ENGG 5501: Foundations of Optimization **2018–19 First Term**

Handout 2: Elements of Convex Analysis

Instructor: Anthony Man–Cho So September 10, 2018

As briefly mentioned in Handout 1, the notion of convexity plays a very important role in both the theoretical and algorithmic aspects of optimization. Before we discuss in depth the relevance of convexity in optimization, however, let us first introduce the notions of convex sets and convex functions and study some of their properties.

1 Affine and Convex Sets

1.1 Basic Definitions and Properties

We begin with some definitions.

Definition 1 *Let $S \subseteq \mathbb{R}^n$ be a set. We say that*

1. *S is* **affine** *if $\alpha x + (1-\alpha)y \in S$ whenever $x, y \in S$ and $\alpha \in \mathbb{R}$;*
2. *S is* **convex** *if $\alpha x + (1-\alpha)y \in S$ whenever $x, y \in S$ and $\alpha \in [0,1]$.*

Given $x, y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, the vector $z = \alpha x + (1-\alpha)y$ is called an **affine combination** of x and y. If $\alpha \in [0,1]$, then z is called a **convex combination** of x and y.

Geometrically, when x and y are distinct points in $\mathbb{R}^n$, the set

$$L = \{z \in \mathbb{R}^n : z = \alpha x + (1-\alpha)y,\ \alpha \in \mathbb{R}\}$$

of all affine combinations of x and y is simply the *line* determined by x and y, and the set

$$S = \{z \in \mathbb{R}^n : z = \alpha x + (1-\alpha)y,\ \alpha \in [0,1]\}$$

is the *line segment* between x and y. By convention, the empty set $\emptyset$ is affine and hence also convex.

The notion of an affine (resp. convex) combination of two points can be easily generalized to any finite number of points. In particular, an affine combination of the points $x_1, \ldots, x_k \in \mathbb{R}^n$ is a point of the form $z = \sum_{i=1}^k \alpha_i x_i$, where $\sum_{i=1}^k \alpha_i = 1$. Similarly, a convex combination of the points $x_1, \ldots, x_k \in \mathbb{R}^n$ is a point of the form $z = \sum_{i=1}^k \alpha_i x_i$, where $\sum_{i=1}^k \alpha_i = 1$ *and* $\alpha_1, \ldots, \alpha_k \geq 0$.

The following proposition reveals the structure of a non–empty affine set.

Proposition 1 *Let $S \subseteq \mathbb{R}^n$ be non–empty. Then, the following are equivalent:*

(a) S is affine.

(b) Any affine combination of points in S belongs to S.

(c) S is the translation of some linear *subspace $V \subseteq \mathbb{R}^n$; i.e., S is of the form $\{x\} + V = \{x + v \in \mathbb{R}^n : v \in V\}$ for some $x \in \mathbb{R}^n$.*

Proof We first establish (a)$\Rightarrow$(b) by induction on the number of points $k \geq 1$ in S used to form the affine combination. The case where $k = 1$ is trivial, and the case where $k = 2$ follows from the assumption that S is affine. For the inductive step, let $k \geq 3$ and suppose that $x = \sum_{i=1}^k \alpha_i x_i$, where $x_1, \ldots, x_k \in S$ and $\sum_{i=1}^k \alpha_i = 1$. If $\alpha_i = 0$ for some $i \in \{1, \ldots, k\}$, then x is an affine combination of $k-1$ points in S and hence $x \in S$ by the inductive hypothesis. On the other hand, if $\alpha_i \neq 0$ for all $i \in \{1, \ldots, n\}$, then we may assume without loss of generality that $\alpha_1 \neq 1$ and write

$$x = \alpha_1 x_1 + (1 - \alpha_1) \sum_{i=2}^k \frac{\alpha_i}{1 - \alpha_1} x_i.$$

Since $\sum_{i=2}^k (\alpha_i / (1 - \alpha_1)) = 1$, the point $\bar{x} = \sum_{i=2}^k (\alpha_i / (1 - \alpha_1)) x_i$ is an affine combination of $k-1$ points in S. Hence, we have $\bar{x} \in S$ by the inductive hypothesis. Since S is affine by assumption and x is an affine combination of $x_1 \in S$ and $\bar{x} \in S$, we conclude that $x \in S$, as desired.

Next, we establish (b)$\Rightarrow$(c). Let $x \in S$ and set $V = S - \{x\}$. Our goal is to show that $V \subseteq \mathbb{R}^n$ is a linear subspace. Towards that end, let $v_1, v_2 \in V$ and $\alpha, \beta \in \mathbb{R}$. By definition of V, there exist $z_1, z_2 \in S$ such that $v_1 = z_1 - x$ and $v_2 = z_2 - x$. Using the fact that $z(t) = tz_1 + (1-t)z_2 \in S$ for any $t \in \mathbb{R}$, we have

$$\begin{aligned}
\alpha v_1 + \beta v_2 &= \alpha(\beta + (1-\beta))z_1 + \beta(\alpha + (1-\alpha))z_2 + (1 - \alpha - \beta)x - x \\
&= \alpha((1-\beta)z_1 + \beta z_2) + \beta(\alpha z_1 + (1-\alpha)z_2) + (1 - \alpha - \beta)x - x \\
&= \alpha z(1-\beta) + \beta z(\alpha) + (1 - \alpha - \beta)x - x \\
&= \bar{z} - x,
\end{aligned}$$

where $\bar{z} = \alpha z(1-\beta) + \beta z(\alpha) + (1-\alpha-\beta)x$ is an affine combination of the points $z(1-\beta), z(\alpha), x \in S$ and hence belongs to S. It follows that $\alpha v_1 + \beta v_2 \in V$, as desired.

Lastly, we establish (c)$\Rightarrow$(a). Suppose that $S = \{x\} + V$ for some vector $x \in \mathbb{R}^n$ and linear subspace $V \subseteq \mathbb{R}^n$. Then, for any $z_1, z_2 \in S$, there exist $v_1, v_2 \in V$ such that $z_1 = x + v_1$ and $z_2 = x + v_2$. Since V is linear, for any $\alpha \in \mathbb{R}$, we have $z(\alpha) = \alpha z_1 + (1-\alpha)z_2 = x + v(\alpha)$ with $v(\alpha) = \alpha v_1 + (1-\alpha)v_2 \in V$. This shows that $z(\alpha) \in S$. $\square$

Proposition 1 provides alternative characterizations of an affine set and furnishes examples of affine sets in $\mathbb{R}^n$. Motivated by the equivalence of (a) and (c) in Proposition 1, we shall also refer to an affine set as an **affine subspace**.

Let us now turn our attention to convex sets. In view of Proposition 1, the following proposition should come as no surprise. We leave its proof to the reader.

Proposition 2 *Let $S \subseteq \mathbb{R}^n$ be arbitrary. Then, the following are equivalent:*

(a) *S is convex.*

(b) *Any convex combination of points in S belongs to S.*

Before we proceed, it is helpful to have some concrete examples of convex sets.

Example 1 (Some Examples of Convex Sets)

1. **Non–Negative Orthant***:* $\mathbb{R}_+^n = \{x \in \mathbb{R}^n : x \geq \mathbf{0}\}$

2. **Hyperplane**: $H(s,c) = \{x \in \mathbb{R}^n : s^T x = c\}$

3. **Halfspaces**: $H^-(s,c) = \{x \in \mathbb{R}^n : s^T x \le c\}$, $H^+(s,c) = \{x \in \mathbb{R}^n : s^T x \ge c\}$

4. **Euclidean Ball**: $B(\bar{x}, r) = \{x \in \mathbb{R}^n : \|x - \bar{x}\|_2 \le r\}$

5. **Ellipsoid**: $E(\bar{x}, Q) = \left\{x \in \mathbb{R}^n : (x - \bar{x})^T Q (x - \bar{x}) \le 1\right\}$, *where* Q *is an* $n \times n$ *symmetric, positive definite matrix (i.e.,* $x^T Q x > 0$ *for all* $x \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ *and denoted by* $Q \in \mathcal{S}^n_{++}$*)*

6. **Simplex**: $\Delta = \{\sum_{i=0}^n \alpha_i x_i : \sum_{i=0}^n \alpha_i = 1,\ \alpha_i \ge 0 \text{ for } i = 0, 1, \ldots, n\}$, *where* $x_0, x_1, \ldots, x_n$ *are vectors in* $\mathbb{R}^n$ *such that the vectors* $x_1 - x_0, x_2 - x_0, \ldots, x_n - x_0$ *are linearly independent (equivalently, the vectors* $x_0, x_1, \ldots, x_n$ *are* affinely *independent)*

7. **Convex Cone**: *A set* $K \subseteq \mathbb{R}^n$ *is called a* **cone** *if* $\{\alpha x : \alpha > 0\} \subseteq K$ *whenever* $x \in K$*. If* K *is also convex, then* K *is called a convex cone.*

8. **Positive Semidefinite Cone**: $\mathcal{S}^n_+ = \{Q \in \mathcal{S}^n : x^T Q x \ge 0 \text{ for all } x \in \mathbb{R}^n\}$

The convexity of the sets in Example 1 can be easily established by first principles. We leave this task to the reader.

It is clear from the definition that the intersection of an arbitrary family of affine (resp. convex) sets is an affine (resp. convex) set. Thus, for any arbitrary $S \subseteq \mathbb{R}^n$, there is a smallest (by inclusion) affine (resp. convex) set containing S; namely, the intersection of all affine (resp. convex) sets containing S. This leads us to the following definitions:

Definition 2 *Let* $S \subseteq \mathbb{R}^n$ *be arbitrary.*

1. *The* **affine hull** *of* S*, denoted by* $\mathrm{aff}(S)$*, is the intersection of all affine subspaces containing* S*. In particular,* $\mathrm{aff}(S)$ *is the smallest affine subspace that contains* S*.*

2. *The* **convex hull** *of* S*, denoted by* $\mathrm{conv}(S)$*, is the intersection of all convex sets containing* S*. In particular,* $\mathrm{conv}(S)$ *is the smallest convex set that contains* S*.*

The above definitions can be viewed as characterizing the affine hull and convex hull of a set S from the *outside*. However, given a point in the affine or convex hull of S, it is not immediately clear from the above definitions how it is related to the points in S. This motivates our next proposition, which in some sense provides a characterization of the affine hull and convex hull of S from the *inside*.

Proposition 3 *Let* $S \subseteq \mathbb{R}^n$ *be arbitrary. Then, the following hold:*

(a) $\mathrm{aff}(S)$ *is the set of all affine combinations of points in* S*.*

(b) $\mathrm{conv}(S)$ *is the set of all convex combinations of points in* S*.*

Proof Let us prove (a) and leave the proof of (b) as a straightforward exercise to the reader. Let T be the set of all affine combinations of points in S. Since $S \subseteq \mathrm{aff}(S)$, every $x \in T$ is an affine combination of points in $\mathrm{aff}(S)$. Hence, by Proposition 1, we have $T \subseteq \mathrm{aff}(S)$.

