Chapter 6
Convex Sets

In this chapter we begin our exploration of convex analysis, which is the mathematical theory essential for analyzing and understanding the theoretical and practical aspects of optimization.

6.1 ▪ Definition and Examples

We begin with the definition of a convex set.

Definition 6.1 (convex sets). *A set $C \subseteq \mathbb{R}^n$ is called* **convex** *if for any* $\mathbf{x},\mathbf{y} \in C$ *and* $\lambda \in [0,1]$, *the point* $\lambda\mathbf{x}+(1-\lambda)\mathbf{y}$ *belongs to* C.

The above definition is equivalent to saying that for any $\mathbf{x},\mathbf{y} \in C$, the line segment $[\mathbf{x},\mathbf{y}]$ is also in C. Examples of convex and nonconvex sets in $\mathbb{R}^2$ are illustrated in Figure 6.1. We will now show some basic examples of convex sets.

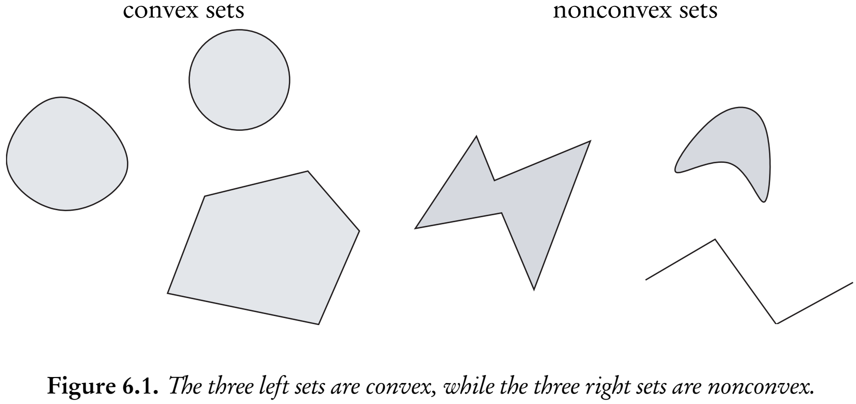

Figure 6.1. *The three left sets are convex, while the three right sets are nonconvex.*

Example 6.2 (convexity of lines). A line in $\mathbb{R}^n$ is a set of the form

$$L = \{\mathbf{z} + t\mathbf{d} : t \in \mathbb{R}\},$$

where $\mathbf{z},\mathbf{d} \in \mathbb{R}^n$ and $\mathbf{d} \neq \mathbf{0}$. To show that L is indeed a convex set, let us take $\mathbf{x},\mathbf{y} \in L$. Then there exist $t_1, t_2 \in \mathbb{R}$ such that $\mathbf{x} = \mathbf{z} + t_1\mathbf{d}$ and $\mathbf{y} = \mathbf{z} + t_2\mathbf{d}$. Therefore, for any

$\lambda \in [0,1]$ we have

$$\lambda\mathbf{x}+(1-\lambda)\mathbf{y} = \lambda(\mathbf{z}+t_1\mathbf{d})+(1-\lambda)(\mathbf{z}+t_2\mathbf{d}) = \mathbf{z}+(\lambda t_1+(1-\lambda)t_2)\mathbf{d} \in L. \qquad \blacksquare$$

Similarly we can show that for any $\mathbf{x},\mathbf{y} \in \mathbb{R}^n$, the closed and open line segments $[\mathbf{x},\mathbf{y}],(\mathbf{x},\mathbf{y})$ are also convex sets. Simpler examples of convex sets are the empty set $\emptyset$ and the entire space $\mathbb{R}^n$. A *hyperplane* is a set of the form $H = \{\mathbf{x}\in\mathbb{R}^n : \mathbf{a}^T\mathbf{x} = b\}$, where $\mathbf{a}\in\mathbb{R}^n\backslash\{\mathbf{0}\}, b\in\mathbb{R}$, and the associated *half-space* is the set $H^- = \{\mathbf{x}\in\mathbb{R}^n : \mathbf{a}^T\mathbf{x}\le b\}$. Both hyperplanes and half-spaces are convex sets.

Lemma 6.3 (convexity of hyperplanes and half-spaces). *Let* $\mathbf{a}\in\mathbb{R}^n\backslash\{\mathbf{0}\}$ *and* $b\in\mathbb{R}$. *Then the following sets are convex:*

(a) *the hyperplane* $H = \{\mathbf{x}\in\mathbb{R}^n : \mathbf{a}^T\mathbf{x} = b\}$,

(b) *the half-space* $H^- = \{\mathbf{x}\in\mathbb{R}^n : \mathbf{a}^T\mathbf{x}\le b\}$,

(c) *the open half-space* $\{\mathbf{x}\in\mathbb{R}^n : \mathbf{a}^T\mathbf{x} < b\}$.

Proof. We will prove only the convexity of the half-space since the proof of convexity of the other two sets is almost identical. Let $\mathbf{x},\mathbf{y}\in H^-$ and let $\lambda\in[0,1]$. We will show that $\mathbf{z} = \lambda\mathbf{x}+(1-\lambda)\mathbf{y}\in H^-$. Indeed,

$$\mathbf{a}^T\mathbf{z} = \mathbf{a}^T[\lambda\mathbf{x}+(1-\lambda)\mathbf{y}] = \lambda(\mathbf{a}^T\mathbf{x})+(1-\lambda)(\mathbf{a}^T\mathbf{y}) \le \lambda b+(1-\lambda)b = b,$$

where the inequality in the above chain of equalities and inequalities follows from the fact that $\mathbf{a}^T\mathbf{x}\le b, \mathbf{a}^T\mathbf{y}\le b$, and $\lambda\in[0,1]$. $\square$

Other important examples of convex sets are the closed and open balls.

Lemma 6.4 (convexity of balls). *Let* $\mathbf{c}\in\mathbb{R}^n$ *and* $r > 0$. *Let* $\|\cdot\|$ *be an arbitrary norm defined on* $\mathbb{R}^n$. *Then the open ball*

$$B(\mathbf{c},r) = \{\mathbf{x}\in\mathbb{R}^n : \|\mathbf{x}-\mathbf{c}\| < r\}$$

and closed ball

$$B[\mathbf{c},r] = \{\mathbf{x}\in\mathbb{R}^n : \|\mathbf{x}-\mathbf{c}\| \le r\}$$

are convex.

Proof. We will show the convexity of the closed ball. The proof of the convexity of the open ball is almost identical. Let $\mathbf{x},\mathbf{y}\in B[\mathbf{c},r]$ and let $\lambda\in[0,1]$. Then

$$\|\mathbf{x}-\mathbf{c}\|\le r, \|\mathbf{y}-\mathbf{c}\|\le r. \tag{6.1}$$

Let $\mathbf{z} = \lambda\mathbf{x}+(1-\lambda)\mathbf{y}$. We will show that $\mathbf{z}\in B[\mathbf{c},r]$. Indeed,

$$\begin{aligned}
\|\mathbf{z}-\mathbf{c}\| &= \|\lambda\mathbf{x}+(1-\lambda)\mathbf{y}-\mathbf{c}\| \\
&= \|\lambda(\mathbf{x}-\mathbf{c})+(1-\lambda)(\mathbf{y}-\mathbf{c})\| \\
&\le \|\lambda(\mathbf{x}-\mathbf{c})\|+\|(1-\lambda)(\mathbf{y}-\mathbf{c})\| && \text{(triangle inequality)} \\
&= \lambda\|\mathbf{x}-\mathbf{c}\|+(1-\lambda)\|\mathbf{y}-\mathbf{c}\| && (0\le\lambda\le 1) \\
&\le \lambda r+(1-\lambda)r = r && \text{(equation (6.1))}.
\end{aligned}$$

Hence, $\mathbf{z}\in B[\mathbf{c},r]$, establishing the result. $\square$

Note that the above result is true for any norm defined on $\mathbb{R}^n$. The *unit-ball* is the ball $B[0,1]$. There are different unit-balls, depending on the norm that is being used. The l_1, l_2, and l_∞ balls are illustrated in Figure 6.2. As always, unless otherwise specified, we assume in this book that the underlying norm is the l_2-norm and that the balls are with respect to the l_2-norm.

