we show that $\lambda\boldsymbol{y}_1 + \boldsymbol{y}_2 \in V$:

$$\begin{aligned}
\lambda\boldsymbol{y}_1 + \boldsymbol{y}_2 \in V &\iff \lambda\boldsymbol{y}_1 + \boldsymbol{y}_2 + \boldsymbol{x}_0 \in L \\
&\iff \lambda(\boldsymbol{x}_1 - \boldsymbol{x}_0) + (\boldsymbol{x}_2 - \boldsymbol{x}_0) + \boldsymbol{x}_0 \in L \\
&\iff \lambda\boldsymbol{x}_1 - \lambda\boldsymbol{x}_0 + \boldsymbol{x}_2 \in L
\end{aligned}$$

The last statement is true, because $\lambda - \lambda + 1 = 1$, so $\lambda\boldsymbol{x}_1 - \lambda\boldsymbol{x}_0 + \boldsymbol{x}_2$ is an affine combination of $\boldsymbol{x}_0, \boldsymbol{x}_1$, and $\boldsymbol{x}_2$, and L is stable by affine combinations, cf. Proposition 1.

Now, we show that V does not depend on the choice of the element $\boldsymbol{x}_0 \in L$. Let $V_1 = L - \boldsymbol{x}_1$ and $V_2 = L - \boldsymbol{x}_2$. We take an element $\boldsymbol{x}_0 \in L$ and we define $\boldsymbol{y}_1 = \boldsymbol{x}_0 - \boldsymbol{x}_1 \in V_1$. Now, observe that $\boldsymbol{y}_1 + \boldsymbol{x}_2 = \boldsymbol{x}_0 - \boldsymbol{x}_1 + \boldsymbol{x}_2$ is an affine combination of $\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2 \in L$, so $\boldsymbol{y}_1 + \boldsymbol{x}_2 \in L$, and $\boldsymbol{y}_1 \in V_2$. This shows $V_1 \subseteq V_2$, and the same reasonning can be used to show the converse inclusion: $V_1 = V_2$.

Finally, the last part of the proposition is a consequence from the basis fact that any subspace of dimension m in $\mathbb{R}^n$ can be writen as $V = \mathbf{Im}\, A$ for some matrix $A \in \mathbb{R}^{n \times m}$, or as $V = \mathbf{Ker}\, F$ for some matrix $F \in \mathbb{R}^{m \times n}$. □

In particular, the above proposition tells us that the affine hull L of a set $S \subseteq \mathbb{R}^n$ is the set of all affine combinations of at most $m+1$ points of S, where the number $m := \dim(\mathbf{aff}\, S) \leq n$ is called the *affine dimension* of S. A similar result gives an upper bound for the number of points of S we need to "combine" to get any point in the convex or conic hull of S, and is due to Caratheodory:

Theorem 3 (Caratheodory). *Let $S \subseteq \mathbb{R}^n$ be of affine dimension $m \leq n$, and let $\boldsymbol{x} \in \mathbf{conv}\, S$. Then, $\boldsymbol{x}$ can be expressed as a convex combination of $k \leq m+1 \leq n+1$ points of S.*

Proof. We know that $\boldsymbol{x} = \sum_{i=1}^k \lambda_i \boldsymbol{x}_i$ for some positive λ_i's summing to 1, and some $\boldsymbol{x}_i \in S$. If $k \leq m+1$, there is nothing to prove. So, let $k > m+1$. We will construct another convex combination of the form $\boldsymbol{x} = \sum_{i=1}^{k-1} \lambda_i' \boldsymbol{x}_i'$. This can be iterated until $\boldsymbol{x}$ is expressed as a combination of at most $m+1$ points.

The vectors $\boldsymbol{x}_2 - \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k - \boldsymbol{x}_1$ must be linearly dependent (this is family of $k-1 > m$ vectors in the subspace $V = L - \boldsymbol{x}_1$, which has dimension m), so there exists some $\mu_i (i = 2, \ldots, k)$, not all equal to zero, such that $\sum_{i=2}^k \mu_i (\boldsymbol{x}_i - \boldsymbol{x}_1) = \mathbf{0}$. Letting $\mu_1 = -\sum_{i=2}^k \mu_i$, we have $\sum_{i=1}^k \mu_i = 0$ and

$$\sum_{i=1}^k \mu_i \boldsymbol{x}_i = \mathbf{0}.$$

And for all $\alpha \in \mathbb{R}$, it holds

$$\sum_{i=1}^k (\lambda_i - \alpha\mu_i)\boldsymbol{x}_i = \boldsymbol{x} \quad \text{and} \quad \sum_{i=1}^k (\lambda_i - \alpha\mu_i) = 1.$$

The above defines a new convex combination of the $\boldsymbol{x}_i$'s if $\forall i,\ \lambda_i - \alpha\mu_i \geq 0$. This is the case if

$$\alpha \leq \alpha^* := \min\{\frac{\lambda_i}{\mu_i} : i \in [k], \mu_i > 0\}.$$

Moreover, for the value $\alpha^* = \alpha$, at least one of the new coefficient $\lambda_i' = \lambda_i - \alpha^* \mu_i$ is equal to zero, i.e., we have expressed $\boldsymbol{x}$ as a convex combination of $\leq k-1$ points. □

The conic version of Caratheodory's theorem is as follows. Contrarily to the convex-hull case, this time n points are sufficient. The proof is similar to the proof of the standard version of the theorem and is left as an exercise.

Theorem 4 (Caratheodory – conic version). *Let $S \subseteq \mathbb{R}^n$, such that $\dim(\mathbf{span}\, S) = m \leq n$, and let $\boldsymbol{x} \in \mathbf{cone}\, S$. Then, $\boldsymbol{x}$ can be expressed as a conic combination of $k \leq m \leq n$ points of S.*

2 Simple examples of convex sets and cones

- The set $\{\boldsymbol{x} \in \mathbb{R}^n : \ \boldsymbol{a}^T \boldsymbol{x} = b\}$, where $\boldsymbol{a} \in \mathbb{R}^n, b \in \mathbb{R}$, is called a *hyperplane*. It is affine and hence convex.
- The set $\{\boldsymbol{x} \in \mathbb{R}^n : \ \boldsymbol{a}^T \boldsymbol{x} \leq b\}$, where $\boldsymbol{a} \in \mathbb{R}^n, b \in \mathbb{R}$, is called a *halfspace*. It is convex, but not affine.
- The set $\{\boldsymbol{x} \in \mathbb{R}^n : \ \|\boldsymbol{x} - \boldsymbol{x}_0\| \leq r\}$, where $\boldsymbol{x}_0 \in \mathbb{R}^n, r \in \mathbb{R}_+$ and $\|\cdot\|$ is *any* norm, is called a *ball* and is convex. (We recall that $\|\cdot\|$ is a norm if it is absolutely homogeneous ($\|\alpha\boldsymbol{u}\| = |\alpha|\,\|\boldsymbol{u}\|$), it satisfies the triangle inequality ($\|\boldsymbol{u}+\boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$), and it is nonnegative definite ($\|\boldsymbol{u}\| \geq 0$, with $\|\boldsymbol{u}\| = 0$ iff $\boldsymbol{u} = \mathbf{0}$).)
- The set $\mathbf{conv}\,\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$, where the $\boldsymbol{x}_i$'s are elements of $\mathbb{R}^n$, is called a *polytope*. It is convex.
- The set $\{\boldsymbol{x} : A\boldsymbol{x} \leq \boldsymbol{b}\}$, where $A \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$, is the intersection of m halfspaces, and is called a *polyhedron*. It is convex.

