Chapter 2

Convex Sets: basic results

In this chapter, we introduce one of the most important tools in the mathematical approach to Economics, namely the theory of convex sets. Almost every situation we will meet will depend on this simple geometric property of a set, whether in simple models of consumer behavior, in the theory of optimization or, indeed, in much of modern analysis and computational mathematics.

As an independent idea, the notion of convexity appeared at the end of the 19th century, particularly in the works of Minkowski who is supposed to have said:

"Everything that is convex interests me."

Since that time, the number of books, papers, and their authors has mushroomed; at this stage it would be impossible to write a survey of the history of the subject of convexity. We can mention only a few names and references, those of T. Bonnesen and W. Fenchel [?], F.A. Valentine[?], of R. T. Rockafellar[?] and of F. H. Clarke[?].

In applications, convexity of often a crucial assumption. For example, when one seeks to model consumer preferences, convexity has the interpretation of diminishing marginal rates of substitution. In optimization models, the underlying convexity of domains is crucial to the existence of optimal solutions as well as to the development of a duality theory, e.g., in the study of multiplier rules which generalize the method of Lagrange multipliers in equality-constrained problems. The closely related notion of convex and concave functions which are so prevalent in economic models, are defined in terms of convex sets; we will discuss such functions in a later chapter.

The geometric definition, as we will see, makes sense in any vector space. Since, for the most of our work we deal only with \( \mathbb{R}^n \), the definitions will be stated in that
CHAPTER 2. CONVEX SETS

context. The interested student may, however, reformulate the definitions either, in the setting of an arbitrary real vector space, or in some concrete vector space as, for example, $C([0, 1]; \mathbb{R})$. One advantage of concentrating on familiar Euclidean space $\mathbb{R}^n$ is that we can take advantage of the interplay between the vector space structure and the topological structure, that is, the notions of open and closed sets, bounded sets, and convergence of sequences in $\mathbb{R}^n$, all of which is familiar territory. We need not, in this book, discuss more abstract structures in which all of these basic ideas are available.

From an intuitive point of view, the notion of a convex set is very simple. In $\mathbb{R}^2$ or $\mathbb{R}^3$ a convex set is one that contains all the points of any line segment joining two points of the set. We think, concretely, of a disk in $\mathbb{R}^2$ or a ball in $\mathbb{R}^3$ as basic examples. Our goal is to make this idea precise and develop some basic theory of such sets, including polyhedral sets, cones, extreme points, representation theorems for convex sets, as well as indicating some applications of this complex of ideas to economics.

2.1 Basics

2.1.1 Definitions and Examples

Convex sets are defined with reference to a line segment joining two points of the set. Here we start with a definition that we use often to check that a set is convex.

**Definition 2.1.1** Let $u, v \in \mathbb{R}^n$. Then the set of all convex combinations of $u$ and $v$ is the set of points

$$\{w_\lambda \in \mathbb{R}^n \mid w_\lambda = (1 - \lambda)u + \lambda v, \ 0 \leq \lambda \leq 1\}.$$  \hspace{1cm} (2.1)

In say $\mathbb{R}^2$ or $\mathbb{R}^3$, this set is exactly the line segment joining the two points $u$ and $v$. (See the examples below.) Convex sets are now defined in terms of these line segments.

**Definition 2.1.2** Let $C \subset \mathbb{R}^n$. Then the set $C$ is said to be convex provided that given two points $u, v \in C$ the set (2.1) is a subset of $C$.

We give some simple examples:

---

1 This symbol stands for the real vector space of continuous, real-valued functions defined on the closed interval $[0, 1]$. 
Examples 2.1.3

(a) An interval of \([a, b] \subset \mathbb{R}\) is a convex set. To see this, let \(c, d \in [a, b]\) and assume, without loss of generality, that \(c < d\). Let \(\lambda \in (0, 1)\). Then,

\[
\begin{align*}
a &\leq c = (1 - \lambda)c + \lambda c < (1 - \lambda)c + \lambda d \\
&< (1 - \lambda)d + \lambda d = d \\
&\leq b.
\end{align*}
\]

(b) A disk with center \((0, 0)\) and radius \(c\) is a convex subset of \(\mathbb{R}^2\). This may be easily checked (Exercise!) by using the usual distance formula in \(\mathbb{R}^2\) namely \(\|x - y\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}\) and the triangle inequality \(\|u + v\| \leq \|u\| + \|v\|\).

(c) In \(\mathbb{R}^n\) the set \(H := \{x \in \mathbb{R}^n \mid a_1x_1 + \ldots + a_nx_n = c\}\) is a convex set. For any particular choice of constants \(a_i\) it is called a hyperplane in \(\mathbb{R}^n\). Its defining equation is a generalization of the usual equation of a plane in \(\mathbb{R}^3\), namely the equation \(ax + by + cz + d = 0\), and hence the name.

To see that \(H\) is indeed a convex set, let \(x^1, x^2 \in H\) and define \(z \in \mathbb{R}^3\) by \(z := (1 - \lambda)x^1 + \lambda x^2\). Then

\[
z = \sum_{i=1}^{n} a_i[(1 - \lambda)x^1_i + \lambda x^2_i] = \sum_{i=1}^{n} [(1 - \lambda)a_i x^1_i + \lambda a_i x^2_i]
\]

\[
= (1 - \lambda) \sum_{i=1}^{n} a_i x^1_i + \lambda \sum_{i=1}^{n} a_i x^2_i = (1 - \lambda)c + \lambda c
\]

Hence \(z \in H\).

(d) As a generalization of the preceding example\(^2\), let \(A\) be an \(m \times n\) matrix, \(b \in \mathbb{R}^m\), and let \(S = \{x \in \mathbb{R}^n \mid Ax = b\}\). The set \(S\) is just the set of all solutions of the linear equation \(Ax = b\). This set \(S\) is a convex subset of \(\mathbb{R}^n\). Indeed, let \(x^1, x^2 \in S\). Then

\[
A \left( (1 - \lambda)x^1 + \lambda x^2 \right) = (1 - \lambda)A(x^1) + \lambda A(x^2) = (1 - \lambda)b + \lambda b = b.
\]

\(^2\)Indeed, just take the matrix \(A\) to be a row matrix \(A = (a_1, a_2, \ldots, a_n)\)
(e) There are always two, so-called *trivial examples*. These are the empty set \( \emptyset \), and the entire space \( \mathbb{R}^n \). Note also that a singleton \( \{x\} \) is convex. In this latter case, as in the case of the empty set, the definition is satisfied vacuously.

As with Example (b) above, one important fact about \( \mathbb{R}^n \) is that the closed unit ball

\[
\overline{B}_1(0) = \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}
\]

is a convex set. This follows from the triangle equality for norms: for any \( x, y \in B_1 \) and any \( \lambda \in [0, 1] \) we have

\[
\|(1 - \lambda)x + \lambda y\| \leq (1 - \lambda)\|x\| + \lambda\|y\| \leq (1 - \lambda) + \lambda = 1.
\]

Now the ball \( \overline{B}_1(0) \) is a closed set. It is easy to see that if we take its interior

\[
\overline{B}_1(0)^\circ = \{x \in \mathbb{R}^n \mid \|x\| < 1\},
\]

this set is also convex. This gives us a hint regarding our next result.

**Proposition 2.1.4** If \( C \subset \mathbb{R}^n \) is convex, the set \( \text{cl}(C) \), the closure of \( C \), is also convex.

**Proof:** Suppose \( x, y \in \text{cl}(C) \). Then there exist sequences \( \{x^k\}_{k=1}^\infty \) and \( \{y^k\}_{k=1}^\infty \) in \( C \) such that \( x^k \to x \) and \( y^k \to y \) as \( k \to \infty \). For some \( 0 \leq \lambda \leq 1 \), define \( z^k := (1 - \lambda)x^k + \lambda y^k \). Then, by convexity of \( C \), \( z^k \in C \). Moreover \( z^k \to (1 - \lambda)x + \lambda y \) as \( k \to \infty \). Hence this latter point lies in \( \text{cl}(C) \). \( \blacksquare \)

We need to be careful here for the statement that if the closure of a set is convex, the original set is as well is false as the next example shows.

**Example 2.1.5** Consider the set

\[
\hat{D} = B_1(0) \setminus \{0\},
\]

the punctured open disk. Then \( \text{cl}(\hat{D}) = \text{cl}(B_1(0)) \) which is certainly convex but clearly the set \( \hat{D} \) is not since the point \((0,0) \notin \hat{D}\). Of course here \( \hat{D} \neq \text{int}(\text{cl}(B_1(0))) \).
Having produced this example, we can nevertheless show that the interior of a convex set is convex. To do this we first show that, for any non-empty convex set $C$ with non-empty interior, a line segment joining an arbitrary point of $C$ with an interior point lies entirely in the interior of $C$. The result is most easily proved by recalling two basic fact about balls in $\mathbb{R}^n$.

**Lemma 2.1.6**  
(a) For all $\alpha \geq 0$, $\alpha B(0) = B_{\alpha \epsilon}(0)$.  
(b) $B(\epsilon x_o) = x_o + B(\epsilon)$.  
(c) For all $\alpha \geq 0$, $\alpha B(\epsilon x_o) = B_{\alpha \epsilon}(x_o)$.

**Proof:** Exercise.

**Lemma 2.1.7** Suppose $C \subset \mathbb{R}^n$ is convex and has non-empty interior. Then if $x_o \in \text{int } (C)$ and $x \in C$, then any $z$ of the form $z = (1 - \lambda) x_o + \lambda x$ with $0 \leq \lambda < 1$ lies in the interior of $C$.

**Proof:** From Proposition 2.1.4 we know that if $C$ is convex, then $\bar{C}$ is convex. We show that

$$\lambda \bar{C} + (1 - \lambda) \text{ int } (C) \subset \text{ int } (C), \text{ for } 0 \leq \lambda < 1.$$  

The first step is to show that the set on the left-hand side of this inclusion is an open set. That means that we need to show that every point of that set is the center of some open ball completely contained in the left-hand side. To do this, choose $x \in \lambda \bar{C} + (1 - \lambda) \text{ int } (C) \subset \hat{C}$. Then $x = \lambda \epsilon + (1 - \lambda) z$ for $\epsilon \in C$ and $z \in \hat{C}$. Since $z \in \hat{C}$ there is an $\epsilon > 0$ such that $B_\epsilon(z) \subset C$ and so, since $(1 - \lambda) \leq 1$, $(1 - \lambda) B_\epsilon(z) = B_{(1 - \lambda)\epsilon}(z) \subset C$. It follows that $x \in \lambda \epsilon + B_{(1 - \lambda)\epsilon}(z) = B_{(1 - \lambda)\epsilon}(z + \lambda \epsilon)$. This latter set is an open ball and hence the result.