To establish the reverse inclusion, we show that T is an affine subspace containing S. As $\mathrm{aff}(S)$ is the smallest affine subspace that contains S, this would show that $\mathrm{aff}(S) \subseteq T$. To begin, we note that $S \subseteq T$. Thus, it remains to show that T is an affine subspace. Towards that end, let

$x_1, x_2 \in T$. By definition, there exist $y_1, \ldots, y_p, y_{p+1}, \ldots, y_q \in S$ and $\alpha_1, \ldots, \alpha_p, \alpha_{p+1}, \ldots, \alpha_q \in \mathbb{R}$ such that

$$x_1 = \sum_{i=1}^{p} \alpha_i y_i, \quad x_2 = \sum_{i=p+1}^{q} \alpha_i y_i$$

and

$$\sum_{i=1}^{p} \alpha_i = \sum_{i=p+1}^{q} \alpha_i = 1.$$

It follows that for any $\beta \in \mathbb{R}$, we have

$$\bar{x} = \beta x_1 + (1-\beta) x_2 = \sum_{i=1}^{p} \beta \alpha_i y_i + \sum_{i=p+1}^{q} (1-\beta) \alpha_i y_i$$

with

$$\sum_{i=1}^{p} \beta \alpha_i + \sum_{i=p+1}^{q} (1-\beta) \alpha_i = 1.$$

In other words, $\bar{x}$ is an affine combination of points in S. Thus, we have $\bar{x} \in T$, which shows that T is an affine subspace, as desired. $\square$

Since an affine subspace is a translation of a linear subspace and the dimension of a linear subspace is a well–defined notion, we can define the dimension of an affine subspace as the dimension of its underlying linear subspace (see Section 1.5 of Handout B). This, together with the definition of affine hull, makes it possible to define the dimension of an arbitrary set in $\mathbb{R}^n$. Specifically, we have the following definition:

Definition 3 *Let $S \subseteq \mathbb{R}^n$ be arbitrary. The* **dimension** *of S, denoted by $\dim(S)$, is the dimension of the affine hull of S.*

Given a non–empty set $S \subseteq \mathbb{R}^n$, we always have $0 \le \dim(S) \le n$. Roughly speaking, $\dim(S)$ is the intrinsic dimension of S. As we shall see, this quantity plays a fundamental role in optimization. To better understand the notion of the dimension of a set, let us consider the following example:

Example 2 (Dimension of a Set) *Consider the two–point set $S = \{(1,1), (3,2)\} \subseteq \mathbb{R}^2$. By Proposition 3(a), we have $\mathrm{aff}(S) = \{\alpha(1,1) + (1-\alpha)(3,2) : \alpha \in \mathbb{R}\} \subseteq \mathbb{R}^2$. It is easy to verify that $\mathrm{aff}(S) = \{(0,1/2)\} + V$, where $V = \{t(1,1/2) : t \in \mathbb{R}\}$ is the linear subspace generated by the vector $(1,1/2)$. Hence, we have $\dim(S) = \dim(V) = 1$.*

1.2 Convexity–Preserving Operations

Although in theory one can establish the convexity of a set from the definition directly, in practice this may not be the easiest route to take. Moreover, in many occasions, we need to apply certain operation to a collection of convex sets, and we would like to know whether the resulting set is convex. Thus, it is natural to ask which set operations are *convexity–preserving.* We have already seen that the intersection of an arbitrary family of convex sets is convex. Thus, set intersection is convexity–preserving. However, it is easy to see that set union is not. In the following, let us introduce some other convexity–preserving operations.

1.2.1 Affine Functions

We say that a map $A : \mathbb{R}^n \to \mathbb{R}^m$ is **affine** if

$$A(\alpha x_1 + (1-\alpha)x_2) = \alpha A(x_1) + (1-\alpha)A(x_2)$$

for all $x_1, x_2 \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$. It can be shown that A is affine iff there exist $A_0 \in \mathbb{R}^{m\times n}$ and $y_0 \in \mathbb{R}^m$ such that $A(x) = A_0 x + y_0$ for all $x \in \mathbb{R}^n$. As the following proposition shows, convexity is preserved under affine mappings.

Proposition 4 *Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be an affine mapping and $S \subseteq \mathbb{R}^n$ be a convex set. Then, the image $A(S) = \{A(x) \in \mathbb{R}^m : x \in S\}$ is convex. Conversely, if $T \subseteq \mathbb{R}^m$ is a convex set, then the inverse image $A^{-1}(T) = \{x \in \mathbb{R}^n : A(x) \in T\}$ is convex.*

Proof The proposition follows from the easily verifiable fact that for any $x_1, x_2 \in \mathbb{R}^n$, $A([x_1, x_2]) = [A(x_1), A(x_2)] \subseteq \mathbb{R}^m$. $\square$

Example 3 (Convexity of the Ellipsoid) *Consider the ball $B(\mathbf{0}, r) = \left\{x \in \mathbb{R}^n : x^T x \le r^2\right\} \subseteq \mathbb{R}^n$, where $r > 0$. Clearly, $B(\mathbf{0}, r)$ is convex. Now, let Q be an $n\times n$ symmetric positive definite matrix. Then, it is well–known that Q is invertible and the $n\times n$ symmetric matrix Q^{-1} is also positive definite. Moreover, there exists an $n \times n$ symmetric matrix $Q^{-1/2}$ such that $Q^{-1} = Q^{-1/2}Q^{-1/2}$. (See [6, Chapter 7] if you are not familiar with these facts.) Thus, we may define an affine mapping $A : \mathbb{R}^n \to \mathbb{R}^n$ by $A(x) = Q^{-1/2}x + \bar{x}$. We claim that*

$$A(B(\mathbf{0}, r)) = \left\{x \in \mathbb{R}^n : (x-\bar{x})^T Q(x-\bar{x}) \le r^2\right\} = E(\bar{x}, Q/r^2).$$

Indeed, let $x \in B(\mathbf{0}, r)$ and consider the point $A(x)$. We compute

$$(A(x)-\bar{x})^T Q(A(x)-\bar{x}) = x^T Q^{-1/2} Q Q^{-1/2} x = x^T x \le r^2;$$

i.e., $A(B(\mathbf{0}, r)) \subseteq E(\bar{x}, Q/r^2)$. Conversely, let $x \in E(\bar{x}, Q/r^2)$. Consider the point $y = Q^{1/2}(x - \bar{x}) = A^{-1}(x)$. Then, we have $y^T y \le r^2$, which implies that $E(\bar{x}, Q/r^2) \subseteq A(B(\mathbf{0}, r))$. Hence, we conclude from the above calculation and Proposition 4 that $E(\bar{x}, Q/r^2)$ is convex.

1.2.2 Perspective Functions

Define the **perspective function** $P : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}^n$ by $P(x, t) = x/t$. The following proposition shows that convexity is preserved by perspective functions.

Proposition 5 *Let $P : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}^n$ be the perspective function and $S \subseteq \mathbb{R}^n \times \mathbb{R}_{++}$ be a convex set. Then, the image $P(S) = \{x/t \in \mathbb{R}^n : (x, t) \in S\}$ is convex. Conversely, if $T \subseteq \mathbb{R}^n$ is a convex set, then the inverse image $P^{-1}(T) = \{(x, t) \in \mathbb{R}^n \times \mathbb{R}_{++} : x/t \in T\}$ is convex.*

Proof For any $x_1 = (\bar{x}_1, t_1) \in \mathbb{R}^n \times \mathbb{R}_{++}$, $x_2 = (\bar{x}_2, t_2) \in \mathbb{R}^n \times \mathbb{R}_{++}$, and $\alpha \in [0, 1]$, we have

$$P(\alpha x_1 + (1-\alpha)x_2) = \frac{\alpha \bar{x}_1 + (1-\alpha)\bar{x}_2}{\alpha t_1 + (1-\alpha)t_2} = \beta P(x_1) + (1-\beta)P(x_2),$$

where

$$\beta = \frac{\alpha t_1}{\alpha t_1 + (1-\alpha)t_2} \in [0, 1].$$

Moreover, as α increases from 0 to 1, β increases from 0 to 1. It follows that $P([x_1, x_2]) = [P(x_1), P(x_2)] \subseteq \mathbb{R}^n$. This completes the proof. $\square$

The following result, which is a straightforward application of Propositions 4 and 5, shows that convexity is also preserved under *linear–fractional mappings* (also known as *projective mappings*).

Corollary 1 *Let $A : \mathbb{R}^n \to \mathbb{R}^{m+1}$ be the affine map given by*

$$A(x) = \begin{bmatrix} Q \\ c^T \end{bmatrix} x + \begin{bmatrix} u \\ d \end{bmatrix},$$

where $Q \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, and $d \in \mathbb{R}$. Furthermore, let $D = \left\{ x \in \mathbb{R}^n : c^T x + d > 0 \right\}$. Define the **linear–fractional map** *$f : D \to \mathbb{R}^m$ by $f = P \circ A$, where $P : \mathbb{R}^m \times \mathbb{R}_{++} \to \mathbb{R}^m$ is the perspective function. If $S \subseteq D$ is convex, then the image $f(S)$ is convex. Conversely, if $T \subseteq \mathbb{R}^m$ is convex, then the inverse image $f^{-1}(T)$ is convex.*

1.3 Extremal Elements of a Convex Set

Let $S \subseteq \mathbb{R}^n$ be a convex set. By Proposition 3(b), any $x \in S$ can be represented as $x = \sum_{i=1}^k \alpha_i x_i$, where $x_1, \ldots, x_k \in S$, $\sum_{i=1}^k \alpha_i = 1$, and $\alpha_1, \ldots, \alpha_k \geq 0$. However, there is no *a priori* bound on k, the number of points needed. The following theorem of Carathéodory remedies this situation:

Theorem 1 (Carathéodory's Theorem) *Let $S \subseteq \mathbb{R}^n$ be arbitrary. Then, any $x \in \mathrm{conv}(S)$ can be represented as a convex combination of at most $n + 1$ points in S.*

Proof Consider an arbitrary convex combination $x = \sum_{i=1}^k \alpha_i x_i$, with $x_1, \ldots, x_k \in S$ and $k \geq n+2$. The plan is to show that one of the coefficients α_i can be set to 0 without changing x. To begin, observe that since $k \geq n+2$, the vectors $x_2 - x_1, x_3 - x_1, \ldots, x_k - x_1$ must be *linearly dependent* in $\mathbb{R}^n$. In particular, there exist $\beta_1, \ldots, \beta_k \in \mathbb{R}$, not all zero, such that $\sum_{i=1}^k \beta_i x_i = \mathbf{0}$ and $\sum_{i=1}^k \beta_i = 0$. For $i = 1, \ldots, k$, define

$$\alpha_i' = \alpha_i - t^* \beta_i, \qquad \text{where } t^* = \min_{j : \beta_j > 0} \frac{\alpha_j}{\beta_j} = \max\left\{ t \geq 0 : \alpha_i - t\beta_i \geq 0 \ \text{ for } i = 1, \ldots, k \right\}.$$

Note that $t^* < \infty$, since there exists at least one index j such that $\beta_j > 0$. Now, it is straightforward to verify that $x = \sum_{i=1}^k \alpha_i' x_i$, $\sum_{i=1}^k \alpha_i' = 1$, and $\alpha_1', \ldots, \alpha_k' \geq 0$, and that $|\{i : \alpha_i' > 0\}| \leq k - 1$. Now, this process can be repeated until there are only at most $n + 1$ non–zero coefficients. This completes the proof. $\square$

Although Carathéodory's theorem asserts that any $x \in \mathrm{conv}(S)$ can be obtained as a convex combination of at most $n + 1$ points in S, it does not imply the existence of a "basis" of size $n + 1$ for the points in S. In other words, there may not exist a *fixed* set of $n + 1$ points $x_0, x_1, \ldots, x_n$ in S such that $\mathrm{conv}(S) = \mathrm{conv}(\{x_0, x_1, \ldots, x_n\})$ (consider, for instance, the unit disk $B(\mathbf{0}, 1)$). This should be contrasted with the case of linear combinations, where a subspace of dimension n can be generated by n fixed basis vectors.

