# Convex Analysis

Jayadev Naram

# Contents

# 1 Convex Sets and Convex Functions

## 1.1 Affine Sets

Let $\mathbb{R}^n$ denote the $n$-dimensional Euclidean real vector space with the inner product defined for any $x, y \in \mathbb{R}^n$ as $\langle x, y \rangle = x^T y = \sum_{i=1}^n x_i y_i$, where $x = [x_1, \ldots, x_n]^T$ and $y = [y_1, \ldots, y_n]^T$ are the coordinates of $x$ and $y$ respectively. Let the norm and the metric on $\mathbb{R}^n$ be defined as $\|x\| = \sqrt{\langle x, x \rangle}$ and $d(x, y) = \|x - y\|$, respectively.

**Definition 1.1.** If $x$ and $y$ are different points in $\mathbb{R}^n$, the set of points of the form

$$(1 - \lambda)x + \lambda y = x + \lambda(y - x), \qquad \lambda \in \mathbb{R},$$

is called the **_line_** through $x$ and $y$. A subset $M$ of $\mathbb{R}^n$ is called an **_affine set_** if $(1-\lambda)x + \lambda y \in M$ for every $x, y \in M$ and $\lambda \in \mathbb{R}$.

The formal geometry of affine sets may be developed from the theorems of linear algebra about subspaces of $\mathbb{R}^n$.

**Theorem 1.2.** *$M$ is a subspace of $\mathbb{R}^n$ iff $M$ is an affine set containing origin.*

*Proof.* ( $\implies$ ) Every subspace contains 0, and being closed under addition and scalar multiplication, a subspace is in particular an affine set.
( $\impliedby$ ) Suppose $M$ is an affine set containing 0. For every $x \in M$ and $\lambda \in \mathbb{R}$, we have

$$\lambda x = (1 - \lambda)0 + \lambda x \in M,$$

so $M$ is closed under scalar multiplication. Now, if $x, y \in M$, we have

$$\tfrac{1}{2}(x + y) = \tfrac{1}{2}x + (1 - \tfrac{1}{2})y \in M,$$

and hence

$$x + y = 2(\tfrac{1}{2}(x + y)) \in M.$$

Thus $M$ is also closed under addition and is a subspace. $\qquad\square$

**Definition 1.3.** For $M \subseteq \mathbb{R}^n$ and $a \in \mathbb{R}^n$, the **_translate_** of $M$ by $a$ is defined to be the set

$$M + a = \{x + a \; : \; x \in M\}.$$

An affine set $M$ is said to be **_parallel_** to an affine set $L$ if $M = L + a$ for some $a$.

**Proposition 1.4.** *A translate of an affine set is another affine set.*

*Proof.* Suppose $M$ is an affine set and consider $L = M + a$, for some $a \in R^n$. For any points $x, y \in M$, $x + a, y + a \in L$, then we have

$$\lambda(x + a) + (1 - \lambda)(y + a) = \underbrace{\lambda x + (1 - \lambda)y}_{\in\, M} + a \in L.$$

$\qquad\square$

**Proposition 1.5.** *The relation "$M$ parallel to $L$" between two affine subsets $M, L \subseteq \mathbb{R}^n$ is an equivalence relation on the collection of affine subsets of $\mathbb{R}^n$.*

*Proof.* Consider affine sets $M, L, N \subseteq R^n$ and $a, b \in R^n$.

  i Reflexive: $M = M + 0$, so $M$ is parallel to $M$.

ii Symmetric: Suppose $M = L + a = \{x + a \ : \ x \in L\}$. Then it is easy to see that $L = M + (-a)$, i.e., $L$ is parallel to $M$.

iii Transistive: Suppose $M = L + a$ and $L = N + b$. Then $M = N + (a + b)$, thus $M$ is parallel to $N$.

$\square$

**Remark 1.6.** From the equivalence relation developed in the previous proposition, if $L_1$ and $L_2$ both are parallel to $M$, then $L_1$ is parallel to $L_2$.

**Theorem 1.7.** *Every non-empty affine set $M$ is parallel to a unique subspace $L$. This $L$ is given by*

$$L = M - M = \{x - y \ : \ x, y \in M\}.$$

*Proof.* Let us show first that $M$ cannot be parallel to two different subspaces. Subspaces $L_1$ and $L_2$ parallel to $M$ would be parallel to each other, so that $L_2 = L_1 + a$ for some $a$. Since $0 \in L_2$, we would then have $-a \in L_1$, and hence $a \in L_1$. But then $L_1 \supseteq L_1 + a = L_2$. By similar argument $L_2 \supseteq L_1$, so $L_1 = L_2$. This establishes the uniqueness.

Now observe that, for any $y \in M$, $M - y = M + (-y)$ is a translate of $M$ containing 0. We considered $M$ to be an affine set and from Prop. 1.4 its translate is also an affine set, in this case containing 0. Then, by Theorem 1.2 and what we have just proved, this affine set must be the unique subspace parallel to $M$. Since $L = M - y$ no matter which $y \in M$ is chosen, we actually have $L = M - M$. $\square$

**Definition 1.8.** The **_dimension_** of a non-empty affine set is defined as the dimension of the subspace parallel to it.

By convention, the dimension of $\emptyset$ is $-1$, and affine sets of dimension $0, 1$ and $2$ are points, lines and planes respectively. An $(n-1)$-dimensional affine set in $\mathbb{R}^n$ is called a **_hyperplane_**.

**Theorem 1.9.** *Given $b \in \mathbb{R}^n$ and an $m \times n$ real matrix $B$, the set*

$$M = \{x \in \mathbb{R}^n \ : \ Bx = b\}$$

*is an affine set in $\mathbb{R}^n$. Moreover, every affine set may be represented in this way.*

*Proof.* ($\implies$) If $x, y \in M$ and $\lambda \in \mathbb{R}$, then we have

$$B((1 - \lambda)x + \lambda y) = (1 - \lambda)Bx + \lambda By = (1 - \lambda)b + \lambda b = b,$$

so $(1 - \lambda)x + \lambda y \in M$. Thus the given $M$ is affine.
($\impliedby$) Starting with an arbitrary non-empty affine set $M$ other than $\mathbb{R}^n$ itself, let $L$ be the subspace parallel to $M$. Let $b_1, \ldots, b_m$ be a basis for $L^\perp$, the orthogonal complement of $L$. Then

$$L = \{x \ : \ \langle x, b_i \rangle = 0, \ i = 1, \ldots, m\} = \{x \ : \ Bx = 0\},$$

where $B$ is the $m \times n$ matrix whose rows are $b_1, \ldots, b_m$. Since $M$ is parallel to $L$, there exists an $a \in \mathbb{R}^n$ such that

$$M = L + a = \{x \ : \ B(x - a) = 0\} = \{x \ : \ Bx = b\},$$

where $b = Ba$. $\square$

**Corollary 1.9.1.** *Given $\beta \in \mathbb{R}$ and a non-zero $b \in \mathbb{R}^n$, the set*

$$H = \{x \in \mathbb{R}^n \ : \ \langle x, b \rangle = \beta\}$$

*is a hyperplane in $\mathbb{R}^n$. Moreover, every hyperplane may be represented in this way, with $b$ and $\beta$ unique up to a common non-zero multiple. The vector $b$ is called a **_normal_** to the hyperplane $H$. Every other normal to $H$ is either a positive or a negative scalar multiple of $b$.*

**Lemma 1.10.** *The intersection of an arbitrary collection of affine sets is affine.*

*Proof.* Consider an arbitrary collection of affine set $\{M_i\}$ and define $M = \cap_i M_i$. Let $x, y \in M$ and $\lambda \in \mathbb{R}$, then we have $(1 - \lambda)x + \lambda y \in M_i$ for each $i$ and $(1 - \lambda)x + \lambda y \in M$, thus $M$ is affine. $\qquad\square$

**Definition 1.11.** Given any $S \subseteq \mathbb{R}^n$, the ***affine hull*** of $S$, denoted by aff $S$, is the intersection of the collection of affine sets containing $S$.

**Proposition 1.12.** *For any $S \subseteq \mathbb{R}^n$, we have*

$$\text{aff } S = \{\lambda_1 x_1 + \cdots + \lambda_m x_m \ : \ x_i \in S \ \text{and} \ \lambda_1 + \cdots + \lambda_m = 1\} \equiv C.$$

*Proof.* ( $\implies$ ) For any $x \in S$, we have the trivial combination $x = 1x \in C$ so that $S \subseteq C$. Suppose $x_1, x_2 \in C$ so that $x = \sum_{i=1}^{m} \mu_i x_i$ and $y = \sum_{i=m+1}^{n} \mu_i x_i$ with $\sum_{i=1}^{m} \mu_i = 1$, $\sum_{i=m+1}^{n} \mu_i = 1$ and $x_i \in S$ for $i = 1, \ldots, n$. Then for any $\lambda \in \mathbb{R}$, we have

$$(1 - \lambda)x + \lambda y = (1 - \lambda)\sum_{i=1}^{m} \mu_i x_i + \lambda \sum_{i=m+1}^{n} \mu_i x_i,$$

such that the sum of the coefficients is

$$(1 - \lambda)\sum_{i=1}^{m} \mu_i + \lambda \sum_{i=m+1}^{n} \mu_i = (1 - \lambda)(1) + \lambda(1) = 1.$$

This shows that $(1 - \lambda)x + \lambda y \in C$ so that $C$ is an affine set containing $S$, thus, aff $S \subseteq C$. ( $\impliedby$ ) Suppose $x \in C$, so that

$$x = \lambda_1 x_1 + \cdots + \lambda_m x_m \ \text{where} \ \lambda_1 + \cdots + \lambda_m = 1.$$

We show that $x \in$ aff $S$ by induction. The base case $m = 2$ holds from the definition of affine set. Assume that for $m > 2$, $x$ belongs to aff $S$. Then consider the case of $m$-point combination in $x$. There is some index $j$ such that $\lambda_j \neq 1$, otherwise $\lambda_1 + \cdots + \lambda_m = m \neq 1$ which is a contradiction. We can assume WLOG $j = 1$, then we have

$$x = \sum_{i=1}^{m} \lambda_i x_i = \lambda_1 x_1 + (1 - \lambda_1)\sum_{i=2}^{m} \frac{\lambda_i}{1 - \lambda_1} x_i \ \text{with} \ \sum_{i=2}^{m} \frac{\lambda_i}{1 - \lambda_1} = 1.$$

This shows that $y = \sum_{i=2}^{m} \frac{\lambda_i}{1 - \lambda_1} x_i \in$ aff $S$ by the induction hypothesis and similarly $x = \lambda_1 x_1 + (1 - \lambda_1)y \in$ aff $S$. $\qquad\square$

**Definition 1.13.** A set of $m + 1$ points $b_0, b_1, \ldots, b_m$ is said to be ***affinely independent*** if aff $\{b_0, b_1, \ldots, b_m\}$ is $m$-dimensional.