Now, since the set $\lambda \bar{C} + (1 - \lambda) \text{ int } (C)$ is open, in order to complete that proof it suffices to show that this set is contained in $C$. To this end, let $z \in \hat{C}$. Then $(1 - \lambda)(\hat{C} - z)$ is an open set containing zero. Hence

$$\lambda \bar{C} = \lambda \bar{C} \subset \lambda C + (1 - \lambda)[\hat{C} - z]$$

$$= \lambda C + (1 - \lambda) \hat{C} - (1 - \lambda) z \subset C - (1 - \lambda) z.$$  

This inclusion then means that $\lambda \bar{C} + (1 - \lambda) z \subset C$ for all $z \in \hat{C}$. ■

From this last lemma, we have the promised proposition whose proof is now trivial.
Proposition 2.1.8 Let $C \subset \mathbb{R}^n$ be a convex set. Then $\text{int}(C)$ is also convex.

Proof: Let $x, y \in \overset{o}{C}$, $\lambda \in [0, 1]$ and take $z = (1 - \lambda)x + \lambda y$. If $\lambda < 1$ we apply the lemma to show $z \in \overset{o}{C}$. For $\lambda = 1$ we have $z = y \in \overset{o}{C}$. ■

The next results concern convexity and the usual set operations.

The simple example of the two intervals $[0, 1]$ and $[2, 3]$ on the real line shows that the union of two convex sets in not necessarily convex. On the other hand, we have the result concerning intersections:

Proposition 2.1.9 The intersection of any number of convex sets is convex.

Proof: Let $\{C_\alpha\}_{\alpha \in A}$ be a family of convex sets, and let $C := \cup_{\alpha \in A} C_\alpha$. Then, for any $x, y \in C$ by definition of the intersection of a family of sets, $x, y \in C_\alpha$ for all $\alpha \in A$ and each of these sets is convex. Hence for any $\alpha \in A$, and $\lambda \in [0, 1]$, $(1 - \lambda)x + \lambda y \in C_\alpha$. Hence $(1 - \lambda)x + \lambda y \in C$. ■

While, by definition, a set is convex provided all convex combinations of two points in the set is again in the set, it is a simple matter to check that we can make a more general statement. This statement is the content of the following proposition. Notice the way in which the proof is constructed; it is often very useful in computations!

Proposition 2.1.10 Let $C$ be a convex set and let $\lambda_1, \lambda_2, \ldots, \lambda_p \geq 0$ and $\sum_{i=1}^{p} \lambda_i = 1$. If $x^1, x^2, \ldots, x^p \in C$ then

$$\sum_{i=1}^{p} \lambda_i x^i \in C.$$  

Proof: We prove the result by induction. Since $C$ is convex, the result is true, trivially, for $p = 1$ and by definition for $p = 2$. Suppose that the proposition is true for $p = r$ (induction hypothesis!) and consider the convex combination $\lambda_1 x^1 + \lambda_2 x^2 + \ldots + \lambda_{r+1} x^{r+1}$. Define $\Lambda := \sum_{i=1}^{r} \lambda_i$. Then since $1 - \Lambda = \sum_{i=1}^{r+1} \lambda_i - \sum_{i=1}^{r} \lambda_i = \lambda_{r+1}$, we have

$$\left(\sum_{i=1}^{r} \lambda_i x^i\right) + \lambda_{r+1} x^{r+1} = \Lambda \left(\sum_{i=1}^{r} \frac{\lambda_i}{\Lambda} x^i\right) + (1 - \Lambda) x^{r+1}.$$
Note that \( \sum_{i=1}^{r} (\lambda_i / \Lambda) = 1 \) and so, by the induction hypothesis, \( \sum_{i=1}^{r} (\lambda_i / \Lambda) x^i \in C \). Since \( x^{r+1} \in C \) it follows that the right hand side is a convex combination of two points of \( C \) and hence lies in \( C \). \[ \square \]

Relative to the vector space operations, we have the following result:

**Proposition 2.1.11** Let \( C, C_1 \) and \( C_2 \) be convex sets in \( \mathbb{R}^n \) and let \( \beta \in \mathbb{R} \). Then

(a) The set \( \beta C := \{ z \in \mathbb{R}^n \mid z = \beta x, x \in C \} \) is convex.

(b) The set \( C_1 + C_2 := \{ x \in \mathbb{R}^n \mid z = x^1 + x^2, x^1 \in C_1, x^2 \in C_2 \} \) is convex.

**Proof:** For part (a), let \( z^1 \) and \( z^2 \) both be elements or \( \beta C \). Then there exists points \( x^1, x^2 \in C \) such that \( z^1 = \beta x^1 \) and \( z^2 = \beta x^2 \). Choose any \( \lambda \in [0, 1] \) and form the convex combination

\[
z = (1 - \lambda) z^1 + \lambda z^2.
\]

But then

\[
z = (1 - \lambda) \beta x^1 + \lambda \beta x^2 = \beta [(1 - \lambda) x^1 + \lambda x^2].
\]

But \( C \) is convex so that \( (1 - \lambda) x^1 + \lambda x^2 \in C \) and hence \( z \in \beta C \). This proves part (a).

Part (b) is proved by a similar argument by simply noting that

\[
(1 - \lambda) (x^1 + x^2) + \lambda (y^1 + y^2) = (1 - \lambda) x^1 + (1 - \lambda) x^2 + \lambda y^1 + \lambda y^2. \quad \square
\]

### 2.1.2 Convex Hulls and Carathéodory’s Theorem

For any given set, \( S \), which is not convex, we often want to “convexify” it, i.e., to find the smallest set which is convex and which contains the given one. Of course it is easy to find a convex set containing the non-convex \( S \) since the entire vector space \( \mathbb{R}^n \) is obviously a convex set. But it is usually too big; we wish to be more economical and find, in a precise sense, the smallest convex set containing our non-convex set. Since we know from Proposition 2.1.9 that the intersection of any family of convex sets is convex, and since the family of all convex sets that contain \( S \) is non-empty, we take the intersection of all such sets. This leads to the notion of the convex hull or convex envelope of a set.
Definition 2.1.12 The convex hull of a set \( C \) is the intersection of all convex sets which contain the set \( C \). We denote the convex hull by \( \text{co}(C) \).

We illustrate this definition with some examples.

Examples 2.1.13

(a) Suppose that \([a,b]\) and \([c,d]\) are two intervals of the real line with \( c < b \) so that the intervals are disjoint. Then the convex hull of the set \([a,b] \cup [c,d]\) is just the interval \([a,d]\).

(b) In \( \mathbb{R}^2 \) consider the annular region \( E \) consisting of the ring
\[
\{(x,y) \in \mathbb{R}^2 : r^2 \leq x^2 + y^2 \leq R\}
\]
for a given positive numbers \( r \) and \( R \).

Clearly the set \( E \) is not convex for the line segment joining the indicated points \( P = (0,R) \) and \( Q = (0,-R) \) has points lying in the “hole” of region and hence not in \( E \). Indeed, this is the case for any line segment joining two points of the region which are, say, symmetric with respect to the origin. Clearly the entire disk of radius \( R \) is convex and indeed is the convex hull, \( \text{co}(E) \).

These examples are typical. In each case, we see that the convex hull is obtained by adjoining all linear combinations of points in the original set. This is indeed a general result.

Theorem 2.1.14 Let \( C \subset \mathbb{R}^n \). Then the set of all convex combinations of points of the set \( C \) is exactly \( \text{co}(C) \).

Proof: Let us denote the set of all convex combinations of points of \( C \) by \( L(C) \). It is clear from Proposition 2.1.10 that \( L(C) \supset \text{co}(C) \). To see that the opposite inclusion holds, simply observe that if \( K \supset C \) is convex, then it must contain all the convex combinations of points of \( C \) and hence \( L(C) \subset K \). From this it follows that \( L(C) \) is a convex set, containing \( C \) and contained in every convex subset that contains \( C \), hence \( L(C) \subset \text{co}(C) \). ■

Convex sets in \( \mathbb{R}^n \) have a very nice characterization discovered by Constantin Carathéodory. His theorem, often called Carathéodory’s Theorem although there are a number of results with this name in different subject areas, has the nature of a representation theorem somewhat analogous to the theorem which says that any vector in a vector space can be represented as a linear combination of the elements of a basis. One thing both theorems do, is to give a finite and minimal representation of all elements of an infinite set.
Theorem 2.1.15 Let $C$ be a subset of $\mathbb{R}^n$. Then every element of $\text{co}(C)$ can be represented as a convex combination of no more than $(n + 1)$ elements of $C$.

Proof: Let $x \in \text{co}(C)$. Then $x$ is a convex combination of points of $C$, and we write

$$x = \sum_{i=1}^{m} \alpha_i x^i, \quad x^i \in C, \quad \alpha_i \geq 0, \quad \sum_{i=1}^{m} \alpha_i = 1.$$ 

Let us assume that $m$ is the minimal number of vectors for which such a representation is possible. In particular, this means that for all $i = 1, \ldots, m$ we have $\alpha_i > 0$, since if not, the number $m$ would not be minimal.

Now, if $m \leq n+1$ there is nothing to prove. On the other hand, suppose that $m > n+1$. Then the vectors $x^i - x^1, i = 2, \ldots, m$, must be linearly dependent since $m - 1 > n$. This means that there are scalars $\beta_i, i = 2, \ldots, m$, not all zero, such that

$$\sum_{i=2}^{m} \beta_i (x^i - x^1) = 0.$$

Now define $\beta_1 := -\sum_{i=2}^{m} \beta_i$. Then,

$$\sum_{i=1}^{m} \beta_i x^i = \beta_1 x^1 + \sum_{i=2}^{m} \beta_i x^i = -\sum_{i=2}^{m} \beta_i x^1 + \sum_{i=2}^{m} \beta_i x^i = 0,$$

and $\sum_{i=1}^{m} \beta_i = 0$. Hence at least one $\beta_i > 0$ since this is a sum of numbers which are not all zero.

