

Convex Analysis

Jayadev Naram

June 6, 2022

Contents

1	Affine Sets	1
2	Convex Sets and Cones	6
3	The Algebra of Convex Sets	9
4	Convex Functions	12
5	Functional Operations	16

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 $x = [x_1, \dots, x_n]^T$ and $y = [y_1, \dots, y_n]^T$ 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

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

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

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

and hence

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

Thus M is also closed under addition and is a subspace. □

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 $a \in \mathbb{R}^n$. For any points $x, y \in M$, $x + a, y + a \in L$, then we have

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

□

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 $M, L, N \subseteq \mathbb{R}^n$ and $a, b \in \mathbb{R}^n$.

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

□

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 $L = M - M$. □

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, \dots, b_m be a basis for L^\perp , the orthogonal complement of L . Then

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

where B is the $m \times n$ matrix whose rows are b_1, \dots, 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 $b = Ba$. □

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 β 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. □

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

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

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

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 $i = 1, \dots, n$. 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, $\text{aff } S \subseteq C$.

(\impliedby) Suppose $x \in C$, so that

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

We show that $x \in \text{aff } S$ by induction. The base case $m = 2$ holds from the definition of affine set. Assume that for $m > 2$, x belongs to $\text{aff } S$. Then consider the case of m -point combination

in x . There is some index j such that $\lambda_j \neq 1$, otherwise $\lambda_1 + \dots + \lambda_m = m \neq 1$ 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 $y = \sum_{i=2}^m \frac{\lambda_i}{1 - \lambda_1} x_i \in \text{aff } S$ by the induction hypothesis and similarly $x = \lambda_1 x_1 + (1 - \lambda_1)y \in \text{aff } S$. \square

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

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

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

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

where $L = \text{aff}\{0, b_1 - b_0, \dots, b_m - b_0\}$ 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. \square

Definition 1.15. A map $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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 $a \in \mathbb{R}^m$.

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.

$$A(\lambda x) = A((1 - \lambda)0 + \lambda x) = (1 - \lambda)A0 + \lambda Ax = \lambda Ax, \quad (1)$$

and

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

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. \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

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

□

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

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

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. □

Theorem 1.19. Let $\{b_0, b_1, \dots, b_m\}$ and $\{b'_0, b'_1, \dots, b'_m\}$ 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 $i = 0, \dots, m$. 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 $b_1 - b_0, \dots, b_n - b_0$ of \mathbb{R}^n onto the basis $b'_1 - b'_0, \dots, b'_n - b'_0$. The desired affine transformation is then given by $Tx = Ax + a$, where $a = b'_0 - Ab_0$. □

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. □

Proposition 1.20. The graph of an affine transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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 $a \in \mathbb{R}^m$. 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} . □

2 Convex Sets and Cones

Definition 2.1. If x and y are different points in \mathbb{R}^n , the set of points of the form

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

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 $0 \leq \lambda \leq 1$.

Remark 2.2. Affine sets are convex sets but not all convex sets are affine sets. For example consider solid cubes in \mathbb{R}^3 .

Example 2.3 (Half-spaces). For any non-zero $b \in \mathbb{R}^n$ and any $\beta \in \mathbb{R}$, the sets

$$\{x : \langle x, b \rangle \leq \beta\}, \quad \{x : \langle x, b \rangle \geq \beta\},$$

are called closed half-spaces. The sets

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

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

Lemma 2.4. *The intersection of an arbitrary collection of convex sets is convex.*

Proof similar to that of Lemma 1.10.

Corollary 2.4.1. *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 2.4.1 will be generalized by Corollary 4.17.1.

Remark 2.5. 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 2.6. 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 2.7. A vector sum

$$\lambda_1 x_1 + \cdots + \lambda_m x_m$$

is called a convex combination of x_1, \dots, x_m if the coefficients λ_i are all nonnegative and $\lambda_1 + \cdots + \lambda_m = 1$.

Definition 2.8. 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 $\text{conv } C$.

Remark 2.9. The convex hull of S is a convex set by Lemma 2.4, the unique smallest one containing S .

Theorem 2.10. *For any $S \subseteq \mathbb{R}^n$, $\text{conv } S$ 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 2.10.1. A subset C of \mathbb{R}^n is convex iff C contains all the convex combinations of its elements.

Corollary 2.10.2. The convex hull of a finite subset $\{b_0, \dots, b_m\}$ of \mathbb{R}^n consists of all the vectors of the form $\lambda_0 b_0 + \dots + \lambda_m b_m$ with $\lambda_i \geq 0$, $i = 0, \dots, m$, $\lambda_0 + \dots + \lambda_m = 1$.

