Part 1

Convex Analysis

Chapter 2

Convex Sets

The concept of convexity is of great importance in the study of optimization problems. Convex sets, polyhedral sets, and separation of disjoint convex sets are used frequently in the analysis of mathematical programming problems, the characterization of their optimal solutions, and the development of computational procedures.

The following is an outline of the chapter.

SECTION 2.1: **Convex Hulls** This section is elementary. It presents some examples of convex sets and defines convex hulls. Readers with previous knowledge of convex sets may skip this section (possibly with the exception of Carathéodory theorem).

SECTION 2.2: **Closure and Interior of a Convex Set** Some topological properties of convex sets, related to interior, boundary, and closure points, are discussed.

SECTION 2.3: **Separation and Support of Convex Sets** This section is important, since the notions of separation and support of convex sets are frequently used in optimization. A careful study of this section is recommended.

SECTION 2.4: **Convex Cones and Polarity** This is a short section mainly dealing with polar cones. This section may be skipped without loss of continuity.

SECTION 2.5: **Polyhedral Sets, Extreme Points, and Extreme Directions** This section treats the special important case of polyhedral sets. Characterization of extreme points and extreme directions of polyhedral sets is developed. Also, the representation of a polyhedral set in terms of its extreme points and extreme directions is proved.

SECTION 2.6: **Linear Programming and the Simplex Method** The well-known simplex method is developed as a natural extension of the material in the previous section. Readers who are familiar with the simplex method may skip this section.

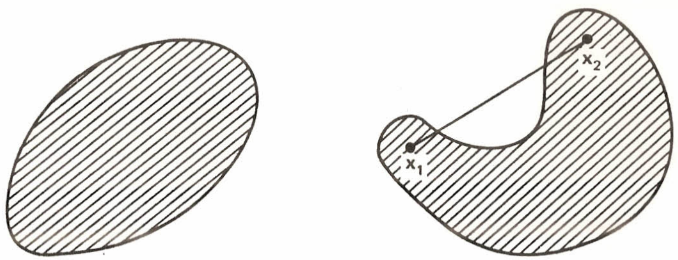
2.1 Convex Hulls

In this section we first introduce the notions of convex sets and convex hulls. We then demonstrate that any point in the convex hull of a set S can be represented in terms of $n+1$ points in the set S.

2.1.1 Definition

A nonempty set S in E_n is said to be *convex* if the *line segment* joining any two points of the set also belongs to the set. In other words, if $\mathbf{x}_1$ and $\mathbf{x}_2$ are in S, then $\lambda\mathbf{x}_1+(1-\lambda)\mathbf{x}_2$ must also belong to S for each $\lambda\in[0, 1]$. Points of the form $\lambda\mathbf{x}_1+(1-\lambda)\mathbf{x}_2$, where $\lambda\in[0, 1]$, are also referred to as *convex conbinations* of $\mathbf{x}_1$ and $\mathbf{x}_2$.

Figure 2.1 below illustrates the notion of a convex set. Note that in Figure 2.1*b*, the line segment joining $\mathbf{x}_1$ and $\mathbf{x}_2$ does not entirely lie in the set.



(*a*) Convex (*b*) Not convex

Figure 2.1 Illustration of convex sets.

The following are some examples of convex sets.

1. $S=\{(x_1, x_2, x_3): x_1+2x_2-x_3=4\}\subset E_3$

   This is an equation of a plane in E_3. In general, $S=\{\mathbf{x}:\mathbf{p}^t\mathbf{x}=\alpha\}$ is called a *hyperplane* in E_n, where $\mathbf{p}$ is a nonzero vector in E_n, usually referred to as the *normal* to the hyperplane, and α is a scalar.

2. $S=\{(x_1, x_2, x_3): x_1+2x_2-x_3\le 4\}\subset E_3$

   These are points on one side of the hyperplane defined above. These points form a *half space*. In general, a half space $S=\{\mathbf{x}:\mathbf{p}^t\mathbf{x}\le\alpha\}$ in E_n is a convex set.

3. $S=\{(x_1, x_2, x_3): x_1+2x_2-x_3\le 4, \quad 2x_1-x_2+x_3\le 6\}\subset E_3$

   This set is the intersection of two half spaces. In general, the set $S=\{\mathbf{x}:\mathbf{A}\mathbf{x}\le\mathbf{b}\}$ is a convex set, where $\mathbf{A}$ is an $m\times n$ matrix, and $\mathbf{b}$ is an m vector.

This set is the intersection of m half spaces and is usually called a *polyhedral set*.

4. $S=\{(x_1, x_2): x_2\ge|x_1|\}\subset E_2$

   This set represents a *convex cone* in E_2.

5. $S=\{(x_1, x_2): x_1^2+x_2^2\le 4\}\subset E_2$

   This set represents points on and inside a circle with center $(0, 0)$ and radius 2.

6. $S=\{\mathbf{x}:\mathbf{x}\,\text{solves}\,\textit{Problem P}\,\text{below}\}$

Problem P

$$\begin{aligned} \text{Minimize}\quad & \mathbf{c}^t\mathbf{x} \\ \text{subject to}\quad & \mathbf{A}\mathbf{x}=\mathbf{b} \\ & \mathbf{x}\ge\mathbf{0} \end{aligned}$$

Here $\mathbf{c}$ is an n vector, $\mathbf{b}$ is an m vector, $\mathbf{A}$ is an $m\times n$ metrix, and $\mathbf{x}$ is an n vector. The set S gives all optimal solutions of the *linear programming problem* of minimizing the linear function $\mathbf{c}^t\mathbf{x}$ over the polyhedral region defined by $\mathbf{A}\mathbf{x}=\mathbf{b}$ and $\mathbf{x}\ge\mathbf{0}$.

The following lemma is an immediate consequence of the definition of convexity. It states that the intersection of two convex sets is convex and that the algebraic sum of two convex sets is also convex. The proof is left as an exercise.

2.1.2 Lemma

Let S_1 and S_2 be convex sets in E_n. Then

1. $S_1\cap S_2$ is convex.
2. $S_1+S_2=\{\mathbf{x}_1+\mathbf{x}_2:\mathbf{x}_1\in S_1, \mathbf{x}_2\in S_2\}$ is convex.
3. $S_1-S_2=\{\mathbf{x}_1-\mathbf{x}_2:\mathbf{x}_1\in S_1, \mathbf{x}_2\in S_2\}$ is convex.

Convex Hulls

Given an arbitrary set S in E_n, different convex sets can be generated from S. In particular, we discuss below the convex hull of S.

2.1.3 Definition

Let S be an arbitrary set in E_n. The *convex hull* of S, denoted by $H(S)$, is the collection of all convex combinations of S. In other words $\mathbf{x}\in H(S)$ if and only if

$\mathbf{x}$ can be represented as

$$\mathbf{x} = \sum_{j=1}^{k} \lambda_j \mathbf{x}_j$$
$$\sum_{j=1}^{k} \lambda_j = 1$$
$$\lambda_j \geq 0, \qquad \text{for } j = 1, \ldots, k$$

where k is a positive integer and $\mathbf{x}_1, \ldots, \mathbf{x}_k \in S$.

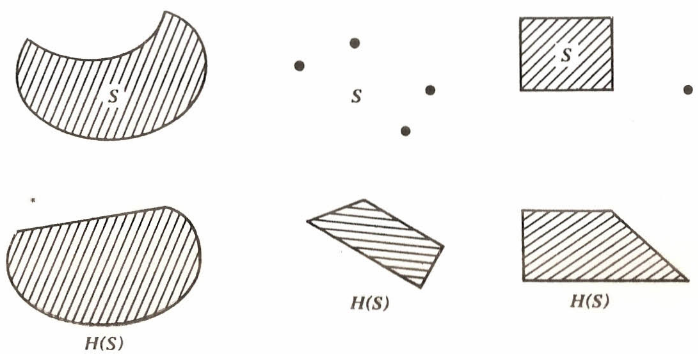


Figure 2.2 Examples of convex hulls.

Figure 2.2 shows some examples of convex hulls. Actually, we see that in each case, $H(S)$ is the minimal convex set that contains S. This is indeed the case in general, as given in Lemma 2.1.4. The proof is left as an exercise.

2.1.4 Lemma

Let S be an arbitrary set in E_n. Then $H(S)$ is the smallest convex set containing S. Indeed, $H(S)$ is the intersection of all convex sets containing S.

We have disucssed above the convex hull of an arbitrary set S. The convex hull of a finite number of points leads to the definitions of a polytope and a simplex.

2.1.5 Definition

The convex hull of a finite number of points $\mathbf{x}_1, \ldots, \mathbf{x}_{k+1}$ in E_n is called a *polytope*. If $\mathbf{x}_2 - \mathbf{x}_1, \mathbf{x}_3 - \mathbf{x}_1, \ldots, \mathbf{x}_{k+1} - \mathbf{x}_1$ are linearly independent, then $H(\mathbf{x}_1, \ldots, \mathbf{x}_{k+1})$, the convex hull of $\mathbf{x}_1, \ldots, \mathbf{x}_{k+1}$, is called a *simplex* with vertices $\mathbf{x}_1, \ldots, \mathbf{x}_{k+1}$.

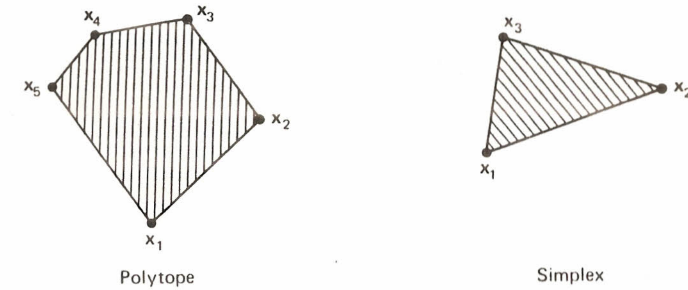


Figure 2.3 Examples of a polytope and a simplex.

Figure 2.3 shows examples of a polytope and a simplex in E_n. Note that the maximum number of linearly independent vectors in E_n is n, and hence, there could be no simplex in E_n with more than $n+1$ vertices.

Carathéodory Theorem

By definition, a point in the convex hull of a set can be represented as a convex combination of a finite number of points in the set. The following theorem shows that any point $\mathbf{x}$ in the convex hull of a set S can be represented as a convex conbination of, at most, $n+1$ points in S. The theorem is trivially true for $\mathbf{x} \in S$.

2.1.6 Theorem

Let S be an arbitrary set in E_n. If $\mathbf{x} \in H(S)$, then $\mathbf{x} \in H(\mathbf{x}_1, \ldots, \mathbf{x}_{n+1})$, where $\mathbf{x}_j \in S$ for $j = 1, \ldots, n+1$. In other words, $\mathbf{x}$ can be represented as

$$\mathbf{x} = \sum_{j=1}^{n+1} \lambda_j \mathbf{x}_j$$
$$\sum_{j=1}^{n+1} \lambda_j = 1$$
$$\lambda_j \geq 0, \qquad \text{for } j = 1, \ldots, n+1$$
$$\mathbf{x}_j \in S, \qquad \text{for } j = 1, \ldots, n+1$$

Proof

Since $\mathbf{x} \in H(S)$, then $\mathbf{x} = \sum_{j=1}^{k} \lambda_j \mathbf{x}_j$, where $\lambda_j > 0$ for $j = 1, \ldots, k$, $\mathbf{x}_j \in S$ for $j = 1, \ldots, k$, and $\sum_{j=1}^{k} \lambda_j = 1$. If $k \leq n+1$, the result is at hand. Now suppose that $k > n+1$. Note that $\mathbf{x}_2 - \mathbf{x}_1, \mathbf{x}_3 - \mathbf{x}_1, \ldots, \mathbf{x}_k - \mathbf{x}_1$ are linearly dependent. Thus there exist scalars $\mu_2, \mu_3, \ldots, \mu_k$ not all zero such that $\sum_{j=2}^{k} \mu_j(\mathbf{x}_j - \mathbf{x}_1) = \mathbf{0}$. Letting $\mu_1 = -\sum_{j=2}^{k} \mu_j$, it follows that $\sum_{j=1}^{k} \mu_j \mathbf{x}_j = \mathbf{0}$, $\sum_{j=1}^{k} \mu_j = 0$, and not all the μ_j's are equal to zero. Note that at least one $\mu_j > 0$. Then

$$\mathbf{x} = \sum_{j=1}^{k} \lambda_j \mathbf{x}_j + \mathbf{0} = \sum_{j=1}^{k} \lambda_j \mathbf{x}_j - \alpha \sum_{j=1}^{k} \mu_j \mathbf{x}_j = \sum_{j=1}^{k} (\lambda_j - \alpha\mu_j)\mathbf{x}_j$$

for any real α. Now choose α as follows:

$$\alpha = \underset{1 \le j \le k}{\text{minimum}} \left\{ \frac{\lambda_j}{\mu_j} : \mu_j > 0 \right\} = \frac{\lambda_i}{\mu_i} \qquad \text{for some } i \in \{1, \ldots, k\}$$

Note that $\alpha > 0$. If $\mu_j \le 0$, then $\lambda_j - \alpha\mu_j > 0$, and if $\mu_j > 0$, then $\lambda_j/\mu_j \ge \lambda_i/\mu_i = \alpha$, and hence $\lambda_j - \alpha\mu_j \ge 0$. In other words, $\lambda_j - \alpha\mu_j \ge 0$ for all $j = 1, \ldots, k$. In particular, $\lambda_i - \alpha\mu_i = 0$ by definition of α. Therefore, $\mathbf{x} = \sum_{j=1}^{k} (\lambda_j - \alpha\mu_j)\mathbf{x}_j$, where $\lambda_j - \alpha\mu_j \ge 0$ for $j = 1, \ldots, k$, $\sum_{j=1}^{k} (\lambda_j - \alpha\mu_j) = 1$, and furthermore, $\lambda_i - \alpha\mu_i = 0$. In other words, $\mathbf{x}$ is represented as a convex combination of, at most, $k-1$ points in S. The process is repeated until $\mathbf{x}$ is represented as a convex combination of $n+1$ points in S. This completes the proof.

2.2 Closure and Interior of a Convex Set

In this section we develop some topological properties of convex sets. Recall that given a point $\mathbf{x}$ in E_n, an ε-neighborhood around it is the set $N_\varepsilon(\mathbf{x}) = \{\mathbf{y} : \|\mathbf{y} - \mathbf{x}\| < \varepsilon\}$. Let us first review the definitions of closure, interior, and boundary of an arbitrary set in E_n.

2.2.1 Definition

Let S be an arbitrary set in E_n. A point $\mathbf{x}$ is said to be in the *closure* of S, denoted by cl S, if $S \cap N_\varepsilon(\mathbf{x}) \ne \varnothing$ for every $\varepsilon > 0$. If $S = \text{cl } S$, then S is called *closed*. $\mathbf{x}$ is said to be in the *interior* of S, denoted by int S, if $N_\varepsilon(\mathbf{x}) \subset S$ for some $\varepsilon > 0$. If $S = \text{int } S$, then S is called *open*. Finally, $\mathbf{x}$ is said to be in the *boundary* of S, denoted by ∂S, if $N_\varepsilon(\mathbf{x})$ contains at least one point in S and one point not in S for every $\varepsilon > 0$.

To illustrate, consider $S = \{(x_1, x_2) : x_1^2 + x_2^2 \le 1\}$, which represents all points within a circle with center $(0, 0)$ and radius 1. It can be easily verified that S is closed; that is, $S = \text{cl } S$. Furthermore, int S consists of all points inside the circle; that is, int $S = \{(x_1, x_2) : x_1^2 + x_2^2 < 1\}$. Finally, ∂S consists of points on the circle; that is, $\partial S = \{(x_1, x_2) : x_1^2 + x_2^2 = 1\}$.

Line Segment Between Closure and Interior Points

Given a convex set with a nonempty interior, the line segment (excluding the endpoints) joining a point in the interior of the set and a point in the closure of the set belongs to the interior of the set. This result is proved below.

2.2.2 Theorem

Let S be a convex set E_n with a nonempty interior. Let $\mathbf{x}_1 \in \text{cl } S$ and $\mathbf{x}_2 \in \text{int } S$. Then $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 \in \text{int } S$ for each $\lambda \in (0, 1)$.

Proof

Since $\mathbf{x}_2 \in \text{int } S$, there exists an $\varepsilon > 0$ such that $\{\mathbf{z} : \|\mathbf{z} - \mathbf{x}_2\| < \varepsilon\} \subset S$. Let $\mathbf{y}$ be such that

$$\mathbf{y} = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 \tag{2.1}$$

where $\lambda \in (0, 1)$. To prove that $\mathbf{y}$ belongs to int S, it suffices to construct a neighborhood about $\mathbf{y}$ that also belongs to S. In particular, we show that $\{\mathbf{z} : \|\mathbf{z} - \mathbf{y}\| < (1-\lambda)\varepsilon\} \subset S$. Let $\mathbf{z}$ be such that $\|\mathbf{z} - \mathbf{y}\| < (1-\lambda)\varepsilon$ and refer to Figure

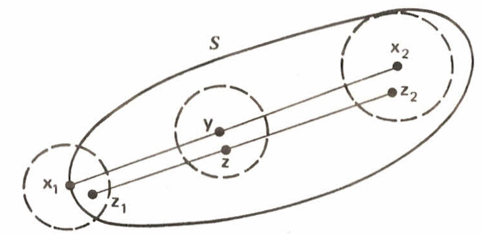


Figure 2.4 Line segment joining closure and interior points.