- The set $\{(\boldsymbol{x},t) \in \mathbb{R}^n \times \mathbb{R} : \ \|\boldsymbol{x}-\boldsymbol{x}_0\| \leq t\}$, where $\boldsymbol{x}_0 \in \mathbb{R}^n$ and $\|\cdot\|$ is *any* norm is called a *norm cone*. It is a convex cone. When $\|\cdot\|$ is the Euclidean norm, this set is reffered as the *Lorentz cone*, or the *second-order cone*.
- The set $\Delta_n := \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} \geq \boldsymbol{0},\ \boldsymbol{1}^T\boldsymbol{x} \leq 1\}$ is called the *unit simplex* of $\mathbb{R}^n$. It is is convex. It is possible to show that the unit simplex is also a polytope. Indeed,

$$\Delta_n = \mathbf{conv}\,\{\boldsymbol{0}, \boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}.$$

- The set $\Delta_n^= := \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} \geq \boldsymbol{0},\ \boldsymbol{1}^T\boldsymbol{x} = 1\}$ is called the *probability simplex* of $\mathbb{R}^n$. It is is convex. It is possible to show that the probability simplex is also a polytope. Indeed,

$$\Delta_n^= = \mathbf{conv}\,\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}.$$

  This set is of affine dimension $n-1$, hence some authors use $n-1$ instead of n as a subscript in the notation for this set. In this course, we'll stick to n in order to emphasize that $\Delta_n^=$ is a subset of $\mathbb{R}^n$.
- The set $\mathbb{R}_+^n$ is called *the nonnegative orhant*. It is a convex cone.
- The set $\mathbb{S}^n$ of $n \times n$ symmetric matrices is a subspace of $\mathbb{R}^{n\times n}$, of dimension $\frac{1}{2}n(n+1)$.

3 Operations that preserve convexity

We list hereafter some operations on sets that preserve convexity. The proofs of these statements are left as exercises.

(a) **Intersection**: The intersection of convex sets is convex (this is also valid for the intersection of an infinite family of convex sets). For example, the set of positive semidefinite matrices is the intersection of (infinitely many) halfspaces, hence it is convex:

$$\{X \in \mathbb{S}^n : \forall \boldsymbol{x} \in \mathbb{R}^n,\ \boldsymbol{x}^T X \boldsymbol{x} \geq 0\} = \bigcap_{\boldsymbol{x} \in \mathbb{R}^n} \{X \in \mathbb{S}^n :\ \langle X, \boldsymbol{x}\boldsymbol{x}^T\rangle \geq 0\}.$$

(b) **Cartesian product**: if $S \subseteq \mathbb{R}^n$ is convex and $T \subseteq \mathbb{R}^m$ is convex, then

$$S \times T = \{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{n+m} :\ \boldsymbol{x} \in S,\ \boldsymbol{y} \in T\}$$

is convex.

(c) **Affine transformation**: if $S \subseteq \mathbb{R}^n$ is convex, $A \in \mathbb{R}^{m\times n}, \boldsymbol{b} \in \mathbb{R}^m$, then $\{A\boldsymbol{x} + \boldsymbol{b} : \boldsymbol{x} \in S\}$ is convex. This includes, as special cases, the following transformation:

  - *scaling* (for $A = \rho I$, $\boldsymbol{b} = \boldsymbol{0}$).
  - *translation* (for $A = O$).
  - *projection over a subset of coordinates* (for A of the form $A = [I, O]$ and $\boldsymbol{b} = \boldsymbol{0}$)
  - *Minkowski sum*: if $S \subseteq \mathbb{R}^n$ is convex and $T \subseteq \mathbb{R}^n$ is convex, then

$$S + T = \{\boldsymbol{x} + \boldsymbol{y} :\ \boldsymbol{x} \in S,\ \boldsymbol{y} \in T\}$$

  is convex. To see this, apply a linear transformation with $A = [I, I]$ over $S \times T$.

(d) **Reverse affine transformation**: if S is convex, then $\{\boldsymbol{x} : A\boldsymbol{x}+\boldsymbol{b} \in S\}$ is convex. For example, the ellipsoid $\{\boldsymbol{x} \in \mathbb{R}^n : \|A\boldsymbol{x}+\boldsymbol{b}\| \leq 1\}$ is the reverse image of the unit ball by some affine transformation, hence it is convex.

(e) **Closure and interior**: If $S \subset \mathbb{R}^n$ is convex, then its closure $\mathbf{cl}\, S$ and its interior $\mathbf{int}\, S$ are convex.

(f) **Perspective transformation**: The perspective function is $P : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}^n$, defined by $P(\boldsymbol{x}, t) = \frac{\boldsymbol{x}}{t}$.

- If $C \subseteq \mathbb{R}^n \times \mathbb{R}_{++}$ is convex, then $P(C)$ is convex.
- If $S \subseteq \mathbb{R}^n$ is convex, then $P^{-1}(S) := \{(\boldsymbol{x}, t) \in \mathbb{R}^n \times \mathbb{R}_{++} : \frac{1}{t}\boldsymbol{x} \in S\}$ is convex.

For example, the Lorrentz cone $\{(\boldsymbol{x}, t) : \|\boldsymbol{x}\| \leq t\}$ is the closure of $P^{-1}(B)$, where B is the unit ball.

In the next two subsections, we study two classes of convex sets which will play an important role in this lecture: ellipsoids, and the cone of positive semidefinite matrices.

4 Positive semidefinite matrices

We now introduce a very important example, the set of positive semidefinite matrices. We recall that a matrix $X \in \mathbb{S}^n$ is called positive semidefinite if

$$\forall \boldsymbol{u} \in \mathbb{R}^n, \boldsymbol{u}^T X \boldsymbol{u} \geq 0.$$

The next proposition gives important equivalent definitions

Proposition 5. *Let $X \in \mathbb{S}^n$. The following statements are equivalent:*

(i) $X \in \mathbb{S}^n_+$ *(S is positive semidefinite)*

(ii) $\forall \boldsymbol{u} \in \mathbb{R}^n, \boldsymbol{u}^T X \boldsymbol{u} \geq 0.$

(iii) All eigenvalues of X are nonnegative.

(iv) $\exists H \in \mathbb{R}^{n \times m}, m \in \mathbb{N} : \; X = HH^T$

(v) $X \in \mathbf{conv}\{\boldsymbol{x}\boldsymbol{x}^T : \boldsymbol{x} \in \mathbb{R}^n\} = \mathbf{cone}\{\boldsymbol{x}\boldsymbol{x}^T : \boldsymbol{x} \in \mathbb{R}^n\}$.

Proof. (i) $\iff$ (ii): This is the definition.

(ii) $\implies$ (iii): Let λ be an eigenvalue of X, and $\boldsymbol{v}$ an associated (normalized) eigenvector. We recall that $\lambda \in \mathbb{R}$, because X is symmetric. Now, we have $\boldsymbol{v}^T X \boldsymbol{v} = \lambda \boldsymbol{v}^T \boldsymbol{v} = \lambda \|\boldsymbol{v}\|^2 = \lambda \geq 0$.

(iii) $\implies$ (iv): X is symmetric, so it admits an eigendecomposition of the form $X = U\Lambda U^T$, where the matrix U is orthogonal (i.e., $UU^T = U^T U = I$), and $\Lambda = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix containing the eigenvalues of X. Setting $H = U\Lambda^{1/2}$, where $\Lambda^{1/2} := \mathrm{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$, we have $X = HH^T$.

(iv) $\implies$ (v): Let $X = HH^T$, and let $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_m \in \mathbb{R}^n$ be the columns of H. Then, we have $X = HH^T = \sum_{i=1}^m \boldsymbol{h}_i \boldsymbol{h}_i^T$, which is a conic combination of the rank-1 matrices $\boldsymbol{h}_i \boldsymbol{h}_i^T$, or $X = \sum_i \frac{1}{m} \frac{\boldsymbol{h}_i}{\sqrt{m}} \frac{\boldsymbol{h}_i}{\sqrt{m}}^T$, a convex combination of matrices of the form $\boldsymbol{x}\boldsymbol{x}^T$, where $\boldsymbol{x} \in \mathbb{R}^n$.

