Chapter 2

Convex sets

2.1 Affine and convex sets

2.1.1 Lines and line segments

Suppose $x_1 \neq x_2$ are two points in $\mathbf{R}^n$. Points of the form

$$y = \theta x_1 + (1-\theta)x_2,$$

where $\theta \in \mathbf{R}$, form the *line* passing through x_1 and x_2. The parameter value $\theta = 0$ corresponds to $y = x_2$, and the parameter value $\theta = 1$ corresponds to $y = x_1$. Values of the parameter θ between 0 and 1 correspond to the (closed) *line segment* between x_1 and x_2.

Expressing y in the form

$$y = x_2 + \theta(x_1 - x_2)$$

gives another interpretation: y is the sum of the *base point* x_2 (corresponding to $\theta = 0$) and the *direction* $x_1 - x_2$ (which points from x_2 to x_1) scaled by the parameter θ. Thus, θ gives the fraction of the way from x_2 to x_1 where y lies. As θ increases from 0 to 1, the point y moves from x_2 to x_1; for $\theta > 1$, the point y lies on the line beyond x_1. This is illustrated in figure 2.1.

2.1.2 Affine sets

A set $C \subseteq \mathbf{R}^n$ is *affine* if the line through any two distinct points in C lies in C, *i.e.*, if for any $x_1, x_2 \in C$ and $\theta \in \mathbf{R}$, we have $\theta x_1 + (1-\theta)x_2 \in C$. In other words, C contains the linear combination of any two points in C, provided the coefficients in the linear combination sum to one.

This idea can be generalized to more than two points. We refer to a point of the form $\theta_1 x_1 + \cdots + \theta_k x_k$, where $\theta_1 + \cdots + \theta_k = 1$, as an *affine combination* of the points $x_1, \ldots, x_k$. Using induction from the definition of affine set (*i.e.*, that it contains every affine combination of two points in it), it can be shown that

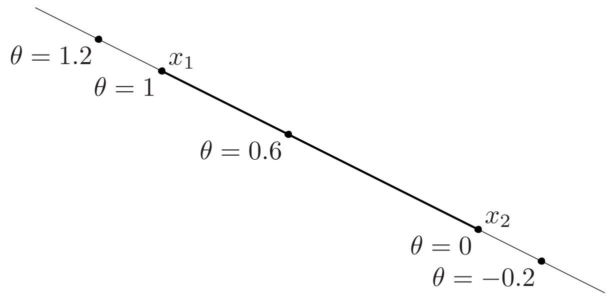

Figure 2.1 The line passing through x_1 and x_2 is described parametrically by $\theta x_1 + (1-\theta)x_2$, where θ varies over $\mathbf{R}$. The line segment between x_1 and x_2, which corresponds to θ between 0 and 1, is shown darker.

an affine set contains every affine combination of its points: If C is an affine set, $x_1, \ldots, x_k \in C$, and $\theta_1 + \cdots + \theta_k = 1$, then the point $\theta_1 x_1 + \cdots + \theta_k x_k$ also belongs to C.

If C is an affine set and $x_0 \in C$, then the set

$$V = C - x_0 = \{x - x_0 \mid x \in C\}$$

is a subspace, *i.e.*, closed under sums and scalar multiplication. To see this, suppose $v_1,\ v_2 \in V$ and $\alpha,\ \beta \in \mathbf{R}$. Then we have $v_1 + x_0 \in C$ and $v_2 + x_0 \in C$, and so

$$\alpha v_1 + \beta v_2 + x_0 = \alpha(v_1 + x_0) + \beta(v_2 + x_0) + (1 - \alpha - \beta)x_0 \in C,$$

since C is affine, and $\alpha + \beta + (1 - \alpha - \beta) = 1$. We conclude that $\alpha v_1 + \beta v_2 \in V$, since $\alpha v_1 + \beta v_2 + x_0 \in C$.

Thus, the affine set C can be expressed as

$$C = V + x_0 = \{v + x_0 \mid v \in V\},$$

i.e., as a subspace plus an offset. The subspace V associated with the affine set C does not depend on the choice of x_0, so x_0 can be chosen as any point in C. We define the *dimension* of an affine set C as the dimension of the subspace $V = C - x_0$, where x_0 is any element of C.