2.4. Since $\mathbf{x}_1 \in \text{cl } S$, then

$$\left\{ \mathbf{x} : \|\mathbf{x} - \mathbf{x}_1\| < \frac{(1-\lambda)\varepsilon - \|\mathbf{z} - \mathbf{y}\|}{\lambda} \right\} \cap S$$

is not empty. In particular, there exists a $\mathbf{z}_1 \in S$ such that

$$\|\mathbf{z}_1 - \mathbf{x}_1\| < \frac{(1-\lambda)\varepsilon - \|\mathbf{z} - \mathbf{y}\|}{\lambda} \tag{2.2}$$

Now let $\mathbf{z}_2 = \dfrac{\mathbf{z} - \lambda\mathbf{z}_1}{1-\lambda}$. From (2.1), the Schwartz inequality, and (2.2), we get

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

Therefore $\mathbf{z}_2 \in S$. By definition of $\mathbf{z}_2$, note that $\mathbf{z} = \lambda \mathbf{z}_1 + (1-\lambda)\mathbf{z}_2$, and since both $\mathbf{z}_1$ and $\mathbf{z}_2$ belong to S, then $\mathbf{z}$ also belongs to S. We showed that any $\mathbf{z}$ with $\|\mathbf{z}-\mathbf{y}\| < (1-\lambda)\varepsilon$ belongs to S. Therefore $\mathbf{y} \in \operatorname{int} S$ and the proof is complete.

Corollary 1

Let S be a convex set. Then $\operatorname{int} S$ is convex.

Corollary 2

Let S be a convex set with a nonempty interior. Then $\operatorname{cl} S$ is convex.

Proof

Let $\mathbf{x}_1, \mathbf{x}_2 \in \operatorname{cl} S$. Pick $\mathbf{z} \in \operatorname{int} S$ (by assumption $\operatorname{int} S \neq \varnothing$). By the theorem, $\lambda \mathbf{x}_2 + (1-\lambda)\mathbf{z} \in \operatorname{int} S$ for each $\lambda \in (0, 1)$. Now fix $\mu \in (0, 1)$. By the theorem, $\mu \mathbf{x}_1 + (1-\mu)[\lambda \mathbf{x}_2 + (1-\lambda)\mathbf{z}] \in \operatorname{int} S \subset S$ for each $\lambda \in (0, 1)$. If we take the limit as λ approaches 1, it follows that $\mu \mathbf{x}_1 + (1-\mu)\mathbf{x}_2 \in \operatorname{cl} S$, and the proof is complete.

Corollary 3

Let S be a convex set with a nonempty interior. Then, $\operatorname{cl}(\operatorname{int} S) = \operatorname{cl} S$.

Proof

Clearly $\operatorname{cl}(\operatorname{int} S) \subset \operatorname{cl} S$. Now let $\mathbf{x} \in \operatorname{cl} S$, and pick $\mathbf{y} \in \operatorname{int} S$ (by assumption $\operatorname{int} S \neq \varnothing$). Then $\lambda \mathbf{x} + (1-\lambda)\mathbf{y} \in \operatorname{int} S$ for each $\lambda \in (0, 1)$. Letting $\lambda \to 1$, it follows that $\mathbf{x} \in \operatorname{cl}(\operatorname{int} S)$.

Corollary 4

Let S be a convex set with a nonempty interior. Then $\operatorname{int}(\operatorname{cl} S) = \operatorname{int} S$.

Proof

Note that $\operatorname{int} S \subset \operatorname{int}(\operatorname{cl} S)$. Let $\mathbf{x}_1 \in \operatorname{int}(\operatorname{cl} S)$. We need to show that $\mathbf{x}_1 \in \operatorname{int} S$. There exists an $\varepsilon > 0$ such that $\|\mathbf{y} - \mathbf{x}_1\| < \varepsilon$ implies that $\mathbf{y} \in \operatorname{cl} S$. Now let $\mathbf{x}_2 \neq \mathbf{x}_1$ belong to $\operatorname{int} S$ and let $\mathbf{y} = (1+\Delta)\mathbf{x}_1 - \Delta \mathbf{x}_2$, where $\Delta = \dfrac{\varepsilon}{2\,\|\mathbf{x}_1 - \mathbf{x}_2\|}$. Since $\|\mathbf{y} - \mathbf{x}_1\| = \varepsilon/2$, then $\mathbf{y} \in \operatorname{cl} S$. But $\mathbf{x}_1 = \lambda \mathbf{y} + (1-\lambda)\mathbf{x}_2$, where $\lambda = 1/(1+\Delta) \in (0, 1)$. Since $\mathbf{y} \in \operatorname{cl} S$ and $\mathbf{x}_2 \in \operatorname{int} S$, then by the theorem, $\mathbf{x}_1 \in \operatorname{int} S$, and the proof is complete.

The above theorem and its corollaries can be considerably strengthened by using the notion of relative interiors (see the Notes and References section at the end of the chapter).

2.3 Separation and Support of Convex Sets

The notions of supporting hyperplanes and separation of disjoint convex sets are very important in optimization. Almost all optimality conditions and duality relationships use some sort of separation or support of convex sets. The results of this section are based on the following geometric fact: given a closed convex set S and a point $\mathbf{y} \notin S$, there exist a unique point $\bar{\mathbf{x}} \in S$ with minimum distance from $\mathbf{y}$ and a hyperplane that separates $\mathbf{y}$ and S.

Minimum Distance from a Point to a Convex Set

In order to establish the above important result, the following *parallelogram law* is needed. Let $\mathbf{a}$ and $\mathbf{b}$ be two vectors in E_n. Then

$$\|\mathbf{a}+\mathbf{b}\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 + 2\mathbf{a}'\mathbf{b}$$
$$\|\mathbf{a}-\mathbf{b}\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - 2\mathbf{a}'\mathbf{b}$$

By adding, we get

$$\|\mathbf{a}+\mathbf{b}\|^2 + \|\mathbf{a}-\mathbf{b}\|^2 = 2\,\|\mathbf{a}\|^2 + 2\,\|\mathbf{b}\|^2$$

This result is illustrated in Figure 2.5 and can be interpreted as follows: the sum of squared norms of the diagonals of a parallelogram is equal to the sum of squared norms of its sides.

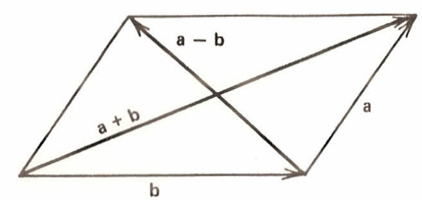


Figure 2.5 Parallelogram law.

2.3.1 Theorem

Let S be a closed convex set in E_n and $\mathbf{y} \notin S$. Then there exists a unique point $\bar{\mathbf{x}} \in S$ with minimum distance from $\mathbf{y}$. Furthermore, $\bar{\mathbf{x}}$ is the minimizing point if and only if $(\mathbf{x}-\bar{\mathbf{x}})'(\bar{\mathbf{x}}-\mathbf{y}) \geq 0$ for all $\mathbf{x} \in S$.

Proof

Let $\inf\{\|\mathbf{y}-\mathbf{x}\| : \mathbf{x} \in S\} = \gamma > 0$. There exists a sequence $\{\mathbf{x}_k\}$ in S such that $\|\mathbf{y}-\mathbf{x}_k\| \to \gamma$. We show that $\{\mathbf{x}_k\}$ has a limit $\bar{\mathbf{x}} \in S$ by showing that $\{\mathbf{x}_k\}$ is a

Cauchy sequence. By the parallelogram law, we have

$$\begin{aligned}\|\mathbf{x}_k-\mathbf{x}_m\|^2&=2\,\|\mathbf{x}_k-\mathbf{y}\|^2+2\,\|\mathbf{x}_m-\mathbf{y}\|^2-\|\mathbf{x}_k+\mathbf{x}_m-2\,\mathbf{y}\|^2\\&=2\,\|\mathbf{x}_k-\mathbf{y}\|^2+2\,\|\mathbf{x}_m-\mathbf{y}\|^2-4\left\|\frac{\mathbf{x}_k+\mathbf{x}_m}{2}-\mathbf{y}\right\|^2\end{aligned}$$

Note that $(\mathbf{x}_k+\mathbf{x}_m)/2\in S$, and by definition of γ we have $\left\|\dfrac{\mathbf{x}_k+\mathbf{x}_m}{2}-\mathbf{y}\right\|^2\geq\gamma^2$. Therefore $\|\mathbf{x}_k-\mathbf{x}_m\|^2\leq 2\,\|\mathbf{x}_k-\mathbf{y}\|^2+2\,\|\mathbf{x}_m-\mathbf{y}\|^2-4\gamma^2$. By choosing k and m sufficiently large, $\|\mathbf{x}_k-\mathbf{y}\|^2$ and $\|\mathbf{x}_m-\mathbf{y}\|^2$ can be made sufficiently close to γ^2, and hence $\|\mathbf{x}_k-\mathbf{x}_m\|^2$ can be made sufficiently close to zero. Therefore $\{\mathbf{x}_k\}$ is a Cauchy sequence and has a limit $\bar{\mathbf{x}}$. Since S is closed, $\bar{\mathbf{x}}\in S$. To show uniqueness, suppose that there is an $\bar{\mathbf{x}}'\in S$ such that $\|\mathbf{y}-\bar{\mathbf{x}}\|=\|\mathbf{y}-\bar{\mathbf{x}}'\|=\gamma$. By convexity of S, $(\bar{\mathbf{x}}+\bar{\mathbf{x}}')/2\in S$. By Schwartz inequality, we get

$$\left\|\mathbf{y}-\frac{\bar{\mathbf{x}}+\bar{\mathbf{x}}'}{2}\right\|\leq\tfrac{1}{2}\,\|\mathbf{y}-\bar{\mathbf{x}}\|+\tfrac{1}{2}\,\|\mathbf{y}-\bar{\mathbf{x}}'\|=\gamma$$

If strict inequality holds, we violate the definition of γ. Therefore equality holds, and we must have $\mathbf{y}-\bar{\mathbf{x}}=\lambda(\mathbf{y}-\bar{\mathbf{x}}')$ for some λ. Since $\|\mathbf{y}-\bar{\mathbf{x}}\|=\|\mathbf{y}-\bar{\mathbf{x}}'\|=\gamma$, $|\lambda|=1$. Clearly, $\lambda\neq-1$, because otherwise $\mathbf{y}=(\bar{\mathbf{x}}+\bar{\mathbf{x}}')/2\in S$, contradicting the assumption that $\mathbf{y}\notin S$. So $\lambda=1$, $\bar{\mathbf{x}}'=\bar{\mathbf{x}}$, and uniqueness is established.

To complete the proof, we need to show that $(\mathbf{x}-\bar{\mathbf{x}})^t(\bar{\mathbf{x}}-\mathbf{y})\geq 0$ for all $\mathbf{x}\in S$ is both a necessary and sufficient condition for $\bar{\mathbf{x}}$ to be the point in S closest to $\mathbf{y}$.

To prove sufficiency, let $\mathbf{x}\in S$. Then,

$$\|\mathbf{y}-\mathbf{x}\|^2=\|\mathbf{y}-\bar{\mathbf{x}}+\bar{\mathbf{x}}-\mathbf{x}\|^2=\|\mathbf{y}-\bar{\mathbf{x}}\|^2+\|\bar{\mathbf{x}}-\mathbf{x}\|^2+2(\bar{\mathbf{x}}-\mathbf{x})^t(\mathbf{y}-\bar{\mathbf{x}})$$

Since $\|\bar{\mathbf{x}}-\mathbf{x}\|^2\geq 0$ and $(\bar{\mathbf{x}}-\mathbf{x})^t(\mathbf{y}-\bar{\mathbf{x}})\geq 0$ by assumption, $\|\mathbf{y}-\mathbf{x}\|^2\geq\|\mathbf{y}-\bar{\mathbf{x}}\|^2$, and $\bar{\mathbf{x}}$ is the minimizing point. Conversely, assume that $\|\mathbf{y}-\mathbf{x}\|^2\geq\|\mathbf{y}-\bar{\mathbf{x}}\|^2$ for all $\mathbf{x}\in S$. Let $\mathbf{x}\in S$ and note that $\bar{\mathbf{x}}+\lambda(\mathbf{x}-\bar{\mathbf{x}})\in S$ for $\lambda>0$ and sufficiently small. Therefore,

$$\|\mathbf{y}-\bar{\mathbf{x}}-\lambda(\mathbf{x}-\bar{\mathbf{x}})\|^2\geq\|\mathbf{y}-\bar{\mathbf{x}}\|^2 \tag{2.3}$$

Also

$$\|\mathbf{y}-\bar{\mathbf{x}}-\lambda(\mathbf{x}-\bar{\mathbf{x}})\|^2=\|\mathbf{y}-\bar{\mathbf{x}}\|^2+\lambda^2\,\|\mathbf{x}-\bar{\mathbf{x}}\|^2+2\lambda(\mathbf{x}-\bar{\mathbf{x}})^t(\bar{\mathbf{x}}-\mathbf{y}) \tag{2.4}$$

From (2.3) and (2.4), we get

$$\lambda^2\,\|\mathbf{x}-\bar{\mathbf{x}}\|^2+2\lambda(\mathbf{x}-\bar{\mathbf{x}})^t(\bar{\mathbf{x}}-\mathbf{y})\geq 0$$

for all $\lambda>0$ and sufficiently small. Dividing by $\lambda>0$ and letting $\lambda\to 0$, the result follows.

The above theorem is illustrated in Figure 2.6. Note that the angle between $\mathbf{x}-\bar{\mathbf{x}}$ and $\bar{\mathbf{x}}-\mathbf{y}$ for any point $\mathbf{x}$ in S is less than or equal to 90° and hence $(\mathbf{x}-\bar{\mathbf{x}})^t(\bar{\mathbf{x}}-\mathbf{y})\geq 0$.

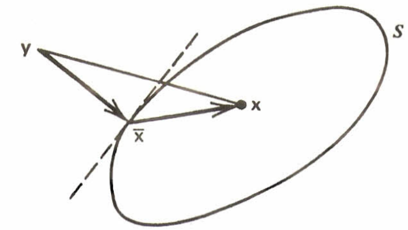


Figure 2.6 Minimum distance to a closed convex set.

Hyperplanes and Separation of Two Sets

Since we shall be dealing with separating and supporting hyperplanes, precise definitions of hyperplanes and half spaces are given below.

2.3.2 Definition

A *hyperplane* H in E_n is a collection of points of the form $\{\mathbf{x}:\mathbf{p}^t\mathbf{x}=\alpha\}$, where $\mathbf{p}$ is a nonzero vector in E_n and α is a scalar. The vector $\mathbf{p}$ is called the *normal vector* of the hyperplane. A hyperplane H defines two *closed half spaces* $H^+=\{\mathbf{x}:\mathbf{p}^t\mathbf{x}\geq\alpha\}$ and $H^-=\{\mathbf{x}:\mathbf{p}^t\mathbf{x}\leq\alpha\}$ and the two *open half spaces* $\{\mathbf{x}:\mathbf{p}^t\mathbf{x}>\alpha\}$ and $\{\mathbf{x}:\mathbf{p}^t\mathbf{x}<\alpha\}$.

Note that any point in E_n lies in H^+, in H^-, or in both. Also, a hyperplane H and the corresponding half spaces can be written in reference to a fixed point, say $\bar{\mathbf{x}}\in H$. If $\bar{\mathbf{x}}\in H$, then $\mathbf{p}^t\bar{\mathbf{x}}=\alpha$, and hence any point $\mathbf{x}\in H$ must satisfy $\mathbf{p}^t\mathbf{x}-\mathbf{p}^t\bar{\mathbf{x}}=\alpha-\alpha=0$; that is, $\mathbf{p}^t(\mathbf{x}-\bar{\mathbf{x}})=0$. Accordingly, $H^+=\{\mathbf{x}:\mathbf{p}^t(\mathbf{x}-\bar{\mathbf{x}})\geq 0\}$ and $H^-=\{\mathbf{x}:\mathbf{p}^t(\mathbf{x}-\bar{\mathbf{x}})\leq 0\}$. Figure 2.7 shows a hyperplane H passing through $\bar{\mathbf{x}}$ and having a normal vector $\mathbf{p}$.

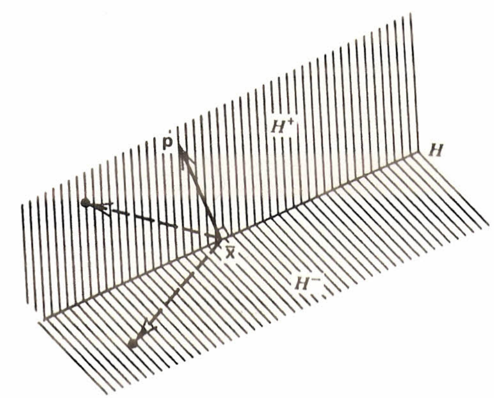


Figure 2.7 An example of a hyperplane and corresponding half spaces.

As an example, consider $H=\{(x_1, x_2, x_3, x_4): x_1+x_2-x_3+2x_4=4\}$. The normal vector $\mathbf{p}=(1,1,-1,2)^t$. Alternatively the hyperplane can be written in reference to any point in H, for example, $\bar{\mathbf{x}}=(0,6,0,-1)^t$. In this case, we write $H=\{(x_1, x_2, x_3, x_4): x_1+(x_2-6)-x_3+2(x_4+1)=0\}$.