(v) $\implies$ (ii): Let $X = \sum_i \lambda_i \boldsymbol{x}_i \boldsymbol{x}_i^T$, with $\boldsymbol{\lambda} \in \mathbb{R}^m_+$ and $\boldsymbol{x}_i \in \mathbb{R}^n$. Then, for all $\boldsymbol{u} \in \mathbb{R}^n$ we have

$$\boldsymbol{u}^T X \boldsymbol{u} = \sum_i \lambda_i \boldsymbol{u}_i^T \boldsymbol{x}_i \boldsymbol{x}_i^T \boldsymbol{u}_i = \sum_i \underbrace{\lambda_i}_{\geq 0} (\boldsymbol{u}_i^T \boldsymbol{x}_i)^2 \geq 0.$$

□

The above proposition implies, in particular, the following

Corollary 6. *The set $\mathbb{S}_+^n$ is a convex cone.*

The next result characterizes the interior of $\mathbb{S}_+^n$, which is itself a convex cone, and consists of all positive definite matrices.

Proposition 7. *The following statements are equivalent:*

(i) $X \in \mathbb{S}_{++}^n$ *(S is positive definite)*

(ii) $X \in \mathbf{int}\,\mathbb{S}_+^n$

(iii) $\forall \boldsymbol{u} \in \mathbb{R}^n, \quad \boldsymbol{u} \neq \boldsymbol{0} \implies \boldsymbol{u}^T X \boldsymbol{u} > 0.$

(iv) *All eigenvalues of X are positive.*

(v) *Sylvester criterion: All leading principal minors of X are positive.*

We now give a few important properties about positive semidefinite matrices. The next lemma can be useful, in particular when we want to show that a matrix is not positive semidefinite.

Lemma 8. *Let $X \in \mathbb{S}_+^n$. Then,*

1. *For any matrix $A \in \mathbb{R}^{m\times n}$, the matrix AXA^T is positive semidefinite.*
2. *If I is a subset of $[n]$, the principal submatrix $X[I,I] = \{X_{i_1,i_2}\}_{i_1\in I, i_2 \in I}$ is positive semidefinite.*
3. *For all $i,j \in [n]$, $|X_{ij}| \leq \sqrt{X_{ii}X_{jj}}$.*
4. $X_{ii} = 0 \implies \forall j \in [n], X_{ij} = 0.$

Proof.

1. X admits a decomposition of the form HH^T. Hence, $AXA^T = (AH)(AH)^T \in \mathbb{S}_+^m$.
2. We construct a matrix A such that $X[I,I] = AXA^T$ and we use the previous result. Let $I = \{i_1, \ldots, i_{|I|}\}$, and define the matrix $A \in \mathbb{R}^{|I|\times n}$ as follows: For $k \in \{1,\ldots,|I|\}$, $A_{k,i_k} = 1$ and $A_{i,j} = 0$ elsewhere.
3. In particular, the 2×2 principal submatrix $\begin{bmatrix} X_{ii} & X_{ij} \\ X_{ij} & X_{jj} \end{bmatrix}$ is positive semidefinite, so its determinant is nonnegative: $X_{ii}X_{jj} - X_{ij}^2 \geq 0$, which implies $|X_{ij}| \leq \sqrt{X_{ii}X_{jj}}$.
4. This is an immediate consequence of the previous inequality.

□

Proposition 9 (Matrix square root)**.** *Let $X \in \mathbb{S}_+^n$. Then, X has a square root, which we denote by $X^{\frac{1}{2}} \in \mathbb{S}_+^n$, and is the only positive semidefinite matrix that satisfies*

$$X = \left(X^{\frac{1}{2}}\right)^2.$$

In particular, the eigenvalues of $X^{\frac{1}{2}}$ are the square roots of the eigenvalues of X.

Proof. Let $X = U\Lambda U^T$ be an eigendecomposition of X, where U is orthogonal and Λ is a diagonal matrix with nonnegative elements. Then, we define $X^{\frac{1}{2}} = U\Lambda^{\frac{1}{2}}U^T \in \mathbb{S}_+^n$, with $\Lambda^{1/2} := \mathrm{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$. Now we can verify that

$$\left(X^{\frac{1}{2}}\right)^2 = U\Lambda^{\frac{1}{2}} \underbrace{U^T U}_{=I} \Lambda^{\frac{1}{2}} U^T = U\Lambda U^T = X.$$

It remains to show that the positive semidefinite square root is unique. A matrix A has the same eigenspaces as its square A^2, and the eigenvalues of A^2 are the square of the eigenvalues of A. Hence, if S_1 and S_2 are positive semidefinite square roots of X, then we have eigendecompositions of the form

$$S_1 = U\Lambda^{\frac{1}{2}}U^T, \qquad S_2 = V\Lambda^{\frac{1}{2}}V^T,$$

and if U_λ, V_λ are the columns of U and V corresponding to some eigenvalue λ of X, then U_λ is an orthogonal transformation of V_λ. This shows that $U = V\Omega$, where Ω is a block diagonal matrix with orthogonal matrices $\Omega_\lambda = V_\lambda^T U_\lambda$ as diagonal blocks. So we have $S_1 = U\Lambda^{\frac{1}{2}}U^T = V\Omega\Lambda^{\frac{1}{2}}\Omega^T V^T$, and the diagonal blocks of $\Omega\Lambda^{\frac{1}{2}}\Omega^T$ are of the form $\Omega_\lambda(\sqrt{\lambda}I)\Omega_\lambda^T = \sqrt{\lambda}\Omega_\lambda\Omega_\lambda^T = \lambda^{1/2}I$. Hence $\Omega\Lambda^{\frac{1}{2}}\Omega^T = \Lambda^{\frac{1}{2}}$, and $S_1 = V\Omega\Lambda^{\frac{1}{2}}\Omega^T V^T = V\Lambda^{\frac{1}{2}}V^T = S_2$. □

We finish this section with another decomposition of positive semidefinite matrices, which we will not prove:

Proposition 10 (Cholesky decomposition)**.** *All positive semidefinite matrices $X \in \mathbb{S}_+^n$ admit a* Cholesky decomposition *of the form $X = LL^T$, where L is a lower triangular matrix. If X is positive definite, then this decomposition is unique.*

5 Ellipsoids

An ellipsoid of $\mathbb{R}^n$ is a set of the form

$$\{\boldsymbol{x} \in \mathbb{R}^n : (\boldsymbol{x} - \boldsymbol{x}_0)^T Q^{-1}(\boldsymbol{x} - \boldsymbol{x}_0) \leq 1\},$$

where $\boldsymbol{x}_0 \in \mathbb{R}^n$ and the matrix Q is positive definite.

Proposition 11 (Affine transformation of a ball)**.** *Let $A \in \mathbb{R}^{n\times n}$ be an invertible matrix, $\boldsymbol{b} \in \mathbb{R}^n$ and $r > 0$. The following sets are ellipsoids (E_1 is the image of a ball by the affine mapping $\boldsymbol{x} \mapsto A\boldsymbol{x} - \boldsymbol{b}$, and E_2 is the reverse image of a ball by the same mapping):*

- $E_1 = \{A\boldsymbol{z} - \boldsymbol{b} : \ \boldsymbol{z} \in \mathbb{R}^n, \|\boldsymbol{z}\| \leq r\} \subset \mathbb{R}^n$
- $E_2 = \{\boldsymbol{x} \in \mathbb{R}^n : \ \|A\boldsymbol{x} - \boldsymbol{b}\| \leq r\} \subset \mathbb{R}^m$

Proof.