Then, introducing a real parameter $\lambda$, and observing that $\sum_{i=1}^{m} \beta_i x^i = 0$,

$$x = \sum_{i=1}^{m} \alpha_i x^i - \lambda \sum_{i=1}^{m} \beta_i x^i = \sum_{i=1}^{m} (\alpha_i - \lambda \beta_i) x^i.$$ 

Now, recalling that all the $\alpha_i > 0$, let $\hat{\lambda}$ be given by

$$\hat{\lambda} := \min_{1 \leq i \leq m} \left\{ \frac{\alpha_i}{\beta_i} \bigg| \beta_i > 0 \right\} = \frac{\alpha_j}{\beta_j}.$$ 

From the definition, $\hat{\lambda} > 0$, and, for every $i, 1 \leq i \leq m$, we have $\alpha_i - \hat{\lambda} \beta_i \geq 0$, with, when $i = j, \alpha_j - \hat{\lambda} \beta_j = 0$.

Therefore,
\[ \mathbf{x} = \sum_{i=1}^{m} (\alpha_i - \hat{\lambda} \beta_i) \mathbf{x}^i, \]

where, for every \( i \), \( (\alpha_i - \hat{\lambda} \beta_i) \geq 0 \) and

\[ \sum_{i=1}^{m} (\alpha_i - \hat{\lambda} \beta_i) = \left( \sum_{i=1}^{m} \alpha_i \right) - \hat{\lambda} \left( \sum_{i=1}^{m} \beta_i \right) = \sum_{i=1}^{m} \alpha_i = 1, \]

so that, since one of these (the \( j^{th} \)) vanishes, we have a convex combination of fewer than \( m \) points which contradicts the minimality of \( m \).

The drawback of Carathéodory’s Theorem, unlike the representation of a vector in a vector space by set of basis vectors, is that the choice of elements used to represent the point is neither uniquely determined for that point, nor does the theorem guarantee that the same set of vectors in \( C \) can be used to represent all vectors in \( C \); the representing vectors will usually change with the point being represented. Nevertheless, the theorem is useful in a number of ways as we will see presently. First, a couple of examples.

**Examples 2.1.16**  
(a) Recalling that intervals of \( \mathbb{R} \) are convex sets, in \( \mathbb{R} \), consider the interval \([0, 1]\) and the subinterval \((1/4, 3/4)\). Then \( \text{co} (1/4, 3/4) = [1/4, 3/4] \). If we take the point \( x = 1/2 \), then we have both

\[ x = \frac{1}{2} \left( \frac{3}{8} \right) + \frac{1}{2} \left( \frac{5}{8} \right) \quad \text{and} \quad x = \frac{3}{4} \left( \frac{7}{16} \right) + \frac{1}{4} \left( \frac{11}{16} \right). \]

So that certainly there is no uniqueness in the representation of \( x = 1/2 \).

(b) In \( \mathbb{R}^2 \) we consider the two triangular regions, \( T_1, T_2 \), joining the points \((0, 0), (1, 4), (0, 2), (3, 4)\) and \((4, 0)\). Joining the apexes of the triangles forms a trapezoid which is a convex set. It is the convex hull of the set \( T_1 \cup T_2 \).

Again, it is clear that two points which both lie in one of the original triangles have more than one representation. Similarly, if we choose two points, one from \( T_1 \) and one from \( T_2 \), say the points \((1, 2)\) and \((3, 2)\), the point

\[ \frac{1}{2} \left( \begin{array}{c} 1 \\ 2 \end{array} \right) + \frac{1}{2} \left( \begin{array}{c} 3 \\ 2 \end{array} \right) = \left( \begin{array}{c} 2 \\ 2 \end{array} \right). \]
2.1. BASICS

does not lie in the original set \( T_1 \cup T_2 \), but does lie in the convex hull. Moreover, this point can also be represented by

\[
\frac{1}{3} \begin{pmatrix} \frac{3}{2} \\ \frac{1}{2} \end{pmatrix} + \frac{2}{3} \begin{pmatrix} \frac{9}{4} \\ \frac{11}{4} \end{pmatrix}
\]

as can easily be checked.

2.1.3 Compact Convex Sets and Minimum Norm Problems

The next results depend on the notion of norm in \( \mathbb{R}^n \) and on the convergence of a sequence of points in \( \mathbb{R}^n \). In particular, it relies on the fact that, in \( \mathbb{R}^n \), or for that matter in any complete metric space, Cauchy sequences converge.

Recall that a set of points in \( \mathbb{R}^n \) is called **compact** provided it is closed and bounded. We have seen in Proposition ?? that this is equivalent to the statement that every sequence in the set contains a convergent subsequence. As a corollary to Carathéodory’s Theorem, we have the next result about compact sets:

**Corollary 2.1.17** The convex hull of a compact set in \( \mathbb{R}^n \) is compact.

**Proof:** Let \( C \subset \mathbb{R}^n \) be compact. We need to show that \( \text{co}(C) \) is also closed and bounded. Since \( C \) is bounded, it is contained in some ball \( B_M(0) \) of radius \( M > 0 \) and this is a convex set containing \( C \). So \( \text{co}(C) \) must also be contained in this ball by definition of the convex hull. Hence the convex hull is indeed bounded.

Notice that the simplex

\[
\sigma := \left\{ (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n : \sum_{i=1}^{n} \lambda_i = 1 \right\}
\]

is also closed and bounded and is therefore compact. (Check!) Now suppose that \( \{v(j)\}_{j=1}^\infty \subset \text{co}(C) \). By Carathéodory’s Theorem, each \( v^j \) can be written in the form

\[
v^k = \sum_{i=1}^{n+1} \lambda_{k,i} x^{k,i}, \quad \text{where} \quad \lambda_{k,i} \geq 0, \sum_{i=1}^{n+1} \lambda_{k,i} = 1, \quad \text{and} \quad x^{k,i} \in C.
\]

Then, since \( C \) and \( \sigma \) are compact, there exists a sequence \( k_1, k_2, \ldots \) such that the limits

\[
\lim_{j \to \infty} \lambda_{k_j,i} = \lambda_i \quad \text{and} \quad \lim_{j \to \infty} x^{k_j,i} = x^i \quad \text{exist for} \quad i = 1, 2, \ldots, n + 1.
\]

Clearly \( \lambda_i \geq 0, \sum_{i=1}^{n+1} \lambda_i = 1 \) and \( x_i \in C \).
Thus, the sequence \( \{v^k\}_{k=1}^\infty \) has a subsequence, \( \{v^{k_j}\}_{j=1}^\infty \) which converges to a point of \( \text{co} (C) \) which shows that this latter set is compact.

The next result shows that if \( C \) is closed and convex (but perhaps not bounded) is has an element with smallest norm. It is a simple result from analysis that involves the facts that the function \( x \rightarrow \|x\| \) is a continuous map from \( \mathbb{R}^n \rightarrow \mathbb{R} \) and, again, that Cauchy sequences in \( \mathbb{R}^n \) converge. It also relies heavily on the parallelogram law for the Euclidean norm. The proof itself foreshadows the "direct method" for minimization problems where a so-called minimizing sequence is used and a convergent subsequence is extracted.

**Theorem 2.1.18** Every closed convex subset of \( \mathbb{R}^n \) has a unique element of minimum norm.

**Proof:**

Let \( C \) be such a set and note that \( \iota := \inf_{x \in C} \|x\| \geq 0 \) so that the function \( x \rightarrow \|x\| \) is bounded below on \( C \). Let \( x^1, x^2, \ldots \) be a sequence of points of \( C \) such that

\[
\lim_{i \to \infty} \|x^i\| = \iota. \tag{3}
\]

Then, by the parallelogram law, \( \|x^i - x^j\|^2 = 2\|x^i\|^2 + 2\|x^j\|^2 - 4\|\frac{1}{2}(x^i + x^j)\|^2 \). Since \( C \) is convex, \( \frac{1}{2}(x^i + x^j) \in C \) so that \( \|\frac{1}{2}(x^i + x^j)\| \geq \iota \). Hence

\[
\|x^i - x^j\|^2 \leq 2\|x^i\|^2 + 2\|x^j\|^2 - 4\iota^2.
\]

As \( i, j \to \infty \), we have \( 2\|x^i\|^2 + 2\|x^j\|^2 - 4\iota \to 0 \). Thus, \( \{x^j\}_{j=1}^\infty \) is a Cauchy sequence and has a limit point \( x^\circ \). Since \( C \) is closed, \( x^\circ \in C \). Moreover, since the function \( x \rightarrow \|x\| \) is a continuous function from \( \mathbb{R}^n \rightarrow \mathbb{R} \),

\[
\iota = \lim_{j \to \infty} \|x^j\| = \|x^\circ\|.
\]

So \( x^\circ \) is the point of minimal norm in \( C \).

In order to show uniqueness of the point with minimal norm, suppose that there were two points, \( x^\circ, y^\circ \in K, x^\circ \neq y^\circ \) such that \( \|x^\circ\| = \|y^\circ\| = \iota \). Then by the parallelogram law,

\[
0 < \|x^\circ - y^\circ\|^2 = 2\|x^\circ\|^2 + 2\|y^\circ\|^2 - 4\left\|\frac{1}{2}(x^\circ + y^\circ)\right\|^2
\]

\[
= 2\iota^2 + 2\iota^2 - 4\left\|\frac{1}{2}(x^\circ + y^\circ)\right\|^2
\]

Here, and throughout this course, we shall call such a sequence a minimizing sequence.
so that \( 4 \iota > 4 \| \frac{1}{2} (x^0 + y^0) \|_2^2 \) or \( \frac{1}{2} (x^0 + y^0) \| < \iota \) which would give a vector in \( C \) of norm less than the infimum \( \iota \) which is impossible. \( \blacksquare \)

**Example 2.1.19** It is easy to illustrate the statement of this last theorem in a concrete case. Suppose that we define three sets in \( \mathbb{R}^2 \) by

\[
H_1^+ := \{(x, y) \in \mathbb{R}^2 : 5x - y \geq 1\}, \quad H_2^+ := \{(x, y) \in \mathbb{R}^2 : 2x + 4y \geq 7\} \quad \text{and} \quad H_3^+ := \{(x, y) \in \mathbb{R}^2 : 2x + 2y \geq 6\}
\]

whose intersection (the intersection of half-spaces) forms a convex set illustrated below. The point of minimal norm is the closest point in this set to the origin. From the projection theorem in \( \mathbb{R}^2 \), that point is determined by the intersection of the boundary line \( 2x + 4y = 6 \) with a line perpendicular to it and which passes through the origin as illustrated here.

We emphasize that minimum norm problems are very important in applications and that a wide variety of problems can be recast as this kind of problem. Here is an important example.