Example 2.1 *Solution set of linear equations.* The solution set of a system of linear equations, $C = \{x \mid Ax = b\}$, where $A \in \mathbf{R}^{m \times n}$ and $b \in \mathbf{R}^m$, is an affine set. To show this, suppose $x_1,\ x_2 \in C$, *i.e.*, $Ax_1 = b$, $Ax_2 = b$. Then for any θ, we have

$$\begin{aligned} A(\theta x_1 + (1-\theta)x_2) &= \theta A x_1 + (1-\theta) A x_2 \\ &= \theta b + (1-\theta) b \\ &= b, \end{aligned}$$

which shows that the affine combination $\theta x_1 + (1-\theta)x_2$ is also in C. The subspace associated with the affine set C is the nullspace of A.

We also have a converse: every affine set can be expressed as the solution set of a system of linear equations.

The set of all affine combinations of points in some set $C \subseteq \mathbf{R}^n$ is called the *affine hull* of C, and denoted $\mathbf{aff}\, C$:

$$\mathbf{aff}\, C = \{\theta_1 x_1 + \cdots + \theta_k x_k \mid x_1, \ldots, x_k \in C,\ \theta_1 + \cdots + \theta_k = 1\}.$$

The affine hull is the smallest affine set that contains C, in the following sense: if S is any affine set with $C \subseteq S$, then $\mathbf{aff}\, C \subseteq S$.

2.1.3 Affine dimension and relative interior

We define the *affine dimension* of a set C as the dimension of its affine hull. Affine dimension is useful in the context of convex analysis and optimization, but is not always consistent with other definitions of dimension. As an example consider the unit circle in $\mathbf{R}^2$, *i.e.*, $\{x \in \mathbf{R}^2 \mid x_1^2 + x_2^2 = 1\}$. Its affine hull is all of $\mathbf{R}^2$, so its affine dimension is two. By most definitions of dimension, however, the unit circle in $\mathbf{R}^2$ has dimension one.

If the affine dimension of a set $C \subseteq \mathbf{R}^n$ is less than n, then the set lies in the affine set $\mathbf{aff}\, C \neq \mathbf{R}^n$. We define the *relative interior* of the set C, denoted $\mathbf{relint}\, C$, as its interior relative to $\mathbf{aff}\, C$:

$$\mathbf{relint}\, C = \{x \in C \mid B(x, r) \cap \mathbf{aff}\, C \subseteq C \text{ for some } r > 0\},$$

where $B(x, r) = \{y \mid \|y - x\| \leq r\}$, the ball of radius r and center x in the norm $\|\cdot\|$. (Here $\|\cdot\|$ is any norm; all norms define the same relative interior.) We can then define the *relative boundary* of a set C as $\mathbf{cl}\, C \setminus \mathbf{relint}\, C$, where $\mathbf{cl}\, C$ is the closure of C.

Example 2.2 Consider a square in the (x_1, x_2)-plane in $\mathbf{R}^3$, defined as

$$C = \{x \in \mathbf{R}^3 \mid -1 \leq x_1 \leq 1,\ -1 \leq x_2 \leq 1,\ x_3 = 0\}.$$

Its affine hull is the (x_1, x_2)-plane, *i.e.*, $\mathbf{aff}\, C = \{x \in \mathbf{R}^3 \mid x_3 = 0\}$. The interior of C is empty, but the relative interior is

$$\mathbf{relint}\, C = \{x \in \mathbf{R}^3 \mid -1 < x_1 < 1,\ -1 < x_2 < 1,\ x_3 = 0\}.$$

Its boundary (in $\mathbf{R}^3$) is itself; its relative boundary is the wire-frame outline,

$$\{x \in \mathbf{R}^3 \mid \max\{|x_1|, |x_2|\} = 1,\ x_3 = 0\}.$$

2.1.4 Convex sets

A set C is *convex* if the line segment between any two points in C lies in C, *i.e.*, if for any $x_1,\ x_2 \in C$ and any θ with $0 \leq \theta \leq 1$, we have