- We first show that E_2 is an ellipsoid. Define $Q = r^2(A^TA)^{-1}$ and $\boldsymbol{x}_0 = A^{-1}\boldsymbol{b}$, so we have

$$\begin{aligned}
\frac{1}{r^2}(\boldsymbol{x} - \boldsymbol{x}_0)^T Q^{-1}(\boldsymbol{x} - \boldsymbol{x}_0) &= \frac{1}{r^2}(\boldsymbol{x}^T Q^{-1}\boldsymbol{x} - 2\boldsymbol{x}_0^T Q^{-1}\boldsymbol{x} + \boldsymbol{x}_0^T Q^{-1}\boldsymbol{x}_0) \\
&= \frac{1}{r^2}(\boldsymbol{x}^T A^T A\boldsymbol{x} - 2\boldsymbol{b}^T A^{-T}A^T A\boldsymbol{x} + \boldsymbol{b}^T A^{-T}A^T AA^{-1}\boldsymbol{b}) \\
&= \frac{1}{r^2}(\boldsymbol{x}^T A^T A\boldsymbol{x} - 2\boldsymbol{b}^T A\boldsymbol{x} + \|\boldsymbol{b}\|^2) \\
&= \frac{1}{r^2}\|A\boldsymbol{x} - \boldsymbol{b}\|^2.
\end{aligned}$$

  This shows that E_2 can be rewritten as $\{\boldsymbol{x} \in \mathbb{R}^n : (\boldsymbol{x} - \boldsymbol{x}_0)^T Q^{-1}(\boldsymbol{x} - \boldsymbol{x}_0) \leq 1\}$, which is an ellipsoid, because $Q = r^2(A^TA)^{-1}$ is positive definite.

- $\boldsymbol{x} \in E_1 \iff \exists \boldsymbol{z} \in \mathbb{R}^n : \|\boldsymbol{z}\| \leq r, \ \boldsymbol{x} = A\boldsymbol{z} - \boldsymbol{b} \iff \|A^{-1}(\boldsymbol{x} + \boldsymbol{b})\| \leq r$. Define $A' = A^{-1}$ and $\boldsymbol{b}' = -A^{-1}\boldsymbol{b}$, so $E_1 = \{\boldsymbol{x} \in \mathbb{R}^n : \ \|A'\boldsymbol{x} - \boldsymbol{b}'\| \leq r\}$, which has the same form as E_2 and is hence an ellipsoid.

□

When the matrix A in the definition of E_1 is not invertible (in particular, if A is non-square), we obtain a "flat ellipsoid", which is contained in the affine space $-\boldsymbol{b}+\mathbf{Im}\, A$. When the matrix A in the definition of E_2 is not invertible, we get a "cylinder", which is infinite in the directions corresponding to singular eigenvectors of A. The next example shows that the directions and lengths of the the semi-axis of an ellipsoid are actually given by the eigenvalue decomposition of its associated matrix Q:

Example:
Let $Q = U\Lambda U^T$ be an eigendecomposition of $Q \in \mathbb{S}^n_{++}$. Denote by $\lambda_1, \ldots, \lambda_n > 0$ the diagonal elements of Λ (the eigenvalues), and by $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ the columns of U (the eigenvectors). The ellipsoid

$$E = \{\boldsymbol{x} \in \mathbb{R}^n : \ (\boldsymbol{x} - \boldsymbol{x}_0)^T Q^{-1} (\boldsymbol{x} - \boldsymbol{x}_0) \leq 1\}$$

can be rewriten as an affine transformation of the unit ball: #1

$$E = \{\boldsymbol{x} : \ \|\Lambda^{-1/2} U^T (\boldsymbol{x} - \boldsymbol{x}_0)^T\| \leq 1\} = \{U\Lambda^{1/2}\boldsymbol{z} + \boldsymbol{x}_0 : \ \boldsymbol{z} \in \mathbb{R}^n, \|\boldsymbol{z}\| \leq 1\}.$$

This can be interpreted by saying that E is an ellipoid centered at $\boldsymbol{x}_0$, with semi-axis of length $\sqrt{\lambda_i}$ in the direction $\boldsymbol{u}_i$. The limit case $\lambda_i \to 0$ gives a "flat ellipsoid", of width 0 in the direction of the singular eigenvector $\boldsymbol{u}_i$.

6 Generalized inequalities and dual cone

Definition 4 (Proper cone)**.** A cone $K \subset \mathbb{R}^n$ is said to be *proper* if it is

- closed;
- convex;
- pointed, i.e., it contains no lines. More precisely,

$$(\boldsymbol{x} \in K, -\boldsymbol{x} \in K) \implies \boldsymbol{x} = \boldsymbol{0};$$

- and it has a nonempty interior.

We can define a partial order *relative to a proper cone* K: we write $\boldsymbol{x} \preceq_K \boldsymbol{y}$ whenever $\boldsymbol{y}$ lies in the affine cone $\boldsymbol{x} + K$. More precisely,

Definition 5 (Generalized order)**.** Let K be a proper cone. Then,

$$\boldsymbol{x} \preceq_K \boldsymbol{y} \iff \boldsymbol{y} - \boldsymbol{x} \in K.$$

We also define the generalized strict inequality

$$\boldsymbol{x} \prec_K \boldsymbol{y} \iff \boldsymbol{y} - \boldsymbol{x} \in \mathbf{int}\, K.$$

Observe that the notion of order relative to a cone generalizes the elementwise inequalities $\boldsymbol{x} \leq \boldsymbol{y}$, which correspond to the case $K = \mathbb{R}^n_+$.

When K is the cone of positive semidefinite matrices, $A \preceq_{\mathbb{S}^n_+} B$ means that $B - A$ is positive semidefinite. So, when we work with matrices, we will often omit the subscript and write $X \succeq 0$ for $X \in \mathbb{S}^n_+$, and $X \succ 0$ for $X \in \mathbb{S}^n_{++}$.

Example:
Let $K \subset \mathbb{R}^{d+1}$ be the cone of coefficients of polynomials of degree d that are nonnegative on $[0,1]$:

$$K = \{\boldsymbol{\alpha} \in \mathbb{R}^{d+1} : \quad \forall x \in [0,1],\ \sum_{i=0}^{d} \alpha_i x^i \geq 0\}.$$

It is easy to check that K is closed, convex, pointed, and has a no empty interior, hence it is proper. The inequality $\boldsymbol{\alpha} \preceq_K \boldsymbol{\beta}$ means that the polynomial with coefficients α_i is dominated by the polynomial with coeffients β_i on the whole interval $[0,1]$: #2

$$\boldsymbol{\alpha} \preceq_K \boldsymbol{\beta} \iff \forall x \in [0,1], \quad \sum_{i=0}^{d} \alpha_i x^i \leq \sum_{i=0}^{d} \beta_i x^i$$

Let us now review a few properties of the generalized order $\preceq_K$. The proof is left as an exercise.

Proposition 12. *Let K be a proper cone. The generalized inequality $\preceq_K$ satisfies following properties:*

1. *transitivity:* $\boldsymbol{x} \preceq_K \boldsymbol{y}$ *and* $\boldsymbol{y} \preceq_K \boldsymbol{z} \implies \boldsymbol{x} \preceq_K \boldsymbol{z}$
2. *reflexivity:* $\boldsymbol{x} \preceq_K \boldsymbol{x}$.
3. *antisymmetry:* $\boldsymbol{x} \preceq_K \boldsymbol{y}$ *and* $\boldsymbol{y} \preceq_K \boldsymbol{x} \implies \boldsymbol{x} = \boldsymbol{y}$.
4. *preservation under addition:* $\boldsymbol{x} \preceq_K \boldsymbol{y}$ *and* $\boldsymbol{u} \preceq_K \boldsymbol{v} \implies \boldsymbol{x} + \boldsymbol{u} \preceq_K \boldsymbol{y} + \boldsymbol{v}$.
5. *preservation under nonnegative scaling:* $\boldsymbol{x} \preceq_K \boldsymbol{y}$ *and* $\alpha \geq 0 \implies \alpha\boldsymbol{x} \preceq_K \alpha\boldsymbol{y}$.

In particular, the properties 1-3 of the proposition above show that $\preceq_K$ is a *partial order*. However, it is not a total order, because some elements can't be compared. This means that in general,

$$\boldsymbol{x} \not\preceq_K \boldsymbol{y} \not\Longleftrightarrow \boldsymbol{x} \succeq_K \boldsymbol{y}.$$

For example, the vectors $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ are not comparable for the elementwise order $\preceq_{\mathbb{R}^2_+}$.