**Example 2.1.20** Here we consider a problem from linear algebra. Consider the equation \( A x = b \) where \( A \) is a real \( n \times m \) matrix. If \( b \) is in the range of \( A \) then either the system has a unique solution, or it has infinitely many. As we have checked in a previous example, the solution set \( S = \{ x \in \mathbb{R}^n \mid A x = b \} \) is a convex set. So, according to Theorem 2.1.18 there is an element of \( S \) of minimum norm. This is called the minimum norm solution of the linear system and it is unique. Note that if the set \( S \) is a singleton, the unique element in \( S \) is the minimum norm solution by default.

Of course, if \( b \notin \mathcal{R}(A) \) then there is no solution of the problem. Then one is often interested in the least squares solution, that is, a solution to the minimization problem for the Euclidean norm

\[
\min_{x \in \mathbb{R}^n} \| A x - b \|.
\]

We will assume here that a solution of this problem exists. But there is certainly no guarantee that such a solution is unique. In fact it usually is not. However, we can seek, among all solutions of this least-squares problem, the minimum norm solution. To see that the minimum norm problem here has a solution, we need only the result of a lemma.

**Lemma 2.1.21** Given an inconsistent linear system \( A x = b \), the set of least-squares solutions is convex.
**Proof:** Let $\Omega$ denote the set of all minimizers of $\|Ax - b\|$ and let $u, v \in \Omega$ and denote the minimum value of the norm by $m$. Choose $\lambda \in [0, 1]$ and look at the convex combination $z = (1 - \lambda)u + \lambda v \in \mathbb{R}^n$. Then we have

\[
m \leq \|Az - b\| = \|(1 - \lambda)u + \lambda v - b\| = \|A((1 - \lambda)u + \lambda v) - (1 - \lambda)b + \lambda b\| \\
\leq (1 - \lambda)\|Au - b\| + \lambda\|Av - b\| = (1 - \lambda)m + \lambda m = m.
\]

Thus we have equality throughout and so $z \in \Omega$. Hence $\Omega$ is convex. 

We conclude that Theorem 2.1.18 guarantees the existence of a minimum norm least-squares solution. This minimum norm least-squares solution is related to the right-hand $b$, of the original equation by a matrix which is called the generalized inverse of the matrix $A$.

In this example we say nothing about how to compute these solutions for which there is a rich literature.

The notion of projection is usually first met in Linear Algebra, but it is certainly a familiar idea even earlier. Briefly, suppose that we are working in $\mathbb{R}^2$ with the usual coordinate axes and are given some point not lying on either axis. Then if $(x, y) \in \mathbb{R}^2$ it is common to call the mapping $P((x, y)) = x$ the projection of the point onto the $x$-axis. Here $P$ is linear map with the properties that $(I - P)((x, y)) = y$ and $P^2 = P$. Notice also that each of the axes constitutes a convex set in $\mathbb{R}^2$ and that the Pythagorean Theorem tells us that the smallest distance between the point $(x, y)$ and the $x$-axis is the distance between the two points $(x, y)$ and $(x, 0)$.

In convex analysis, this notion of projection is generalized. This is a fundamental theorem called the Projection Theorem which defines and proves the existence of a point of a closed convex set closest to a given point outside the set. The theorem is closely related to the theorem on minimum norms, Theorem 2.1.18. We will use it in the next sections. As we will see, it is also central to our proof of the all-important Separation Theorem (and its corollaries) which depends on the idea of the projection of a point onto a convex set. Here is the theorem:

**Theorem 2.1.22** Let $C \subset \mathbb{R}^n$ be a closed, convex set. Then
(a) For every \( x \in \mathbb{R}^n \) there exists a unique vector \( z^* \in C \) that minimizes \( \|z - x\| \) over all \( z \in C \). We call \( z^* \) the projection of \( x \) onto \( C \).

(b) \( z^* \) is the projection of \( x \) onto \( C \) if and only if

\[
\langle y - z^*, x - z^* \rangle \leq 0, \text{ for all } y \in C.
\]

**Proof:** Fix \( x \in \mathbb{R}^n \) and let \( w \in C \). Then minimizing \( \|x - z\| \) over all \( z \in C \) is equivalent to minimizing the same function over the set \( \{ z \in C \mid \|x - z\| \leq \|x - w\| \} \). This latter set is both closed and bounded and therefore the continuous function \( g(z) = \|z - x\| \), according to the theorem of Weierstrass, takes on its minimum at some point of the set.

We use the parallelogram identity to prove uniqueness as follows. Suppose that there are two distinct points, \( z^1 \) and \( z^2 \), which both minimize \( \|z - x\| \) and denote this minimum by \( \iota \). Then we have

\[
0 < \|(z^1 - x) - (z^2 - x)\|^2 = 2\|z^1 - x\|^2 + 2\|z^2 - x\|^2 - 4\left|\frac{1}{2}[(z^1 - x) + (z^2 - x)]\right|^2
\]

\[
= 2\|z^1 - x\|^2 + 2\|z^2 - x\|^2 - 4\left\|\frac{z^1 + z^2}{2} - x\right\|^2 = 2\iota^2 + 2\iota^2 - 4\|\hat{z} - x\|^2,
\]

where \( \hat{z} = (z^1 + z^2)/2 \in C \) since \( C \) is convex. Rearranging, and taking square roots, we have

\[
\|\hat{z} - x\| < \iota
\]

which is a contradiction of the fact that \( z^1 \) and \( z^2 \) give minimal values to the distance. Thus uniqueness is established.

To prove the inequality in part (b) we have, for all \( y, z \in C \), the inequality

\[
\|y - x\|^2 = \|y - z\|^2 + \|z - x\|^2 - 2\langle (y - z), (x - z) \rangle
\]

\[
\geq \|z - x\|^2 - 2\langle (y - z), (x - z) \rangle.
\]

Hence, if \( z \) is such that \( \langle (y - z), (x - z) \rangle \leq 0 \) for all \( y \in C \), then \( \|y - x\|^2 \geq \|z - x\|^2 \) for all \( y \in C \) and so, by definition \( z = z^* \).

To prove the necessity of the condition, let \( z^* \) be the projection of \( x \) onto \( C \) and let \( y \in C \) be arbitrary. For \( \alpha > 0 \) define \( y_\alpha = (1 - \alpha)z^* + \alpha y \) then
\[ \|x - y_\alpha\|^2 = \|(1 - \alpha)(x - z^*) + \alpha(x - y)\|^2 \]
\[ = (1 - \alpha)^2\|x - z^*\|^2 + \alpha^2\|x - y\|^2 + 2(1 - \alpha)\alpha \langle (x - z^*), (x - y) \rangle. \]

Now consider the function \( \varphi(\alpha) := \|x - y_\alpha\|^2. \) Then we have from the preceding result
\[ \frac{\partial \varphi}{\partial \alpha} \bigg|_{\alpha=0} = -2\|x - z^*\|^2 + 2 \langle (x - z^*), (x - y) \rangle = -2 \langle (y - z^*), (x - z^*) \rangle. \]

Therefore, if \( \langle (y - z^*), (x - z^*) \rangle > 0 \) for some \( y \in C, \) then
\[ \frac{\partial}{\partial \alpha} \left\{ \|x - y_\alpha\|^2 \right\} \bigg|_{\alpha=0} < 0 \]
and, for positive but small enough \( \alpha, \) we have \( \|x - z^*\|. \) This contradicts the fact that \( z^* \) is the projection of \( x \) onto \( C \) and shows that \( \langle (y - z^*), (x - z^*) \rangle \leq 0 \) for all \( y \in C. \]

\section*{2.2 Cones}

Certain subsets of \( \mathbb{R}^n \) (or of any vector space) occur with a frequency that merits singling out for separate study. They are \textit{cones}, and cones have crucial roles to play, for example, in the establishment of partial orderings and hence utility structures in economics, are central to the theory of necessary conditions in optimality problems, and have suprising roles to play in numerical analysis and the theory of probability.

\subsection*{2.2.1 Basic Definitions and Examples}

In this section we present some basic definitions and properties and show how cones essentially define preference structures. As usual, we start with some definitions.

\textbf{Definition 2.2.1} A set \( K \subset \mathbb{R}^n \) is called a cone with vertex \( x_o \) provided it is invariant under all maps of the form \( x \rightarrow x_o + \alpha (x - x_o) \) for any real \( \alpha > 0. \)
2.2. CONES

In most of the immediate applications, it suffices to select \( x_0 = 0 \) as the vertex, but that choice is not always useful as we shall see when we discuss necessary conditions. Nevertheless, we shall assume, unless otherwise stated, that 0 is the vertex. The implication of this assumption is then that \( K \) is a cone provided, for any \( x \in K \) and \( \alpha > 0 \), \( \alpha x \in K \).

If \( x \in K \) and \( x \neq 0 \), the set of points of the form \( \alpha x \) is called the open half-line; the set \( \{ \alpha x \in K \mid x \in K \setminus \{0\}, \alpha \geq 0 \} \) is then called the closed half-line.

Note that the cone may, or many not, contain its vertex. If \( 0 \in K \) then the cone is called pointed; otherwise it is called non-pointed. Now, it is trivially true that \( \{0\} \) is a pointed cone. Likewise the familiar set \( \mathbb{R}_\geq = \{ x \in \mathbb{R} \mid x > 0 \} \) is a cone which is not pointed, while \( \mathbb{R}_\geq \) is clearly a pointed cone. Likewise, the set \( Q = \mathbb{R}^2_\geq \cup \mathbb{R}^2_\leq \) is a cone with vertex 0 but it is not pointed. Of course a non-pointed cone \( C \) can be made pointed by adding the vertex. Thus, in this last example, \( \tilde{Q} = Q \cup \{0\} \) (or, equivalently \( \tilde{Q} = \mathbb{R}^2_\geq \cup \mathbb{R}^2_\leq \)) is a pointed cone.

**Example 2.2.2** We list here some further simple examples. Notice that the last entry involves a special class of square matrices\(^4\):

<table>
<thead>
<tr>
<th>(1) ( \mathbb{R}^n_\geq )</th>
<th>(2) ( \mathbb{R}^n_\geq )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3) ( \mathbb{R}^n )</td>
<td>(4) ( {0} )</td>
</tr>
<tr>
<td>(5) ( \emptyset )</td>
<td>(6) ( {x \mid Ax \leq 0, A \in \mathcal{M}_{mn}} )</td>
</tr>
<tr>
<td>(7) ( {x \in \mathbb{R}^2 \mid ax_1 + bx_1x_2 + cx_2^2 = 0} )</td>
<td>(8) ( S_+(n) \subset \mathcal{M}_{nn} ), the set of all positive semi-definite matrices.</td>
</tr>
</tbody>
</table>

In light of the examples (3) and (4), we will refer to a cone \( C \neq \emptyset \) as non-trivial provided \( C \neq \{0\} \) and \( C \neq \mathbb{R}^n \).