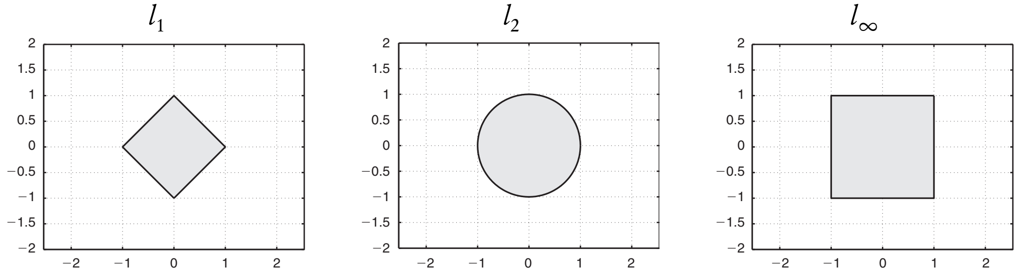

Figure 6.2. *l_1, l_2, and l_∞ balls in $\mathbb{R}^2$.*

Another important example of convex sets are ellipsoids.

Example 6.5 (convexity of ellipsoids). An *ellipsoid* is a set of the form

$$E = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{b}^T\mathbf{x} + c \leq 0\},$$

where $\mathbf{Q} \in \mathbb{R}^{n\times n}$ is positive semidefinite, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$. Denoting

$$f(\mathbf{x}) \equiv \mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{b}^T\mathbf{x} + c,$$

the set E can be rewritten as

$$E = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \leq 0\}.$$

To prove the convexity of E, we take $\mathbf{x}, \mathbf{y} \in E$ and $\lambda \in [0,1]$. Then $f(\mathbf{x}) \leq 0, f(\mathbf{y}) \leq 0$, and thus the vector $\mathbf{z} = \lambda\mathbf{x} + (1-\lambda)\mathbf{y}$ satisfies

$$\begin{aligned}\mathbf{z}^T\mathbf{Q}\mathbf{z} &= (\lambda\mathbf{x} + (1-\lambda)\mathbf{y})^T\mathbf{Q}(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \\ &= \lambda^2\mathbf{x}^T\mathbf{Q}\mathbf{x} + (1-\lambda)^2\mathbf{y}^T\mathbf{Q}\mathbf{y} + 2\lambda(1-\lambda)\mathbf{x}^T\mathbf{Q}\mathbf{y}.\end{aligned} \tag{6.2}$$

Now, note that $\mathbf{x}^T\mathbf{Q}\mathbf{y} = (\mathbf{Q}^{1/2}\mathbf{x})^T(\mathbf{Q}^{1/2}\mathbf{y})$, and hence by the Cauchy–Schwarz inequality, it follows that

$$\mathbf{x}^T\mathbf{Q}\mathbf{y} \leq \|\mathbf{Q}^{1/2}\mathbf{x}\| \cdot \|\mathbf{Q}^{1/2}\mathbf{y}\| = \sqrt{\mathbf{x}^T\mathbf{Q}\mathbf{x}}\sqrt{\mathbf{y}^T\mathbf{Q}\mathbf{y}} \leq \frac{1}{2}(\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{y}^T\mathbf{Q}\mathbf{y}), \tag{6.3}$$

where the last inequality follows from the fact that $\sqrt{ac} \leq \frac{1}{2}(a+c)$ for any two nonnegative scalars a, c. Plugging inequality (6.3) into (6.2), we obtain that

$$\mathbf{z}^T\mathbf{Q}\mathbf{z} \leq \lambda\mathbf{x}^T\mathbf{Q}\mathbf{x} + (1-\lambda)\mathbf{y}^T\mathbf{Q}\mathbf{y},$$

and hence

$$\begin{aligned} f(\mathbf{z}) &= \mathbf{z}^T\mathbf{Q}\mathbf{z} + 2\mathbf{b}^T\mathbf{z} + c \\ &\le \lambda\mathbf{x}^T\mathbf{Q}\mathbf{x} + (1-\lambda)\mathbf{y}^T\mathbf{Q}\mathbf{y} + 2\lambda\mathbf{b}^T\mathbf{x} + 2(1-\lambda)\mathbf{b}^T\mathbf{y} + c \\ &= \lambda(\mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{b}^T\mathbf{x} + c) + (1-\lambda)(\mathbf{y}^T\mathbf{Q}\mathbf{y} + 2\mathbf{b}^T\mathbf{y} + c) \\ &= \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) \le 0, \end{aligned}$$

establishing the desired result that $\mathbf{z} \in E$. ∎

6.2 ▪ Algebraic Operations with Convex Sets

An important property of convexity is that it is preserved under the intersection of sets.

Lemma 6.6. *Let $C_i \subseteq \mathbb{R}^n$ be a convex set for any $i \in I$, where I is an index set (possibly infinite). Then the set $\bigcap_{i\in I} C_i$ is convex.*

Proof. Suppose that $\mathbf{x},\mathbf{y} \in \bigcap_{i\in I} C_i$ and let $\lambda \in [0,1]$. Then $\mathbf{x},\mathbf{y} \in C_i$ for any $i \in I$, and since C_i is convex, it follows that $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in C_i$ for any $i \in I$. Therefore, $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in \bigcap_{i\in I} C_i$. □

Example 6.7 (convex polytopes). A direct consequence of the above result is that a set defined by a set of linear inequalities, specifically,

$$P = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} \le \mathbf{b}\},$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$ is convex. The convexity of P follows from the fact that it is an intersection of half-spaces:

$$P = \bigcap_{i=1}^{m} \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}_i\mathbf{x} \le b_i\},$$

where $\mathbf{A}_i$ is the ith row of $\mathbf{A}$. Since half-spaces are convex (Lemma 6.3), the convexity of P follows. Sets of the form P are called *convex polytopes*. ∎

Convexity is also preserved under addition, Cartesian product, linear mappings, and inverse linear mappings. This result is now stated, and its simple proof is left as an exercise (see Exercise 6.1).

Theorem 6.8 (preservation of convexity under addition, intersection and linear mappings).

(a) *Let $C_1, C_2, \ldots, C_k \subseteq \mathbb{R}^n$ be convex sets and let $\mu_1, \mu_2, \ldots, \mu_k \in \mathbb{R}$. Then the set*

$$\mu_1 C_1 + \mu_2 C_2 + \cdots + \mu_k C_k = \left\{ \sum_{i=1}^{k} \mu_i \mathbf{x}_i : \mathbf{x}_i \in C_i, i = 1,2,\ldots,k \right\}$$

is convex.

(b) *Let $C_i \subseteq \mathbb{R}^{k_i}$ be a convex set for any $i = 1,2,\ldots,m$. Then the Cartesian product*

$$C_1 \times C_2 \times \cdots \times C_m = \{(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m) : \mathbf{x}_i \in C_i, i = 1,2,\ldots,m\}$$

is convex.

(c) *Let $M \subseteq \mathbb{R}^n$ be a convex set and let $\mathbf{A} \in \mathbb{R}^{m\times n}$. Then the set*

$$\mathbf{A}(M) = \{\mathbf{Ax} : \mathbf{x} \in M\}$$

is convex.

(d) *Let $D \subseteq \mathbb{R}^m$ be a convex set, and let $\mathbf{A} \in \mathbb{R}^{m\times n}$. Then the set*

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

is convex.

A direct result of part (a) of Theorem 6.8 is that if $C \subseteq \mathbb{R}^n$ is a convex set and $\mathbf{b} \in \mathbb{R}^n$, then the set

$$C + \mathbf{b} = \{\mathbf{x} + \mathbf{b} : \mathbf{x} \in C\}$$

is also convex.

6.3 ▪ The Convex Hull

Definition 6.9 (convex combinations). *Given k vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in \mathbb{R}^n$, a* **convex combination** *of these k vectors is a vector of the form $\lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2 + \ldots + \lambda_k\mathbf{x}_k$, where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are nonnegative numbers satisfying $\lambda_1 + \lambda_2 + \cdots + \lambda_k = 1$.*

A convex set is defined by the property that any convex combination of two points from the set is also in the set. We will now show that a convex combination of *any* number of points from a convex set is in the set.