We now define a notion that will be very important later in this course:

Definition 6 (Dual cone)**.** Let K be a cone. The dual cone of $K \subset \mathbb{R}^n$ is

$$K^* := \{\boldsymbol{y} \in \mathbb{R}^n : \ \langle \boldsymbol{x}, \boldsymbol{y} \rangle \geq 0,\ \forall \boldsymbol{x} \in K\}.$$

We next present a fundamental property of the dual cone, which we will use later in this course to define the Lagrangian dual of conic programming problems.

Proposition 13. *Let K be a cone. Then,*

$$\inf_{\boldsymbol{x} \in K} \ \boldsymbol{c}^T \boldsymbol{x} = \begin{cases} 0 & \text{if } \boldsymbol{c} \in K^* \\ -\infty & \text{otherwise.} \end{cases}$$

Similarly,

$$\sup_{\boldsymbol{x} \in K} \ \boldsymbol{c}^T \boldsymbol{x} = \begin{cases} 0 & \text{if } \boldsymbol{c} \in -K^* \\ +\infty & \text{otherwise.} \end{cases}$$

Proof. We simply use the definition of K^*:

$$\boldsymbol{c} \succeq_{K^*} \boldsymbol{0} \iff \forall \boldsymbol{x} \in K, \boldsymbol{c}^T\boldsymbol{x} \geq 0.$$

Hence, if $\boldsymbol{c} \succeq_{K^*} \boldsymbol{0}$, it holds $\inf_{\boldsymbol{x}\in K}\ \boldsymbol{c}^T\boldsymbol{x} \geq 0$. On the other hand, since $\boldsymbol{0} \in K$, we have $\inf_{\boldsymbol{x}\in K}\ \boldsymbol{c}^T\boldsymbol{x} \leq \boldsymbol{c}^T\boldsymbol{0} = 0$. Conversely, if $\boldsymbol{c} \not\succeq_{K^*} \boldsymbol{0}$, then $\exists \boldsymbol{x}_0 \in K : \boldsymbol{c}^T\boldsymbol{x}_0 < 0$. Setting $\boldsymbol{x} = t\boldsymbol{x}_0 \in K$ and letting $t \to \infty$ yields $\boldsymbol{c}^T\boldsymbol{x} \to -\infty$.

The proof of the statement with the sup is obtained by changing $\boldsymbol{c}$ to $-\boldsymbol{c}$. □

The next proposition gives some important properties of dual cones. We will prove them in exercise.

Proposition 14 (Properties of dual cones)**.** *Let K be a convex cone.*

1. *K^* is a convex cone.*
2. *K^* is closed (even if K is not).*
3. *$K_1 \subseteq K_2 \implies K_2^* \subseteq K_1^*$.*
4. *K has a nonempty interior $\implies$ K^* pointed.*
5. *$K^{**} = \mathbf{cl}\, K$ (so, in particular, K closed $\implies K = K^{**}$).*
6. *$\mathbf{cl}\, K$ is pointed $\implies$ K^* has a nonempty interior.*

Observe that the above proposition shows that if K is a proper cone, then K^* is a proper cone, too. And in this case, we always have $K = (K^*)^*$, which justifies the term "dual cone".

We next define the notions of *minimal* and *minimum* element in a set relative to a proper cone K

Definition 7 (Minimum and minimal elements)**.** Let K be a proper cone. We say that $\boldsymbol{x}$ is the *minimum* element of S (with respect to the generalized order $\preceq_K$) if $S \subseteq \boldsymbol{x} + K$. In other words,

$$\boldsymbol{x} \text{ is the minimum of } S \iff \forall \boldsymbol{y} \in S, \boldsymbol{x} \preceq_K \boldsymbol{y}.$$

If S has a minimum element, then this minimum is unique.

We say that $\boldsymbol{x}$ is a *minimal* element of S (with respec to $\preceq_K$) if $(\boldsymbol{x} - K) \cap S = \{\boldsymbol{x}\}$, that is:

$$\boldsymbol{x} \text{ is minimal in } S \iff \Big(\boldsymbol{y} \in S,\ \boldsymbol{y} \preceq_K \boldsymbol{x} \implies \boldsymbol{y} = \boldsymbol{x}\Big).$$

A minimum element of S is always minimal, but minimal elements need not be unique.

We should prove that the minimum element of S is necessarily unique. Assume that for all $\boldsymbol{y} \in S$, $\boldsymbol{y} \succeq_K \boldsymbol{x}_1$ and $\boldsymbol{y} \succeq_K \boldsymbol{x}_2$, with $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$. Then, in particular it holds $\boldsymbol{x}_2 \succeq_K \boldsymbol{x}_1$ and $\boldsymbol{x}_1 \succeq_K \boldsymbol{x}_2$, so by antisymmetry of $\preceq_K$ it holds $\boldsymbol{x}_1 = \boldsymbol{x}_2$. The fact that minimum elements are minimal can be seen in a similar manner. If $\boldsymbol{x} \preceq_K \boldsymbol{y}$ for all $\boldsymbol{y} \in S$, then any element $\boldsymbol{z} \in S$ satisfying $\boldsymbol{z} \preceq_K \boldsymbol{x}$ also satisfies $\boldsymbol{z} \succeq_K \boldsymbol{x}$, hence $\boldsymbol{z} = \boldsymbol{x}$ by antisymmetry of $\preceq_K$.

The generalized order $\preceq_K$ is related to the order $\preceq_{K^*}$ by the following proposition, which also gives an equivalent characterization of minimum and minimal elements.

Proposition 15. *Let K be a proper cone. Then,*

$$\boldsymbol{x} \preceq_K \boldsymbol{y} \iff \forall \boldsymbol{\lambda} \succeq_{K^*} \boldsymbol{0}, \quad \boldsymbol{\lambda}^T\boldsymbol{x} \leq \boldsymbol{\lambda}^T\boldsymbol{y};$$

$$\boldsymbol{x} \prec_K \boldsymbol{y} \iff \forall \boldsymbol{\lambda} \in K^* \setminus \{\mathbf{0}\}, \quad \boldsymbol{\lambda}^T \boldsymbol{x} < \boldsymbol{\lambda}^T \boldsymbol{y}.$$

In particular, $\boldsymbol{x}$ is the minimum element of S if it minimizes the linear form $\boldsymbol{x} \mapsto \langle \boldsymbol{x}, \boldsymbol{\lambda} \rangle$ over S, for all $\boldsymbol{\lambda} \in K^*$:

$$\boldsymbol{x} \text{ is the minimum of } S \iff \forall \boldsymbol{\lambda} \succeq_{K^*} \mathbf{0}, \forall \boldsymbol{y} \in S,\ \boldsymbol{\lambda}^T \boldsymbol{x} \leq \boldsymbol{\lambda}^T \boldsymbol{y}.$$

Geometrically, this means that $\{\boldsymbol{y} \in \mathbb{R}^n : \boldsymbol{\lambda}^T(\boldsymbol{y} - \boldsymbol{x}) = 0\}$ is a supporting hyperplane of S at $\boldsymbol{x}$, for all $\boldsymbol{\lambda} \in K^*$. For minimal elements however, we only obtain a sufficient condition in general:

$$\boldsymbol{x} \text{ is minimal in } S \impliedby \exists \boldsymbol{\lambda} \succ_{K^*} \mathbf{0}, \forall \boldsymbol{y} \in S,\ \boldsymbol{\lambda}^T \boldsymbol{x} \leq \boldsymbol{\lambda}^T \boldsymbol{y}.$$

But a weaker form of the converse implication holds when S is convex:

$$\boldsymbol{x} \text{ is minimal in the convex set } S \implies \exists \boldsymbol{\lambda} \succeq_{K^*} \mathbf{0}, \forall \boldsymbol{y} \in S,\ \boldsymbol{\lambda}^T \boldsymbol{x} \leq \boldsymbol{\lambda}^T \boldsymbol{y}.$$

