

# CONVEX SETS AND CONVEX FUNCTIONS

CHEE-HAN TAN

ABSTRACT. We define convex sets and convex functions, and explore the intricate relationships between these two concepts. The fundamental idea is that convex functions on  $\mathbb{R}^n$  can be identified with certain convex subsets of  $\mathbb{R}^{n+1}$ , while convex sets in  $\mathbb{R}^n$  can be identified with certain convex functions on  $\mathbb{R}^n$ . This provides a bridge between a geometric approach and an analytical approach in dealing with convex functions. In particular, one should be acquainted with the geometric connection between convex functions and epigraphs.

## PREFACE

The structure of these notes follows closely Chapter 1 of the book “Convex Analysis” by R. T. Rockafellar [Roc70]. All the theorems and corollaries are numbered independently of [Roc70], and additional examples or remarks or results from other sources are added as I see fit, mainly to facilitate my understanding. Any mistakes here are of course my own. Please report any typographical errors or mathematical fallacy to me by email.

## 1. AFFINE SETS

A problem of practical interest is studying geometrical properties of sets of points that are invariant under certain transformations such as translations and rotations. One approach would be to model the space of points as a vector space, but this is not very satisfactory. One reason is that the point corresponds to the zero vector  $\mathbf{0}$ , called the origin, plays a special role, when there is really no reason to have a privileged origin [Gal11].

For two distinct points  $x, y \in \mathbb{R}^n$ , 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  $M \subseteq \mathbb{R}^n$  is called an *affine set* if

$$(1 - \lambda)x + \lambda y \in M \quad \text{for every } x \in M, y \in M, \text{ and } \lambda \in \mathbb{R}.$$

For the remaining section, we exploit the fact that almost every affine concept is the analogue of certain concepts in linear algebra.



Roughly speaking, affine sets are vector spaces whose origin we try to forget about. To distinguish the underlying structure, given any vector space  $V$ , elements of  $V$  as a vector space are called *vectors*, and elements of  $V$  as an affine set are called *points*.

---

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF UTAH, SALT LAKE CITY, UT 84112

*E-mail address:* [tan@math.utah.edu](mailto:tan@math.utah.edu).

*Date:* February 5, 2019.

Take any two points  $x, y$  of an affine set  $M$  and  $\lambda \in \mathbb{R}$ . The combination  $(1 - \lambda)x + \lambda y = x + \lambda(y - x) =: z$  yields another point in  $M$ . There are two crucial interpretations of  $z$  here.

- (1) Writing  $z = x + \lambda(y - x)$ , we see that  $z$  is obtained from translating the point  $x$  by the vector  $\lambda(y - x)$ . With  $\lambda = 1$ , this says that it makes sense to *subtract* two points of the affine set, in the sense that for every  $x, y \in M$ , there exists a unique (translation) vector  $v = y - x$  such that  $y = x + v$ .
- (2) Writing  $z = (1 - \lambda)x + \lambda y$ , we say that  $z$  is obtained by taking linear combination of  $x$  and  $y$  where the coefficients  $1 - \lambda$  and  $\lambda$  sum to 1. This suggests that linear combination of points of  $M$  is well-defined only if the coefficients sum to 1.

We postpone the explanation of “well-defined” in the last sentence to Section 1.2, where we justify the necessity of defining a suitable notion of linear combination of points of affine sets, called *affine combination*.

**1.1. Parallelism and characterisation of affine sets in terms of hyperplanes.** Almost every concept in affine sets is the counterpart of certain concept in linear algebra. We begin with the exact correspondence between affine sets and subspaces of  $\mathbb{R}^n$ .

**Theorem 1.1.** *The subspaces of  $\mathbb{R}^n$  are the affine sets which contain the origin.*

*Proof.* The forward direction is clear, since every subspace of  $\mathbb{R}^n$  contains the origin  $\mathbf{0}$  and is closed under addition and scalar multiplication. Conversely, suppose  $M$  is an affine set containing  $\mathbf{0}$ . Choose any  $x, y \in M$  and  $\lambda \in \mathbb{R}$ . Then  $M$  is closed under scalar multiplication since

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

and is closed under addition since

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

Hence  $M$  is a subspace of  $\mathbb{R}^n$ . ♥

Given any set  $M \subseteq \mathbb{R}^n$  and vector  $a \in \mathbb{R}^n$ , the *translate* of  $M$  by  $a$  is defined to be the set

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

It is clear that a translate of an affine set is another affine set. An affine set  $M$  is said to be *parallel* to an affine set  $L$  if  $M = L + a$  for some  $a \in \mathbb{R}^n$ . One can readily check that “ $M$  is parallel to  $L$ ” is an equivalence relation on the family of affine subsets of  $\mathbb{R}^n$ , *i.e.*,

- (1)  $M$  is parallel to  $M$ ;
- (2) if  $M$  is parallel to  $L$ , then  $L$  is parallel to  $M$ ; and
- (3) if  $M$  is parallel to  $L$  and  $L$  is parallel to  $N$ , then  $M$  is parallel to  $N$ .



This definition of parallelism is restrictive in the sense that it does not include the idea of a line being parallel to a plane.

**Theorem 1.2.** *Each nonempty affine set  $M \subseteq \mathbb{R}^n$  is a parallel to a unique subspace  $L \subseteq \mathbb{R}^n$ . This  $L$  is given by*

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

*Proof.* We first show that  $M$  cannot be parallel to two different subspaces. Suppose there are two subspaces  $L_1, L_2$  parallel to  $M$ . Then  $L_2 = L_1 + a$  for some vector  $a \in \mathbb{R}^n$  from the equivalence relation of parallelism. Since  $L_2$  is a subspace of  $\mathbb{R}^n$ , we have  $\mathbf{0} \in L_2$  and so  $-a \in L_1$  and  $a = -(-a) \in L_1$  since  $L_1$  is also a subspace of  $\mathbb{R}^n$ . In particular, we have that  $L_2 = L_1 + a \subseteq L_1$ . A similar argument shows that  $L_1 \subseteq L_2$  and so  $L_1 = L_2$ , establishing the uniqueness. Now, observe that for any  $x \in M$ , the affine set  $M - x$  is a translate of  $M$  containing the origin  $\mathbf{0}$ . It follows from Theorem 1.1 and the uniqueness proof above that this set must be the unique subspace  $L$  parallel to  $M$ . Since  $L = M - x$  no matter which  $x \in M$  is chosen, we actually have  $L = M - M$ . ♥

Theorem 1.2 simply says that an affine set  $M \subseteq \mathbb{R}^n$  is a translation of some subspace  $L \subseteq \mathbb{R}^n$ . Moreover,  $L$  is uniquely determined by  $M$  and independent of the choice of  $x \in M$ , so  $M = L + x$  for any  $x \in M$ . We can now define the *dimension* of a non-empty affine set as the dimension of the subspace parallel to it, which is well-defined from Theorem 1.2. Affine sets of dimension 0, 1, and 2 are called *points*, *lines*, and *planes*, respectively. An  $(n - 1)$ -dimensional (or 1-codimensional) affine set in  $\mathbb{R}^n$  is called a *hyperplane*.

**Theorem 1.3.** *Given  $\beta \in \mathbb{R}$  and a nonzero  $b \in \mathbb{R}^n$ , the set*

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

*is a hyperplane in  $\mathbb{R}^n$ . Moreover, every hyperplane in  $\mathbb{R}^n$  may be represented in this way, with  $b$  and  $\beta$  unique up to a common nonzero scalar multiple.*

*Proof.* For the forward direction, observe that  $H$  is an  $(n - 1)$ -dimensional subset of  $\mathbb{R}^n$  since it is the solution set of a one-dimensional linear system in  $n$  variables. To see that  $H$  is affine, take any  $x, y \in H$  and  $\lambda \in \mathbb{R}$ . Then

$$\langle b, (1 - \lambda)x + \lambda y \rangle = (1 - \lambda)\langle b, x \rangle + \lambda\langle b, y \rangle = (1 - \lambda)\beta + \lambda\beta = \beta,$$

which gives  $(1 - \lambda)x + \lambda y \in H$ . Conversely, let  $H \subset \mathbb{R}^n$  be a hyperplane and  $L \subset \mathbb{R}^n$  the subspace parallel to  $H$ , *i.e.*,  $H = L + a$  for some  $a \in H$ . Since  $\dim(L) = n - 1$ , we can write  $L$  as  $\text{span}(b)^\perp \subset \mathbb{R}^n$  for some nonzero vector  $b \in \mathbb{R}^n$  (unique up to a nonzero scalar multiple). Consequently,

$$\begin{aligned} H &= \text{span}(b)^\perp + a = \{x + a \in \mathbb{R}^n : \langle b, x \rangle = 0\} \\ &= \{y \in \mathbb{R}^n : \langle b, y - a \rangle = 0\} \\ &= \{y \in \mathbb{R}^n : \langle b, y \rangle = \langle b, a \rangle =: \beta\}. \end{aligned}$$

♥

The vector  $b \in \mathbb{R}^n \setminus \{\mathbf{0}\}$  in Theorem 1.3 is called a *normal* to the corresponding hyperplane  $H$ . Any hyperplane of  $\mathbb{R}^n$  (or more generally a Euclidean space) has exactly two unit normal vectors, and it separates  $\mathbb{R}^n$  into two half-spaces; see Figure 1. We generalise Theorem 1.3 to any affine subset of  $\mathbb{R}^n$ , characterising it as the solution set of an inhomogeneous linear system.

**Theorem 1.4.** *Given  $b \in \mathbb{R}^m$  and  $B \in \mathbb{R}^{m \times n}$ , the set*

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

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

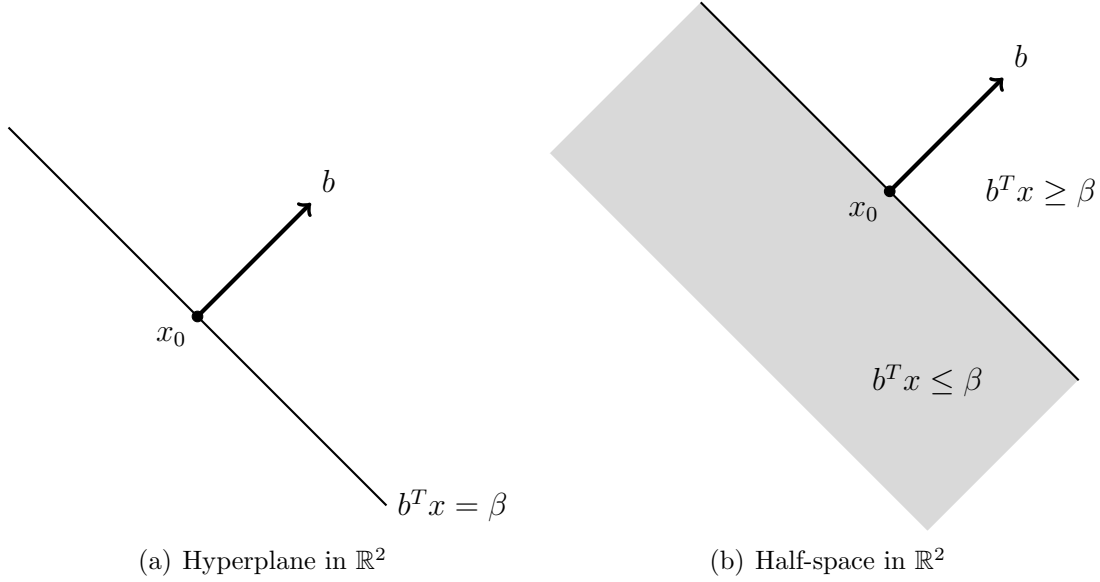


FIGURE 1. Hyperplane and half-space in  $\mathbb{R}^2$ , with a normal vector  $b \in \mathbb{R}^2$  and a point  $x_0$  in the hyperplane.