Let us check that (6) and (7) do, in fact, describe cones. To do this, we simply check

\(^4\)Recall that this is the set of all symmetric \( n \times n \) matrices with the property that \( \langle x, Ax \rangle \geq 0 \) for all \( x \in \mathbb{R}^n \).
the definition. In the case of (6), take any real \( \alpha > 0 \) and \( x \in \mathbb{R}^n \) with the property that \( Ax \leq 0 \). Then the result follows from the trivial observation that \( A(\alpha x) = \alpha A(x) \leq 0 \).

Likewise, if \((x_1, x_2)^T \in \mathbb{R}^2\) satisfies satisfies \( ax_1^2 + bx_1x_2 + cx_2^2 = 0 \) and if \( \alpha > 0 \), then \((\alpha x_1, \alpha x_2)^T\) satisfies

\[
a(\alpha x_1)^2 + b(\alpha x_1 \alpha x_2) + c(\alpha x_2)^2 = \alpha^2 [ax_1^2 + bx_1x_2 + cx_2^2] = \alpha 0 = 0.
\]

With regard to the set \( \tilde{Q} \) defined previously, there are two important differences between that set and the set \( \mathbb{R}^2_\geq \). The first is that \( \tilde{Q} \) contains an entire line, namely the line \( x_2 = x_1 \) while \( \mathbb{R}^2_\geq \) does not contain any such line. The second is that \( \mathbb{R}^n_\geq \) is a convex set while \( \tilde{Q} \) is not convex. Convex cones, that is cones which are also convex sets, are particularly important in applications. Certain properties of such cones, as we shall see, have profound influence in economic applications. One such property is that of being pointed. Another is described by the next definition.

**Definition 2.2.3** A convex cone is called line-free or proper provided it does not contain any line passing through 0.

We now list some simple propositions that are useful in various ways.

**Proposition 2.2.4** A pointed convex cone, \( C \), is line-free if and only if \( C \setminus \{0\} \) is convex.

**Proof:** If \( C \) contains a line through the origin, then \( C \setminus \{0\} \) is clearly not convex. Suppose, conversely, that \( C \) is line-free and let \( x, y \in C \setminus \{0\} \). Let \( \lambda \in [0,1] \). Then, since \( C \) is convex, \((1 - \lambda)x + \lambda y \in C \). Suppose that this line segment passes through 0. Then, for some \( \lambda_o \in (0,1) \), we have \((1 - \lambda_o)x + \lambda_o y = 0\) so that

\[
x = \left( \frac{-\lambda_o}{(1 - \lambda_o)} \right) y, \quad \text{and} \quad \frac{-\lambda_o}{(1 - \lambda_o)} < 0.
\]

Hence the cone \( C \) contains the line through 0 and \( x \) and so is not line-free, a contradiction \( \blacksquare \)

**Proposition 2.2.5** A subset \( C \subset \mathbb{R}^n \) is a convex cone if and only if \( C + C \subset C \) and \( \alpha C \subset C \) for all real \( \alpha > 0 \).
2.2. CONES

Proof: Exercise.

It should be clear that it is always possible to express any vector in $\mathbb{R}^2$ as the sum of a vector with all non-negative components and a vector all of whose components are negative. Looked at another way the cone $\mathbb{R}_+^2 \cup \mathbb{R}_-^2$ in fact generates all of $\mathbb{R}^2$ in the sense that $\mathbb{R}^2$ is the smallest vector space containing the union. In fact, this is a very general result as we will now see.

**Proposition 2.2.6** If $C \neq \emptyset$ is a convex cone in $\mathbb{R}^n$. Then $C - C$ is a vector subspace of $\mathbb{R}^n$ and, indeed, is the smallest vector space containing $C$.

Proof: If $V = C - C$ then $V \neq \emptyset$ since it contains the vector 0. We first show that $V$ is a vector subspace of $\mathbb{R}^n$ by showing that $V$ is algebraically closed with respect to scalar multiplication and addition. To show the first we show that $\lambda V \subset V$ for all $\lambda \neq 0$. Indeed, suppose first that $\lambda > 0$. Then $\lambda V \lambda C - \lambda C \subset C - C = V$. On the other hand, if $\lambda < 0$, set $\mu = -\lambda$ and then $\lambda V = -\mu C + \mu C \subset C - C = V$. So $V$ is invariant with respect to multiplication by scalars.

Likewise, $V + V = (C - C) + (C - C) = 2C - 2C \subset C - C = V$. So $V$ is invariant with respect to addition. This shows that $V$ is a vector subspace.

Finally, if $W$ is a subspace of $V$ and if $C \subset W$, then $V = C - C \subset W$. ■

Hence the smallest subspace containing $C$ is $C - C$ and is called the** vector space generated by $C$.** Now we ask about the largest subspace contained in $C$? For this, of course, we need to ensure that this largest vector space contains the zero element, hence we must add an hypothesis to $C$. It might seem odd that such a subspace should exist, so before starting, we give a simple example.

**Example 2.2.7** In $\mathbb{R}^2$ let $C$ be the closed upper half-plane $\{x \in \mathbb{R}^2 | x_2 \geq 0\}$ Then $-C = \{x \in \mathbb{R}^2 | x_2 \leq 0\}$. Clearly, $C$ is a cone in $\mathbb{R}^2$ and $C \cap (-C) = \{x \in \mathbb{R}^2 | x_2 = 0\}$, the real $x$-axis which is a two dimensional vector space contained in $C$.

Now we have the proposition.

**Proposition 2.2.8** If $C$ is a pointed convex cone, then the largest vector subspace contained in $C$ is the set $C \cap (-C)$.\(^{5}\)

\(^{5}\)Note that if $C$ is pointed, then $C \neq \emptyset$. 
Proof: If $W = C \cap (-C)$ then $W \neq \emptyset$ and $\lambda W = W$ for all $\lambda \neq 0$. Furthermore, since $W + W = W$,

$$W + W = (C + C) \cap -(C + C) \subset C \cap (-C) = W.$$ 

Hence $W$ is a subspace of $\mathbb{R}^n$. Finally, let $V$ be a vector subspace of $C$. Then $V \subset C \cap (-C) = W$. □

Corollary 2.2.9 A pointed convex cone is line-free if and only if $C \cap (-C) = \{0\}$.

Proof: Clearly, if $C \cap (-C) = \{0\}$ then $C$ cannot contain a line which would be a subspace larger than $\{0\}$. On the other hand, if $C$ is line-free then the only subspace contained in $C$ is the trivial one. □

As a simple exercise that will help to focus these ideas the reader should try

Exercise 2.2.10 Show that the set $\mathbb{R}^2_+ \cup \{(x, y) \in \mathbb{R}^2 \mid x < 0, y \geq -x\}$ is a cone, is not convex, and is not line-free.

Since linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ preserve scalar multiplication and addition, the next result and its corollary are not surprising. We leave both proposition and corollary as exercises. Note that the corollary depends on the result of Proposition 2.2.5.

Proposition 2.2.11 Suppose $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is linear and that $C \subset \mathbb{R}^n$ is a cone. Then the set $T(C)$ is a cone in $\mathbb{R}^m$.

Corollary 2.2.12 If, in addition, the cone $C$ is convex, then so is $T(C)$.

2.2.2 Conical Hulls

It is easy to check that if $C_1$ and $C_2$ are convex cones in $\mathbb{R}^n$ which are both pointed, then $C_1 \cap C_2$ is also a pointed cone. In fact, just as with convex sets, more is true. The proof is similar to the corresponding result for convex sets and we leave it as an exercise.
Proposition 2.2.13 Let $A$ be an arbitrary index set and for each $\alpha \in A$ let $C_\alpha$ be a pointed convex cone with vertex zero. Then

$$\bigcap_{\alpha \in A} C_\alpha$$

is also a convex cone with vertex 0.

Proof: Exercise.

You will recall that the corresponding result for convex sets leads to the introduction of the convex hull of an arbitrary set. We have a parallel construction here for cones. Now, suppose that $A \subset \mathbb{R}^n$ is an arbitrary set. Then the entire space $\mathbb{R}^n$ is a cone containing the set $A$. This means that the family of all convex cones containing the set $A$ is a non-empty family. The Proposition 2.2.13 now leads to the definition.

Definition 2.2.14 Let $A \subset \mathbb{R}^n$ be arbitrary and let $\{C_\alpha\}_{\alpha \in A}$ be the family of all convex cones containing $A$. Then $\bigcap_{\alpha \in A} C_\alpha$ is called the convex cone generated by $A$ or the conical hull of $A$. It will be denoted by $\text{con}(A)$.

It should be clear that the cone generated by $A$ is the smallest convex cone containing the set $A$ hence the name conical hull. The cone which is the conical hull of a finite set of vectors is said to be finitely generated. More generally we can characterize the conical hull of a set in the following way.

Proposition 2.2.15 Let $\{C_\alpha\}_{\alpha \in A}$ be a family of convex cones in $\mathbb{R}^n$. Then the convex cone generated by the union of the $C_\alpha$ is identical with the set of points of the form $\sum_{i \in I} x_i$, where $I$ is any finite subset of $A$ and $x_i \in C_i$, for all $i \in I$.

Proof: It is clear that set set of all such points, C, forms a cone. It is also clearly convex since

$$(1 - \lambda) \sum_{i \in I} x_i + \lambda \sum_{\omega \in \Omega} y_\omega = \sum_{i \in I} (1 - \lambda) x_i + \sum_{\omega \in \Omega} \lambda y_\omega,$$

which, when combined, is, again, a finite sum of elements of the respective convex cones and so contains the union of the $C_\alpha$. Finally, it is clear that it is contained in any convex cone that contains the union. \qed

As a corollary of this result, we have
Corollary 2.2.16 For any subset $A \subset \mathbb{R}^n$, the convex cone generated by $A$ is identical with the set of finite linear combinations of the form $\sum_{i \in I} \lambda_i \mathbf{x}_i$ where $\{\mathbf{x}_i\}_{i \in I}$ is any finite non-empty family of points of $A$ and where $\lambda_i > 0$ for all $i \in I$. Moreover, the $\mathbf{x}_i$ may be chosen to be linearly independent and so, in $\mathbb{R}^n$, the sum can be written $\sum_{i=1}^{k} \lambda_i \mathbf{x}_i$ with $k \leq n$.