However, not all is lost. It turns out that under suitable conditions, there exists a *fixed*, though possibly infinite, set S' of points in S such that $S = \mathrm{conv}(S')$. In order to introduce this result, we need a definition.

Definition 4 *Let $S \subseteq \mathbb{R}^n$ be non–empty and convex. We say that $x \in \mathbb{R}^n$ is an* **extreme point** *of S if $x \in S$ and there do not exist two different points $x_1, x_2 \in S$ such that $x = \frac{1}{2}(x_1 + x_2)$.*

The above definition can be rephrased as follows: $x \in S$ is an extreme point of S if $x = x_1 = x_2$ whenever $x_1, x_2 \in S$ are such that $x = \alpha x_1 + (1-\alpha)x_2$ for some $\alpha \in (0,1)$.

Example 4 (Extreme Points of a Ball) *Consider the unit ball $B(\mathbf{0},1)$. We claim that any $x \in B(\mathbf{0},1)$ with $\|x\|_2 = 1$ is an extreme point of $B(\mathbf{0},1)$. Indeed, let $x \in B(\mathbf{0},1)$ be such that $\|x\|_2 = 1$, and let $x_1, x_2 \in B(\mathbf{0},1)$ be such that $x = \frac{1}{2}(x_1 + x_2)$. We compute*

$$1 = \|x\|_2^2 = \frac{1}{4}\|x_1 + x_2\|_2^2 = \frac{1}{2}\|x_1\|_2^2 + \frac{1}{2}\|x_2\|_2^2 - \frac{1}{4}\|x_1 - x_2\|_2^2 \le 1 - \frac{1}{4}\|x_1 - x_2\|_2^2.$$

This shows that $x_1 = x_2$.

We are now ready to state the result alluded to earlier.

Theorem 2 (Minkowski's Theorem) *Let $S \subseteq \mathbb{R}^n$ be compact and convex. Then, S is the convex hull of its extreme points.*

In particular, when combined with Carathéodory's theorem, we see that if $S \subseteq \mathbb{R}^n$ is a compact convex set, then any element in S can be expressed as a convex combination of at most $n+1$ *extreme points* of S. We remark that Theorem 2 is false without the compactness assumption. For instance, consider the set $\mathbb{R}^n_+$.

We shall defer the proof of Theorem 2 to Section 1.6, after we have developed some additional machinery.

Note that extreme points are zero–dimensional objects. By generalizing Definition 4, we can define higher dimensional analogues of extreme points.

Definition 5 *Let $S \subseteq \mathbb{R}^n$ be non–empty and convex. We say that a non–empty convex set $F \subseteq \mathbb{R}^n$ is a* **face** *of S if $F \subseteq S$ and there do not exist two points $x_1, x_2 \in S$ such that (i) $x_1 \in S \setminus F$ or $x_2 \in S \setminus F$, and (ii) $\frac{1}{2}(x_1 + x_2) \in F$.*

In particular, a non–empty convex set $F \subseteq S$ is a face of S if $[x_1, x_2] \subseteq F$ whenever $x_1, x_2 \in S$ and $\alpha x_1 + (1-\alpha)x_2 \in F$ for some $\alpha \in (0,1)$.

The collection of faces of a convex set S contains important geometric information about S and has implications in the design and analysis of optimization algorithms. As an illustration of the notion of face, consider the following example.

Example 5 (Faces of a Simplex) *Consider the simplex*

$$\Delta = \left\{ \sum_{i=0}^{n} \alpha_i e_i : \sum_{i=0}^{n} \alpha_i = 1,\ \alpha_i \ge 0 \ \text{for}\ i = 0, 1, \ldots, n \right\},$$

where $e_i \in \mathbb{R}^n$ is the i–th standard basis vector for $i = 1, \ldots, n$ and $e_0 = \mathbf{0} \in \mathbb{R}^n$. Let I be a non–empty subset of $\{0, 1, \ldots, n\}$ and define

$$\Delta_I = \left\{ \sum_{i \in I} \alpha_i e_i : \sum_{i \in I} \alpha_i = 1,\ \alpha_i \ge 0 \ \text{for}\ i \in I,\ \alpha_i = 0 \ \text{for}\ i \notin I \right\}.$$

Clearly, Δ_I is a non–empty convex subset of Δ. We claim that Δ_I is a face of Δ. Indeed, let $x_1, x_2 \in \Delta$ and suppose that $\beta x_1 + (1-\beta)x_2 \in \Delta_I$ for some $\beta \in (0,1)$. Then, there exist multipliers $\{\alpha_i^1\}_{i=0}^n$, $\{\alpha_i^2\}_{i=0}^n$, and $\{\alpha_i\}_{i=0}^n$ such that

$$\sum_{i=0}^n \alpha_i^1 = \sum_{i=0}^n \alpha_i^2 = \sum_{i \in I} \alpha_i = 1, \quad \alpha_i^1, \alpha_i^2, \alpha_i \geq 0 \text{ for } i = 0, 1, \ldots, n, \quad \alpha_i = 0 \text{ for } i \notin I,$$

$$x_1 = \sum_{i=0}^n \alpha_i^1 e_i, \quad x_2 = \sum_{i=0}^n \alpha_i^2 e_i,$$

and

$$\beta \sum_{i=0}^n \alpha_i^1 e_i + (1-\beta) \sum_{i=0}^n \alpha_i^2 e_i = \sum_{i \in I} \alpha_i e_i. \tag{1}$$

Upon comparing coefficients in (1), we have

$$\beta \alpha_i^1 + (1-\beta)\alpha_i^2 = \alpha_i \qquad \text{for } i = 1, \ldots, n. \tag{2}$$

In particular, we see that $\alpha_i^1 = \alpha_i^2 = 0$ whenever $i \notin I \cup \{0\}$, since $\beta \in (0,1)$ and $\alpha_i^1, \alpha_i^2 \geq 0$ for $i = 0, 1, \ldots, n$. Now, in order to complete the proof of the claim, it suffices to show that $x_1, x_2 \in \Delta_I$. Towards that end, observe that if $0 \in I$, then we have $\alpha_i^1 = \alpha_i^2 = 0$ for $i \notin I$, which implies that $x_1, x_2 \in \Delta_I$. Otherwise, we have $0 \notin I$, and our goal is to show that $\alpha_0^1 = \alpha_0^2 = 0$. To achieve that, we first sum the equations in (2) and use the fact that $\alpha_0 = 0$ to get

$$\beta \sum_{i=1}^n \alpha_i^1 + (1-\beta) \sum_{i=1}^n \alpha_i^2 = \sum_{i=1}^n \alpha_i = 1.$$

Now, upon adding $\beta \alpha_0^1 + (1-\beta)\alpha_0^2$ on both sides, we have

$$1 = 1 + \beta \alpha_0^1 + (1-\beta)\alpha_0^2.$$

Since $\beta \in (0,1)$ and $\alpha_0^1, \alpha_0^2 \geq 0$, this implies that $\alpha_0^1 = \alpha_0^2 = 0$, as desired. In particular, we conclude that Δ_I is a face of Δ.

A useful property of the notion of extreme points, which is captured in the following proposition, is that it is transitive. The result should be intuitively obvious and we leave the proof to the reader.

Proposition 6 *Let $S \subseteq \mathbb{R}^n$ be non–empty and convex. Then, the following hold:*

(a) *If S' is a convex subset of S and $x \in S'$ is an extreme point of S, then x is an extreme point of S'.*

(b) *If $F \subseteq S$ is a face of S, then any extreme point of F is an extreme point of S.*

1.4 Topological Properties

Given an arbitrary set $S \subseteq \mathbb{R}^n$, recall that its **interior** is defined by

$$\text{int}(S) = \{x \in S : B(x, \epsilon) \subseteq S \text{ for some } \epsilon > 0\}.$$

The notion of interior is intimately related to the space in which the set S lies. For instance, consider the set $S = [0,1]$. When viewed as a set in $\mathbb{R}$, then $\text{int}(S) = (0,1)$. However, if we treat S as a set in $\mathbb{R}^2$, then $\text{int}(S) = \emptyset$, since no 2–dimensional ball of positive radius is contained in S. Such ambiguity motivates the following definition.

Definition 6 *Let $S \subseteq \mathbb{R}^n$ be arbitrary. We say that $x \in S$ belongs to the* **relative interior** *of S, denoted by $x \in \operatorname{rel\,int}(S)$, if there exists an $\epsilon > 0$ such that $B(x, \epsilon) \cap \operatorname{aff}(S) \subseteq S$. The* **relative boundary** *of S, denoted by $\operatorname{rel\,bd}(S)$, is defined by $\operatorname{rel\,bd}(S) = \operatorname{cl}(S) \setminus \operatorname{rel\,int}(S)$.*

The following result demonstrates the relevance of the above definition when S is convex.