*Proof.* The forward direction is clear. Conversely, let  $M \subseteq \mathbb{R}^n$  be any affine set. If  $M = \mathbb{R}^n$  or  $M = \emptyset$ , then we can choose  $B = \mathbf{0}_{m \times n} \in \mathbb{R}^{m \times n}$  with  $b = \mathbf{0} \in \mathbb{R}^n$  or  $b \neq \mathbf{0}$ , respectively. Otherwise, let  $L \subset \mathbb{R}^n$  be the subspace parallel to  $M$  and  $b_1, \dots, b_m$  a basis for  $L^\perp$ . For some  $a \in M$  we have

$$\begin{aligned} M = L + a &= L^{\perp\perp} + a = \{x + a \in \mathbb{R}^n : \langle b_j, x \rangle = 0, j = 1, \dots, m\} \\ &= \{x + a \in \mathbb{R}^n : Bx = \mathbf{0}\} \\ &= \{y \in \mathbb{R}^n : By = Ba =: b\}, \end{aligned}$$

where  $B \in \mathbb{R}^{m \times n}$  is the matrix with rows  $b_1, \dots, b_m$ . ♥

**Corollary 1.5.** *Every affine subset of  $\mathbb{R}^n$  is an intersection of a finite collection of hyperplanes.*

*Proof.* Let  $M$  be any affine subset of  $\mathbb{R}^n$ . Theorem 1.4 asserts that

$$M = \{x \in \mathbb{R}^n : \langle b_j, x \rangle = \beta_j, j = 1, \dots, m\} = \bigcap_{j=1}^m H_j,$$

where  $b_j \in \mathbb{R}^n$  and  $\beta_j \in \mathbb{R}$  are the  $j$ th row of  $B \in \mathbb{R}^{m \times n}$  and the  $j$ th component of  $b \in \mathbb{R}^n$ , respectively, and

$$H_j = \{x \in \mathbb{R}^n : \langle b_j, x \rangle = \beta_j\}, \quad j = 1, \dots, m.$$

Each  $H_j$  is a hyperplane ( $b_j \neq \mathbf{0}$ ), or the empty set ( $b_j = \mathbf{0}, \beta_j \neq 0$ ) which can be regarded as the intersection of two different parallel hyperplanes, or  $\mathbb{R}^n$  ( $b_j = \mathbf{0}, \beta_j = 0$ ) which can be regarded as the intersection of the empty collection of hyperplanes. ♥

**1.2. Affine combinations and affine hulls.** Attempting to extend the notion of linear combination of vectors in vector spaces to that of points in affine sets in  $\mathbb{R}^n$  is a nontrivial task. The naive approach, where we define linear combination of points by associating points with vectors is problematic, because vector addition depends crucially on the choice of coordinate systems. It turns out that if we impose the additional constraint that the coefficients sum to 1, then the above definition is intrinsic, in the sense that it is independent from the choice of the origin  $\mathbf{o}$  in the affine set.

**Lemma 1.6.** *Let  $M \subseteq \mathbb{R}^n$  be any affine set, and  $x_1, \dots, x_m$  be points in  $M$ . Let  $\lambda_1, \dots, \lambda_m$  be a sequence of scalars in  $\mathbb{R}$ . For any two points  $\mathbf{o}, \mathbf{o}' \in M$ ,*

(a) *if  $\lambda_1 + \dots + \lambda_m = 1$ , then*

$$\mathbf{o} + \sum_{j=1}^m \lambda_j(x_j - \mathbf{o}) = \mathbf{o}' + \sum_{j=1}^m \lambda_j(x_j - \mathbf{o}')$$

*and this is written as  $\lambda_1 x_1 + \dots + \lambda_m x_m$ ;*

(b) *if  $\lambda_1 + \dots + \lambda_m = 0$ , then*

$$\sum_{j=1}^m \lambda_j(x_j - \mathbf{o}) = \sum_{j=1}^m \lambda_j(x_j - \mathbf{o}').$$

*Proof.* For part (a),

$$\begin{aligned} \mathbf{o} + \sum_{j=1}^m \lambda_j(x_j - \mathbf{o}) &= \mathbf{o} + \sum_{j=1}^m \lambda_j(x_j - \mathbf{o}' + \mathbf{o}' - \mathbf{o}) \\ &= \mathbf{o} + \sum_{j=1}^m \lambda_j(\mathbf{o}' - \mathbf{o}) + \sum_{j=1}^m \lambda_j(x_j - \mathbf{o}') \\ &= \mathbf{o} + (\mathbf{o}' - \mathbf{o}) + \sum_{j=1}^m \lambda_j(x_j - \mathbf{o}') && \left[ \text{Since } \sum_{j=1}^m \lambda_j = 1. \right] \\ &= \mathbf{o}' + \sum_{j=1}^m \lambda_j(x_j - \mathbf{o}') \end{aligned}$$

Part (b) follows from a similar computation just given. ♥

Lemma 1.6 essentially says that the combination of points  $\lambda_1 x_1 + \dots + \lambda_m x_m$  is well-defined if and only if either the coefficients sum to 1 which results in a point or the coefficients sum to 0 which results in a vector. An *affine combination or barycenter* of  $m$  points  $x_1, \dots, x_m \in \mathbb{R}^n$  is a sum  $\lambda_1 x_1 + \dots + \lambda_m x_m$ , where  $\lambda_1, \dots, \lambda_m \in \mathbb{R}$  and  $\lambda_1 + \dots + \lambda_m = 1$ .

**Theorem 1.7.** *A nonempty subset of  $\mathbb{R}^n$  is affine if and only if it contains all the affine combinations of its elements.*

*Proof.* By definition, a set  $M \subseteq \mathbb{R}^n$  is affine if and only if  $M$  is closed under taking affine combinations with  $m = 2$ . We must show that affineness of  $M$  also implies that  $M$  is closed under taking affine combinations with  $m > 2$ . Let us proceed by induction on  $m$ . The base case  $m = 2$  follows from affineness of  $M$ . Suppose the statement holds for some  $m > 2$ . Let  $x_1, \dots, x_{m+1} \in M$  and  $\lambda_1, \dots, \lambda_{m+1}$  be scalars in  $\mathbb{R}$  such that  $\lambda_1 + \dots + \lambda_{m+1} = 1$ . The

statement coincides with the induction hypothesis if one of the scalars is zero, so suppose  $\lambda_1, \dots, \lambda_{m+1}$  are all nonzero and assume WLOG that  $\lambda_{m+1} \neq 1$ . Then

$$\begin{aligned} z &:= \sum_{j=1}^{m+1} \lambda_j x_j = \sum_{j=1}^m \lambda_j x_j + \lambda_{m+1} x_{m+1} \\ &= (1 - \lambda_{m+1}) \sum_{j=1}^m \left( \frac{\lambda_j}{1 - \lambda_{m+1}} \right) x_j + \lambda_{m+1} x_{m+1} \\ &= (1 - \lambda_{m+1}) y + \lambda_{m+1} x_{m+1}. \end{aligned}$$

Since

$$\sum_{j=1}^m \frac{\lambda_j}{1 - \lambda_{m+1}} = \frac{\lambda_1 + \dots + \lambda_m}{1 - \lambda_{m+1}} = \frac{1 - \lambda_{m+1}}{1 - \lambda_{m+1}} = 1,$$

it follows that  $y$  is an affine combination of  $m$  points in  $M$  and we have  $y \in M$  by induction hypothesis. Since  $M$  is affine and  $z$  is an affine combination of  $y$  and  $x_{m+1}$ , we conclude that  $z \in M$ , as desired.  $\heartsuit$

**Theorem 1.8.** *The intersection of an arbitrary collection of affine sets is an affine set.*

*Proof.* Let  $M_i \subseteq \mathbb{R}^n$  be an affine sets for all  $i \in I$ , where  $I$  is an arbitrary index set. Consider  $M = \bigcap_{i \in I} M_i$  and let  $x, y \in M$  and  $\lambda \in \mathbb{R}$ . Then  $x, y \in M_i$  for any  $i \in I$  and since  $M_i$  is affine, it follows that  $z := (1 - \lambda)x + \lambda y \in M_i$  for any  $i \in I$ , i.e.,  $z \in M$ .  $\heartsuit$

Theorem 1.8 motivates the following definition: Given any  $S \subseteq \mathbb{R}^n$ , the *affine hull* of  $S$ , denoted by  $\text{aff}(S)$ , is the unique smallest affine set containing  $S$ ; namely, the intersection of all affine sets containing  $S$ . In other words, if  $M \subseteq \mathbb{R}^n$  is any affine set containing  $S$ , then  $\text{aff}(S) \subseteq M$ . Below are some examples:

- ♣ For two distinct points  $x, y \in \mathbb{R}^n$ ,  $\text{aff}\{x, y\}$  is the line through  $x$  and  $y$ .
- ♣ For three distinct points  $x, y, z \in \mathbb{R}^n$  that are not collinear,  $\text{aff}\{x, y, z\}$  is the plane passing through them.
- ♣ The affine hull of a set of  $(n + 1)$  points not in a hyperplane in  $\mathbb{R}^n$  is  $\mathbb{R}^n$ .
- ♣ The affine hull of the set  $\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1, z = 1\}$  is the plane  $\{z = 1\}$ .

The definition above can be viewed as a characterisation of  $\text{aff}(S)$  from the *outside*, but it is not immediately clear how an element in  $\text{aff}(S)$  relates to elements of  $S$ . This motivates the next theorem and in some sense it provides a characterisation of  $\text{aff}(S)$  from the *inside*.