2.3.3 Definition

Let S_1 and S_2 be nonempty sets in E_n. A hyperplane $H=\{\mathbf{x}:\mathbf{p}^t\mathbf{x}=\alpha\}$ is said to *separate* S_1 and S_2 if $\mathbf{p}^t\mathbf{x}\geq\alpha$ for each $\mathbf{x}\in S_1$ and $\mathbf{p}^t\mathbf{x}\leq\alpha$ for each $\mathbf{x}\in S_2$. If, in addition, $S_1\cup S_2\not\subset H$, then H is said to *properly separate* S_1 and S_2. The hyperplane H is said to *strictly separate* S_1 and S_2 if $\mathbf{p}^t\mathbf{x}>\alpha$ for each $\mathbf{x}\in S_1$ and $\mathbf{p}^t\mathbf{x}<\alpha$ for each $\mathbf{x}\in S_2$. The hyperplane H is said to *strongly separate* S_1 and S_2 if $\mathbf{p}^t\mathbf{x}\geq\alpha+\varepsilon$ for each $\mathbf{x}\in S_1$ and $\mathbf{p}^t\mathbf{x}\leq\alpha$ for each $\mathbf{x}\in S_2$, where ε is a positive scalar.

Figure 2.8 shows various types of separation. Of course, strong separation implies strict separation, which implies proper separation, which in turn implies separation. Improper separation is usually of little value, since it corresponds to a hyperplane containing both S_1 and S_2 as shown in Figure 2.8.

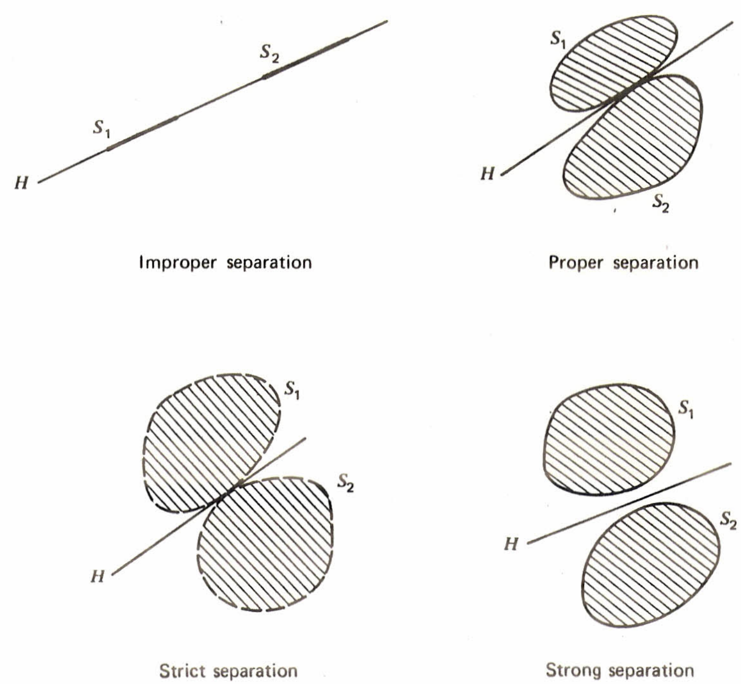


Figure 2.8 Various types of separation.

Separation of a Convex Set and a Point

We shall now present the first and most fundamental separation theorem. Other separation and support theorems will follow from this basic result.

2.3.4 Theorem

Let S be a nonempty closed convex set in E_n and $\mathbf{y}\notin S$. Then there exists a nonzero vector $\mathbf{p}$ and a scalar α such that $\mathbf{p}^t\mathbf{y}>\alpha$ and $\mathbf{p}^t\mathbf{x}\leq\alpha$ for each $\mathbf{x}\in S$.

Proof

The set S is a nonempty closed convex set and $\mathbf{y}\notin S$. Hence by Theorem 2.3.1, there exists a unique minimizing point $\bar{\mathbf{x}}\in S$ such that $(\mathbf{x}-\bar{\mathbf{x}})^t(\mathbf{y}-\bar{\mathbf{x}})\leq 0$ for each $\mathbf{x}\in S$. Note that

$$\|\mathbf{y}-\bar{\mathbf{x}}\|^2=(\mathbf{y}-\bar{\mathbf{x}})^t(\mathbf{y}-\bar{\mathbf{x}})=\mathbf{y}^t(\mathbf{y}-\bar{\mathbf{x}})-\bar{\mathbf{x}}^t(\mathbf{y}-\bar{\mathbf{x}}) \tag{2.5}$$

But since $-\bar{\mathbf{x}}^t(\mathbf{y}-\bar{\mathbf{x}})\leq-\mathbf{x}^t(\mathbf{y}-\bar{\mathbf{x}})$ for any $\mathbf{x}\in S$, then (2.5) implies that $\mathbf{p}^t(\mathbf{y}-\mathbf{x})\geq\|\mathbf{y}-\bar{\mathbf{x}}\|^2$ for each $\mathbf{x}\in S$, where $\mathbf{p}=\mathbf{y}-\bar{\mathbf{x}}\neq\mathbf{0}$. This shows that $\mathbf{p}^t\mathbf{y}\geq\mathbf{p}^t\mathbf{x}+\|\mathbf{y}-\bar{\mathbf{x}}\|^2$ for each $\mathbf{x}\in S$. Letting $\alpha=\sup\{\mathbf{p}^t\mathbf{x}:\mathbf{x}\in S\}$, the result follows.

Corollary

Let S be a closed convex set in E_n. Then S is the intersection of all half spaces containing S.

Proof

Obviously S is contained in the intersection of all half spaces containing it. To the contrary of the desired result, suppose that there is a point $\mathbf{y}$ in the intersection of these half spaces but not in S. By the theorem, there exists a half space that contains S but not $\mathbf{y}$. This contradiction proves the corollary.

The following statements are equivalent to the conclusion of the theorem. The reader is asked to verify this equivalence. Note that statements 1 and 2 are equivalent only in this special case since $\mathbf{y}$ is a point.

1. There exists a hyperplane that *strictly* separates S and $\mathbf{y}$.
2. There exists a hyperplane that *strongly* separates S and $\mathbf{y}$.
3. There exists a vector $\mathbf{p}$ such that $\mathbf{p}^t\mathbf{y}>\sup\{\mathbf{p}^t\mathbf{x}:\mathbf{x}\in S\}$.
4. There exists a vector $\mathbf{p}$ such that $\mathbf{p}^t\mathbf{y}<\inf\{\mathbf{p}^t\mathbf{x}:\mathbf{x}\in S\}$.

Farkas' Theorem as a Consequence of Theorem 2.3.4

Farkas' theorem has been used extensively in the derivation of optimality conditions of linear and nonlinear programming problems. The theorem can be

Faltam
Páginas 46 e 47

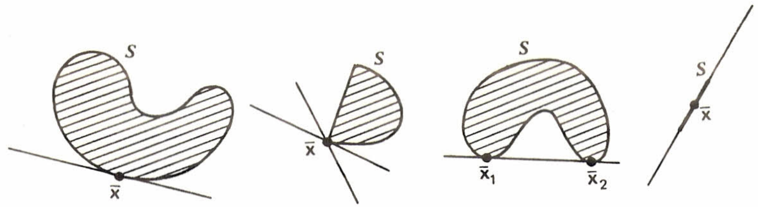


Figure 2.10 Examples of supporting hyperplanes.

Figure 2.10 shows some examples of supporting hyperplanes. The figure illustrates the cases of a unique supporting hyperplane at a boundary point, an infinite number of supporting hyperplanes at a boundary point, a hyperplane that supports the set at more than one point, and finally an improper supporting hyperplane that contains the whole set.

We now prove that a convex set has a supporting hyperplane at each boundary point (see Figure 2.11). As a corollary, a result similar to Theorem 2.3.4, where S is not required to be closed, follows.

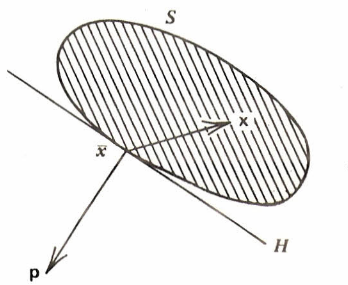


Figure 2.11 Supporting hyperplane.

2.3.7 Theorem

Let S be a nonempty convex set in E_n, and let $\bar{\mathbf{x}} \in \partial S$. Then, there exists a hyperplane that supports S at $\bar{\mathbf{x}}$; that is, there exists a nonzero vector $\mathbf{p}$ such that $\mathbf{p}'(\mathbf{x}-\bar{\mathbf{x}}) \leq 0$ for each $\mathbf{x} \in \text{cl } S$.

Proof

Since $\bar{\mathbf{x}} \in \partial S$, there exists a sequence $\{\mathbf{y}_k\}$ not in cl S such that $\mathbf{y}_k \to \bar{\mathbf{x}}$. By Theorem 2.3.4, corresponding to each $\mathbf{y}_k$ there exists a $\mathbf{p}_k$ with norm 1 such that $\mathbf{p}_k'\mathbf{y}_k > \mathbf{p}_k'\mathbf{x}$ for each $\mathbf{x} \in \text{cl } S$. (In Theorem 2.3.4, the normal vector can be normalized by dividing it by its norm, so that $\|\mathbf{p}_k\| = 1$.) Since $\{\mathbf{p}_k\}$ is bounded, it has a convergent subsequence $\{\mathbf{p}_k\}_{\mathscr{K}}$ with limit $\mathbf{p}$ whose norm is equal to one.

Considering this subsequence we have $\mathbf{p}_k'\mathbf{y}_k > \mathbf{p}_k'\mathbf{x}$ for each $\mathbf{x} \in \text{cl } S$. Fix $\mathbf{x} \in \text{cl } S$ and take the limit as $k \in \mathscr{K}$ approaches ∞. Then $\mathbf{p}'(\mathbf{x}-\bar{\mathbf{x}}) \leq 0$. Since this is true for each $\mathbf{x} \in \text{cl } S$, the result follows.

Corollary

Let S be a nonempty convex set in E_n and $\bar{\mathbf{x}} \notin S$. Then there is a nonzero vector $\mathbf{p}$ such that $\mathbf{p}'(\mathbf{x}-\bar{\mathbf{x}}) \leq 0$ for each $\mathbf{x} \in \text{cl } S$.

Proof

If $\bar{\mathbf{x}} \notin \text{cl } S$, then the corollary follows from Theorem 2.3.4. On the other hand, if $\bar{\mathbf{x}} \in \partial S$, the corollary reduces to the above theorem.

Separation of Two Convex Sets

So far, we have discussed separation of a convex set and a point not in the set and have also discussed support of convex sets at boundary points. In addition, if we have two disjoint convex sets, they can be separated by a hyperplane H such that one of the sets belongs to H^+ and the other set belongs to H^-. In fact, this result holds true even if the two sets have some points in common, as long as their interiors are disjoint. This result is made precise by the following theorem.

2.3.8 Theorem

Let S_1 and S_2 be nonempty convex sets in E_n and suppose that $S_1 \cap S_2$ is empty. Then there exists a hyperplane that separates S_1 and S_2; that is, there exists a nonzero vector $\mathbf{p}$ in E_n such that

$$\inf \{\mathbf{p}'\mathbf{x} : \mathbf{x} \in S_1\} \geq \sup \{\mathbf{p}'\mathbf{x} : \mathbf{x} \in S_2\}$$

Proof

Let $S = S_1 - S_2 = \{\mathbf{x}_1 - \mathbf{x}_2 : \mathbf{x}_1 \in S_1 \text{ and } \mathbf{x}_2 \in S_2\}$. Note that S is convex. Furthermore, $\mathbf{0} \notin S$ because otherwise $S_1 \cap S_2$ will be nonempty. By the corollary of Theorem 2.3.7, there exists a nonzero $\mathbf{p} \in E_n$ such that $\mathbf{p}'\mathbf{x} \geq 0$ for all $\mathbf{x} \in S$. This means that $\mathbf{p}'\mathbf{x}_1 \geq \mathbf{p}'\mathbf{x}_2$ for all $\mathbf{x}_1 \in S_1$ and $\mathbf{x}_2 \in S_2$, and the result follows.

Corollary 1

Let S_1 and S_2 be nonempty convex sets in E_n. Suppose that int S_2 is not empty and that $S_1 \cap \text{int } S_2$ is empty. Then there exists a nonzero $\mathbf{p}$ such that

$$\inf \{\mathbf{p}'\mathbf{x} : \mathbf{x} \in S_1\} \geq \sup \{\mathbf{p}'\mathbf{x} : \mathbf{x} \in S_2\}$$

Proof

Replace S_2 by int S_2, apply the theorem, and note that

$$\sup\{\mathbf{p}'\mathbf{x}:\mathbf{x}\in S_2\}=\sup\{\mathbf{p}'\mathbf{x}:\mathbf{x}\in \operatorname{int} S_2\}$$

Corollary 2

Let S_1 and S_2 be convex sets with nonempty interiors. Suppose that int $S_1 \cap$ int $S_2=\varnothing$. Then there exists a nonzero $\mathbf{p}$ such that

$$\inf\{\mathbf{p}'\mathbf{x}:\mathbf{x}\in S_2\}\geq\sup\{\mathbf{p}'\mathbf{x}:\mathbf{x}S_2\}.$$

Gordan's Theorem as a consequence of Theorem 2.3.8

We shall now derive a theorem credited to Gordan based on the existence of a hyperplane that separates two disjoint convex sets. This theorem is frequently used in nonlinear programming. Like the Farkas' theorem, Gordan's theorem states that exactly one of two systems only has a solution.

2.3.9 Theorem (Gordan's Theorem)

Let $\mathbf{A}$ be an $m\times n$ matrix. Exactly one of the following systems has a solution:

System 1 $\quad \mathbf{Ax}<\mathbf{0} \qquad$ for some $\mathbf{x}\in E_n$

System 2 $\quad \mathbf{A}'\mathbf{p}=\mathbf{0}$ and $\mathbf{p}\geq\mathbf{0} \qquad$ for some nonzero $\mathbf{p}\in E_m$

Proof

We shall first prove that if *System 1* has a solution $\hat{\mathbf{x}}$, then we cannot have a solution to $\mathbf{A}'\mathbf{p}=\mathbf{0}, \mathbf{p}\geq\mathbf{0}, \mathbf{p}$ nonzero. Suppose, on the contrary, that a solution $\hat{\mathbf{p}}$ exists. Then, since $\mathbf{A}\hat{\mathbf{x}}<\mathbf{0}, \hat{\mathbf{p}}\geq\mathbf{0}$, and $\hat{\mathbf{p}}\neq\mathbf{0}$, we have $\hat{\mathbf{p}}'\mathbf{A}\hat{\mathbf{x}}<0$, that is, $\hat{\mathbf{x}}'\mathbf{A}'\hat{\mathbf{p}}<0$. But $\mathbf{A}'\hat{\mathbf{p}}=\mathbf{0}$ by assumption. Hence, $\hat{\mathbf{x}}'\mathbf{A}'\hat{\mathbf{p}}=0$, a contradiction. Hence, *System 2* cannot have a solution.

Now assume that *System 1* has no solution. Consider the following two sets:

$$S_1=\{\mathbf{z}:\mathbf{z}=\mathbf{Ax}, \mathbf{x}\in E_n\}$$
$$S_2=\{\mathbf{z}:\mathbf{z}<\mathbf{0}\}$$

Note that S_1 and S_2 are nonempty convex sets such that $S_1\cap S_2=\varnothing$. Then, by Theorem 2.3.8, there exists a hyperplane that separates S_1 and S_2; that is, there exists a nonzero vector $\mathbf{p}$ such that

$$\mathbf{p}'\mathbf{Ax}\geq\mathbf{p}'\mathbf{z} \qquad \text{for each } \mathbf{x}\in E_n \text{ and } \mathbf{z}\in \operatorname{cl} S_2$$

Since each component of $\mathbf{z}$ could be made an arbitrarily large negative number, it follows that $\mathbf{p}\geq\mathbf{0}$. Also, by letting $\mathbf{z}=\mathbf{0}$, we must have $\mathbf{p}'\mathbf{Ax}\geq 0$ for each $\mathbf{x}\in E_n$. By choosing $\mathbf{x}=-\mathbf{A}'\mathbf{p}$, it then follows that $-\|\mathbf{A}'\mathbf{p}\|^2\geq 0$, and thus $\mathbf{A}'\mathbf{p}=\mathbf{0}$. Hence, *System 2* has a solution, and the proof is complete.

The separation theorem 2.3.8 can be strengthened to avoid trivial separation where both S_1 and S_2 are contained in the separating hyperplane.

2.3.10 Theorem (Strong Separation)

Let S_1 and S_2 be a closed convex sets, and suppose that S_1 is bounded. If $S_1\cap S_2$ is empty, then there exists a nonzero $\mathbf{p}$ and $\varepsilon>0$ such that

$$\inf\{\mathbf{p}'\mathbf{x}:\mathbf{x}\in S_1\}\geq\varepsilon+\sup\{\mathbf{p}'\mathbf{x}:\mathbf{x}\in S_2\}$$

Proof

Let $S=S_1-S_2$ and note that S is convex and that $\mathbf{0}\notin S$. We shall show that S is closed. Let $\{\mathbf{x}_k\}$ in S converge to $\mathbf{x}$. By definition of S, $\mathbf{x}_k=\mathbf{y}_k-\mathbf{z}_k$, where $\mathbf{y}_k\in S_1$ and $\mathbf{z}_k\in S_2$. Since S_1 is compact, there is a subsequence $\{\mathbf{y}_k\}_{\mathscr{K}}$ with limit $\mathbf{y}$ in S_1. Since $\mathbf{y}_k-\mathbf{z}_k\to\mathbf{x}$ and $\mathbf{y}_k\to\mathbf{y}$ for $k\in\mathscr{K}$, then $\mathbf{z}_k\to\mathbf{z}$. Since S_2 is closed, $\mathbf{z}\in S_2$. Therefore $\mathbf{x}=\mathbf{y}-\mathbf{z}$ with $\mathbf{y}\in S_1$ and $\mathbf{z}\in S_2$. Therefore, $\mathbf{x}\in S$, and hence S is closed. By Theorem 2.3.4, there is a nonzero $\mathbf{p}$ and an ε such that $\mathbf{p}'\mathbf{x}\geq\varepsilon$ for each $\mathbf{x}\in S$ and $\mathbf{p}'\mathbf{0}<\varepsilon$. Therefore, $\varepsilon>0$. By the definition of S, we conclude that $\mathbf{p}'\mathbf{x}_1\geq\varepsilon+\mathbf{p}'\mathbf{x}_2$ for each $\mathbf{x}_1\in S_1$ and $\mathbf{x}_2\in S_2$, and the result follows.

