ECE285: Semidefinite and sum-of-squares optimization **Winter 2024**

Lecture 3 & 4: Review of Convexity (I)-(II)

Lecturer: Yang Zheng *Scribe: Yang Zheng*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. Any typos should be sent to* `zhengy@eng.ucsd.edu`.

Learning goals:

1. Basic topology in $\mathbb{R}^n$
2. Convex sets and examples
3. Separating hyperplane theorems
4. External and internal representations

The notion of convexity is essential in semidefinite and sum-of-squares optimization. In Lectures 3 and 4, we review some basic elements of convex analysis. More comprehensive treatments can be found in many convex optimization books, e.g., [2, Chapter 2], [3, Chapter 2].

1 Basic topology in $\mathbb{R}^n$

1.1 Open and closed sets

The (Euclidean) ball with center $x \in \mathbb{R}^n$ and radius $r \in \mathbb{R}$ is

$$B(x, r) = \{y \in \mathbb{R}^n \mid \|y - x\|_2 \leq r\}.$$

Given a set $C \subseteq \mathbb{R}^n$, a point $x \in C$ is an *interior point* of C if there exists an $\epsilon > 0$ such that $B(x, \epsilon) \subseteq C$, i.e., there exists a ball centered at x that belongs to C. We denote the set of all interior points of C as $\operatorname{int} C$.

A set $C \subseteq \mathbb{R}^n$ is called *open* if $C = \operatorname{int} C$, i.e., all points in C are interior points. A set C is closed if its complement $\mathbb{R}^n \setminus C = \{x \in \mathbb{R}^n \mid x \notin C\}$ is open. The closure of C, denoted as $\operatorname{cl} C$, is the smallest closed set that contains C. We can also characterize closed sets in terms of convergent sequences and limit points. A set $C \in \mathbb{R}^n$ is closed if and only if every converging sequence of points in C has its limit point in C. The closure of C is the closed set that contains all limit points of convergent sequences in C.

The *boundary* of the set C is defined as $\partial C = \operatorname{cl} C \setminus \operatorname{int} C$. In other words, for any point $x \in \partial C$, an arbitrarily small ball $B(x, \epsilon), \forall \epsilon > 0$ contains points in C and points outside C. For example, the boundary of the unit ball with a center 0 is the unit sphere

$$\partial B(0, 1) = \{x \in \mathbb{R}^n \mid x^{\mathsf{T}} x = 1\}.$$

This is often called the $(n-1)$-dimensional unit sphere, denoted as $\mathcal{S}^{n-1}$.

A set $C \in \mathbb{R}^n$ is *compact* if every sequence in C has a subsequence that converges to a point in C. The set C is compact if and only if it is closed and bounded.

1.2 Affine subspaces and relative interior

A subspace in $\mathbb{R}^n$ is a non-empty subset $V \subseteq \mathbb{R}^n$ that is closed under sums and scalar multiplication (i.e., if $u, v \in V$, then $u + v \in V$, and $\alpha u \in V, \forall \alpha \in \mathbb{R}$). For example, the span (or *linear hull*) of a set of vectors $x_i \in \mathbb{R}^n, i = 1, \ldots, m$, defined as

$$\operatorname{span}(x_1, \ldots, x_m) = \left\{ \sum_{i=1}^{m} \alpha_i x_i \mid \alpha_i \in \mathbb{R}, i = 1, \ldots, m \right\} \tag{1}$$

is a subspace in $\mathbb{R}^n$. The basis of a subspace $V \subseteq \mathbb{R}^n$ is a set of independent vectors whose span equals to V. The number of vectors in the basis is called the *dimension* of V.

A subset $A \subseteq \mathbb{R}^n$ is called an affine subspace if we can write it as

$$A = x_0 + V = \{x_0 + y \mid y \in V\}, \tag{2}$$

where $x_0 \in \mathbb{R}^n$ and $V \subseteq \mathbb{R}^n$ is a subspace. The dimension of an affine subspace A is defined as the dimension of V.

Example 3.1. *One-dimensional affine subspaces are lines: a line is a set of the form* $\{\theta x + (1 - \theta) y \in \mathbb{R}^n \mid \theta \in \mathbb{R}\}$, *where* $x, y \in \mathbb{R}^n$ *are two points, and* $(n-1)$*-dimensional affine subspaces are hyperplanes: a hyperplane is a set of the form*

$$H = \{x \in \mathbb{R}^n \mid c^{\mathsf{T}} x = b\}, \tag{3}$$

where $c \in \mathbb{R}^n \setminus \{0\}$ *is the normal of* H *and* $b \in \mathbb{R}$. *Given any point* $x_0 \in H$, (3) *can also be represented as* $H = \{x \in \mathbb{R}^n \mid c^{\mathsf{T}}(x - x_0) = 0\}$. □

From (2), it is clear that an affine set contains every *affine combination* of its points: If A is an affine set, $x_1, \ldots, x_k \in A$, and $\alpha_1 + \ldots + \alpha_k = 1$, then $\alpha_1 x_1 + \ldots + \alpha_k x_k \in A$. The *affine hull* of an arbitrary set $C \subseteq \mathbb{R}^n$, denoted as $\operatorname{aff} C$, is the set of all affine combinations of its points

$$\operatorname{aff} C = \{\alpha_1 x_1 + \ldots + \alpha_k x_k \mid x_1, \ldots, x_k \in C, \alpha_1 + \cdots + \alpha_k = 1, k \geq 1\}.$$

The affine hull is the smallest affine subspace that contains C. The dimension of $C \subseteq \mathbb{R}^n$ is defined as the dimension of its affine hull.

A set $C \subseteq \mathbb{R}^n$ is called *full dimensional* if its dimension equals to n, i.e., $\operatorname{aff} C = \mathbb{R}^n$. A set C is full dimensional if and only if its interior is non-empty. For example, any ball $B(x, r)$ with radius $r > 0$ is full dimensional. If the dimension of C is smaller than n, then C does not have any interior point, i.e., $\operatorname{int} C = \emptyset$. In this case, we are interested in its interior relative to its affine subspace $\operatorname{aff} C$. The relative interior of C is defined as

$$\operatorname{relint} C = \{x \in C \mid B(x, \epsilon) \cap \operatorname{aff} C \subseteq C, \text{ for some } \epsilon > 0\}.$$

2 Convex sets

2.1 Basic properties

Definition 3.1. *A set* $C \subseteq \mathbb{R}^n$ *is called convex if for any* $x, y \in C$ *and any* θ *with* $0 \leq \theta \leq 1$, *we have*

$$\theta x + (1 + \theta) y \in C.$$

In other words, a convex set contains the line segment between any two of its points.