**Proposition 1.14.** *The points $b_0, b_1, \ldots, b_m$ are affinely independent iff $b_1 - b_0, \ldots, b_m - b_0$ are linearly independent.*

*Proof.* By definition, the points $b_0, b_1, \ldots, b_m$ are affinely independent iff $M = $ aff $\{b_0, b_1, \ldots, b_m\}$ is $m$-dimensional. But then

$$M = \text{aff } \{b_0, b_1, \ldots, b_m\} = L + b_0,$$

where $L = \text{aff}\{0, b_1 - b_0, \ldots, b_m - b_0\}$ is the smallest subspace parallel to $M$, by Theorem 1.2 and 1.7. The dimension of $L$ is $m$ iff these vectors are linearly independent. $\qquad\square$

**Definition 1.15.** A map $T : \mathbb{R}^n \to \mathbb{R}^m$ is called an ***affine transformation*** if

$$T((1 - \lambda)x + \lambda y) = (1 - \lambda)Tx + \lambda Ty,$$

for every $x, y \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.

**Theorem 1.16.** *The affine transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ are the mappings $T$ of the form $Tx = Ax + a$, where $A$ is a linear transformation and $a \in \mathbb{R}^n$.*

*Proof.* ( $\implies$ ) Suppose $T$ is affine and define the map $Ax = Tx - T0$, where $0 \in \mathbb{R}^n$ is the origin. We show that $A$ is linear. Note that

$$
\begin{aligned}
A((1 - \lambda)x + \lambda y) &= T((1 - \lambda)x + \lambda y) - T0 \\
&= (1 - \lambda)Tx + \lambda Ty - T0 \\
&= (1 - \lambda)Tx + \lambda Ty - (1 - \lambda)T0 - \lambda T0 \\
&= (1 - \lambda)(Tx - T0) + \lambda(Ty - T0) \\
&= (1 - \lambda)Ax + \lambda Ay.
\end{aligned}
$$

This shows that $A$ is affine. Also note that $A0 = T0 - T0 = 0$. Now, we proceed to show the linearity of $A$ using these two facts.

$$A(\lambda x) = A((1 - \lambda)0 + \lambda x) = (1 - \lambda)A0 + \lambda Ax = \lambda Ax, \tag{1}$$

and

$$
\begin{aligned}
A(\tfrac{1}{2}(x+y)) &= A(\tfrac{1}{2}x + (1 - \tfrac{1}{2})y) = \tfrac{1}{2}Ax + \tfrac{1}{2}Ay, \\
A(x + y) &= A(2(\tfrac{1}{2}(x + y))) = 2A(\tfrac{1}{2}(x + y)) = Ax + Ay.
\end{aligned} \tag{2}
$$

From (1) and (2), we have proved that $A$ is linear.
( $\impliedby$ ) If $Tx = Ax + a$, where $A$ is linear, one has

$$T((1 - \lambda)x + \lambda y) = (1 - \lambda)Ax + \lambda Ay + a = (1 - \lambda)Tx + \lambda Ty,$$

thus $T$ is affine. $\qquad\square$

**Proposition 1.17.** *Consider an affine transformation $T$ on $\mathbb{R}^n$ defined as $Tx = Ax + a$, where $A$ is a linear transformation on $\mathbb{R}^n$ and $a \in \mathbb{R}^n$. If $A^{-1}$ exists, then $T^{-1}$ exists and it is affine on $\mathbb{R}^n$ which is given by*

$$T^{-1}(x) = A^{-1}(x - a).$$

*Proof.* First we verify that $T^{-1}$ defined above is the inverse of $T$. For any $x \in \mathbb{R}^n$, we have

$$
\begin{aligned}
T^{-1}(T(x)) &= T^{-1}(Ax + a) = A^{-1}(Ax + a - a) = x \\
T(T^{-1}(x)) &= T(A^{-1}(x - a)) = A(A^{-1}(x - a)) + a = x.
\end{aligned}
$$

Now we show that $T^{-1}$ is affine. For any $x, y \in \mathbb{R}^n$, we have

$$
\begin{aligned}
T^{-1}((1 - \lambda)x + \lambda y) &= A^{-1}((1 - \lambda)x + \lambda y - a) \\
&= (1 - \lambda)A^{-1}x + \lambda A^{-1}y - A^{-1}a \\
&= (1 - \lambda)A^{-1}(x - a) + \lambda A^{-1}(y - a) \\
&= (1 - \lambda)T^{-1}x + \lambda T^{-1}y.
\end{aligned}
$$

$\qquad\square$

**Proposition 1.18.** *If a mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is affine, then the image set $TM = \{Tx \; : \; x \in M\}$ is affine in $\mathbb{R}^n$ for every affine set $M$ in $\mathbb{R}^n$. In particular, then, affine transformations preserve affine hulls:*

$$T(\mathrm{aff}\, S) = \mathrm{aff}\,(TS).$$

*Proof.* Let $M \subseteq \mathbb{R}^n$ be an affine set. Then for any $x, y \in M$ and $\lambda \in \mathbb{R}$, we have $T(x), T(y) \in TM$ and since $T$ is affine, we get

$$(1 - \lambda)T(x) + \lambda T(y) = T((1 - \lambda)x + \lambda y).$$

But $(1 - \lambda)x + \lambda y \in M$, so that $(1 - \lambda)T(x) + \lambda T(y) \in TM$. Thus, $TM$ is affine. $\qquad\square$

**Theorem 1.19.** *Let $\{b_0, b_1, \ldots, b_m\}$ and $\{b_0', b_1', \ldots, b_m'\}$ be affinely independent sets in $\mathbb{R}^n$. Then there exists a one-to-one affine transformation $T$ on $\mathbb{R}^n$, such that $Tb_i = b_i'$ for $i = 0, \ldots, m$. If $m = n$, $T$ is unique.*

*Proof.* Enlarging the given affinely independent sets if necessary, we can reduce the question to the case when $m = n$. Then, there exists a unique one-to-one linear transformation $A$ on $\mathbb{R}^n$ carrying the basis $b_1 - b_0, \ldots, b_n - b_0$ of $\mathbb{R}^n$ onto the basis $b_1' - b_0', \ldots, b_n' - b_0'$. The desired affine transformation is then given by $Tx = Ax + a$, where $a = b_0' - Ab_0$. $\qquad\square$

**Corollary 1.19.1.** *Let $M_1$ and $M_2$ be any two affine sets in $\mathbb{R}^n$ of the same dimension. Then there exists a one-to-one affine transformation $T$ on $\mathbb{R}^n$ such that $TM_1 = M_2$.*

*Proof.* Any $m$-dimensional affine set can be expressed as the affine hull of an affinely independent set of $m + 1$ points, and affine hulls are preserved by affine transformations. $\qquad\square$

**Proposition 1.20.** *The graph of an affine transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is an affine subset of $\mathbb{R}^{m+n}$.*

*Proof.* Let $G = \{z \in \mathbb{R}^{m+n} \; : \; z = (x, y), \; Tx = y\}$ denote the graph of $T$. From Theorem 1.16, we know that $Tx = Ax + a$ for some linear transformation $A$ and $a \in \mathbb{R}^n$. Then consider a linear map $Bz = Ax - y$, for $z = (x, y) \in \mathbb{R}^{m+n}$. Then for all $z \in G$, we have $Bz = b$ with $b = -a$, i.e., $G = \{z \in \mathbb{R}^{m+n} \; : \; Bz = b\}$. Then by Theorem 1.9, $G$ is affine subset of $\mathbb{R}^{m+n}$. $\quad\square$

## 1.2 Convex Sets and Cones

**Definition 1.21.** If $x$ and $y$ are different points in $\mathbb{R}^n$, the set of points of the form

$$(1 - \lambda)x + \lambda y, \quad 0 \le \lambda \le 1,$$

is called the ***(closed) line segment*** between $x$ and $y$. A subset $C$ of $\mathbb{R}^n$ is called a ***convex set*** if $(1 - \lambda)x + \lambda y \in C$ for every $x, y \in C$ and $0 \le \lambda \le 1$.

**Remark 1.22.** Affine sets are convex sets but not all convex sets are affine sets. For example consider solid cubes in $\mathbb{R}^3$.

**Example 1.23** (Half-spaces)**.** For any non-zero $b \in \mathbb{R}^n$ and any $\beta \in \mathbb{R}$, the sets

$$\{x \; : \; \langle x, b \rangle \le \beta\}, \qquad \{x \; : \; \langle x, b \rangle \ge \beta\},$$

are called ***closed half-spaces***. The sets

$$\{x \; : \; \langle x, b \rangle < \beta\}, \qquad \{x \; : \; \langle x, b \rangle > \beta\},$$

are called ***open half-spaces***. All four sets are convex and correspond to the same hyperplane $H = \{x \; : \; \langle x, b \rangle = \beta\}$.

**Lemma 1.24.** *The intersection of an arbitrary collection of convex sets is convex.*

Proof similar to that of Lemma 1.10.

**Corollary 1.24.1.** *Let $b_i \in \mathbb{R}^n$ and $\beta_i \in \mathbb{R}$ for $i \in I$, where $I$ is an arbitrary index set. Then the set*
$$C = \{x \in \mathbb{R}^n \ : \ \langle x, b_i \rangle \leq \beta_i, \ \forall \ i \in I\}$$
*is convex.*

Corollary 1.24.1 will be generalized by Corollary 1.80.1.

**Remark 1.25.** The conclusion of the corollary would still be valid if some of the inequalities $\leq$ were replaced by $\geq, <, >$ or $=$. Thus, given any system of simultaneous linear inequalities and equations in $n$ variables, the set $C$ of solutions is a convex set $\mathbb{R}^n$.

**Definition 1.26.** A set which can be expressed as the intersection of finitely many closed half spaces of $\mathbb{R}^n$ is called a ***polyhedral*** convex set.

**Definition 1.27.** A vector sum
$$\lambda_1 x_1 + \cdots + \lambda_m x_m$$
is called a ***convex combination*** of $x_1, \ldots, x_m$ if the coefficients $\lambda_i$ are all nonnegative and $\lambda_1 + \cdots + \lambda_m = 1$.

**Definition 1.28.** The intersection of all the convex sets containing a given subset $S$ of $\mathbb{R}^n$ is called the ***convex hull*** of $S$ and is denoted by conv $C$.