Proof. For the first equivalence, we use the fact that K is equal to its bidual K^{**}:

$$\boldsymbol{x} \preceq_K \boldsymbol{y} \iff \boldsymbol{y} - \boldsymbol{x} \in K \iff \boldsymbol{y} - \boldsymbol{x} \in K^{**} \iff \forall \boldsymbol{\lambda} \in K^*, \langle \boldsymbol{y} - \boldsymbol{x}, \boldsymbol{\lambda} \rangle \geq 0.$$

Then, the characterization of the minimum element is obtained by plugging the above equivalence in the definition of the minimum of S. For minimal elements, assume that $\boldsymbol{x}$ minimizes $\boldsymbol{x}^T\boldsymbol{\lambda}$ over S, where $\lambda \succeq_{K^*} \mathbf{0}$. Then, if some element $\boldsymbol{y} \in S$ satisfies $\boldsymbol{y} \preceq_K \boldsymbol{x}$, we must have $\boldsymbol{y}^T\boldsymbol{\lambda} \leq \boldsymbol{x}^T\boldsymbol{\lambda}$, so $\boldsymbol{y}^T\boldsymbol{\lambda} = \boldsymbol{x}^T\boldsymbol{\lambda}$. But then, $(\boldsymbol{x} - \boldsymbol{y}) \in K \setminus \{\mathbf{0}\}$ would imply $(\boldsymbol{x} - \boldsymbol{y})^T\boldsymbol{\lambda} > 0$, a contradiction. Hence, $\boldsymbol{x} = \boldsymbol{y}$, and $\boldsymbol{x}$ is minimal in S. We leave the proof of the last statement for the case of a convex set S as an exercise, as it requires the separating hyperplane theorem from the next section. □

7 Separating hyperplane theorems

This section proves a major result of convex geometry, which will be important for the notion of duality in convex optimization. It basically states that if two convex sets do not intersect, then they can be separated by some hyperplane:

Theorem 16 (Separating hyperplane)**.** *Let X, Y be two disjoint, nonempty convex sets of $\mathbb{R}^n$. Then, there exist a scalar c and a vector $\boldsymbol{v} \in \mathbb{R}^n$, $\boldsymbol{v} \neq \mathbf{0}$, such that*

$$\forall \boldsymbol{x} \in X, \quad \langle \boldsymbol{x}, \boldsymbol{v} \rangle \leq c \quad \text{and} \quad \forall \boldsymbol{y} \in Y, \quad \langle \boldsymbol{y}, \boldsymbol{v} \rangle \geq c.$$

In other words, the hyperplane $\{\boldsymbol{x} : \langle \boldsymbol{x}, \boldsymbol{v} \rangle = c\}$ separates *X and Y.*

Proof. We first prove the following intermediate result: if S is a closed convex set, then it has a unique vector of minimal norm. To see this, let $\delta := \inf_{\boldsymbol{x} \in S} \|\boldsymbol{x}\|$. This infimum is clearly reached, since we minimize a continuous function (the norm) over the compact set $S \cap \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\| \leq \delta + 1\}$. Now, assume that $\|\boldsymbol{x}\| = \|\boldsymbol{y}\| = \delta$. By convexity of S, we have $\frac{1}{2}(\boldsymbol{x} + \boldsymbol{y}) \in S$, so $\|\frac{\boldsymbol{x}+\boldsymbol{y}}{2}\|^2 \geq \delta^2 \iff \|\boldsymbol{x} + \boldsymbol{y}\|^2 \geq 4\delta^2$. Now, we use the parallelogram law: $\|\boldsymbol{x} - \boldsymbol{y}\|^2 = 2\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{y}\|^2 - \|\boldsymbol{x} + \boldsymbol{y}\|^2 \leq 2\delta^2 + 2\delta^2 - 4\delta^2 = 0$. This shows $\boldsymbol{x} = \boldsymbol{y}$.

The set $S = \{\boldsymbol{y} - \boldsymbol{x} :\ \boldsymbol{x} \in X, \boldsymbol{y} \in Y\}$ is convex (it is the Minkowski sum of Y and $-X$, which are both convex), and it does not contain $\mathbf{0}$ (because $X \cap Y = \emptyset$). So, now, we use the above intermediate result, and we denote by $\boldsymbol{v}$ the (unique) vector of minimal norm in $\mathbf{cl}\, S$.

Let $\boldsymbol{z} \in S$. We now show that $\langle \boldsymbol{z}, \boldsymbol{v} \rangle \geq \|\boldsymbol{v}\|^2$. To see this, define $f(t) = \|\boldsymbol{v} + t(\boldsymbol{z} - \boldsymbol{v})\|^2$, and observe that $f(t) \geq f(0) = \|\boldsymbol{v}\|^2$ for all $t \in [0, 1]$, because by convexity, the point $\boldsymbol{v} + t(\boldsymbol{z} - \boldsymbol{v})$ belongs to S. So $\frac{1}{t}(f(t) - f(0)) \geq 0$ for all $t \in (0, 1]$. We have:

$$\frac{1}{t}(f(t) - f(0)) = \frac{1}{t}(\|v\|^2 + 2t\boldsymbol{v}^T(\boldsymbol{z} - \boldsymbol{v}) + t^2\|\boldsymbol{z} - \boldsymbol{v}\|^2 - \|\boldsymbol{v}\|^2) = 2(\boldsymbol{v}^T\boldsymbol{z} - \|\boldsymbol{v}\|^2) + O(t) \geq 0.$$

So, letting $t \to 0$, we obtain $\langle \boldsymbol{z}, \boldsymbol{v} \rangle \geq \|\boldsymbol{v}\|^2$.

If $\boldsymbol{v} \neq \boldsymbol{0}$, this readily proves the theorem. Indeed, we have

$$\begin{aligned}\forall \boldsymbol{z} \in S, \langle \boldsymbol{z}, \boldsymbol{v} \rangle \geq \|\boldsymbol{v}\|^2 &\implies \forall \boldsymbol{x} \in X, \forall \boldsymbol{y} \in Y, \quad \langle \boldsymbol{y}, \boldsymbol{v} \rangle - \langle \boldsymbol{x}, \boldsymbol{v} \rangle \geq \|\boldsymbol{v}\|^2 > 0\\ &\implies \forall \boldsymbol{x} \in X, \forall \boldsymbol{y} \in Y, \quad \langle \boldsymbol{y}, \boldsymbol{v} \rangle > \langle \boldsymbol{x}, \boldsymbol{v} \rangle.\end{aligned}$$

Hence, we can set the scalar c to any number between $\sup_{\boldsymbol{x} \in X} \langle \boldsymbol{x}, \boldsymbol{v} \rangle$ and $\inf_{\boldsymbol{y} \in Y} \langle \boldsymbol{y}, \boldsymbol{v} \rangle$ to get the desired inequality.

To handle the case where $\boldsymbol{v} = \boldsymbol{0}$, let S_n be an increasing sequence of compact sets converging to $\mathbf{int}\, S$, and denote by $\boldsymbol{v}_n$ the unique vector of minimal norm in S_n. Then, as before we have $\forall \boldsymbol{z}_n \in S_n, \langle \boldsymbol{z}_n, \boldsymbol{v}_n \rangle \geq \|\boldsymbol{v}_n\|^2 \iff \langle \boldsymbol{z}_n, \frac{\boldsymbol{v}_n}{\|\boldsymbol{v}_n\|} \rangle \geq \|\boldsymbol{v}_n\| \geq 0$. This shows that for all $\boldsymbol{z} \in S$ and for all n large enough, $\langle \boldsymbol{z}, \frac{\boldsymbol{v}_n}{\|\boldsymbol{v}_n\|} \rangle \geq 0$. Now, the vectors $\frac{\boldsymbol{v}_n}{\|\boldsymbol{v}_n\|}$ live in the unit ball, which is compact, so we can consider a convergent subsequence, with $\lim_{i \to \infty} \frac{\boldsymbol{v}_{n_i}}{\|\boldsymbol{v}_{n_i}\|} = \bar{\boldsymbol{v}}$, where $\bar{\boldsymbol{v}}$ is a vector of unit norm. Taking the limit, we obtain that for all $\boldsymbol{z} \in S$, $\langle \boldsymbol{z}, \bar{\boldsymbol{v}} \rangle \geq 0$, where the vector $\bar{\boldsymbol{v}}$ is nonzero. Then, we conclude as above: $\inf_{\boldsymbol{y} \in Y} \langle \boldsymbol{y}, \bar{\boldsymbol{v}} \rangle \geq \sup_{\boldsymbol{x} \in X} \langle \boldsymbol{x}, \bar{\boldsymbol{v}} \rangle$. □