A *convex combination* of points $x_1, x_2, \ldots, x_m$ is a point of the form

$$\alpha_1 x_1 + \cdots + \alpha_m x_m, \quad \text{where } \alpha_1 + \cdots + \alpha_m = 1, \alpha_i \geq 0, i = 1, \ldots, m.$$

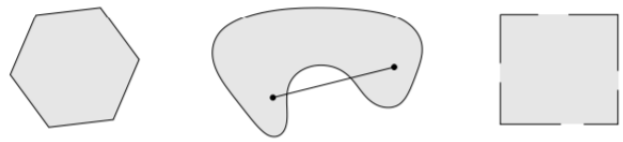


Figure 1: Some simple convex and nonconvex sets (images from [2, Figgure 2.2]).

A convex set contains all convex combinations of its points.

The *convex hull* of an arbitrary set $C \subseteq \mathbb{R}^n$, denoted as $\operatorname{conv} C$, is the set of all convex combinations of its points

$$\operatorname{conv} C = \{\alpha_1 x_1 + \ldots + \alpha_m x_m \mid x_i \in C, \alpha_i \geq 0, i = 1, \ldots, m, \alpha_1 + \cdots + \alpha_m = 1\}. \tag{4}$$

The convex hull of the set C is the smallest convex set that contains C, i.e., the intersection of all convex sets containing C,

$$\operatorname{conv} C = \bigcap_{C \subseteq B,\ B \text{ is convex}} B. \tag{5}$$

Figure 2 illustrates the definition of the convex hull.

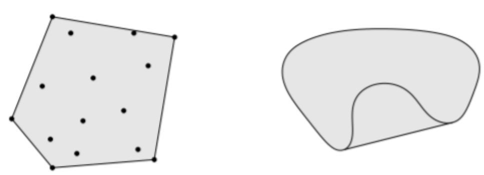


Figure 2: The convex hulls of two sets in $\mathbb{R}^2$ (images from [2, Figure 2.3]).

In (4) ((5), resp.), we present an "internal" ("external", resp.) representation of the convex hull of the set $C \in \mathbb{R}^n$. In Section 4, we will further discuss these two aspects of convex sets. Note that in (4) we considered all convex combinations of its points of C, i.e., any number m and any selections of m points. In fact, the value of m can be restricted. The Carathéodory theorem asserts that the number of points m need not be larger than $n + 1$.

Theorem 3.1 (Carathéodory)**.** *Let $C \subseteq \mathbb{R}^n$. Then, every point in $\operatorname{conv} C$ is a convex combination of at most $n+1$ points in C, i.e., if $x \in \operatorname{conv} C$, there exist $x_i \in C$ and $\alpha_i \geq 0, i = 1, \ldots, n+1$ with $\alpha_1 + \cdots + \alpha_{n+1} = 1$, such that*

$$x = \alpha_1 x_1 + \cdots + \alpha_{n+1} x_{n+1}.$$

In Section 4, we will further show that every point in $\operatorname{conv} C$ can be represented by some special points if $\operatorname{conv} C$ is compact (i.e., closed and bounded).

Another useful property is that a convex set with an empty interior can be embeded into a lower dimensional affine subspace.

Lemma 3.1 ([3, Lemma 2.9])**.** *Let $C \subseteq \mathbb{R}^n$ be a convex set. Then $\operatorname{int} C = \emptyset$ if and only if it is contained in an affine subspace with dimension at most $n - 1$.*

2.2 Examples

We start with some simple examples that are convex.

- By definition, any subspace (1) and affine subspace (2) contain the entire line between any of its two points, so they are convex (of course they contain the line segment between any of its two points). For example, $\mathbb{R}^n$ is convex.
- Any line, which has the form of $\{\theta x + (1-\theta)y \mid \theta \in \mathbb{R}\}$ where $x, y \in \mathbb{R}^n$, is a one-dimensional affine subspace, so it is convex.
- A half line, i.e., a ray, which has the form of $\{x_0 + \theta v \mid \theta \geq 0\}$ where $x_0, v \in \mathbb{R}^n, v \neq 0$, is convex.
- Any hyperplane $H = \{x \in \mathbb{R}^n \mid c^\mathsf{T} x = b\}$ where $c \in \mathbb{R}^n, c \neq 0, b \in \mathbb{R}$ is an $(n-1)$-dimensional affine subspace, so it is convex.
- A hyperplane H splits $\mathbb{R}^n$ into two halfspaces $H^+ = \{x \in \mathbb{R}^n \mid c^\mathsf{T} x \geq b\}$ and $H^- = \{x \in \mathbb{R}^n \mid c^\mathsf{T} x \leq b\}$. The interior of H^- is $\{x \in \mathbb{R}^n \mid c^\mathsf{T} x < b\}$, which is called an *open halfspace*. Halfspaces (closed or open) are convex.

We next discuss a few other common examples that are convex.

- Given any norm $\|\cdot\|$ on $\mathbb{R}^n$, a *norm ball* with radius $r > 0$ and center x_c, given by $\{x \in \mathbb{R}^n \mid \|x - x_\mathrm{c}\| \leq r\}$ is convex: if $\|x_1 - x_\mathrm{c}\| \leq r$, $\|x_2 - x_\mathrm{c}\| \leq r$, and $0 \leq \theta \leq 1$, we have

$$\begin{aligned}\|\theta x_1 + (1-\theta)x_2 - x_\mathrm{c}\| &= \|\theta(x_1 - x_\mathrm{c}) + (1-\theta)(x_2 - x_\mathrm{c})\| \\ &\leq \theta\|x_1 - x_\mathrm{c}\| + (1-\theta)\|x_2 - x_\mathrm{c}\| \\ &\leq r,\end{aligned}$$

  where we have applied the homogeneity property and triangle inequality for norms.

- In particular, the unit l_p-ball, defined as

$$B_p^n = \{x \in \mathbb{R}^n \mid \|x\|_p \leq 1\},$$

  is convex for $p \geq 1$. When $p = 1$, B_1^n is called the *cross-polytope*; when $p = 2$, B_2^n is known as the *unit Euclidean ball*, and if $p \to \infty$, $B_\infty^n = [-1, 1]^n$ is called the *n-dimensional cube*.

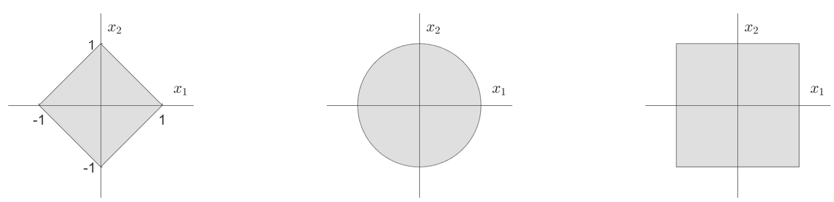


Figure 3: The unit balls in $\mathbb{R}^2$ with norms l_1 (left), l_2 (center), and l_∞ (right).