$$\theta x_1 + (1 - \theta) x_2 \in C.$$

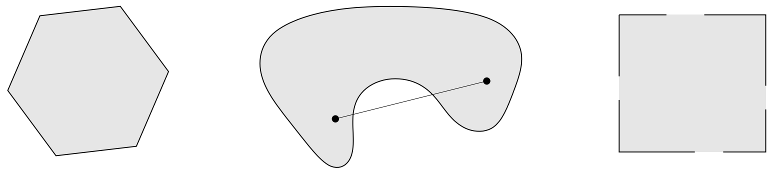

Figure 2.2 Some simple convex and nonconvex sets. *Left.* The hexagon, which includes its boundary (shown darker), is convex. *Middle.* The kidney shaped set is not convex, since the line segment between the two points in the set shown as dots is not contained in the set. *Right.* The square contains some boundary points but not others, and is not convex.

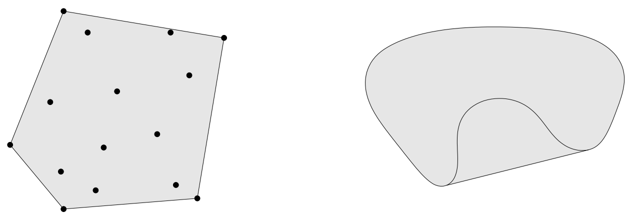

Figure 2.3 The convex hulls of two sets in $\mathbf{R}^2$. *Left.* The convex hull of a set of fifteen points (shown as dots) is the pentagon (shown shaded). *Right.* The convex hull of the kidney shaped set in figure 2.2 is the shaded set.

Roughly speaking, a set is convex if every point in the set can be seen by every other point, along an unobstructed straight path between them, where unobstructed means lying in the set. Every affine set is also convex, since it contains the entire line between any two distinct points in it, and therefore also the line segment between the points. Figure 2.2 illustrates some simple convex and nonconvex sets in $\mathbf{R}^2$.

We call a point of the form $\theta_1 x_1 + \cdots + \theta_k x_k$, where $\theta_1 + \cdots + \theta_k = 1$ and $\theta_i \geq 0$, $i = 1, \ldots, k$, a *convex combination* of the points $x_1, \ldots, x_k$. As with affine sets, it can be shown that a set is convex if and only if it contains every convex combination of its points. A convex combination of points can be thought of as a *mixture* or *weighted average* of the points, with θ_i the fraction of x_i in the mixture.

The *convex hull* of a set C, denoted $\mathbf{conv}\, C$, is the set of all convex combinations of points in C:

$$\mathbf{conv}\, C = \{\theta_1 x_1 + \cdots + \theta_k x_k \mid x_i \in C,\ \theta_i \geq 0,\ i = 1, \ldots, k,\ \theta_1 + \cdots + \theta_k = 1\}.$$

As the name suggests, the convex hull $\mathbf{conv}\, C$ is always convex. It is the smallest convex set that contains C: If B is any convex set that contains C, then $\mathbf{conv}\, C \subseteq B$. Figure 2.3 illustrates the definition of convex hull.

The idea of a convex combination can be generalized to include infinite sums, integrals, and, in the most general form, probability distributions. Suppose $\theta_1, \theta_2, \ldots$

satisfy

$$\theta_i \geq 0, \quad i = 1, 2, \ldots, \qquad \sum_{i=1}^{\infty} \theta_i = 1,$$

and $x_1, x_2, \ldots \in C$, where $C \subseteq \mathbf{R}^n$ is convex. Then

$$\sum_{i=1}^{\infty} \theta_i x_i \in C,$$

if the series converges. More generally, suppose $p : \mathbf{R}^n \to \mathbf{R}$ satisfies $p(x) \geq 0$ for all $x \in C$ and $\int_C p(x)\, dx = 1$, where $C \subseteq \mathbf{R}^n$ is convex. Then

$$\int_C p(x) x \, dx \in C,$$

if the integral exists.

In the most general form, suppose $C \subseteq \mathbf{R}^n$ is convex and x is a random vector with $x \in C$ with probability one. Then $\mathbf{E}\, x \in C$. Indeed, this form includes all the others as special cases. For example, suppose the random variable x only takes on the two values x_1 and x_2, with $\mathbf{prob}(x = x_1) = \theta$ and $\mathbf{prob}(x = x_2) = 1 - \theta$, where $0 \leq \theta \leq 1$. Then $\mathbf{E}\, x = \theta x_1 + (1 - \theta) x_2$, and we are back to a simple convex combination of two points.

