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 Convex Sets

1.1 Basic Definitions and Properties

We begin with some definitions.

**Definition 1** Let $S \subset \mathbb{R}^n$ 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 convex.

It is clear that one can generalize the notion of affine (resp. convex) combination of two points 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 $z = \sum_{i=1}^{k} \alpha_i x_i$, 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 $z = \sum_{i=1}^{k} \alpha_i x_i$, where $\sum_{i=1}^{k} \alpha_i = 1$ and $\alpha_1, \ldots, \alpha_k \geq 0$.

Here are some sets in Euclidean space whose convexity can be easily established by first principles:

**Example 1** (Some Examples of Convex Sets)

1. **Non-Negative Orthant**: $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x \geq 0\}$
2. **Hyperplane**: $H(s, c) = \{x \in \mathbb{R}^n : s^T x = c\}$
3. **Halfspaces**: $H^+(s, c) = \{x \in \mathbb{R}^n : s^T x \leq c\}, H^-(s, c) = \{x \in \mathbb{R}^n : s^T x \geq c\}$
4. **Euclidean Ball**: $B(\bar{x}, r) = \{x \in \mathbb{R}^n : \|x - \bar{x}\|_2 \leq r\}$
5. **Ellipsoid:** $E(\bar{x}, Q, r) = \{x \in \mathbb{R}^n : (x - \bar{x})^T Q (x - \bar{x}) \leq r^2\}$, where $Q$ is an $n \times n$ symmetric, positive definite matrix (i.e., $x^T Q x > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$)

6. **Simplex:** $\Delta = \{\sum_{i=0}^n \alpha_i x_i : \sum_{i=0}^n \alpha_i = 1, \alpha_i \geq 0$ 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. **Positive Semidefinite Cone:** $S^n_+ = \{X \in \mathbb{R}^{n \times n} : X \text{ symmetric and positive semidefinite}\}$

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

Although in theory one can always establish the convexity of a set from the definition directly, such an approach could potentially be quite difficult. 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, we are led to the following question: which set operations are convexity-preserving? Clearly, set intersection is convexity-preserving, while set union is not. In other words, the intersection of an arbitrary family of convex sets is convex. However, the union of two convex sets need not be convex. Now, let us prove that convexity is also preserved by affine mappings. To begin, we say that a function $A : \mathbb{R}^n \to \mathbb{R}^m$ 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 if there exist an $m \times n$ matrix $A_0$ and a $y_0 \in \mathbb{R}^m$ such that $A(x) = A_0 x + y_0$ for all $x \in \mathbb{R}^n$. We then have the following theorem:

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

**Proof** Let $x_1, x_2 \in \mathbb{R}^n$. Then, we have $A([x_1, x_2]) = [A(x_1), A(x_2)] \subset \mathbb{R}^m$. This completes the proof.

As an application of Theorem 1, consider the following example:

**Example 2** Consider the ball $B(0, r) = \{x \in \mathbb{R}^n : x^T x \leq r^2\} \subset \mathbb{R}^n$, where $r > 0$. Clearly, $B(0, r)$ 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

$$A(B(0, r)) = \{x \in \mathbb{R}^n : (x - \bar{x})^T Q (x - \bar{x}) \leq r^2\} \equiv E(\bar{x}, Q, r).$$

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

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

i.e., $A(B(0, r)) \subset E(\bar{x}, Q, r)$. Conversely, let $x \in E(\bar{x}, Q, r)$. Consider the point $y = Q^{1/2} (x - \bar{x}) = A^{-1}(x)$. Then, we have $y^T y \leq r^2$, which implies that $E(\bar{x}, Q, r) \subset A(B(0, r))$. Hence, we conclude from the above calculation and Theorem 1 that $E(\bar{x}, Q, r)$ is convex.
1.2 Projection onto Closed Convex Sets

Let $S \subset \mathbb{R}^n$ be a non-empty closed convex set, and let $x \in \mathbb{R}^n$ be arbitrary. We call the point in $S$ that is closest (in the Euclidean norm) to $x$ the \textit{projection} of $x$ on $S$ and denote it by $\Pi_S(x)$. One of the many nice things about convex sets is the following projection property:

\textbf{Theorem 2} Let $S \subset \mathbb{R}^n$ be a non-empty closed convex set. Then, for every $x \in \mathbb{R}^n$, there exists a unique point $z^* \in S$ that is closest to $x$.