Theorem 3 *Let $S \subseteq \mathbb{R}^n$ be non–empty and convex. Then, $\operatorname{rel\,int}(S)$ is non–empty.*

Proof Let $k = \dim(S) \geq 0$. Then, S contains $k+1$ affinely independent points $x_0, x_1, \ldots, x_k$, which generate the simplex $\Delta = \operatorname{conv}(\{x_0, \ldots, x_k\})$. Clearly, we have $\Delta \subseteq S$. Moreover, since $\dim(\Delta) = k$, we have $\operatorname{aff}(\Delta) = \operatorname{aff}(S)$. Thus, it suffices to show that $\operatorname{rel\,int}(\Delta)$ is non–empty. Towards that end, let

$$\bar{x} = \frac{1}{k+1} \sum_{i=0}^{k} x_i.$$

Clearly, we have $\bar{x} \in \operatorname{aff}(\Delta) = \operatorname{aff}(S)$. Now, define $V^i = \operatorname{aff}(\{x_0, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k\})$ and let $\epsilon_i = \min_{x \in V^i} \|x - \bar{x}\|_2$. Then, we have $\epsilon_i > 0$ for $i = 0, 1, \ldots, k$. Upon setting $\epsilon = \min_{0 \leq i \leq m} \epsilon_i$, we conclude that $B(\bar{x}, \epsilon) \cap \operatorname{aff}(S) \subseteq S$, as required. $\square$

As one would expect, if we move from a point in the relative interior of a non–empty convex set S to any other point in S, then all the intermediate points should remain in the relative interior of S. Such observation is made precise in the following proposition.

Proposition 7 *Let $S \subseteq \mathbb{R}^n$ be non–empty and convex. For any $x \in \operatorname{cl}(S)$ and $x' \in \operatorname{rel\,int}(S)$, we have*

$$(x, x'] = \big\{ \alpha x + (1-\alpha) x' \in \mathbb{R}^n : \alpha \in [0,1) \big\} \subseteq \operatorname{rel\,int}(S).$$

Proof Let $\alpha \in [0,1)$ be fixed and consider $\bar{x} = \alpha x + (1-\alpha) x'$. Since $x \in \operatorname{cl}(S)$, for any $\epsilon > 0$, we have $x \in S + (B(\mathbf{0}, \epsilon) \cap \operatorname{aff}(S))$. Now, we compute

$$\begin{aligned}
B(\bar{x}, \epsilon) \cap \operatorname{aff}(S) &= \big\{ \alpha x + (1-\alpha) x' \big\} + (B(\mathbf{0}, \epsilon) \cap \operatorname{aff}(S)) \\
&\subseteq \big\{ (1-\alpha) x' \big\} + \alpha S + (1+\alpha)(B(\mathbf{0}, \epsilon) \cap \operatorname{aff}(S)) \\
&= \alpha S + (1-\alpha) \left[B\left(x', \frac{1+\alpha}{1-\alpha} \epsilon \right) \cap \operatorname{aff}(S) \right].
\end{aligned}$$

Since $x' \in \operatorname{rel\,int}(S)$, by taking $\epsilon > 0$ to be sufficiently small, we have

$$B\left(x', \frac{1+\alpha}{1-\alpha} \epsilon \right) \cap \operatorname{aff}(S) \subseteq S.$$

It follows that for sufficiently small $\epsilon > 0$, we have $B(\bar{x}, \epsilon) \cap \operatorname{aff}(S) \subseteq \alpha S + (1-\alpha) S = S$, where the last equality is due to the convexity of S. This implies that $\bar{x} \in \operatorname{rel\,int}(S)$, as desired. $\square$

Another topological concept of interest is that of compactness. Recall that a set $S \subseteq \mathbb{R}^n$ is compact if it is closed and bounded. Moreover, the Weierstrass theorem asserts that if $S \subseteq \mathbb{R}^n$ is a non–empty compact set and $f : S \to \mathbb{R}$ is a continuous function, then f attains its maximum and minimum on S (see, e.g., [10, Chapter 7, Theorem 18] or Section 3.1 of Handout C). Thus, in the context of optimization, it would be good to know when a convex set $S \subseteq \mathbb{R}^n$ is compact. The following proposition provides one sufficient criterion.

Proposition 8 *Let $S \subseteq \mathbb{R}^n$ be compact. Then,* $\text{conv}(S)$ *is compact.*

Proof By Carathéodory's theorem (Theorem 1), for every $x \in \text{conv}(S)$, there exist $x_1, \ldots, x_{n+1} \in S$ and $\alpha = (\alpha_1, \ldots, \alpha_{n+1}) \in \Delta$, where

$$\Delta = \left\{ \alpha \in \mathbb{R}^{n+1} : \sum_{i=1}^{n+1} \alpha_i = 1,\, \alpha \geq \mathbf{0} \right\},$$

such that $x = \sum_{i=1}^{n+1} \alpha_i x_i$. Conversely, it is obvious that any convex combination of $n+1$ points in S belongs to $\text{conv}(S)$. Thus, if we define the continuous function $f : S^{n+1} \times \Delta \to S$ by

$$f(x_1, \ldots, x_{n+1}; \alpha_1, \ldots, \alpha_{n+1}) = \sum_{i=1}^{n+1} \alpha_i x_i,$$

then $\text{conv}(S) = f(S^{n+1} \times \Delta)$. Now, observe that $S^{n+1} \times \Delta$ is compact, as both S and Δ are compact. Since the continuous image of a compact set is compact (see, e.g., [10, Chapter 7, Proposition 24] or Section 3.1 of Handout C), we conclude that $\text{conv}(S)$ is compact, as desired. $\square$

1.5 Projection onto Closed Convex Sets

Given a non–empty set $S \subseteq \mathbb{R}^n$ and a point $x \in \mathbb{R}^n \setminus S$, a natural problem is to find a point in S that is closest, say, in the Euclidean norm, to x. However, it is easy to see that such a point may not exist (e.g., when S is open) or may not be unique (e.g., when S is an arc on a circle and x is the center of the circle). Nevertheless, as the following result shows, the aforementioned difficulties do not arise when S is closed and convex.

Theorem 4 *Let $S \subseteq \mathbb{R}^n$ be non–empty, closed, and convex. Then, for every $x \in \mathbb{R}^n$, there exists a unique point $z^* \in S$ that is closest (in the Euclidean norm) to x.*

Proof Let $x' \in S$ be arbitrary and consider the set $T = \{z \in S : \|x - z\|_2 \leq \|x - x'\|_2\}$. Note that T is compact, as it is closed and bounded. Since the function $z \mapsto \|x - z\|_2^2$ is continuous, we conclude by Weierstrass' theorem that its minimum over the compact set T is attained at some $z^* \in T$. Clearly, z^* is a point in S that is closest to x. This establishes the existence.

Now, let $\mu^* = \|x - z^*\|_2$ and suppose that $z_1, z_2 \in S$ are such that $\mu^* = \|x - z_1\|_2 = \|x - z_2\|_2$. Consider the point $\bar{z} = \frac{1}{2}(z_1 + z_2)$. By Pythagoras' theorem, we have

$$\|\bar{z} - x\|_2^2 = (\mu^*)^2 - \|z_1 - \bar{z}\|_2^2 = (\mu^*)^2 - \frac{1}{4}\|z_1 - z_2\|_2^2.$$

In particular, if $z_1 \neq z_2$, then $\|\bar{z} - x\|_2^2 < (\mu^*)^2$, which is a contradiction. This establishes the uniqueness and completes the proof of the theorem. $\square$

In the sequel, we shall refer to the point z^* in Theorem 4 as the **projection** of x on S and denote it by $\Pi_S(x)$. In other words, we have

$$\Pi_S(x) = \arg\min_{z \in S} \|x - z\|_2^2.$$

The following theorem provides a useful characterization of the projection.

Theorem 5 *Let $S \subseteq \mathbb{R}^n$ be non–empty, closed, and convex. Given any $x \in \mathbb{R}^n$, we have $z^* = \Pi_S(x)$ iff $z^* \in S$ and $(z - z^*)^T(x - z^*) \le 0$ for all $z \in S$.*

Proof Let $z^* = \Pi_S(x)$ and $z \in S$. Consider points of the form $z(\alpha) = \alpha z + (1-\alpha)z^*$, where $\alpha \in [0,1]$. By convexity, we have $z(\alpha) \in S$. Moreover, we have $\|z^* - x\|_2 \le \|z(\alpha) - x\|_2$ for all $\alpha \in [0,1]$. On the other hand, note that

$$\begin{aligned} \|z(\alpha) - x\|_2^2 &= (z^* + \alpha(z - z^*) - x)^T(z^* + \alpha(z - z^*) - x) \\ &= \|z^* - x\|_2^2 + 2\alpha(z - z^*)^T(z^* - x) + \alpha^2\|z - z^*\|_2^2. \end{aligned}$$

Thus, we see that $\|z(\alpha) - x\|_2^2 \ge \|z^* - x\|_2^2$ for all $\alpha \in [0,1]$ iff $(z - z^*)^T(z^* - x) \ge 0$. This is precisely the stated condition.

Conversely, suppose that for some $z' \in S$, we have $(z - z')^T(x - z') \le 0$ for all $z \in S$. Upon setting $z = \Pi_S(x)$, we have

$$(\Pi_S(x) - z')^T(x - z') \le 0. \tag{3}$$

On the other hand, by our argument in the preceding paragraph, the point $\Pi_S(x)$ satisfies

$$(z' - \Pi_S(x))^T(x - \Pi_S(x)) \le 0. \tag{4}$$

Upon adding (3) and (4), we obtain

$$(\Pi_S(x) - z')^T(\Pi_S(x) - z') = \|\Pi_S(x) - z'\|_2^2 \le 0,$$

which is possible only when $z' = \Pi_S(x)$. $\square$

We remark that the projection operator Π_S plays an important role in many optimization algorithms. In particular, the efficiency of those algorithms depends in part on the efficient computability of Π_S. We refer the interested reader to the recent paper [4] for details and further references.

1.6 Separation Theorems

The results in the previous sub–section allow us to establish various separation theorems of convex sets, which are of fundamental importance in convex analysis and optimization. We begin with the following simple yet powerful result.

Theorem 6 *Let $S \subseteq \mathbb{R}^n$ be non–empty, closed, and convex. Furthermore, let $x \in \mathbb{R}^n \setminus S$ be arbitrary. Then, there exists a $y \in \mathbb{R}^n$ such that*

$$\max_{z \in S} y^T z < y^T x.$$

Proof Since S is a non–empty closed convex set, by Theorem 4, there exists a unique point $z^* \in S$ that is closest to x. Set $y = x - z^*$. By Theorem 5, for all $z \in S$, we have $(z - z^*)^T y \le 0$, which implies that

$$y^T z \le y^T z^* = y^T x + y^T(z^* - x) = y^T x - \|y\|_2^2.$$

Since $x \notin S$, we have $y \ne \mathbf{0}$. It follows that

$$\max_{z \in S} y^T z = y^T z^* = y^T x - \|y\|_2^2 < y^T x,$$

as desired. $\square$

To demonstrate the power of Theorem 6, we shall use it to establish three useful results. The first states that every closed convex set can be represented as an intersection of halfspaces that contain it.