2.1.5 Cones

A set C is called a *cone*, or *nonnegative homogeneous*, if for every $x \in C$ and $\theta \geq 0$ we have $\theta x \in C$. A set C is a *convex cone* if it is convex and a cone, which means that for any $x_1,\ x_2 \in C$ and $\theta_1,\ \theta_2 \geq 0$, we have

$$\theta_1 x_1 + \theta_2 x_2 \in C.$$

Points of this form can be described geometrically as forming the two-dimensional pie slice with apex 0 and edges passing through x_1 and x_2. (See figure 2.4.)

A point of the form $\theta_1 x_1 + \cdots + \theta_k x_k$ with $\theta_1, \ldots, \theta_k \geq 0$ is called a *conic combination* (or a *nonnegative linear combination*) of $x_1, \ldots, x_k$. If x_i are in a convex cone C, then every conic combination of x_i is in C. Conversely, a set C is a convex cone if and only if it contains all conic combinations of its elements. Like convex (or affine) combinations, the idea of conic combination can be generalized to infinite sums and integrals.

The *conic hull* of a set C is the set of all conic combinations of points in C, *i.e.*,

$$\{\theta_1 x_1 + \cdots + \theta_k x_k \mid x_i \in C,\ \theta_i \geq 0,\ i = 1, \ldots, k\},$$

which is also the smallest convex cone that contains C (see figure 2.5).

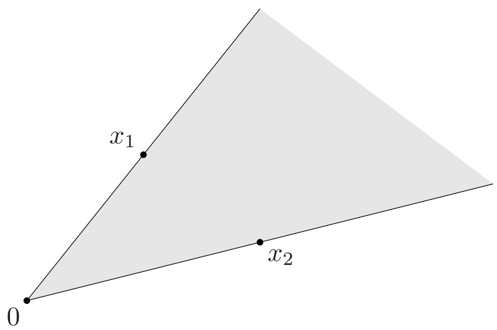

Figure 2.4 The pie slice shows all points of the form $\theta_1 x_1 + \theta_2 x_2$, where $\theta_1,\ \theta_2 \geq 0$. The apex of the slice (which corresponds to $\theta_1 = \theta_2 = 0$) is at 0; its edges (which correspond to $\theta_1 = 0$ or $\theta_2 = 0$) pass through the points x_1 and x_2.

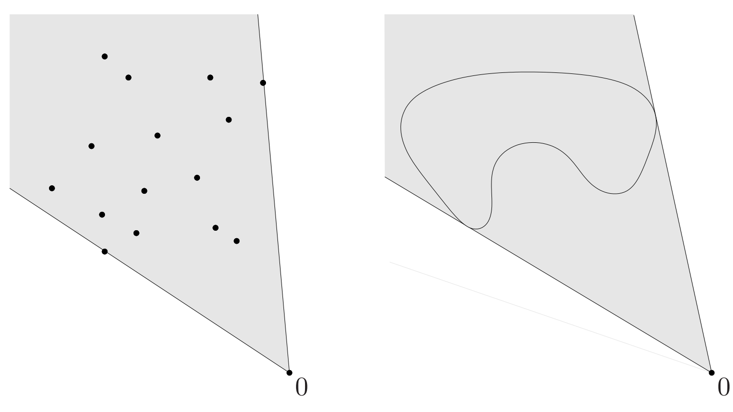

Figure 2.5 The conic hulls (shown shaded) of the two sets of figure 2.3.

2.2 Some important examples

In this section we describe some important examples of convex sets which we will encounter throughout the rest of the book. We start with some simple examples.

- The empty set $\emptyset$, any single point (*i.e.*, singleton) $\{x_0\}$, and the whole space $\mathbf{R}^n$ are affine (hence, convex) subsets of $\mathbf{R}^n$.
- Any line is affine. If it passes through zero, it is a subspace, hence also a convex cone.
- A line segment is convex, but not affine (unless it reduces to a point).
- A *ray*, which has the form $\{x_0 + \theta v \mid \theta \geq 0\}$, where $v \neq 0$, is convex, but not affine. It is a convex cone if its base x_0 is 0.
- Any subspace is affine, and a convex cone (hence convex).