Theorem 6.10. *Let $C \subseteq \mathbb{R}^n$ be a convex set and let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m \in C$. Then for any $\lambda \in \Delta_m$, the relation $\sum_{i=1}^m \lambda_i \mathbf{x}_i \in C$ holds.*

Proof. We will prove the result by induction on m. For $m = 1$ the result is obvious (it essentially says that $\mathbf{x}_1 \in C$ implies that $\mathbf{x}_1 \in C$...). The induction hypothesis is that for any m vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m \in C$ and any $\lambda \in \Delta_m$, the vector $\sum_{i=1}^m \lambda_i \mathbf{x}_i$ belongs to C. We will now prove the theorem for $m+1$ vectors. Suppose that $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{m+1} \in C$ and that $\lambda \in \Delta_{m+1}$. We will show that $\mathbf{z} \equiv \sum_{i=1}^{m+1} \lambda_i \mathbf{x}_i \in C$. If $\lambda_{m+1} = 1$, then $\mathbf{z} = \mathbf{x}_{m+1} \in C$ and the result obviously follows. If $\lambda_{m+1} < 1$, then

$$\begin{aligned}
\mathbf{z} &= \sum_{i=1}^{m+1} \lambda_i \mathbf{x}_i \\
&= \sum_{i=1}^{m} \lambda_i \mathbf{x}_i + \lambda_{m+1}\mathbf{x}_{m+1} \\
&= (1-\lambda_{m+1}) \underbrace{\sum_{i=1}^{m} \frac{\lambda_i}{1-\lambda_{m+1}} \mathbf{x}_i}_{\mathbf{v}} + \lambda_{m+1}\mathbf{x}_{m+1}.
\end{aligned}$$

Since $\sum_{i=1}^m \frac{\lambda_i}{1-\lambda_{m+1}} = \frac{1-\lambda_{m+1}}{1-\lambda_{m+1}} = 1$, it follows that $\mathbf{v}$ (as defined in the above equation) is a convex combination of m points from C, and hence by the induction hypothesis we have that $\mathbf{v} \in C$. Thus, by the definition of a convex set, $\mathbf{z} = (1-\lambda_{m+1})\mathbf{v} + \lambda_{m+1}\mathbf{x}_{m+1} \in C$. $\square$

Definition 6.11 (convex hulls). *Let $S \subseteq \mathbb{R}^n$. Then the* **convex hull** *of S, denoted by* $\text{conv}(S)$, *is the set comprising all the convex combinations of vectors from S:*

$$\text{conv}(S) \equiv \left\{ \sum_{i=1}^{k} \lambda_i \mathbf{x}_i : \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in S, \lambda \in \Delta_k, k \in \mathbb{N} \right\}.$$

Note that in the definition of the convex hull, the number of vectors k in the convex combination representation can be any positive integer. The convex hull $\text{conv}(S)$ is the "smallest" convex set containing S meaning that if another convex set T contains S, then $\text{conv}(S) \subseteq T$. This property is stated and proved in the following lemma.

Lemma 6.12. *Let $S \subseteq \mathbb{R}^n$. If $S \subseteq T$ for some convex set T, then* $\text{conv}(S) \subseteq T$.

Proof. Suppose that indeed $S \subseteq T$ for some convex set T. To prove that $\text{conv}(S) \subseteq T$, take $\mathbf{z} \in \text{conv}(S)$. Then by the definition of the convex hull, there exist $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in S \subseteq T$ (where k is a positive integer) and $\lambda \in \Delta_k$ such that $\mathbf{z} = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i$. By Theorem 6.10 and the convexity of T, it follows that any convex combination of elements from T is in T, and therefore, since $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in T$, it follows that $\mathbf{z} \in T$, showing the desired result. $\square$

An example of a convex hull of a nonconvex polytope is given in Figure 6.3.

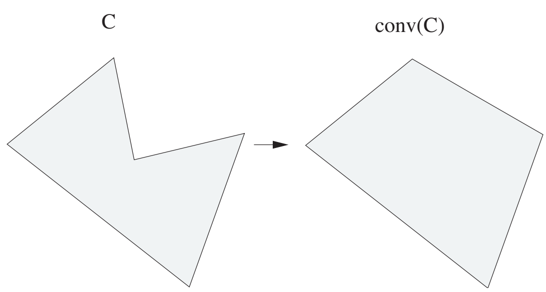

Figure 6.3. *A nonconvex set and its convex hull.*

The following well-known result, called the Carathéodory theorem, states that any element in the convex hull of a subset of a given set $S \subseteq \mathbb{R}^n$ can be represented as a convex combination of no more than $n+1$ vectors from S.

Theorem 6.13 (Carathéodory theorem). *Let $S \subseteq \mathbb{R}^n$ and let $\mathbf{x} \in \text{conv}(S)$. Then there exist $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n+1} \in S$ such that $\mathbf{x} \in \text{conv}(\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n+1}\})$; that is, there exist $\lambda \in \Delta_{n+1}$ such that*

$$\mathbf{x} = \sum_{i=1}^{n+1} \lambda_i \mathbf{x}_i.$$

Proof. Let $\mathbf{x} \in \text{conv}(S)$. By the definition of the convex hull, there exist $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \in S$ and $\lambda \in \Delta_k$ such that

$$\mathbf{x} = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i.$$

We can assume that $\lambda_i > 0$ for all $i = 1, 2, \dots, k$, since otherwise the vectors corresponding to the zero coefficients can be omitted. If $k \leq n+1$, the result is proven. Otherwise, if $k \geq n+2$, then the vectors $\mathbf{x}_2 - \mathbf{x}_1, \mathbf{x}_3 - \mathbf{x}_1, \dots, \mathbf{x}_k - \mathbf{x}_1$, being more than n vectors in $\mathbb{R}^n$, are necessarily linearly dependent, which means that there exist $\mu_2, \mu_3, \dots, \mu_k$ which are not all zeros such that

$$\sum_{i=2}^{k} \mu_i (\mathbf{x}_i - \mathbf{x}_1) = \mathbf{0}.$$

Defining $\mu_1 = -\sum_{i=2}^{k} \mu_i$, we obtain that

$$\sum_{i=1}^{k} \mu_i \mathbf{x}_i = \mathbf{0},$$

where not all of the coefficients $\mu_1, \mu_2, \dots, \mu_k$ are zeros and in addition they satisfy $\sum_{i=1}^{k} \mu_i = 0$. In particular, there exists an index i for which $\mu_i < 0$. Let $\alpha \in \mathbb{R}_+$. Then

$$\mathbf{x} = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i + \alpha \sum_{i=1}^{k} \mu_i \mathbf{x}_i = \sum_{i=1}^{k} (\lambda_i + \alpha \mu_i) \mathbf{x}_i. \tag{6.4}$$

We have $\sum_{i=1}^{k} (\lambda_i + \alpha \mu_i) = 1$, so the representation (6.4) is a convex combination representation if and only if

$$\lambda_i + \alpha \mu_i \geq 0 \text{ for all } i = 1, \dots, k. \tag{6.5}$$

Since $\lambda_i > 0$ for all i, it follows that the set of inequalities (6.5) is satisfied for all $\alpha \in [0, \varepsilon]$ where $\varepsilon = \min_{i:\mu_i<0} \{-\frac{\lambda_i}{\mu_i}\}$. The scalar ε is well-defined since, as was already mentioned, there exists an index for which $\mu_i < 0$. If we substitute $\alpha = \varepsilon$, then (6.5) still holds, but $\lambda_j + \varepsilon \mu_j = 0$ for $j \in \text{argmin}_{i:\mu_i<0} \{-\frac{\mu_i}{\lambda_i}\}$. This means that we have found a representation of $\mathbf{x}$ as a convex combination of $k-1$ vectors. This process can be carried on until a representation of $\mathbf{x}$ as a convex combination of no more than $n+1$ vectors is derived. ☐

Example 6.14. For $n = 2$, consider the following four vectors:

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \mathbf{x}_4 = \begin{pmatrix} 2 \\ 2 \end{pmatrix},$$

and let $\mathbf{x} \in \text{conv}(\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\})$ be given by

$$\mathbf{x} = \frac{1}{8}\mathbf{x}_1 + \frac{1}{4}\mathbf{x}_2 + \frac{1}{2}\mathbf{x}_3 + \frac{1}{8}\mathbf{x}_4 = \begin{pmatrix} \frac{13}{8} \\ \frac{11}{8} \end{pmatrix}.$$