Proof: The result follows from the observation that, if a convex cone contains some $\mathbf{x} \neq 0$ then it contains that half-line $C_{\mathbf{x}}$ of points $\lambda \mathbf{x}$ where $\lambda$ varies in $\mathbb{R}_>$. The sets $C_{\mathbf{x}}$ are clearly convex cones and we may now apply the proposition to the indexed family $\{C_{\mathbf{x}}\}$.

To see that the vectors in the representation

$$\mathbf{x} = \lambda_1 \mathbf{x}_1 + \cdots + \lambda_r \mathbf{x}_r$$

may be chosen to be linearly independent, assume that they are not. Then there exist constants, $\mu_1, \ldots, \mu_r$ not all zero, such that

$$\mu_1 \mathbf{x}_1 + \cdots + \mu_r \mathbf{x}_r = 0.$$ 

Without loss of generality, we may assume that some of the numbers $\mu_1, \ldots, \mu_r$ are positive (otherwise we just replace all the numbers with their negatives). Denote by $I$ the set of those indices $i$ with $1 \leq i \leq r$ for which $\mu_i > 0$, and set

$$\beta = \min_{i \in I} \frac{\lambda_i}{\mu_i}.$$ 

Further, let

$$\hat{\lambda}_i = \lambda_i - \beta \mu_i.$$ 

Then all the numbers $\hat{\lambda}_i$ are non-negative, and at least one of them is zero. On the other hand

$$\sum_{i=1}^{r} \hat{\lambda}_i \mathbf{x}_i = \sum_{i=1}^{r} \lambda_i \mathbf{x}_i - \sum_{i=1}^{r} \mu_i \mathbf{x}_i = \beta \sum_{i=1}^{r} \mu_i \mathbf{x}_i = \sum_{i=1}^{r} \lambda_i \mathbf{x}_i = \mathbf{x}.$$ 

Thus we have represented $\mathbf{x}$ in the form of the sum of no more than $n - 1$ non-zero terms.

Finally we have another characterization of the conical hull in the case that the set $A$ is, itself, convex.
Theorem 2.2.17 If the set \( A \subset \mathbb{R}^n \) is convex, then the convex cone generated by \( A \) is identical with \( C = \bigcup_{\lambda>0} \lambda A \).

**Proof:** It is clear that the set \( C \) is a cone. Moreover, it is convex. To see this, let \( \mu > 0, \nu > 0 \) and \( x, y \in A \). Take \( \lambda > 0 \). Then if we define \( z \in \mathbb{R}^n \) by

\[
z = \left( \frac{(1-\lambda)\mu}{(1-\lambda)\mu + \lambda\nu} \right) x + \left( \frac{\lambda\nu}{(1-\lambda)\mu + \lambda\nu} \right) y,
\]

then

\[
(1-\lambda)\mu x + \lambda\nu y = ((1-\lambda)\mu + \lambda\nu) z,
\]

Since \( A \) is convex, \( z \in A \) and \( (1-\lambda)\mu + \lambda\nu > 0 \), hence \( (1-\lambda)\mu x + \lambda\nu y \in C \).

This shows that \( C \subset \text{con} (A) \). On the other hand, since \( \text{con}(A) \) contains all sets of the form \( \lambda A, \lambda > 0 \), \( \text{con}(A) \subset C \) and so the sets are identical. ■

**Exercise 2.2.18** Show that if \( A \) is a convex set and \( 0 \not\in A \) then the cone, \( C \), generated by \( A \) is not pointed and that \( C \cup \{0\} \) is line-free.

### 2.2.3 Cones and Preferences: a relationship

We pause here to point out, explicitly, the relationship between cones and partial orderings. This relationship was mentioned earlier. It shows the relevance of cones to the introduction of preference relations in, for example, models of exchange economies where points in \( \mathbb{R}^n \) are interpreted as bundles of consumer goods. Moreover, such relations lie at the foundation of the study of Pareto optimality which we will discuss later.

Let us recall that a partial ordering on a set \( A \) is a reflexive, transitive, and antisymmetric relation on the product set \( A \times A \). In what follows we will denote a partial ordering by the symbol \( \prec \). The interesting and important fact is that specifying a cone automatically specifies a partial order provided that the cone has certain properties. In this context, it is easy to introduce a binary operation. Indeed, given a cone \( C \subset \mathbb{R}^n \) with vertex \( 0 \), we may define a binary relation \( \prec \) on \( \mathbb{R}^n \) by

\[
x \prec y \quad \text{provided} \quad y - x \in C.
\] (2.2)
CHAPTER 2. CONVEX SETS

With this definition $\prec$ we can easily check that this binary relation is a partial ordering of the vector space $\mathbb{R}^n$ provided $C$ is convex, contains the origin, and is line free. Each of these properties implies one of the properties that will make $\prec$ a partial order.

(a) If $0 \in C$, then $\prec$ is reflexive. This follows from the observation that for any $x \in \mathbb{R}^n$, $x - x = 0 \in C$ which implies that $x \prec x$.

(b) If $C$ is convex then $\prec$ is transitive, for if $x, y, z \in \mathbb{R}^n$, and if $x \prec y$ and $y \prec z$ then $y - x \in C$ and $z - y \in C$. Since $C$ is convex,

$$\frac{1}{2}(y - x) + \frac{1}{2}(z - y) \in C,$$

and so $\frac{1}{2}(z - x) \in C$ from which it follows that $z - x \in C$. Hence $x \prec z$.

(c) If $C$ is line-free, then $\prec$ is antisymmetric. Indeed, if $x \prec y$ and $y \prec x$ then $y - x \in C \cap (-C) = \{0\}$ so that $x = y$.

To summarize, these three observations show that the following theorem is true.

**Theorem 2.2.19** If $C$ is a line-free, convex cone with $0 \in C$, then the binary relation $\prec$ defined by

$$x \prec y \quad \text{if and only if} \quad y - x \in C,$$

defines a partial order on the vector space $\mathbb{R}^n$.

There is a partial converse of this theorem. If $\prec$ is a partial order on $\mathbb{R}^n$ which respects the operations, i.e. $x \prec y \Rightarrow x + z \prec y + z$ and $\lambda x \prec \lambda y$ for all $x, y, z \in \mathbb{R}^n$ and $\lambda > 0$, then $C := \{x \in \mathbb{R}^n | 0 \prec x\}$ is a line-free, convex cone, and contains 0. This is easily proven by arguments similar to those above.

### 2.2.4 More on Cones

We will begin this section with two important definitions.

**Definition 2.2.20** Given an non-empty set $C$, the **polar cone** of $C$, $C^*$, is defined by

$$C^* = \{y \in \mathbb{R}^n | \langle y, x \rangle \leq 0, \quad \text{for all } x \in C\}$$
Note that, whether $C$ is a cone or not, the set $C^\star$ is a cone. Below, we will check that this cone is both closed and convex. But first, we introduce another type of cone that will play a significant role in what follows.

**Definition 2.2.21** A cone $C$ is said to be **polyhedral** provided it has the form

$$ C = \{ x \in \mathbb{R}^n \mid \langle a_j, x \rangle \leq 0, j = 1, \ldots, r \}, $$

where $a_1, a_2, \ldots, a_r$ are a finite set of vectors in $\mathbb{R}^n$.

Observe that both these cones are defined as the intersection of a family of half-spaces. Specifically, for $C^\star$ we may define for each $x \in C$ the set $C^\star_x = \{ y \in \mathbb{R}^n \mid \langle y, x \rangle \}$. This is a half-space since the defining inequality is $\sum_{i=1}^n x_i y_i \leq 0$ for fixed $x$. Then clearly $C^\star = \bigcap_{x \in C} C^\star_x$. Since each half-space is a closed convex set, we see that $C^\star$ is always closed and convex regardless of the set $C$. These same remarks can be applied to the polyhedral cone of Definition 2.2.21.

We also recall the following definition (see the paragraph following Definition 2.2.14).

**Definition 2.2.22** A cone $C$ is said to be **finitely generated** provided it has the form

$$ C = \left\{ x \in \mathbb{R}^n \mid x = \sum_{j=1}^r \mu_j a_j \mu_j > 0, j = 1, 2, \ldots, r \right\}, $$

where $a_1, a_2, \ldots, a_r$ are a finite set of vectors in $\mathbb{R}^n$.

It can be easily checked that finitely generated cones are convex but it is not evident that they are closed. To prove that they are is one of our next tasks. Another important job is to show that there is a close connection between polyhedral cones and finitely generated cones; both are defined in terms of a finite set of vectors and, as it turns out, each is an instance of the other. This result is one of the basic structure theorems in the theory of cones.

We start our investigation by proving a result which is often called the **Polar Cone Theorem**.

**Theorem 2.2.23** For any non-empty, closed, pointed convex cone $C$, we have $(C^\star)^\star = C$. 


**Proof:** First we show that \( C \subset (C^*)^* \). Indeed, if \( x \in C \) then, for all \( y \in C^* \), \( \langle x, y \rangle \leq 0 \) and so, by definition of polars, \( x \in (C^*)^* \). Thus we have the result that \( C \subset (C^*)^* \).

For the reverse inclusion, we will use the Projection Theorem 2.1.22. Start by choosing any \( z \in (C^*)^* \). By hypothesis the cone \( C \) is closed and we let \( \hat{z} \) be the projection of \( z \) onto \( C \). Not only does the Projection Theorem guarantee the existence of a unique projection but it also implies that

\[
\langle (z - \hat{z}), (x - \hat{z}) \rangle \leq 0, \text{ for all } x \in C. \tag{2.3}
\]

By taking \( x = 0 \) this inequality becomes: (a) \( \langle (z - \hat{z}), -\hat{z} \rangle \leq 0 \) or \( \langle (z - \hat{z}), \hat{z} \rangle \geq 0 \); and by taking \( x = 2\hat{z} \) the inequality becomes: (b) \( \langle (z - \hat{z}), \hat{z} \rangle \leq 0 \). Then, combining (a) and (b) we may conclude that

\[
\langle (z - \hat{z}), \hat{z} \rangle = 0.
\]

Now this result, combined with that earlier inequality 2.3 yields

\[
\langle (z - \hat{z}), (x - \hat{z}) \rangle = \langle (z - \hat{z}), x \rangle - \langle (z - \hat{z}), \hat{z} \rangle,
\]

so that

\[
\langle (z - \hat{z}), x \rangle \leq 0, \text{ for all } x \in C.
\]

Considering the definition of the polar cone, this last inequality implies that \( (z - \hat{z}) \in C^* \) and, since \( z \in (C^*)^* \) this means that \( \langle (z - \hat{z}), z \rangle \leq 0 \), which, when added to \( \langle (z - \hat{z}), \hat{z} \rangle = 0 \) yields

\[
\langle (z - \hat{z}), z \rangle + \langle (z - \hat{z}), \hat{z} \rangle = \langle (z - \hat{z}), (z - \hat{z}) \rangle = \|z - \hat{z}\|^2 \leq 0
\]

and it follows that \( z = \hat{z} \) so that \( z \in C \). This shows that \( (C^*)^* \subset C \) and hence, together with the first part of the proof, that \( C = (C^*)^* \). □

We can now answer one of the questions we asked above: is a finitely generated cone closed? The answer is “yes” although it takes a little work to prove that it is. The proof uses induction and also identifies the polar cone.