Definition 2.11. A set which is the convex hull of finitely many points is called a **polytope**. If $\{b_0, \dots, b_m\}$ is affinely independent, its convex hull is called an **m -dimensional simplex**, and b_0, \dots, b_m are called the **vertices** of the simplex.

Definition 2.12. The **dimension of a convex set** C is defined to be the dimension of the affine hull of C .

Remark 2.13. The dimension of an affine hull or simplex as already defined agrees with its dimension as a convex set.

Theorem 2.14. 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 $\{b_0, \dots, b_m\}$ be such a set with m maximal, and let M be its affine hull. Then $\dim M = m$ and $M \subseteq \text{aff } C$. Furthermore, $C \subseteq M$, for if $C - M$ contained an element b , the set of $m + 2$ elements b_0, \dots, b_m, b in C would be affinely independent, contrary to the maximality of m . Since $\text{aff } C$ is the smallest affine set which includes C , it follows that $\text{aff } C \subseteq M$, or equivalently $\text{aff } C = M$ and hence $\dim C = m$. \square

Definition 2.15. 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 2.16 (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

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

and the **positive orthant**

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

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.,

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

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

Lemma 2.17. The intersection of an arbitrary collection of convex cones is a convex cone.

Corollary 2.17.1. 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 2.5 hold for the above corollary. The following characterization of convex cones highlights an analogy between convex cones and subspaces.

Theorem 2.18. 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 λy belong to K for $x, y \in K$, and hence $(1 - \lambda)x + \lambda y \in K$. \square

Corollary 2.18.1. *A subset of \mathbb{R}^n is a convex cone iff it contains all the positive linear combinations of its elements.*

Corollary 2.18.2. *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 2.18.3. *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 2.19. *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 2.18, 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 $\lambda_1 z_1 + \lambda_2 z_2 \in K - K$. By Cor. 2.18.1, $K - K$ 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 $\text{aff } K$ 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 2.17 $(-K) \cap K$ 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 . \square

Definition 2.20. 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 2.21. 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 2.22. 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 simplifying notation in the following proposition. But it will be used more than that in context of Variational Analysis.

Proposition 2.23. 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 2.18. □

Definition 2.24. 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 2.25. 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 . □

3 The Algebra of Convex Sets

Theorem 3.1. 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

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

□

Remark 3.2 (Symmetric Convex Set). The **symmetric reflection** of C across origin is $-C = (-1)C$. A convex set is 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 2.19).

Theorem 3.3. 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

$$(1 - \lambda)(x_1 + x_2) + \lambda(x'_1 + x'_2) = ((1 - \lambda)x_1 + \lambda x'_1) + ((1 - \lambda)x_2 + \lambda x'_2) \in C_1 + C_2.$$

□

Example 3.4. If C_1 is any convex set and C_2 is the nonnegative orthant (see Example 2.16), 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 3.3 when C_1 is convex.

Theorem 3.5. A set C is a convex set iff $(1 - \lambda)C + \lambda C \subseteq C$, where $0 \leq \lambda \leq 1$.

Proposition 3.6. Let C_1, \dots, C_m be convex sets, then so is the linear combination

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

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

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

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

□

Remark 3.7 (Set Algebra). The following algebraic laws are valid for the addition and scalar multiplication of sets even without convexity being involved:

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

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 3.8. 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. (\implies) 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

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

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

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

hence $D \subseteq C$. On multiplying through by $\lambda_1 + \lambda_2$, we get $\lambda_1 C + \lambda_2 C \subseteq (\lambda_1 + \lambda_2)C$.

(\impliedby) The distributive law implies that $\lambda C + (1 - \lambda)C \subseteq C$ when $0 \leq \lambda \leq 1$. Then by Theorem 3.5, this implies that C is a convex set. □

Remark 3.9. It follows from this theorem, for instance, that $C + C = 2C$, $C + C + C = 3C$, and so forth, when C is convex.

Proposition 3.10. *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 $\text{conv}(\bigcup_{i \in I} C_i)$.*

Remark 3.11. 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 3.12. *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., $C = \text{conv}(\bigcup_{i \in I} C_i)$. 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 λ_i such that only finitely many are non-zero and these add up to 1.)