By the Carathéodory theorem, $\mathbf{x}$ can be expressed as a convex combination of three of the four vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$. To find such a convex combination, let us employ the process described in the proof of the theorem. The vectors

$$\mathbf{x}_2 - \mathbf{x}_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \mathbf{x}_3 - \mathbf{x}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{x}_4 - \mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

are linearly dependent, and the linear dependence is given by the equation

$$(\mathbf{x}_2-\mathbf{x}_1)+(\mathbf{x}_3-\mathbf{x}_1)-(\mathbf{x}_4-\mathbf{x}_1)=0.$$

We thus have the linear dependence relation

$$-\mathbf{x}_1+\mathbf{x}_2+\mathbf{x}_3-\mathbf{x}_4=0.$$

Therefore, we can write the following for any $\alpha \geq 0$:

$$\mathbf{x}=\left(\frac{1}{8}-\alpha\right)\mathbf{x}_1+\left(\frac{1}{4}+\alpha\right)\mathbf{x}_2+\left(\frac{1}{2}+\alpha\right)\mathbf{x}_3+\left(\frac{1}{8}-\alpha\right)\mathbf{x}_4.$$

The weights in the above representation add up to one, so we need only guarantee that they are nonnegative, meaning that

$$\frac{1}{8}-\alpha\geq 0,\quad \frac{1}{4}+\alpha\geq 0,\quad \frac{1}{2}+\alpha\geq 0,\quad \frac{1}{8}-\alpha\geq 0,$$

which combined with $\alpha \geq 0$ yields that $0 \leq \alpha \leq \frac{1}{8}$. Substituting $\alpha = \frac{1}{8}$, we obtain the convex combination

$$\mathbf{x}=\frac{3}{8}\mathbf{x}_2+\frac{5}{8}\mathbf{x}_3.$$

Note that in this example two coefficients were turned into zero, so we obtained a representation with only two vectors, while the Carathéodory theorem can guarantee only a presentation by at most three vectors. ∎

6.4 ▪ Convex Cones

A set S is called a *cone* if it satisfies the following property: for any $\mathbf{x}\in S$ and $\lambda \geq 0$, the inclusion $\lambda\mathbf{x}\in S$ is satisfied. The following lemma shows that there is a very simple and elegant characterization of convex cones.

Lemma 6.15. *A set S is a convex cone if and only if the following properties hold:*

A. $\mathbf{x},\mathbf{y}\in S \Rightarrow \mathbf{x}+\mathbf{y}\in S$.

B. $\mathbf{x}\in S, \lambda\geq 0 \Rightarrow \lambda\mathbf{x}\in S$.

Proof. (*convex cone* $\Rightarrow$ A,B). Suppose that S is a convex cone. Then property B follows from the definition of a cone. To prove property A, assume that $\mathbf{x},\mathbf{y}\in S$. Then by the convexity of S we have that $\frac{1}{2}(\mathbf{x}+\mathbf{y})\in S$, and hence, since S is a cone, it follows that $\mathbf{x}+\mathbf{y}=2\cdot\frac{1}{2}(\mathbf{x}+\mathbf{y})\in S$.

(A,B $\Rightarrow$ *convex cone*). Now assume that S satisfies properties A and B. Then S is a cone by property B. To prove the convexity, let $\mathbf{x},\mathbf{y}\in S$ and $\lambda\in[0,1]$. Then $\lambda\mathbf{x},(1-\lambda)\mathbf{y}\in S$ by property B, and thus by property A, $\lambda\mathbf{x}+(1-\lambda)\mathbf{y}\in S$, establishing the convexity of S. □

The following are some well-known examples of convex cones.

Example 6.16. Consider the convex polytope

$$C = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} \le \mathbf{0}\},$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$. The set C is clearly a convex set since it is a convex polytope (see Example 6.7). It is also a cone since

$$\mathbf{x} \in C, \lambda \ge 0 \Rightarrow \mathbf{A}\mathbf{x} \le \mathbf{0}, \lambda \ge 0 \Rightarrow \mathbf{A}(\lambda\mathbf{x}) \le \mathbf{0} \Rightarrow \lambda\mathbf{x} \in C.$$

Taking, for example, $m = n$ and $\mathbf{A} = -\mathbf{I}$, the set C reduces to the nonnegative orthant $\mathbb{R}^n_+$. ∎

Example 6.17 (Lorentz cone). The *Lorentz cone*, or *ice cream cone* whose boundary is described in Figure 6.4, is given by

$$L^n = \left\{ \begin{pmatrix} \mathbf{x} \\ t \end{pmatrix} \in \mathbb{R}^{n+1} : \|\mathbf{x}\| \le t, \mathbf{x} \in \mathbb{R}^n, t \in \mathbb{R} \right\}.$$

The Lorentz cone is in fact a convex cone. To show this, let us take $\binom{\mathbf{x}}{t}, \binom{\mathbf{y}}{s} \in L^n$. Then $\|\mathbf{x}\| \le t, \|\mathbf{y}\| \le s$, which combined with the triangle inequality implies that

$$\|\mathbf{x}+\mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\| \le t + s,$$

showing that $\binom{\mathbf{x}}{t} + \binom{\mathbf{y}}{s} \in L^n$, and hence that property A holds. To show property B, take $\binom{\mathbf{x}}{t} \in L^n$ and $\lambda \ge 0$, then since $\|\mathbf{x}\| \le t$, it readily follows that $\|\lambda\mathbf{x}\| \le \lambda t$, so that $\lambda\binom{\mathbf{x}}{t} \in L^n$. ∎

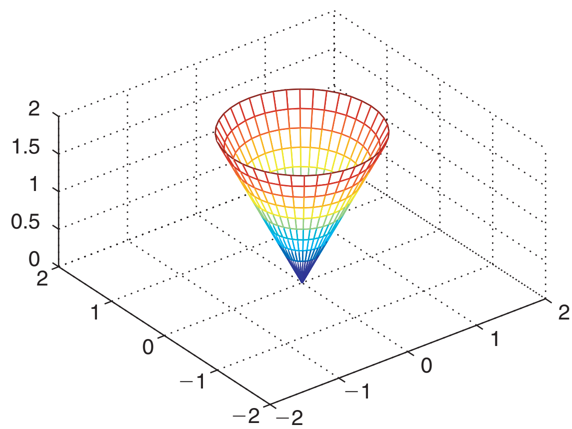

Figure 6.4. *The boundary of the ice cream cone L^2.*

Example 6.18 (nonnegative polynomials). Consider the set of all coefficients of polynomials of degree of at most $n-1$ which are nonnegative over $\mathbb{R}$:

$$K^n = \{\mathbf{x} \in \mathbb{R}^n : x_1 t^{n-1} + x_2 t^{n-2} + \cdots + x_{n-1}t + x_n \ge 0 \text{ for all } t \in \mathbb{R}\}.$$

It is easy to verify that this is a convex cone. Let us consider two special cases. When $n = 2$, then clearly

$$K^2 = \{(x_1, x_2)^T : x_1 t + x_2 \ge 0 \text{ for all } t \in \mathbb{R}\} = \{(x_1, x_2) : x_1 = 0, x_2 \ge 0\},$$

so K^2 is the nonnegative part of the x_2-axis. For $n=3$ we have

$$K^3 = \{(x_1,x_2,x_3)^T : x_1 t^2 + x_2 t + x_3 \geq 0\}.$$

A quadratic polynomial $\varphi(t) = at^2 + bt + c$ is nonnegative over $\mathbb{R}$ if and only if $a,c \geq 0$ and the discriminant $\Delta = b^2 - 4ac$ is nonpositive. We thus conclude that

$$K^3 = \{(x_1,x_2,x_3)^T : x_1,x_3 \geq 0, x_2^2 \leq 4x_1x_3\}. \quad \blacksquare$$

Similarly to the notion of a convex combination, we will now define the concept of a *conic combination*.

Definition 6.19 (conic combination). *Given k points* $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_k \in \mathbb{R}^n$, *a* **conic combination** *of these k points is a vector of the form* $\lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2 + \cdots + \cdots + \lambda_k\mathbf{x}_k$, *where* $\lambda \in \mathbb{R}^k_+$.

It is easy to show that any conic combination of points in a convex cone C belong to C. The proof of this elementary result is left as an exercise (Exercise 6.14).

Lemma 6.20. *Let C be a convex cone, and let* $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_k \in C$ *and* $\lambda_1,\lambda_2,\ldots,\lambda_k \geq 0$. *Then* $\sum_{i=1}^k \lambda_i\mathbf{x}_i \in C$.