**Lemma 2.2.24** Let \( \{a_1, a_2, \ldots, a_r\} \) be a finite set of vectors and let \( F \) be the cone generated by this finite set of vectors. (\( F \) is thus a finitely generated cone.) Then \( F \) is closed and its polar cone \( F^* \) is the polyhedral cone given by

\[
F^* = \{x | \langle x, a_j \rangle \leq 0, j = 1, 2, \ldots, r \}.
\]
2.2. CONES

Proof: First let us establish the result concerning the polar cone. To do this is simply a matter of applying the definitions. Indeed

\[ F^\ast = \{ y \in \mathbb{R}^n \mid \langle x, y \rangle \leq 0, \text{ for all } x \in F \} = \{ y \in \mathbb{R}^m \mid \sum_{j=1}^{r} \mu_j \langle a_j, y \rangle \leq 0, \text{ for all } \mu_j \geq 0 \} \]

\[ = \{ y \in \mathbb{R}^n \mid \langle a_j, y \rangle \leq 0, j = 1, \ldots, r \} . \]

It remains to show that the finitely generated cone \( F \) is closed. For this, we proceed by induction on \( r \), the number of generators. For \( r = 1 \) the cone \( F \) is just a closed half-line. Suppose that the conclusion is true for \( r = \ell \). Thus we assume the \( \text{con}F_\ell = \text{con}(\{a_1, \ldots, a_\ell\}) \) is closed and show that the cone \( F_{\ell+1} = \text{con}(\{a_1, \ldots, a_{\ell+1}\}) \) is closed. Without loss of generality, we assume that, for all indices \( j \), \( \|a_j\| = 1 \). Now let \( x_o \) be any limit point of the set \( F_{\ell+1} \) and let \( \{x^k\} \) be a subsequence in \( F_{\ell+1} \) that converges to \( x_o \). We will have proved the result if we can show that \( x_o \in F_{\ell+1} \).

Now, write \( x^k \) as a sum \( x^k = y^k + \lambda_k a_{\ell+1} \) where \( y^k \in F_\ell \) and \( \lambda_k \geq 0 \). Clearly, the sequence \( \{\lambda_k\} \) is a bounded sequence and we may assume, taking a subsequence if necessary and renumbering, that it converges to some \( \lambda_o \geq 0 \). Now, rewrite the expression for \( x^k \):

\[ x^k = y^k + \lambda_o a_{\ell+1} + (\lambda_k - \lambda_o) a_{\ell+1} . \]

Since the \( x^k \) converge to \( x_o \) and the term \( \lambda_k - \lambda_o \to 0 \) as \( n \to \infty \), the \( y^k \) must converge to some vector \( y_o \). Since the \( y^k \in F_\ell \) and the induction hypothesis is that this latter set is closed, we have \( y_o \in F_\ell \). Hence

\[ x_o = y_o + \lambda_o a_{\ell+1} \text{ with } y_o \in F_\ell, \lambda_o \geq 0 . \]

Hence \( x_o \in F_{\ell+1} \) and the proof is complete. □

This result shows another relationship between finitely generated cones and polyhedral cones; finitely generated cones have polyhedral cones as polars.

The next result is one form of a famous theorem due to Farkas and Minkowski. We will discuss it from a different point of view in a later chapter where we also will see some of its applications to mathematical economics. As with any major result, the theorem has several different interpretations and several different proofs. With the machinery we have developed so far, we can give a quick proof of this particular version using the Polar Cone Theorem.
Theorem 2.2.25 (Farkas-Minkowski Lemma) Let \(x, e_1, \ldots, e_m, a_1, \ldots, a_r\) be vectors in \(\mathbb{R}^n\). Then for all vectors \(y \in \mathbb{R}^n\) such that
\[
\langle y, e_i \rangle = 0, \quad i = 1, \ldots, m, \quad \text{and} \quad \langle y, a_j \rangle \leq 0, \quad j = 1, \ldots, r,
\]
we have
\[
\langle x, y \rangle \leq 0
\]
if and only if \(x\) can be written in the form
\[
x = \sum_{i=1}^m \lambda_i e_i + \sum_{j=1}^r \mu_j a_j, \quad \text{where} \quad \lambda_i, \mu_i \in \mathbb{R}, \quad \mu_j \geq 0.
\]

Proof: Notice that, since \(\lambda_i\) has no sign restriction, we can decompose it into its non-negative and negative parts\(^6\) \(\lambda_i = \lambda_i^+ - \lambda_i^-\) and then we can rewrite \(C\) as
\[
C = \left\{ x \in \mathbb{R}^n \mid x = \sum_{i=1}^m \lambda_i^+ e_i + \sum_{i=1}^m \lambda_i^- (-e_i) + \sum_{j=1}^r \mu_j a_j, \mu_j, \lambda_i^+, \lambda_i^- \geq 0 \right\}
\]
\[= \text{con} \left( \{e_1, -e_1, \ldots, e_m, -e_m, a_1, \ldots, a_r\} \right) \quad \text{by Corollary 2.2.16.}
\]
and
\[
P = \{ y \in \mathbb{R}^m \mid \langle y, b_j \rangle \leq 0, \quad j = 1, \ldots, r + 2m \},
\]
\[\{b_j\}_{j=1}^{r+2m} = \{ e_1, -e_1, \ldots, e_m, -e_m, a_1, \ldots, a_r \}
\]
Then the statement of the Farkas-Minkowski Lemma becomes:
\[
x \in C \text{ if and only if } x \in P^*.
\]
Since, by the previous proposition, \(P = C^*\) and \(C\) is closed, we have by the Polar Cone Theorem (2.2.23), that \(P^* = (C^*)^* = C\). \(\blacksquare\)

Remark: One useful way of thinking about this particular result as well as other forms of the Farkas-Minkowski Lemma is as an Alternative Theorem. Such theorems state that either one type of relation is true, or another (the “alternative”) is true, but not both. Perhaps the most famous such theorem is called the Fredholm Alternative Theorem which we will discuss later. Here is a restatement of the theorem as an alternative theorem.

\(^6\)To do this, set \(\lambda_i^+ = \max\{\lambda_i(x), 0\}\) and \(\lambda_i^- = \lambda_i^+ - \lambda\).
Corollary 2.2.26 Let \( x, e_1, \ldots, e_m, \) and \( a_1, \ldots, a_r \) be vectors in \( \mathbb{R}^n \). Then for all vectors \( y \in \mathbb{R}^n \) such that
\[
\langle y, e_i \rangle = 0, i = 1, \ldots, m, \text{ and } \langle y, a_j \rangle \leq 0, j = 1, \ldots, r,
\]
we have
\[
\langle x, y \rangle \geq 0
\]
or \( x \) can be written in the form
\[
x = \sum_{i=1}^{m} \lambda_i e_i + \sum_{j=1}^{r} \mu_j a_j, \text{ where } \lambda_i, \mu_i \in \mathbb{R}, \mu_j \geq 0.
\]
but not both.

We will find theorems written in the form of alternatives are particularly useful in applications.

The next result is, in a way, analogous of the basis theorem of linear algebra that the set of all linear combinations of a finite set of vectors is a vector subspace and that every finite dimensional vector subspace can be described as the set of all linear combintaion of a finite set of linearly independent vectors called the basis. The result here is that the cone of all non-negative linear combinations of a finite number of vectors is a polyhedral cone. This is the well-known theorem of Minkowski and Weyl. Both names are attached to this statement. In fact it is a combination of two theorems, the first, due to Minkowski, states that polyhedral cones are finitely generated, while Weyl’s result is the converse. Here is the statement and proof.

Proposition 2.2.27 (Minkowski, Weyl) A cone is polyhedral if and only if it is finitely generated.

Proof: We first show\(^7\) that if a cone \( C \subset \mathbb{R}^n \) is finitely generated then it is polyhedral. To this end, consider the finitely generated cone
\[
C = \{ x \in \mathbb{R}^n | x = \sum_{j=1}^{r} \mu_j a_j, \mu_j \geq 0, j = 1, \ldots, r \}.
\]
Without loss of generality we may assume that \( a_1 = 0 \) (if not, just add a new \( a_1 = 0 \) and renumber the others). The proof now proceeds by induction on the number of generators \( r \).

For \( r = 1 \), \( C = \{ 0 \} \) which is polyhedral since it can be expressed as the set
\[
\{ x \in \mathbb{R}^n | \langle e_i, x \rangle \leq 0, \langle -e_i, x \rangle \leq 0, i = 1, \ldots, n \},
\]
\(^7\)This part of the proof is due to Wets[?].
where the $e_i$ are the standard basis vectors in $\mathbb{R}^n$. Now, suppose that, for some $r \geq 2$, the set
\[
C_{r-1} = \{ x \in \mathbb{R}^n \mid \sum_{i=1}^{r-1} \mu_i a_i, \mu_i \geq 0, i = 1, \ldots, (r-1) \},
\]
has a polyhedral representation
\[
P_{r-1} = \{ x \in \mathbb{R}^n \mid \langle b_j, x \rangle \leq 0, j = 1, \ldots, m \}.
\]

We will complete the induction by showing that the finitely generated cone
\[
P_r = \{ x \in \mathbb{R}^n \mid \langle b_j, x \rangle \leq 0, j = 1, \ldots, m \}
\]
has the particular polyhedral representation
\[
P_r = \{ x \in \mathbb{R}^n \mid \langle b_j, x \rangle \leq 0, j \in J^- \cup J^+, \langle b_{k,\ell}, x \rangle \leq 0, \ell \in J^+, k \in J^- \}.
\]

First, we show that $C_r \subset P_r$. Notice that, in order to show that an $x \in C_r$ lies in $P_r$, it suffices (since all the $\mu_j \geq 0$) to show that the inequalities defining $P_r$ are satisfied by the generators of $C_r$. Now, it is clear from the induction hypothesis that, for all $j = 1, 2, \ldots, (r-1)$ we have $\langle b_j, a_j \rangle \leq 0$ while for $a_r$, we have $\beta_k = \langle b_k, a_r \rangle \leq 0, k \in J^0 \cup J^-$. Hence the first set of inequalities defining $P_r$ is satisfied. To check the second set, $\langle b_{k,\ell}, a_r \rangle \leq 0$ for $\ell \in J^+, k \in J^-$ we compute using the definitions
\[
\langle b_{k,\ell}, a_r \rangle = \langle b_{k,\ell}, a_r \rangle - \frac{\beta_k}{\beta_{k,\ell}} \langle b_{k,\ell}, a_r \rangle = 0.
\]

Hence $C_r \subset P_r$.