\textbf{Proof} Let $x' \in S$ be arbitrary, and consider the set $T = \{z \in S : \|x - z\|_2 \leq \|x - x'\|_2\}$. Then, $T$ is compact, since it is closed and bounded. Now, since the function $z \mapsto \|x - z\|_2$ is continuous, we conclude by Weierstrass’ theorem that its minimum over the compact set $T$ is attained at some $z^* \in T$. 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

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

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

The above theorem establishes the existence and uniqueness of the projection $\Pi_S(x)$. However, it would be nice if we can have some condition that checks whether a given $z \in S$ actually equals $\Pi_S(x)$. The following theorem supplies such a condition:

\textbf{Theorem 3} Let $S \subset \mathbb{R}^n$ be a non-empty closed convex set, and let $x \in \mathbb{R}^n$. Then, $z^* = \Pi_S(x)$ iff $z^* \in S$ and $(z - z^*)^T (x - z^*) \leq 0$ for all $z \in S$.

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

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

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

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

$$
(\Pi_S(x) - z')^T (x - z') \leq 0. \tag{1}
$$

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

$$
(z' - \Pi_S(x))^T (x - \Pi_S(x)) \leq 0. \tag{2}
$$

Upon adding (1) and (2), we obtain

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

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

The projection operator $\Pi_S$ plays an important role in many optimization algorithms. In particular, the efficiency of those algorithms depends in part on the efficient computability of $\Pi_S$. We refer the interested reader to the recent paper [4] for details and further references.
1.3 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.

**Theorem 4** Let $S \subset \mathbb{R}^n$ be a non-empty closed convex set, and let $x \in \mathbb{R}^n$ be such that $x \notin S$. Then, there exists a $y \in \mathbb{R}^n$ such that

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

**Proof** Since $S$ is a non-empty closed convex set, by Theorem 2, there exists a point $z^* \in S$ that is closest to $x$. By Theorem 3, we have $(z - z^*)^T(x - z^*) \leq 0$ for all $z \in S$. Set $y = x - z^*$. Note that $y \neq 0$, since $x \notin S$. Now, for each $z \in S$, since $(z - z^*)^T y \leq 0$, we have

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

This completes the proof. \qed

We can easily generalize the above theorem to the following:

**Corollary 1** Let $S_1, S_2 \subset \mathbb{R}^n$ be two non-empty closed convex sets with $S_1 \cap S_2 = \emptyset$. If $S_2$ is bounded, then there exists a $y \in \mathbb{R}^n$ such that

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

**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 4 to the non-empty closed convex set $S_1 - S_2$. Indeed, since $S_1 \cap S_2 = \emptyset$, we see that $0 \notin S_1 - S_2$. By Theorem 4, there exists a $y \in \mathbb{R}^n$ such that $\sup_{z \in S_1 - S_2} y^T z < 0$. Equivalently, we have

$$0 > \sup_{z \in S_1, u \in S_2} y^T (z - u) = \sup_{z \in S_1} y^T z + \sup_{u \in S_2} (-y^T u) = \sup_{z \in S_1} y^T z - \inf_{u \in S_2} y^T u.$$  

Since $S_2$ is compact, we can replace the inf by min. This completes the proof. \qed

Note that the closedness assumption in Theorem 4 is crucial. If the set $S$ is not closed, then we may not have strict separation. However, as the following theorem demonstrates, not all is lost.

**Theorem 5** Let $S \subset \mathbb{R}^n$ be a non-empty convex set, and let $x \in \mathbb{R}^n$ be such that $x \notin S$. Then, there exists a non-zero $y \in \mathbb{R}^n$ such that $\sup_{z \in S} y^T z \leq y^T x$.

**Proof** Let $x_1, x_2, \ldots$ be a sequence in $\mathbb{R}^n$ such that $x_k \to x$, with $x_k \notin S$, where $\bar{S}$ is the closure of $S$. By Theorem 4, there exists a $y_k \in \mathbb{R}^n$ such that $y_k^T z < y_k^T x_k$ for all $z \in S$. In particular, we have $y_k \neq 0$, so we may assume without loss that $\|y_k\|_2 = 1$. Since $\{y_k\}$ is a sequence on the compact set $\{x \in \mathbb{R}^n : \|x\|_2 = 1\}$, there exists a subsequence $\{y_{k_i}\}$ such that $y_{k_i} \to y$, with $\|y\|_2 = 1$. It follows that by taking limit over the subsequence, we have $y^T z \leq y^T x$ for all $z \in S$, as desired. \qed