**Remark 1.29.** The convex hull of $S$ is a convex set by Lemma 1.24, the unique smallest one containing $S$.

**Theorem 1.30.** *For any $S \subseteq \mathbb{R}^n$, conv $S$ consists of all the convex combinations of the elements of $S$.*

The proof is same as that of Prop. 1.12 with the additional nonnegativity constraint added throughout.

**Corollary 1.30.1.** *A subset $C$ of $\mathbb{R}^n$ is convex iff $C$ contains all the convex combinations of its elements.*

**Corollary 1.30.2.** *The convex hull of a finite subset $\{b_0, \ldots, b_m\}$ of $\mathbb{R}^n$ consists of all the vectors of the form $\lambda_0 b_0 + \cdots + \lambda_m b_m$ with $\lambda_i \geq 0$, $i = 0, \ldots, m$, $\lambda_0 + \cdots + \lambda_m = 1$.*

**Definition 1.31.** A set which is the convex hull of finitely many points is called a ***polytope***. If $\{b_0, \ldots, b_m\}$ is affinely independent, its convex hull is called an ***m-dimensional simplex***, and $b_0, \ldots, b_m$ are called the ***vertices*** of the simplex.

**Definition 1.32.** The ***dimension of a convex set*** $C$ is defined to be the dimension of the affine hull of $C$.

**Remark 1.33.** The dimension of an affine hull or simplex as already defined agrees with its dimension as a convex set.

**Theorem 1.34.** *The dimension of a convex set $C$ is the maximum of the dimensions of the various simplices included in $C$.*

*Proof.* The convex hull of any subset of $C$ is included in $C$ by the convexity of $C$. The maximum dimension of the various simplices included in $C$ is thus the largest $m$ such that $C$ contains an affinely independent set of $m + 1$ elements. Let $\{b_0, \ldots, b_m\}$ be such a set with $m$ maximal, and let $M$ be its affine hull. Then dim $M = m$ and $M \subseteq$ aff $C$. Furthermore, $C \subseteq M$, for if $C - M$ contained an element $b$, the set of $m + 2$ elements $b_0, \ldots, b_m, b$ in $C$ would be affinely independent, contrary to the maximality of $m$. Since aff $C$ is the smallest affine set which includes $C$, it follows that aff $C \subseteq M$, or equivalently aff $C = M$ and hence dim $C = m$. $\qquad\square$

**Definition 1.35.** A subset $K$ of $\mathbb{R}^n$ is called a **_cone_** if it closed under positive scalar multiplication, i.e., $\lambda x \in K$ when $x \in K$ and $\lambda > 0$. A **_convex cone_** is a cone which is a convex set.

**Example 1.36** (Convex cone)**.** Subspaces of $\mathbb{R}^n$; open and closed half-spaces corresponding to a hyperplane passing through the origin, i.e., $\beta = 0$; the **_nonnegative orthant_** of $\mathbb{R}^n$

$$\{x = (\xi_1, \ldots, \xi_n) \ : \ \xi_1 \geq 0, \ldots, \xi_n \geq 0\},$$

and the **_positive orthant_**

$$\{x = (\xi_1, \ldots, \xi_n) \ : \ \xi_1 > 0, \ldots, \xi_n > 0\}.$$

It is customary to write $x \geq x'$ for $x, x' \in \mathbb{R}^n$, if $x - x'$ belongs to the nonnegative orthant, i.e.,

$$\xi_i \geq \xi_i' \quad \text{for} \quad i = 1, \ldots, n.$$

In this notation, the nonnegative orthant consits of the vectors $x$ such that $x \geq 0$.

**Lemma 1.37.** *The intersection of an arbitrary collection of convex cones is a convex cone.*

**Corollary 1.37.1.** *Let $b_i \in \mathbb{R}^n$ for $i \in I$, where $I$ is an arbitrary index set. Then*

$$K = \{x \in \mathbb{R}^n \ : \ \langle x, b_i \rangle \leq 0, \ i \in I\}$$

*is a convex cone.*

Similar remarks to Remark 1.25 hold for the above corollary. The following characterization of convex cones highlights an analogy between convex cones and subspaces.

**Theorem 1.38.** *A subset of $\mathbb{R}^n$ is a convex cone iff it is closed under addition and positive scalar multiplication.*

*Proof.* ( $\implies$ ) Let $K$ be a convex cone. By definition it is closed under positive scalar multiplication, so we need to show that it is closed under addition. Let $x, y \in K$, then by the convexity of $K$, $z = (1/2)(x + y) \in K$, and $x + y = 2z \in K$ since $K$ is a cone.
( $\impliedby$ ) Suppose $K$ is closed under addition and positive scalar multiplication, and let $0 < \lambda < 1$. Then, the vectors $(1 - \lambda)x$ and $\lambda y$ belong $K$ for $x, y \in K$, and hence $(1 - \lambda)x + \lambda y \in K$. $\qquad\square$

**Corollary 1.38.1.** *A subset of $\mathbb{R}^n$ is a convex cone iff it contains all the positive linear combinations of its elements.*

**Corollary 1.38.2.** *Let $S$ be an arbitrary subset of $\mathbb{R}^n$, and let $K$ be the set of all positive linear combinations of $S$. Then $K$ is the smallest convex cone which includes $S$.*

**Corollary 1.38.3.** *Let $C$ be a convex set, and let*

$$K = \{\lambda x \ : \ \lambda > 0, x \in C\}.$$

*Then $K$ is the smallest convex cone which includes $C$.*

**Theorem 1.39.** *Let $K$ be a convex cone containing $0$. Then there is the smallest subspace containing $K$, namely*

$$K - K = \{x - y \ : \ x \in K, y \in K\} = \text{aff } K,$$

*and there is the largest subspace contained within $K$, namely $(-K) \cap K$.*

*Proof.* To show that each of these sets are subspaces, we first show that they are convex cones. Then by Theorem 1.38, it suffices to show that they contain $0$ and are closed under multiplication by $-1$.

Consider the set $K - K = \{x - y \ : \ x, y \in K\}$. For any $z_1, z_2 \in K - K$ and $\lambda_1, \lambda_2 > 0$, we have $z_1 = x_1 - y_1$ and $z_2 = x_2 - y_2$ for some $x_1, x_2, y_1.y_2 \in K$. Then we have

$$\lambda_1 z_1 + \lambda_2 z_2 = (\lambda_1 x_1 + \lambda_2 x_2) - (\lambda_1 y_1 + \lambda_2 y_2).$$

But $\lambda_1 x_1 + \lambda_2 x_2, \lambda_1 y_1 + \lambda_2 y_2 \in K$ as $K$ is a convex cone, thus $\lambda z_1 + \lambda_2 z_2 \in K - K$. By Cor. 1.38.1, $K - K$ is also a convex cone. For every $x \in K$, $x - x = 0 \in K - K$. For every $z \in K - K$, there are $x, y \in K$ such that $z = x - y$, but then we also have $-z = y - x \in K - K$, showing that $K - K$ is a subspace. Since $0 \in K - K$, $K - 0 \subseteq K - K$, i.e., $K - K$ contains $K$. Let $L$ be a subspace containing $K$, then for any $z = x - y \in K - K$, $x, y \in K$, we have $x - y \in L$ or equivalently $z \in L$ showing that $K - K \subseteq L$, i.e., $K - K$ is the smallest subspace containing $K$. Since $K$ is a set containing $0$, its affine hull aff $K$ must also contain $0$, which by Theorem 1.2 is a subspace. But we have proved that $K - K$ is the smallest subspace containing $K$, hence $K - K = \text{aff } K$.

Now consider the set $(-K) \cap K$. By Lemma 1.37 $(-K) \cap K$ is also a convex cone containing $0$, since $0 \in K$ and $0 \in (-K)$. For any $x \in (-K) \cap K$, we have $x \in K$ and $x \in (-K)$, i.e., $-x \in K$, which shows that $-x \in K$ and $-x \in (-K)$ or $-x \in (-K) \cap K$. This shows that $(-K) \cap K$ is a subspace. Since $K \subseteq K$ and $K \cap (-K) \subseteq K$, $(-K) \cap K$ is a subspace contained in $K$. Suppose $M$ is a subspace contained in $K$ and let $x \in M$. Since $M$ is a subspace $-x \in M$, i.e., $x \in (-K) \cap K$ as $M \subseteq K$. Thus $(-K) \cap K$ is the largest subspace contained in $K$. $\qquad\square$

**Definition 1.40.** Let $S$ be an arbitrary subset of $\mathbb{R}^n$. The convex cone obtained by adjoining the origin to the set of all positive linear combinations of $S$ is called as the ***convex cone generated by $S$*** and is denoted by cone $S$.

**Remark 1.41.** The convex cone generated by $S$ is not, under our terminology, the same as the smallest convex cone containing $S$ unless the latter cone happens to contain the origin. If $S \neq \emptyset$, cone $S$ consists of all non-negative (rather than positive) linear combinations of elements of $S$. Clearly, cone $S = \text{conv } (\text{ray } S)$, where ray $S$ is the union of the origin and the various rays (half-lines of the form $\{\lambda y \ : \ \lambda \geq 0\}$) generated by the non-zero vectors $y \in S$.

**Definition 1.42.** A vector $x^*$ is said to be ***normal*** to a convex set $C$ at a point $a$, where $a \in C$, if $x^*$ does not make an acute angle with any line segment in $C$ with $a$ as endpoint, i.e., if $\langle x - a, x^* \rangle \leq 0$ for every $x \in C$. The set of all vectors $x^*$ normal to $C$ at $a$ is called the ***normal cone*** to $C$ at $a$. The ***normal cone map*** associated with the convex set $C$ is a set-valued map $N_C : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ which is defined as

$$N_C(a) = \begin{cases} \{x^* \in \mathbb{R}^n \ : \ \langle x - a, x^* \rangle \leq 0 \ \forall \ x \in C\}, & \text{if } a \in C \\ \emptyset, & \text{if } a \notin C. \end{cases}$$

At this point, the normal cone map definition only helps in simplfying notation in the following proposition. But it will be used more than that in context of Variational Analysis.

**Proposition 1.43.** *For a convex set $C$ and $a \in C$, $N_C(a)$ is a convex cone.*

*Proof.* Suppose that $x^*, y^* \in N_C(a)$ and $\lambda, \mu > 0$. Then $\langle x - a, x^* \rangle \leq 0$ and $\langle x - a, y^* \rangle \leq 0$, which implies that

$$\langle x - a, \lambda x^* + \mu y^* \rangle \leq 0, \text{ i.e., } \lambda x^* + \mu y^* \in N_C(a),$$

where we have used Theorem 1.38. $\qquad\square$

**Definition 1.44.** The ___barrier cone___ of a convex set $C$ is defined as the set of all vectors $x^*$ such that, for some $\beta \in \mathbb{R}$, $\langle x, x^* \rangle \leq \beta$ for every $x \in C$.