In some cases, it is possible to ensure the existence of a strict separating hyperplane. A sufficient condition is given in the following theorem:

Theorem 17 (Strict separating hyperplane)**.** *Let X, Y be two disjoint, nonempty, closed convex sets of $\mathbb{R}^n$. If X or Y is compact, then there exist a scalar c a vector $\boldsymbol{v} \in \mathbb{R}^n$, $\boldsymbol{v} \neq \boldsymbol{0}$, such that*

$$\forall \boldsymbol{x} \in X, \quad \langle \boldsymbol{x}, \boldsymbol{v} \rangle < c \quad \text{ and } \quad \forall \boldsymbol{y} \in Y, \quad \langle \boldsymbol{y}, \boldsymbol{v} \rangle > c.$$

In other words, the hyperplane $\{\boldsymbol{x} : \langle \boldsymbol{x}, \boldsymbol{v} \rangle = c\}$ strictly separates *X and Y.*

Proof. Assume that X is closed and Y is compact. Then, it can be seen that the Minkowski sum $S = Y + (-X)$ is closed (but not necessarily compact). We know from the previous proof that S has a vector $\boldsymbol{v}$ of minimal norm and does not contain $\{0\}$, so $\boldsymbol{v} \neq \boldsymbol{0}$. (Contrarily to the previous proof, we do not need to take the closure of S, since S is already closed.)

Following the same line of reasoning as in the previous proof, we now obtain:

$$\forall \boldsymbol{x} \in X, \forall \boldsymbol{y} \in Y, \quad \langle \boldsymbol{y}, \boldsymbol{v} \rangle > \langle \boldsymbol{x}, \boldsymbol{v} \rangle + \|\boldsymbol{v}\|^2.$$

Finally, set $c = \sup_{\boldsymbol{x} \in X} \langle \boldsymbol{x}, \boldsymbol{v} \rangle + \frac{1}{2}\|\boldsymbol{v}\|^2$, so it holds

$$\sup_{\boldsymbol{x} \in X} \langle \boldsymbol{x}, \boldsymbol{v} \rangle < c < \inf_{\boldsymbol{y} \in Y} \langle \boldsymbol{y}, \boldsymbol{v} \rangle.$$

□

As an exercise, you can try to find a counterexample showing that we can not get rid off the compactness assumption of either X or Y.

We can specialize the separation theorem for the case in which one of the two sets is a cone: In this case, the constant c can be taken equal to 0:

Theorem 18 (Separating hyperplane theorem for a cone)**.** *Let $C \subseteq \mathbb{R}^n$ be a nonempty convex cone, and $Y \subseteq \mathbb{R}^n$ be a nonempty convex set which does not intersect C. Then, there exists a vector $\boldsymbol{v} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ such that*

$$\forall \boldsymbol{x} \in C, \quad \langle \boldsymbol{x}, \boldsymbol{v} \rangle \leq 0 \quad \text{ and } \quad \forall \boldsymbol{y} \in Y, \quad \langle \boldsymbol{y}, \boldsymbol{v} \rangle \geq 0.$$

In other words, the hyperplane $\{\boldsymbol{x} : \langle \boldsymbol{x}, \boldsymbol{v} \rangle = 0\}$ separates separates *C and Y. If in addition, the cone C is closed and the set Y is compact, then there is a strict separating hyperplane:*

$$\forall \boldsymbol{x} \in C, \quad \langle \boldsymbol{x}, \boldsymbol{v} \rangle \leq 0 \quad \text{ and } \quad \forall \boldsymbol{y} \in Y, \quad \langle \boldsymbol{y}, \boldsymbol{v} \rangle > 0.$$

We leave the proof of this result as an exercise. Note that we often encounter the last theorem for the case where $Y = \{\boldsymbol{y}\}$ is a singleton: If $C \subseteq \mathbb{R}^n$ is a closed convex cone and $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{y} \notin C$, then

$$\exists \boldsymbol{v} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}: \quad \forall \boldsymbol{x} \in C,\ \boldsymbol{x}^T \boldsymbol{v} \leq 0 \quad \text{and} \quad \boldsymbol{y}^T \boldsymbol{v} > 0.$$

Applying the above result to the cone $C = \{A\boldsymbol{x} : \boldsymbol{x} \succeq_K \boldsymbol{0}\}$ for a proper cone K and a singleton $Y = \{\boldsymbol{b}\}$, we obtain the conic version of the Farkas lemma:

Lemma 19 (Farkas Lemma)**.** *Let K be a proper cone (or any cone such that $C = \{A\boldsymbol{x} : \boldsymbol{x} \succeq_K 0\}$ is closed). Then, exactly one of the following alternatices is true:*

1. $\exists \boldsymbol{x} \succeq \boldsymbol{0}: \ A\boldsymbol{x} = \boldsymbol{b}$;

2. $\exists \boldsymbol{y}: \ \boldsymbol{b}^T \boldsymbol{y} < 0$ *and* $A^T \boldsymbol{y} \succeq_{K^*} \boldsymbol{0}$.

Proof. We prove the equivalence (not(*1.*) $\iff$ *2.*). We have $\boldsymbol{b} \notin C$ if and only if there exists a strict separating hyperplane between C and $\{\boldsymbol{b}\}$: $\exists \boldsymbol{y} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}: \quad \forall \boldsymbol{x} \succeq_K \boldsymbol{0},\ \boldsymbol{y}^T A\boldsymbol{x} \geq 0$ and $\boldsymbol{b}^T \boldsymbol{y} < 0$. The condition $\forall \boldsymbol{x} \succeq_K \boldsymbol{0}$, $\boldsymbol{y}^T A\boldsymbol{x} = \langle \boldsymbol{x}, A^T \boldsymbol{y} \rangle \geq 0$ is equivalent to $A^T \boldsymbol{y} \succeq_{K^*} \boldsymbol{0}$. □

Let us now define the notion of supporting hyperplane:

Definition 8 (Supporting hyperplane)**.** Let $S \subseteq \mathbb{R}^n$ be nonempty, $\boldsymbol{a} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ and $b \in \mathbb{R}$. We say that $H = \{\boldsymbol{x} : \boldsymbol{a}^T \boldsymbol{x} = b\}$ is a supporting hyperplane of S if

- S is contained in one of the two halfspaces defined by H, i.e, $\forall \boldsymbol{x} \in S, \boldsymbol{a}^T \boldsymbol{x} - b \leq 0$ or $\forall \boldsymbol{x} \in S, \boldsymbol{a}^T \boldsymbol{x} - b \geq 0$.
- S has at least one boundary point on the hyperplane, i.e., $H \cap \partial S \neq \emptyset$, where $\partial S := \mathbf{cl}\, S \setminus \mathbf{int}\, S$ is the boundary of S.

The supporting hyperplane theorem is an important consequence of the separating hyperplane theorem:

Theorem 20 (Supporting hyperplane theorem)**.** *Let S be a convex set and $\boldsymbol{x}_0$ be a boundary point of S. Then, S has a supporting hyperplane at $\boldsymbol{x}_0$, that is,*

$$\exists \boldsymbol{a} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}: \quad \forall \boldsymbol{x} \in S,\ \boldsymbol{a}^T \boldsymbol{x} \leq \boldsymbol{a}^T \boldsymbol{x}_0.$$

Conversely, if S is closed , has nonempty interior, and has (at least) one supporting hyperplane in each of its boundary points, then S is convex.