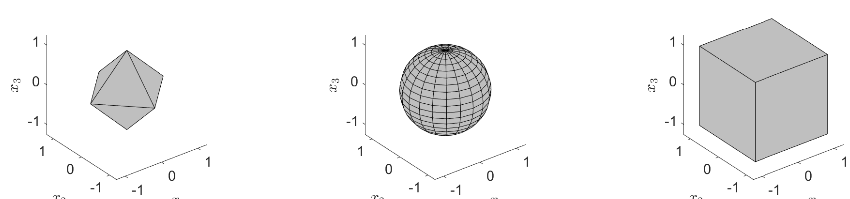


Figure 4: The unit balls in $\mathbb{R}^3$ with norms l_1 (left), l_2 (center), and l_∞ (right).

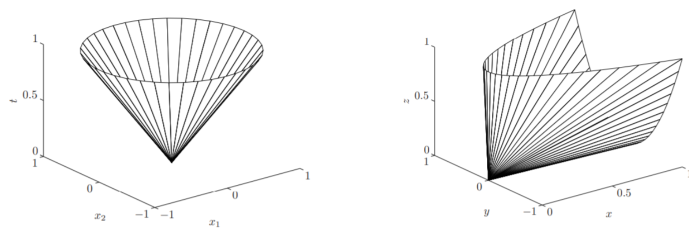
- Similarly, given any norm $\|\cdot\|$ on $\mathbb{R}^{m\times n}$, the norm ball $\{X \in \mathbb{R}^{m\times n} \mid \|X - X_0\| \leq r\}$ is convex.

As one central object, the set of positive semidefinite matrices $\mathbb{S}^n_+$ is convex: if $X_1 \in \mathbb{S}^n_+, X_2 \in \mathbb{S}^n_+$, $0 \leq \theta \leq 1$ then we have

$$x^\mathsf{T}(\theta X_1 + (1-\theta)X_2)x = \theta x^\mathsf{T} X_1 x + (1-\theta)x^\mathsf{T} X_2 x \geq 0, \ \ \forall x \in \mathbb{R}^n,$$

confirming $\theta X_1 + (1-\theta)X_2 \in \mathbb{S}^n_+$. Similarly, the set of positive definite matrices $\mathbb{S}^n_{++}$ is convex.

The second-order cone $\{(x,t) \in \mathbb{R}^{n+1} \mid \|x\|_2 \leq t\}$ is convex. This is also known as "ice-cream cone".



(a) (b)

Figure 5: (a) Boundary of second-order cone in $\mathbb{R}^3$: $\{(x_1,x_2,t) \mid (x_1^2+x_2^2)^{1/2} \leq t\}$. (b) Boundary of positive semidefinite cone $\mathbb{S}^2_+ := \{X = \begin{bmatrix} x & y \\ y & z \end{bmatrix} \mid x \geq 0, z \geq 0, xz - y^2 \geq 0\}$ (images from [2, Figures 2.10 & 2.12]).

As another central object, the set of univariate nonnegative polynomials, defined as

$$\{(a_0, a_1, \ldots, a_n) \in \mathbb{R}^{n+1} \mid a_0 + a_1 x + \ldots + a_n x^n \geq 0, \forall x \in \mathbb{R}\},$$

is convex (the verification is straightforward). For example, when $n = 2$, the set of univariate nonnegative quadratic polynomials is $\{(a_0,a_1,a_2) \in \mathbb{R}^3 \mid a_0 + a_1 x + a_2 x^2 \geq 0, \forall x \in \mathbb{R}\} = \{(a_0,a_1,a_2) \in \mathbb{R}^3 \mid a_2 \geq 0, a_1^2 - 4a_2 a_0 \leq 0\}$.

Finally, a *polytope* in $\mathbb{R}^n$ is the convex hull of finitely many points $x_0, \ldots, x_m \in \mathbb{R}^n$, which is convex by definition.

- If these $m+1$ points are affinely independent, meaning that $x_1 - x_0, \ldots, x_m - x_0$ are linearly independent (thus $m \leq n$), the convex hull $\mathrm{conv}\{x_0, \ldots, x_m\}$ is called an *m-dimensional simplex* in $\mathbb{R}^n$.
- The *unit simplex* is the n-dimensional simplex generated by the zero vector and the unit vectors, i.e., $0, e_1, \ldots, e_n \in \mathbb{R}^n$, which can be expressed as $\{x \in \mathbb{R}^n \mid x_1 + \cdots + x_n \leq 1, x_i \geq 0, i = 1, \ldots, n\}$.
- The *probability simplex* is the $(n-1)$-dimensional simplex generated by the unit vectors, which is $\{x \in \mathbb{R}^n \mid x_1 + \cdots + x_n = 1, x_i \geq 0, i = 1, \ldots, n\}$

2.3 Intersection and affine functions

We here discuss two important operations that preserve convexity. First, convexity is preserved under the operation of intersection: if C_1 and C_2 are convex, then $C_1 \cap C_2$ is convex. This property extends to the intersection of an arbitrary number of sets.

Lemma 3.2. *Let I be an arbitrary index set. If the sets $C_i \subseteq \mathbb{R}^n, i \in I$ are convex, then $C = \bigcap_{i\in I} C_i$ is convex.*

This property allows us to easily establish the convexity of some common sets.

Example 3.2 (Polyhedra)**.** *A polyhedron is defined as the solution set of a finite number of affine inequalities and equalities*

$$\mathcal{P} = \{x \in \mathbb{R}^n \mid a_j^\mathsf{T} x \leq b_j, j = 1, \ldots, m, c_j^\mathsf{T} x = d_j, j = 1, \ldots, p\}, \tag{6}$$

which is convex, since it is an intersection of a finite number of halfspaces and hyperplanes

$$\mathcal{P} = \bigcap_{j=1}^{m} \{x \in \mathbb{R}^n \mid a_j^\mathsf{T} x \leq b_j\} \bigcap_{j=1}^{p} \{x \in \mathbb{R}^n \mid c_j^\mathsf{T} x = d_j\}.$$

When a polyhedron $\mathcal{P}$ is bounded, then it is a polytope, i.e., it can also be represented by a convex hull of finitely many points, i.e., $\mathcal{P} = \mathrm{conv}\{x_1, \ldots, x_k\}$ for some value of k. In general, any polyhedron $\mathcal{P}$ in (6) *can also be represented in the form (and vice versa)*

$$\{\theta_1 x_1 + \cdots + \theta_k x_k \mid \theta_1 + \cdots + \theta_m = 1, \theta_i \geq 0, i = 1, \ldots, k, m \leq k\}. \tag{7}$$