**Proposition 1.45.** *The barrier cone of a convex set is a convex cone.*

*Proof.* Suppose $C$ is a convex set and $x^*, y^*$ belong to the barrier cone of $C$ and $\lambda, \mu > 0$. Then there are some $\beta, \gamma \in \mathbb{R}$ such that $\langle x, x^* \rangle \leq \beta$ and $\langle x, y^* \rangle \leq \gamma$ for every $x \in C$. Using this, we see that for all $x \in C$

$$\langle x, \lambda x^* + \mu y^* \rangle \leq \lambda \beta + \mu \gamma \in \mathbb{R},$$

which shows that $\lambda x^* + \mu y^*$ also belongs to the barrier cone of $C$. $\qquad\square$

## 1.3 The Algebra of Convex Sets

**Theorem 1.46.** *Let $C$ be a convex set in $\mathbb{R}^n$. Then so is every translate $C + a$ and every* ___scalar multiple___ *$\lambda C = \{\lambda x \ : \ x \in C\}$, where $a \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.*

*Proof.* Let $x + a, y + a \in C + a$ and $0 \leq \mu \leq 1$. Then using convexity of $C$, we get

$$(1 - \mu)(x + a) + \mu(y + a) = ((1 - \mu)x + \mu y) + a \in C + a.$$

Similarly, let $\lambda x, \lambda y \in \lambda C$ and $0 \leq \mu \leq 1$. Then using convexity of $C$, we get

$$(1 - \mu)\lambda x + \mu \lambda y = \lambda((1 - \mu)x + \mu y) \in \lambda C.$$

$\qquad\square$

**Remark 1.47** (Symmetric Convex Set)**.** The ___symmetric reflection___ of $C$ across origin is $-C = (-1)C$. A convex set if said to be ___symmetric___ if $-C = C$. Such a set (if non-empty) must contain the origin, since it must contain along with each vector $x$, not only $-x$, but the entire line segment between $x$ and $-x$. The non-empty convex cones which are symmetric are the subspaces (Theorem 1.39).

**Theorem 1.48.** *If $C_1$ and $C_2$ are convex sets in $\mathbb{R}^n$, then so is their sum $C_1 + C_2$, where*

$$C_1 + C_2 = \{x_1 + x_2 \ : \ x_1 \in C_1, x_2 \in C_2\}.$$

*Proof.* Let $x_1 + x_2, x_1' + x_2' \in C_1 + C_2$ and $0 \leq \lambda \leq 1$. Then using convexity of $C_1$ and $C_2$, we obtain

$$(1 - \lambda)(x_1 + x_2) + \lambda(x_1' + x_2') = ((1 - \lambda)x_1 + \lambda x_1') + ((1 - \lambda)x_2 + \lambda x_2') \in C_1 + C_2.$$

$\qquad\square$

**Example 1.49.** If $C_1$ is any convex set and $C_2$ is the nonnegative orthant (see Example 1.36), then

$$C_1 + C_2 = \{x_1 + x_2 \ : \ x_1 \in C_1, x_2 \geq 0\} = \{x \ : \ \exists x_1 \in C_1, x_1 \leq x\}.$$

Thus, $C_1 + C_2$ is convex by Theorem 1.48 when $C_1$ is convex.

**Theorem 1.50.** *A set $C$ is a convex set iff $(1 - \lambda)C + \lambda C \subseteq C$, where $0 \leq \lambda \leq 1$.*

**Proposition 1.51.** *Let $C_1, \ldots, C_m$ be convex sets, then so is the linear combination*

$$C = \lambda_1 C_1 + \cdots + \lambda_m C_m,$$

*where $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$.*

*Proof.* Suppose $x = \lambda_1 x_1 + \cdots + \lambda_m x_m$, $x' = \lambda_1 x'_1 + \cdots + \lambda_m x'_m$ with $x_i, x'_i \in C_i$ so that $x, x' \in C$ and assume $0 \leq \lambda \leq 1$. Then

$$(1 - \lambda)x + \lambda x' = \lambda_1((1 - \lambda)x_1 + \lambda x'_1) + \cdots + \lambda_m((1 - \lambda)x_m + \lambda x'_m) \in C.$$

$\square$

**Remark 1.52** (Set Algebra)**.** The following algebraic laws are valid for the addition and scalar multiplication of sets even without convexity being involved:

$$
\begin{aligned}
C_1 + C_2 &= C_2 + C_1, \\
(C_1 + C_2) + C_3 &= C_1 + (C_2 + C_3), \\
\lambda_1(\lambda_2 C) &= (\lambda_1 \lambda_2)C, \\
\lambda(C_1 + C_2) &= \lambda C_1 + \lambda C_2, \\
(\lambda_1 + \lambda_2)C &\subseteq \lambda_1 C + \lambda_2 C, \tag{3}
\end{aligned}
$$

where $\lambda, \lambda_1, \lambda_2 \in \mathbb{R}$. The set $\{0\}$ is the identity element for the addition operation. Additive inverses do not exist for sets containing more than one point; the best one can say in general is that $0 \in [C + (-C)]$ when $C \neq \emptyset$. As can be noted from (3), to see that the reverse inclusion is not true in general, consider $C = \{1, 2\} \subseteq \mathbb{R}$ and $\lambda_1 = 1, \lambda_2 = 2$. Then $3C = \{3, 6\} \neq C + 2C = \{3, 4, 5, 6\}$. However, under the convexity assumption along with $\lambda_1, \lambda_2 \geq 0$, the next theorem shows that we can retain the reverse inclusion.

**Theorem 1.53.** *A set $C$ is a convex set iff $(\lambda_1 + \lambda_2)C = \lambda_1 C + \lambda_2 C$, where $\lambda_1, \lambda_2 \geq 0$.*

*Proof.* ( $\Longrightarrow$ ) The inclusion $(\lambda_1 + \lambda_2)C \subseteq \lambda_1 C + \lambda_2 C$ is true whether $C$ were convex or not. Now consider the reverse inclusion. Assume that $\lambda_1 + \lambda_2 > 0$ and define the set

$$D \equiv \tfrac{\lambda_1}{\lambda_1 + \lambda_2}C + \tfrac{\lambda_2}{\lambda_1 + \lambda_2}C.$$

Let $x \in D$, so that there exists $x_1, x_2 \in C$ such that

$$x = \tfrac{\lambda_1}{\lambda_1 + \lambda_2}x_1 + \tfrac{\lambda_2}{\lambda_1 + \lambda_2}x_2 = \tfrac{\lambda_1}{\lambda_1 + \lambda_2}x_1 + \left(1 - \tfrac{\lambda_1}{\lambda_1 + \lambda_2}\right)x_2 \in C,$$

hence $D \subseteq C$. On multiplying through by $\lambda_1 + \lambda_2$, we get $\lambda_1 C + \lambda_2 C \subseteq (\lambda_1 + \lambda_2)C$.
( $\Longleftarrow$ ) The distributive law implies that $\lambda C + (1 - \lambda)C \subseteq C$ when $0 \leq \lambda \leq 1$. Then by Theorem 1.50, this implies that $C$ is a convex set. $\square$

**Remark 1.54.** It follows from this theorem, for instance, that $C + C = 2C$, $C + C + C = 3C$, and so forth, when $C$ is convex.

**Proposition 1.55.** *Given an arbitrary family of non-empty convex sets $\{C_i \ : \ i \in I\}$ in $\mathbb{R}^n$, there is a unique largest convex set included in all of them, namely $\bigcap_{i \in I} C_i$, and a unique smallest convex set including all of them, namely $\mathrm{conv}\left(\bigcup_{i \in I} C_i\right)$.*

**Remark 1.56.** From the above proposition, the collection of convex subsets in $\mathbb{R}^n$ is a complete lattice under the natural partial order corresponding to inclusion.

**Theorem 1.57.** *Let $\{C_i \ : \ i \in I\}$ be an arbitrary collection of non-empty convex sets of $\mathbb{R}^n$, and let $C$ be the convex hull of the union of the collection, i.e., $C = \text{conv}\left(\bigcup_{i \in I} C_i\right)$. Then*

$$C = \bigcup \left\{ \sum_{i \in I} \lambda_i C_i \right\},$$

*where the sum is taken over all finite convex combinations (i.e., over all nonnegative choices of the coefficients $\lambda_i$ such that only finitely many are non-zero and these add up to 1.)*

*Proof.* By Theorem 1.30, $C$ is the set of all convex combinations

$$x = \mu_1 y_1 + \cdots + \mu_m y_m,$$

where $y_1, \ldots, y_m \in \bigcup_{i \in I} C_i$, $\sum_{i=1}^{m} \mu_i = 1$, and $m < \infty$ as $x \in \mathbb{R}^n$. Actually, we can get $C$ just by taking those combinations in which the coefficients are non-zero and vectors are taken from different sets $C_i$. Indeed, vectors with zero coefficients can be omitted from the combination, and if two of the vectors with positive coefficients belong to the same set $C_i$, say $y_1$ and $y_2$, then the term $\mu_1 y_1 + \mu_2 y_2$ in $x$ can be replaced by $\mu y$, where $\mu = \mu_1 + \mu_2$ and $y = (\mu_1/\mu)y_1 + (\mu_2/\mu)y_2 \in C_i$ so that we have

$$x = \mu y + \sum_{i=3}^{m} \mu_i x_i, \quad \mu + \sum_{i=3}^{m} \mu_i = \sum_{i=1}^{m} \mu_i = 1.$$

Thus by Theorem 1.30, $C$ is the union of the finite convex combinations of the form

$$\mu_1 C_{i_1} + \cdots + \mu_m C_{i_m},$$

where the indices $i_1, \ldots, i_m$ are all different. $\square$

**Theorem 1.58.** *Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then $AC$, image of $C$ under $A$, is a convex set in $\mathbb{R}^m$ for every convex set $C$ in $\mathbb{R}^n$, and $A^{-1}D$, the inverse image of $D$ under $A$, is a convex set in $\mathbb{R}^n$ for every convex set $D$ in $\mathbb{R}^m$.*

*Proof.* Let $C$ be a convex subset of $\mathbb{R}^n$ and let $x, y \in C$ so that $Ax, Ay \in AC$. Then for $0 \leq \lambda \leq 1$, we have

$$\lambda Ax + (1 - \lambda)Ay = A(\lambda x + (1 - \lambda)y) \in AC.$$

Now, suppose that $D$ is a convex subset of $\mathbb{R}^m$ and let $A^{-1}x, A^{-1}y \in A^{-1}D$. Then for $0 \leq \lambda \leq 1$, we have

$$\lambda A^{-1}x + (1 - \lambda)A^{-1}y = A^{-1}(\lambda x + (1 - \lambda)y) \in A^{-1}D.$$

$\square$

**Corollary 1.58.1.** *The orthogonal projection of a convex set $C$ on a subspace $L$ is convex.*

**Theorem 1.59.** *Let $C$ and $D$ be convex sets in $\mathbb{R}^m$ and $\mathbb{R}^p$, respectively. Then, the direct sum*

$$C \oplus D = \{x = (y, z) \ : \ y \in C, z \in D\}$$

*is a convex set in $\mathbb{R}^{m+p}$.*

**Remark 1.60.** When dealing with finite collection of sets, the direct sum and the cartesian product of the collection coincide. So the previous result also applies for cartesian product of $C$ and $D$.