The definition of the *conic hull* is now quite natural.

Definition 6.21 (conic hulls). *Let* $S \subseteq \mathbb{R}^n$. *Then the* **conic hull** *of S, denoted by* $\text{cone}(S)$, *is the set comprising all the conic combinations of vectors from S:*

$$\text{cone}(S) \equiv \left\{ \sum_{i=1}^k \lambda_i\mathbf{x}_i : \mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_k \in S, \lambda \in \mathbb{R}^k_+, k \in \mathbb{N} \right\}.$$

Similar to the convex hull, the conic hull of a set S is the smallest convex cone containing S. The proof of this result is left as an exercise (Exercise 6.15).

Lemma 6.22. *Let* $S \subseteq \mathbb{R}^n$. *If* $S \subseteq T$ *for some convex cone* T, *then* $\text{cone}(S) \subseteq T$.

A natural question that arises is whether we can establish a result similar to the Carathéodory theorem on the representation of vectors in the conic hull of a set. Interestingly, we can establish an even stronger result for conic hulls: each vector in the conic hull of a set $S \subseteq \mathbb{R}^n$ can be represented as a convex combination of at most n vectors from S (recall that in Carathéodory's theorem $n+1$ vectors are required).

Theorem 6.23 (conic representation theorem). *Let* $S \subseteq \mathbb{R}^n$ *and let* $\mathbf{x} \in \text{cone}(S)$. *Then there exist k linearly independent vectors* $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_k \in S$ *such that* $\mathbf{x} \in \text{cone}(\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_k\})$; *that is, there exists* $\lambda \in \mathbb{R}^k_+$ *such that*

$$\mathbf{x} = \sum_{i=1}^k \lambda_i\mathbf{x}_i.$$

In addition, $k \leq n$.

Proof. The proof is similar to the proof of the Carathéodory theorem. Let $\mathbf{x} \in \text{cone}(S)$. By the definition of the conic hull, there exist $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in S$ and $\lambda \in \mathbb{R}^k_+$ such that

$$\mathbf{x} = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i.$$

We can assume that $\lambda_i > 0$ for all $i = 1, 2, \ldots, k$, since otherwise the vectors corresponding to the zero λ_i's can be omitted. If the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ are linearly independent, then the result is proven. Otherwise, if the vectors are linearly dependent, then there exist $\mu_1, \mu_2, \ldots, \mu_k \in \mathbb{R}$ which are not all zeros such that

$$\sum_{i=1}^{k} \mu_i \mathbf{x}_i = \mathbf{0}.$$

Then for any $\alpha \in \mathbb{R}$

$$\mathbf{x} = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i + \alpha \sum_{i=1}^{k} \mu_i \mathbf{x}_i = \sum_{i=1}^{k} (\lambda_i + \alpha \mu_i) \mathbf{x}_i. \tag{6.6}$$

The representation (6.6) is a conic representation if and only if

$$\lambda_i + \alpha \mu_i \geq 0 \text{ for all } i = 1, \ldots, k. \tag{6.7}$$

Since $\lambda_i > 0$ for all i, it follows that the set of inequalities (6.7) is satisfied for all $\alpha \in I$ where I is a closed interval with a nonempty interior. Note that one (but not both) of the endpoints of I might be infinite. If we substitute one of the finite endpoints of I, call it $\tilde{\alpha}$, into α, then we still get that (6.7) holds, but in addition $\lambda_j + \tilde{\alpha} \mu_j = 0$ for some index j. Thus we obtain a representation of $\mathbf{x}$ as a conic combination of at most $k-1$ vectors. This process can be carried on until we obtain a representation of $\mathbf{x}$ as a conic combination of linearly independent vectors. Since the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ are linearly independent vectors in $\mathbb{R}^n$ it follows that $k \leq n$. $\square$

The latter representation theorem has an important application to convex polytopes of the form

$$P = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\},$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. We will assume without loss of generality that the rows of $\mathbf{A}$ are linearly independent. Linear systems consisting of linear equalities and nonnegativity constraints often appear as constraints in standard formulations of linear programming problems. An important property of nonempty convex polytopes of the form P is that they contain at least one vector with at most m nonzero elements (m being the number of constraints). An important notion in this respect is the one of a *basic feasible solution*.

Definition 6.24 (basic feasible solutions). *Let $P = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Suppose that the rows of $\mathbf{A}$ are linearly independent. Then $\bar{\mathbf{x}}$ is a* **basic feasible solution** *(abbreviated bfs) of P if the columns of $\mathbf{A}$ corresponding to the indices of the positive values of $\bar{\mathbf{x}}$ are linearly independent.*

Obviously, since the columns of $\mathbf{A}$ reside in $\mathbb{R}^m$, it follows that a bfs has at most m nonzero elements.

Example 6.25. Consider the linear system

$$\begin{aligned} x_1 + x_2 + x_3 &= 6, \\ x_2 + x_3 &= 3, \\ x_1, x_2, x_3 &\geq 0. \end{aligned}$$

An example of a bfs of the system is $(x_1, x_2, x_3) = (3, 3, 0)$. This vector is indeed a bfs since it satisfies all the constraints and the columns corresponding to the positive elements, meaning columns 1 and 2

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

are linearly independent. ∎

The existence of a bfs in P, provided that it is nonempty, follows directly from Theorem 6.23.

Theorem 6.26. *Let* $P = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0\}$, *where* $\mathbf{A} \in \mathbb{R}^{m \times n}$ *and* $\mathbf{b} \in \mathbb{R}^m$. *If* $P \neq \emptyset$, *then it contains at least one bfs.*

Proof. Since $P \neq \emptyset$, it follows that $\mathbf{b} \in \text{cone}(\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\})$, where $\mathbf{a}_i$ denotes the ith column of $\mathbf{A}$. By the conic representation theorem (Theorem 6.23), we have that $\mathbf{b}$ can be represented as a conic combination of k linearly independent vectors from $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$; that is, there exist indices $i_1 < i_2 < \cdots < i_k$ and k numbers $y_{i_1}, y_{i_2}, \ldots, y_{i_k} \geq 0$ such that $\mathbf{b} = \sum_{j=1}^k y_{i_j} \mathbf{a}_{i_j}$ and $\mathbf{a}_{i_1}, \mathbf{a}_{i_2}, \ldots, \mathbf{a}_{i_k}$ are linearly independent. Denote $\bar{\mathbf{x}} = \sum_{j=1}^k y_{i_j} \mathbf{e}_{i_j}$. Then obviously $\bar{\mathbf{x}} \geq 0$ and in addition

$$\mathbf{A}\bar{\mathbf{x}} = \sum_{j=1}^k y_{i_j} \mathbf{A}\mathbf{e}_{i_j} = \sum_{j=1}^k y_{i_j} \mathbf{a}_{i_j} = \mathbf{b}.$$

Therefore, $\bar{\mathbf{x}}$ is contained in P and satisfies that the columns of $\mathbf{A}$ corresponding to the indices of the positive components of $\bar{\mathbf{x}}$ are linearly independent, meaning that P contains a bfs. □

6.5 ▪ Topological Properties of Convex Sets

We begin by proving that the closure of a convex set is a convex set.

Theorem 6.27 (convexity preservation under closure). *Let* $C \subseteq \mathbb{R}^n$ *be a convex set. Then* $\text{cl}(C)$ *is a convex set.*

Proof. Let $\mathbf{x}, \mathbf{y} \in \text{cl}(C)$ and let $\lambda \in [0, 1]$. Then by the definition of the closure set, it follows that there exist sequences $\{\mathbf{x}_k\}_{k \geq 0} \subseteq C$ and $\{\mathbf{y}_k\}_{k \geq 0} \subseteq C$ for which $\mathbf{x}_k \to \mathbf{x}$ and $\mathbf{y}_k \to \mathbf{y}$ as $k \to \infty$. By the convexity of C, it follows that $\lambda \mathbf{x}_k + (1 - \lambda)\mathbf{y}_k \in C$ for any $k \geq 0$. Since $\lambda \mathbf{x}_k + (1 - \lambda)\mathbf{y}_k \to \lambda \mathbf{x} + (1 - \lambda)\mathbf{y}$, it follows that there exists a sequence in C that converges to $\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}$, implying that $\lambda \mathbf{x} + (1 - \lambda)\mathbf{y} \in \text{cl}(C)$. □

Proving that the interior of a convex set is also a convex set is more tricky, and we will require the following technical and quite useful result called the *line segment principle*.