Theorem 7 *A closed convex set $S \subseteq \mathbb{R}^n$ is the intersection of all the halfspaces containing S.*

Proof We may assume that $\emptyset \subsetneq S \subsetneq \mathbb{R}^n$, for otherwise the theorem is trivial. Let $x \in \mathbb{R}^n \setminus S$ be arbitrary. Then, by Theorem 6, there exist $y \in \mathbb{R}^n$ and $c = \max_{z \in S} y^T z \in \mathbb{R}$ such that the halfspace $H^-(y,c) = \{z \in \mathbb{R}^n : y^T z \leq c\}$ contains S but not x. It follows that the intersection of all the halfspaces containing S is precisely S itself. $\square$

Theorem 7 suggests that at any boundary point x of a closed convex set S, there should be a hyperplane supporting S at x. To formalize this intuition, we need some definitions.

Definition 7 *Let $S \subseteq \mathbb{R}^n$ be non–empty. We say that the hyperplane $H(s,c)$*

1. **supports** *S if S is contained in one of the halfspaces defined by $H(s,c)$, say, $S \subseteq H^-(s,c)$;*
2. **supports** *S* **at** *$x \in S$ if $H(s,c)$ supports S and $x \in H(s,c)$;*
3. **supports** *S* **non–trivially** *if $H(s,c)$ supports S and $S \not\subseteq H(s,c)$.*

We are now ready to establish the second result, which concerns the existence of supporting hyperplanes to convex sets.

Theorem 8 *Let $S \subseteq \mathbb{R}^n$ be non–empty, closed, and convex. Then, for any $x \in \operatorname{rel}\operatorname{bd}(S)$, there exists a hyperplane supporting S non–trivially at x.*

Proof Let $x \in \operatorname{rel}\operatorname{bd}(S)$ be fixed. Set $S' = S - \{x\}$ and $V = \operatorname{aff}(S) - \{x\}$. Then, S' is a non–empty closed convex set in the Euclidean space V with $\mathbf{0} \in \operatorname{rel}\operatorname{bd}(S')$. Moreover, since $\operatorname{rel}\operatorname{bd}(S) \neq \emptyset$, we have $S \neq \operatorname{aff}(S)$, which implies that $V \neq S'$. Now, consider a sequence $\{x_k\}_{k \geq 1}$ in $V \setminus S'$ such that $x_k \to \mathbf{0}$. By Theorem 6, for $k = 1, 2, \ldots$, there exists a $y_k \in V$ such that $\|y_k\|_2 = 1$ and $y_k^T w < y_k^T x_k$ for all $w \in S'$. The latter is equivalent to

$$y_k^T (z - x) < y_k^T x_k \quad \text{for all } z \in S. \tag{5}$$

Since the sequence $\{y_k\}_{k \geq 1}$ is bounded, by considering a subsequence if necessary, we may assume that $y_k \to y$ with $\|y\|_2 = 1$. It follows from (5) that $y^T z \leq y^T x$ for all $z \in S$; i.e., $S \subseteq H^-(y, y^T x)$. In addition, if $S \subseteq H(y, y^T x)$, or equivalently, $y^T z = y^T x$ for all $z \in S$, then $y^T z = y^T x$ for all $z \in \operatorname{aff}(S)$. This implies that $y \in V^\perp$, which is impossible because $y \in V$ and $y \neq \mathbf{0}$. Hence, we have $S \not\subseteq H(y, y^T x)$. Consequently, we deduce that the hyperplane $H(y, y^T x)$ supports S non–trivially at x, as desired. $\square$

Armed with Theorem 8, we can complete the proof of Minkowski's theorem (Theorem 2).

Proof of Minkowski's Theorem We proceed by induction on $\dim(S)$. The base case, which corresponds to $\dim(S) = 0$, is trivial, because in this case S consists of a single point, which by definition is an extreme point of S. Now, suppose that $\dim(S) = k \geq 1$ and let $x \in S$ be arbitrary. We consider two cases:

Case 1: $x \in \operatorname{rel\,bd}(S)$. By Theorem 8, we can find a hyperplane H supporting S at x. Observe that $H \cap S$ is a compact convex set whose dimension is at most $k-1$. Thus, by the inductive hypothesis, $x \in H \cap S$ can be expressed as a convex combination of extreme points of $H \cap S$. Now, it remains to show that the extreme points of $H \cap S$ are also the extreme points of S. Towards that end, it suffices to show that $H \cap S$ is a face of S and then invoke Proposition 6. Let $x_1, x_2 \in S$ and $\alpha \in (0,1)$ be such that $\alpha x_1 + (1-\alpha)x_2 \in H \cap S$. Suppose that H takes the form $H = \{z \in \mathbb{R}^k : s^T z = s^T x\}$ for some $s \in \mathbb{R}^k \setminus \{\mathbf{0}\}$. Since H supports S, we may assume without loss of generality that $s^T x_1 \leq s^T x$ and $s^T x_2 \leq s^T x$. This, together with the fact that $\alpha x_1 + (1-\alpha)x_2 \in H$, implies that $x_1, x_2 \in H$. It follows that $[x_1, x_2] \subseteq H \cap S$, as desired.

Case 2: $x \in \operatorname{rel\,int}(S)$. Since $\dim(S) \geq 1$, there exists an $x' \in S$ such that $x' \neq x$. Then, Proposition 7 and the compactness of S imply that the line joining x and x' intersects $\operatorname{rel\,bd}(S)$ at two points, say, $y, z \in \operatorname{rel\,bd}(S)$. By the result in Case 1, both y and z can be expressed as a convex combination of the extreme points of S. Since x is a convex combination of y and z, it follows that x can also be expressed as a convex combination of the extreme points of S. □

To motivate the third result, observe that Theorem 6 is a result on point–set separation. A natural question is whether we can derive an analogous result for set–set separation. In other words, given two non–empty and non–intersecting convex sets, is it possible to separate them using a hyperplane? As it turns out, under suitable conditions, we can reduce this question to that of point–set separation. Consequently, we can apply Theorem 6 to obtain the desired separation result.

Theorem 9 *Let $S_1, S_2 \subseteq \mathbb{R}^n$ be non–empty, closed, and convex with $S_1 \cap S_2 = \emptyset$. Furthermore, suppose that S_2 is bounded. Then, there exists a $y \in \mathbb{R}^n$ such that*

$$\max_{z \in S_1} y^T z < \min_{u \in S_2} y^T u.$$

Proof First, note that the set $S_1 - S_2 = \{z - u \in \mathbb{R}^n : z \in S_1,\, u \in S_2\}$ is non–empty and convex. Moreover, we claim that it is closed. To see this, let $x_1, x_2, \ldots$ be a sequence in $S_1 - S_2$ such that $x_k \to x$. We need to show that $x \in S_1 - S_2$. Since $x_k \in S_1 - S_2$, there exist $z_k \in S_1$ and $u_k \in S_2$ such that $x_k = z_k - u_k$ for $k = 1, 2, \ldots$. Since S_2 is compact, there exists a subsequence $\{u_{k_i}\}$ such that $u_{k_i} \to u \in S_2$. Since $x_{k_i} \to x$, we conclude that $z_{k_i} \to x + u$. Since S_1 is closed, we conclude that $x + u \in S_1$. It then follows that $x = (x+u) - u \in S_1 - S_2$, as desired.

We are now in a position to apply Theorem 6 to the non–empty closed convex set $S_1 - S_2$. Indeed, since $S_1 \cap S_2 = \emptyset$, we see that $\mathbf{0} \notin S_1 - S_2$. By Theorem 6, there exist $y \in \mathbb{R}^n$, $z^* \in S_1$, and $u^* \in S_2$ such that

$$y^T(z^* - u^*) = \max_{v \in S_1 - S_2} y^T v < 0.$$

Since S_2 is compact, we have $y^T u^* = \min_{u \in S_2} y^T u$. This implies that $y^T z^* = \max_{z \in S_1} y^T z$. Hence, we obtain

$$\max_{z \in S_1} y^T z < \min_{u \in S_2} y^T u,$$

as desired. □

2 Convex Functions

2.1 Basic Definitions and Properties

Let us now turn to the notion of a convex function.

Definition 8 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be an extended real–valued function that is not identically $+\infty$.*

1. *We say that f is* **convex** *if*
$$f(\alpha x_1 + (1-\alpha)x_2) \le \alpha f(x_1) + (1-\alpha)f(x_2) \tag{6}$$
for all $x_1, x_2 \in \mathbb{R}^n$ and $\alpha \in [0,1]$. We say that f is **concave** *if $-f$ is convex.*

2. *The* **epigraph** *of f is the set $\mathrm{epi}(f) = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : f(x) \le t\}$.*

3. *The* **effective domain** *of f is the set $\mathrm{dom}(f) = \{x \in \mathbb{R}^n : f(x) < +\infty\}$.*

A major advantage of allowing f to take the value $+\infty$ is that we do not need to explicitly specify the domain of f. Indeed, if the function f is defined on a set $S \subseteq \mathbb{R}^n$, then we may simply extend f to $\mathbb{R}^n$ by setting $f(x) = +\infty$ for all $x \notin S$ and obtain $\mathrm{dom}(f) = S$.

The relationship between convex sets and convex functions is explained in the following proposition, whose proof is left as an exercise to the reader.

Proposition 9 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be as in Definition 8. Then, f is convex (as a function) iff $\mathrm{epi}(f)$ is convex (as a set).*

A simple consequence of Proposition 9 is that if f is convex, then $\mathrm{dom}(f)$ is convex. This follows from Proposition 4 and the observation that $\mathrm{dom}(f)$ is the image of the convex set $\mathrm{epi}(f)$ under a projection (which is a linear mapping). Another useful consequence of Proposition 9 is the following inequality.

Corollary 2 (Jensen's Inequality) *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be as in Definition 8. Then, f is convex iff*
$$f\left(\sum_{i=1}^k \alpha_i x_i\right) \le \sum_{i=1}^k \alpha_i f(x_i)$$
for any $x_1, \ldots, x_k \in \mathbb{R}^n$ and $\alpha_1, \ldots, \alpha_k \in [0,1]$ such that $\sum_{i=1}^k \alpha_i = 1$.