**Theorem 1.9.** *For any  $S \subseteq \mathbb{R}^n$ ,  $\text{aff}(S)$  is the set of all affine combinations of elements of  $S$ . That is,*

$$\text{aff}(S) = \left\{ \sum_{j=1}^m \lambda_j x_j : x_1, \dots, x_m \in S, \sum_{j=1}^m \lambda_j = 1, m \in \mathbb{N} \right\}.$$

*Proof.* Let  $C$  denote the set of all affine combinations of elements of  $S$ . Since elements of  $S$  belong to  $\text{aff}(S)$ , we have  $C \subseteq \text{aff}(S)$  by Theorem 1.8. For the reverse inclusion, it suffices to show that  $C$  is an affine set since  $C$  contains  $S$  and  $\text{aff}(S) \subseteq M$  for any affine set  $M$

containing  $S$  from the definition of affine hull. To this end, take any  $y, z \in C$ . There exists  $x_1, \dots, x_k, x_{k+1}, \dots, x_m \in S$  and scalars  $\lambda_1, \dots, \lambda_k, \lambda_{k+1}, \dots, \lambda_m$  such that

$$y = \sum_{j=1}^k \lambda_j x_j, \quad z = \sum_{j=k+1}^m \lambda_j x_j, \quad \text{and} \quad \sum_{j=1}^k \lambda_j = \sum_{j=k+1}^m \lambda_j = 1.$$

For any  $\mu \in \mathbb{R}$  we have

$$(1 - \mu)y + \mu z = \sum_{j=1}^k (1 - \mu)\lambda_j x_j + \sum_{j=k+1}^m \mu\lambda_j x_j = \sum_{j=1}^m \Lambda_j x_j,$$

with

$$\sum_{j=1}^m \Lambda_j = \sum_{j=1}^k (1 - \mu)\lambda_j + \sum_{j=k+1}^m \mu\lambda_j = (1 - \mu) + \mu = 1.$$

This means that  $(1 - \mu)y + \mu z$  is an affine combination of elements of  $C$  and since  $y, z \in C$  were arbitrary, this shows that  $C$  is affine, as desired. ♥

**1.3. Affine independence and affine basis.** The analogue of the concept of linear independence in vector spaces is affine independence in affine sets of  $\mathbb{R}^n$ . A set of  $(m + 1)$  points  $b_0, b_1, \dots, b_m \in \mathbb{R}^n$  is said to be *affinely independent* if  $\text{aff}\{b_0, b_1, \dots, b_m\}$  is  $m$ -dimensional. The next theorem shows that this definition is equivalent to the condition of linear independence of certain set of  $m$  vectors.

**Theorem 1.10.** *Let  $I$  be the index set  $\{0, 1, \dots, m\}$ . A set of  $(m + 1)$  points  $b_0, b_1, \dots, b_m \in \mathbb{R}^n$  is affinely independent if and only if the set of  $m$  vectors  $\{b_j - b_k\}_{j \in I \setminus \{k\}}$  is linearly independent for some  $k \in I$ .*

*Proof.* Let  $M = \text{aff}\{b_0, b_1, \dots, b_m\}$ . By Theorem 1.1, we can write  $M = L + b_k$  for some  $k \in I$ , where  $L$  is the subspace given by

$$L = \text{aff}\{b_0 - b_k, b_1 - b_k, \dots, b_m - b_k\}.$$

Since  $\dim(M) = \dim(L)$ , it follows (upon excluding the zero vector  $b_k - b_k$ ) that the set of points  $b_0, b_1, \dots, b_m \in \mathbb{R}^n$  is affinely independent if and only if the set of vectors  $\{b_j - b_k\}_{j \in I \setminus \{k\}}$  is linearly independent. ♥

**Theorem 1.11.** *An  $m$ -dimensional affine set  $M$  can be represented as the affine hull of  $(m + 1)$  affinely independent points of  $M$ .*

*Proof.* Let  $M$  be any  $m$ -dimensional affine set and  $L$  be the  $m$ -dimensional subspace parallel to  $M$ . Choose a basis  $\{c_1, \dots, c_m\}$  of  $L$ . Since  $M = L + b_0$  for some  $b_0 \in M$ , the chosen basis of  $L$  can be written as  $b_1 - b_0, \dots, b_m - b_0$  for some  $b_1, \dots, b_m \in M$ . For any  $x \in M$ ,  $x - b_0$  is a vector of  $L$  and it can be written (uniquely) as

$$x - b_0 = \lambda_1(b_1 - b_0) + \dots + \lambda_m(b_m - b_0),$$

for some sequence of scalars  $\lambda_1, \dots, \lambda_m$ . Thus

$$\begin{aligned} x &= b_0 + (x - b_0) = (1 - (\lambda_1 + \dots + \lambda_m))b_0 + \lambda_1 b_1 + \dots + \lambda_m b_m \\ &= \lambda_0 b_0 + \lambda_1 b_1 + \dots + \lambda_m b_m. \end{aligned}$$

Since  $\lambda_0 + \lambda_1 + \dots + \lambda_m = 1$ , this shows that any  $x \in M$  can be written as an affine combination of  $(m + 1)$  points  $b_0, b_1, \dots, b_m$  of  $M$ . The desired statement follows from Theorem 1.9. ♥

We provide another characterisation for the set of affinely independent points, which is a key ingredient in several beautiful and deep theorems about convex sets, such as Carathéodory's theorem, Radon's theorem, and Helly's theorem.

**Theorem 1.12.** *A set of  $(m + 1)$  points  $b_0, b_1, \dots, b_m \in \mathbb{R}^n$  is affinely independent if and only if the equations*

$$(1.1) \quad \begin{cases} \lambda_0 b_0 + \lambda_1 b_1 + \dots + \lambda_m b_m = 0 \\ \lambda_0 + \lambda_1 + \dots + \lambda_m = 0 \end{cases}$$

*can only be satisfied by the scalars  $\lambda_0 = \lambda_1 = \dots = \lambda_m = 0$ .*

*Proof.* By Theorem 1.10, it suffices to show that the set of  $m$  vectors  $S = \{b_1 - b_0, \dots, b_m - b_0\}$  is linearly independent if and only if the equations (1.1) can only be satisfied by the scalars  $\lambda_0 = \lambda_1 = \dots = \lambda_m = 0$ . To prove the forward direction, suppose (1.1) holds. Then

$$\begin{aligned} 0 &= \lambda_0 b_0 + \lambda_1 b_1 + \dots + \lambda_m b_m \\ &= -(\lambda_1 + \dots + \lambda_m) b_0 + \lambda_1 b_1 + \dots + \lambda_m b_m \\ &= \lambda_1 (b_1 - b_0) + \lambda_2 (b_2 - b_0) + \dots + \lambda_m (b_m - b_0). \end{aligned}$$

Since  $S$  is linearly independent, this shows that  $\lambda_1 = \dots = \lambda_m = 0$  and  $\lambda_0 = -(\lambda_1 + \dots + \lambda_m) = 0$ . The reverse direction is straightforward using a similar argument to the one just given. ♥

An important property of the set of linearly independent vectors  $S = \{v_1, \dots, v_m\}$  is that any  $x \in \text{span}(S)$  can be written uniquely as

$$x = \lambda_1 v_1 + \dots + \lambda_m v_m$$

for some scalars  $\lambda_1, \dots, \lambda_m$ . A similar result holds for affinely independent points.

**Theorem 1.13.** *Given any  $(m + 1)$  points  $b_0, b_1, \dots, b_m \in \mathbb{R}^n$ , let  $M = \text{aff}\{b_0, b_1, \dots, b_m\}$ . Any  $x \in M$  can be written uniquely as the affine combination of  $b_0, b_1, \dots, b_m$  for some sequence of scalars in  $\mathbb{R}$  if and only if the set of points  $b_0, b_1, \dots, b_m$  is affinely independent.*

*Proof.* By Theorem 1.9, any  $x \in M = \text{aff}\{b_0, b_1, \dots, b_m\}$  can be written as

$$x = \lambda_0 b_0 + \lambda_1 b_1 + \dots + \lambda_m b_m, \quad \text{with } \lambda_0 + \lambda_1 + \dots + \lambda_m = 1.$$

Define the index set  $I = \{0, 1, \dots, m\}$ . We must prove that the sequence of scalars  $\{\lambda_j\}_{j \in I}$  in such an expression of  $x \in M$  is unique if and only if the set of points  $\{b_j\}_{j \in I}$  is affinely independent. Suppose the sequence of scalars is unique, and that

$$\mu_1 (b_1 - b_0) + \dots + \mu_m (b_m - b_0) = \mathbf{0}$$

for some scalars  $\mu_1, \dots, \mu_m \in \mathbb{R}$ . Since

$$\mu_1 (b_1 - b_0) + \dots + \mu_m (b_m - b_0) = \mathbf{0} = 0(b_1 - b_0) + \dots + 0(b_m - b_0),$$

it follows from the uniqueness assumption that  $\mu_1 = \dots = \mu_m = 0$ . Thus the set of vectors  $b_1 - b_0, \dots, b_m - b_0$  is linearly independent, and the affine independence of  $\{b_j\}_{j \in I}$  follows from Theorem 1.10. Conversely, suppose the set of points  $\{b_j\}_{j \in I}$  is affinely independent, and the sequence of scalars  $\{\lambda_j\}_{j \in I}$  is not unique, *i.e.*, there exists another sequence of scalars  $\{\mu_j\}_{j \in I}$  such that

$$x = \mu_0 b_0 + \mu_1 b_1 + \dots + \mu_m b_m, \quad \text{with } \mu_0 + \mu_1 + \dots + \mu_m = 1.$$



Then we have -

$$0 = x - x = (\lambda_0 - \mu_0)b_0 + (\lambda_1 - \mu_1)b_1 + \cdots + (\lambda_m - \mu_m)b_m,$$

with

$$\sum_{j=0}^m \lambda_j - \mu_j = \sum_{j=0}^m \lambda_j - \sum_{j=0}^m \mu_j = 1 - 1 = 0.$$

By Theorem 1.12,  $\lambda_j - \mu_j = 0$  for every  $j \in I$ , i.e., the sequence of scalars  $\{\lambda_j\}_{j \in I}$  is unique. ♥

Given an affine set  $M \subseteq \mathbb{R}^n$ , the set of  $(m+1)$  points  $\{b_0, b_1, \dots, b_m\}$  in  $M$  is an *affine basis* of  $M$  if any  $x \in M$  can be written uniquely as

$$x = \lambda_0 b_0 + \lambda_1 b_1 + \cdots + \lambda_m b_m,$$

for some sequence of scalars  $\lambda_0, \lambda_1, \dots, \lambda_m$  that sums up to 1. The scalars  $\{\lambda_0, \lambda_1, \dots, \lambda_m\}$  are called the *barycentric coordinates* of  $x$  over  $\{b_0, b_1, \dots, b_m\}$ . The origin of the term barycentric coordinates stems from the following physical interpretation: If the  $b_j$ 's are viewed as bodies having weights  $\lambda_j$ 's, then the point  $x$  is the *barycenter* of the  $b_j$ 's where the weights have been normalised so that  $\lambda_0 + \lambda_1 + \cdots + \lambda_m = 1$ .