Lemma 6.28 (line segment principle). *Let C be a convex set, and assume that* $\text{int}(C) \neq \emptyset$. *Suppose that* $\mathbf{x} \in \text{int}(C)$ *and* $\mathbf{y} \in \text{cl}(C)$. *Then* $(1-\lambda)\mathbf{x} + \lambda\mathbf{y} \in \text{int}(C)$ *for any* $\lambda \in (0,1)$.

Proof. Since $\mathbf{x} \in \text{int}(C)$, there exists $\varepsilon > 0$ such that $B(\mathbf{x}, \varepsilon) \subseteq C$. Let $\mathbf{z} = (1-\lambda)\mathbf{x} + \lambda\mathbf{y}$. To prove that $\mathbf{z} \in \text{int}(C)$, we will show that in fact $B(\mathbf{z}, (1-\lambda)\varepsilon) \subseteq C$. Let then $\mathbf{w}$ be a vector satisfying $\|\mathbf{w} - \mathbf{z}\| < (1-\lambda)\varepsilon$. Since $\mathbf{y} \in \text{cl}(C)$, it follows that there exists $\mathbf{w}_1 \in C$ such that

$$\|\mathbf{w}_1 - \mathbf{y}\| < \frac{(1-\lambda)\varepsilon - \|\mathbf{w} - \mathbf{z}\|}{\lambda}. \tag{6.8}$$

Set $\mathbf{w}_2 = \frac{1}{1-\lambda}(\mathbf{w} - \lambda\mathbf{w}_1)$. Then

$$\begin{aligned}
\|\mathbf{w}_2 - \mathbf{x}\| &= \left\| \frac{\mathbf{w} - \lambda\mathbf{w}_1}{1-\lambda} - \mathbf{x} \right\| \\
&= \frac{1}{1-\lambda} \|(\mathbf{w} - \mathbf{z}) + \lambda(\mathbf{y} - \mathbf{w}_1)\| \\
&\leq \frac{1}{1-\lambda} (\|\mathbf{w} - \mathbf{z}\| + \lambda\|\mathbf{w}_1 - \mathbf{y}\|) \\
&\overset{(6.8)}{<} \varepsilon,
\end{aligned}$$

and hence, since $B(\mathbf{x}, \varepsilon) \subseteq C$, it follows that $\mathbf{w}_2 \in C$. Finally, since $\mathbf{w} = \lambda\mathbf{w}_1 + (1-\lambda)\mathbf{w}_2$ with $\mathbf{w}_1, \mathbf{w}_2 \in C$, we have that $\mathbf{w} \in C$, and the line segment principle is thus proved. $\square$

The immediate consequence of the line segment principle is the convexity of interiors of convex sets.

Theorem 6.29 (convexity of interiors of convex sets). *Let* $C \subseteq \mathbb{R}^n$ *be a convex set. Then* $\text{int}(C)$ *is convex.*

Proof. If $\text{int}(C) = \emptyset$, then the theorem is obviously true. Otherwise, let $\mathbf{x}_1, \mathbf{x}_2 \in \text{int}(C)$, and let $\lambda \in (0,1)$. Then by the line segment principle we have that $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 \in \text{int}(C)$, establishing the convexity of $\text{int}(C)$. $\square$

Other topological properties that are implied by the line segment property are given in the next result.

Lemma 6.30. *Let C be a convex set with a nonempty interior. Then*

(a) $\text{cl}(\text{int}(C)) = \text{cl}(C)$,

(b) $\text{int}(\text{cl}(C)) = \text{int}(C)$.

Proof. (a) Obviously, since $\text{int}(C) \subseteq C$, the inclusion $\text{cl}(\text{int}(C)) \subseteq \text{cl}(C)$ holds. To prove the opposite, let $\mathbf{x} \in \text{cl}(C)$ and let $\mathbf{y} \in \text{int}(C)$. Then by the line segment principle, $\mathbf{x}_k = \frac{1}{k}\mathbf{y} + (1-\frac{1}{k})\mathbf{x} \in \text{int}(C)$ for any $k \geq 1$. Since $\mathbf{x}$ is the limit of the sequence $\{\mathbf{x}_k\}_{k\geq 1} \subseteq \text{int}(C)$, it follows that $\mathbf{x} \in \text{cl}(\text{int}(C))$.

(b) The inclusion $\text{int}(C) \subseteq \text{int}(\text{cl}(C))$ follows immediately from the inclusion $C \subseteq \text{cl}(C)$. To show the reverse inclusion, we take $\mathbf{x} \in \text{int}(\text{cl}(C))$ and show that $\mathbf{x} \in \text{int}(C)$. Since $\mathbf{x} \in \text{int}(\text{cl}(C))$, there exists $\varepsilon > 0$ such that $B(\mathbf{x}, \varepsilon) \subseteq \text{cl}(C)$. Let $\mathbf{y} \in \text{int}(C)$. If $\mathbf{y} = \mathbf{x}$, then the result is proved. Otherwise, define

$$\mathbf{z} = \mathbf{x} + \alpha(\mathbf{x} - \mathbf{y}),$$

where $\alpha = \frac{\varepsilon}{2\|\mathbf{x}-\mathbf{y}\|}$. Since $\|\mathbf{z}-\mathbf{x}\| = \frac{\varepsilon}{2}$, it follows that $\mathbf{z} \in \mathrm{cl}(C)$. Therefore, $(1-\lambda)\mathbf{y}+\lambda\mathbf{z} \in \mathrm{int}(C)$ for any $\lambda \in [0,1)$ and specifically for $\lambda = \lambda_\alpha = \frac{1}{1+\alpha}$. Noting that $(1-\lambda_\alpha)\mathbf{y}+\lambda_\alpha\mathbf{z} = \mathbf{x}$, we conclude that $\mathbf{x} \in \mathrm{int}(C)$. $\square$

In general, the convex hull of a closed set is not necessarily a closed set. A classical example for this fact is the closed set

$$S = \{(0,0)^T\} \cup \{(x,y)^T : xy \geq 1, x \geq 0, y \geq 0\},$$

whose convex hull is given by the set

$$\mathrm{conv}(S) = \{(0,0)^T\} \cup \mathbb{R}^2_{++},$$

which is neither closed nor open. However, closedness is preserved under the convex hull operation if the set is compact. As will be shown in the proof of the following result, this nontrivial fact follows from the Carathéodory theorem.

Proposition 6.31 (closedness of convex hulls of compact sets). *Let $S \subseteq \mathbb{R}^n$ be a compact set. Then* $\mathrm{conv}(S)$ *is compact.*

Proof. To prove the boundedness of $\mathrm{conv}(S)$, note that since S is compact, there exists $M > 0$ such that $\|\mathbf{x}\| \leq M$ for any $\mathbf{x} \in S$. Now, let $\mathbf{y} \in \mathrm{conv}(S)$. Then by the Carathéodory theorem it follows that there exist $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n+1} \in S$ and $\lambda \in \Delta_{n+1}$ for which $\mathbf{y} = \sum_{i=1}^{n+1} \lambda_i \mathbf{x}_i$, and therefore

$$\|\mathbf{y}\| = \left\| \sum_{i=1}^{n+1} \lambda_i \mathbf{x}_i \right\| \leq \sum_{i=1}^{n+1} \lambda_i \|\mathbf{x}_i\| \leq M \sum_{i=1}^{n+1} \lambda_i = M,$$

establishing the boundedness of $\mathrm{conv}(S)$. To prove the closedness of $\mathrm{conv}(S)$, let $\{\mathbf{y}_k\}_{k\geq 1} \subseteq \mathrm{conv}(S)$ be a sequence of vectors from $\mathrm{conv}(S)$ converging to $\mathbf{y} \in \mathbb{R}^n$. Our objective is to show that $\mathbf{y} \in \mathrm{conv}(S)$. By the Carathéodory theorem we have that for any $k \geq 1$ there exist vectors $\mathbf{x}_1^k, \mathbf{x}_2^k, \ldots, \mathbf{x}_{n+1}^k \in S$ and $\lambda^k \in \Delta_{n+1}$ such that