2.2.1 Hyperplanes and halfspaces

A *hyperplane* is a set of the form

$$\{x \mid a^T x = b\},$$

where $a \in \mathbf{R}^n$, $a \neq 0$, and $b \in \mathbf{R}$. Analytically it is the solution set of a nontrivial linear equation among the components of x (and hence an affine set). Geometrically, the hyperplane $\{x \mid a^T x = b\}$ can be interpreted as the set of points with a constant inner product to a given vector a, or as a hyperplane with *normal vector* a; the constant $b \in \mathbf{R}$ determines the offset of the hyperplane from the origin. This geometric interpretation can be understood by expressing the hyperplane in the form

$$\{x \mid a^T (x - x_0) = 0\},$$

where x_0 is any point in the hyperplane (*i.e.*, any point that satisfies $a^T x_0 = b$). This representation can in turn be expressed as

$$\{x \mid a^T (x - x_0) = 0\} = x_0 + a^\perp,$$

where $a^\perp$ denotes the orthogonal complement of a, *i.e.*, the set of all vectors orthogonal to it:

$$a^\perp = \{v \mid a^T v = 0\}.$$

This shows that the hyperplane consists of an offset x_0, plus all vectors orthogonal to the (normal) vector a. These geometric interpretations are illustrated in figure 2.6.

A hyperplane divides $\mathbf{R}^n$ into two *halfspaces*. A (closed) halfspace is a set of the form

$$\{x \mid a^T x \leq b\}, \tag{2.1}$$

where $a \neq 0$, *i.e.*, the solution set of one (nontrivial) linear inequality. Halfspaces are convex, but not affine. This is illustrated in figure 2.7.

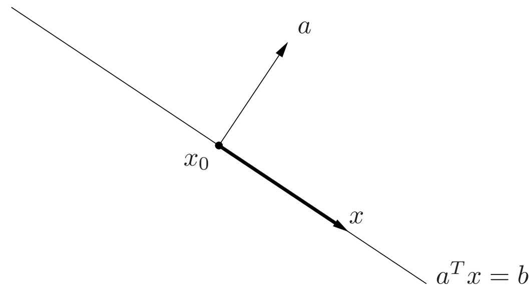

Figure 2.6 Hyperplane in $\mathbf{R}^2$, with normal vector a and a point x_0 in the hyperplane. For any point x in the hyperplane, $x - x_0$ (shown as the darker arrow) is orthogonal to a.

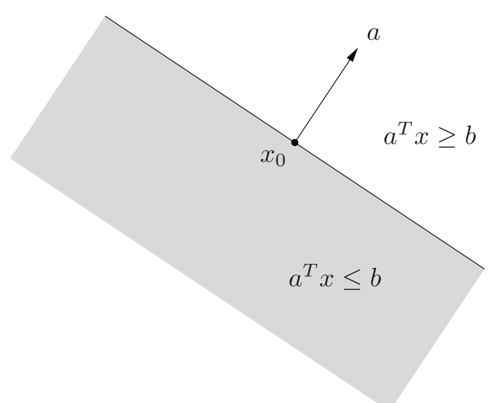

Figure 2.7 A hyperplane defined by $a^T x = b$ in $\mathbf{R}^2$ determines two halfspaces. The halfspace determined by $a^T x \geq b$ (not shaded) is the halfspace extending in the direction a. The halfspace determined by $a^T x \leq b$ (which is shown shaded) extends in the direction $-a$. The vector a is the outward normal of this halfspace.

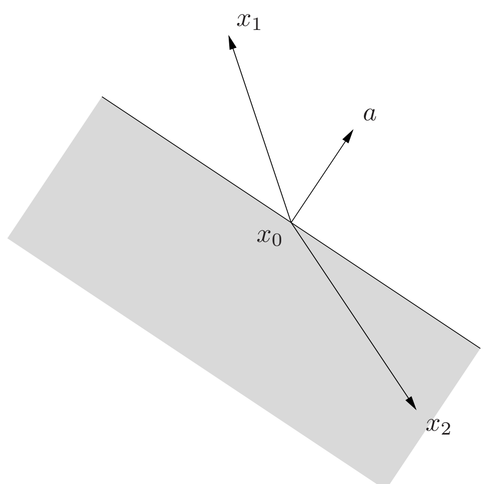

Figure 2.8 The shaded set is the halfspace determined by $a^T(x - x_0) \leq 0$. The vector $x_1 - x_0$ makes an acute angle with a, so x_1 is not in the halfspace. The vector $x_2 - x_0$ makes an obtuse angle with a, and so is in the halfspace.