Given a polyhedron $\mathcal{P}$, (6) *can be considered as an external representation and* (7) *can be considered as an internal representation. These two representations can, however, differ significantly in size for the same polyhedron $\mathcal{P}$.* □

Example 3.3. *The set of positive semidefinite matrices can be represented as*

$$\mathbb{S}_+^n = \bigcap_{x \neq 0} \{A \in \mathbb{S}^n \mid x^\mathsf{T} A x \geq 0\},$$

which is an intersection of infinitely many halfspaces in $\mathbb{S}^n$, indexed by each $x \neq 0$ as $\{A \in \mathbb{S}^n \mid x^\mathsf{T} A x \geq 0\}$ (note that $f(A) = x^\mathsf{T} A x$ is a linear function in $A \in \mathbb{S}^n$). Thus, $\mathbb{S}_+^n$ is convex. Similarly, the set of univariate nonnegative polynomials can be written as

$$\bigcap_{x \in \mathbb{R}} \{(a_0, a_1, \ldots, a_n) \in \mathbb{R}^{n+1} \mid a_0 + a_1 x + \ldots + a_n x^n \geq 0\},$$

which is an intersection of infinitely many halfspaces in $\mathbb{R}^{n+1}$, and thus it is convex. □

In the examples above, we establish convexity of a set by expressing it as an intersection of (possibly infinite) halfspaces. Indeed, we will discuss in Section 4 that every closed convex set C is an intersection of (usually infinite) halfspaces (Theorem 3.8).

Another useful operation that preserves convexity is via *affine functions*. Recall that a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is affine if it has the form $f(x) = Ax + b$, where $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$.

Lemma 3.3. *Let $C \subseteq \mathbb{R}^n$ be a convex set.*

1. *If $f : \mathbb{R}^n \to \mathbb{R}^m$ is an affine function, then the image of C under f, defined as $f(C) = \{f(x) \in \mathbb{R}^m \mid x \in C\}$, is convex in $\mathbb{R}^m$.*
2. *If $f : \mathbb{R}^k \to \mathbb{R}^n$ is an affine function, then the pre-image of C under f, defined as $f^{-1}(C) = \{z \in \mathbb{R}^k \mid f(z) \in C\}$, is convex in $\mathbb{R}^k$.*

Proof. For statement (1), let $y_1, y_2 \in f(C)$. There exist $x_1, x_2 \in C$ such that $y_1 = f(x_1)$, $y_2 = f(x_2)$. Then, for any $0 \leq \theta \leq 1$, we have

$$\theta y_1 + (1 - \theta) y_2 = \theta f(x_1) + (1 - \theta) f(x_2) = f(\theta x_1 + (1 - \theta) x_2) \in f(C),$$

where we have applied the affine property of $f(x)$ and the convexity of C.

For statement (2), let $z_1, z_2 \in f^{-1}(C)$, meaning that $f(z_1), f(z_2) \in C$. Then, for any $0 \leq \theta \leq 1$, we have

$$f(\theta z_1 + (1 - \theta) z_2) = \theta f(z_1) + (1 - \theta) f(z_2) \in C.$$

This finishes the proof. □

Let us discuss a few affine operations below.

- *Scaling* and *translation* are two simple affine operations: if $C \subseteq \mathbb{R}^n$ is convex, then for any $\alpha \in \mathbb{R}, a \in \mathbb{R}^n$, the sets
$$\alpha C = \{\alpha x \mid x \in C\}, \qquad C + a = \{x + a \mid x \in C\},$$
are both convex.
- The *Minkowski sum* of two convex sets $C_1, C_2 \subseteq \mathbb{R}^n$ preserves convexity:
$$C_1 + C_2 = \{x_1 + x_2 \in \mathbb{R}^n \mid x_1 \in C_1, x_2 \in C\}.$$
The set $C_1 + C_2$ is the image of the direct (or Cartesian) product $C_1 \times C_2 = \{(x_1, x_2) \in \mathbb{R}^{2n} \mid x_1 \in C_1, x_2 \in C_2\}$ (that is convex) under the linear function $f(x_1, x_2) = x_1 + x_2$, thus it is convex.
- The *projection* of a convex set onto some of its coordinates is convex: if $C \subseteq \mathbb{R}^m \times \mathbb{R}^n$ is convex, then its projection onto the first m coordinates
$$S = \{x_1 \in \mathbb{R}^m \mid (x_1, x_2) \in C \text{ for some } x_2 \in \mathbb{R}^n\}$$
is convex. The set S can be seen as the image of C under the linear function $f(x_1, x_2) = x_1 + 0_{m\times n} \times x_2$, and it is thus convex.

We end this section by noting that the solution set of a linear matrix inequality, given by
$$A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n \preceq 0,$$
where $A_i \in \mathbb{S}^n, i = 0, \ldots, n$ are given symmetric matrices, is convex. The set $\{x \in \mathbb{R}^n \mid A(x) \preceq 0\}$ can be viewed as the pre-image of $\mathbb{S}^n_+$ under the affine function $f(x) = -A(x)$.

3 Separating hyperplane theorems

In this section, we introduce several fundamental results on separating hyperplanes for convex sets. Let us first present the definitions of separating hyperplanes.

Recall that a hyperplane is in the form of $H = \{x \in \mathbb{R}^n \mid c^\mathsf{T} x = b\}$, where $c \in \mathbb{R}^n \setminus \{0\}$ is the normal vector of H and $b \in \mathbb{R}$. Note that when H passes through a point z_0, we have $b = c^\mathsf{T} z_0$. This hyperplane H splits $\mathbb{R}^n$ into two closed halfspaces
$$H^+ = \{x \in \mathbb{R}^n \mid c^\mathsf{T} x \geq b\}, \text{ and } H^- = \{x \in \mathbb{R}^n \mid c^\mathsf{T} x \leq b\}.$$
A hyperplane H is said to separate two sets $C \subseteq \mathbb{R}^n$ and $D \subseteq \mathbb{R}^n$ if they lie on different sides of H, i.e., $C \subseteq H^+$, $D \subseteq H^-$ (or $C \subseteq H^-$, $D \subseteq H^+$). In other words, the sets $C \subseteq \mathbb{R}^n$ and $D \subseteq \mathbb{R}^n$ are separated by a hyperplane if there exists a non-zero vector $c \in \mathbb{R}^n$ and a scalar $b \in \mathbb{R}$ such that
$$c^\mathsf{T} x - b \geq 0, \forall x \in C, \quad \text{and} \quad c^\mathsf{T} x - b \leq 0, \forall x \in D. \tag{8}$$
The separation is said to be *strict* if the inequalities (8) are both strict, i.e.,
$$c^\mathsf{T} x - b > 0, \forall x \in C, \quad \text{and} \quad c^\mathsf{T} x - b < 0, \forall x \in D. \tag{9}$$
Note that (8) and (9) can also be written as $c^\mathsf{T} x \geq c^\mathsf{T} y, \forall x \in C, y \in D$ and $c^\mathsf{T} x > c^\mathsf{T} y, \forall x \in C, y \in D$, respectively.