$$\mathbf{y}_k = \sum_{i=1}^{n+1} \lambda_i^k \mathbf{x}_i^k. \tag{6.9}$$

By the compactness of S and Δ_{n+1}, it follows that $\{(\lambda^k, \mathbf{x}_1^k, \mathbf{x}_2^k, \ldots, \mathbf{x}_{n+1}^k)\}_{k\geq 1}$ has a convergent subsequence $\{(\lambda^{k_j}, \mathbf{x}_1^{k_j}, \mathbf{x}_2^{k_j}, \ldots, \mathbf{x}_{n+1}^{k_j})\}_{j\geq 1}$ whose limit will be denoted by

$$(\lambda, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n+1})$$

with $\lambda \in \Delta_{n+1}, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n+1} \in S$. Taking the limit $j \to \infty$ in

$$\mathbf{y}_{k_j} = \sum_{i=1}^{n+1} \lambda_i^{k_j} \mathbf{x}_i^{k_j},$$

we obtain that

$$\mathbf{y} = \sum_{i=1}^{n+1} \lambda_i \mathbf{x}_i,$$

meaning that $\mathbf{y} \in \text{conv}(S)$ as required. $\square$

Another topological result which might seem simple but actually requires the conic representation theorem is the closedness of the conic hull of a finite set of points.

Lemma 6.32 (closedness of the conic hull of a finite set). *Let* $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k \in \mathbb{R}^n$. *Then* $\text{cone}(\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k\})$ *is closed.*

Proof. By the conic representation theorem, each element of $\text{cone}(\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k\})$ can be represented as a conic combination of a linearly independent subset of $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k\}$. Therefore, if $S_1, S_2, \ldots, S_N$ are all the subsets of $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k\}$ comprising linearly independent vectors, then

$$\text{cone}(\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k\}) = \bigcup_{i=1}^{N} \text{cone}(S_i).$$

It is enough to show that $\text{cone}(S_i)$ is closed for any $i \in \{1, 2, \ldots, N\}$. Indeed, let $i \in \{1, 2, \ldots, N\}$. Then

$$S_i = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_m\}$$

for some linearly independent vectors $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_m$. We can write $\text{cone}(S_i)$ as

$$\text{cone}(S_i) = \{\mathbf{B}\mathbf{y} : \mathbf{y} \in \mathbb{R}^m_+\},$$

where $\mathbf{B}$ is the matrix whose columns are $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_m$. Suppose that $\mathbf{x}_k \in \text{cone}(S_i)$ for all $k \geq 1$ and that $\mathbf{x}_k \to \bar{\mathbf{x}}$. We need to show that $\bar{\mathbf{x}} \in \text{cone}(S_i)$. Since $\mathbf{x}_k \in \text{cone}(S_i)$, it follows that there exists $\mathbf{y}_k \in \mathbb{R}^m_+$ such that

$$\mathbf{x}_k = \mathbf{B}\mathbf{y}_k. \tag{6.10}$$

Therefore, using the fact that the columns of $\mathbf{B}$ are linearly independent, we can deduce that

$$\mathbf{y}_k = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{x}_k.$$

Taking the limit as $k \to \infty$ in the last equation, we obtain that $\mathbf{y}_k \to \bar{\mathbf{y}}$ where $\bar{\mathbf{y}} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\bar{\mathbf{x}}$, and since $\mathbf{y}_k \in \mathbb{R}^m_+$ for all k, we also have $\bar{\mathbf{y}} \in \mathbb{R}^m_+$. Thus, taking the limit in (6.10), we conclude that $\bar{\mathbf{x}} = \mathbf{B}\bar{\mathbf{y}}$ with $\bar{\mathbf{y}} \in \mathbb{R}^m_+$, and hence $\bar{\mathbf{x}} \in \text{cone}(S_i)$. $\square$

6.6 ▪ Extreme Points

Definition 6.33 (extreme points). *Let* $S \subseteq \mathbb{R}^n$ *be a convex set. A point* $\mathbf{x} \in S$ *is called an* **extreme point** *of* S *if there do not exist* $\mathbf{x}_1, \mathbf{x}_2 \in S (\mathbf{x}_1 \neq \mathbf{x}_2)$ *and* $\lambda \in (0, 1)$, *such that* $\mathbf{x} = \lambda\mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2$.

That is, an extreme point is a point in the set that cannot be represented as a nontrivial convex combination of two different points in S. The set of extreme points is denoted by $\text{ext}(S)$. The set of extreme points of a convex polytope consists of all its vertices; see, for example, Figure 6.5.

We can fully characterize the extreme points of convex polytopes of the form $P = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$ has linearly independent rows and $\mathbf{b} \in \mathbb{R}^m$.

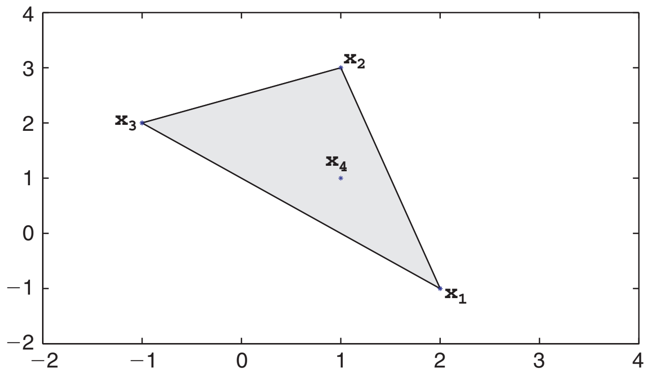

Figure 6.5. *The filled area is the convex set* $S = \mathrm{conv}\{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4\}$, *where* $\mathbf{x}_1 = (2,-1)^T, \mathbf{x}_2 = (1,3)^T, \mathbf{x}_3 = (-1,2)^T, \mathbf{x}_4 = (1,1)^T$. *The extreme points set is* $\mathrm{ext}(S) = \{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3\}$.

Recall (see Section 6.4) that $\bar{\mathbf{x}}$ is called a basic feasible solution of P if the columns of $\mathbf{A}$ corresponding to the indices of the positive values of $\bar{\mathbf{x}}$ are linearly independent. In Section 6.4 it was shown that if P is not empty, then it has at least one basic feasible solution. Interestingly, the extreme points of P are *exactly* the basic feasible solutions of P, which means that the linear independence of the columns of $\mathbf{A}$ corresponding to the positive variables is an algebraic characterization of extreme points.

Theorem 6.34 (equivalence between extreme points and basic feasible solutions). *Let* $P = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$, *where* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *has linearly independent rows and* $\mathbf{b} \in \mathbb{R}^m$. *Then* $\bar{\mathbf{x}}$ *is a basic feasible solution of* P *if and only if it is an extreme point of* P.

Proof. Suppose that $\bar{\mathbf{x}}$ is a basic feasible solution and assume without loss of generality that its first k components are positive while the others are zeros: $\bar{x}_1 > 0, \bar{x}_2 > 0, \ldots, \bar{x}_k > 0, \bar{x}_{k+1} = \bar{x}_{k+2} = \cdots = \bar{x}_n = 0$. Since $\bar{\mathbf{x}}$ is a basic feasible solution, the first k columns of $\mathbf{A}$ denoted by $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ are linearly independent. Suppose in contradiction that $\bar{\mathbf{x}} \notin \mathrm{ext}(P)$. Then there exist two different vectors $\mathbf{y}, \mathbf{z} \in P$ and $\lambda \in (0,1)$ such that $\bar{\mathbf{x}} = \lambda\mathbf{y} + (1-\lambda)\mathbf{z}$. Combining this with the fact that $\mathbf{y}, \mathbf{z} \geq \mathbf{0}$, we can conclude that the last $n-k$ variables in $\mathbf{y}$ and $\mathbf{z}$ are zeros. We can therefore write

$$\sum_{i=1}^{k} y_i \mathbf{a}_i = \mathbf{b},$$
$$\sum_{i=1}^{k} z_i \mathbf{a}_i = \mathbf{b}.$$