Using a similar argument as in the proof of Corollary 1, we have the following:
Corollary 2 Let \( S_1, S_2 \subseteq \mathbb{R}^n \) be two non-empty convex sets with \( S_1 \cap S_2 = \emptyset \). Then, there exists a non-zero \( y \in \mathbb{R}^n \) such that
\[
\sup_{z \in S_1} y^T z \leq \inf_{u \in S_2} y^T u.
\]

1.4 Cones and Their Dual

In Example 1 we introduced the notion of a cone, which is defined as a set \( K \) that contains the open half-line \( \{\alpha x : \alpha > 0\} \) whenever \( x \in K \). An important object associated with a cone is its dual cone:

Definition 2 Let \( K \subseteq \mathbb{R}^n \) be a cone. The dual cone of \( K \) is defined as the set
\[
K^* = \{ y \in \mathbb{R}^n : y^T x \geq 0 \text{ for all } x \in K \}.
\]

The cone \( K^0 \equiv -K^* \) is usually known as the polar of \( K \).

As we shall see later, dual cones play an important role in optimization. Note that by definition, \( K^* \) is always closed and convex, regardless of whether \( K \) has those properties or not. Furthermore, we have the following proposition:

Proposition 1 Let \( K \subseteq \mathbb{R}^n \) be a non-empty closed convex cone. Then, we have \( K = K^{**} \).

**Proof** It is clear that \( K \subseteq K^{**} \). To establish the converse, let \( v \in K^{**} \). If \( v \notin K \), then by Theorem 4, there exists a \( y \in \mathbb{R}^n \) such that \( \inf_{x \in K} y^T x > y^T v \). We claim that \( \theta^* \equiv \inf_{x \in K} y^T x = 0 \). Clearly, we have \( \theta^* \leq 0 \), since \( 0 \in K \). Now, if \( \theta^* < 0 \), then there exists an \( x' \in K \) such that \( 0 > y^T x' > y^T v \). However, since \( \alpha x' \in K \) for all \( \alpha > 0 \), we see that \( \alpha y^T x' > y^T v \) for all \( \alpha \geq 1 \), which is impossible. Thus, the claim is established. In particular, this shows that \( y \in K^* \). However, we then have the inequality \( 0 > y^T v \), which contradicts the fact that \( v \in K^{**} \). Hence, we conclude that \( v \in K \). \( \square \)

Given a cone \( K \subseteq \mathbb{R}^n \), a natural question is to determine whether a given vector belongs to the dual cone \( K^* \). The following proposition gives one such criterion:

Proposition 2 Let \( K \subseteq \mathbb{R}^n \) be a cone, and let \( y \in \mathbb{R}^n \) be given. Suppose that there exists a constant \( B \in \mathbb{R} \) such that \( y^T x \geq B \) for all \( x \in K \). Then, we have \( y \in K^* \).

**Proof** Suppose that \( y \notin K^* \). Then, there exists an \( x \in K \) such that \( y^T x < 0 \). Since \( K \) is a cone, we have \( \alpha x \in K \) for all \( \alpha > 0 \). However, \( y^T (\alpha x) = \alpha y^T x \) cannot be bounded below for all \( \alpha > 0 \). This gives the desired contradiction. \( \square \)

In many cases, the dual cone can be explicitly computed. Here are two examples.

Example 3 Consider the non-negative orthant \( \mathbb{R}_+^n = \{ x \in \mathbb{R}^n : x \geq 0 \} \). Its dual is given by \( (\mathbb{R}_+^n)^* = \{ y \in \mathbb{R}^n : y^T x \geq 0 \text{ for all } x \in \mathbb{R}_+^n \} \). It is clear that \( \mathbb{R}_+^n \subseteq (\mathbb{R}_+^n)^* \), because \( y^T x \geq 0 \) if \( x, y \geq 0 \). Now, suppose that \( y \in (\mathbb{R}_+^n)^* \). Then, we have \( x^T y \geq 0 \) for all \( x \in \mathbb{R}_+^n \). In particular, we have \( e_i^T y = y_i \geq 0 \). This shows that \( y \geq 0 \); i.e., \( y \in \mathbb{R}_+^n \). Hence, we have \( \mathbb{R}_+^n = (\mathbb{R}_+^n)^* \).