Proof. By Theorem 2.10, C is the set of all convex combinations

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

where $y_1, \dots, y_m \in \bigcup_{i \in I} C_i$, $\sum_{i=1}^m \mu_i = 1$, 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 μy , 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 2.10, 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, \dots, i_m are all different. □

Theorem 3.13. *Let $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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

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

□

Corollary 3.13.1. *The orthogonal projection of a convex set C on a subspace L is convex.*

Theorem 3.14. *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 3.15. 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 3.16. 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 3.17 (Algebra of Convex cones). The sets $K_1 + K_2$, $\text{conv}(K_1 \cup K_2)$, $K_1 \oplus K_2$, AK , $A^{-1}K$ and λK 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 3.18. If K_1 and K_2 are convex cones containing origin, then

$$K_1 + K_2 = \text{conv}(K_1 \cup K_2).$$

Proof. By Theorem 3.12, $\text{conv}(K_1 \cup K_2)$ 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 $\text{conv}(K_1 \cup K_2)$ coincides with $K_1 + K_2$. \square

4 Convex Functions

Definition 4.1. Let $f : S \subseteq \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ 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 $\text{epi } f$. We define f to be a convex function on S if $\text{epi } f$ 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 4.2. The effective domain of a convex function f on S , which we denote by $\text{dom } f$, 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 4.3. 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 infinite values which is defined throughout the space of \mathbb{R}^n .” This leads to extended arithmetic rules involving $\pm\infty$ where the only forbidden calculations are $\infty - \infty$ and $-\infty + \infty$.

Definition 4.4. 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 4.5. 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, μ) and (y, ν) belong to $\text{epi } f$ and $0 \leq \lambda \leq 1$. By setting $\mu = f(x)$ and $\nu = f(y)$, we get the required result. \square

Corollary 4.5.1. *Let f be a function from \mathbb{R}^n to $[-\infty, +\infty]$. Then f is convex on C iff*

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

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

Corollary 4.5.2 (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 $\lambda_1 \geq 0, \dots, \lambda_m \geq 0, \lambda_1 + \cdots + \lambda_m = 1$.

Remark 4.6. The inequality in Theorem 4.5 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 4.5.1 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 4.7. *Let f be a twice continuously differentiable real-valued function on an open interval (α, β) . Then f is convex iff its second derivative f'' is nonnegative throughout (α, β) .*

Proof. (\Leftarrow) Suppose first that f'' is nonnegative on (α, β) . Then f' is non-decreasing on (α, β) . 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

$$\begin{aligned} f(z) &\leq f(x) + \lambda f'(z)(y - x), \\ f(z) &\leq f(y) - (1 - \lambda) f'(z)(y - x). \end{aligned}$$

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

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

(\Rightarrow) Suppose that f'' were not nonnegative on (α, β) . Then f'' would be negative on a certain subinterval (α', β') 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 (α, β) . \square

Example 4.8. Here are some functions on \mathbb{R} convexity is a consequence of Theorem 4.7.

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 4.9. 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 matrix $\text{Hess } f(x)$ 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 4.7, 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 $y \in C$. \square

Definition 4.10. 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 4.11. The epigraph of the indicator function is $C \times [0, \infty)$ which is a “half-cylinder with cross-section C ”.

Proposition 4.12. A set C in \mathbb{R}^n is a convex set iff δ_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 3.14 and Remark 3.16. But then $\text{epi } \delta_C$ is convex which means that δ_C is convex. \square

Definition 4.13. The support function δ_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 4.14. The gauge function γ_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 4.15. The (Euclidean) distance function d_C is defined by

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

Remark 4.16. The convexity of these functions (in Def. 4.13, 4.14 and 4.15) 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 4.17. 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 4.5.1, 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 $\text{epi } f$ 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 $\text{epi } f$. \square

Corollary 4.17.1. *Let f_i be a convex function on \mathbb{R}^n and α_i be a real number for each $i \in I$, where I is an arbitrary index set. Then*

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

is a convex set.

Proof. Like Corollary 2.4.1. \square

Definition 4.18. 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 4.19. Positive homogeneity is equivalent to the epigraph being a cone in \mathbb{R}^{n+1} .

Theorem 4.20. *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 2.18, because the subadditivity condition of f in (4) is equivalent to $\text{epi } f$ being closed under addition. \square

Corollary 4.20.1. *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 $\lambda_1 > 0, \dots, \lambda_m > 0$.