**1.4. Affine transformations.** Corresponding to linear transformation is the concept of *affine transformation*, i.e., any single-valued mapping  $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$  satisfying

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

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

**Theorem 1.14.** *An affine transformation  $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$  preserves affine combinations of points in  $\mathbb{R}^n$ . That is, for any  $x_1, \dots, x_k \in \mathbb{R}^n$  and  $\lambda_1, \dots, \lambda_k \in \mathbb{R}$  satisfying  $\lambda_1 + \cdots + \lambda_k = 1$ , we have*

$$T\left(\sum_{j=1}^k \lambda_j x_j\right) = \sum_{j=1}^k \lambda_j T x_j.$$

*Proof.* We proceed by induction on  $k$ . The base case  $k = 1$  is trivial, and the case  $k = 2$  follows from the definition of affine transformation. Suppose the statement holds for some  $k > 2$ . Let  $x_1, \dots, x_{k+1} \in \mathbb{R}^n$  and  $\lambda_1, \dots, \lambda_{k+1}$  be scalars in  $\mathbb{R}$  such that  $\lambda_1 + \cdots + \lambda_{k+1} = 1$ . The statement coincides with the induction hypothesis if one of the scalars is zero, so suppose  $\lambda_1, \dots, \lambda_{k+1}$  are all nonzero and assume WLOG that  $\lambda_{k+1} \neq 1$ . Then

$$\begin{aligned} T\left(\sum_{j=1}^{k+1} \lambda_j x_j\right) &= T\left(\sum_{j=1}^k \lambda_j x_j + \lambda_{k+1} x_{k+1}\right) \\ &= T\left((1-\lambda_{k+1}) \sum_{j=1}^k \left(\frac{\lambda_j}{1-\lambda_{k+1}}\right) x_j + \lambda_{k+1} x_{k+1}\right) \\ &= (1-\lambda_{k+1})T\left(\sum_{j=1}^k \left(\frac{\lambda_j}{1-\lambda_{k+1}}\right) x_j\right) + \lambda_{k+1} T x_{k+1} \\ &= (1-\lambda_{k+1})T y + \lambda_{k+1} T x_{k+1}. \end{aligned}$$

Since

$$\sum_{j=1}^k \frac{\lambda_j}{1 - \lambda_{k+1}} = \frac{\lambda_1 + \cdots + \lambda_k}{1 - \lambda_{k+1}} = \frac{1 - \lambda_{k+1}}{1 - \lambda_{k+1}} = 1,$$

it follows that  $y$  is an affine combination of  $k$  points in  $\mathbb{R}^n$  and using the induction hypothesis, we obtain

$$\begin{aligned} T \left( \sum_{j=1}^{k+1} \lambda_j x_j \right) &= (1 - \lambda_{k+1})Ty + \lambda_{k+1}Tx_{k+1} \\ &= (1 - \lambda_{k+1}) \sum_{j=1}^k \left( \frac{\lambda_j}{1 - \lambda_{k+1}} \right) Tx_j + \lambda_{k+1}Tx_{k+1} \\ &= \sum_{j=1}^k \lambda_j Tx_j + \lambda_{k+1}Tx_{k+1} = \sum_{j=1}^{k+1} \lambda_j Tx_j. \end{aligned}$$

♥

**Theorem 1.15.** *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 from  $\mathbb{R}^n$  to  $\mathbb{R}^m$  and  $a \in \mathbb{R}^m$ .*

*Proof.* Suppose  $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$  is affine. Let  $a = T\mathbf{0}$  and  $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$  be the mapping given by  $Ax = Tx - a$ . Then  $A\mathbf{0} = \mathbf{0}$  by construction. We claim that  $A$  is a linear transformation. Indeed, for any  $x, y \in \mathbb{R}^n$  and  $\alpha \in \mathbb{R}$  we have

$$\begin{aligned} A(\alpha x) &= T(\alpha x) - a = T(\alpha x + (1 - \alpha)\mathbf{0}) - a \\ &= \alpha Tx + (1 - \alpha)T\mathbf{0} - a \\ &= \alpha Tx - \alpha a \\ &= \alpha(Tx - a) = \alpha Ax \\ A(x + y) &= A \left( 2 \left( \frac{1}{2}(x + y) \right) \right) = 2 \left( T \left( \frac{1}{2}(x + y) \right) - a \right) \\ &= 2 \left( \frac{1}{2}Tx + \frac{1}{2}Ty - a \right) \\ &= Tx + Ty - 2a \\ &= (Tx - a) + (Ty - a) = Ax + Ay. \end{aligned}$$

Conversely, suppose  $Tx = Ax + a$  where  $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$  is any linear transformation and  $a \in \mathbb{R}^m$ . Choose any  $x, y \in \mathbb{R}^n$  and  $\lambda \in \mathbb{R}$ . Then

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

Hence  $T$  is an affine transformation from  $\mathbb{R}^n$  to  $\mathbb{R}^m$ .

♥

**Theorem 1.16.** *Let  $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$  be an affine transformation.*

- (a) The inverse of  $T$ , if it exists, is affine.  
(b) For any affine set  $M \subseteq \mathbb{R}^n$ , the image of  $M$  under  $T$ ,

$$T(M) = \{Tx \in \mathbb{R}^m : x \in M\},$$

is an affine set in  $\mathbb{R}^m$ .

- (c)  $T$  preserve affine hulls, i.e.,  $T(\text{aff}(S)) = \text{aff}(TS)$  for any  $S \subseteq \mathbb{R}^n$ , where  $TS$  is the image of  $S$  under  $T$ .

*Proof.* For part (a), we assume that  $T^{-1}: \mathbb{R}^m \rightarrow \mathbb{R}^n$  exists and show that  $T^{-1}$  is affine. Consider  $y_1, y_2 \in \mathbb{R}^m$ . There exists corresponding points  $x_1, x_2 \in \mathbb{R}^n$  such that  $Tx_1 = y_1$  and  $Tx_2 = y_2$ , or equivalently,  $T^{-1}y_1 = x_1$  and  $T^{-1}y_2 = x_2$ . Since  $T$  is affine, for any  $\lambda \in \mathbb{R}$  we have

$$(1 - \lambda)y_1 + \lambda y_2 = (1 - \lambda)Tx_1 + \lambda Tx_2 = T[(1 - \lambda)x_1 + \lambda x_2]$$

which implies that

$$T^{-1}[(1 - \lambda)y_1 + \lambda y_2] = (1 - \lambda)x_1 + \lambda x_2 = (1 - \lambda)T^{-1}y_1 + \lambda T^{-1}y_2,$$

establishing the affinity of  $T^{-1}$ . For part (b), let  $y_1, y_2 \in T(M)$  and  $\lambda \in \mathbb{R}$ . There exists corresponding points  $x_1, x_2 \in M$  such that  $Tx_1 = y_1$  and  $Tx_2 = y_2$ . We will show that the point  $z = (1 - \lambda)y_1 + \lambda y_2 \in T(M)$ . Indeed,

$$z = (1 - \lambda)Tx_1 + \lambda Tx_2 = T[(1 - \lambda)x_1 + \lambda x_2] \in T(M),$$

since  $M$  is affine and  $x_1, x_2 \in M$ . For part (c), first take any  $x \in \text{aff}(S)$ . From Theorem 1.9, there exists elements  $x_1, \dots, x_k \in S$  and scalars  $\lambda_1, \dots, \lambda_k \in \mathbb{R}$  such that

$$x = \sum_{j=1}^k \lambda_j x_j \quad \text{and} \quad \sum_{j=1}^k \lambda_j = 1.$$

By Theorem 1.14, we have that

$$Tx = T\left(\sum_{j=1}^k \lambda_j x_j\right) = \sum_{j=1}^k \lambda_j Tx_j \in \text{aff}(TS),$$

since  $Tx_j \in TS$  for all  $j = 1, \dots, k$ . This shows that  $T(\text{aff}(S)) \subseteq \text{aff}(TS)$ . For the reverse inclusion, it suffices to show that  $T(\text{aff}(S))$  is an affine set since  $T(\text{aff}(S))$  contains  $TS$  and  $\text{aff}(TS) \subseteq M$  for any affine set  $M \subseteq \mathbb{R}^m$  containing  $TS$  from the definition of affine hull. The affinity of  $T(\text{aff}(S))$  follows from part (b). ♥

**Theorem 1.17.** 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  $T$  of  $\mathbb{R}^n$  onto itself, such that  $Tb_i = b'_i$  for  $i = 0, 1, \dots, m$ . If  $m = n$ , then such  $T$  is unique.

*Proof.* Enlarging the given affinely independent sets if necessary, we can reduce the problem to the case where  $m = n$ . By Theorem 1.10, the sets of  $n$  vectors  $V = \{b_1 - b_0, \dots, b_n - b_0\}$  and  $W = \{b'_1 - b'_0, \dots, b'_n - b'_0\}$  are linearly independent in  $\mathbb{R}^n$ , hence they are bases of  $\mathbb{R}^n$ . It is well-known that there exists a unique one-to-one linear transformation  $A$  from  $\mathbb{R}^n$  to itself, mapping the basis  $V$  to the basis  $W$ . The desired affine transformation is then given by  $Tx = Ax + a$ , where  $a = b'_0 - Ab_0$ . Indeed,

$$Tb_j = Ab_j + (b'_0 - Ab_0) = A(b_j - b_0) + b'_0 = b'_j - b'_0 + b'_0 = b'_j \quad \text{for every } j = 1, \dots, n.$$

♥

**Corollary 1.18.** *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  $T$  of  $\mathbb{R}^n$  onto itself such that  $TM_1 = M_2$ .*

*Proof.* Let  $M_1, M_2$  be any two affine sets in  $\mathbb{R}^n$  of the same dimension  $m \leq n$ . By Theorem 1.11,  $M_1$  and  $M_2$  can be written as affine hulls of two affinely independent sets of  $(m + 1)$  points, respectively. The desired result follows immediately from Theorem 1.17 and the fact that affine transformation preserves affine hull (see Theorem 1.16). ♥

## 2. CONVEX SETS AND CONES

**2.1. Convex sets.** A set  $C \subseteq \mathbb{R}^n$  is said to be convex if  $(1 - \lambda)x + \lambda y \in C$  for any  $x, y \in C$  and  $\lambda \in [0, 1]$ . All affine sets are convex, but the converse is not true. For example, ellipsoids and cubes in  $\mathbb{R}^3$  are convex but not affine. **Half-spaces** are important examples of convex sets. For any  $b \in \mathbb{R}^n \setminus \{0\}$  and any  $\beta \in \mathbb{R}$ , **closed half-spaces** are the sets

$$\{x \in \mathbb{R}^n : \langle x, b \rangle \leq \beta\} \quad \text{and} \quad \{x \in \mathbb{R}^n : \langle x, b \rangle \geq \beta\},$$

and **open half-spaces** are the sets

$$\{x \in \mathbb{R}^n : \langle x, b \rangle < \beta\} \quad \text{and} \quad \{x \in \mathbb{R}^n : \langle x, b \rangle > \beta\}.$$

These sets depend only on the hyperplane  $H = \{x \in \mathbb{R}^n : \langle x, b \rangle = \beta\}$  and they are all nonempty and convex.