Example 4 Let \( C \subseteq \mathbb{R}^n \) be a closed convex cone and \( A \in \mathbb{R}^{m \times n} \). Consider the set \( K = \{ x \in \mathbb{R}^n : Ax \in C \} \). We claim that \( K \) is a closed convex cone, and that \( K^* = \{ A^T y : y \in C^* \} \). The first statement follows directly from the closedness and convexity of \( C \). To prove the second statement, let \( y \in C^* \) and \( x \in K \). Then, since \( Ax \in C \), we have \( (A^T y)^T x = y^T Ax \geq 0 \), thus showing that
\{A^T y : y \in C^*\} \subset K^*. Now, suppose that there exists an \(u \in K^* \setminus \{A^T y : y \in C^*\}\). Observe that the set \(\{A^T y : y \in C^*\}\) is closed and convex, since \(C^*\) is closed and convex. Thus, by Theorem 4, there exists a \(z \in \mathbb{R}^n\) and \(\epsilon > 0\) such that

\[u^T z + \epsilon \leq z^T A^T y \quad \text{for all } y \in C^*.
\]

By Proposition 2 (taking \(K = C^*\) and \(B = u^T z + \epsilon\)), we conclude that \(Az \in C^{**} = C\), which implies that \(z \in K\). Now, since \(u \in K^*\), we have \(u^T z \geq 0\). On the other hand, since \(0 \in C^*\), inequality (3) implies that \(0 < \epsilon \leq u^T z + \epsilon \leq 0\), which is a contradiction. Thus, we have \(K^* = \{A^T y : y \in C^*\}\), as desired.

## 2 Convex Functions

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

**Definition 3** Let \(f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}\) be an extended-valued function that is not identically \(+\infty\).

1. We say that \(f\) is convex if

\[1f(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha f(x_1) + (1 - \alpha)f(x_2)
\]

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

3. The epigraph of \(f\) is the set \(\text{epi}(f) = \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq r\}\).

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 3** Let \(f\) be as in Definition 3. Then, \(f\) is convex (as a function) iff \(\text{epi}(f)\) is convex (as a set in \(\mathbb{R}^n \times \mathbb{R}\)).

Let \(r \in \mathbb{R}\) be arbitrary. A set closely related to the epigraph is the so-called \(r\)-level set of \(f\), which is defined as \(L(r) = \{x \in \mathbb{R}^n : f(x) \leq r\}\). It is clear that if \(f\) is convex, then \(L(r)\) is convex for all \(r \in \mathbb{R}\). However, the converse is not true. For instance, consider the function \(f : \mathbb{R} \to \mathbb{R}\) given by \(f(x) = x^3\). A function \(f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}\) whose domain is convex and whose \(r\)-level sets are convex for all \(r \in \mathbb{R}\) is called quasi-convex.

Sometimes we may want to restrict the domain of a function so that it is convex on that domain. This motivates the following definition:

**Definition 4** Let \(S \subset \mathbb{R}^n\) be a non-empty convex set. A function \(f : S \to \mathbb{R}\) is said to be convex on \(S\) if (4) holds for all \(x_1, x_2 \in S\) and \(\alpha \in [0, 1]\).

For instance, the function \(x \mapsto x^3\) is convex on \(\mathbb{R}_+\). We remark that Definitions 3 and 4 are in fact equivalent. Given the function \(f : S \to \mathbb{R}\), we simply extend it to \(\tilde{f} : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}\) by

\[
\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in S, \\ +\infty & \text{if } x \notin S. \end{cases}
\]

One can then check that \(\tilde{f}\) is convex in the sense of Definition 3 iff \(f\) is convex on \(S\) (i.e., in the sense of Definition 4).

One of the most desirable features of convex functions is the following:
Proposition 4 Consider the optimization problem

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad x \in S,
\end{align*}
\]

where \(S \subseteq \mathbb{R}^n\) is a convex set and \(f : \mathbb{R}^n \to \mathbb{R}\) is convex. Then, any local minimizer of \(f\) is also a global minimizer.

We leave the proof as an exercise to the reader.

2.1 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. We shall restrict our attention to finite-valued functions, although the following results can be generalized to extended-valued functions.

Theorem 6 The following hold:

1. (Non-Negative Combinations) Let \(f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}\) be convex functions, and let \(\alpha_1, \ldots, \alpha_m \geq 0\). Then, the function \(\sum_{i=1}^{m} \alpha_i f_i\) is convex.

2. (Pointwise Supremum) Let \(\{f_i\}_{i \in I}\) be an arbitrary family of convex functions on \(\mathbb{R}^n\). Then, the pointwise supremum \(f = \sup_{i \in I} f_i\) is convex.