2.4 Convex Cones and Polarity

In this section, we briefly discuss the notions of convex cones and polar cones. This section may be skipped without loss of continuity.

2.4.1 Definition

A nonempty set C in E_n is called a *cone* with vertex zero if $\mathbf{z}\in C$ implies that $\lambda\mathbf{x}\in C$ for all $\lambda\geq 0$. If, in addition, C is convex, then C is called a *convex cone.*

Figure 2.12 shows an example of a convex cone and an example of a nonconvex cone.

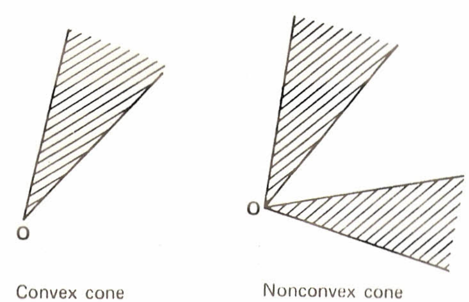


Figure 2.12 Examples of cones.

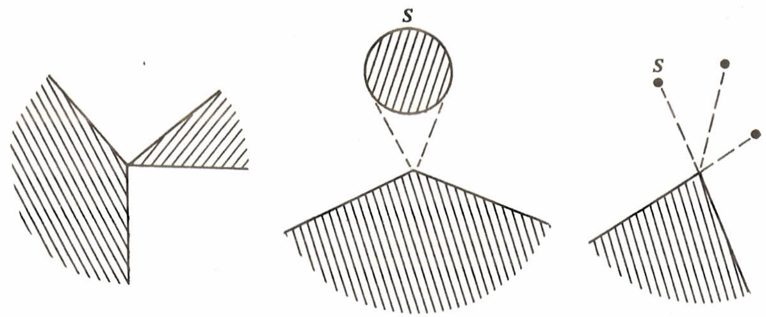


Figure 2.13 Examples of polar cones.

An important special class of convex cones is that of polar cones defined below and illustrated in Figure 2.13.

2.4.2 Definition

Let S be a nonempty set in E_n. Then the *polar cone* of S, denoted by S^*, is given by $\{\mathbf{p}:\mathbf{p}^t\mathbf{x}\leq 0 \text{ for all } \mathbf{x}\in S\}$. If S is empty, S^* will be interpreted as E_n.

The following lemma, the proof of which is left as an exercise, summarizes some facts about polar cones.

2.4.3 Lemma

Let S, S_1, and S_2 be nonempty sets in E_n. Then the following statements hold true.

1. S^* is a closed convex cone.
2. $S\subset S^{**}$, where S^{**} is the polar cone of S^*.
3. $S_1\subset S_2$ implies that $S_2^*\subset S_1^*$.

We now prove an important theorem for closed convex cones. As an application of the theorem, we give another derivation of Farkas' theorem.

2.4.4 Theorem

Let C be a nonempty closed convex cone. Then $C=C^{**}$.

Proof

Clearly $C\subset C^{**}$. Now let $\mathbf{x}\in C^{**}$, and suppose, by contradiction, that $\mathbf{x}\notin C$. By Theorem 2.3.4, there exists a nonzero vector $\mathbf{p}$ and a scalar α such that $\mathbf{p}^t\mathbf{y}\leq\alpha$ for all $\mathbf{y}\in C$ and $\mathbf{p}^t\mathbf{x}>\alpha$. But since $\mathbf{y}=\mathbf{0}\in C$, then $\alpha\geq 0$, and so $\mathbf{p}^t\mathbf{x}>0$. We now show that $\mathbf{p}\in C^*$. If not, then $\mathbf{p}^t\bar{\mathbf{y}}>0$ for some $\bar{\mathbf{y}}\in C$, and $\mathbf{p}^t(\lambda\bar{\mathbf{y}})$ can be

made arbitrarily large by choosing λ arbitrarily large. This contradicts the fact that $\mathbf{p}^t\mathbf{y}\leq\alpha$ for all $\mathbf{y}\in C$. Therefore, $\mathbf{p}\in C^*$. Since $\mathbf{x}\in C^{**}$, then $\mathbf{p}^t\mathbf{x}\leq 0$. This contradicts the fact that $\mathbf{p}^t\mathbf{x}>0$, and we conclude that $\mathbf{x}\in C$. This completes the proof.

Farkas' Theorem as a consequence of Theorem 2.4.4

Let $\mathbf{A}$ be an $m\times n$ matrix, and let $C=\{\mathbf{A}^t\mathbf{y}:\mathbf{y}\geq\mathbf{0}\}$. Note that C is a closed convex cone. It can be easily verified that $C^*=\{\mathbf{x}:\mathbf{A}\mathbf{x}\leq\mathbf{0}\}$. By the theorem, $\mathbf{c}\in C^{**}$ if and only if $\mathbf{c}\in C$. But $\mathbf{c}\in C^{**}$ means that whenever $\mathbf{x}\in C^*$ then $\mathbf{c}^t\mathbf{x}\leq 0$, or equivalently, $\mathbf{A}\mathbf{x}\leq\mathbf{0}$ implies that $\mathbf{c}^t\mathbf{x}\leq 0$. By definition of C, $\mathbf{c}\in C$ means that $\mathbf{c}=\mathbf{A}^t\mathbf{y}$ and $\mathbf{y}\geq\mathbf{0}$. Thus the result $C=C^{**}$ could be stated as follows: System 1 below is consistent if and only if System 2 has a solution $\mathbf{y}$.

System 1 $\mathbf{A}\mathbf{x}\leq\mathbf{0}$ implies $\mathbf{c}^t\mathbf{x}\leq 0$.

System 2 $\mathbf{A}^t\mathbf{y}=\mathbf{c}$, $\mathbf{y}\geq\mathbf{0}$.

This statement can be put in the more usual and equivalent form of Farkas' theorem. Exactly one of the following two systems has a solution:

System 1 $\mathbf{A}\mathbf{x}\leq\mathbf{0}$, $\mathbf{c}^t\mathbf{x}>0$ (that is, $\mathbf{c}\notin C^{**}=C$).

System 2 $\mathbf{A}^t\mathbf{y}=\mathbf{c}$, $\mathbf{y}\geq\mathbf{0}$ (that is, $\mathbf{c}\in C$).

2.5 Polyhedral Sets, Extreme Points, and Extreme Directions

In this section we introduce the notions of extreme points and extreme directions for convex sets. We then discuss in more detail their use for the special important case of polyhedral sets.

Polyhedral Sets

Polyhedral sets represent an important special case of convex sets. We have seen from the corollary to Theorem 2.3.4 that any closed convex set is the intersection of all closed half spaces containing it. In the case of polyhedral sets, only a finite number of half spaces is needed to represent the set.

2.5.1 Definition

A nonempty set S in E_n is called a *polyhedral set* if it is the intersection of a finite number of closed half spaces; that is, $S=\{\mathbf{x}:\mathbf{p}_i^t\mathbf{x}\leq\alpha_i \text{ for } i=1,\ldots,m\}$, where $\mathbf{p}_i$ is a nonzero vector and α_i is a scalar for $i=1,\ldots,m$.

Note that a polyhderal set is a closed convex set. Since an equation can be represented by two inequalities, a polyhedral set can be represented by a finite number of inequalities and/or equations. The following are some typical examples of polyhedral sets, where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{b}$ is an m vector.

$$S = \{\mathbf{x} : \mathbf{A}\mathbf{x} \le \mathbf{b}\}$$
$$S = \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \ge \mathbf{0}\}$$
$$S = \{\mathbf{x} : \mathbf{A}\mathbf{x} \ge \mathbf{b}, \mathbf{x} \ge \mathbf{0}\}$$

Figure 2.14 illustrates the polyhedral set

$$S = \{(x_1, x_2) : -x_1 + x_2 \le 2, x_2 \le 4, x_1 \ge 0, x_2 \ge 0\}$$

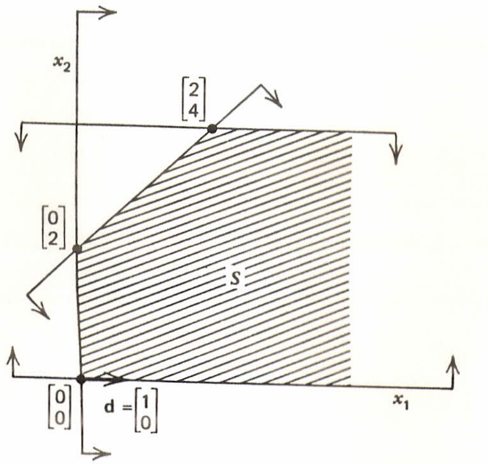


Figure 2.14 A polyhedral set.

Extreme Points and Extreme Directions

We now introduce the concepts of extreme points and extreme directions for convex sets. We then give their full characterizations in the case of polyhedral sets.

2.5.2 Definition

Let S be a nonempty convex set in E_n. A vector $\mathbf{x} \in S$ is called an *extreme point* of S if $\mathbf{x} = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ with $\mathbf{x}_1, \mathbf{x}_2 \in S$, and $\lambda \in (0, 1)$ implies that $\mathbf{x} = \mathbf{x}_1 = \mathbf{x}_2$.

The following are some examples of extreme points of convex sets. We denote the set of extreme points by E and illustrate them in Figure 2.15 by

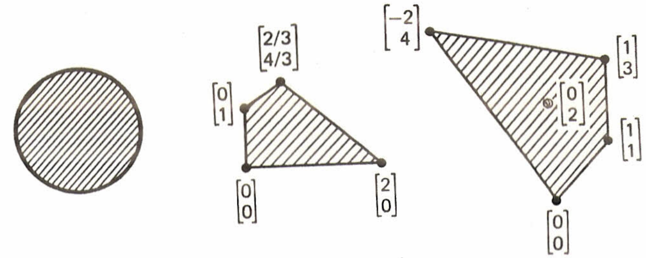


Figure 2.15 Examples of extreme points.

dark points or dark lines.

1. $S = \{(x_1, x_2) : x_1^2 + x_2^2 \le 1\}$
   $E = \{(x_1, x_2) : x_1^2 + x_2^2 = 1\}$
2. $S = \{(x_1, x_2) : x_1 + x_2 \le 2, -x_1 + 2x_2 \le 2, x_1, x_2 \ge 0\}$
   $E = \{(0, 0)^t, (0, 1)^t, (2/3, 4/3)^t, (2, 0)^t\}$
3. S is the polytope generated by $(0, 0)^t, (1, 1)^t, (1, 3)^t, (-2, 4)^t$, and $(0, 2)^t$
   $E = \{(0, 0)^t, (1, 1)^t, (1, 3)^t, (-2, 4)^t\}$

From Figure 2.15, we see that any point of the convex set S can be represented as a convex combination of the extreme points. This turns out to be true for compact convex sets. However, for unbounded sets, we may not be able to represent every point in the set as a convex combination of its extreme points. To illustrate, let $S = \{(x_1, x_2) : x_2 \ge |x_1|\}$. Note that S is convex and closed. However, S contains only one extreme point, namely the origin, and obviously S is not equal to the collection of convex combinations of its extreme points. In order to deal with unbounded sets, the notion of extreme directions is needed.

2.5.3 Definition

Let S be a closed convex set in E_n. A nonzero vector $\mathbf{d}$ in E_n is called a *direction* of S if for each $\mathbf{x} \in S$, $\mathbf{x} + \lambda \mathbf{d} \in S$ for all $\lambda \ge 0$. Two directions $\mathbf{d}_1$ and $\mathbf{d}_2$ of S are called *distinct* if $\mathbf{d}_1 \ne \alpha \mathbf{d}_2$ for any $\alpha > 0$. A direction $\mathbf{d}$ of S is called an *extreme direction* if it cannot be written as a positive linear combination of two distinct directions, that is, if $\mathbf{d} = \lambda_1 \mathbf{d}_1 + \lambda_2 \mathbf{d}_2$ for $\lambda_1, \lambda_2 > 0$ then $\mathbf{d}_1 = \alpha \mathbf{d}_2$ for some $\alpha > 0$.

To illustrate, consider $S = \{(x_1, x_2) : x_2 \ge |x_1|\}$ shown in Figure 2.16. The directions of S are nonzero vectors that make an angle less than or equal to 45° with the vector $(0, 1)^t$. In particular, $\mathbf{d}_1 = (1, 1)^t$ and $\mathbf{d}_2 = (-1, 1)^t$ are two extreme directions of S. Any other direction of S can be represented as a positive linear combination of $\mathbf{d}_1$ and $\mathbf{d}_2$.

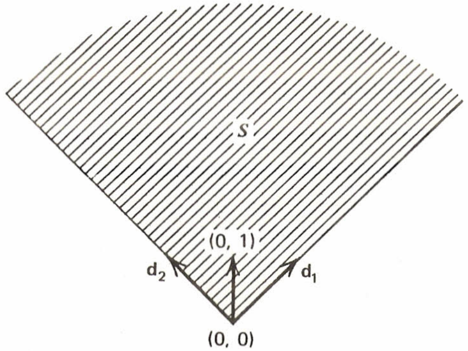


Figure 2.16 Example of extreme directions.

Characterization of Extreme Points and Extreme Directions for Polyhedral Sets

Consider the polyhedral set $S=\{\mathbf{x}:\mathbf{A}\mathbf{x}=\mathbf{b}, \mathbf{x}\geq\mathbf{0}\}$, where $\mathbf{A}$ is an $m\times n$ matrix and $\mathbf{b}$ is an m vector. We assume that the rank of $\mathbf{A}$ is m. If not, we can throw away any redundant equations.

Extreme Points Rearrange the columns of $\mathbf{A}$ so that $\mathbf{A}=[\mathbf{B},\mathbf{N}]$, where $\mathbf{B}$ is an $m\times m$ matrix of full rank, and $\mathbf{N}$ is an $m\times n-m$ matrix. Let $\mathbf{x}_B$ and $\mathbf{x}_N$ be the vectors corresponding to $\mathbf{B}$ and $\mathbf{N}$, respectively. Then $\mathbf{A}\mathbf{x}=\mathbf{b}$ and $\mathbf{x}\geq\mathbf{0}$ can be rewritten as follows:

$$\mathbf{B}\mathbf{x}_B+\mathbf{N}\mathbf{x}_N=\mathbf{b} \quad \text{and} \quad \mathbf{x}_B\geq\mathbf{0}, \mathbf{x}_N\geq\mathbf{0}$$

The following theorem gives a necessary and sufficient characterization of an extreme point of S.

2.5.4 Theorem (Characterization of Extreme Points)

Let $S=\{\mathbf{x}:\mathbf{A}\mathbf{x}=\mathbf{b}, \mathbf{x}\geq\mathbf{0}\}$, where $\mathbf{A}$ is an $m\times n$ matrix of rank m, and $\mathbf{b}$ is an m vector. A point $\mathbf{x}$ is an extreme point of S if and only if $\mathbf{A}$ can be decomposed into $[\mathbf{B},\mathbf{N}]$ such that:

$$\mathbf{x}=\begin{bmatrix}\mathbf{x}_B\\ \mathbf{x}_N\end{bmatrix}=\begin{bmatrix}\mathbf{B}^{-1}\mathbf{b}\\ \mathbf{0}\end{bmatrix}$$

where $\mathbf{B}$ is an $m\times m$ invertible matrix satisfying $\mathbf{B}^{-1}\mathbf{b}\geq\mathbf{0}$.

Proof

Suppose that $\mathbf{A}$ can be decomposed into $[\mathbf{B},\mathbf{N}]$ with $\mathbf{x}=\begin{bmatrix}\mathbf{B}^{-1}\mathbf{b}\\ \mathbf{0}\end{bmatrix}$ and $\mathbf{B}^{-1}\mathbf{b}\geq\mathbf{0}$.