Corollary 4.20.2. *If f is a positively homogeneous proper convex function, then $f(-x) \geq -f(x)$ for every x .*

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

Theorem 4.21. *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, \dots, 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, \dots, 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

$$\begin{aligned} f(\lambda_1 b_1) + \cdots + f(\lambda_m b_m) &\geq f(x) \geq -f(-x) \geq -(f(-\lambda_1 b_1) + \cdots + f(-\lambda_m b_m)) \\ &= f(\lambda_1 b_1) + \cdots + f(\lambda_m b_m), \end{aligned}$$

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 . \square

5 Functional Operations

Theorem 5.1. *Let f be a convex function from \mathbb{R}^n to $(-\infty, +\infty]$, and let φ 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 4.5 we have

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

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

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

□

Example 5.2. Motivated from Example 4.8 we give some examples of convex function from \mathbb{R}^n to \mathbb{R} as a consequence of Theorem 5.1.

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

Remark 5.4. The properness in the hypothesis of Theorem 5.3 is to avoid $\infty - \infty$ when $f_1 + f_2$ is formed. Also note that the $\text{dom } f_1 + f_2 = \text{dom } f_1 \cap \text{dom } f_2$, where $\text{dom } f$ is the effective domain defined in Def. 4.2. Thus, $f_1 + f_2$ is improper is when the intersection of the effective domains is empty.

Corollary 5.4.1. *A linear combination $\lambda_1 f_1 + \dots + \lambda_m f_m$ of proper convex functions with nonnegative coefficients is convex.*

Theorem 5.5. *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. 4.5.1, f is convex. □

Definition 5.6. The infimal convolution of f_1, \dots, f_m (functions on \mathbb{R}^n) is the function on \mathbb{R}^n defined as

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

Remark 5.7. 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 5.8. Let f_1, \dots, f_m be proper convex functions on \mathbb{R}^n , then $f_1 \square \dots \square f_m$ is a convex function on \mathbb{R}^n .

Proof. Let $F_i = \text{epi } f_i$ and $F = F_1 + \dots + F_m$. 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 $\mu_i \geq f_i(x_i)$, $\mu = \mu_1 + \dots + \mu_m$ and $x = x_1 + \dots + x_m$. Thus f defined in the theorem is the convex function obtained from F by the construction in Theorem 5.5. \square

Remark 5.9. 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 point $a \in \mathbb{R}^n$, 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 $\text{dom } f$ and $\text{dom } g$.
- 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. 4.15.

- 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 5.10. 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 5.11. 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 5.5, we see that $f\lambda$ is convex.
- From the definition, when $\lambda > 0$ we have

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

When $\lambda = 0$, $(f0)(x) = \delta_{\{0\}}(x)$ for $f \not\equiv +\infty$, otherwise $f0 = f$, for $f \equiv +\infty$.

- The function f is positively homogeneous iff $f\lambda = f$ for each $\lambda > 0$. This is evident from (6) and Def. 4.18.

Remark 5.12. Let h be a convex function on \mathbb{R}^n and define a function f as follows:

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

Convex Cones - Fenchel Notes

Cones

Definition 5.13. 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 λ . The particular cones consisting of a non-zero vector x and all its multiples λx ($\lambda \geq 0$) is a ray, denoted by (x) .

Remark 5.14. 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 5.15. 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 5.16. Suppose for any $x, y \in S^n$, we define the map $[\cdot, \cdot] : S^n \times S^n \rightarrow \mathbb{R}$ 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 5.17. We review definitions regarding the metric topology induced by $[\cdot, \cdot]$ on S^n .

- (a) A sequence of rays (x^ν) is said to converge to a ray (x) if $[x^\nu, x] \rightarrow 0$.
- (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 ε -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 ε -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 5.18. *A set $M \subseteq S^n$ is closed in $S^n \subseteq \mathbb{R}^n$ iff M is closed in \mathbb{R}^n .*

Proof. (\Leftarrow) If M is closed in \mathbb{R}^n , then it is closed in S^n by subspace topology.

(\Rightarrow) First we show that S^n itself is closed in \mathbb{R}^n . Consider the norm $F : \mathbb{R}^n \rightarrow \mathbb{R}$ 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. \square

Proposition 5.19. *An alternate definition of open and closed cones are given below.*

(a) *A cone is closed in the sense of Def. 5.17(c) iff it is closed in \mathbb{R}^n .*

(b) *A cone is open in the sense of Def. 5.17(e) iff it is open in \mathbb{R}^n .*

Proof. (a) It follows from Lemma 5.18 and the fact that we identified the set of all rays in \mathbb{R}^n with S^n .

(b) \square

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

Definition 5.20. 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 5.21. 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 5.22. A set C of vectors is a convex cone iff it contains all vectors

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

Lemma 5.23. *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 λ and μ 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.