3. (Affine Composition) Let \(f : \mathbb{R}^n \to \mathbb{R}\) be a convex function and \(A : \mathbb{R}^m \to \mathbb{R}^n\) be an affine mapping. Then, the function \(f \circ A : \mathbb{R}^m \to \mathbb{R}^n\) given by \((f \circ A)(x) = f(A(x))\) is convex on \(\mathbb{R}^m\).

4. (Composition with an Increasing Convex Function) Let \(f : \mathbb{R}^n \to \mathbb{R}\) be a convex function, and let \(g : \mathbb{R} \to \mathbb{R}\) be an increasing convex function. Then, the function \(g \circ f : \mathbb{R}^n \to \mathbb{R}\) defined by \((g \circ f)(x) = g(f(x))\) is convex on \(\mathbb{R}^n\).

5. (Restriction on Lines) Let \(S \subseteq \mathbb{R}^n\) be a convex set and \(f : S \to \mathbb{R}\) be a function. Given \(x_0 \in S\) and \(h \in \mathbb{R}^n\), let \(\tilde{S}_{x_0, h} = \{t \in \mathbb{R} : x_0 + th \in S\}\) and define the function \(\tilde{f}_{x_0, h} : \tilde{S}_{x_0, h} \to \mathbb{R}\) by \(\tilde{f}_{x_0, h}(t) = f(x_0 + th)\). Then, \(\tilde{S}_{x_0, h}\) is convex. Moreover, \(f\) is convex on \(S\) iff \(\tilde{f}_{x_0, h}\) is convex on \(\tilde{S}_{x_0, h}\) for any \(x_0 \in S\) and \(h \in \mathbb{R}^n\).

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

Proof We prove item 5. The convexity of \(\tilde{S}_{x_0, h}\) follows from the convexity of \(S\). Suppose that \(f\) is convex on \(S\). Let \(x_0 \in S\) and \(h \in \mathbb{R}^n\) be arbitrary, and let \(\alpha \in [0, 1]\). Then, for \(t_1, t_2 \in \tilde{S}_{x_0, h}\), we have

\[
\begin{align*}
\tilde{f}_{x_0, h}(\alpha t_1 + (1 - \alpha)t_2) & = f(x_0 + (\alpha t_1 + (1 - \alpha)t_2)h) \\
& = f(\alpha(x_0 + t_1 h) + (1 - \alpha)(x_0 + t_2 h)) \\
& \leq \alpha f(x_0 + t_1 h) + (1 - \alpha)f(x_0 + t_2 h) \\
& = \alpha \tilde{f}_{x_0, h}(t_1) + (1 - \alpha)\tilde{f}_{x_0, h}(t_2);
\end{align*}
\]
\[ f((1 - \alpha)x_1 + \alpha x_2) = \tilde{f}_{x_0,h}(\alpha) \]
\[ = \tilde{f}_{x_0,h}(\alpha \cdot 1 + (1 - \alpha) \cdot 0) \]
\[ \leq \alpha \tilde{f}_{x_0,h}(1) + (1 - \alpha) \tilde{f}_{x_0,h}(0) \]
\[ = \alpha f(x_2) + (1 - \alpha)f(x_1); \]

i.e., \( f \) is convex on \( S \). This completes the proof. \( \square \)

2.2 Differentiable Convex Functions

When \( f \) is a differentiable function, we can characterize its convexity via its gradient.

**Theorem 7** Let \( f : \Omega \to \mathbb{R} \) be a differentiable function on the open set \( \Omega \subset \mathbb{R}^n \), and let \( S \subset \Omega \) be convex. Then, \( f \) is convex on \( S \) iff

\[ f(x_1) \geq f(x_2) + (\nabla f(x_2))^T(x_1 - x_2) \]

for all \( x_1, x_2 \in S \).

**Proof** Suppose that \( f \) is convex on \( S \). Let \( x_1, x_2 \in S \) and \( \alpha \in (0, 1) \). Then, we have

\[ f(x_1) \geq \frac{f(\alpha x_1 + (1 - \alpha)x_2) - (1 - \alpha)f(x_2)}{\alpha} = f(x_2) + \frac{f(x_2 + \alpha(x_1 - x_2)) - f(x_2)}{\alpha}, \quad (5) \]

Now, recall that

\[ \lim_{\alpha \searrow 0} \frac{f(x_2 + \alpha(x_1 - x_2)) - f(x_2)}{\alpha} \]

is the directional derivative of \( f \) at \( x_2 \) in the direction \( x_1 - x_2 \), and is equal to \((\nabla f(x_2))^T(x_1 - x_2)\) (see Section 3.2.2 of Handout C). Hence, upon letting \( \alpha \searrow 0 \) in (5), we have

\[ f(x_1) \geq f(x_2) + (\nabla f(x_2))^T(x_1 - x_2), \]

as desired.