**Remark 1.61.** Suppose $C, D \subseteq \mathbb{R}^n$. Then the direct sum $C \oplus D$ coincides with the ordinary sum $C + D$, if each $x \in C + D$ can be expressed uniquely in the form $x = y + z$, $y \in C$ and $z \in D$. This happens iff the symmetric convex sets $C - C$ and $D - D$ have only the zero vector of $\mathbb{R}^n$ in common. (It can be shown that then $\mathbb{R}^n$ may be expressed as a direct sum of two subspaces, one containing $C$ and the other containing $D$.)

**Remark 1.62** (Algebra of Convex cones)**.** The sets $K_1 + K_2$, conv $(K_1 \cup K_2)$, $K_1 \oplus K_2$, $AK$, $A^{-1}K$ and $\lambda K$ are convex cones when $K_1$, $K_2$ and $K$ are convex cones. Positive scalar multiplication is a trivial operation for cones: one has $\lambda K = K$ for every $\lambda > 0$.

**Theorem 1.63.** *If $K_1$ and $K_2$ are convex cones containing origin, then*

$$K_1 + K_2 = conv\ (K_1 \cup K_2).$$

*Proof.* By Theorem 1.57, conv $(K_1 \cup K_2)$ is the union over $\lambda \in [0,1]$ of $(1 - \lambda)K_1 + \lambda K_2$, i.e.,

$$\text{conv}\ (K_1 \cup K_2) = \bigcup_{\lambda \in [0,1]} (1 - \lambda)K_1 + \lambda K_2.$$

The latter set is $K_1 + K_2$ when $0 < \lambda < 1$ (since $\lambda K = K$ for a cone with $\lambda > 0$), $K_1$ when $\lambda = 0$, $K_2$ when $\lambda = 1$. Since $0 \in K_1$ and $0 \in K_2$, $K_1 + K_2$ includes both $K_1$ and $K_2$. Thus conv $(K_1 \cup K_2)$ coincides with $K_1 + K_2$. $\qquad \square$

## 1.4 Convex Functions

**Definition 1.64.** Let $f : S \subseteq \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ be a function. The set

$$\{(x, \mu)\ :\ x \in S, \mu \in \mathbb{R}, \mu \geq f(x)\}$$

is called the **_epigraph_** of $f$ and is denoted by epi $f$. We define $f$ to be a **_convex function_** on $S$ if epi $f$ is convex as a subset of $\mathbb{R}^{n+1}$. A **_concave function_** on $S$ is a function whose negative is convex. An **_affine function_** on $S$ is a function which is finite, convex and concave.

**Definition 1.65.** The **_effective domain_** of a convex function $f$ on $S$, which we denote by dom $f$, is the projection on $\mathbb{R}^n$ of the epigraph of $f$:

$$\text{dom}\ f = \{x\ :\ \exists \mu, (x, \mu) \in \text{epi}\ f\} = \{x\ :\ f(x) < +\infty\}.$$

**Remark 1.66.** Since a convex function $f$ on $S$ can always be extended to a convex function on all of $\mathbb{R}^n$ by setting $f(x) = +\infty$ for $x \notin S$, we shall henceforth refer by a "convex function" a "convex function with possibly infinte values which is defined throughout the space of $\mathbb{R}^n$." This leads to extened arithmetic rules involving $\pm\infty$ where the only forbidden calculations are $\infty - \infty$ and $-\infty + \infty$.

**Definition 1.67.** A convex function $f$ is said to be **_proper_** if its epigraph is non-empty and contains no vertical lines, i.e., $f(x) < +\infty$ for at least one $x$ and $f(x) > -\infty$ for every $x$. A convex function which is not proper is **_improper_**.

**Theorem 1.68.** *Let $f$ be a function from $C$ to $(-\infty, +\infty]$, where $C$ is a convex set. Then $f$ is convex on $C$ iff*

$$f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y), \quad 0 < \lambda < 1,$$

*for every $x, y \in C$.*

*Proof.* By definition, $f$ is convex on $C$ iff

$$(1 - \lambda)(x, \mu) + \lambda(y, \nu) = ((1 - \lambda)x + \lambda y, (1 - \lambda)\mu + \lambda \nu) \in \text{epi}\ f$$

whenever $(x, \mu)$ and $(y, \nu)$ belong to epi $f$ and $0 \leq \lambda \leq 1$. By setting $\mu = f(x)$ and $\nu = f(y)$, we get the required result. $\qquad \square$

**Corollary 1.68.1.** *Let $f$ be a function from $\mathbb{R}^n$ to $[-\infty, +\infty]$. Then $f$ is convex on $C$ iff*

$$f((1-\lambda)x + \lambda y) < (1-\lambda)\alpha + \lambda\beta, \quad 0 < \lambda < 1,$$

*whenever $f(x) < \alpha$ and $f(y) < \beta$.*

**Corollary 1.68.2** (Jensen's Inequality). *Let $f$ be a function from $\mathbb{R}^n$ to $(-\infty, +\infty]$. Then $f$ is convex iff*

$$f(\lambda_1 x_1 + \cdots + \lambda_m x_m) \leq \lambda_1 f(x_1) + \cdots + \lambda_m f(x_m)$$

*whenever $\lambda_1 \geq 0, \ldots, \lambda_m \geq 0, \lambda_1 + \cdots + \lambda_m = 1$.*

**Remark 1.69.** The inequality in Theorem 1.68 is often taken as the definition of the convexity of a function $f$ from a convex set $C$ to $(-\infty, +\infty]$. This approach causes difficulties, however, when $f$ can have both $+\infty$ and $-\infty$ among its values, since the expression $\infty - \infty$ could arise. One could instead use the condition in Corollary 1.68.1 as the definition of convexity in the general case, but the definition given at the beginning of this section seems preferable because it emphasizes the geometry which is fundamental to the theory of convex functions.

**Theorem 1.70.** *Let $f$ be a twice continuously differentiable real-valued function on an open interval $(\alpha, \beta)$. Then $f$ is convex iff its second derivative $f''$ is nonnegative throughout $(\alpha, \beta)$.*

*Proof.* ( $\Longleftarrow$ ) Suppose first that $f''$ is nonnegative on $(\alpha, \beta)$. Then $f'$ is non-decreasing on $(\alpha, \beta)$. For $\alpha < x < y < \beta$, $0 < \lambda < 1$ and $z = (1-\lambda)x + \lambda y$, we have $f'(x) \leq f'(z) \leq f'(y)$ and

$$f(z) - f(x) = \int_x^z f'(t)dt \leq f'(z)(z-x),$$

$$f(y) - f(z) = \int_z^y f'(t)dt \geq f'(z)(y-z).$$

Since $z - x = \lambda(y-x)$ and $y - z = (1-\lambda)(y-x)$, we have

$$f(z) \leq f(x) + \lambda f'(z)(y-x),$$
$$f(z) \leq f(y) - (1-\lambda)f'(z)(y-x).$$

Multiplying the two inequalities by $(1-\lambda)$ and $\lambda$ respectively and adding them together, we get

$$f(z) = f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y).$$

( $\Longrightarrow$ ) Suppose that $f''$ were not nonnegative on $(\alpha, \beta)$. Then $f''$ would be negative on a certain subinterval $(\alpha', \beta')$ by continuity. Then by similar arguments as above, we would have

$$f(z) - f(x) > f'(z)(z-x),$$

$$f(y) - f(z) < f'(z)(y-z),$$

and hence

$$f((1-\lambda)x + \lambda y) > (1-\lambda)f(x) + \lambda f(y).$$

Thus, $f$ would not be convex on $(\alpha, \beta)$. $\qquad\square$

**Example 1.71.** Here are some functions on $\mathbb{R}$ convexity is a consequence of Theorem 1.70.

1. $f(x) = e^{\alpha x}$, where $-\infty < \alpha < +\infty$.

2. $f(x) = x^p$ if $x \geq 0$ and $f(x) = \infty$ if $x < 0$, where $1 \leq p < \infty$.

3. $f(x) = -x^p$ if $x \geq 0$, $f(x) = \infty$ if $x \leq 0$, where $0 \leq p \leq 1$.

4. $f(x) = x^p$ if $x > 0$ and $f(x) = \infty$ if $x \leq 0$, where $p \in (-\infty, 0] \cup [1, \infty)$.

5. $f(x) = (\alpha^2 - x^2)^{-1/2}$ if $|x| < \alpha$, $f(x) = \infty$ if $|x| \geq \alpha$, where $\alpha > 0$.

6. $f(x) = -\log x$ if $x > 0$, $f(x) = \infty$ if $x \leq 0$.

**Theorem 1.72.** *Let $f$ be a twice continuously differentiable real-valued function on an open convex set $C$ in $\mathbb{R}^n$. Then $f$ is convex iff its Hessian matix Hess $f(x)$ is positive semi-definite for every $x \in C$.*

*Proof.* The convexity of $f$ on $C$ is equivalent to the convexity of the restriction of $f$ to each line segment in $C$. This is the same as the convexity of the function $g(\lambda) = f(y + \lambda z)$ on the open real interval $\{\lambda \; : \; y + \lambda z \in C\}$ for each $y \in C$ and $z \in \mathbb{R}^n$. Note that

$$g''(\lambda) = \langle z, \text{Hess } f(y + \lambda z)z \rangle.$$

Thus, by Theorem 1.70, $g$ is convex for each $y \in C$ and $z \in \mathbb{R}^n$ iff $\langle z, \text{Hess } f(y + \lambda z)z \rangle \geq 0$ for every $z \in \mathbb{R}^n$ and $y \in C$. $\qquad\square$

**Definition 1.73.** For a set $C$ in $\mathbb{R}^n$ the **_indicator function_** $\delta_C(x)$ of $C$ is defined as

$$\delta_C(x) = \begin{cases} 0, & \text{if } x \in C \\ +\infty, & \text{if } x \notin C. \end{cases}$$

**Remark 1.74.** The epigraph of the indicator function is $C \times [0, \infty)$ which is a "half-cylinder with cross-section $C$".