Proof The "if" part is clear. To prove the "only if" part, we may assume without loss of generality that $x_1, \ldots, x_k \in \mathrm{dom}(f)$. Then, we have $(x_i, f(x_i)) \in \mathrm{epi}(f)$ for $i = 1, \ldots, k$. Since $\mathrm{epi}(f)$ is convex by Proposition 9, we may invoke Proposition 2 to conclude that
$$\sum_{i=1}^k \alpha_i (x_i, f(x_i)) = \left(\sum_{i=1}^k \alpha_i x_i, \sum_{i=1}^k \alpha_i f(x_i)\right) \in \mathrm{epi}(f).$$
However, this is equivalent to
$$f\left(\sum_{i=1}^k \alpha_i x_i\right) \le \sum_{i=1}^k \alpha_i f(x_i),$$

which completes the proof. $\square$

The epigraph epi(f) of f is closely related to, but not the same as, the t**–level set** $L_t(f)$ of f, where $L_t(f) = \{x \in \mathbb{R}^n : f(x) \leq t\}$ and $t \in \mathbb{R}$ is arbitrary. One obvious difference is that epi(f) is a subset of $\mathbb{R}^n \times \mathbb{R}$, but $L_t(f)$ is a subset of $\mathbb{R}^n$. However, there is another, more subtle, difference: Even if $L_t(f)$ is convex for all $t \in \mathbb{R}$, the function f may not be convex. This can be seen, e.g., from the function $x \mapsto x^3$. A function whose domain is convex and whose t–level sets are convex for all $t \in \mathbb{R}$ is called a **quasi–convex** function. The class of quasi–convex functions possesses nice properties and is important in its own right. We refer the interested reader to [5, 8] for details.

Now, let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function, so that epi(f) is a non–empty convex set by Proposition 9. Suppose in addition that epi(f) is closed. By Theorem 8, we know that epi(f) can be represented as an intersection of all the halfspaces containing it. Such halfspaces take the form $H^-((y, y_0), c) = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : y^T x + y_0 t \leq c\}$ for some $y \in \mathbb{R}^n$ and $y_0, c \in \mathbb{R}$. In fact, it suffices to use only those halfspaces with $y_0 = -1$ in the representation. To see this, suppose that epi$(f) \subseteq H^-((y, y_0), c)$ and consider the following cases:

Case 1: $y_0 \neq 0$. Then, we must have $y_0 < 0$ (since we can fix $x \in \mathrm{dom}(f)$ and take $t \geq f(x)$ to be arbitrarily large), in which case we may assume without loss that $y_0 = -1$ (by scaling y and c).

Case 2: $y_0 = 0$. We claim that there exists another halfspace $H^-((\bar{y}, -1), \bar{c})$ such that epi$(f) \subseteq H^-((\bar{y}, -1), \bar{c})$. Indeed, if this is not the case, then every halfspace containing epi(f) is of the form $H^-((y, 0), c)$ for some $y \in \mathbb{R}^n$ and $c \in \mathbb{R}$. This implies that $f = -\infty$, which is impossible.

Next, we show that for every $(\bar{x}, \bar{t}) \notin H^-((y, 0), c)$, there exists a halfspace $H^-((y', -1), c')$ with $y' \in \mathbb{R}^n$, $c' \in \mathbb{R}$ such that epi$(f) \subseteq H^-((y', -1), c')$ and $(\bar{x}, \bar{t}) \notin H^-((y', -1), c')$. This would imply that halfspaces of the form $H^-((y, 0), c)$ with $y \in \mathbb{R}^n$ and $c \in \mathbb{R}$ are not needed in the representation of epi(f). To begin, observe that for any $(x, t) \in$ epi(f), we have $(x, t) \in H^-((y, 0), c)$ and $(x, t) \in H^-((\bar{y}, -1), \bar{c})$, which means that $y^T x - c \leq 0$ and $\bar{y}^T x - \bar{c} \leq t$. It follows that for any $\lambda \geq 0$,

$$\lambda(y^T x - c) + \bar{y}^T x - \bar{c} \leq t.$$

Moreover, since $y^T \bar{x} - c > 0$, for sufficiently large $\lambda \geq 0$, we have $\lambda(y^T \bar{x} - c) + \bar{y}^T \bar{x} - \bar{c} > \bar{t}$. Thus, by setting $y' = \lambda y + \bar{y}$ and $c' = \lambda c + \bar{c}$, we conclude that epi$(f) \subseteq H^-((y', -1), c')$ and $(\bar{x}, \bar{t}) \notin H^-((y', -1), c')$, as desired.

It is not hard to see that the halfspace $H^-((y, -1), c)$ is the epigraph of the affine function $h : \mathbb{R}^n \to \mathbb{R}$ given by $h(x) = y^T x - c$. Moreover, if epi$(f) \subseteq H^-((y, -1), c)$, then $h(x) \leq f(x)$ for all $x \in \mathbb{R}^n$. (This is obvious for $x \notin \mathrm{dom}(f)$. For $x \in \mathrm{dom}(f)$, note that $(x, f(x)) \in$ epi$(f) \subseteq H^-((y, -1), c)$ implies $y^T x - f(x) \leq c$, or equivalently, $h(x) \leq f(x)$.) Since the intersection of halfspaces yields the epigraph of the pointwise supremum of the affine functions induced by those halfspaces, we obtain the following theorem:

Theorem 10 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function such that* epi(f) *is closed. Then, f can be represented as the pointwise supremum of all affine functions $h : \mathbb{R}^n \to \mathbb{R}$ satisfying $h \leq f$.*

Motivated by Theorem 10, given a convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, we consider the set

$$S_f = \left\{(y, c) \in \mathbb{R}^n \times \mathbb{R} : y^T x - c \leq f(x) \text{ for all } x \in \mathbb{R}^n\right\},$$

which consists of the coefficients of those affine functions $h : \mathbb{R}^n \to \mathbb{R}$ satisfying $h \leq f$. Clearly, we have $y^T x - c \leq f(x)$ for all $x \in \mathbb{R}^n$ iff $\sup_{x \in \mathbb{R}^n} \left\{y^T x - f(x)\right\} \leq c$. This shows that S_f is the epigraph of the function $f^* : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ given by

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \left\{ y^T x - f(x) \right\}.$$

Moreover, observe that S_f is closed and convex. This implies that f^* is convex. The function f^* is called the **conjugate** of f. It plays a very important role in convex analysis and optimization. We shall study it in greater detail later.

2.2 Convexity–Preserving Transformations

As in the case of convex sets, it is sometimes difficult to check directly from the definition whether a given function is convex or not. In this sub–section we describe some transformations that preserve convexity.

Theorem 11 *The following hold:*

(a) **(Non–Negative Combinations)** *Let $f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex functions satisfying $\cap_{i=1}^m \mathrm{dom}(f_i) \neq \emptyset$. Then, for any $\alpha_1, \ldots, \alpha_m \geq 0$, the function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by*

$$f(x) = \sum_{i=1}^m \alpha_i f_i(x)$$

is convex.

(b) **(Pointwise Supremum)** *Let I be an index set and $\{f_i\}_{i \in I}$, where $f_i : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ for all $i \in I$, be a family of convex functions. Define the* **pointwise supremum** *$f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ of $\{f_i\}_{i \in I}$ by*

$$f(x) = \sup_{i \in I} f_i(x).$$

Suppose that $\mathrm{dom}(f) \neq \emptyset$. Then, the function f is convex.

(c) **(Affine Composition)** *Let $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function and $A : \mathbb{R}^m \to \mathbb{R}^n$ be an affine mapping. Suppose that $\mathrm{range}(A) \cap \mathrm{dom}(g) \neq \emptyset$. Then, the function $f : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ defined by $f(x) = g(A(x))$ is convex.*

(d) **(Composition with an Increasing Convex Function)** *Let $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $h : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be convex functions that are not identically $+\infty$. Suppose that h is increasing on $\mathrm{dom}(h)$. Define the function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ by $f(x) = h(g(x))$, with the convention that $h(+\infty) = +\infty$. Suppose that $\mathrm{dom}(f) \neq \emptyset$. Then, the function f is convex.*

(e) **(Restriction on Lines)** *Given a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ that is not identically $+\infty$, a point $x_0 \in \mathbb{R}^n$, and a direction $h \in \mathbb{R}^n$, define the function $\tilde{f}_{x_0,h} : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ by $\tilde{f}_{x_0,h}(t) = f(x_0 + th)$. Then, the function f is convex iff the function $\tilde{f}_{x_0,h}$ is convex for any $x_0 \in \mathbb{R}^n$ and $h \in \mathbb{R}^n$.*

The above results can be derived directly from the definition. We shall prove (e) and leave the rest as exercises to the reader.

Proof of Theorem 11(e) Suppose that f is convex. Let $x_0, h \in \mathbb{R}^n$ be arbitrary. Then, for any $t_1, t_2 \in \mathbb{R}$ and $\alpha \in [0,1]$, we have

$$\begin{aligned}
\tilde{f}_{x_0,h}(\alpha t_1 + (1-\alpha)t_2) &= f(x_0 + (\alpha t_1 + (1-\alpha)t_2)h) \\
&= f(\alpha(x_0 + t_1 h) + (1-\alpha)(x_0 + t_2 h)) \\
&\le \alpha f(x_0 + t_1 h) + (1-\alpha) f(x_0 + t_2 h) \\
&= \alpha \tilde{f}_{x_0,h}(t_1) + (1-\alpha)\tilde{f}_{x_0,h}(t_2);
\end{aligned}$$

i.e., $\tilde{f}_{x_0,h}$ is convex. Conversely, suppose that $\tilde{f}_{x_0,h}$ is convex for any $x_0, h \in \mathbb{R}^n$. Let $x_1, x_2 \in S$ and $\alpha \in [0,1]$. Upon setting $x_0 = x_1 \in \mathbb{R}^n$ and $h = x_2 - x_1 \in \mathbb{R}^n$, we have

$$\begin{aligned}
f((1-\alpha)x_1 + \alpha x_2) &= \tilde{f}_{x_0,h}(\alpha) \\
&= \tilde{f}_{x_0,h}(\alpha \cdot 1 + (1-\alpha)\cdot 0) \\
&\le \alpha \tilde{f}_{x_0,h}(1) + (1-\alpha)\tilde{f}_{x_0,h}(0) \\
&= \alpha f(x_2) + (1-\alpha) f(x_1);
\end{aligned}$$

i.e., f is convex. This completes the proof. $\square$

2.3 Differentiable Convex Functions

When a given function is differentiable, it is possible to characterize its convexity via its derivatives. We begin with the following result, which makes use of the gradient of the given function.

Theorem 12 *Let $f : \Omega \to \mathbb{R}$ be a differentiable function on the open set $\Omega \subseteq \mathbb{R}^n$ and $S \subseteq \Omega$ be a convex set. Then, f is convex on S (i.e., the inequality (6) holds for all $x_1, x_2 \in S$ and $\alpha \in [0,1]$) iff*

$$f(x) \ge f(\bar{x}) + (\nabla f(\bar{x}))^T (x - \bar{x})$$

for all $x, \bar{x} \in S$.