It is obvious that $\mathbf{x}\in S$. Now suppose that $\mathbf{x}=\lambda\mathbf{x}_1+(1-\lambda)\mathbf{x}_2$ with $\mathbf{x}_1,\mathbf{x}_2\in S$ for some $\lambda\in(0,1)$. In particular, let $\mathbf{x}_1^t=(\mathbf{x}_{11}^t,\mathbf{x}_{12}^t)$ and $\mathbf{x}_2^t=(\mathbf{x}_{21}^t,\mathbf{x}_{22}^t)$. Then

$$\begin{bmatrix}\mathbf{B}^{-1}\mathbf{b}\\ \mathbf{0}\end{bmatrix}=\lambda\begin{bmatrix}\mathbf{x}_{11}\\ \mathbf{x}_{12}\end{bmatrix}+(1-\lambda)\begin{bmatrix}\mathbf{x}_{21}\\ \mathbf{x}_{22}\end{bmatrix}.$$

Since $\mathbf{x}_{12},\mathbf{x}_{22}\geq\mathbf{0}$ and $\lambda\in(0,1)$, it follows that $\mathbf{x}_{12}=\mathbf{x}_{22}=\mathbf{0}$. But this implies that $\mathbf{x}_{11}=\mathbf{x}_{21}=\mathbf{B}^{-1}\mathbf{b}$, and hence $\mathbf{x}=\mathbf{x}_1=\mathbf{x}_2$. This shows that $\mathbf{x}$ is an extreme point of S. Conversely, suppose that $\mathbf{x}$ is an extreme point of S. Without loss of generality, suppose that $\mathbf{x}=(x_1,\ldots,x_k,0,\ldots,0)^t$, where $x_1,\ldots,x_k$ are positive. We shall first show that $\mathbf{a}_1,\ldots,\mathbf{a}_k$ are linearly independent. By contradiction, suppose that there exist scalars $\lambda_1,\ldots,\lambda_k$ not all zero such that $\sum_{j=1}^k \lambda_j\mathbf{a}_j=\mathbf{0}$. Let $\boldsymbol{\lambda}=(\lambda_1,\ldots,\lambda_k,0,\ldots,0)^t$. Construct the following two vectors where $\alpha>0$ is chosen such that $\mathbf{x}_1,\mathbf{x}_2\geq\mathbf{0}$:

$$\mathbf{x}_1=\mathbf{x}+\alpha\boldsymbol{\lambda} \quad \text{and} \quad \mathbf{x}_2=\mathbf{x}-\alpha\boldsymbol{\lambda}$$

Note that

$$\mathbf{A}\mathbf{x}_1=\sum_{j=1}^k (x_j+\alpha\lambda_j)\mathbf{a}_j=\sum_{j=1}^k x_j\mathbf{a}_j+\alpha\sum_{j=1}^k \lambda_j\mathbf{a}_j=\mathbf{b}$$

and similarly $\mathbf{A}\mathbf{x}_2=\mathbf{b}$. Therefore $\mathbf{x}_1,\mathbf{x}_2\in S$ and since $\alpha>0$, $\mathbf{x}_1$ and $\mathbf{x}_2$ are distinct. Moreover, $\mathbf{x}=\frac{1}{2}\mathbf{x}_1+\frac{1}{2}\mathbf{x}_2$. This contradicts the fact that $\mathbf{x}$ is an extreme point. Thus $\mathbf{a}_1,\ldots,\mathbf{a}_k$ are linearly independent and $m-k$ columns out of the last $n-k$ columns may be chosen such that they, together with the first k columns, form a linearly independent set of vectors. To simplify the notation, suppose that these columns are $\mathbf{a}_{k+1},\ldots,\mathbf{a}_m$. Thus, $\mathbf{A}$ can be written as $\mathbf{A}=[\mathbf{B},\mathbf{N}]$, where $\mathbf{B}=[\mathbf{a}_1,\ldots,\mathbf{a}_m]$ is of full rank. Furthermore $\mathbf{B}^{-1}\mathbf{b}=(x_1,\ldots,x_k,0,\ldots,0)^t$, and since $x_j>0$ for $j=1,\ldots,k$, then $\mathbf{B}^{-1}\mathbf{b}\geq\mathbf{0}$. This completes the proof.

Corollary

The number of extreme points of S is finite.

Proof

The number of extreme points is less than or equal to $\binom{n}{m}=\dfrac{n!}{m!\,(n-m)!}$, which is the maximum number of possible ways to choose m columns of $\mathbf{A}$ to form $\mathbf{B}$.

From the above theorem, it is clear that a polyhedral set of the form $\{\mathbf{x}:\mathbf{A}\mathbf{x}=\mathbf{b}, \mathbf{x}\geq\mathbf{0}\}$ has a finite number of extreme points. The following theorem shows that every nonempty polyhedral set of this form must have at least one extreme point.

2.5.5 Theorem (Existence of Extreme Points)

Let $S=\{\mathbf{x}:\mathbf{A}\mathbf{x}=\mathbf{b}, \mathbf{x}\geq \mathbf{0}\}$ be nonempty, where $\mathbf{A}$ is an $m\times n$ matrix of rank m, and $\mathbf{b}$ is an m vector. Then S has at least one extreme point.

Proof

Let $\mathbf{x}\in S$, and without loss of generality, suppose that $\mathbf{x}=(x_1,\ldots,x_k,0,\ldots,0)^t$, where $x_j>0$ for $j=1,\ldots,k$. If $\mathbf{a}_1,\ldots,\mathbf{a}_k$ are linearly independent, then $k\leq m$, and $\mathbf{x}$ is an extreme point. Otherwise, there exist scalars $\lambda_1,\ldots,\lambda_k$ with at least one positive component such that $\sum_{j=1}^k \lambda_j\mathbf{a}_j=\mathbf{0}$. Define $\alpha>0$ as follows.

$$\alpha=\underset{1\leq j\leq k}{\text{minimum}}\left\{\frac{x_j}{\lambda_j}:\lambda_j>0\right\}=\frac{x_i}{\lambda_i}$$

Consider the point $\mathbf{x}'$ whose jth component x_j' is given by

$$x_j'=\begin{cases}x_j-\alpha\lambda_j & \text{for } j=1,\ldots,k\\ 0 & \text{for } j=k+1,\ldots,n\end{cases}$$

Note $x_j'\geq 0$ for $j=1,\ldots,k$ and $x_j'=0$ for $j=k+1,\ldots,n$. Moreover, $x_i'=0$, and

$$\sum_{j=1}^{n}\mathbf{a}_jx_j'=\sum_{j=1}^{k}\mathbf{a}_j(x_j-\alpha\lambda_j)=\sum_{j=1}^{k}\mathbf{a}_jx_j-\alpha\sum_{j=1}^{k}\mathbf{a}_j\lambda_j=\mathbf{b}-\mathbf{0}=\mathbf{b}$$

So far, we have constructed a new point $\mathbf{x}'$ with, at most, $k-1$ positive components. The process is continued until the positive components correspond to linearly independent columns, which results in an extreme point. Thus, we have shown that S has at least one extreme point, and the proof is complete.

Extreme Directions Let $S=\{\mathbf{x}:\mathbf{A}\mathbf{x}=\mathbf{b},\mathbf{x}\geq\mathbf{0}\}$, where $\mathbf{A}$ is an $m\times n$ matrix of rank m. By definition, a nonzero vector $\mathbf{d}$ is a direction of S if $\mathbf{x}+\lambda\mathbf{d}\in S$ for each $\mathbf{x}\in S$ and each $\lambda\geq 0$. Noting the structure of S, it is clear that $\mathbf{d}\neq\mathbf{0}$ is a direction of S if and only if

$$\mathbf{A}\mathbf{d}=\mathbf{0},\qquad \mathbf{d}\geq\mathbf{0}$$

In particular, we are interested in the characterization of extreme directions of S.

2.5.6 Theorem (Characterization of Extreme Directions)

Let $S=\{\mathbf{x}:\mathbf{A}\mathbf{x}=\mathbf{b},\mathbf{x}\geq\mathbf{0}\}$, where $\mathbf{A}$ is an $m\times n$ matrix of rank m, and $\mathbf{b}$ is an m vector. A vector $\bar{\mathbf{d}}$ is an extreme direction of S if and only if $\mathbf{A}$ can be

decomposed into $[\mathbf{B},\mathbf{N}]$ such that $\mathbf{B}^{-1}\mathbf{a}_j\leq\mathbf{0}$ for some column $\mathbf{a}_j$ of $\mathbf{N}$, and $\bar{\mathbf{d}}$ is a positive multiple of $\mathbf{d}=\begin{pmatrix}-\mathbf{B}^{-1}\mathbf{a}_j\\ \mathbf{e}_j\end{pmatrix}$, where $\mathbf{e}_j$ is an $n-m$ vector of zeros except for a 1 in position j.

Proof

If $\mathbf{B}^{-1}\mathbf{a}_j\leq\mathbf{0}$, then $\mathbf{d}\geq\mathbf{0}$. Furthermore, $\mathbf{A}\mathbf{d}=\mathbf{0}$, so that $\mathbf{d}$ is a direction of S. We now show that $\mathbf{d}$ is indeed an extreme direction. Suppose that $\mathbf{d}=\lambda_1\mathbf{d}_1+\lambda_2\mathbf{d}_2$, where $\lambda_1,\lambda_2>0$ and $\mathbf{d}_1,\mathbf{d}_2$ are directions of S. Noting that $n-m-1$ components of $\mathbf{d}$ are equal to zero, then the corresponding components of $\mathbf{d}_1$ and $\mathbf{d}_2$ must also be equal to zero. Thus, $\mathbf{d}_1$ and $\mathbf{d}_2$ could be written as follows:

$$\mathbf{d}_1=\alpha_1\begin{pmatrix}\mathbf{d}_{11}\\ \mathbf{e}_j\end{pmatrix}\qquad \mathbf{d}_2=\alpha_2\begin{pmatrix}\mathbf{d}_{21}\\ \mathbf{e}_j\end{pmatrix}$$

where $\alpha_1,\alpha_2>0$. Noting that $\mathbf{A}\mathbf{d}_1=\mathbf{A}\mathbf{d}_2=\mathbf{0}$, it can easily be verified that $\mathbf{d}_{11}=\mathbf{d}_{21}=-\mathbf{B}^{-1}\mathbf{a}_j$. Thus, $\mathbf{d}_1$ and $\mathbf{d}_2$ are not distinct, which implies that $\mathbf{d}$ is an extreme direction. Since $\bar{\mathbf{d}}$ is a positive multiple of $\mathbf{d}$, it is also an extreme direction.

Conversely, suppose that $\bar{\mathbf{d}}$ is an extreme direction of S. Without loss of generality, suppose that

$$\bar{\mathbf{d}}=(\bar{d}_1,\ldots,\bar{d}_k,0,\ldots,\bar{d}_j,\ldots,0)^t$$

where $\bar{d}_i>0$ for $i=1,\ldots,k$ and for $i=j$. We claim that $\mathbf{a}_1,\ldots,\mathbf{a}_k$ are linearly independent. By contradiction, suppose that this were not the case. Then there would exist scalars $\lambda_1,\ldots,\lambda_k$ not all zero such that $\sum_{i=1}^k\lambda_i\mathbf{a}_i=\mathbf{0}$. Let $\boldsymbol{\lambda}=(\lambda_1,\ldots,\lambda_k,0,\ldots,0)^t$ and choose $\alpha>0$ sufficiently small such that both

$$\mathbf{d}_1=\bar{\mathbf{d}}+\alpha\boldsymbol{\lambda}\qquad\text{and}\qquad \mathbf{d}_2=\bar{\mathbf{d}}-\alpha\boldsymbol{\lambda}$$

are nonnegative. Note that

$$\mathbf{A}\mathbf{d}_1=\mathbf{A}\bar{\mathbf{d}}+\alpha\mathbf{A}\boldsymbol{\lambda}=\mathbf{0}+\alpha\sum_{i=1}^{k}\mathbf{a}_i\lambda_i=\mathbf{0}$$

Similarly $\mathbf{A}\mathbf{d}_2=\mathbf{0}$. Since $\mathbf{d}_1,\mathbf{d}_2\geq 0$, they are both directions of S. Note also that they are distinct, since $\alpha>0$ and $\boldsymbol{\lambda}\neq\mathbf{0}$. Furthermore, $\bar{\mathbf{d}}=\frac{1}{2}\mathbf{d}_1+\frac{1}{2}\mathbf{d}_2$, contradicting the assumption that $\bar{\mathbf{d}}$ is an extreme direction. Thus, $\mathbf{a}_1,\ldots,\mathbf{a}_k$ are linearly independent, and since rank $\mathbf{A}$ is equal to m, it is clear that $k\leq m$. Then there must exist $m-k$ vectors from among the set of vectors $\{\mathbf{a}_i:i=k+1,\ldots,n;\,i\neq j\}$ which, together with $\mathbf{a}_1,\ldots,\mathbf{a}_k$, form a linearly independent set of vectors. Without loss of generality, suppose that these are $\mathbf{a}_1,\ldots,\mathbf{a}_m$. Denote $[\mathbf{a}_1,\ldots,\mathbf{a}_m]$ by $\mathbf{B}$, and note that $\mathbf{B}$ is invertible. Thus, $\mathbf{0}=\mathbf{A}\bar{\mathbf{d}}=\mathbf{B}\hat{\mathbf{d}}+\mathbf{a}_j\bar{d}_j$, where $\hat{\mathbf{d}}$ is the first m components of $\bar{\mathbf{d}}$. Therefore, $\hat{\mathbf{d}}=-\bar{d}_j\mathbf{B}^{-1}\mathbf{a}_j$, and hence the vector

$\bar{\mathbf{d}}$ is of the form $\bar{\mathbf{d}} = \bar{d}_j\begin{pmatrix} -\mathbf{B}^{-1}\mathbf{a}_j \\ \mathbf{e}_j \end{pmatrix}$. Noting that $\bar{\mathbf{d}} \geq \mathbf{0}$ and that $\bar{d}_j > 0$, then $\mathbf{B}^{-1}\mathbf{a}_j \leq \mathbf{0}$, and the proof is complete.

Corollary

The number of extreme directions of S is finite.

Proof

For each choice of a matrix $\mathbf{B}$ from $\mathbf{A}$, there is $n-m$ possible ways to extract the column $\mathbf{a}_j$ from $\mathbf{N}$. Therefore, the maximum number of extreme directions is bounded by $(n-m)\binom{n}{m} = \dfrac{n!}{m!\,(n-m-1)!}$.

The Representation of Polyhedral Sets in Terms of Extreme Points and Extreme Directions

By definition, a polyhedral set is the intersection of a finite number of half spaces. This representation may be thought of as an *outer representation*. A polyhedral set can also be described fully by an *inner representation* by means of its extreme points and extreme directions. This fact is fundamental to several linear and nonlinear programming procedures.

The main result can be stated as follows. Let S be a polyhedral set of the form $\{\mathbf{x}: \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$. Then, any point in S can be represented as a convex combination of its extreme points plus a nonnegative linear combination of its extreme directions. Of course, if S is bounded, then it contains no directions, and so any point in S can be described as a convex combination of its extreme points.

In Theorem 2.5.7 below, it is implicitly assumed that the extreme points and the extreme directions of S are finite in number. This fact follows from the corollaries to Theorems 2.5.4 and 2.5.6.

2.5.7 Theorem (Representation Theorem)

Let S be a nonempty polyhedral set in E_n of the form $\{\mathbf{x}: \mathbf{A}\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}\}$, where $\mathbf{A}$ is an $m \times n$ matrix with rank m. Let $\mathbf{x}_1, \ldots, \mathbf{x}_k$ be the extreme points of S and $\mathbf{d}_1, \ldots, \mathbf{d}_l$ be the extreme directions of S. Then $\mathbf{x} \in S$ if and only if $\mathbf{x}$ can be written as

$$\mathbf{x} = \sum_{j=1}^{k} \lambda_j \mathbf{x}_j + \sum_{j=1}^{l} \mu_j \mathbf{d}_j$$

$$\sum_{j=1}^{k} \lambda_j = 1 \tag{2.6}$$

$$\lambda_j \geq 0 \quad \text{for } j = 1, \ldots, k \tag{2.7}$$

$$\mu_j \geq 0 \quad \text{for } j = 1, \ldots, l \tag{2.8}$$

Proof

Construct the following set:

$$\Lambda = \left\{\sum_{j=1}^{k} \lambda_j \mathbf{x}_j + \sum_{j=1}^{l} \mu_j \mathbf{d}_j : \sum_{j=1}^{k} \lambda_j = 1, \lambda_j \geq 0 \text{ for all } j, \mu_j \geq 0 \text{ for all } j\right\}$$

Note that Λ is a closed convex set. Furthermore, by Theorem 2.5.5, S has at least one extreme point, and hence Λ is not empty. Also note that $\Lambda \subset S$. To show that $S \subset \Lambda$, suppose by contradiction that there is a $\mathbf{z} \in S$ such that $\mathbf{z} \notin \Lambda$. By Theorem 2.3.4, there exists a scalar α and a nonzero vector $\mathbf{p}$ in E_n such that

$$\mathbf{p}^t\mathbf{z} > \alpha$$

$$\mathbf{p}^t\left(\sum_{j=1}^{k} \lambda_j \mathbf{x}_j + \sum_{j=1}^{l} \mu_j \mathbf{d}_j\right) \leq \alpha \tag{2.9}$$

for λ_j's and μ_j's, satisfying (2.6), (2.7), and (2.8). Since μ_j can be made arbitrarily large, (2.9) holds true only if $\mathbf{p}^t\mathbf{d}_j \leq 0$ for $j = 1, \ldots, l$. From (2.9), by letting $\mu_j = 0$ for all j, $\lambda_j = 1$, and $\lambda_i = 0$ for $i \neq j$, it follows that $\mathbf{p}^t\mathbf{x}_j \leq \alpha$ for each $j = 1, \ldots, k$. Since $\mathbf{p}^t\mathbf{z} > \alpha$, we have $\mathbf{p}^t\mathbf{z} > \mathbf{p}^t\mathbf{x}_j$ for all j. Summarizing, there exists a nonzero vector $\mathbf{p}$ such that

$$\mathbf{p}^t\mathbf{z} > \mathbf{p}^t\mathbf{x}_j \quad \text{for } j = 1, \ldots, k \tag{2.10}$$

$$\mathbf{p}^t\mathbf{d}_j \leq 0 \quad \text{for } j = 1, \ldots, l \tag{2.11}$$

Consider the extreme point $\bar{\mathbf{x}}$ defined as follows:

$$\mathbf{p}^t\bar{\mathbf{x}} = \underset{1 \leq j \leq k}{\text{maximum}}\ \mathbf{p}^t\mathbf{x}_j \tag{2.12}$$