**Proposition 1.75.** *A set $C$ in $\mathbb{R}^n$ is a convex set iff $\delta_C$ is a convex function in $\mathbb{R}^n$.*

*Proof.* The set $C \subseteq \mathbb{R}^n$ is convex iff $C \times [0, \infty) \subseteq \mathbb{R}^{n+1}$ is convex by Theorem 1.59 and Remark 1.61. But then epi $\delta_C$ is convex which means that $\delta_C$ is convex. $\qquad\square$

**Definition 1.76.** The **_support function_** $\delta_C^*$ of a convex set $C$ in $\mathbb{R}^n$ is defined by

$$\delta_C^*(x) = \sup\{\langle x, y \rangle \; : \; y \in C\}.$$

**Definition 1.77.** The **_gauge function_** $\gamma_C$ of a convex set $C$ in $\mathbb{R}^n$ is defined by

$$\gamma_C(x) = \inf\{\lambda \geq 0 \; : \; x \in \lambda C\}, \quad C \neq \emptyset.$$

**Definition 1.78.** The (Euclidean) **_distance function_** $d_C$ is defined by

$$d_C(x) = \inf\{\|x - y\| \; : \; y \in C\}.$$

**Remark 1.79.** The convexity of these functions (in Def. 1.76, 1.77 and 1.78) on $\mathbb{R}^n$ could be verified now directly, but we shall wait until the next section, where the convexity can be shown to follow from general principles.

**Theorem 1.80.** *For any convex function $f$ and any $\alpha \in [-\infty, +\infty]$, the level sets $\{x \; : \; f(x) < \alpha\}$ and $\{x \; : \; f(x) \leq \alpha\}$ are convex.*

*Proof.* In the case of strict inequality the result is immediate from Corollary 1.68.1, with $x = y$ and $\beta = \alpha$. The convexity of $\{x \; : \; f(x) \leq \alpha\}$ then follows from the fact that it is the intersection of the convex sets $\{x \; : \; f(x) \leq \mu\}$ for $\mu > \alpha$, i.e.,

$$\{x \; : \; f(x) \leq \alpha\} = \bigcap_{\mu > \alpha} \{x \; : \; f(x) < \mu\}.$$

A more geometric way of seeing this convexity is to observe that $\{x \; : \; f(x) \leq \alpha\}$ is the projection on $\mathbb{R}^n$ of the intersection of epi $f$ and the horizontal hyperplane $\{(x, \mu) \; : \; \mu = \alpha\}$ in $\mathbb{R}^{n+1}$, so that $\{x \; : \; f(x) \leq \alpha\}$ can be regarded as a horizontal cross-section of epi $f$. $\qquad\square$

**Corollary 1.80.1.** *Let $f_i$ be a convex function on $\mathbb{R}^n$ and $\alpha_i$ be a real number for each $i \in I$, where $I$ is an arbitrary index set. Then*

$$C = \{x \ : \ f_i(x) \leq \alpha, \ \forall i \in I\}$$

*is a convex set.*

*Proof.* Like Corollary 1.24.1. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

**Definition 1.81.** A function $f$ on $\mathbb{R}^n$ is said to be ***positively homogeneous*** (of degree 1) if for every $x$ one has

$$f(\lambda x) = \lambda f(x), \quad 0 < \lambda < \infty.$$

**Remark 1.82.** Positive homogeneity is equivalent to the epigraph being a cone in $\mathbb{R}^{n+1}$.

**Theorem 1.83.** *A positively homogeneous function $f$ from $\mathbb{R}^n$ to $(-\infty, +\infty]$ is convex iff*

$$f(x + y) \leq f(x) + f(y) \tag{4}$$

*for every $x, y \in \mathbb{R}^n$.*

*Proof.* This is implied by Theorem 1.38, because the subadditivity condition of $f$ in (4) is equivalent to epi $f$ being closed under addition. $\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

**Corollary 1.83.1.** *If $f$ is a positively homogeneous proper convex function, then*

$$f(\lambda_1 x_1 + \cdots + \lambda_m x_m) \leq \lambda_1 f(x_1) + \cdots + \lambda_m f(x_m)$$

*whenever $\lambda_1 > 0, \ldots, \lambda_m > 0$.*

**Corollary 1.83.2.** *If $f$ is a positively homogeneous proper convex function, then $f(-x) \geq -f(x)$ for every $x$.*

*Proof.* We have $f(x) + f(-x) \geq f(x - x) = f(0) \geq 0$. $\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

**Theorem 1.84.** *A positively homogeneous proper convex function $f$ is linear on a subspace $L$ iff $f(-x) = -f(x)$ for every $x \in L$. The converse is true if merely $f(-b_i) = -f(b_i)$ for all the vectors in some basis $b_1, \ldots, b_m$ for $L$.*

*Proof.* ( $\implies$ ) Linearity of $f$ implies $f(-x) = -f(x)$ for every $x \in L$.
( $\impliedby$ ) Assume that $f$ is positively homogeneous proper convex function and that $f(-b_i) = -f(b_i)$ for all the vectors in some basis $b_1, \ldots, b_m$ for $L$. Then $f(\lambda_i b_i) = \lambda_i f(b_i)$ for every $\lambda_i \in \mathbb{R}$, not just for $\lambda_i > 0$. For any $x = \lambda_1 b_1 + \cdots + \lambda_m b_m \in L$ we have

$$f(\lambda_1 b_1) + \cdots + f(\lambda_m b_m) \geq f(x) \geq -f(-x) \geq -(f(-\lambda_1 b_1) + \cdots + f(-\lambda_m b_m))$$
$$= f(\lambda_1 b_1) + \cdots + f(\lambda_m b_m),$$

and hence

$$f(x) = f(\lambda_1 b_1) + \cdots + f(\lambda_m b_m) = \lambda_1 f(b_1) + \cdots + \lambda_m f(b_m).$$

Thus $f$ is linear on $L$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

## 1.5 Functional Operations

**Theorem 1.85.** *Let $f$ be a convex function from $\mathbb{R}^n$ to $(-\infty, +\infty]$, and let $\varphi$ be a convex function from $\mathbb{R}$ to $(-\infty, +\infty]$ which is non-decreasing. Then $h(x) = \varphi(f(x))$ is convex on $\mathbb{R}^n$ (where one sets $\varphi(+\infty) = +\infty$).*

*Proof.* For any $x, y \in \mathbb{R}^n$ and $0 < \lambda < 1$, from Theorem 1.68 we have

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y).$$

Applying $\varphi$ to both sides of this inequality, we retain the inequality sign due to the non-decreasing property of $\varphi$ to get

$$h((1-\lambda)x + \lambda y) \leq \varphi((1-\lambda)f(x) + \lambda f(y)) \leq (1-\lambda)h(x) + \lambda h(y).$$

$\square$

**Example 1.86.** Motivated from Example 1.71 we give some examples of convex function from $\mathbb{R}^n$ to $\mathbb{R}$ as a consequence of Theorem 1.85.

1. $h(x) = e^{f(x)}$ is a proper convex function when $f$ is proper convex.

2. $h(x) = f(x)^p$ is convex for $p \geq 1$ when $f$ is convex and nonnegative.

3. $h(x) = 1/g(x)$ is convex on $C = \{x \ : \ g(x) > 0\}$ when $g$ is concave. To see this, take $f = -g$ and $\varphi(\xi) = -1/\xi$ for $\xi < 0$, and $\varphi(\xi) = +\infty$ for $\xi \geq 0$.

4. $h(x) = \lambda f(x) + \alpha$ is a proper convex function when $f$ is proper convex and $\lambda > 0$.

**Theorem 1.87.** *If $f_1$ and $f_2$ are proper convex functions on $\mathbb{R}^n$, then $f_1 + f_2$ is convex.*

**Remark 1.88.** The properness in the hypothesis of Theorem 1.87 is to avoid $\infty - \infty$ when $f_1 + f_2$ is formed. Also note that the dom $f_1 + f_2 = $ dom $f_1 \cap$ dom $f_2$, where dom $f$ is the effective domain defined in Def. 1.65. Thus, $f_1 + f_2$ is improper is when the intersection of the effective domains is empty.

**Corollary 1.88.1.** *A linear combination $\lambda_1 f_1 + \cdots + \lambda_m f_m$ of proper convex functions with nonnegative coefficients is convex.*

**Theorem 1.89.** *Let $F$ be any convex set in $\mathbb{R}^{n+1}$, and let*

$$f(x) = \inf\{\mu \ : \ (x, \mu) \in F\}.$$

*Then $f$ is a convex function on $\mathbb{R}^n$.*

*Proof.* Let $(x, \mu), (y, \gamma) \in F$ so that $((1-\lambda)x + \lambda y, (1-\lambda)\mu + \lambda\gamma) \in F$. This implies that

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)\mu + \lambda\gamma, \tag{5}$$

such that $f(x) \leq \mu$ and $f(y) \leq \gamma$. Then by Cor. 1.68.1, $f$ is convex. $\square$

**Remark 1.90.** It is noteworthy to compare epi $f$ with $F$. Now suppose $(x, \mu) \in F$, then $f(x) \leq \mu$ by definition of $f$, so that $(x, \mu) \in$ epi $f$. This shows that $F \subseteq$ epi $f$.

**Proposition 1.91.** *Let $f$ be a function on $\mathbb{R}^n$. Then*

$$f(x) = \inf\{\mu \ : \ (x, \mu) \in epi \ f\}.$$

*Proof.* We need to show that $f(x)$ is the greatest lower bound of the set $M = \{\mu \ : \ (x, \mu) \in \text{epi } f\}$. Note that for all $(x, \mu) \in \text{epi } f$, we have $f(x) \leq \mu$ from the definition of the epigraph. For any $\epsilon > 0$, we have $\mu = f(x) \in M$ such that $\mu - \epsilon < f(x)$ showing that $f(x)$ is the greatest lower bound of $M$. $\qquad\square$

**Definition 1.92.** The ***infimal convolution*** of $f_1, \ldots, f_m$ (functions on $\mathbb{R}^n$) is the function on $\mathbb{R}^n$ defined as

$$(f_1 \square \cdots \square f_m)(x) = \inf\{f_1(x_1) + \cdots + f_m(x_m) \ : \ x_i \in R^n, x_1 + \cdots + x_m = x\}.$$

**Remark 1.93.** This terminology arises from the fact that, when only two functions are involved, $\square$ can be expressed by

$$(f \square g)(x) = \inf_y \{f(x - y) + g(y)\},$$

and this is analogous to the classical formula for integral convolution.