To appreciate the geometric content of the above theorem, observe that $x \mapsto f(\bar{x}) + (\nabla f(\bar{x}))^T (x - \bar{x})$ is an affine function whose level sets are hyperplanes with normal $\nabla f(\bar{x})$ and takes the value $f(\bar{x})$ at $\bar{x}$. Thus, Theorem 12 stipulates that at every $\bar{x} \in S$, the function f is minorized by an affine function that coincides with f at $\bar{x}$.

Proof Suppose that f is convex on S. Let $x, \bar{x} \in S$ and $\alpha \in (0,1)$. Then, we have

$$f(x) \ge \frac{f(\alpha x + (1-\alpha)\bar{x}) - (1-\alpha) f(\bar{x})}{\alpha} = f(\bar{x}) + \frac{f(\bar{x} + \alpha(x - \bar{x})) - f(\bar{x})}{\alpha}. \tag{7}$$

Now, recall that

$$\lim_{\alpha \searrow 0} \frac{f(\bar{x} + \alpha(x - \bar{x})) - f(\bar{x})}{\alpha}$$

is the directional derivative of f at $\bar{x}$ in the direction $x - \bar{x}$ and is equal to $(\nabla f(\bar{x}))^T (x - \bar{x})$ (see Section 3.2.2 of Handout C). Hence, upon letting $\alpha \searrow 0$ in (7), we have

$$f(x) \ge f(\bar{x}) + (\nabla f(\bar{x}))^T (x - \bar{x}),$$

as desired.

Conversely, let $x_1, x_2 \in S$ and $\alpha \in (0,1)$. Then, we have $\bar{x} = \alpha x_1 + (1-\alpha)x_2 \in S$, which implies that

$$f(x_1) \geq f(\bar{x}) + (1-\alpha)(\nabla f(\bar{x}))^T(x_1 - x_2), \tag{8}$$

$$f(x_2) \geq f(\bar{x}) + \alpha(\nabla f(\bar{x}))^T(x_2 - x_1). \tag{9}$$

Upon multiplying (8) by α and (9) by $1-\alpha$ and summing, we obtain the desired result. □

In the case where f is twice continuously differentiable, we have the following characterization:

Theorem 13 *Let $f : S \to \mathbb{R}$ be a twice continuously differentiable function on the open convex set $S \subseteq \mathbb{R}^n$. Then, f is convex on S iff $\nabla^2 f(x) \succeq \mathbf{0}$ for all $x \in S$.*

Proof Suppose that $\nabla^2 f(x) \succeq \mathbf{0}$ for all $x \in S$. Let $x_1, x_2 \in S$. By Taylor's theorem, there exists an $\bar{x} \in [x_1, x_2] \subseteq S$ such that

$$f(x_2) = f(x_1) + (\nabla f(x_1))^T(x_2 - x_1) + \frac{1}{2}(x_2 - x_1)^T \nabla^2 f(\bar{x})(x_2 - x_1). \tag{10}$$

Since $\nabla^2 f(\bar{x}) \succeq \mathbf{0}$, we have $(x_2 - x_1)^T \nabla^2 f(\bar{x})(x_2 - x_1) \geq 0$. Upon substituting this inequality into (10) and invoking Theorem 12, we conclude that f is convex on S.

Conversely, suppose that $\nabla^2 f(\bar{x}) \not\succeq \mathbf{0}$ for some $\bar{x} \in S$. Then, there exists a $v \in \mathbb{R}^n$ such that $v^T \nabla^2 f(\bar{x}) v < 0$. Since S is open and $\nabla^2 f$ is continuous, there exists an $\epsilon > 0$ such that $\bar{x}' = \bar{x} + \epsilon v \in S$ and $v^T \nabla^2 f(\bar{x} + \alpha(\bar{x}' - \bar{x}))v < 0$ for all $\alpha \in [0,1]$. This implies (by taking $x_1 = \bar{x}$ and $x_2 = \bar{x}'$ in (10)) that $f(\bar{x}') < f(\bar{x}) + (\nabla f(\bar{x}))^T(\bar{x}' - \bar{x})$. Hence, by Theorem 12, we conclude that f is not convex on S. This completes the proof. □

We point out that Theorem 13 only applies to functions f that are twice continuously differentiable on an *open* convex set S. This should be contrasted with Theorem 12, where the function f needs only be differentiable on a *convex subset* S' of an open set. In particular, the set S' need not be open. To see why S must be open in Theorem 13, consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x,y) = x^2 - y^2$. This function is convex on the set $S = \mathbb{R} \times \{0\}$. However, its Hessian, which is given by

$$\nabla^2 f(x,y) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \qquad \text{for } (x,y) \in \mathbb{R}^2,$$

is nowhere positive semidefinite.

2.4 Establishing Convexity of Functions

Armed with the tools developed in previous sub–sections, we are already able to establish the convexity of many functions. Here are some examples.

Example 6 (Some Examples of Convex Functions)

1. *Let $f : \mathbb{R}^n \times \mathcal{S}^n_{++} \to \mathbb{R}$ be given by $f(x,Y) = x^T Y^{-1} x$. We compute*

$$\begin{aligned} \mathrm{epi}(f) &= \left\{ (x,Y,r) \in \mathbb{R}^n \times \mathcal{S}^n_{++} \times \mathbb{R} : Y \succ \mathbf{0},\ x^T Y^{-1} x \leq r \right\} \\ &= \left\{ (x,Y,r) \in \mathbb{R}^n \times \mathcal{S}^n_{++} \times \mathbb{R} : \begin{bmatrix} Y & x \\ x^T & r \end{bmatrix} \succeq \mathbf{0},\ Y \succ \mathbf{0} \right\}, \end{aligned}$$

where the last equality follows from the Schur complement (see, e.g., [2, Section A.5.5]). This shows that $\mathrm{epi}(f)$ *is a convex set, which implies that* f *is convex on* $\mathbb{R}^n \times \mathcal{S}^n_{++}$.

2. *Let* $f : \mathbb{R}^{m\times n} \to \mathbb{R}_+$ *be given by* $f(X) = \|X\|_2$, *where* $\|\cdot\|_2$ *denotes the* spectral norm *or* largest singular value *of the* $m \times n$ *matrix* X. *It is well known that (see, e.g., [6])*

$$f(X) = \sup\left\{u^T X v : \|u\|_2 = 1,\ \|v\|_2 = 1\right\}.$$

This shows that f *is a pointwise supremum of a family of linear functions of* X. *Hence,* f *is convex.*

3. *Let* $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_+$ *be a norm on* $\mathbb{R}^n$ *and* $f : \mathbb{R}^n \to \mathbb{R}_+$ *be given by* $f(x) = \|x\|^p$, *where* $p \ge 1$. *Then, for any* $x \in \mathbb{R}^n$, *we have* $f(x) = g(\|x\|)$, *where* $g : \mathbb{R}_+ \to \mathbb{R}_+$ *is given by* $g(z) = z^p$. *Clearly,* g *is increasing on* $\mathbb{R}_+$ *and convex on* $\mathbb{R}_{++}$ *(e.g., by verifying* $g''(z) \ge 0$ *for all* $z > 0$*). To show that* g *is convex on* $\mathbb{R}_+$, *it remains to verify that for any* $z \in \mathbb{R}_+$ *and* $\alpha \in [0, 1]$,

$$g(\alpha \cdot 0 + (1-\alpha)z) = (1-\alpha)^p z^p \le (1-\alpha)z^p = \alpha g(0) + (1-\alpha)g(z).$$

Hence, by Theorem 11(d), we conclude that f *is convex.*

4. *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be given by* $f(x) = \log\left(\sum_{i=1}^n \exp(x_i)\right)$. *We compute*

$$\frac{\partial^2 f}{\partial x_i x_j} = \begin{cases} \dfrac{\exp(x_i)}{\sum_{i=1}^n \exp(x_i)} - \dfrac{\exp(2x_i)}{\left(\sum_{i=1}^n \exp(x_i)\right)^2} & \text{if } i = j, \\[2ex] -\dfrac{\exp(x_i + x_j)}{\left(\sum_{i=1}^n \exp(x_i)\right)^2} & \text{if } i \ne j. \end{cases}$$

This gives

$$\nabla^2 f(x) = \frac{1}{(e^T z)^2}\left(\left(e^T z\right) \mathit{diag}(z) - zz^T\right),$$

where $z = (\exp(x_1), \ldots, \exp(x_n))$. *Now, for any* $v \in \mathbb{R}^n$, *we have*

$$\begin{aligned} v^T \nabla^2 f(x) v &= \frac{1}{(e^T z)^2}\left[\left(\sum_{i=1}^n z_i\right)\left(\sum_{i=1}^n z_i v_i^2\right) - \left(\sum_{i=1}^n z_i v_i\right)^2\right] \\ &= \frac{1}{(e^T z)^2}\left[\left(\sum_{i=1}^n (\sqrt{z_i})^2\right)\left(\sum_{i=1}^n (\sqrt{z_i} v_i)^2\right) - \left(\sum_{i=1}^n \sqrt{z_i} \cdot (\sqrt{z_i} v_i)\right)^2\right] \\ &\ge 0 \end{aligned}$$

by the Cauchy–Schwarz inequality. Hence, f *is convex.*

5. *([2, Chapter 3, Exercise 3.17]) Suppose that* $p \in (0, 1)$. *Let* $f : \mathbb{R}^n_{++} \to \mathbb{R}$ *be given by* $f(x) = \left(\sum_{i=1}^n x_i^p\right)^{1/p}$. *We compute*

$$\frac{\partial^2 f}{\partial x_i x_j} = \begin{cases} (1-p)\left(\sum_{i=1}^n x_i^p\right)^{p^{-1}-2}\left[-\left(\sum_{i=1}^n x_i^p\right) x_i^{p-2} + x_i^{2(p-1)}\right] & \text{if } i = j, \\[2ex] (1-p)\left(\sum_{i=1}^n x_i^p\right)^{p^{-1}-2} x_i^{p-1} x_j^{p-1} & \text{if } i \ne j. \end{cases}$$

This gives

$$\nabla^2 f(x) = (1-p)\left(\sum_{i=1}^n x_i^p\right)^{p^{-1}-2}\left[-\left(\sum_{i=1}^n x_i^p\right)\mathit{diag}\left(x_1^{p-2},\ldots,x_n^{p-2}\right)+zz^T\right],$$

where $z_i = x_i^{p-1}$ *for* $i=1,\ldots,n$. *Now, for any* $v \in \mathbb{R}^n$, *we have*

$$v^T\nabla^2 f(x)v = (1-p)\left(\sum_{i=1}^n x_i^p\right)^{p^{-1}-2}\left[-\left(\sum_{i=1}^n x_i^p\right)\left(\sum_{i=1}^n v_i^2 x_i^{p-2}\right)+\left(\sum_{i=1}^n v_i x_i^{p-1}\right)^2\right] \le 0,$$

since

$$-\left(\sum_{i=1}^n x_i^p\right)\left(\sum_{i=1}^n v_i^2 x_i^{p-2}\right)+\left(\sum_{i=1}^n \left(v_i x_i^{(p-2)/2}\right)\left(x_i^{p/2}\right)\right)^2 \le 0$$

by the Cauchy–Schwarz inequality. It follows that f *is concave on* $\mathbb{R}^n_{++}$.