Conversely, let \( x_1, x_2 \in S \) and \( \alpha \in (0, 1) \). Then, we have \( \alpha x_1 + (1 - \alpha)x_2 \in S \), which implies that

\[ f(x_1) \geq f(\alpha x_1 + (1 - \alpha)x_2) + (1 - \alpha)(\nabla f(\alpha x_1 + (1 - \alpha)x_2))^T(x_1 - x_2), \quad (6) \]
\[ f(x_2) \geq f(\alpha x_1 + (1 - \alpha)x_2) + \alpha(\nabla f(\alpha x_1 + (1 - \alpha)x_2))^T(x_2 - x_1). \quad (7) \]

Upon multiplying (6) by \( \alpha \) and (7) by \( 1 - \alpha \) and summing, we obtain the desired result. \( \square \)

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

**Theorem 8** Let \( f : S \to \mathbb{R} \) be a twice continuously differentiable function on the open convex set \( S \subset \mathbb{R}^n \). Then, \( f \) is convex on \( S \) iff \( \nabla^2 f(\bar{x}) \) is positive semidefinite for all \( \bar{x} \in S \).
Proof Suppose that \( f \) is twice continuously differentiable on \( S \), and let \( x_1, x_2 \in S \). Then, by Taylor’s theorem, we have

\[
f(x_2) = f(x_1) + (\nabla f(x_1))^T(x_2 - x_1) + \frac{1}{2}(x_2 - x_1)^T\nabla^2 f(x(\alpha))(x_2 - x_1),
\]

(8)

where \( x(\alpha) = x_1 + \alpha(x_2 - x_1) \in S \) for some \( \alpha \in [0, 1] \). If \( \nabla^2 f(\bar{x}) \) is positive semidefinite for all \( \bar{x} \in S \), then \( (x_2 - x_1)^T\nabla^2 f(x(\alpha))(x_2 - x_1) \geq 0 \), which, together with Theorem 7, implies that \( f \) is convex.

Conversely, suppose that \( \nabla^2 f \) is not positive semidefinite on \( S \). Then, there exists a \( v \in \mathbb{R}^n \) such that \( v^T\nabla^2 f(\bar{x})v < 0 \). Consider the point \( x' = \bar{x} + \epsilon v \in S \) for some \( \epsilon > 0 \). Now, for sufficiently small \( \epsilon > 0 \), the point \( \bar{x} + \epsilon av = \bar{x} + \alpha(x' - \bar{x}) \) will be close to \( \bar{x} \) for any \( \alpha \in [0, 1] \). Since \( \nabla^2 f(\cdot) \) is continuous, we have \( v^T\nabla^2 f(\bar{x} + \epsilon av)v < 0 \) for all \( \alpha \in [0, 1] \). It follows from (8) that

\[
f(x') < f(\bar{x}) + (\nabla f(\bar{x}))^T(x' - \bar{x}).
\]

This, together with Theorem 7, shows that \( f \) is not convex. This completes the proof.

We point out that Theorem 8 only applies to functions \( f \) that are twice continuously differentiable on an open convex set \( S \). This is to be contrasted with Theorem 7, 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 8, 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 \( S = \mathbb{R}^2 \times \{0\} \). However, its Hessian, which is given by

\[
\nabla^2 f(x, y) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}
\]

for all \( (x, y) \in \mathbb{R}^2 \),

is nowhere positive semidefinite.

2.3 Some Useful Inequalities

Let us begin with Jensen’s inequality, which can be viewed as a generalization of (4).

**Proposition 5 (Jensen’s Inequality)** Let \( f \) be a convex function (as in Definition 3). Then, for any \( x_1, \ldots, x_k \in \text{dom}(f) \) and \( \alpha_1, \ldots, \alpha_k \in [0, 1] \) such that \( \sum_{i=1}^k \alpha_i = 1 \), we have

\[
f\left(\sum_{i=1}^k \alpha_i x_i\right) \leq \sum_{i=1}^k \alpha_i f(x_i).
\]

The proof of Jensen’s inequality is straightforward and we leave it as an exercise to the reader.