Another closely related notion is *supporting hyperplanes*. A hyperplane is said to support a set $C \subseteq \mathbb{R}^n$ at a point $x \in C$ if $x \in H$ and the set C lies entirely in one of the halfspaces H^+ or H^-. In this case, H is called a *supporting hyperplane* of C at point x. The corresponding halfspace (H^+ or H^-) that contains C is called a supporting halfspace.

3.1 Projections

We here introduce a projection operator that allows us to construct separating hyperplanes in the next section.

Consider a closed convex set $C \in \mathbb{R}^n$, a point $x \in \mathbb{R}^n$, and a norm $\|\cdot\|$. The distance of the point x to the set C is defined as

$$\mathrm{dist}(x, C) = \min_{y \in C} \|x - y\|. \tag{10}$$

Any point $z \in C$ that achieves the minimum $\|z - x\| = \mathrm{dist}(x, C)$, i.e., closest to x, is called a *projection* of x on C. Note that for l_1 and l_∞ norms, the minimizers of (10) may be non-unique (since the l_1 and l_∞ norms are piece-wise linear functions), i.e., there may be more than one points in C that are closest to x.

Example 3.4. *Consider a convex set $C = \{(t, -t) \in \mathbb{R}^2 \mid t \in \mathbb{R}\}$, and $x = (1, 1)$. Then*

$$\min_{y \in C} \|x - y\|_1 = \min_{t \in \mathbb{R}} \|(1 - t, 1 + t)\|_1 = \min_{t \in \mathbb{R}} |1 - t| + |1 + t| \geq |1 - t + 1 + t| = 2.$$

Meanwhile, we have $\|(1, -1) - x\|_1 = 2$, $\|(0, 0) - x\|_1 = 2$, and $\|(-1, 1) - x\|_1 = 2$. All three points $(1, -1)$, $(0, 0)$ and $(-1, 1)$ are projections of x on C under the l_1 norm. □

Throughout this course, we only consider the Euclidean norm in (10). It can then be shown that the projection is unique, denoted as $P_C(x)$, i.e.,

$$P_C(x) = \arg\min_{y \in C} \|x - y\|_2. \tag{11}$$

Theorem 3.2 ([3, Section 2.1.2])**.** *Let $C \subseteq \mathbb{R}^n$ be a non-empty, closed and convex set. Then, for any $x \in \mathbb{R}^n$, there exists a unique point $z \in C$ that is closest to x under the Euclidean norm. Further, $z = P_C(x)$ if and only if*

$$\langle y - z, x - z \rangle \leq 0, \quad \forall y \in C. \tag{12}$$

If C is an affine subspace, then $\forall y \in C$, we have $P_C(x) - y + P_C(x) = 2P_C(x) - y \in C$. In this case, (12) reads as

$$\begin{aligned} &\langle y - P_C(x), x - P_C(x) \rangle \leq 0 \\ &\langle P_C(x) - y, x - P_C(x) \rangle \leq 0, \forall y \in C, \end{aligned}$$

which implies $\langle y - P_C(x), x - P_C(x) \rangle = 0, \forall y \in C$, i.e., $x - P_C(x) \perp C$. In the case of Example 3.4, we have $P_C(x) = (0, 0)$, and $\langle x - 0, (t, -t) - 0 \rangle = t - t = 0, \forall t \in \mathbb{R}$.

One useful property is that the Euclidean projection (11) is nonexpansive.

Lemma 3.4. *Let $C \subseteq \mathbb{R}^n$ be a non-empty closed convex set. Then, we have*

$$\|P_C(x) - P_C(y)\|_2 \leq \|x - y\|_2, \quad \forall x, y \in \mathbb{R}^n.$$

3.2 Separating hyperplanes

A closed convex set and a point outside it can be strictly separated by a hyperplane, and this hyperplane can be constructed by using the Euclidean projection.

Theorem 3.3. *Let $C \subseteq \mathbb{R}^n$ be a closed convex set, and $x \notin C$. Let $P_C(x) \in \mathbb{R}^n$ be the projection of x on C. Then, the hyperplane passing through point $z_0 = \frac{1}{2}(x + P_C(x))$ with normal vector $c = x - P_C(x)$ strictly separates $\{x\}$ and C, i.e.*

$$c^{\mathsf{T}} x > c^{\mathsf{T}} z_0, \quad \text{and} \quad c^{\mathsf{T}} y < c^{\mathsf{T}} z_0, \ \forall y \in C.$$

Proof. It is clear that $c = x - P_C(x)$ is nonzero since $x \notin C$ and $P_C(x) \in C$. We have

$$c^{\mathsf{T}}(x - z_0) = c^{\mathsf{T}}\left(x - \frac{1}{2}(x + P_C(x))\right) = \frac{1}{2}\|c\|_2^2 > 0.$$

By Theorem 3.2, we have

$$\langle c, y - P_C(x)\rangle \leq 0, \quad \forall y \in C. \tag{13}$$

Meanwhile, we have

$$c^{\mathsf{T}}(P_C(x) - z_0) = c^{\mathsf{T}}\left(P_C(x) - \frac{1}{2}(x + P_C(x))\right) = -\frac{1}{2}\|c\|_2^2 < 0.$$

Combining this with (13) leads to the desired result $c^{\mathsf{T}}y < c^{\mathsf{T}}z_0, \ \forall y \in C$. □

Theorem 3.4. *Let C and D be two closed convex sets in $\mathbb{R}^n$, and let C be bounded. If $C \cap D = \emptyset$, then there exists a hyperplane separating them strictly, i.e., there exist $c \in \mathbb{R}^n \setminus \{0\}$ and $b \in \mathbb{R}$ such that*

$$c^{\mathsf{T}}x - b > 0, \forall x \in C, \quad \text{and} \quad c^{\mathsf{T}}y - b < 0, \forall y \in D.$$

Proof. Since C and D are closed and C is bounded, the set $F = C - D$ is closed and convex. We further know that $0 \notin F$ because $C \cap D = \emptyset$. By Theorem 3.3, there exists $c \in \mathbb{R}^n \setminus \{0\}$ and $b_0 \in \mathbb{R}$ such that

$$0 = \langle c, 0\rangle < b_0, \quad \text{and} \quad c^{\mathsf{T}}(x - y) > b_0, \quad \forall x \in C, y \in D.$$

This implies that

$$\inf_{x\in C} c^{\mathsf{T}}x \geq b_0 + \sup_{y\in D} c^{\mathsf{T}}y > \frac{b_0}{2} + \sup_{y\in D} c^{\mathsf{T}}y > \sup_{y\in D} c^{\mathsf{T}}y.$$

Thus, we can choose $b = \frac{b_0}{2} + \sup_{y\in D} c^{\mathsf{T}}y$. This completes the proof. □

The boundness assumption in Theorem 3.4 cannot be removed in general. The strict separation may not be possible even when C and D are closed. For example, consider

$$C = \left\{(x, y) \in \mathbb{R}^2 \mid x > 0, y \geq \frac{1}{x}\right\}, \ D = \{(x, y) \in \mathbb{R}^2 \mid y \leq 0\},$$

which are closed and disjoint, but they cannot be separated strictly.