The halfspace (2.1) can also be expressed as

$$\{x \mid a^T(x - x_0) \leq 0\}, \tag{2.2}$$

where x_0 is any point on the associated hyperplane, *i.e.*, satisfies $a^T x_0 = b$. The representation (2.2) suggests a simple geometric interpretation: the halfspace consists of x_0 plus any vector that makes an obtuse (or right) angle with the (outward normal) vector a. This is illustrated in figure 2.8.

The boundary of the halfspace (2.1) is the hyperplane $\{x \mid a^T x = b\}$. The set $\{x \mid a^T x < b\}$, which is the interior of the halfspace $\{x \mid a^T x \leq b\}$, is called an *open halfspace*.

2.2.2 Euclidean balls and ellipsoids

A *(Euclidean) ball* (or just ball) in $\mathbf{R}^n$ has the form

$$B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\} = \{x \mid (x - x_c)^T(x - x_c) \leq r^2\},$$

where $r > 0$, and $\|\cdot\|_2$ denotes the Euclidean norm, *i.e.*, $\|u\|_2 = (u^T u)^{1/2}$. The vector x_c is the *center* of the ball and the scalar r is its *radius*; $B(x_c, r)$ consists of all points within a distance r of the center x_c. Another common representation for the Euclidean ball is

$$B(x_c, r) = \{x_c + ru \mid \|u\|_2 \leq 1\}.$$

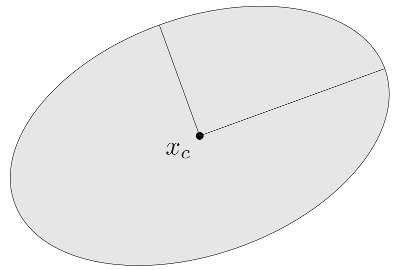

Figure 2.9 An ellipsoid in $\mathbf{R}^2$, shown shaded. The center x_c is shown as a dot, and the two semi-axes are shown as line segments.

A Euclidean ball is a convex set: if $\|x_1 - x_c\|_2 \leq r$, $\|x_2 - x_c\|_2 \leq r$, and $0 \leq \theta \leq 1$, then

$$\begin{array}{rcl} \|\theta x_1 + (1-\theta)x_2 - x_c\|_2 & = & \|\theta(x_1 - x_c) + (1-\theta)(x_2 - x_c)\|_2 \\ & \leq & \theta\|x_1 - x_c\|_2 + (1-\theta)\|x_2 - x_c\|_2 \\ & \leq & r. \end{array}$$

(Here we use the homogeneity property and triangle inequality for $\|\cdot\|_2$; see §A.1.2.)

A related family of convex sets is the *ellipsoids*, which have the form

$$\mathcal{E} = \{x \mid (x - x_c)^T P^{-1} (x - x_c) \leq 1\}, \tag{2.3}$$

where $P = P^T \succ 0$, *i.e.*, P is symmetric and positive definite. The vector $x_c \in \mathbf{R}^n$ is the *center* of the ellipsoid. The matrix P determines how far the ellipsoid extends in every direction from x_c; the lengths of the semi-axes of $\mathcal{E}$ are given by $\sqrt{\lambda_i}$, where λ_i are the eigenvalues of P. A ball is an ellipsoid with $P = r^2 I$. Figure 2.9 shows an ellipsoid in $\mathbf{R}^2$.

Another common representation of an ellipsoid is

$$\mathcal{E} = \{x_c + Au \mid \|u\|_2 \leq 1\}, \tag{2.4}$$

where A is square and nonsingular. In this representation we can assume without loss of generality that A is symmetric and positive definite. By taking $A = P^{1/2}$, this representation gives the ellipsoid defined in (2.3). When the matrix A in (2.4) is symmetric positive semidefinite but singular, the set in (2.4) is called a *degenerate ellipsoid*; its affine dimension is equal to the rank of A. Degenerate ellipsoids are also convex.