**Proposition 1.94.** *Let $f_1, \ldots, f_m$ be proper convex functions on $\mathbb{R}^n$, then $f_1 \square \cdots \square f_m$ is a convex function on $\mathbb{R}^n$.*

*Proof.* Let $F_i = \text{epi } f_i$ and $F = F_1 + \cdots + F_m$. Then $F$ is a convex set in $\mathbb{R}^{n+1}$. By definition, $(x, \mu) \in F$ iff there exists $x_i \in \mathbb{R}^n$, $\mu_i \in \mathbb{R}$, such that $\mu_i \geq f_i(x_i)$, $\mu = \mu_1 + \cdots + \mu_m$ and $x = x_1 + \cdots + x_m$. Thus $f$ defined in the theorem is the convex function obtained from $F$ by the construction in Theorem 1.89. $\qquad\square$

**Remark 1.95.** We note few properties of infimal convolution operator below. Consider two function $f, g$ on $\mathbb{R}^n$.

- If $g = \delta_{\{a\}}$ for a certain poinr $a \in \mathbb{R}^n$, then $(f \square \delta_{\{a\}})(x) = f(x - a)$, which is the function whose graph is obtained by translating the graph of $f$ horizontally by $a$.

- The effective domain of $f \square g$ is the sum of dom $f$ and dom $g$.

- Taking $f$ to be the Euclidean norm and $g$ to be the indicator function of a convex set $C$, we get
$$(f \square g)(x) = \inf_y \{\|x - y\| + \delta_C(y)\} = \inf_{y \in C} \|x - y\| = d_C(x).$$
This establishes the convexity of the distance function $d_C$ defined in Def. 1.78.

- Properness of convex functions is not always preserved by infimal convolution, since the infimum in the definition may be $-\infty$. Nor is infimal convolution of improper functions defined by this formula, because of the rule of avoiding $\infty - \infty$. However, $f_1 \square f_2$ can be defined for any functions $f_1$ and $f_2$ from $\mathbb{R}^n$ to $[-\infty, +\infty]$ directly in terms of addition of epigraphs:
$$(f_1 \square f_2)(x) = \inf\{\mu \ : \ (x, \mu) \in (\text{epi } f_1 + \text{epi } f_2)\}.$$

- As an operation on the collection of all functions from $\mathbb{R}^n$ to $[-\infty, +\infty]$, infimal convolution is commutative, associative and convexity-preserving. The function $\delta_{\{0\}}$ acts as the identity element for this operation.

**Definition 1.96.** For any function $f$ on $\mathbb{R}^n$ and $0 \leq \lambda < \infty$ we define its ***right scalar multiple***, denoted by $f\lambda$, to be the following function

$$(f\lambda)(x) = \inf\{\mu \ : \ (x, \mu) \in \lambda(\text{epi } f)\}.$$

**Remark 1.97.** We note few properties of the right scalar multiple operator below. Let $f$ be a function on $\mathbb{R}^n$ and $0 \leq \lambda < \infty$.

- If $f$ is convex, then $\lambda(\text{epi } f)$ is a convex set. From the definition of $f\lambda$ and Theorem 1.89, we see that $f\lambda$ is convex.

- From the definition, when $\lambda > 0$ we have

$$
\begin{aligned}
(f\lambda)(x) &= \inf\{\mu \ : \ (x,\mu) \in \lambda(\text{epi } f)\} \\
&= \inf\{\lambda(\mu/\lambda) \ : \ (x/\lambda, \mu/\lambda) \in (\text{epi } f)\} \\
&\overset{(\dagger)}{=} \lambda f(\lambda^{-1}x),
\end{aligned} \tag{6}
$$

where we used Prop. 1.91 to obtain ($\dagger$). When $\lambda = 0$, $(f0)(x) = \delta_{\{0\}}(x)$ for $f \not\equiv +\infty$, otherwise $f0 = f$, for $f \equiv +\infty$.

- Consider the epigraph of $f\lambda$ for $\lambda > 0$

$$
\begin{aligned}
\text{epi } (f\lambda) &= \{(x,\mu) \ : \ \mu \geq (f\lambda)(x)\} \\
&= \{(x,\mu) \ : \ \mu \geq \lambda f(\lambda^{-1}x)\} \\
&= \lambda\{(\lambda^{-1}x, \lambda^{-1}\mu) \ : \ \lambda^{-1}\mu \geq f(\lambda^{-1}x)\} \\
&= \lambda(\text{epi } f).
\end{aligned} \tag{7}
$$

- The function $f$ is positively homogeneous iff $f\lambda = f$ for each $\lambda > 0$. This is evident from (6) and Def. 1.81.

**Remark 1.98.** Let $h$ be a convex function on $\mathbb{R}^n$ and define a function $f$ as follows:

$$
f(x) = \inf\{\mu \ : \ (x,\mu) \in \text{cone}(\text{epi } h)\}.
$$

Then epi $f \overset{(?)}{=}$ cone (epi $h$) $= \bigcup_{\lambda \geq 0} \lambda(\text{epi } h)$. Thus, epi $f$ is a convex cone in $\mathbb{R}^{n+1}$ containing origin. This shows that $f$ is positively homogeneous convex function with the property that $f(0) \leq 0$ and $f(x) \leq h(x)$ for all $x \in \text{dom } h$. From the definition of $f$ it is also true that $f$ is the *greatest* such function satisfying these conditions. We call this $f$ to be the ***positively homogeneous convex function generated by h***. Since cone (epi $h$) $= \bigcup_{\lambda \geq 0} \lambda(\text{epi } h)$, using the right scalar multiple definition here we also have

$$
f(x) = \inf\{(h\lambda)(x) \ : \ \lambda \geq 0\}, \quad \text{when } h \not\equiv +\infty.
$$

Consider the function $h(x) = \delta_C(x) + 1$ for a non-empty convex set in $\mathbb{R}^n$. Then the positively homogeneous convex function generated by $h$ at $x \in R^n$ is

$$
\inf\{(h\lambda)(x) \ : \ \lambda \geq 0\} = \inf\{\delta_{\lambda C}(x) + \lambda \ : \ \lambda \geq 0\} = \inf\{\lambda \geq 0 \ : \ x \in \lambda C\} = \gamma_C(x).
$$

Thus the gauge function defined in Def. 1.77 is positively homogeneous convex function on $\mathbb{R}^n$.

**Theorem 1.99.** *The pointwise supremum of an arbitrary collection of convex functions is convex.*

*Proof.* Let $f_i$ be an arbitrary collection of convex sets indexed by $I$. Then the pointwise supremum of this collection is defined as

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

It is easy to set that epi $f = \bigcap_{i \in I} \text{epi } f_i$. Then by the definition of convexity and Lemma 1.37, $f$ is convex. $\qquad\square$

**Remark 1.100.** Here are few applications of Theorem 1.99.

- The convexity of the support function $\delta_C^*$ defined in Def. 1.76 is implied by Theorem 1.99, because this function is by definition the pointwise supremum of the collection of linear function $\langle \cdot, y \rangle$ as $y$ ranges over $C$.

- Consider the function $f$ which assigns to each $x = (\xi_1, \ldots, \xi_n)$ the greatest of the components $\xi_i$ of $x$. This function is convex by Theorem 1.99, because it is the pointwise supremum of the linear function $\langle x, e_j \rangle$, $j = 1, \ldots, n$, where $e_j$ is the $j$th column of the $n \times n$ identity matrix. The function $f$ is also positively homogeneous and in fact $f$ is the support function of the simplex

$$C = \{y = (\eta_1, \ldots, \eta_n) \ : \ \eta_j \geq 0, \eta_1 + \cdots + \eta_n = 1\}.$$

  as for an $x \in \mathbb{R}^n$, $\langle x, y \rangle \leq \max\{\xi_j \ : \ j = 1, \ldots, n\}$ which is attained by setting $y = e_j$ when $\xi_j$ is maximum.

- The **_Tchebycheff norm_** or the $\ell_\infty$ norm on $\mathbb{R}^n$ is defined as

$$k(x) = \max\{|\xi_j| \ : \ j = 1, \ldots, n\},$$

  which is a convex function by Theorem 1.99. This function is the support function of the convex set

$$D = \{y = (\eta_1, \ldots, \eta_n) \ : \ |\eta_1| + \cdots + |\eta_n| = 1\},$$

  which can been following similar arguments to the previous case and at the same time the gauge of the $n$-dimensional cube

$$E = \{x = (\xi_1, \ldots, \xi_n) \ : \ -1 \leq \xi_j \leq 1, j = 1, \ldots, n\}.$$

  This can be seen by noting that

$$\inf\{\lambda \geq 0 \ : \ |\xi_j| \leq \lambda, \ j = 1, \ldots, n\} = \max\{|\xi_j| \ : \ j = 1, \ldots, n\} = k(x).$$

**Definition 1.101.** The convex hull of a non-convex function $g$, denoted by conv $g$, is the _greatest convex function majorized by_ $g$ which is defined as

$$\text{conv } g = \inf\{\mu \ : \ (x, \mu) \in \text{conv}(\text{epi } g)\}.$$

The convex hull of an arbitrary collection of functions $\{f_i \ : \ i \in I\}$ on $\mathbb{R}^n$ is the _greatest convex function_ $f$ defined as

$$f(x) = \inf\{\mu \ : \ (x, \mu) \in \text{conv}(\cup_{i \in I} \text{ epi } f_i)\}.$$

**Remark 1.102.** Another way to represent the convex hull of $g$ is as follows. By Theorem 1.30 $(x, \mu) \in \text{conv}(\text{epi } g)$ iff there exist $(x_i, \mu_i) \in \text{epi } g$ and $\lambda_i \geq 0$ for $i = 1, \ldots, m$ such that

$$(x, \mu) = \lambda_1(x_1, \mu_1) + \cdots + \lambda_m(x_m, \mu_m) = (\lambda_1 x_1 + \cdots + \lambda_m x_m, \lambda_1 \mu_1 + \cdots + \lambda_m \mu_m),$$

and $\lambda_1 + \cdots + \lambda_m = 1$. Then

$$\text{conv } g = \inf\{\lambda_1 g(x_1) + \cdots + \lambda_m g(x_m) \ : \ \lambda_1 x_1 + \cdots + \lambda_m x_m = x, \lambda_1 + \cdots + \lambda_m = 1\},$$

provided $g$ does not take on the value $-\infty$, so that the summation is unambiguous.