Since $\bar{\mathbf{x}}$ is an extreme point, by Theorem 2.5.4, $\bar{\mathbf{x}} = \begin{pmatrix} \mathbf{B}^{-1}\mathbf{b} \\ \mathbf{0} \end{pmatrix}$ where $\mathbf{A} = [\mathbf{B}, \mathbf{N}]$ and $\mathbf{B}^{-1}\mathbf{b} \geq \mathbf{0}$. Without loss of generality assume that $\mathbf{B}^{-1}\mathbf{b} > \mathbf{0}$ (see Exercise 2.42). Since $\mathbf{z} \in S$, then $\mathbf{A}\mathbf{z} = \mathbf{b}$ and $\mathbf{z} \geq \mathbf{0}$. Therefore, $\mathbf{B}\mathbf{z}_B + \mathbf{N}\mathbf{z}_N = \mathbf{b}$ and hence $\mathbf{z}_B = \mathbf{B}^{-1}\mathbf{b} - \mathbf{B}^{-1}\mathbf{N}\mathbf{z}_N$, where $\mathbf{z}^t$ is decomposed into $(\mathbf{z}_B^t, \mathbf{z}_N^t)$. From (2.10), we have $\mathbf{p}^t\mathbf{z} - \mathbf{p}^t\bar{\mathbf{x}} > 0$, and decomposing $\mathbf{p}^t$ into $(\mathbf{p}_B^t, \mathbf{p}_N^t)$, we get

$$\begin{aligned} 0 &< \mathbf{p}^t\mathbf{z} - \mathbf{p}^t\bar{\mathbf{x}} \\ &= \mathbf{p}_B^t(\mathbf{B}^{-1}\mathbf{b} - \mathbf{B}^{-1}\mathbf{N}\mathbf{z}_N) + \mathbf{p}_N^t\mathbf{z}_N - \mathbf{p}_B^t\mathbf{B}^{-1}\mathbf{b} \\ &= (\mathbf{p}_N^t - \mathbf{p}_B^t\mathbf{B}^{-1}\mathbf{N})\mathbf{z}_N \end{aligned} \tag{2.13}$$

Since $\mathbf{z}_N \geq \mathbf{0}$, from (2.13) it follows that there is a component $j \geq m+1$ such that $z_j > 0$ and $p_j - \mathbf{p}_B^t\mathbf{B}^{-1}\mathbf{a}_j > 0$. We first show that $\mathbf{y}_j = \mathbf{B}^{-1}\mathbf{a}_j \nleq \mathbf{0}$. By contradiction, suppose that $\mathbf{y}_j \leq \mathbf{0}$. Consider the vector $\mathbf{d}_j = \begin{bmatrix} -\mathbf{y}_j \\ \mathbf{e}_j \end{bmatrix}$, where $\mathbf{e}_j$ is an $n-m$

dimensional unit vector with 1 at position j. By Theorem 2.5.6, $\mathbf{d}_j$ is an extreme direction of S. From (2.11) $\mathbf{p}'\mathbf{d}_j \le 0$, that is, $-\mathbf{p}_B^t\mathbf{B}^{-1}\mathbf{a}_j + p_j \le 0$, which contradicts the assertion that $p_j - \mathbf{p}_B^t\mathbf{B}^{-1}\mathbf{a}_j > 0$. Therefore $\mathbf{y}_j \not\le \mathbf{0}$, and we can construct the following vector:

$$\mathbf{x} = \begin{pmatrix} \bar{\mathbf{b}} \\ \mathbf{0} \end{pmatrix} + \lambda \begin{pmatrix} -\mathbf{y}_j \\ \mathbf{e}_j \end{pmatrix}$$

where $\bar{\mathbf{b}}$ is given by $\mathbf{B}^{-1}\mathbf{b}$, and λ is given by

$$\lambda = \underset{1\le i\le m}{\text{minimum}} \left\{ \frac{\bar{b}_i}{y_{ij}} : y_{ij} > 0 \right\} = \frac{\bar{b}_r}{y_{rj}} > 0$$

Note that $\mathbf{x} \ge \mathbf{0}$ has, at most, m positive components, where the rth component drops to zero and the jth component is given by λ. The vector $\mathbf{x}$ belongs to S, since $\mathbf{A}\mathbf{x} = \mathbf{B}(\mathbf{B}^{-1}\mathbf{b} - \lambda\mathbf{B}^{-1}\mathbf{a}_j) + \lambda\mathbf{a}_j = \mathbf{b}$. Since $y_{rj} \ne 0$, it can be shown that the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_{r-1}, \mathbf{a}_{r+1}, \ldots, \mathbf{a}_m, \mathbf{a}_j$ are linearly independent. Therefore, by Theorem 2.5.4, $\mathbf{x}$ is an extreme point; that is, $\mathbf{x} \in \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$. Furthermore,

$$\begin{aligned} \mathbf{p}'\mathbf{x} &= (\mathbf{p}_B^t, \mathbf{p}_N^t)\begin{pmatrix} \bar{\mathbf{b}} - \lambda\mathbf{y}_j \\ \lambda\mathbf{e}_j \end{pmatrix} \\ &= \mathbf{p}_B^t\bar{\mathbf{b}} - \lambda\mathbf{p}_B^t\mathbf{y}_j + \lambda p_j \\ &= \mathbf{p}'\bar{\mathbf{x}} + \lambda(p_j - \mathbf{p}_B^t\mathbf{B}^{-1}\mathbf{a}_j) \end{aligned}$$

Since $\lambda > 0$ and $p_j - \mathbf{p}_B^t\mathbf{B}^{-1}\mathbf{a}_j > 0$, then $\mathbf{p}'\mathbf{x} > \mathbf{p}'\bar{\mathbf{x}}_j$. Thus, we have constructed an extreme point $\mathbf{x}$ such that $\mathbf{p}'\mathbf{x} > \mathbf{p}'\bar{\mathbf{x}}$, which contradicts (2.12). This contradiction asserts that $\mathbf{z}$ must belong to Λ, and the proof is complete.

Corollary (Existence of Extreme Directions)

Let S be a nonempty polyhedral set of the form $\{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \ge \mathbf{0}\}$ where $\mathbf{A}$ is an $m \times n$ matrix with rank m. Then S has at least one extreme direction if and only if it is unbounded.

Proof

If S has an extreme direction, then it is obviously unbounded. Now suppose that S is unbounded, and by contradiction, suppose that S has no extreme directions. Using the theorem and Schwartz inequality, it follows that

$$\|\mathbf{x}\| = \left\| \sum_{j=1}^{k} \lambda_j \mathbf{x}_j \right\| \le \sum_{j=1}^{k} \lambda_j \|\mathbf{x}_j\| \le \sum_{j=1}^{k} \|\mathbf{x}_j\|$$

for any $\mathbf{x} \in S$. However, this violates the unboundedness assumption. Therefore, S has at least one extreme direction and the proof is complete.

2.6 Linear Programming and the Simplex Method

A linear programming problem is the minimization or the maximization of a linear function over a polyhedral set. Many problems can be formulated as, or approximated by, linear programs. Also, linear programming is often used in the process of solving nonlinear and discrete problems. In this section, we describe the well-known simplex method for solving linear programming problems. The method is mainly based on exploiting the extreme points and directions of the polyhedral set defining the problem.

Consider the following linear programming problem:

$$\begin{aligned} &\text{Minimize} && \mathbf{c}'\mathbf{x} \\ &\text{subject to} && \mathbf{x} \in S \end{aligned}$$

where S is a polyhedral set in E_n. The set S is called the *constraint set* or the *feasible region*, and the linear function $\mathbf{c}'\mathbf{x}$ is called the *objective function*.

The optimum objective function value of a linear programming problem may be finite or unbounded. We give below a necessary and sufficient condition for a finite optimal solution. The importance of the concepts of extreme points and extreme directions in linear programming will be evident from the theorem.

2.6.1 Theorem (Optimality Conditions in Linear Programming)

Consider the following linear programming problem: Minimize $\mathbf{c}'\mathbf{x}$, subject to $\mathbf{A}\mathbf{x} = \mathbf{b}$, $\mathbf{x} \ge \mathbf{0}$. Here, $\mathbf{c}$ is an n vector, $\mathbf{A}$ is an $m \times n$ matrix of rank m, and $\mathbf{b}$ is an m vector. Suppose that the feasible region is not empty, and let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ be the extreme points and $\mathbf{d}_1, \ldots, \mathbf{d}_l$ be the extreme directions of the feasible region. A necessary and sufficient condition for a finite optimal solution is that $\mathbf{c}'\mathbf{d}_j \ge 0$ for $j = 1, \ldots, l$. If this were the case, then there exists an extreme point $\mathbf{x}_i$ that solves the problem.

Proof

By Theorem 2.5.7, $\mathbf{A}\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \ge \mathbf{0}$ if and only if

$$\begin{aligned} &\mathbf{x} = \sum_{j=1}^{k} \lambda_j \mathbf{x}_j + \sum_{j=1}^{l} \mu_j \mathbf{d}_j \\ &\sum_{j=1}^{k} \lambda_j = 1 \\ &\lambda_j \ge 0 \qquad \text{for } j = 1, \ldots, k \\ &\mu_j \ge 0 \qquad \text{for } j = 1, \ldots, l \end{aligned}$$

Therefore the linear programming problem can be stated as follows:

$$\begin{aligned}
\text{Minimize} \quad & \mathbf{c}^t\left(\sum_{j=1}^{k} \lambda_j \mathbf{x}_j + \sum_{j=1}^{l} \mu_j \mathbf{d}_j\right)\\
\text{subject to} \quad & \sum_{j=1}^{k} \lambda_j = 1\\
& \lambda_j \geq 0 \qquad \text{for } j=1,\ldots,k\\
& \mu_j \geq 0 \qquad \text{for } j=1,\ldots,l
\end{aligned}$$

Note that if $\mathbf{c}^t\mathbf{d}_j < 0$ for some j, then μ_j can be chosen arbitrarily large, leading to an unbounded optimal objective value. This shows that a necessary and sufficient condition for a finite optimal is $\mathbf{c}^t\mathbf{d}_j \geq 0$ for $j=1,\ldots,l$. If this were the case, then in order to minimize the objective function, we may choose $\mu_j = 0$ for $j=1,\ldots,l$, and the problem reduces to minimizing $\mathbf{c}^t(\sum_{j=1}^k \lambda_j \mathbf{x}_j)$ subject to $\sum_{j=1}^k \lambda_j = 1$ and $\lambda_j \geq 0$ for $j=1,\ldots,k$. It is clear that the optimal solution to this latter problem is finite and found by letting $\lambda_i = 1$ and $\lambda_j = 0$ for $j \neq i$, where the index i is given by $\mathbf{c}^t\mathbf{x}_i = \text{minimum}_{1\leq j\leq k}\, \mathbf{c}^t\mathbf{x}_j$. Thus, there exists an optimal extreme point, and the proof is complete.

From the above theorem, at least for the case in which the feasible region is bounded, one may be tempted to calculate $\mathbf{c}^t\mathbf{x}_j$ for $j=1,\ldots,k$ and then find $\text{minimum}_{1\leq j\leq k}\, \mathbf{c}^t\mathbf{x}_j$. Even though this is theoretically possible, it is computationally not feasible because the number of extreme points is usually large.

The Simplex Method

The simplex method is a systematic procedure for solving a linear programming problem by moving from an extreme point to an extreme point with a better (at least not worse) objective function value. This process continues until an optimal extreme point is reached or else until an extreme direction $\mathbf{d}$ with $\mathbf{c}^t\mathbf{d} < 0$ is found. In the latter case, we conclude that the optimal objective value is unbounded.

Consider the following linear programming problem in which the polyhedral set is defined in terms of equations and variables that are restricted to be nonnegative.

$$\begin{aligned}
\text{Minimize} \quad & \mathbf{c}^t\mathbf{x}\\
\text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}\\
& \mathbf{x} \geq \mathbf{0}
\end{aligned}$$

Note that any polyhedral set can be put in the above *standard format*. For example, an inequality of the form $\sum_{j=1}^n a_{ij}x_j \leq b_i$ can be transformed into an

equation by adding the nonnegative *slack variable* s_i, so that $\sum_{j=1}^n a_{ij}x_j + s_i = b_i$. Also, an unrestricted variable x_j can be replaced by the difference of two nonnegative variables; that is, $x_j = x_j^+ - x_j^-$, where $x_j^+, x_j^- \geq 0$. These and other manipulations could be used to put the problem in the above format. We shall assume for the time being that the constraint set admits at least one feasible point and that the rank of $\mathbf{A}$ is equal to m.

By Theorem 2.6.1, at least in the case of a finite optimal solution, it suffices to concentrate on extreme points. Suppose that we have an extreme point $\bar{\mathbf{x}}$. By Theorem 2.5.4, this point is characterized by a decomposition of $\mathbf{A}$ into $[\mathbf{B}, \mathbf{N}]$, where $\mathbf{B} = [\mathbf{a}_{B_1}, \ldots, \mathbf{a}_{B_m}]$ is an $m \times m$ matrix of full rank called the *basis*, and $\mathbf{N}$ is an $m \times n - m$ matrix. By Theorem 2.5.4, note that $\bar{\mathbf{x}}$ could be written as $\bar{\mathbf{x}}^t = (\bar{\mathbf{x}}_B^t, \bar{\mathbf{x}}_N^t) = (\bar{\mathbf{b}}^t, \mathbf{0}^t)$, where $\bar{\mathbf{b}} = \mathbf{B}^{-1}\mathbf{b} \geq \mathbf{0}$. The variables corresponding to the basis $\mathbf{B}$ are called *basic variables* and are denoted by $x_{B_1}, \ldots, x_{B_m}$, whereas the variables corresponding to $\mathbf{N}$ are called *nonbasic variables*. Now let us consider a point $\mathbf{x}$ satisfying $\mathbf{A}\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$. Decompose $\mathbf{x}^t$ into $(\mathbf{x}_B^t, \mathbf{x}_N^t)$ and note that $\mathbf{x}_B, \mathbf{x}_N \geq \mathbf{0}$. Also $\mathbf{A}\mathbf{x} = \mathbf{b}$ can be written as $\mathbf{B}\mathbf{x}_B + \mathbf{N}\mathbf{x}_N = \mathbf{b}$. Hence,

$$\mathbf{x}_B = \mathbf{B}^{-1}\mathbf{b} - \mathbf{B}^{-1}\mathbf{N}\mathbf{x}_N \tag{2.14}$$

Then, using (2.14),

$$\begin{aligned}
\mathbf{c}^t\mathbf{x} &= \mathbf{c}_B^t\mathbf{x}_B + \mathbf{c}_N^t\mathbf{x}_N\\
&= \mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{b} + (\mathbf{c}_N^t - \mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{N})\mathbf{x}_N\\
&= \mathbf{c}^t\bar{\mathbf{x}} + (\mathbf{c}_N^t - \mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{N})\mathbf{x}_N
\end{aligned} \tag{2.15}$$

Hence $\mathbf{c}^t\mathbf{x} \geq \mathbf{c}^t\bar{\mathbf{x}}$ if $\mathbf{c}_N^t - \mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{N} \geq \mathbf{0}$, since $\mathbf{x}_N \geq \mathbf{0}$, and $\bar{\mathbf{x}}$ is an optimal extreme point. On the other hand, suppose $\mathbf{c}_N^t - \mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{N} \ngeq \mathbf{0}$. In particular, suppose that the jth component $c_j - \mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{a}_j$ is negative. Consider $\mathbf{x} = \bar{\mathbf{x}} + \lambda\mathbf{d}_j$, where

$$\mathbf{d}_j = \begin{pmatrix} -\mathbf{B}^{-1}\mathbf{a}_j \\ \mathbf{e}_j \end{pmatrix}$$

where $\mathbf{e}_j$ is an $n-m$ unit vector with a 1 at position j. Then, from (2.15),

$$\mathbf{c}^t\mathbf{x} = \mathbf{c}^t\bar{\mathbf{x}} + \lambda(c_j - \mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{a}_j) \tag{2.16}$$

and $\mathbf{c}^t\mathbf{x} < \mathbf{c}^t\bar{\mathbf{x}}$ for $\lambda > 0$, since $c_j - \mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{a}_j < 0$. We now consider the following two cases, where $\mathbf{y}_j = \mathbf{B}^{-1}\mathbf{a}_j$.

Case 1: $\mathbf{y}_j \leq \mathbf{0}$. Note that $\mathbf{A}\mathbf{d}_j = \mathbf{0}$ and since $\mathbf{A}\bar{\mathbf{x}} = \mathbf{b}$, then $\mathbf{A}\mathbf{x} = \mathbf{b}$ for $\mathbf{x} = \bar{\mathbf{x}} + \lambda\mathbf{d}_j$ and for all values of λ. Hence, $\mathbf{x}$ is feasible if and only if $\mathbf{x} \geq \mathbf{0}$. This obviously holds true for all $\lambda \geq 0$ if $\mathbf{y}_j \leq \mathbf{0}$. Thus, from (2.16), the objective function value is unbounded. In this case we have found an extreme direction $\mathbf{d}_j$ with $\mathbf{c}^t\mathbf{d}_j = c_j - \mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{a}_j < 0$ (see Theorems 2.6.1 and 2.5.6).

Case 2: $\mathbf{y}_j \not\leq \mathbf{0}$. Let $\mathbf{B}^{-1}\mathbf{b} = \bar{\mathbf{b}}$, and let λ be defined by

$$\lambda = \underset{1 \le i \le m}{\text{minimum}} \left\{ \frac{\bar{b}_i}{y_{ij}} : y_{ij} > 0 \right\} = \frac{\bar{b}_r}{y_{rj}} \ge 0 \tag{2.17}$$

where y_{ij} is the ith component of $\mathbf{y}_j$. In this case the components of $\mathbf{x} = \bar{\mathbf{x}} + \lambda \mathbf{d}_j$ are given by

$$\begin{aligned} x_{B_i} &= \bar{b}_i - \frac{\bar{b}_r}{y_{rj}} y_{ij} \qquad \text{for } i = 1, \ldots, m \\ x_j &= \bar{b}_r / y_{rj} \\ &\text{all other } x_i\text{'s are equal to zero.} \end{aligned} \tag{2.18}$$

The positive components of $\mathbf{x}$ can only be $x_{B_1}, \ldots, x_{B_{r-1}}, x_{B_{r+1}}, \ldots, x_{B_m}$ and x_j. Hence, at most, m components of $\mathbf{x}$ are positive. It is easy to verify that their corresponding columns in $\mathbf{A}$ are linearly independent. Therefore, by Theorem 2.5.4, the point $\mathbf{x}$ is itself an extreme point. In this case we say that the basic variable x_{B_r} left the basis and the nonbasic variable x_j entered instead.