**Theorem 2.1** (Closed under arbitrary intersections). *Let  $C_i \subseteq \mathbb{R}^n$  be a convex set for any  $i \in I$ , where  $I$  is an arbitrary index set. Then the set  $C = \bigcap_{i \in I} C_i$  is convex.*

*Proof.* Suppose  $x, y \in C$  and  $\lambda \in [0, 1]$ . Then  $x, y \in C_i$  for any  $i \in I$  and since  $C_i$  is convex, it follows that  $\lambda x + (1 - \lambda)y \in C_i$  for any  $i \in I$ , i.e.,  $\lambda x + (1 - \lambda)y \in \bigcap_{i \in I} C_i = C$ . ♥

**Example 2.2.** Consider the convex **polyhedral**  $C$  defined by

$$C = \{x \in \mathbb{R}^n : \langle x, b_i \rangle \leq \beta_i \text{ for all } i \in I\},$$

where  $b_i \in \mathbb{R}^n$  and  $\beta_i \in \mathbb{R}$  for  $i \in I$ ,  $I$  a finite index set.  $C$  is convex since it can be written as the intersection of closed half spaces:

$$C = \bigcap_{i \in I} \{x \in \mathbb{R}^n : \langle x, b_i \rangle \leq \beta_i\}.$$

**Definition 2.3.** 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  $\lambda_i$  are all nonnegative and  $\lambda_1 + \cdots + \lambda_m = 1$ .

**Theorem 2.4.** *A subset of  $\mathbb{R}^n$  is convex if and only if it contains all the convex combinations of its elements.*

*Proof.* By definition, a set  $C$  is convex if and only if  $C$  is closed under taking convex combinations with  $m = 2$ . We must show that this implies  $C$  is also closed under taking convex combinations with  $m > 2$ . Let us prove this by induction on  $m$ . The base case  $m = 1$  is trivially true. Suppose the statement holds for some  $m \geq 1$ . Let  $x_1, \dots, x_{m+1} \in C$  and

$\lambda_1, \dots, \lambda_m$  be such that  $\lambda_1 + \dots + \lambda_{m+1} = 1$ . The statement collapses to the induction hypothesis if one of the scalars is 1, so assume WLOG that  $\lambda_{m+1} < 1$ . Then

$$\begin{aligned} z &= \sum_{j=1}^{m+1} \lambda_j x_j = \sum_{j=1}^m \lambda_j x_j + \lambda_{m+1} x_{m+1} \\ &= (1 - \lambda_{m+1}) \sum_{j=1}^m \left( \frac{\lambda_j}{1 - \lambda_{m+1}} \right) x_j + \lambda_{m+1} x_{m+1} \\ &= (1 - \lambda_{m+1}) y + \lambda_{m+1} x_{m+1}. \end{aligned}$$

Since  $\frac{\lambda_j}{1 - \lambda_{m+1}} \geq 0$  for any  $j = 1, \dots, m$  and

$$\sum_{j=1}^m \frac{\lambda_j}{1 - \lambda_{m+1}} = \frac{\lambda_1 + \dots + \lambda_m}{1 - \lambda_{m+1}} = \frac{1 - \lambda_{m+1}}{1 - \lambda_{m+1}} = 1,$$

it follows that  $y$  is a convex combination of  $m$  elements in  $C$  and we have  $y \in C$  by induction hypothesis. Since  $C$  is convex and  $z = (1 - \lambda_{m+1}) y + \lambda_{m+1} x_{m+1}$ , we conclude that  $z \in C$  as desired. ♥

**2.2. Convex hull.** Let  $S \subseteq \mathbb{R}^n$ . The **convex hull** of  $S$ , denoted by  $\text{conv}(S)$ , is the intersection of all the convex sets containing  $S$ , or equivalently, the unique smallest convex set containing  $S$ . This definition can be viewed as a characterisation of  $\text{conv}(S)$  from the *outside*, but it is not immediately clear how an element in  $\text{conv}(S)$  relates to an element in  $S$ . This next theorem provides an answer to this question, and it provides a characterisation of  $\text{conv}(S)$  from the *inside*.

**Theorem 2.5.** *For any  $S \subseteq \mathbb{R}^n$ ,  $\text{conv}(S)$  is the set of all convex combinations of elements of  $S$ .*

*Proof.* Let  $C$  denote the set of all convex combinations of elements of  $S$ . Since all elements of  $S$  belong to  $\text{conv}(S)$ , it follows from Theorem 2.4 that  $C \subset \text{conv}(S)$ . On the other hand, take any  $y, z \in C$ . There exists  $x_1, \dots, x_m, x_{m+1}, x_n \in S$  and scalars  $\lambda_1, \dots, \lambda_m, \lambda_{m+1}, \dots, \lambda_n \geq 0$  such that

$$y = \sum_{j=1}^m \lambda_j x_j, \quad z = \sum_{j=m+1}^n \lambda_j x_j, \quad \text{and} \quad \sum_{j=1}^m \lambda_j = \sum_{j=m+1}^n \lambda_j = 1.$$

For any  $\mu \in [0, 1]$ , we have

$$(1 - \mu)y + \mu z = (1 - \mu) \sum_{j=1}^m \lambda_j x_j + \mu \sum_{j=m+1}^n \lambda_j x_j = \sum_{j=1}^n \Lambda_j x_j$$

with

$$\sum_{j=1}^n \Lambda_j = \sum_{j=1}^m (1 - \mu)\lambda_j + \sum_{j=m+1}^n \mu\lambda_j = (1 - \mu) + \mu = 1.$$

Thus  $C$  is a convex set. Moreover,  $S \subset C$  from the definition of  $C$  and so  $\text{conv}(S) \subset C$ . The desired result follows. ♥

A set which is the convex hull of finitely many points is called a **polytope**. If  $\{b_0, b_1, \dots, b_m\}$  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. In terms of the barycentric coordinates on  $\text{aff}\{b_0, b_1, \dots, b_m\}$ , each point of the simplex is uniquely expressible as a convex combination of the vertices. The point

$$\lambda_0 b_0 + \lambda_1 b_1 + \dots + \lambda_m b_m \quad \text{with } \lambda_0 = \lambda_1 = \dots = \lambda_m = \frac{1}{1+m},$$

is called the **midpoint** or **barycenter** of the simplex. When  $m = 0, 1, 2$  or  $3$ , the simplex is a point, (closed) line segment, triangle or tetrahedron, respectively.

In general, the dimension of a convex set  $C$  refers to the dimension of the affine hull of  $C$ .

**Theorem 2.6.** *The dimension of a convex set  $C$  is the maximum of the dimensions of the various simplices included in  $C$ .*

**2.3. Convex cone.** A set  $K \subseteq \mathbb{R}^n$  is a **cone** if it is closed under positive scalar multiplication, *i.e.*,  $\lambda x \in K$  whenever  $x \in K$  and  $\lambda > 0$ . By definition, a cone is a union of half-lines emanating from the origin, but the origin itself may or may not be included. A set  $K \subseteq \mathbb{R}^n$  is a **convex cone** if it is a cone which is a convex set. Convex cones are not necessarily “pointed”. Examples of convex cones are subspaces of  $\mathbb{R}^n$ , halfspaces corresponding to a hyperplane through the origin, nonnegative orthant  $\mathbb{R}_+^n$  and positive orthant  $\mathbb{R}_{++}^n$ .

**Theorem 2.7.** *The intersection of an arbitrary collection of convex cones is a convex cone.*

*Proof.* Let  $K_i \subseteq \mathbb{R}^n$  be a convex cone for all  $i \in I$ , where  $I$  is an arbitrary index set. Consider  $K = \bigcap_{i \in I} K_i$ . The convexity of  $K$  follows from Theorem 2.1. It remains to show that  $K$  is a cone. Let  $x \in K$  and  $\lambda > 0$ . Then  $x \in K_i$  for any  $i \in I$  and since  $K_i$  is a cone, it follows that  $\lambda x \in K_i$  for any  $i \in I$ , *i.e.*,  $\lambda x \in K$ . ♥

The following characterisation of convex cones highlights an analogy between convex cones and subspaces.

**Theorem 2.8.** *A set  $K \subseteq \mathbb{R}^n$  is a convex cone if and only if the following properties hold:*

- (a)  $x, y \in K \implies x + y \in K$ .
- (b)  $x \in K, \lambda > 0 \implies \lambda x \in K$ .

*Proof.* Suppose  $K$  is a convex cone. Property (b) follows from the definition of a cone. Let  $x, y \in K$ . By the convexity of  $K$  we have  $z = \frac{1}{2}(x + y) \in K$  and hence  $x + y = 2z \in K$ , establishing property (a). Conversely, suppose  $K$  satisfies properties (a) and (b). Then  $K$  is a cone by property (b). Let  $x, y \in K$  and  $\lambda \in [0, 1]$ . Then  $(1 - \lambda)x, \lambda y \in K$  by property (b) and thus  $(1 - \lambda)x + \lambda y \in K$  by property (a), establishing the convexity of  $K$ . ♥

**Example 2.9.** Consider the set  $K$  defined by

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

where  $b_i \in \mathbb{R}^n$  for  $i \in I$ ,  $I$  an arbitrary index set. It is a convex cone since it can be written as

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

and each of these  $K_i$  is a convex cone since it is a closed half-space corresponding to a hyperplane through the origin. Indeed, suppose  $x \in K_i$  and  $\lambda > 0$ . Then

$$\langle \lambda x, b_i \rangle = \lambda \langle x, b_i \rangle \leq 0 \implies \lambda x \in K_i.$$

**Theorem 2.10.** *A set  $K \subseteq \mathbb{R}^n$  is a convex cone if and only if it contains all the positive linear combinations of its elements.*

*Proof.* We first show that  $K \subseteq \mathbb{R}^n$  is a convex cone if and only if  $\alpha x + \beta y \in K$  for any  $\alpha, \beta > 0$  and  $x, y \in K$ . Suppose  $K$  is a convex cone. Let  $\alpha, \beta > 0$  and  $x, y \in K$ . Then both  $\frac{\alpha}{\alpha + \beta}x$  and  $\frac{\beta}{\alpha + \beta}y$  are elements of  $K$  since  $K$  is a cone. By the convexity of  $K$  we have

$$z = \frac{\alpha}{\alpha + \beta}x + \frac{\beta}{\alpha + \beta}y = \frac{\alpha x + \beta y}{\alpha + \beta} \in K$$

and hence  $\alpha x + \beta y = (\alpha + \beta)z \in K$ . On the other hand, suppose  $\alpha x + \beta y \in K$  for any  $\alpha, \beta > 0$  and  $x, y \in K$ . Convexity of  $K$  follows immediately by choosing  $\alpha = (1 - \lambda)$  and  $\beta = \lambda$  for  $\lambda \in [0, 1]$ . To show that  $K$  is a cone, let  $x \in K$  and  $\lambda > 0$ . Then setting  $\alpha = \beta = \lambda/2$  yields

$$\lambda x = \frac{\lambda}{2}x + \frac{\lambda}{2}x \in K.$$

The general case follows from an induction argument. We point out that this result can be viewed as a corollary of Theorem 2.8. ♥

**Corollary 2.11.** *Let  $S \subseteq \mathbb{R}^n$  be arbitrary and  $K$  be the set of all positive linear combinations of  $S$ . Then  $K$  is the smallest convex cone which includes  $S$ .*

*Proof.* It is clear that  $S \subset K$  from the definition of  $K$  and  $K$  is a convex cone since it is closed under addition and positive scalar multiplication. Suppose  $S \subseteq T$  for some convex cone  $T$ . We need to show that  $K \subseteq T$ . To this end, take  $z \in K$ . There exists elements  $x_1, \dots, x_m \in S \subseteq T$  and  $\lambda_1, \dots, \lambda_m > 0$  such that

$$z = \lambda_1 x_1 + \dots + \lambda_m x_m.$$

Since  $T$  is a convex cone, it follows from Theorem 2.10 that  $z \in T$ . ♥

A simpler description is possible if  $S$  is convex.