2.2.3 Norm balls and norm cones

Suppose $\|\cdot\|$ is any norm on $\mathbf{R}^n$ (see §A.1.2). From the general properties of norms it can be shown that a *norm ball* of radius r and center x_c, given by $\{x \mid \|x - x_c\| \leq r\}$, is convex. The *norm cone* associated with the norm $\|\cdot\|$ is the set

$$C = \{(x, t) \mid \|x\| \leq t\} \subseteq \mathbf{R}^{n+1}.$$

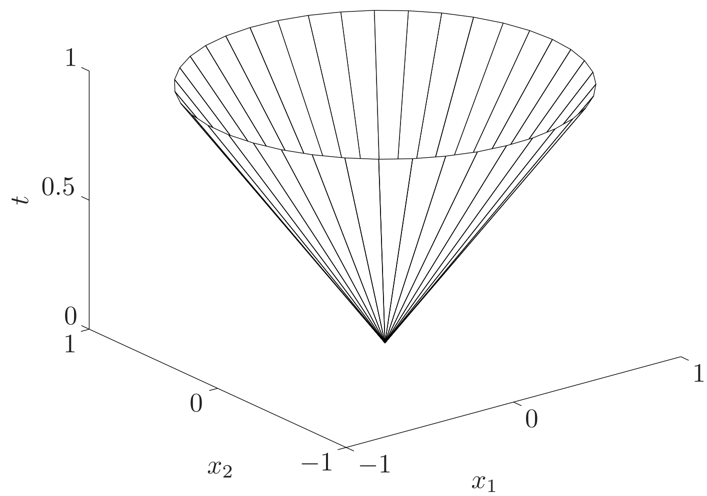

Figure 2.10 Boundary of second-order cone in $\mathbf{R}^3$, $\{(x_1, x_2, t) \mid (x_1^2+x_2^2)^{1/2} \leq t\}$.

It is (as the name suggests) a convex cone.

Example 2.3 *The second-order cone* is the norm cone for the Euclidean norm, *i.e.*,

$$\begin{aligned} C &= \{(x,t) \in \mathbf{R}^{n+1} \mid \|x\|_2 \leq t\} \\ &= \left\{ \left[\begin{array}{c} x \\ t \end{array}\right] \;\middle|\; \left[\begin{array}{c} x \\ t \end{array}\right]^T \left[\begin{array}{cc} I & 0 \\ 0 & -1 \end{array}\right] \left[\begin{array}{c} x \\ t \end{array}\right] \leq 0,\ t \geq 0 \right\}. \end{aligned}$$

The second-order cone is also known by several other names. It is called the *quadratic cone*, since it is defined by a quadratic inequality. It is also called the *Lorentz cone* or *ice-cream cone*. Figure 2.10 shows the second-order cone in $\mathbf{R}^3$.

2.2.4 Polyhedra

A *polyhedron* is defined as the solution set of a finite number of linear equalities and inequalities:

$$\mathcal{P} = \{x \mid a_j^T x \leq b_j,\ j = 1, \ldots, m,\ c_j^T x = d_j,\ j = 1, \ldots, p\}. \tag{2.5}$$

A polyhedron is thus the intersection of a finite number of halfspaces and hyperplanes. Affine sets (*e.g.*, subspaces, hyperplanes, lines), rays, line segments, and halfspaces are all polyhedra. It is easily shown that polyhedra are convex sets. A bounded polyhedron is sometimes called a *polytope*, but some authors use the opposite convention (*i.e.*, polytope for any set of the form (2.5), and polyhedron

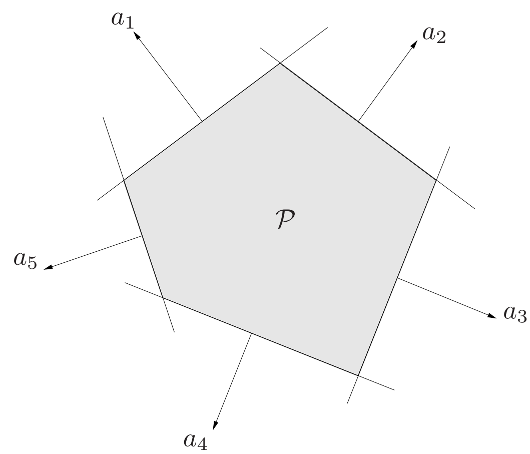

Figure 2.11 The polyhedron $\mathcal{P}$ (shown shaded) is the intersection of five halfspaces, with outward normal vectors $a_1, \ldots, a_5$.

when it is bounded). Figure 2.11 shows an example of a polyhedron defined as the intersection of five halfspaces.