So far we have shown that, given an extreme point, we can check its optimality and stop, or find an extreme direction leading to an unbounded solution, or find an extreme point with a better objective value. The process is then repeated.

Summary of the Simplex Algorithm

Outlined below is a summary of the simplex algorithm for a minimization problem of the form to minimize $\mathbf{c}^t\mathbf{x}$ subject to $\mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \ge \mathbf{0}$. A maximization problem can be either transformed into a minimization problem or else we have to modify step 1 such that we stop if $\mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{N} - \mathbf{c}_N^t \ge \mathbf{0}$, and introduce x_j into the basis if $\mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{a}_j - c_j < 0$.

Initialization Step Find a starting extreme point $\mathbf{x}$ with basis $\mathbf{B}$. If such a point is not readily available, then use artificial variables as discussed later in the section.

Main Step 1. Let $\mathbf{x}$ be an extreme point with basis $\mathbf{B}$. Calculate $\mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{N} - \mathbf{c}_N^t$. If this vector is nonpositive, then stop; $\mathbf{x}$ is an optimal extreme point. Otherwise pick the most positive component $\mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{a}_j - c_j$. If $\mathbf{y}_j = \mathbf{B}^{-1}\mathbf{a}_j \le \mathbf{0}$, then stop; the optimal objective value is unbounded along the *ray*

$$\left\{ \mathbf{x} + \lambda \begin{pmatrix} -\mathbf{y}_j \\ \mathbf{e}_j \end{pmatrix} : \lambda \ge 0 \right\}$$

where $\mathbf{e}_j$ is a vector of zeros except for a 1 in position j. If on the other hand, $\mathbf{y}_j \not\leq \mathbf{0}$, then go to step 2.

2. Compute the index r from (2.17) and form the new extreme point $\mathbf{x}$ in (2.18). Form the new basis by deleting the column $\mathbf{a}_{B_r}$ from $\mathbf{B}$ and introducing $\mathbf{a}_j$ instead. Repeat step 1.

Finite Convergence of the Simplex Method

If at each iteration, that is, one pass through the main step, we have $\bar{\mathbf{b}} = \mathbf{B}^{-1}\mathbf{b} > \mathbf{0}$, then λ, defined by (2.17), would be strictly positive, and the objective value at the current extreme point would be strictly less than that at any of the previous iterations. This would imply that the current point is distinct from those previously generated. Since we have a finite number of extreme points, the simplex algorithm must stop in a finite number of iterations. If, on the other hand, $\bar{b}_r = 0$, then $\lambda = 0$, and we would remain at the same extreme point but with a different basis. In theory, this could happen an infinite number of times and may cause nonconvergence. This phenomena is called *cycling* and rarely occurs in practice. The problem of cycling can be overcome, but this topic will not be discussed here. Most textbooks on linear programming give detailed procedures for avoiding cycling.

Tableau Format of the Simplex Method

Suppose that we have the starting basis $\mathbf{B}$ corresponding to an initial extreme point. The objective function and the constraints could be written as

$$\begin{array}{ll} \text{Objective row:} & f - \mathbf{c}_B^t\mathbf{x}_B - \mathbf{c}_N^t\mathbf{x}_N = 0 \\ \text{Constraint rows:} & \mathbf{B}\mathbf{x}_B + \mathbf{N}\mathbf{x}_N = \mathbf{b} \end{array}$$

These equations could be displayed in the following *simplex tableau* where the entries in the *RHS* column are the right-hand-side constants.

f	$\mathbf{x}_B^t$	$\mathbf{x}_N^t$	RHS
1	$-\mathbf{c}_B^t$	$-\mathbf{c}_N^t$	0
$\mathbf{0}$	$\mathbf{B}$	$\mathbf{N}$	$\mathbf{b}$

The constraint rows are updated by multiplying by $\mathbf{B}^{-1}$, and the objective row is updated by adding to it $\mathbf{c}_B^t$ times the new constraint rows. We then get the following updated tableau. Note that the basic variables are indicated on the left-hand side.

	f	$\mathbf{x}_B^t$	$\mathbf{x}_N^t$	RHS
f	1	$\mathbf{0}^t$	$\mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{N} - \mathbf{c}_N^t$	$\mathbf{c}_B^t\bar{\mathbf{b}}$
$\mathbf{x}_B$	$\mathbf{0}$	$\mathbf{I}$	$\mathbf{B}^{-1}\mathbf{N}$	$\bar{\mathbf{b}}$

Note that the values of the basic variables and that of f are recorded on the right-hand side of the tableau. Also, the vector $\mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{N}-\mathbf{c}_N^t$ and the matrix $\mathbf{B}^{-1}\mathbf{N}$ are conveniently stored under the nonbasic variables.

The above tableau displays all the information needed to perform step 1 of the simplex method. If $\mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{N}-\mathbf{c}_N^t\leq\mathbf{0}$, then we stop; the current extreme point is optimal. Otherwise, upon examining the objective row, we can pick a nonbasic variable with negative $\mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{a}_j-c_j$. If $\mathbf{B}^{-1}\mathbf{a}_j\leq\mathbf{0}$, then we stop; the optimal solution is unbounded. Now suppose that $\mathbf{y}_j=\mathbf{B}^{-1}\mathbf{a}_j\nleq\mathbf{0}$. Since $\bar{\mathbf{b}}$ and $\mathbf{y}_j$ are recorded under RHS and x_j, respectively, then λ in (2.17) can be easily calculated from the tableau. The basic variable x_{B_r} corresponding to the minimum ratio of (2.17) leaves the basis and x_j enters the basis.

Now we would like to update the tableau to reflect the new basis. This can be done by *pivoting* at the x_{B_r} row and the x_j column, that is, at y_{rj}, as follows:

1. Divide the rth row corresponding to x_{B_r} by y_{rj}.
2. Multiply the new rth row by y_{ij} and subtract from the ith constraint row, for $i=1,\ldots,m$, $i\neq r$.
3. Multiply the new rth row by $\mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{a}_j-c_j$ and subtract from the objective row.

The reader can easily verify that the above pivoting operation will update the tableau to reflect the new basis (see Exercise 2.48).

2.6.2 Example

$$\begin{array}{ll}\text{Minimize} & x_1-3x_2\\ \text{subject to} & -x_1+2x_2\leq 6\\ & x_1+x_2\leq 5\\ & x_1,\quad x_2\geq 0\end{array}$$

The problem is illustrated in Figure 2.17. It is clear that the optimal point is $(\frac{4}{3},\frac{11}{3})$ and that the corresponding value of the objective function is $-\frac{29}{3}$.

In order to use the simplex method, we now introduce the two *slack variables* x_3 and $x_4\geq 0$. This leads to the following standard format.

$$\begin{array}{ll}\text{Minimize} & x_1-3x_2\\ \text{subject to} & -x_1+2x_2+x_3 \qquad =6\\ & x_1+x_2 \qquad +x_4=5\\ & x_1,\quad x_2,\quad x_3,\quad x_4\geq 0\end{array}$$

Note that $\mathbf{c}=(1,-3,0,0)^t$, $\mathbf{b}=\begin{pmatrix}6\\5\end{pmatrix}$, and $\mathbf{A}=\begin{bmatrix}-1 & 2 & 1 & 0\\ 1 & 1 & 0 & 1\end{bmatrix}$. By choosing

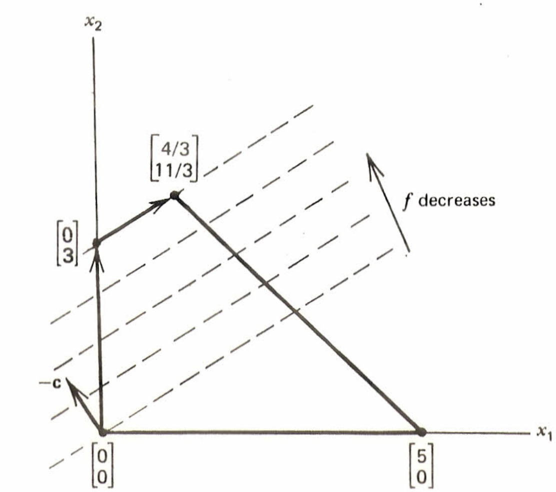


Figure 2.17 A linear programming example.

$\mathbf{B}=[a_3,a_4]=\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$, we note that $\mathbf{B}^{-1}\mathbf{b}=\mathbf{b}\geq\mathbf{0}$, and hence we have a starting extreme point. The corresponding tableau is displayed below.

	f	x_1	x_2	x_3	x_4	RHS
f	1	-1	3	0	0	0
x_3	0	-1	2	1	0	6
x_4	0	1	1	0	1	5

Note that x_2 enters and x_3 leaves the basis. The new basis $\mathbf{B}=[\mathbf{a}_2,\mathbf{a}_4]$.

	f	x_1	x_2	x_3	x_4	RHS
f	1	$\frac{1}{2}$	0	$-\frac{3}{2}$	0	-9
x_2	0	$-\frac{1}{2}$	1	$\frac{1}{2}$	0	3
x_4	0	$\frac{3}{2}$	0	$-\frac{1}{2}$	1	2

Now x_1 enters and x_4 leaves the basis. The new basis $\mathbf{B}=[\mathbf{a}_2,\mathbf{a}_1]$.

	f	x_1	x_2	x_3	x_4	RHS
f	1	0	0	$-\frac{4}{3}$	$-\frac{1}{3}$	$-\frac{29}{3}$
x_2	0	0	1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{11}{3}$
x_1	0	1	0	$-\frac{1}{3}$	$\frac{2}{3}$	$\frac{4}{3}$

This solution is optimal since $\mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{N}-\mathbf{c}_N^t\le\mathbf{0}$. The three points corresponding to the three tableaux are shown in the (x_1, x_2) space in Figure 2.17. We see that the simplex method moved from one extreme point to another until the optimal point is reached.

The Initial Extreme Point

Recall that the simplex method starts with an initial extreme point. From Theorem 2.5.4, finding an initial extreme point of the set $S=\{\mathbf{x}:\mathbf{A}\mathbf{x}=\mathbf{b}, \mathbf{x}\ge\mathbf{0}\}$ involves decomposing $\mathbf{A}$ into $\mathbf{B}$ and $\mathbf{N}$ with $\mathbf{B}^{-1}\mathbf{b}\ge\mathbf{0}$. In Example 2.6.2 above, an initial extreme point was immediately available. However, in many cases, an initial extreme point may not be conveniently available. This difficulty can be overcome by introducing *artificial variables.*

We discuss briefly two procedures for obtaining the initial extreme point. These are the two-phase method and the big-M method. For both methods the problem is first put in the standard format $\mathbf{A}\mathbf{x}=\mathbf{b}$ and $\mathbf{x}\ge\mathbf{0}$, with the additional requirement that $\mathbf{b}\ge\mathbf{0}$ (if $b_i<0$, then the ith constraint is multiplied by -1).

The Two-Phase Method In this method, the constraints of the problem are altered by the use of artificial variables so that an extreme point of the new system is at hand. In particular the constraint system is modified to

$$\begin{aligned} \mathbf{A}\mathbf{x}+\mathbf{x}_a &= \mathbf{b} \\ \mathbf{x},\ \mathbf{x}_a &\ge \mathbf{0} \end{aligned}$$

where $\mathbf{x}_a$ is an artificial vector. Obviously $\mathbf{x}=\mathbf{0}$ and $\mathbf{x}_a=\mathbf{b}$ represent an extreme point of the above system. Since a feasible solution of the original system will be obtained only if $\mathbf{x}_a=\mathbf{0}$, we can use the simplex method itself to minimize the sum of the artificial variables starting from the above extreme point. This leads to the following *Phase I* problem.

$$\begin{aligned} &\text{Minimize} && \mathbf{l}^t\mathbf{x}_a \\ &\text{subject to} && \mathbf{A}\mathbf{x}+\mathbf{x}_a=\mathbf{b} \\ & && \mathbf{x},\ \mathbf{x}_a\ge\mathbf{0} \end{aligned}$$

where $\mathbf{l}$ is a vector of ones. At the end of Phase I, either $\mathbf{x}_a\ne\mathbf{0}$ or $\mathbf{x}_a=\mathbf{0}$. In the former case we conclude that the original system is inconsistent; that is, the feasible region is empty. In the latter case the artificial variables would drop from the basis,† and hence we would obtain an extreme point of the original system. Starting with this extreme point, *Phase II* of the simplex method minimizes the original objective $\mathbf{c}^t\mathbf{x}$.

†It is possible that some of the artificial variables remain in the basis at zero level at the end of Phase I. This case can be easily treated (see Charnes and Cooper [1961] and Dantzig [1963]).

The Big-M Method As in the two-phase method, the constraints are modified by the use of artificial variables so that an extreme point of the new system is immediately available. A large positive cost coefficient M is assigned to each artificial variable so that they will drop to zero level. This leads to the following problem.

$$\begin{aligned} &\text{Minimize} && \mathbf{c}^t\mathbf{x}+M\mathbf{l}^t\mathbf{x}_a \\ &\text{subject to} && \mathbf{A}\mathbf{x}+\mathbf{x}_a=\mathbf{b} \\ & && \mathbf{x},\ \mathbf{x}_a\ge\mathbf{0} \end{aligned}$$

If at termination $\mathbf{x}_a=\mathbf{0}$, then we have an optimal solution of the original problem. Otherwise, if $\mathbf{x}_a\ne\mathbf{0}$ at termination of the simplex method, and provided that the variable entering the basis is the one with the most positive coefficient in the objective row, we conclude that the system $\mathbf{A}\mathbf{x}=\mathbf{b}$ and $\mathbf{x}\ge\mathbf{0}$ admits no feasible solutions.

2.1 Let S be a nonempty set in E_n. Show that S is convex if and only if for each integer $k \geq 2$, the following holds true: $\mathbf{x}_1, \ldots, \mathbf{x}_k \in S$ implies that $\sum_{j=1}^{k} \lambda_j \mathbf{x}_j \in S$, where $\sum_{j=1}^{k} \lambda_j = 1$ and $\lambda_j \geq 0$ for $j = 1, \ldots, k$.

2.2 Let S be a convex set in E_n, $\mathbf{A}$ be an $m \times n$ matrix, and α be a scalar. Show that the following two sets are convex.

a. $\mathbf{A}S = \{\mathbf{y} : \mathbf{y} = \mathbf{A}\mathbf{x}, \mathbf{x} \in S\}$

b. $\alpha S = \{\alpha \mathbf{x} : \mathbf{x} \in S\}$

2.3 Let $S_1 = \{\mathbf{x} : x_1 = 0, 0 \leq x_2 \leq 1\}$ and $S_2 = \{\mathbf{x} : 0 \leq x_1 \leq 1, x_2 = 2\}$. Describe $S_1 + S_2$ and $S_1 - S_2$.

2.4 Prove Lemma 2.1.2.

2.5 Let S be a closed set. Is it necessarily true that $H(S)$ is also closed? If it is not true in general, specify a sufficient condition so that $H(S)$ is closed. (*Hint*: Suppose that S is compact.)

2.6 Let S_1 and S_2 be nonempty sets in E_n. Show that $H(S_1 \cap S_2) \subset H(S_1) \cap H(S_2)$. Is $H(S_1 \cap S_2) = H(S_1) \cap H(S_1)$ true in general? If not, give a counter example.

2.7 Prove Lemma 2.1.4.

2.8 Let S be a polytope in E_n. Show that S is a closed, bounded convex set.

2.9 Let S_1 and S_2 be closed convex sets. Prove that $S_1 + S_2$ is convex. Show by an example that $S_1 + S_2$ is not necessarily closed. Prove that compactness of S_1 or S_2 is a sufficient condition for $S_1 + S_2$ to be closed.

2.10 Let $S_1 = \{\lambda \mathbf{d}_1 : \lambda \geq 0\}$ and $S_2 = \{\lambda \mathbf{d}_2 : \lambda \geq 0\}$, where $\mathbf{d}_1$ and $\mathbf{d}_2$ are nonzero vectors in E_n. Show that $S_1 + S_2$ is a closed convex set.

2.11 A *linear subspace* L of E_n is a subset of E_n such that $\mathbf{x}_1, \mathbf{x}_2 \in L$ implies that $\lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 \in L$ for all scalars λ_1 and λ_2. The *orthogonal complement* $L^\perp$ is defined by $L^\perp = \{\mathbf{y} : \mathbf{y}^t \mathbf{x} = 0 \text{ for all } \mathbf{x} \in L\}$. Show that any vector $\mathbf{x}$ in E_n could be represented uniquely as $\mathbf{x}_1 + \mathbf{x}_2$, where $\mathbf{x}_1 \in L$ and $\mathbf{x}_2 \in L^\perp$. Illustrate by writing the vector $(1, 2, 3)$ as the sum of two vectors in L and $L^\perp$, respectively, where $L = \{(x_1, x_2, x_3) : 2x_1 + x_2 - x_3 = 0\}$.

2.12 Let S be a polytope in E_n and let $S_j = \{\lambda \mathbf{d}_j : \lambda \geq 0\}$, where $\mathbf{d}_j$ is a nonzero vector in E_n for $j = 1, 2, \ldots, k$. Show that $S + \sum_{j=1}^{k} S_j$ is a closed convex set. (Note that Exercises 2.8 and 2.12 show that the set Λ in the proof of Theorem 2.5.7 is closed.)