**Corollary 2.12.** *Let  $C \subseteq \mathbb{R}^n$  be convex and let*

$$K = \{\lambda x \in \mathbb{R}^n : \lambda > 0, x \in C\}.$$

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

*Proof.* This follows from Corollary 2.11, since every positive linear combination of elements of  $C$  is a positive scalar multiple of a convex combination of elements of  $C$  and hence is an element of  $K$ . ♥

The convex cone obtained by adjoining the origin to the cone in Corollary 2.11 (or Corollary 2.12) is known as the **convex cone generated by  $S$  (or  $C$ )** (conic hull?) and is denoted by  $\text{cone}(S)$ . Thus 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  $S$  is nonempty, then  $\text{cone}(S)$  consists of all nonnegative linear combinations of elements of  $S$ . Clearly

$$\text{cone}(S) = \text{conv}(\text{ray}(S)),$$

where  $\text{ray}(S)$  is the union of the origin and the various rays (half-lines of the form  $\{\lambda y \in \mathbb{R}^n : \lambda \geq 0\}$ ) generated by the nonzero vectors  $y \in S$ .

Convex cones are usually simpler to handle than general convex sets, so it is useful to be able to convert a question about convex sets into a question about convex cones. The following proposition provides a way of doing this, and it is essentially a generalisation of that fact that circle is a cross-section of a three-dimensional cone.

**Proposition 2.13.** *Every convex set  $C \subseteq \mathbb{R}^n$  can be regarded as a cross-section of some convex cone  $K$  in  $\mathbb{R}^{n+1}$ .*

*Proof.* Consider the set of pairs  $S = \{(x, 1) \in \mathbb{R}^{n+1} : x \in C\}$  and let  $K = \text{cone}(S)$  be the convex cone generated by  $S$ . Consider the hyperplane  $H = \{(x, \lambda) \in \mathbb{R}^{n+1} : \lambda = 1\}$ . Since  $K$  consists of pairs  $(\lambda x, \lambda)$  with  $\lambda \geq 0$ , intersecting  $K$  with  $H$  may be regarded as  $C$ , upon dropping the extra dimension. ♥

A vector  $x^*$  is said to be **normal** to a convex set  $C$  at a point  $a \in C$  if  $x^*$  does not make an acute angle with any line segment in  $C$  with  $a$  as endpoint, *i.e.*,

$$\langle x - a, x^* \rangle \leq 0 \quad \text{for every } x \in C.$$

Take  $C$  to be the half-space  $\{x \in \mathbb{R}^n : \langle x, b \rangle \leq \beta\}$  for some  $\beta \in \mathbb{R}$  and  $b \in \mathbb{R}^n \setminus \{0\}$ . If  $a \in \mathbb{R}^n$  is such that  $\langle a, b \rangle = \beta$ , then  $b$  is normal to  $C$  at  $a$ . Indeed,

$$\langle x - a, b \rangle = \langle x, b \rangle - \langle a, b \rangle = \langle x, b \rangle - \beta \leq 0 \quad \text{for any } x \in C.$$

In general, the set of all vectors  $x^*$  normal to  $C$  at  $a$  is called the **normal cone** to  $C$  at  $a$ . One can verify that the normal cone is always convex.

Another example of a convex cone is the **barrier cone** of a convex set  $C$ . This 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  $x \in C$ . Each convex cone containing  $0$  is associated with a pair of subspaces as follows.

**Theorem 2.14.** *Let  $K \subseteq \mathbb{R}^n$  be a convex cone containing  $0$ . There is a smallest subspace containing  $K$ , namely*

$$K - K = \{x - y \in \mathbb{R}^n : x, y \in K\} = \text{aff}(K),$$

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

*Proof.* By Theorem 2.8,  $K$  is closed under addition and positive scalar multiplication. To be a subspace, a set must further contain  $0$  and be closed under multiplication by  $-1$ . Clearly  $K - K$  is the smallest such set containing  $K$ , and  $(-K) \cap K$  is the largest such set contained within  $K$ . The former must coincide with  $\text{aff}(K)$ , since the affine hull of a set containing  $0$  is a subspace by Theorem 1.1. ♥



### 3. THE ALGEBRA OF CONVEX SETS

Establishing convexity of sets directly from the definition of a convex set can be tedious and often requires a cunning observation. In this section we will describe some operations that preserve convexity of sets, and these operations allow us to prove that a set is convex by constructing it from simple sets for which convexity is known.

If  $C \subseteq \mathbb{R}^n$  is convex,  $\lambda \in \mathbb{R}$ , and  $a \in \mathbb{R}^n$ , then the sets  $C + a$  and  $\lambda C$  are convex, where

$$C + a = \{x + a : x \in C\} \quad \text{and} \quad \lambda C = \{\lambda x : x \in C\}.$$

A convex set  $C \subseteq \mathbb{R}^n$  is said to be **symmetric** if  $-C = C$ . Such a set, if nonempty, 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 nonempty convex cones which are symmetric are the subspaces.

**Theorem 3.1.** *If  $C_1, C_2 \subseteq \mathbb{R}^n$  are convex, 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  $x, y \in C_1 + C_2$ . There exists elements  $x_1, y_1 \in C_1$  and  $x_2, y_2 \in C_2$  such that  $x = x_1 + x_2$  and  $y = y_1 + y_2$ . For any  $\lambda \in [0, 1]$ , we have

$$(1 - \lambda)x + \lambda y = \left[ (1 - \lambda)x_1 + \lambda y_1 \right] + \left[ (1 - \lambda)x_2 + \lambda y_2 \right] \in C_1 + C_2,$$

since  $(1 - \lambda)x_1 + \lambda y_1 \in C_1$  and  $(1 - \lambda)x_2 + \lambda y_2 \in C_2$  by the convexity of  $C_1$  and  $C_2$ , respectively. ♥

The convexity of a set  $C$  means by definition that

$$(1 - \lambda)C + \lambda C \subset C, \quad \lambda \in (0, 1).$$

It turns out that equality actually holds for convex sets. A set  $K$  is a convex cone if and only if  $\lambda K \subset K$  for every  $\lambda > 0$  and  $K + K \subset K$ .

If  $C_1, \dots, C_m$  are convex sets, then so is the linear combination

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

Such  $C$  is called a **convex combination** of  $C_1, \dots, C_m$  when all the scalars are nonnegative and  $\lambda_1 + \dots + \lambda_m = 1$ . In that case, it is appropriate to think of  $C$  geometrically as a sort of mixture of  $C_1, \dots, C_m$ . For the sake of geometric intuition, it is sometimes helpful to regard  $C_1 + C_2$  as the union of all the translates  $x_1 + C_2$  as  $x_1$  varies over  $C_1$ .

Without convexity being involved, one has the following algebraic laws of sets:

$$\begin{aligned} C_1 + C_2 &= C_2 + C_1 \\ (C_1 + C_2) + C_3 &= C_1 + (C_2 + C_3) \\ \lambda_1(\lambda_2 C) &= (\lambda_1 \lambda_2)C \\ \lambda(C_1 + C_2) &= \lambda C_1 + \lambda C_2. \end{aligned}$$

The convex set consisting of  $\mathbf{0}$  alone is the identity element for the addition operation. Additive inverses do not exist for sets containing more than one element, the best one can say in general is that  $\mathbf{0} \in C + (-C)$  when  $C \neq \emptyset$ .

**Theorem 3.2.** *If  $C \subseteq \mathbb{R}^n$  is convex and  $\alpha_1, \alpha_2 \geq 0$ , then*

$$(\alpha_1 + \alpha_2)C = \alpha_1 C + \alpha_2 C.$$

*Proof.* If either  $\lambda_1$  or  $\lambda_2$  is 0, then the result is trivial, so suppose  $\lambda_1, \lambda_2 > 0$ . The inclusion  $\subset$  would be true without the convexity of  $C$ . The reverse inclusion follows from the convexity relation

$$\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)C + \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)C \subset C$$

which is well-defined since  $\lambda_1 + \lambda_2 > 0$ . ♥

Given any two convex sets  $C_1$  and  $C_2$  in  $\mathbb{R}^n$ , there is a unique largest convex set  $C_1 \cap C_2$  contained in both  $C_1$  and  $C_2$ , and a unique smallest convex set  $\text{conv}(C_1 \cup C_2)$  containing both  $C_1$  and  $C_2$ . This holds for an arbitrary family of convex sets as well, which means that the collection of all convex subsets of  $\mathbb{R}^n$  is a **complete lattice** under the natural partial ordering corresponding to inclusion.

**Theorem 3.3.** *Let  $\{C_i: i \in I\}$  be an arbitrary collection of nonempty convex sets in  $\mathbb{R}^n$ , and let  $C$  be the convex hull of the union of the collection. Then*

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

where the union is taken over all finite convex combinations, i.e., over all nonnegative choices of the coefficients  $\lambda_i$  such that only finitely many are nonzero and these add up to 1.