What is more curious is that Jensen’s inequality can be used to prove the following well-known inequalities:

**Proposition 6** The following hold:

1. (Arithmetic–Geometric Mean Inequality) For every \( x \in \mathbb{R}^n_+ \), we have

\[
\left(\prod_{i=1}^n x_i\right)^{1/n} \leq \frac{1}{n} \sum_{i=1}^n x_i.
\]

2. (Cauchy–Schwarz’s Inequality) For every \( x, y \in \mathbb{R}^n \), we have \( |x^T y| \leq \|x\|_2 \cdot \|y\|_2 \).
3. (Hölder’s Inequality) Let $p \in (1, \infty)$ and $q \in (1, \infty)$ be such that $p^{-1} + q^{-1} = 1$. Then, for every $x, y \in \mathbb{R}^n$, we have $|x^Ty| \leq \|x\|_p \cdot \|y\|_q$.

**Proof** The key is to observe that the function $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ given by $f(x) = -\log x$ is convex (this can be verified directly or by checking $f''(x) \geq 0$ for all $x \in \mathbb{R}^+$). To prove the Arithmetic–Geometric Mean Inequality, let $x_1, \ldots, x_n \in \mathbb{R}^+$, and consider an arbitrary convex combination $x = \sum_{i=1}^n \alpha_i x_i$. By Jensen’s inequality, we have

$$-\log \left( \sum_{i=1}^n \alpha_i x_i \right) \leq -\sum_{i=1}^n \alpha_i \log(x_i).$$

Upon multiplying both sides by $-1$ and taking exponentials, we obtain

$$\sum_{i=1}^n \alpha_i x_i \geq \prod_{i=1}^n x_i^{\alpha_i}. \quad (9)$$

The desired result then follows by taking $\alpha_1 = \cdots = \alpha_n = 1/n$. Now, it is clear that Hölder’s inequality includes Cauchy–Schwarz’s inequality as a special case, so it suffices to establish the former. By (9), we have $u^\alpha v^{1-\alpha} \leq \alpha u + (1-\alpha)v$ for $u, v \in \mathbb{R}_+$ and $\alpha \in [0, 1]$. Upon taking $\alpha = 1/p$,

$$u = \sum_{i=1}^n |x_i|^p, \quad v = \sum_{i=1}^n |y_i|^q$$

and summing the resulting inequalities over $i = 1, \ldots, n$, we obtain the desired result. \qed

### 2.4 Examples of Convex 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 5**

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

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \begin{cases} \frac{\exp(x_i)}{\sum_{i=1}^{n} \exp(x_i)} - \frac{\exp(2x_i)}{(\sum_{i=1}^{n} \exp(x_i))^2} & \text{if } i = j, \\ -\frac{\exp(x_i + x_j)}{(\sum_{i=1}^{n} \exp(x_i))^2} & \text{if } i \neq j. \end{cases}$$

This gives

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

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

$$v^T \nabla^2 f(x) v = \frac{1}{(e^T z)^2} \left[ \left( \sum_{i=1}^{n} z_i \right)^2 - \left( \sum_{i=1}^{n} z_i v_i \right)^2 \right]$$

$$= \frac{1}{(e^T z)^2} \left[ \left( \sum_{i=1}^{n} (\sqrt{z_i})^2 \right) \left( \sum_{i=1}^{n} (\sqrt{z_i} v_i)^2 \right) - \left( \sum_{i=1}^{n} \sqrt{z_i} \cdot (\sqrt{z_i} v_i) \right)^2 \right]$$

$$\geq 0$$

by Cauchy–Schwarz’s inequality. Hence, $f$ is convex.
2. Let $f : \mathbb{R}_{++}^n \to \mathbb{R}$ be given by $f(x) = (\prod_{i=1}^{n} x_i)^{1/n}$. We compute

\[
\frac{\partial^2 f}{\partial x_i \partial x_j} = \begin{cases} 
-(n-1) \frac{(\prod_{i=1}^{n} x_i)^{1/n}}{n^2 x_i^2} & \text{if } i = j, \\
\frac{(\prod_{i=1}^{n} x_i)^{1/n}}{n^2 x_i x_j} & \text{if } i \neq j.
\end{cases}
\]

This gives

\[
\nabla^2 f(x) = -\frac{(\prod_{i=1}^{n} x_i)^{1/n}}{n^2} \left[ n \cdot \text{diag}(x_1^{-2}, \ldots, x_n^{-2}) - q q^T \right],
\]

where $q_i = x_i^{-1}$ for $i = 1, \ldots, n$. Using a similar argument as above, one can show that $\nabla^2 f(x) \preceq 0$. Hence, $f$ is concave.