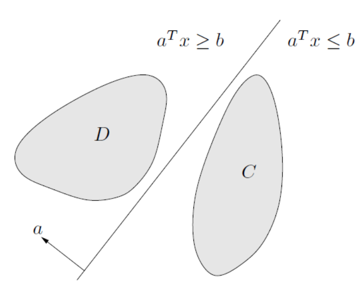


Figure 6: The hyperplane $\{x \mid a^{\mathsf{T}}x = b\}$ separates the disjoint convex sets C and D. The affine function $a^{\mathsf{T}}x - b$ is nonpositive on C and nonnegative on D (images from [2, Figure 2.19]).

This issue is related to the case where the set C in Theorem 3.3 is open and the point $x \notin C$ lies on the boundary of C. In this case, we have a non-strict separation result.

Theorem 3.5. *Let $C \subseteq \mathbb{R}^n$ be a convex set, and $x \notin C$. Then, there exists a nonzero vector $c \in \mathbb{R}^n$ such that*

$$c^{\mathsf{T}}y \leq c^{\mathsf{T}}x, \quad \forall y \in C.$$

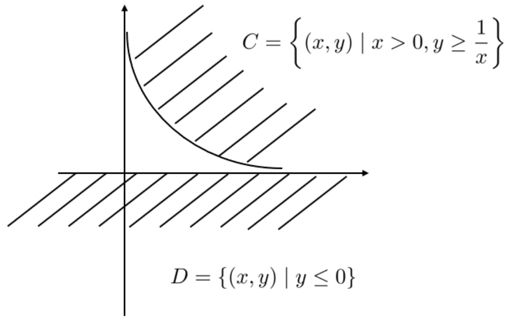


Figure 7: The sets $C = \left\{(x, y) \mid x > 0, y \geq \frac{1}{x}\right\}$, $D = \{(x, y) \mid y \leq 0\}$ are closed and disjoint, but they cannot be separated strictly.

Proof. If x is outside the closure of C, i.e., $\operatorname{cl} C$ (which is closed and convex), then this result follows from Theorem 3.3 (indeed, the strict separation holds).

Assume that x is on the boundary of C. Let us consider a sequence of points $x^1, \ldots, x^k, \ldots$ outside $\operatorname{cl} C$, which converges to x. By Theorem 3.3, for each x^k, there exists a non-zero $c^k \in \mathbb{R}^n$ such that

$$\langle c^k, x^k \rangle > \langle c^k, y \rangle, \qquad \forall y \in C. \tag{14}$$

Without loss of generality, we can assume $\|c^k\|_2 = 1, \forall k \in \mathbb{N}$ (since we can divide $\|c^k\|_2$ on both sides of (14)). Then, the bounded sequence c^k contains a convergent subsequence, and we denote the limit as $c \in \mathbb{R}^n$. Taking such a limit over (14), we get the desired result $c^\mathsf{T} x \geq c^\mathsf{T} y, \forall y \in C$. □

The non-strict separating hyperplane result above holds for two convex sets C, D if they are not intersected. The proof is similar to that of Theorem 3.4.

Theorem 3.6. *Let C and D be two convex sets in $\mathbb{R}^n$. If $C \cap D = \emptyset$, then there exists $c \in \mathbb{R}^n \setminus \{0\}$ such that*

$$c^\mathsf{T} x \geq c^\mathsf{T} y, \quad \forall x \in C, \forall y \in D.$$

In Theorem 3.6, the sets C and D can be both open and unbounded.

3.3 Supporting hyperplanes

Following the proof of Theorem 3.5, for any convex set, we can construct a supporting hyperplane at every boundary point.

Theorem 3.7. *Let $C \subseteq \mathbb{R}^n$ be a non-empty convex set. For any x_0 on the boundary of C, there exists a hyperplane that supports C at x_0, i.e., there exists a non-zero vector $c \in \mathbb{R}^n$ such that $c^\mathsf{T} x \leq c^\mathsf{T} x_0, \forall x \in C$.*

In the next section, we will further show that any closed convex set is the intersection of all its supporting halfspaces (Theorem 3.8).

4 External and internal representations of convex sets

In this section, we present external and internal representations of closed convex sets. The external representation gives an implicit description that allows us to verify whether a point belongs to the convex set. On the other hand, the internal representation gives an explicit description which gives a simple way to generate points in the convex set.

4.1 External (implicit) representation

A direct application of the strict separation result in Theorem 3.3 can show that any closed convex set can be expressed as an intersection of halfspaces (more precisely, its supporting halfspaces).

Theorem 3.8. *Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set. Then, C is an intersection of its supporting halfspaces, i.e.,*

$$C = \bigcap_{x \in \partial C} H_x^-,$$

where H_x^- denotes a supporting halfspace $\{z \in \mathbb{R}^n \mid c^\mathsf{T} z \leq b\}$ of C at point x.

Proof. Let us denote

$$D = \bigcap_{x \in \partial C} H_x^-.$$

Since $C \subseteq H_x^-$ for each $x \in \partial C$, it is clear that $C \subseteq D$.

We only need to prove that $D \subseteq C$. It suffices to prove that if $y \notin C$, then $y \notin D$. Consider any point $y \notin C$. Let $x_0 = P_C(y)$ be the projection of y on C, which is on the boundary of C. Following the argument in Theorem 3.3, the hyperplane $H_{x_0} = \{z \in \mathbb{R}^n \mid c^\mathsf{T} z = c^\mathsf{T} x_0\}$ with normal $c = y - x_0$ is a supporting hyperplane such that

$$c^\mathsf{T} x \leq c^\mathsf{T} x_0, \forall x \in C, \quad \text{and} \quad c^\mathsf{T} y > c^\mathsf{T} x_0.$$

Thus, $y \notin H_{x_0}^-$. By the definition of D, we know $y \notin D$. This completes the proof. □

This theorem gives a representation of a closed convex set C as the intersection of its supporting halfspaces. To verify whether a point x belongs to C, we can check whether x lies in all these supporting halfspaces. If the number of supporting halfspaces is finite, the set C is a polyhedron in the form of (6). A simple membership test for C is to just verify whether all the inequalities and equalities in (6) are satisfied.