Proof. If S has a nonempty interior, then the supporting hyperplane at $\boldsymbol{x}_0$ is found by applying the separating hyperplane theorem between the sets $\mathbf{int}\, S$ and $\{\boldsymbol{x}_0\}$. Otherwise, S is contained in some affine space $\mathcal{A}$ of dimension less than n, so in particular S is contained in some hyperplane, which is a (trivial) supporting hyperplane at $\boldsymbol{x}_0$.

We leave the proof of the (partial) converse statement as an exercise. □

8 Extreme points, extreme rays, and Minkowski theorems

Definition 9 (Face)**.** Let $S \subseteq \mathbb{R}^n$ be a convex set. We say that F is a face of S if F is convex, and, for any $\boldsymbol{x} \in F$, if $\boldsymbol{x} = \lambda \boldsymbol{a} + (1-\lambda)\boldsymbol{b}$, where $\boldsymbol{a}, \boldsymbol{b} \in S$ and $\lambda \in (0,1)$, (i.e., $\boldsymbol{x}$ is a convex combination of two points $\boldsymbol{a}$ and $\boldsymbol{b}$ in S), then $\boldsymbol{a}, \boldsymbol{b} \in F$.

Definition 10 (Extreme point)**.** Let $S \subseteq \mathbb{R}^n$ be a convex set. The point $\boldsymbol{x}$ is an extreme point of S if $\{\boldsymbol{x}\}$ is a face of S, that is, extreme points are faces of dimension 0:

$$\boldsymbol{x} \text{ is an extreme point of } S \iff \big(\exists \boldsymbol{a}, \boldsymbol{b} \in S, \lambda \in (0,1) : \ \boldsymbol{x} = \lambda \boldsymbol{a} + (1-\lambda)\boldsymbol{b} \implies \boldsymbol{a} = \boldsymbol{b} = \boldsymbol{x}\big).$$

The Minkowski theorem tells us that any point in a compact convex set can be written as a convex combination of its extreme points. We do not prove this result, but we mention it can be proved by induction on the dimension of S.

Theorem 21 (Minkowski theorem)**.** *Let S be a compact convex set of $\mathbb{R}^n$. Denote by* $\operatorname{ext}(S)$ *the set of extreme points of S. Then,* $S = \mathbf{conv}\Big(\operatorname{ext}(S)\Big)$.

There is another, dual characterization of convex sets, which relies on the notion of supporting hyperplanes: Every convex set is the intersection of all halfspaces that contain it.

Theorem 22. *Let S be a closed convex set of $\mathbb{R}^n$. Then,*

$$S = \bigcap_{\substack{H \text{ halfspace} \\ H \supseteq S}} H.$$

Proof. Let T denote the set in the right hand side of the statement. The inclusion $S \subseteq T$ is trivial. For the converse inclusion, we show $\boldsymbol{x} \notin S \implies \boldsymbol{x} \notin T$. If $\boldsymbol{x} \notin S$, then we know that there exists a strict separating hyperplane between $\{\boldsymbol{x}\}$ and S:

$$\exists \boldsymbol{a} \in \mathbb{R}^n \setminus \{0\}, \exists b \in \mathbb{R} : \quad \boldsymbol{a}^T \boldsymbol{x} < b \quad \text{and} \quad \forall \boldsymbol{y} \in S, \boldsymbol{a}^T \boldsymbol{y} > b.$$

This shows that $\boldsymbol{x}$ is not included in the halfspace $H := \{\boldsymbol{y} : \boldsymbol{a}^T \boldsymbol{y} > b\}$, which contains S. Hence, $\boldsymbol{x} \notin T$. ☐

Let us now define the notion of *extreme rays* of a cone, which play a similar role as extreme points for bounded convex sets.

Definition 11 (Extreme ray)**.** Let K be a convex cone. A set of the form $R = \{\lambda \boldsymbol{x} : \lambda \geq 0\}$, where $\boldsymbol{x} \in K \setminus \{\boldsymbol{0}\}$, is called an *extreme ray* if

$$(\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z}, \ \boldsymbol{x} \in R, \ \boldsymbol{y}, \boldsymbol{z} \in K) \implies \boldsymbol{y}, \boldsymbol{z} \in R.$$

The Minkowski theorem for closed, convex, pointed cones follow:

Theorem 23. *Let K be a closed convex pointed cone in $\mathbb{R}^n$. Then, K is the conic hull of its extreme rays.*

Proof. We must show that any $\boldsymbol{x} \in K$ can be written as a conic combination of extreme rays of K.

The interior of the dual cone K^* is $\mathbf{int}\, K^* = \{\boldsymbol{y} : \ \boldsymbol{x}^T \boldsymbol{y} > 0, \ \forall \boldsymbol{x} \in K \setminus \{\boldsymbol{0}\}\}$. We know from Proposition 14 that since K is closed and pointed, K^* has a nonempty interior. Let $\boldsymbol{a} \in \mathbf{int}\, K^*$, so $K \setminus \{\boldsymbol{0}\}$ is included in the open halfspace $\{\boldsymbol{x} : \ \boldsymbol{a}^T \boldsymbol{x} > 0\}$.

We claim that $K \cap \{\boldsymbol{x} : \boldsymbol{a}^T \boldsymbol{x} = 1\}$ is bounded, hence compact. Otherwise, there is a sequence such that $\|\boldsymbol{x}_n\| \to \infty$ and $\boldsymbol{a}^T \boldsymbol{x}_n = 1 \implies \boldsymbol{a}^T \frac{\boldsymbol{x}_n}{\|\boldsymbol{x}_n\|} \to 0$. Denote by $\boldsymbol{v}$ a limit point of the sequence $\frac{\boldsymbol{x}_n}{\|\boldsymbol{x}_n\|}$, so $\boldsymbol{a}^T \boldsymbol{v} = 0$. The points $\frac{\boldsymbol{x}_n}{\|\boldsymbol{x}_n\|}$ are in K, so since K is closed we have $\boldsymbol{v} \in K$, and $\boldsymbol{v} \neq \boldsymbol{0}$ because it is a vector of unit norm. But $\boldsymbol{v} \in K \setminus \{\boldsymbol{0}\}$ contradicts $\boldsymbol{a}^T \boldsymbol{v} = 0$.

Now, $C := K \cap \{\boldsymbol{x} : \boldsymbol{a}^T \boldsymbol{x} = 1\}$ is a compact convex set, so we can apply the Minkowski theorem: if $\boldsymbol{x} \in C$, then $\boldsymbol{x}$ can be written as a convex combination of extreme points of C. It remains to show that each of these extreme points are on extreme rays of K, and the result will follow. Let $\boldsymbol{z} \in C$, and assume that $\boldsymbol{z} = \boldsymbol{u} + \boldsymbol{v}$ for some vectors $\boldsymbol{u}, \boldsymbol{v} \in K$. Then,

$$\boldsymbol{z} = \underbrace{(\boldsymbol{a}^T \boldsymbol{u})}_{\lambda_1} \underbrace{\frac{\boldsymbol{u}}{\boldsymbol{a}^T \boldsymbol{u}}}_{\boldsymbol{z}_1} + \underbrace{(\boldsymbol{a}^T \boldsymbol{v})}_{\lambda_2} \underbrace{\frac{\boldsymbol{v}}{\boldsymbol{a}^T \boldsymbol{v}}}_{\boldsymbol{z}_2}.$$

It is easy to see that $\boldsymbol{z}_1, \boldsymbol{z}_2 \in C$. Moreover, $\lambda_1 + \lambda_2 = \boldsymbol{a}^T(\boldsymbol{u} + \boldsymbol{v}) = \boldsymbol{a}^T \boldsymbol{z} = 1$, so $\boldsymbol{z}$ is a convex combination of $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$. Since $\boldsymbol{z}$ is an extreme point of C, then we must have $\boldsymbol{z}_1 = \boldsymbol{z}_2 = \boldsymbol{z}$, which implies that $\boldsymbol{u}$ and $\boldsymbol{v}$ lie on the ray $R = \{\lambda \boldsymbol{z} : \ \lambda \geq 0\}$, hence R is an extreme ray of K.

□