To show the reverse inclusion, $P_r \subset C_r$, start with an $x \in P_r$. It suffices to show, by the induction hypothesis, that there is a $\mu_r \geq 0$ such that $x - \mu_r a_r \in P_{r-1}$. Such a $\mu_r$ will exist if and only if
\[
\langle b_j, x - \mu_r a_r \rangle = \langle b_j, x \rangle - \mu_r \langle b_j, a_r \rangle \leq 0.
\]

This means that $\mu_r \langle b_j, a_r \rangle \geq \langle b_j, x \rangle$. We then have
\[
\mu_r \begin{cases} 
\geq \frac{\langle b_j, x \rangle}{\langle b_j, a_r \rangle} = \frac{\langle b_j, x \rangle}{\beta_j} & \text{if } \langle b_j, a_r \rangle \geq 0 \\
\leq \frac{\langle b_j, x \rangle}{\langle b_j, a_r \rangle} = \frac{\langle b_j, x \rangle}{\beta_j} & \text{if } \langle b_j, a_r \rangle \leq 0
\end{cases}
\]

Now, define
\[
\gamma = \max \left\{ 0, \max_{j \in J^+} \frac{\langle b_j, x \rangle}{\beta_j} \right\}, \quad \text{and } \delta = \min_{j \in J^-} \frac{\langle b_j, x \rangle}{\beta_j}.
\]

Then, by the inequalities above, $\mu_r \geq \gamma$ and $\mu_r \leq \delta$.

Finally, since $x \in P_r$, we have
\[
0 \leq \frac{\langle b_k, x \rangle}{\beta_k}, \text{ for all } k \in J^-,
\]
2.2. CONES

as well as \( \langle b_{\ell,k}, x \rangle \leq 0 \) for all \( \ell \in J^+, k \in J^- \), or, equivalently

\[
\frac{\langle b_{\ell}, x \rangle}{\beta_{\ell}} \leq \frac{\langle b_{k}, x \rangle}{\beta_{k}} \quad \text{for all} \quad \ell \in J^+, k \in J^-.
\]

Hence \( \gamma \leq \delta \) and so \( C_r \supset P_r \) which was to be proved. So every finitely generated cone is polyhedral.

For the second part of the proof, the statement that every polyhedral cone is finitely generated, we will use Lemma 2.2.24 and the Polar Cone Theorem. Consider the typical polyhedral set

\[
C = \{ x \in \mathbb{R}^n | \langle x, a_j \rangle \leq 0, j = 1, \ldots, r \},
\]

where the \( a_j \in \mathbb{R}^n \) are given vectors. Next, using the same set of vectors, form the finitely generated cone

\[
\hat{C} = \{ x \in \mathbb{R}^n | x = \sum_{i=1}^{r} \mu_j a_j, \mu_j \geq 0, j = 1, \ldots, r \}.
\]

Then, by definition of the polar cone, \( \hat{C}^* = C \), and, by the Polar Cone Theorem, \( \hat{C} = (\hat{C}^*)^* = C^* \). Hence the polar of any polyhedral cone is finitely generated.

It remains to show that \( C \) is the polar of a polyhedral cone. But this follows from the fact that \( C = \hat{C} \) and the observation that, from the first part of the proof, \( \hat{C} \), being finitely generated, is also polyhedral.

To conclude this section, we use this last result to give a basic structural result for polyhedral sets with are not necessarily cones. These sets often arise as sets of feasible points, i.e., sets of points which satisfy all the given constraints, in constrained optimization problems, particularly in linear and quadratic programming. First, we need to decide what a polyhedral set is.

**Definition 2.2.28** A non-empty subset \( P \) of \( \mathbb{R}^n \) is said to be a polyhedral set, or simply a polyhedron provided it is of the form

\[
P = \{ x \in \mathbb{R}^n | \langle a_j, x \rangle \leq b_j, j = 1, \ldots, r \},
\]

where, for all indices \( j \), \( a_j \in \mathbb{R}^n \) and \( j \in \mathbb{R} \).

Again, as with the Minkowski-Weyl theorem above, this last result, variously called the Minkowski-Farkas-Weyl Theorem or the Resolution Theorem, is analogous to the result on bases in a vector space. While in the first of these two theorems, we are concerned with all non-negative linear combinations of a finite set of vectors, here we are interested in convex combinations of a finite number of vectors. A careful reading reveals that it says that every convex combination of finitely many vectors is a polyhedron and every polyhedron, provided it is bounded, can be expressed as the convex hull of a finite set of vectors.
Theorem 2.2.29 A set $P \subset \mathbb{R}^n$ is a polyhedron if and only if there exist a non-empty and finite set of vectors $\{v^1, v^2, \ldots, v^m\}$ and a finitely generated cone $C$ such that

$$P = \left\{ x \in \mathbb{R}^n \mid x = y + \sum_{j=1}^{m} \mu_j v^j, \ y \in C, \mu_j \geq 0, \sum_{j=1}^{m} \mu_j = 1 \right\}.$$ 

Proof: Assume that $P$ is polyhedral. Then it has the form

$$P = \{ x \in \mathbb{R}^n \mid (a_j, x) \leq b_j, \ j = 1, \ldots, r \}$$

for some vectors $a_j$ and some scalars $b_j$. Consider the polyhedral cone in $\mathbb{R}^{n+1}$

$$\hat{P} = \{ (x, w) \mid 0 \leq w, (a_j, x) \leq b_j w, \ j = 1, \ldots, r \},$$

and note that $P = \{ x \in \mathbb{R}^n \mid (x, 1) \in \hat{P} \}$.

By the Minkowski-Weyl Theorem (Theorem 2.2.27) the polyhedral cone $\hat{P}$ is finitely generated, so it has the form

$$\hat{P} = \left\{ (x, w) \in \mathbb{R}^{n+1} \mid x = \sum_{j=1}^{m} \mu_j v^j + \sum_{j \in J^+} \mu_j d_j, \mu_j \geq 0, j = 1, \ldots, m \right\},$$

for some vectors $v^j$ and scalars $d_j$. Since $w \geq 0$ for all vectors $(x, w) \in \hat{P}$ we see that $d_j \geq 0$ for all indices $j$ for otherwise we could produce a negative $w$ with appropriate choice of the $\mu_j$. Now define the two sets of indices

$$J^+ = \{ j \mid d_j > 0 \}, \quad J^0 = \{ j \mid d_j = 0 \}.$$ 

By replacing $\mu_j$ by $\mu_j/d_j$ for all $j \in J^+$, we obtain the equivalent representation

$$\hat{P} = \left\{ (x, w) \in \mathbb{R}^{n+1} \mid x = \sum_{j=1}^{m} \mu_j v^j + \sum_{j \in J^+} \mu_j v_j, \mu_j \geq 0, j = 1, \ldots, m \right\}.$$ 

Now we observe that the set

$$K = \left\{ \sum_{j \in J^+} \mu_j v_j \mid \mu_j \geq 0, j \in J^0 \right\}$$

is a cone. So we may finally rewrite $P = \{ x \in \mathbb{R}^n \mid (x, 1) \in \hat{P} \}$ as

$$P = \left\{ x \in \mathbb{R}^n \mid x = \sum_{j \in J^+} \mu_j v_j + \sum_{j \in J^+} \mu_j v_j, \sum_{j \in J^+} \mu_j = 1, \mu_j \geq 0, j = 1, \ldots, m \right\}.$$ 

Thus we have written $P$ as the vector sum of the convex hull of the vectors $v_j, j \in J^+$, and of the finitely generated cone $K$.

Conversely, suppose that $P = Q + C$ where $Q = \{ v^1, \ldots, v^m \}$ and $Q = \text{con} \{ u^1, \ldots, u^\ell \}$. Then $x \in P$ if and only if $(x, 1)$ is in the cone generated by $\{ (v^1, 1), \ldots, (v^m, 1), (u^1, 0), \ldots, (u^\ell, 0) \}$. Again, invoking the Minkowski-Weyl Theorem, this cone is polyhedral. \[\blacksquare\]

Example 2.2.30 Here we give a simple example to illustrate this last theorem. Let $P = \{ bx \in \mathbb{R}^2 \mid x_1 + x_2 \leq 4 \}$. Certainly $P$ is a polyhedron. We take $Q = \left\{ \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\}$ and take

$$C = \text{con} \left\{ \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\}.$$
We want to show that $P = Q + C$. Now, every element of $Q + C$ has the form

$$\begin{pmatrix} 2 \\ 2 \end{pmatrix} + \mu_1 \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \mu_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \mu_3 \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 - \mu_1 + \mu_2 - \mu_3 \\ 2 + \mu_1 - \mu_2 - \mu_3 \end{pmatrix}, \quad \mu_i \geq 0.$$ 

Adding the two components we get $4 - 2\mu_3 \leq 4$ so this vector is in $P$. Hence $Q + C \subset P$.

To prove the reverse inclusion, suppose $(x_1, x_2)^\top \in P$. Then $(x_1, x_2)^\top \in Q + C$ provided

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} + \mu_1 \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \mu_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \mu_3 \begin{pmatrix} -1 \\ -1 \end{pmatrix},$$

or, as above,

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 - \mu_1 + \mu_2 - \mu_3 \\ 2 + \mu_1 - \mu_2 - \mu_3 \end{pmatrix}.$$ 

This implies that $4 - 2\mu_3 = x_1 + x_2 \leq 4$ which is true if and only if $-2\mu_3 \leq 0$ or $\mu_3 \geq 0$. So, in particular, if we choose $\mu_2 = \mu_1 = 0$. Then

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 - \mu_3 \\ 2 - \mu_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} + \mu_3 \begin{pmatrix} -1 \\ -1 \end{pmatrix},$$

and so is an element of $Q + C$. 