*Proof.* By Theorem (convex combinations),  $C$  is the set of all convex combinations of elements of union of the sets  $C_i$ . Actually, we can get  $C$  just by taking those convex combinations in which the coefficients are nonzero and vectors are taken from different sets  $C_i$ . Indeed, elements with zero coefficients can be omitted, and if two of the elements with positive coefficients belong to the same  $C_i$ , say  $x_1$  and  $x_2$ , then the term  $\lambda_1 x_1 + \lambda_2 x_2$  can be written as  $\lambda x$ , where

$$\lambda = \lambda_1 + \lambda_2 \quad \text{and} \quad x = \left(\frac{\lambda_1}{\lambda}\right)x_1 + \left(\frac{\lambda_2}{\lambda}\right)x_2 \in C_i.$$

Thus  $C$  is the union of the finite convex combinations of the form

$$\lambda_1 C_{i_1} + \cdots + \mu_m C_{i_m},$$

where the indices  $i_1, \dots, i_m$  are distinct. ♥

**Theorem 3.4.** *Let  $A$  be a linear transformation from  $\mathbb{R}^n$  to  $\mathbb{R}^m$ .*

(a) *If  $C \subseteq \mathbb{R}^n$  is convex, then the image of  $C$  under  $A$ ,*

$$A(C) = \{Ax \in \mathbb{R}^m: x \in C\},$$

*is convex.*

(b) *If  $D \subseteq \mathbb{R}^m$  is convex, then the inverse image of  $D$  under  $A$ ,*

$$A^{-1}(D) = \{x \in \mathbb{R}^n: Ax \in D\},$$

*is convex.*

*Proof.* Let  $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a linear transformation. To prove part (a), let  $C \subseteq \mathbb{R}^n$  be convex,  $y_1, y_2 \in A(C)$  and  $\lambda \in [0, 1]$ . There exists elements  $x_1, x_2 \in C$  such that  $Ax_1 = y_1$  and  $Ax_2 = y_2$ . We will show that the point  $z = (1 - \lambda)y_1 + \lambda y_2 \in A(C)$ . Indeed,

$$z = (1 - \lambda)Ax_1 + \lambda Ax_2 = A\left[(1 - \lambda)x_1 + \lambda x_2\right] \in A(C),$$

since  $(1 - \lambda)x_1 + \lambda x_2 \in C$  by the convexity of  $C$ . To prove part (b), let  $D \subseteq \mathbb{R}^m$  be convex,  $x_1, x_2 \in A^{-1}(D)$  and  $\lambda \in [0, 1]$ . There exists elements  $y_1, y_2 \in D$  such that  $Ax_1 = y_1$  and  $Ax_2 = y_2$ . We will show that the point  $z = (1 - \lambda)x_1 + \lambda x_2 \in A^{-1}(D)$ , or equivalently,  $Az \in D$ . Indeed,

$$Az = A\left[(1 - \lambda)x_1 + \lambda x_2\right] = (1 - \lambda)Ax_1 + \lambda Ax_2 = (1 - \lambda)y_1 + \lambda y_2 \in D,$$

since  $D$  is convex. ♥

**Theorem 3.5.** *The orthogonal projection of a convex set  $C \subseteq \mathbb{R}^n$  onto a subspace  $L$  is another convex set.*

*Proof.* The orthogonal projection onto  $L$  is a linear transformation, the one which assigns to each point  $x$  the unique  $y \in L$  such that  $(x - y) \perp L$ . ♥

#### 4. CONVEX FUNCTIONS

Let  $f$  be an *extended real-valued function*  $f: S \rightarrow \mathbb{R} \cup \{\pm\infty\}$  whose domain is a subset  $S$  of  $\mathbb{R}^n$ . The **epigraph** of  $f$  is the subset of  $\mathbb{R}^{n+1}$  given by

$$\text{epi}(f) = \{(x, \mu) \in S \times \mathbb{R}: f(x) \leq \mu\} \subseteq \mathbb{R}^{n+1}.$$

We said that  $f$  is 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. Observe that  $\text{epi}(f) = \emptyset$  if and only if  $f$  is identically equal to  $+\infty$ .

The **effective domain** of a convex function  $f$  on  $S$ , denoted by  $\text{dom}(f)$ , is the projection of  $\text{epi}(f)$  onto  $\mathbb{R}^n$ :

$$\begin{aligned} \text{dom}(f) &= \{x \in \mathbb{R}^n: (x, \mu) \in \text{epi}(f) \text{ for some } \mu\} \\ &= \{x \in \mathbb{R}^n: f(x) < +\infty\}. \end{aligned}$$

This is a convex set in  $\mathbb{R}^n$  since it is the image of the convex set  $\text{epi}(f)$  under a linear transformation. Its dimension is called the dimension of  $f$ . Trivially, the convexity of  $f$  is equivalent to that of the restriction of  $f$  to  $\text{dom}(f)$ .

At this point, there are two technical approaches in terms of handling the effective domain.

- (1) One could limit attention to functions which are nowhere  $+\infty$ , so that  $S$  coincides with  $\text{dom}(f)$  but would vary with  $f$ .
- (2) One could limit attention to functions which given on all of  $\mathbb{R}^n$ , 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  $x \notin S$ .

The second approach will be adopted and so unless stated otherwise, by a convex function we shall always mean a “convex function with possibly infinite values which is defined throughout

the space  $\mathbb{R}^{n''}$ . Unfortunately, this approach leads to arithmetic calculations involving  $+\infty$  and  $-\infty$ . The rules we adopt are the obvious ones:

$$\begin{aligned}
\alpha + \infty &= \infty + \alpha = \infty && \text{for } \alpha \in (-\infty, +\infty] \\
\alpha - \infty &= -\infty + \alpha = -\infty && \text{for } \alpha \in [-\infty, +\infty) \\
\alpha \infty &= \infty \alpha = \infty && \text{for } \alpha \in (0, +\infty] \\
\alpha(-\infty) &= (-\infty)\alpha = -\infty && \text{for } \alpha \in (0, +\infty] \\
\alpha \infty &= \infty \alpha = -\infty && \text{for } \alpha \in [-\infty, 0) \\
\alpha(-\infty) &= (-\infty)\alpha = \infty && \text{for } \alpha \in [-\infty, 0) \\
0 \infty &= \infty 0 = 0 = 0(-\infty) = (-\infty)0 \\
-(-\infty) &= \infty \\
\inf \emptyset &= +\infty \\
\sup \emptyset &= -\infty.
\end{aligned}$$

The combinations  $\infty - \infty$  and  $-\infty + \infty$  are undefined. One can verify that under these rules, the familiar laws of arithmetic are still valid provided none of the binary sums is the forbidden  $\infty - \infty$  (or  $-\infty + \infty$ ).

A convex function  $f$  is said to be **proper** if  $\text{epi}(f)$  is nonempty and contains no vertical lines, *i.e.*,  $f(x) < +\infty$  for at least one  $x$  and  $f(x) > -\infty$  for every  $x$ . Thus  $f$  is proper if and only if the convex set  $C = \text{dom}(f)$  is nonempty and the restriction of  $f$  to  $C$  is finite. Equivalently, a proper convex function on  $\mathbb{R}^n$  are the function obtained by taking a finite convex function  $f$  on a nonempty convex set  $C$  and then extending it to all of  $\mathbb{R}^n$  by setting  $f(x) = +\infty$  for  $x \notin C$ .

A convex function which is not proper is *improper*. Improper convex functions are of interest mainly as possible by-products of various constructions. Examples of improper convex function which are not identically  $+\infty$  or  $-\infty$  is the function  $f$  on  $\mathbb{R}$  defined by

$$f(x) = \begin{cases} -\infty & \text{if } |x| < 1, \\ 0 & \text{if } |x| = 1, \\ +\infty & \text{if } |x| > 1. \end{cases}$$

or the function  $g$  on  $\mathbb{R}$  defined by

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

By definition,  $f$  is convex on  $S$  if and only if

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

whenever  $(x, \mu), (y, \nu) \in \text{epi}(f)$  and  $\lambda \in [0, 1]$ . In other words, we must have  $(1 - \lambda)x + \lambda y \in S$  and

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

whenever  $x, y \in S$ ,  $f(x) \leq \mu \in \mathbb{R}$ ,  $f(y) \leq \nu \in \mathbb{R}$  and  $\lambda \in [0, 1]$ .

**Theorem 4.1.** Let  $f$  be a function from  $C$  to  $(-\infty, +\infty]$ , where  $C \subseteq \mathbb{R}^n$  is a convex set. Then  $f$  is convex on  $C$  if and only if

$$f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y) \quad \text{for all } x, y \in C \text{ and } \lambda \in (0, 1).$$

**Theorem 4.2.** Let  $f$  be a function from  $\mathbb{R}^n$  to  $[-\infty, +\infty]$ . Then  $f$  is convex if and only if for every  $\lambda \in (0, 1)$  we have

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

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

*Proof.* Suppose  $f$  is convex, then  $\text{epi}(f)$  is convex as a subset of  $\mathbb{R}^{n+1}$ . Let  $\lambda \in (0, 1)$  and  $x, y \in \mathbb{R}^n$  be such that  $f(x) < \alpha$  and  $f(y) < \beta$ . Then

♥

**Theorem 4.3** (Jensen's inequality). Let  $f$  be a function from  $\mathbb{R}^n$  to  $(-\infty, +\infty]$ . Then  $f$  is convex if and only if

$$f\left(\sum_{j=1}^m \lambda_j x_j\right) \leq \sum_{j=1}^m \lambda_j f(x_j),$$

whenever  $\lambda_1, \dots, \lambda_m$  are nonnegative and  $\lambda_1 + \dots + \lambda_m = 1$ .

Concave functions satisfy the opposite inequalities under similar hypotheses. Affine functions satisfy the inequalities as equations. Thus the affine functions on  $\mathbb{R}^n$  are the affine transformations from  $\mathbb{R}^n$  to  $\mathbb{R}$ . In the multidimensional case, it follows easily from Theorem 4.1 that every function of the form

$$f(x) = \langle x, b \rangle + \alpha, \quad b \in \mathbb{R}^n, \quad \alpha \in \mathbb{R},$$

is finite, convex and concave, thus affine in  $\mathbb{R}^n$ . In fact, every affine function on  $\mathbb{R}^n$  is of this form.

#### 4.1. Second-order characterisation of convex functions.

**Theorem 4.4.** Let  $f \in C^2(\alpha, \beta)$ . Then  $f$  is convex if and only if its second derivative  $f''$  is nonnegative throughout  $(\alpha, \beta)$ .