**Theorem 1.103.** _Let $\{f_i \ : \ i \in I\}$ be an arbitrary collection of proper convex functions on $\mathbb{R}^n$, and let $f$ be the convex hull of the collection. Then_

$$f(x) = \inf\left\{\sum_{i \in I} \lambda_i f_i(x_i) \ : \ \sum_{i \in I} \lambda_i x_i = x\right\},$$

_where the infimum is taken over all representation of $x$ as a convex combinations of elements $x_i$, such that only finitely many coefficients $\lambda_i$ are non-zero. (The formula is also valid if one actually restricts $x_i$ to lie in dom $f_i$.)_

*Proof.* By definition $f(x)$ is the infimum of the values of $\mu$ such that $(x, \mu) \in F$, where $F$ is the convex hull of the union of convex sets $C_i = \text{epi } f_i$. Each $C_i$ is non-empty as they are assumed to be proper. By Theorem 1.57, $(x, \mu) \in F$ iff $(x, \mu)$ can be expressed as a finite combination of the form

$$(x, \mu) = \sum_{i \in I} \lambda_i(x_i, \mu_i) = \left( \sum_{i \in I} \lambda_i x_i, \sum_{i \in I} \lambda_i \mu_i \right),$$

where $(x_i, \mu_i) \in C_i$ (only finitely many of the coefficients being non-zero). Thus $f(x)$ is the infimum of $\sum_{i \in I} \lambda_i \mu_i$ over all expressions of $x$ as a finite convex combinations $\sum_{i \in I} \lambda_i x_i$ with $\mu_i \geq f_i(x_i)$ for every $i$. This is same as the infimum in the theorem statement. $\square$

**Remark 1.104.** Similar to Remark 1.56 for convex sets, the collection of all convex functions on $\mathbb{R}^n$, regarded as a partially ordered set relative to the pointwise ordering (where $f \geq g$ iff $f(x) \geq g(x)$ for every $x$), is a complete lattice. The greatest lower bound of a family of convex functions is the convex hull of the collection, while the least upper bound is $\sup\{f_i : i \in I\}$.

**Theorem 1.105.** *Let $A$ be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. Then, for each convex function $g$ on $\mathbb{R}^m$, the function $gA$ defined by*

$$(gA)(x) = g(Ax)$$

*is convex on $\mathbb{R}^n$. For each convex function $h$ on $\mathbb{R}^n$, the function $Ah$ defined by*

$$(Ah)(y) = \inf\{h(x) : Ax = y\}$$

*is convex on $\mathbb{R}^m$.*

*Proof.* Let $x, y \in \mathbb{R}^n$ and $0 \leq \lambda \leq 1$. Then

$$(gA)((1 - \lambda)x + \lambda y) = g((1 - \lambda)Ax + \lambda Ay) \overset{(\dagger)}{=} (1 - \lambda)(gA)(x) + \lambda(gA)(y),$$

where we used the convexity of $g$ to obtain $(\dagger)$. This shows the convexity of $gA$.
The convexity of $f = Ah$ also follows from applying Theorem 1.89 to the image $F$ of the epigraph of $h$ under the linear transformation $(x, \mu) \to (Ax, \mu)$ from $\mathbb{R}^{n+1}$ to $\mathbb{R}^{m+1}$. $\square$

**Remark 1.106.** The function $Ah$ in Theorem 1.105 is called the ***image of h under A***, while $gA$ is called the ***inverse image of g under A***. This terminology is suggested by the case where $g$ and $h$ are indicators of convex sets.

**Proposition 1.107.** *Let $h$ be any map on $\mathbb{R}^n$ and $A$ is a linear transformation on $\mathbb{R}^n$. If $A$ is non-singular, then $Ah = hA^{-1}$.*

*Proof.* Consider the expressions of $Ah$

$$(Ah)(y) = \inf\{h(x) : Ax = y\} = \inf\{h(A^{-1}y)\} \overset{(\dagger)}{=} h(A^{-1}y) = (hA^{-1})(y),$$

where we used the uniqueness of $A^{-1}y$ in $(\dagger)$. $\square$

## Convex Cones - Fenchel Notes

### Cones

**Definition 1.108.** A subset $M$ of $\mathbb{R}^n$ is called a ***cone*** if 0 is in $M$ and $x \in M$ implies $\lambda x \in M$ for every non-negative real scalar $\lambda$. The particular cones consisting of a non-zero vector $x$ and all its multiples $\lambda x$ ($\lambda \geq 0$) is a ***ray***, denoted by $(x)$.

**Remark 1.109.** A cone which contains atleast one non-zero vector is therefore just the union of the rays it contains.

Since all non-trivial cones may be thought of as set of rays, it is desirable to introduce a metric-topology on the set of all rays from the metric-topology of $\mathbb{R}^n$ induced from the canonical metric $d(\cdot, \cdot)$.

**Remark 1.110.** Let $S^n = \{x \in \mathbb{R}^n : \|x\| = 1\}$ be the unit sphere in $\mathbb{R}^n$. Then for any $x \in \mathbb{R}^n$, we have

$$(x) \text{ is a ray in } \mathbb{R}^n \Leftrightarrow \frac{x}{\|x\|} \text{ is in } S^n.$$

Thus, from here on we will represent $S^n$ to be the set of all rays in $\mathbb{R}^n$.

**Proposition 1.111.** *Suppose for any $x, y \in S^n$, we define the map $[\cdot, \cdot] : S^n \times S^n \to \mathbb{R}$ as*

$$[x, y] = \sqrt{2(1 - x^T y)},$$

*which is the chord distance between $x$ and $y$. Then, $[\cdot, \cdot]$ is a metric on $S^n$ and consequently, $(S^n, [\cdot, \cdot])$ is a metric space.*

*Proof.* Note that

$$\begin{aligned}
[x, y] &= \sqrt{2(1 - x^T y)} \\
&= \sqrt{1 + 1 - 2x^T y} \\
&= \sqrt{\|x\|^2 + \|y\|^2 - 2x^T y} \\
&= \|x - y\| = d(x, y).
\end{aligned}$$

Then, $[\cdot, \cdot]$ is a restriction of $d(\cdot, \cdot)$ on $S^n$, in other words, $[\cdot, \cdot] = d(\cdot, \cdot)|_{S^n}$. □

**Definition 1.112.** We review definitions regarding the metric topology induced by $[\cdot, \cdot]$ on $S^n$.

(a) A ***sequence of rays*** $(x^\nu)$ is said to ***converge*** to a ray $(x)$ if $[x^\nu, x] \to 0$.

(b) A ray $(x)$ is called a ***limit ray*** of a cone $M$ if there is a sequence of rays of the cone which are different from $(x)$ and which converges to $(x)$.

(c) A ***closed cone*** or a closed set of rays is a cone which contains all its limit points.

(d) For any $\varepsilon > 0$, the ***$\varepsilon$-neighborhood*** of $(x)$ $\mathbb{B}_\varepsilon(x)$ is the set of all rays $(y)$ with $[x, y] < \varepsilon$.

(e) $M$ is ***open*** if for every $(x) \in M$, there is an $\varepsilon > 0$, such that $\mathbb{B}_\varepsilon(x) \subseteq M$.

(f) A ray $(x)$ is called an ***interior ray*** of a cone $M$ if $M$ contains an $\varepsilon$-neighborhood of $(x)$ for some $\varepsilon > 0$.

(g) A ray $(x)$ such that the complementary cone to the cone $M$ contains a neighborhood of $(x)$ is called an ***exterior ray*** of $M$.

(h) A ***boundary ray*** of a cone $M$ is a limit ray of $M$ which is not an interior ray of $M$.

**Lemma 1.113.** *A set $M \subseteq S^n$ is closed in $S^n \subseteq \mathbb{R}^n$ iff $M$ is closed in $\mathbb{R}^n$.*

*Proof.* ($\Longleftarrow$) If $M$ is closed in $\mathbb{R}^n$, then it is closed in $S^n$ by subspace topology.
($\Longrightarrow$) First we show that $S^n$ itself is closed in $\mathbb{R}^n$. Consider the norm $F : \mathbb{R}^n \to \mathbb{R}$ defined as $F(x) = \|x\|$. Since $F$ is continuous and $S^n = F^{-1}(\{1\})$, i.e., $S^n$ is the preimage of the closed set $\{1\}$ under $F$, $S^n$ is closed in $\mathbb{R}^n$. Suppose if $M$ is closed in $S^n$, then there exists a closed set $M'$ in $\mathbb{R}^n$ such that $M = S^n \cap M'$, which is intersection of two closed sets in $\mathbb{R}^n$. Hence $M$ is closed in $\mathbb{R}^n$ as well. □

**Proposition 1.114.** *An alternate definition of open and closed cones are given below.*

*(a) A cone is closed in the sense of Def. 1.112(c) iff it is closed in $\mathbb{R}^n$.*

*(b) A cone is open in the sense of Def. 1.112(e) iff it is open in $\mathbb{R}^n$.*

*Proof.* (a) It follows from Lemma 1.113 and the fact that we identified the set of all rays in $\mathbb{R}^n$ with $S^n$.
   (b) □

With any cone $M$ there is associated a smallest linear subspace of $\mathbb{R}^n$ which contains $M$.

**Definition 1.115.** The intersection of all subspaces containing the cone $M$ is denoted by $S(M)$. The dimension $d(M)$ of the space $S(M)$ is called ***linear dimension*** of the cone $M$.

For the results on cones in later sections, open or closed relative to $S(M)$, and interior, exterior, and boundary rays relative to $S(M)$ will be considered rather than their counterparts in the topology of the full space $\mathbb{R}^n$. They will be called for simplicity relative interior, relative exterior, relative boundary rays.

## Convex Cones

**Definition 1.116.** A cone $C$ is ***convex*** if the ray $(x+y)$ is in $C$ whenever $(x)$ and $(y)$ are rays in $C$. The largest subspace $s(C)$ contained in a convex cone $C$ is called the ***lineality space*** of $C$ and the dimension $l(C)$ of $s(C)$ is called the lineality of $C$.

**Remark 1.117.** A set $C$ of vectors is a convex cone iff it contains all vectors

$$\lambda x + \mu y \ (\lambda, \mu \geq 0; x, y \in C).$$

**Lemma 1.118.** *If $(x)$ is an interior ray of a convex cone $C$ relative to $S(C)$ and $(y)$ is a boundary or interior ray of $C$ relative to $S(C)$, then every ray $(\lambda x + \mu y)$, where $\lambda$ and $\mu$ are positive real numbers, is an interior ray of $C$ relative to $S(C)$.*

# References

[1] Rockafellar, R.T. Convex Analysis. Princeton University Press, 1972.

[2] Fenchel, W. Convex Cones, Sets, and Functions. Princeton University, 1953.