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

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

This gives

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

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

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

since

\[
-\left( \sum_{i=1}^{n} x_i^p \right) \left( \sum_{i=1}^{n} v_i x_i^p \right) + \left( \sum_{i=1}^{n} v_i x_i^{p-1} \right)^2 \leq 0
\]

by Cauchy-Schwarz’s inequality. It follows that $f$ is concave.

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

\[
\text{epi}(f) = \{(x, Y, r) \in \mathbb{R}^n \times S_{++}^n \times \mathbb{R} : Y > 0, x^T Y^{-1} x \leq r \}
\]

\[
= \{(x, Y, r) \in \mathbb{R}^n \times S_{++}^n \times \mathbb{R} : \begin{bmatrix} Y & x \\ x^T & r \end{bmatrix} \succeq 0, Y > 0 \},
\]

where the last equality follows from the Schur complement (see, e.g., [2, Section A.5.5]). This shows that $\text{epi}(f)$ is a convex set, which implies that $f$ is convex.
5. 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])

\[
    f(X) = \sup \{ u^T X v : \|u\|_2 = 1, \|v\|_2 = 1 \}.
\]

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

6. Let \( f : S^n_{++} \to \mathbb{R} \) be given by \( f(X) = \log \det X \). Let \( X_0 \in S^n_{++} \) and \( H \in S^n \), and consider the function \( \tilde{f}_{X_0,H} : \mathbb{R} \to \mathbb{R} \) given by \( \tilde{f}_{X_0,H}(t) = f(X_0 + tH) \). We shall restrict \( \tilde{f}_{X_0,H} \) to the interval of values of \( t \) such that \( X_0 + tH \succ 0 \). Now, we compute

\[
    \tilde{f}_{X_0,H}(t) = \log \det (X_0 + tH) = \log \det \left( X_0^{1/2} \left( I + tX_0^{-1/2} H X_0^{-1/2} \right) X_0^{1/2} \right) = \sum_{i=1}^n \log(1 + t\lambda_i) + \log \det X_0,
\]

where \( \lambda_1, \ldots, \lambda_n \) are the eigenvalues of \( X_0^{-1/2} H X_0^{-1/2} \). Since we have

\[
    \frac{\partial^2 f}{\partial t^2}_{X_0,H}(t) = -\sum_{i=1}^n \frac{\lambda_i^2}{(1 + t\lambda_i)^2} \leq 0,
\]

it follows that \( f \) is concave on \( S^n_{++} \).

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 if and only if 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 graph of \( f \) lies above its tangent hyperplane at every \( \bar{x} \in \mathbb{R}^n \). 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 5** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a function. A vector \( s \in \mathbb{R}^n \) is called a subgradient of \( f \) at \( \bar{x} \) if

\[
    f(x) \geq f(\bar{x}) + s^T(x - \bar{x}) \quad \text{for all } x \in \mathbb{R}^n.
\]

The set of vectors \( s \) such that (10) 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 we have \( \partial f(0) = [-1, 1] \).

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 [8, Chapter 2, Section 2.5].
Theorem 9 Let \( f : \mathbb{R}^n \to \mathbb{R} \cup \{ +\infty \} \) be a convex function.

(a) (Subgradient and Directional Derivative) Let
\[
f'(x, d) = \lim_{t \downarrow 0} \frac{f(x + td) - f(x)}{t}
\]
be the directional derivative of \( f \) at \( x \in \mathbb{R}^n \) in the direction \( d \in \mathbb{R}^n \setminus \{0\} \), and let \( x \in \text{int dom}(f) \). Then, \( \partial f(x) \) is a non-empty compact convex set. Moreover, for any \( d \in \mathbb{R}^n \), we have \( f'(x, d) = \max_{g \in \partial f(x)} g^T d \).

(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 is 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 \( x_0 \in \text{dom}(f) \) such that \( f_1 \) is continuous at \( x_0 \). Then, we have
\[
\partial f(x) = \partial f_1(x) + \partial f_2(x)
\]
for all \( x \in \text{dom}(f) \).

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, 9, 7, 5, 1, 8].

References


13