2.13 Identify the closure, interior, and boundary of each of the following convex sets.

a. $S = \{\mathbf{x} : x_1^2 + x_2^2 \leq x_3\}$

b. $S = \{\mathbf{x} : 1 \leq x_1 \leq 2, x_2 = 3\}$

c. $S = \{\mathbf{x} : x_1 + x_2 \leq 3, -x_1 + x_2 + x_3 \leq 5, x_1, x_2, x_3 \geq 0\}$

d. $S = \{\mathbf{x} : x_1 + x_2 = 3, x_1 + x_2 + x_3 \leq 6\}$

e. $S = \{\mathbf{x} : x_1^2 + x_2^2 + x_3^2 \leq 4, x_1 + x_3 = 1\}$

2.14 Let $S = \{\mathbf{x} : x_1^2 + x_2^2 + x_3^2 \leq 1, x_1^2 - x_2 \leq 0\}$ and $\mathbf{y} = (1, 0, 2)^t$. Find the minimum distance from $\mathbf{y}$ to S, the unique minimizing point, and a separating hyperplane.

2.15 Prove that exactly one of the following two systems has a solution.

a. $\mathbf{A}\mathbf{x} \geq \mathbf{0}, \mathbf{x} \geq \mathbf{0}$, and $\mathbf{c}^t \mathbf{x} > 0$

b. $\mathbf{A}^t \mathbf{y} \geq \mathbf{c}$ and $\mathbf{y} \leq \mathbf{0}$

(*Hint*: Use Farkas' theorem.)

2.16 Show that the system $\mathbf{A}\mathbf{x} \leq \mathbf{0}$ and $\mathbf{c}^t \mathbf{x} > 0$ has a solution $\mathbf{x}$ in E_3, where $\mathbf{A} = \begin{bmatrix} 1 & -1 & -1 \\ 2 & 2 & 0 \end{bmatrix}$ and $\mathbf{c} = (1, 0, 5)^t$.

2.17 Let $\mathbf{A}$ be an $m \times n$ matrix. Using Farkas' theorem, prove that exactly one of the following two systems has a solution.

System 1 $\mathbf{A}\mathbf{x} > \mathbf{0}$

System 2 $\mathbf{A}^t \mathbf{y} = \mathbf{0}, \mathbf{y} \geq \mathbf{0}, \mathbf{y} \neq \mathbf{0}$

(This is Gordan's theorem developed in the text using Theorem 2.3.8.)

2.18 Let $\mathbf{A}$ be an $m \times n$ matrix and $\mathbf{c}$ be an n vector. Show that exactly one of the following two systems has a solution.

System 1 $\mathbf{A}\mathbf{x} = \mathbf{c}$

System 2 $\mathbf{A}^t \mathbf{y} = \mathbf{0}, \mathbf{c}^t \mathbf{y} = 1$

(This is a theorem of the alternative credited to *Gale*.)

2.19 Let $\mathbf{A}$ be an $m \times n$ matrix. Show that the following two systems have solutions $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ such that $\mathbf{A}\bar{\mathbf{x}} + \bar{\mathbf{y}} > \mathbf{0}$.

System 1 $\mathbf{A}\mathbf{x} \geq \mathbf{0}$

System 2 $\mathbf{A}^t \mathbf{y} = \mathbf{0}, \mathbf{y} \geq \mathbf{0}$

(This is an existence theorem credited to *Tucker*.)

2.20 Let $\mathbf{A}$ be a $p \times n$ matrix and $\mathbf{B}$ be a $q \times n$ matrix. Show that if *System 1* below has no solution, then *System 2* has a solution.

System 1 $\mathbf{A}\mathbf{x} < \mathbf{0}$ $\quad \mathbf{B}\mathbf{x} = \mathbf{0}$ $\quad$ for some $\mathbf{x} \in E_n$

System 2 $\mathbf{A}^t \mathbf{u} + \mathbf{B}^t \mathbf{v} = \mathbf{0}$ $\quad$ for some nonzero $(\mathbf{u}, \mathbf{v})$ with $\mathbf{u} \geq \mathbf{0}$.

Furthermore, show that if $\mathbf{B}$ has full rank, then exactly one of the systems has a solution. Is this necessarily true if $\mathbf{B}$ is not of full rank? Prove or give a counter example.

2.21 Let $\mathbf{A}$ be a $p \times n$ matrix, and $\mathbf{B}$ be a $q \times n$ matrix. Show that exactly one of the following systems has a solution.

System 1 $\mathbf{A}\mathbf{x} < \mathbf{0}$ $\quad \mathbf{B}\mathbf{x} = \mathbf{0}$ $\quad$ for some $\mathbf{x} \in E_n$

System 2 $\mathbf{A}^t \mathbf{u} + \mathbf{B}^t \mathbf{v} = \mathbf{0}$ $\quad$ for some $(\mathbf{u}, \mathbf{v})$, $\mathbf{u} \neq \mathbf{0}, \mathbf{u} \geq \mathbf{0}$.

2.22 Let S_1 and S_2 be convex sets in E_n. Show that there exists a hyperplane that strongly separates S_1 and S_2 if and only if

$$\inf \{\|\mathbf{x}_1 - \mathbf{x}_2\| : \mathbf{x}_1 \in S_2, \mathbf{x}_2 \in S_2\} > 0$$

2.23 Let $S_1 = \{\mathbf{x} : x_2 \geq e^{-x_1}\}$ and $S_2 = \{\mathbf{x} : x_2 \leq -e^{-x_1}\}$. Show that S_1 and S_2 are disjoint convex sets, and then find a hyperplane that separates them. Does there exist a hyperplane that strongly separates S_1 and S_2?

2.24 Let S_1 and S_2 be nonempty disjoint convex sets in E_n. Prove that there exist two nonzero vectors $\mathbf{p}_1$ and $\mathbf{p}_2$ such that

$$\mathbf{p}_1^t \mathbf{x}_1 + \mathbf{p}_2^t \mathbf{x}_2 \geq 0 \qquad \text{for all } \mathbf{x}_1 \in S_1 \text{ and all } \mathbf{x}_2 \in S_2$$

Can you generalize the result for three or more disjoint convex sets?

2.25 Consider $S = \{\mathbf{x} : x_1^2 + x_2^2 \leq 1\}$. Represent S as the intersection of a collection of half spaces. Find the half spaces explicitly.

2.26 Let C be a nonempty set in E_n. Show that C is a convex cone if and only if $\mathbf{x}_1, \mathbf{x}_2 \in C$ implies that $\lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 \in C$ for all $\lambda_1, \lambda_2 \geq 0$.

2.27 Let C_1 and C_2 be convex cones in E_n. Show that $C_1 + C_2$ is also a convex cone and that $C_1 + C_2 = H(C_1 \cup C_2)$.

2.28 Let S be a nonempty set in E_n and let $\bar{\mathbf{x}} \in S$. Consider the set $C = \{\mathbf{y} : \mathbf{y} = \lambda(\mathbf{x} - \bar{\mathbf{x}}), \lambda \geq 0, \mathbf{x} \in S\}$.
a. Show that C is a cone and interpret it geometrically.
b. Show that C is convex if S is convex.
c. Suppose that S is closed. Is it necessarily true that C is closed? If not, under what conditions would C be closed?

2.29 Let $C_\varepsilon = \{\mathbf{y} : \mathbf{y} = \lambda(\mathbf{x} - \bar{\mathbf{x}}), \lambda \geq 0, \mathbf{x} \in S \cap N_\varepsilon(\bar{\mathbf{x}})\}$, where $N_\varepsilon(\bar{\mathbf{x}})$ is an ε-neighborhood around $\bar{\mathbf{x}}$. Let T be the intersection of all such cones, that is, $T = \cap\{C_\varepsilon : \varepsilon > 0\}$. Interpret the cone T geometrically.
(T is called the *cone of tangents* of S at $\bar{\mathbf{x}}$ and will be discussed in more detail in Chapter 5.)

2.30 Derive an explicit form of the polar C^* of the following cones:
a. $C = \{(x_1, x_2) : 0 \leq x_2 \leq x_1\}$
b. $C = \{(x_1, x_2) : x_2 \geq -|x_1|\}$
c. $C = \{\mathbf{x} : \mathbf{x} = \mathbf{A}\mathbf{p}, \mathbf{p} \geq \mathbf{0}\}$

2.31 Let S be a nonempty set in E_n. The *polar set* of S, denoted by S_p is given by $\{\mathbf{y} : \mathbf{y}^t\mathbf{x} \leq 1 \text{ for all } \mathbf{x} \in S\}$.
a. Find the polar sets of the following two sets:

$$\{(x_1, x_2) : x_1^2 + x_2^2 \leq 1\}, \text{ and } \{(x_1, x_2) : x_1 + x_2 \leq 2, -x_1 + 2x_2 \leq 1, x_1, x_2 \geq 0\}$$

b. Show that S_p is a convex set. Is it necessarily closed?
c. If S is a polyhedral set, is it necessarily true that S_p is also a polyhedral set?
d. Show that if S is a polyhedral set containing the origin then $S = S_{pp}$.

2.32 Let C be a nonempty convex cone in E_n. Show that $C + C^* = E_n$; that is, any point in E_n can be written as a point in the cone C plus a point in its polar cone C^*. Is the representation unique? What if C is a linear subspace?

2.33 Identify the extreme points and extreme directions of the following sets.
a. $S = \{\mathbf{x} : x_2 \geq x_1^2, x_1 + x_2 + x_3 \leq 1\}$
b. $S = \{\mathbf{x} : x_1 + x_2 + x_3 \leq 2, x_1 + x_2 = 1, x_1, x_2, x_3 \geq 0\}$
c. $S = \{\mathbf{x} : x_2 \geq |x_1|, x_1^2 + x_2^2 \leq 1\}$

2.34 Consider the set $S = \{\mathbf{x} : -x_1 + 2x_2 \leq 3, x_1 + x_2 \leq 2, x_2 \leq 1, x_1, x_2 \geq 0\}$. Identify all extreme points and extreme directions. Represent the point $(1, \frac{1}{2})$ as a convex combination of the extreme points plus a nonnegative combination of the extreme directions.

2.35 Let S be a simplex in E_n with vertices $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{k+1}$. Show that the extreme points of S consist of its vertices.

2.36 Establish the set of directions for each of the following convex sets:
a. $S = \{(x_1, x_2) : x_2 \geq x_1^2\}$
b. $S = \{(x_1, x_2) : x_1x_2 \geq 1, x_1 > 0\}$
c. $S = \{(x_1, x_2) : |x_1| + |x_2| \leq 1\}$

2.37 Let S be a closed convex set in E_n and let $\bar{\mathbf{x}} \in S$. Suppose that $\mathbf{d}$ is a nonzero vector in E_n and that $\bar{\mathbf{x}} + \lambda\mathbf{d} \in S$ for all $\lambda \geq 0$. Show that $\mathbf{d}$ is a direction of S.

2.38 Find the extreme points and directions of the following polyhedral sets:
a. $S = \{\mathbf{x} : x_1 + x_2 + x_3 \leq 10, -x_1 + 2x_2 = 4, x_1, x_2, x_3 \geq 0\}$
b. $S = \{\mathbf{x} : x_1 + 2x_2 \geq 2, -x_1 + x_2 = 4, x_1, x_2 \geq 0\}$

2.39 Show that $C = \{\mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{0}\}$, where $\mathbf{A}$ is an $m \times n$ matrix, has at most one extreme point, namely the origin.

2.40 Let $S = \{\mathbf{x} : x_1 + x_2 \leq 1\}$. Find the extreme points and directions of S. Can you represent any point in S as a convex combination of its extreme points plus a nonnegative linear combination of its extreme directions? If not, discuss in relation to Theorem 2.5.7.

2.41 Consider the nonempty unbounded polyhedral set $S = \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$, where $\mathbf{A}$ is an $m \times n$ matrix of rank m. Prove directly that S has at least one extreme direction.
(*Hint*: Starting with a direction, use the characterization of Theorem 2.5.6 to construct an extreme direction.)

2.42 Prove Theorem 2.5.7 if the nondegeneracy assumption $\mathbf{B}^{-1}\mathbf{b} > 0$ is dropped.

2.43 Consider the following problem.

$$\begin{array}{ll} \text{Minimize} & \mathbf{c}^t\mathbf{x} \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{array}$$

where $\mathbf{A}$ is an $m \times n$ matrix with rank m. Let $\mathbf{x}$ be an extreme point with corresponding basis $\mathbf{B}$. Furthermore, suppose that $\mathbf{B}^{-1}\mathbf{b} > \mathbf{0}$. Use Farkas' theorem to show that $\mathbf{x}$ is an optimal point if and only if $\mathbf{c}_N^t - \mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{N} \geq \mathbf{0}$.

2.44 Consider the following problem.

$$\begin{array}{ll} \text{Minimize} & \mathbf{c}^t\mathbf{x} \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{array}$$

where $\mathbf{A}$ is an $m \times n$ matrix with rank m. Let $\mathbf{x}$ be an extreme point with basis $\mathbf{B}$, and let $\bar{\mathbf{b}} = \mathbf{B}^{-1}\mathbf{b}$. Furthermore, suppose that $\bar{b}_i = 0$ for some component i. Is it possible that $\mathbf{x}$ is an optimal solution even if $c_j - \mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{a}_j < 0$ for some nonbasic x_j? Discuss and give an example if this were possible.

2.45 Solve the following problem by the simplex method.

$$\begin{array}{ll} \text{Minimize} & x_1 + 3x_2 + x_3 \\ \text{subject to} & x_1 + 4x_2 + 3x_3 \leq 12 \\ & -x_1 + 2x_2 - x_3 \leq 4 \\ & x_1, \quad x_2, \quad x_3 \geq 0 \end{array}$$

2.46 Consider the set $\{\mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$, where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{b}$ is an m vector. Show that a nonzero vector $\mathbf{d}$ is a direction of the set if and only if $\mathbf{A}\mathbf{d} \leq \mathbf{0}$ and $\mathbf{d} \geq \mathbf{0}$. Show how the simplex method can be used to generate such a direction.

2.47 Consider the following problem:

$$\begin{array}{ll} \text{Minimize} & x_1 - 6x_2 \\ \text{subject to} & x_1 + x_2 \leq 12 \\ & -x_1 + 2x_2 \leq 4 \\ & x_2 \leq 6 \end{array}$$

Find the optimal solution geometrically and verify its optimality by showing that $\mathbf{c}_N^t - \mathbf{c}_B^t\mathbf{B}^{-1}\mathbf{N} \geq \mathbf{0}$.

2.48 Show in detail that pivoting at y_{rj} updates the simplex tableau.

2.49 Solve the following problem by the two-phase simplex method and by the big-M method.

$$\begin{array}{ll} \text{Maximize} & -x_1-2x_2+x_3 \\ \text{subject to} & x_1+3x_2+x_3\geq 4 \\ & x_1+2x_2-x_3\geq 6 \\ & x_1 \qquad\;\; +x_3\leq 12 \\ & x_1, \quad x_2, \quad x_3\geq 0 \end{array}$$

Notes and References

In this chapter we treat the topic of convex sets. This subject was first studied systematically by Minkowski [1911] whose work contains the essence of the important results in this area. The topic of convexity is fully developed in a variety of good texts, and the interested reader may refer to Eggleston [1958], Rockafellar [1970], Stoer and Witzgall [1970], and Valentine [1964] for a more detailed anlysis of convex sets.

Section 2.1 presents some basic definitions and develops the Carathéodory theorem, which states that each point in the convex hull of any given set can be represented as the convex combination of $n+1$ points in the set. This result can be sharpened by using the notion of *dimension of the set.* Using this notion, several Carathéodory-type theorems can be developed. See, for example, Bazaraa and Shetty [1976], Eggleston [1958], and Rockafellar [1970].

In Section 2.2 we develop some topological properties of convex sets related to interior and closure points. Section 2.3 presents various types of theorems that separate disjoint convex sets. Support and separation theorems are of special importance in the area of optimization, and are also widely used in game theory, functional analysis, and optimal control theory. An interesting application is the use of these results in coloring problems in graph theory. For further reading on support and separation of convex sets, see Eggleston [1958], Klee [1969], Mangasarian [1969a], Rockafellar [1970], Stoer and Witzgall [1970], and Valentine [1964]. Many of the results in Sections 2.2 and 2.3 can be strengthened by using the notion of *relative interior.* For example, every nonempty convex set has a nonempty relative interior. Furthermore, a hyperplane that properly separates two convex sets exists provided that they have disjoint relative interiors. Also Theorem 2.2.2 and several of its corollaries can be sharpened using this concept. For a good discussion of relative interiors, see Eggleston [1958], Rockafellar [1970], and Valentine [1964].

In Section 2.4, a brief introduction to polar cones is given. For more details, see Rockafellar [1970]. In Section 2.5 we treat the important special case of polyhedral sets and prove the representation theorem, which states that every point in the set can be represented as a convex combination of the extreme points plus a nonnegative linear combination of the extreme directions. This result was first provided by Motzkin [1936] using a different approach. The representation theorem is also true for closed convex sets that contain no lines. For a proof of this result, see Bazaraa and Shetty [1976] and Rockafellar [1970]. An exhaustive treatment of convex polytopes is given by Grünbaum [1967].

In Section 2.6 we present the simplex algorithm for solving linear programming problems. The simplex algorithm was developed by Dantzig in 1947. The efficiency of the simplex algorithm, the advances in computer technology, and the ability of linear programming to model large and complex problems led to the popularity of the simplex method and linear programming. The presentation of the simplex method in Section 2.6 is a natural extension of the material