Subtracting the second inequality from the first, we obtain

$$\sum_{i=1}^{k} (y_i - z_i)\mathbf{a}_i = \mathbf{0},$$

and since $\mathbf{y} \neq \mathbf{z}$, we obtain that the vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ are linearly dependent, which is a contradiction to the assumption that they are linearly independent. To prove the reverse direction, let us suppose that $\tilde{\mathbf{x}} \in P$ is an extreme point and assume in contradiction that $\tilde{\mathbf{x}}$ is not a basic feasible solution. This means that the columns corresponding to the positive components are linearly dependent. Without loss of generality, let us assume that the positive variables are exactly the first k components. The linear dependance of the first k columns means that there exists a nonzero vector $\mathbf{y} \in \mathbb{R}^k$ for which

$$\sum_{i=1}^{k} y_i \mathbf{a}_i = \mathbf{0}.$$

The above identity can also be written as $\mathbf{A}\tilde{\mathbf{y}} = \mathbf{0}$, where $\tilde{\mathbf{y}} = \binom{\mathbf{y}}{\mathbf{0}_{n-k}}$. Since the first k components of $\tilde{\mathbf{x}}$ are positive, it follows that there exist an $\varepsilon > 0$ for which $\mathbf{x}_1 = \tilde{\mathbf{x}} + \varepsilon\tilde{\mathbf{y}} \geq \mathbf{0}$ and $\mathbf{x}_2 = \tilde{\mathbf{x}} - \varepsilon\tilde{\mathbf{y}} \geq \mathbf{0}$. In addition $\mathbf{A}\mathbf{x}_1 = \mathbf{A}\tilde{\mathbf{x}} + \varepsilon\mathbf{A}\tilde{\mathbf{y}} = \mathbf{b} + \varepsilon \cdot \mathbf{0} = \mathbf{b}$, and similarly $\mathbf{A}\mathbf{x}_2 = \mathbf{b}$. We thus conclude that $\mathbf{x}_1, \mathbf{x}_2 \in P$; these two vectors are different since $\tilde{\mathbf{y}}$ is not the zeros vector, and finally, $\tilde{\mathbf{x}} = \frac{1}{2}\mathbf{x}_1 + \frac{1}{2}\mathbf{x}_2$, contradicting the assumption that $\tilde{\mathbf{x}}$ is an extreme point. $\square$

We finish this section with a very important and well-known theorem called the Krein–Milman theorem, stating that a compact convex set is the convex hull of its extreme points. We will state this theorem without a proof.

Theorem 6.35 (Krein–Milman). *Let $S \subseteq \mathbb{R}^n$ be a compact convex set. Then*

$$S = \operatorname{conv}(\operatorname{ext}(S)).$$

Exercises

6.1. Prove Theorem 6.8.

6.2. Give an example of two convex sets C_1, C_2 whose union $C_1 \cup C_2$ is not convex.

6.3. Show that the following set is not convex:

$$S = \{\mathbf{x} \in \mathbb{R}^2 : x_1^2 - x_2^2 + x_1 + x_2 \leq 4\}.$$

6.4. Prove that

$$\operatorname{conv}\{\mathbf{e}_1, \mathbf{e}_2, -\mathbf{e}_1, -\mathbf{e}_2\} = \{\mathbf{x} \in \mathbb{R}^2 : |x_1| + |x_2| \leq 1\},$$

where $\mathbf{e}_1 = (1,0)^T, \mathbf{e}_2 = (0,1)^T$.

6.5. Let K be a convex, bounded, and symmetric[2] set such that $\mathbf{0} \in \operatorname{int} K$. Define the Minkowski functional as

$$p(\mathbf{x}) = \inf\left\{\lambda > 0 : \frac{\mathbf{x}}{\lambda} \in K\right\}, \quad \mathbf{x} \in \mathbb{R}^n.$$

Show that $p(\cdot)$ is a norm.

6.6. Show the following properties of the convex hull:

(i) If $S \subseteq T$, then $\operatorname{conv}(S) \subseteq \operatorname{conv}(T)$.

(ii) For any $S \subseteq \mathbb{R}^n$, the identity $\operatorname{conv}(\operatorname{conv}(S)) = \operatorname{conv}(S)$ holds.

(iii) For any $S_1, S_2 \subseteq \mathbb{R}^n$ the following holds:

$$\operatorname{conv}(S_1 + S_2) = \operatorname{conv}(S_1) + \operatorname{conv}(S_2).$$

[2]A set S is called *symmetric* if $\mathbf{x} \in S$ implies $-\mathbf{x} \in S$.

6.7. Let C be a convex set. Prove that cone(C) is a convex set.

6.8. Show that the conic hull of the set

$$S = \left\{(x_1, x_2) : (x_1 - 1)^2 + x_2^2 = 1\right\}$$

is the set

$$\{(x_1, x_2) : x_1 > 0\} \cup \{(0, 0)\}.$$

Remark: This is an example illustrating the fact that the conic hull of a closed set is not necessarily a closed set.

6.9. Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n (\mathbf{a} \neq \mathbf{b})$. For what values of μ is the set

$$S_\mu = \{\mathbf{x} \in \mathbb{R}^n : ||\mathbf{x} - \mathbf{a}|| \leq \mu ||\mathbf{x} - \mathbf{b}||\}$$

convex?

6.10. Let $C \subseteq \mathbb{R}^n$ be a nonempty convex set. For each $\mathbf{x} \in C$ define the *normal cone* of C at $\mathbf{x}$ by

$$N_C(\mathbf{x}) = \{\mathbf{w} \in \mathbb{R}^n : \langle \mathbf{w}, \mathbf{y} - \mathbf{x} \rangle \leq 0 \text{ for all } \mathbf{y} \in C\},$$

and define $N_C(\mathbf{x}) = \emptyset$ when $\mathbf{x} \notin C$. Show that $N_C(\mathbf{x})$ is a closed convex cone.

6.11. Let $C \subseteq \mathbb{R}^n$ be cone. The **dual cone** is defined by

$$C^* = \{\mathbf{y} \in \mathbb{R}^n : \langle \mathbf{y}, \mathbf{x} \rangle \geq 0 \text{ for all } \mathbf{x} \in C\}.$$

(i) Prove that C^* is a closed convex cone (even if C is nonconvex).

(ii) Prove that if C_1, C_2 are cones satisfying $C_1 \subseteq C_2$, then $C_2^* \subseteq C_1^*$.

(iii) Show that $(L^n)^* = L^n$.

(iv) Show that the dual cone of $K = \{\binom{\mathbf{x}}{t} : ||\mathbf{x}||_1 \leq t\}$ is

$$K^* = \left\{ \begin{pmatrix} \mathbf{x} \\ t \end{pmatrix} : ||\mathbf{x}||_\infty \leq t \right\}.$$

6.12. A cone K is called *pointed* if it contains no lines, meaning that $\mathbf{x}, -\mathbf{x} \in K \Rightarrow \mathbf{x} = \mathbf{0}$. Show that if K has a nonempty interior, then K^* is pointed.

6.13. Consider the optimization problem

$$(P_\mathbf{a}) \quad \min\{\mathbf{a}^T \mathbf{x} : \mathbf{x} \in S\},$$

where $S \subseteq \mathbb{R}^n$. Let $\mathbf{x}^* \in S$ and let $K \subseteq \mathbb{R}^n$ be the set of all vectors $\mathbf{a}$ for which $\mathbf{x}^*$ is an optimal solution of $(P_\mathbf{a})$. Show that K is a convex cone.

6.14. Prove Lemma 6.20.

6.15. Prove Lemma 6.22.

6.16. Find all the basic feasible solutions of the system

$$\begin{aligned} -4x_2 + x_3 &= 6, \\ 2x_1 - 2x_2 - x_4 &= 1, \\ x_1, x_2, x_3, x_4 &\geq 0. \end{aligned}$$

6.17. Let S be a convex set. Prove that $\mathbf{x} \in S$ is an extreme point of S if and only if $S \backslash \{\mathbf{x}\}$ is convex.

6.18. Let $S = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_\infty \leq 1\}$. Show that

$$\text{ext}(S) = \{\mathbf{x} \in \mathbb{R}^n : x_i^2 = 1, i = 1, 2, \ldots, n\}.$$

6.19. Let $S = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 \leq 1\}$. Show that

$$\text{ext}(S) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 = 1\}.$$

6.20. Let $X_i \subseteq \mathbb{R}^{n_i}, i = 1, 2, \ldots, k$. Prove that

$$\text{ext}(X_1 \times X_2 \times \cdots \times X_k) = \text{ext}(X_1) \times \text{ext}(X_2) \times \cdots \times \text{ext}(X_k).$$