6. *Let* $f : \mathcal{S}^n_{++} \to \mathbb{R}$ *be given by* $f(X) = -\ln\det X$. *For those readers who are well versed in matrix calculus (see, e.g., [7] for a comprehensive treatment), the following formulas should be familiar:*

$$\nabla f(X) = -X^{-1}, \quad \nabla^2 f(X) = X^{-1}\otimes X^{-1}.$$

Here, $\otimes$ *denotes the Kronecker product. Since* $X^{-1} \succ \mathbf{0}$, *it can be shown that* $X^{-1}\otimes X^{-1} \succ \mathbf{0}$. *It follows that* f *is convex on* $\mathcal{S}^n_{++}$.

Alternatively, we can establish the convexity of f *on* $\mathcal{S}^n_{++}$ *by applying Theorem 11(e). To begin, let* $X_0 \in \mathcal{S}^n_{++}$ *and* $H \in \mathcal{S}^n$. *Define the set* $\mathcal{D} = \{t \in \mathbb{R} : X_0 + tH \succ \mathbf{0}\} = \{t \in \mathbb{R} : \lambda_{\min}(X_0+tH) > 0\}$. *Since* $\lambda_{\min}$ *is continuous (see, e.g., [12, Chapter IV, Theorem 4.11]), we see that* $\mathcal{D}$ *is open and convex. Now, consider the function* $\tilde{f}_{X_0,H} : \mathcal{D} \to \mathbb{R}$ *given by* $\tilde{f}_{X_0,H}(t) = f(X_0+tH)$. *For any* $t \in \mathcal{D}$, *we compute*

$$\begin{aligned}\tilde{f}_{X_0,H}(t) &= -\ln\det(X_0+tH)\\ &= -\ln\det\left(X_0^{1/2}\left(I+tX_0^{-1/2}HX_0^{-1/2}\right)X_0^{1/2}\right)\\ &= -\left(\sum_{i=1}^n \ln(1+t\lambda_i)+\ln\det X_0\right)\end{aligned}$$

and

$$\tilde{f}''_{X_0,H}(t) = \sum_{i=1}^n \frac{\lambda_i^2}{(1+t\lambda_i)^2} \ge 0,$$

where $\lambda_1,\ldots,\lambda_n$ *are the eigenvalues of* $X_0^{-1/2}HX_0^{-1/2}$. *It follows that* $\tilde{f}_{X_0,H}$ *is convex on* $\mathcal{D}$. *This, together with Theorem 11(e), implies that* f *is convex on* $\mathcal{S}^n_{++}$.

2.5 Non–Differentiable Convex Functions

In previous sub–sections we developed techniques to check whether a differentiable convex function is convex or not. In particular, we showed that a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex iff at every $\bar{x} \in \mathbb{R}^n$, we have $f(x) \geq f(\bar{x}) + (\nabla f(\bar{x}))^T(x - \bar{x})$. The latter condition has a geometric interpretation, namely, the epigraph of f is supported by its tangent hyperplane at every $(\bar{x}, f(\bar{x})) \in \mathbb{R}^n \times \mathbb{R}$. Of course, as can be easily seen from the function $x \mapsto |x|$, not every convex function is differentiable. However, the above geometric interpretation seems to hold for such functions as well. In order to make such an observation rigorous, we need to first generalize the notion of a gradient to that of a *subgradient*.

Definition 9 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be as in Definition 8. A vector $s \in \mathbb{R}^n$ is called a* **subgradient** *of f at $\bar{x}$ if*

$$f(x) \geq f(\bar{x}) + s^T(x - \bar{x}) \qquad \text{for all } x \in \mathbb{R}^n. \tag{11}$$

The set of vectors s such that (11) is called the **subdifferential** *of f at $\bar{x}$ and is denoted by $\partial f(\bar{x})$.*

Note that $\partial f(\bar{x})$ may not be empty even though f is not differentiable at $\bar{x}$. For example, consider the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = |x|$. Clearly, f is not differentiable at the origin, but it can be easily verified from the definition that $\partial f(0) = [-1, 1]$.

In many ways, the subdifferential behaves like a derivative, although one should note that the former is a set, while the latter is a unique element. Here we state some important properties of the subdifferential without proof. For details, we refer the reader to [11, Chapter 2, Section 2.5].

Theorem 14 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function that is not identically $+\infty$.*

(a) **(Subgradient and Directional Derivative)** *Let*

$$f'(x, d) = \lim_{t \searrow 0} \frac{f(x + td) - f(x)}{t}$$

be the directional derivative of f at $x \in \mathbb{R}^n$ in the direction $d \in \mathbb{R}^n \backslash \{\mathbf{0}\}$, and let $x \in \operatorname{int}\operatorname{dom}(f)$. Then, $\partial f(x)$ is a non–empty compact convex set. Moreover, for any $d \in \mathbb{R}^n$, we have $f'(x, d) = \max_{s \in \partial f(x)} s^T d$.

(b) **(Subdifferential of a Differentiable Function)** *The convex function f is differentiable at $x \in \mathbb{R}^n$ iff the subdifferential $\partial f(x)$ is a singleton, in which case it consists of the gradient of f at x.*

(c) **(Additivity of Subdifferentials)** *Suppose that $f = f_1 + f_2$, where $f_1 : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $f_2 : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ are convex functions that are not identically $+\infty$. Furthermore, suppose that there exists an $x_0 \in \operatorname{dom}(f)$ such that f_1 is continuous at x_0. Then, we have*

$$\partial f(x) = \partial f_1(x) + \partial f_2(x) \qquad \text{for all } x \in \operatorname{dom}(f).$$

Theorem 14(a) states that the subdifferential ∂f of a convex function f is non–empty at any $x \in \operatorname{int}\operatorname{dom}(f)$. This leads to the natural question of what happens when $x \in \operatorname{bd}\operatorname{dom}(f)$. The following example shows that ∂f can be empty at such points.

Example 7 (A Convex Function with Empty Subdifferential at a Point) *Let* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *be given by*

$$f(x) = \begin{cases} -\sqrt{1 - \|x\|_2^2} & \text{if } \|x\|_2 \leq 1, \\ +\infty & \text{otherwise.} \end{cases}$$

It is clear that f *is a convex function with* $\mathrm{dom}(f) = B(\mathbf{0}, 1)$ *and is differentiable on* $\mathrm{int}\,\mathrm{dom}(f) = \{x \in \mathbb{R}^n : \|x\|_2 < 1\}$*. It is also clear that* $\partial f(x) = \emptyset$ *for all* $x \notin \mathrm{dom}(f)$*. Now, let* $\bar{x} \in \mathrm{bd}\,\mathrm{dom}(f) = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ *be arbitrary. If* $s \in \partial f(\bar{x})$*, then we must have*

$$f(x) = -\sqrt{1 - \|x\|_2^2} \geq s^T(x - \bar{x}) \quad \text{for all } x \in B(\mathbf{0}, 1). \tag{12}$$

Without loss of generality, we may assume that $\bar{x} = e_1$*. For* $\alpha \in [-1, 1]$*, define* $x(\alpha) = \alpha e_1 \in B(\mathbf{0}, 1)$*. From* (12)*, we see that* $s \in \mathbb{R}^n$ *satisfies*

$$f(x(\alpha)) = -\sqrt{1 - \alpha^2} \geq \alpha s_1 - s_1 \quad \text{for all } \alpha \in [-1, 1].$$

However, this implies that $s_1 \geq \sqrt{(1 + \alpha)/(1 - \alpha)}$ *for all* $\alpha \in [-1, 1)$*, which is impossible. Hence, we have* $\partial f(\bar{x}) = \emptyset$ *for all* $\bar{x} \in \mathrm{bd}\,\mathrm{dom}(f)$*.*

To further illustrate the concepts and results discussed above, let us compute the subdifferential of the Euclidean norm.

Example 8 (Subdifferential of the Euclidean Norm) *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be given by* $f(x) = \|x\|_2$*. Note that* f *is differentiable whenever* $x \neq \mathbf{0}$*. Hence, by Theorem 14(b) we have* $\partial f(x) = \{\nabla f(x)\} = \{x/\|x\|_2\}$ *for all* $x \in \mathbb{R}^n \setminus \{\mathbf{0}\}$*.*

Now, recall from Definition 9 that $s \in \mathbb{R}^n$ *is a subgradient of* f *at* $\mathbf{0}$ *iff* $\|x\|_2 \geq s^T x$ *for all* $x \in \mathbb{R}^n$*. It follows that* $\partial f(\mathbf{0}) = B(\mathbf{0}, 1)$*.*

3 Further Reading

Convex analysis is a rich subject with many deep and beautiful results. For more details, one can consult the general references listed on the course website and/or the books [3, 13, 9, 5, 1, 11].

References

[1] A. Barvinok. *A Course in Convexity*, volume 54 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, Rhode Island, 2002.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2004. Available online at `http://www.stanford.edu/~boyd/cvxbook/`.

[3] A. Brøndsted. *An Introduction to Convex Polytopes*, volume 90 of *Graduate Texts in Mathematics*. Springer–Verlag, New York, 1983.

[4] C. Ding, D. Sun, and K.-C. Toh. An Introduction to a Class of Matrix Cone Programming. *Mathematical Programming, Series A*, 144(1–2):141–179, 2014.

[5] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of Convex Analysis*. Grundlehren Text Editions. Springer–Verlag, Berlin/Heidelberg, 2001.

[6] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1985.

[7] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Chichester, England, revised edition, 1999.

[8] E. A. Papa Quiroz, L. Mallma Ramirez, and P. R. Oliveira. An Inexact Proximal Method for Quasiconvex Minimization. *European Journal of Operational Research*, 246(3):721–729, 2015.

[9] R. T. Rockafellar. *Convex Analysis*. Princeton Landmarks in Mathematics and Physics. Princeton University Press, Princeton, New Jersey, 1997.

[10] H. L. Royden. *Real Analysis*. Macmillan Publishing Company, New York, third edition, 1988.

[11] A. Ruszczyński. *Nonlinear Optimization*. Princeton University Press, Princeton, New Jersey, 2006.

[12] G. W. Stewart and J. Sun. *Matrix Perturbation Theory*. Academic Press, Boston, 1990.

[13] G. M. Ziegler. *Lectures on Polytopes*, volume 152 of *Graduate Texts in Mathematics*. Springer–Verlag, New York, revised first edition, 1995.