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

$$\begin{aligned} f(z) - f(x) &= \int_x^z f'(t) dt \leq f'(z)(z - x) f'(z) \left[ (1 - \lambda)x + \lambda y - x \right] \\ &= f'(z) \left[ \lambda(y - x) \right] \end{aligned} \tag{4.1}$$

$$\begin{aligned} f(z) - f(y) &= \int_y^z f'(t) dt \leq f'(z)(z - y) = f'(z) \left[ (1 - \lambda)x + \lambda y - y \right] \\ &= f'(z) \left[ (1 - \lambda)(x - y) \right]. \end{aligned} \tag{4.2}$$

Multiplying (4.1) by  $(1 - \lambda)$  and (4.2) by  $\lambda$  and adding them together, we obtain

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

but the LHS is just  $f(z) = f((1 - \lambda)x + \lambda y)$ . Conversely, suppose  $f''$  is not nonnegative on  $(\alpha, \beta)$ . Then  $f'' < 0$  on some subinterval  $(\alpha', \beta')$  by continuity of  $f''$ . A similar argument as before shows that for  $\alpha' < x < y < \beta'$ ,  $\lambda \in (0, 1)$  and  $z = (1 - \lambda)x + \lambda y$ , we have

$$\begin{aligned} f(z) - f(x) &> f'(z)(z - x) \\ f(z) - f(y) &> f'(z)(z - y) \end{aligned}$$

and hence  $f(z) > (1 - \lambda)f(x) + \lambda f(y)$ . Thus  $f$  is not convex on  $(\alpha, \beta)$ . ♥

**Example 4.5.** Below are some functions on  $\mathbb{R}$  whose convexity is a consequence of Theorem 4.4.

- (1)  $f(x) = e^{\alpha x}$ , where  $\alpha \in \mathbb{R}$ .
- (2)  $f(x) = \begin{cases} x^p & \text{if } x \geq 0, \\ +\infty & \text{if } x < 0, \end{cases}$  where  $p \in [1, \infty)$ .
- (3)  $f(x) = \begin{cases} -x^p & \text{if } x \geq 0, \\ +\infty & \text{if } x < 0, \end{cases}$  where  $p \in [0, 1]$ .
- (4)  $f(x) = \begin{cases} x^p & \text{if } x > 0, \\ +\infty & \text{if } x \leq 0, \end{cases}$  where  $p \in (-\infty, 0]$ .
- (5)  $f(x) = \begin{cases} (\alpha^2 - x^2)^{-1/2} & \text{if } |x| < \alpha, \\ +\infty & \text{if } |x| \geq \alpha, \end{cases}$  where  $\alpha > 0$ .
- (6)  $f(x) = \begin{cases} -\log x & \text{if } x > 0, \\ \infty & \text{if } x \leq 0, \end{cases}$

**Theorem 4.6.** Let  $C \subseteq \mathbb{R}^n$  be convex and  $f \in C^2(C)$ . Then  $f$  is convex on  $C$  if and only if its Hessian matrix

$$Q_x = (q_{ij}(x)), \quad q_{ij}(x) = \frac{\partial^2 f}{\partial \xi_i \partial \xi_j}(\xi_1, \dots, \xi_n),$$

is positive semidefinite 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  $C$ , i.e., the convexity of the one-dimensional function  $g_{x,d}(\lambda) = f(x + \lambda d)$  on the open real interval  $\{\lambda \in \mathbb{R} : y + \lambda z \in C\}$  for each  $y \in C$  and  $z \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ . Let  $y = x + \lambda d$ . A straightforward calculation reveals that

$$\begin{aligned} g'(\lambda) &= \langle \nabla_y f, d \rangle \\ g''(\lambda) &= \langle \nabla_y^2 f d, d \rangle = \langle Q_y d, d \rangle. \end{aligned}$$

It follows from Theorem 4.4 that  $g$  is convex for each  $x \in C$  and  $d \in \mathbb{R}^n \setminus \{\mathbf{0}\}$  if and only if  $\langle Q_y d, d \rangle \geq 0$  for every  $y \in C$  and  $d \in \mathbb{R}^n$ . ♥

**Theorem 4.7.** Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  be the quadratic function defined by

$$f(x) = \frac{1}{2} \langle x, Ax \rangle + \langle x, b \rangle + \alpha,$$

where  $A \in \mathbb{R}^{n \times n}$  is symmetric,  $b \in \mathbb{R}^n$ , and  $\alpha \in \mathbb{R}$ . Then  $f$  is convex on  $\mathbb{R}^n$  if and only if  $A$  is positive semidefinite on  $\mathbb{R}^n$ .

**Example 4.8.** Consider the negative of the geometric mean:

$$f(x) = f(\xi_1, \dots, \xi_n) = \begin{cases} -(\xi_1 \xi_2 \dots \xi_n)^{1/n} & \text{if } \xi_1, \dots, \xi_n \geq 0, \\ +\infty & \text{otherwise.} \end{cases}$$

A straightforward computation shows that

**4.2. Correspondence between convex sets and convex functions.** Given any set  $C \subseteq \mathbb{R}^n$ , we associate with  $C$  the **indicator function**  $\delta_C: \mathbb{R}^n \rightarrow (-\infty, +\infty]$  defined by

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

The epigraph of  $\delta_C$  is a “half-cylinder with cross-section  $C$ ”.

**Theorem 4.9.** A set  $C \subseteq \mathbb{R}^n$  is convex if and only if  $\delta_C(x)$  is convex.

*Proof.* Suppose that  $C \subseteq \mathbb{R}^n$  is convex. Let  $x, y \in \mathbb{R}^n$  and  $\lambda \in (0, 1)$ . For the “only if” direction, by Theorem 4.1 it suffices to show that

$$(4.3) \quad \delta_C((1 - \lambda)x + \lambda y) \leq (1 - \lambda)\delta_C(x) + \lambda\delta_C(y).$$

This inequality is trivially satisfied if either  $x \notin C$  or  $y \notin C$ , so suppose  $x, y \in C$ . By convexity of  $C$ ,  $(1 - \lambda)x + \lambda y \in C$  and thus

$$\delta_C((1 - \lambda)x + \lambda y) = 0 = (1 - \lambda)\delta_C(x) + \lambda\delta_C(y).$$

To prove the “if” direction, suppose  $\delta_C$  is convex on  $\mathbb{R}^n$ . Let  $x, y \in C$  and  $\lambda \in (0, 1)$ . We need to show that  $z = (1 - \lambda)x + \lambda y \in C$ . The convexity of  $\delta_C$  implies that

$$\delta_C(z) \leq (1 - \lambda)\delta_C(x) + \lambda\delta_C(y) = 0.$$

This immediately shows that  $\delta_C(z) = 0$  since  $\delta_C$  only takes the values 0 and  $+\infty$ . Thus  $z \in C$  and the convexity of  $C$  follows. ♥

Let  $C \subseteq \mathbb{R}^n$  be convex. The **support function**  $\delta_C^*$  of  $C$  is defined by

$$\delta_C^*(x) = \sup_{y \in C} \langle x, y \rangle, \quad x \in \mathbb{R}^n.$$

The **gauge function**  $\gamma_C$  of  $C$  is defined by

$$\gamma_C(x) = \inf_{x \in \lambda C} \lambda, \quad C \neq \emptyset, \quad x \in \mathbb{R}^n.$$

The **Euclidean distance function** of  $C$  is defined by

$$d(x, C) = \inf_{y \in C} |x - y|.$$

### 4.3. Sublevel sets of convex functions.

**Theorem 4.10.** For any convex function  $f$  and any  $\alpha \in [-\infty, +\infty]$ , the level sets  $\{x \in \mathbb{R}^n : f(x) < \alpha\}$  and  $\{x \in \mathbb{R}^n : f(x) \leq \alpha\}$  are convex.

*Proof.* The case of strict inequality is immediately from Theorem 4.2, with  $\beta = \alpha$ . The convexity of the set  $M = \{x \in \mathbb{R}^n : f(x) \leq \alpha\}$  follows from the fact that

$$M = \bigcap_{\mu > \alpha} \{x \in \mathbb{R}^n : f(x) < \mu\}.$$

Geometrically,  $M$  is the projection on  $\mathbb{R}^n$  of the intersection of  $\text{epi}(f)$  and the horizontal hyperplane  $\{(x, \alpha)\} \subseteq \mathbb{R}^{n+1}$ , so that  $M$  can be regarded as a horizontal cross-section of  $\text{epi}(f)$ . ♥

**Corollary 4.11.** Let  $f_i$  be a convex function on  $\mathbb{R}^n$  and  $\alpha_i \in \mathbb{R}$  for each  $i \in I$ , where  $I$  is an arbitrary index set. Then the set

$$C = \bigcap_{i \in I} \{x \in \mathbb{R}^n : f_i(x) \leq \alpha_i\}$$

is a convex set.

**Example 4.12.** Ellipsoids, paraboloids, and spherical balls are convex since each of these can be realised as the set of points satisfying a quadratic inequality of the form

$$\frac{1}{2} \langle x, Ax \rangle + \langle x, b \rangle + \alpha \leq 0$$

for some symmetric positive semidefinite matrix  $Q \in \mathbb{R}^{n \times n}$  and some  $b \in \mathbb{R}^n$  and  $\alpha \in \mathbb{R}$ .

Theorem 4.10 and Corollary 4.11 have a clear significance for the theory of systems of nonlinear inequalities. But convexity proves to be a powerful tool for establishing classical inequalities such as arithmetic geometric mean inequality, Young's inequality, Hölder's inequality and Minkowski's inequality.

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

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

Positive homogeneity is equivalent to the epigraph being a cone in  $\mathbb{R}^{n+1}$ . An example of a positively homogeneous convex function that is not a linear function is  $f(x) = \lambda|x|$ .

**Theorem 4.13.** A positively homogeneous function  $f: \mathbb{R}^n \rightarrow (-\infty, +\infty]$  is convex if and only if

$$f(x + y) \leq f(x) + f(y) \quad \text{for every } x, y \in \mathbb{R}^n.$$

*Proof.* This follows from the characterisation of convex cone, since the subadditivity condition of  $f$  is equivalent to  $\text{epi}(f)$  being closed under addition. ♥



**Corollary 4.14.** *If  $f$  is a positively homogeneous proper convex function, then*

$$f\left(\sum_{j=1}^m \lambda_j x_j\right) \leq \sum_{j=1}^m \lambda_j f(x_j),$$

whenever  $\lambda_1, \dots, \lambda_m > 0$ .

**Corollary 4.15.** *If  $f$  is a positively homogeneous proper convex function, then  $f(-x) \geq -f(x)$  for every  $x$ .*

*Proof.*

$$f(-x) + f(x) \geq f(x - x) = f(0) \geq 0.$$

♥

**Theorem 4.16.** *A positively homogeneous proper convex function  $f$  is linear on a subspace  $L$  if and only if  $f(-x) = -f(x)$  for every  $x \in L$ . This 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.* The “only if” direction is clear. Suppose  $f(-b_i) = -f(b_i)$  for all the vectors in some basis  $\{b_1, \dots, b_m\}$  for  $L$ . Then

$$f(\lambda_i b_i) = \lambda_i f(b_i) \quad \text{for all } \lambda_i \in \mathbb{R}.$$

For any  $x = \lambda_1 b_1 + \dots + \lambda_m b_m \in L$ , we have

$$\begin{aligned} \sum_{j=1}^m f(\lambda_j b_j) &\geq f(x) && \left[ \text{From Theorem 4.13.} \right] \\ &\geq -f(-x) && \left[ \text{From Corollary 4.15.} \right] \\ &\geq -\sum_{j=1}^m f(-\lambda_j b_j) && \left[ \text{From Theorem 4.13.} \right] \\ &= \sum_{j=1}^m f(\lambda_j b_j). \end{aligned}$$

Thus  $f$  is linear on  $L$ , and in particular  $f(-x) = -f(x)$  for every  $x \in L$ .

♥

## REFERENCES

- [BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [Gal11] J. Gallier. *Geometric methods and applications: for computer science and engineering*. 2nd. Vol. 38. Texts in Applied Mathematics. Springer Science & Business Media, 2011. DOI: [10.1007/978-1-4419-9961-0](https://doi.org/10.1007/978-1-4419-9961-0).
- [Roc70] R. T. Rockafellar. *Convex analysis*. Princeton university press, 1970.