4.2 Internal (explicit) representation

We here give an internal representation of compact and convex sets. For this, we define a concept of extreme points.

Definition 3.2 (Extreme point)**.** *Let $C \subseteq \mathbb{R}^n$ be a convex set. A point $x \in C$ is called extreme if there exist no other two distinct points $x_1, x_2 \in C$ and a scalar $0 < \lambda < 1$ such that $x = \lambda x_1 + (1-\lambda) x_2$.*

In other words, an extreme point $x \in C$ does not lie in the line segment between any two distinct points of C. For example, the extreme points of a line segment $\{\theta x + (1-\theta) y \in \mathbb{R}^n \mid 0 \leq \theta \leq 1\}$ are the boundary points x and y. The set of all extreme points of C is denoted as $\operatorname{extr} C$.

Theorem 3.9 (Minkowski's theorem [3, Theorem 2.44])**.** *Let $C \subseteq \mathbb{R}^n$ be a compact and convex set. Then*

$$C = \operatorname{conv}(\operatorname{extr} C). \tag{15}$$

The internal representation (15) gives an explicit way to generate points inside C. This result can be extended to unbounded convex sets; see [3, Theorem 2.46] for details.

5 Cones and dual cones

An important class of convex sets – convex cones plays a significant role in convex optimization, especially conic programming which is a central framework in this course.

Definition 3.3 (Cones)**.** *A set $C \subseteq \mathbb{R}^n$ is called a cone if, for any $x \in C$ and $\theta \geq 0$, we have $\theta x \in C$.*

Definition 3.4 (Convex cones)**.** *A set $C \subseteq \mathbb{R}^n$ is called a convex cone if it is convex and a cone. Equivalently, for any $x_1, x_2 \in C$ and $\theta_1, \theta_2 \geq 0$, we have $\theta_1 x_1 + \theta_2 x_2 \in C$.*

A *conic combination* (or nonnegative linear combination) of points $x_1, x_2, \ldots, x_m$ is a point of the form

$$\alpha_1 x_1 + \cdots + \alpha_m x_m, \quad \text{where } \alpha_i \geq 0, i = 1, \ldots, m.$$

A convex cone contains all conic combinations of its points. The converse is also true (i.e., a set that contains all conic combinations of its points is a convex cone). The *conic hull* of a set $C \subseteq \mathbb{R}^n$, denoted as $\operatorname{cone} C$, is the set of all conic combinations of its points

$$\operatorname{cone} C = \{\alpha_1 x_1 + \cdots + \alpha_m x_m \mid x_i \in C, \alpha_i \geq 0, i = 1, \ldots, m\}. \tag{16}$$

The conic hull of the set C is the smallest convex cone that contains C.

Example 3.5. *Some simple examples of convex cones are 1) any line passing through zero in the form of $\{\theta x_0 \in \mathbb{R}^n \mid \theta \in \mathbb{R}^n\}$ where $x_0 \in \mathbb{R}^n$ is a given point; 2) any hyperplane passing through zero in the form of $\{x \in \mathbb{R}^n \mid c^\mathsf{T} x = 0\}$; 3) any halfspace passing through zero in the form of $\{x \in \mathbb{R}^n \mid c^\mathsf{T} x \leq 0\}$; 4) the span (or linear hull) of a set of vectors $x_i \in \mathbb{R}^n, i = 1, \ldots, m$ (this includes any subspace).*

The strict separating result in Theorem 3.3 can be specialized to convex cones as follows.

Theorem 3.10. *Let $C \subseteq \mathbb{R}^n$ be a closed convex cone and let $x \notin C$. Then, there is a hyperplane that strictly separates $\{x\}$ and C. Even stronger, there exists $c \in \mathbb{R}^n \setminus \{0\}$ such that*

$$c^\mathsf{T} x > 0, \quad \text{and} \quad c^\mathsf{T} y \leq 0, \ \forall y \in C.$$

The proof follows directly by choosing $c = x - P_C(x)$ and applying Theorem 3.2.

5.1 Three important proper cones

For practical optimization methods, we focus on a class of proper cones.

Definition 3.5 (Proper cones)**.** *A cone $K \subseteq \mathbb{R}^n$ is called a proper cone if it is convex, closed, has a non-empty interior (i.e., full dimensional), and pointed (i.e., it contains no line or equivalently $x, -x \in K \Rightarrow x = 0$).*

The convex cones in Example 3.5 are not proper since they are not pointed. Here, we discuss three important classes of proper cones which will be used extensively in this course.

- The *non-negative orthant* is defined as

  $$\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x_i \geq 0, i = 1, \ldots, n\}.$$

  It is convex, closed, has a non-empty interior, and is pointed. Thus, it is a proper cone. It is easy to see that $\mathbb{R}^n_+$ is generated by the standard basis $e_1, \ldots, e_n \in \mathbb{R}^n$

  $$\mathbb{R}^n_+ = \operatorname{cone}\{e_1, \ldots, e_n\}.$$

- The *second-order cone* is defined as

  $$\mathcal{L}^{n+1} = \{(x, t) \in \mathbb{R}^{n+1} \mid \|x\|_2 = (x_1^2 + \cdots + x_n^2)^{1/2} \leq t\}.$$

  This is also a proper cone. Unlike the non-negative orthant $\mathbb{R}^n_+$, the second-order cone $\mathcal{L}^{n+1}$ is not a polyhedron. Very often, $\mathcal{L}^{n+1}$ is also called the ice cream cone or the Lorentz cone. It can be verified that

  $$\mathcal{L}^{n+1} = \operatorname{cone}\{(x, 1) \mid \|x\|_2 = 1\}.$$

- The cone of positive semidefinite matrices

$$\mathbb{S}^n_+ = \{X \in \mathbb{S}^n \mid X \text{ is positive semidefinite}\}.$$

It is convex, closed, pointed, and has a non-empty interior (see Theorem 2.3 in Lecture 2 for characterizations). Theorem 2.3 (2) suggests that the cone of positive semidefinite matrices $\mathbb{S}^n_+$ is generated by rank-1 matrices, i.e.,

$$\mathbb{S}^n_+ = \text{cone}\{xx^\mathsf{T} \mid x \in \mathbb{R}^n, \|x\|_2 = 1\}.$$

A proper convex cone $K \subseteq \mathbb{R}^n$ can define a partial ordering on $\mathbb{R}^n$ as

$$x \succeq_K y \quad \Leftrightarrow \quad x - y \in K.$$

We also write $y \preceq_K x$ for $x \succeq_K y$. If $K = \mathbb{R}^n_+$, the associated partial ordering $x \succeq_K y$ corresponds to component-wise inequalities, i.e., $x_i \geq y_i, i = 1, \ldots, n$. When $K = \mathbb{S}^n_+$, the associated partial ordering $X \succeq_K Y$ means that $X - Y$ is positive semidefinite. Very often, we drop the subscript and simply write $X \succeq Y$.

5.2 Dual cones

The concept of dual cones is also very important in conic programming (Lectures 5 & 6) and its duality analysis (Lectures 7 & 8).

Definition 3.6. *Let $K \subseteq \mathbb{R}^n$ be a cone. The set*

$$K^* = \{y \in \mathbb{R}^n \mid y^\mathsf{T} x \geq 0, \ \ \forall x \in K\}$$

is called the dual cone of K.

The negative of the dual cone $K^\circ = -K^*$ is called the *polar cone*. It is easy to see that the dual cone K^* is indeed a cone (as the name suggests). Further, K^* is always closed and convex since it can be viewed as an intersection of (infinitely) closed halfspaces:

$$K^* = \bigcap_{x \in K} \{y \in \mathbb{R}^n \mid y^\mathsf{T} x \geq 0\}. \tag{17}$$

From (17), it is also easy to see that if $K_1 \subseteq K_2$, then $K_2^* \subseteq K_1^*$.

Geometrically, $y \in K^*$ if and only if $-y$ is the normal of a hyperplane that supports K at the origin. This is shown in Figure 8.

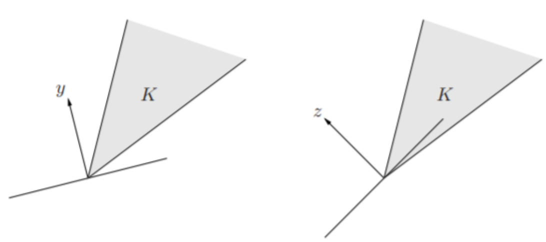


Figure 8: Left: The halfspace with inward normal y contains the cone K, so $y \in K^*$. Right: The halfspace with inward normal z does not contain K, so $z \notin K^*$ (images from [2, Figure 2.22]).

For example, the dual cone of a ray (which is a cone) in the form of $K = \{\theta x_0 \in \mathbb{R}^n \mid \theta \geq 0\}$ is a halfspace $K^* = \{y \in \mathbb{R}^n \mid x_0^\mathsf{T} y \leq 0\}$. The dual cone of a line passing through zero $K = \{\theta x_0 \in \mathbb{R}^n \mid \theta \in \mathbb{R}\}$ is a hyperplane $K^* = \{y \in \mathbb{R}^n \mid x_0^\mathsf{T} y = 0\}$. Generally, the dual cone of a subspace $V \subseteq \mathbb{R}^n_+$ is its orthogonal complement $V^\perp = \{y \in \mathbb{R}^n \mid x^\mathsf{T} y = 0, \forall x \in V\}$.

Let K^{**} denote the dual cone of the dual cone of K, i.e.$K^{**} = (K^*)^*$. The following result is known as the *biduality (or bipolar)* theorem.

Theorem 3.11 ([3, Theorem 2.27]). *Let $K \subseteq \mathbb{R}^n$ be a closed convex cone. Then,*

$$K^{**} = K.$$

The direction $K \subseteq K^{**}$ is based on the definition of dual cones, while the other direction $K^{**} \subseteq K$ requires the machinery of separation theorems, especially Theorem 3.10.

Proof. Let first prove $K \subseteq K^{**}$ by showing that $\forall b \in K \ \Rightarrow \ b \in K^{**}$. By definition, we have $K^* = \bigcap_{x \in K} \{y \in \mathbb{R}^n \mid y^\mathsf{T} x \geq 0\}$. Let b be any point in K. We thus have

$$K^* \subseteq \{y \in \mathbb{R}^n \mid y^\mathsf{T} b \geq 0\},$$

meaning that $y^\mathsf{T} b \geq 0, \forall y \in K^*$. This in turn implies that $b \in (K^*)^*$.

We then prove $K^{**} \subseteq K$ by showing that $\forall b \notin K \ \Rightarrow \ b \notin K^{**}$. Let any $b \notin K$. Since K is a closed convex cone, Theorem 3.10 ensures that there exists $c \in \mathbb{R}^n \setminus \{0\}$ such that

$$c^\mathsf{T} b > 0, \quad \text{and} \quad c^\mathsf{T} y \leq 0, \ \forall y \in K.$$

This means that $-c \in K^*$. Since $(-c)^\mathsf{T} b < 0$, we know that $b \notin K^{**}$ (again recall the definition $K^{**} = \{y \in \mathbb{R}^n \mid y^\mathsf{T} x \geq 0, \forall x \in K^*\}$). □

We call a cone K *self-dual* if $K^* = K$, i.e. the dual cone is itself. All the three classes of proper cones in Section 5.1 are self-dual (note that there are other self-dual cones [1]).

- *Nonnegative orthant.* The cone $\mathbb{R}^n_+$ is self dual $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+$:
$$x^\mathsf{T} y \geq 0, \forall x \in \mathbb{R}^n_+ \ \Leftrightarrow \ y \in \mathbb{R}^n_+. \tag{18}$$
- *Second-order cone.* The second-order cone $\mathcal{L}^{n+1} = \{(x,t) \in \mathbb{R}^{n+1} \mid \|x\|_2 \leq t\}$ is self dual $(\mathcal{L}^{n+1})^* = \mathcal{L}^{n+1}$, i.e.,
$$x^\mathsf{T} u + tv \geq 0, \forall (x,t) \in \mathcal{L}^{n+1} \ \Leftrightarrow \ (u,v) \in \mathcal{L}^{n+1}. \tag{19}$$
- *Positive semidefinite cone.* The positive semidefinite cone $\mathbb{S}^n_+$ is self-dual $(\mathbb{S}^n_+)^* = \mathbb{S}^n_+$, i.e.
$$\text{trace}(XY) \geq 0, \forall X \in \mathbb{S}^n_+ \ \Leftrightarrow \ Y \in \mathcal{S}^n. \tag{20}$$

The fact (18) is straightforward. We leave the proofs of (19) and (20) as exercises. In fact, it can be shown that the dual of a norm cone is the cone defined by the dual norm [2, Example 2.25]: let $\|\cdot\|$ be a norm on $\mathbb{R}^n$, then the dual cone of $K = \{(x,t) \in \mathbb{R}^{n+1} \mid \|x\| \leq t\}$ is

$$K^* = \{(u,v) \in \mathbb{R}^{n+1} \mid \|u\|_* \leq v\}.$$

References

[1] George Philip Barker and James Foran. Self-dual cones in Euclidean spaces. *Linear Algebra and its Applications*, 13(1-2):147–155, 1976.

[2] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization.* Cambridge university press, 2004.

[3] Andrzej Ruszczynski. *Nonlinear optimization.* Princeton university press, 2011.