It will be convenient to use the compact notation

$$\mathcal{P} = \{x \mid Ax \preceq b,\ Cx = d\} \tag{2.6}$$

for (2.5), where

$$A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix}, \qquad C = \begin{bmatrix} c_1^T \\ \vdots \\ c_p^T \end{bmatrix},$$

and the symbol $\preceq$ denotes *vector inequality* or *componentwise inequality* in $\mathbf{R}^m$: $u \preceq v$ means $u_i \leq v_i$ for $i = 1, \ldots, m$.

Example 2.4 The *nonnegative orthant* is the set of points with nonnegative components, *i.e.*,

$$\mathbf{R}_+^n = \{x \in \mathbf{R}^n \mid x_i \geq 0,\ i = 1, \ldots, n\} = \{x \in \mathbf{R}^n \mid x \succeq 0\}.$$

(Here $\mathbf{R}_+$ denotes the set of nonnegative numbers: $\mathbf{R}_+ = \{x \in \mathbf{R} \mid x \geq 0\}$.) The nonnegative orthant is a polyhedron and a cone (and therefore called a *polyhedral cone*).

Simplexes

Simplexes are another important family of polyhedra. Suppose the $k+1$ points $v_0, \ldots, v_k \in \mathbf{R}^n$ are *affinely independent*, which means $v_1 - v_0, \ldots, v_k - v_0$ are linearly independent. The simplex determined by them is given by

$$C = \mathbf{conv}\{v_0, \ldots, v_k\} = \{\theta_0 v_0 + \cdots + \theta_k v_k \mid \theta \succeq 0,\ \mathbf{1}^T\theta = 1\}, \tag{2.7}$$

where $\mathbf{1}$ denotes the vector with all entries one. The affine dimension of this simplex is k, so it is sometimes referred to as a k-dimensional simplex in $\mathbf{R}^n$.

Example 2.5 *Some common simplexes.* A 1-dimensional simplex is a line segment; a 2-dimensional simplex is a triangle (including its interior); and a 3-dimensional simplex is a tetrahedron.

The *unit simplex* is the n-dimensional simplex determined by the zero vector and the unit vectors, *i.e.*, $0,\ e_1, \ldots, e_n \in \mathbf{R}^n$. It can be expressed as the set of vectors that satisfy

$$x \succeq 0, \qquad \mathbf{1}^T x \leq 1.$$

The *probability simplex* is the $(n-1)$-dimensional simplex determined by the unit vectors $e_1, \ldots, e_n \in \mathbf{R}^n$. It is the set of vectors that satisfy

$$x \succeq 0, \qquad \mathbf{1}^T x = 1.$$

Vectors in the probability simplex correspond to probability distributions on a set with n elements, with x_i interpreted as the probability of the ith element.

To describe the simplex (2.7) as a polyhedron, *i.e.*, in the form (2.6), we proceed as follows. By definition, $x \in C$ if and only if $x = \theta_0 v_0 + \theta_1 v_1 + \cdots + \theta_k v_k$ for some $\theta \succeq 0$ with $\mathbf{1}^T \theta = 1$. Equivalently, if we define $y = (\theta_1, \ldots, \theta_k)$ and

$$B = \left[\begin{array}{ccc} v_1 - v_0 & \cdots & v_k - v_0 \end{array}\right] \in \mathbf{R}^{n \times k},$$

we can say that $x \in C$ if and only if

$$x = v_0 + By \tag{2.8}$$

for some $y \succeq 0$ with $\mathbf{1}^T y \leq 1$. Now we note that affine independence of the points $v_0, \ldots, v_k$ implies that the matrix B has rank k. Therefore there exists a nonsingular matrix $A = (A_1, A_2) \in \mathbf{R}^{n \times n}$ such that

$$AB = \left[\begin{array}{c} A_1 \\ A_2 \end{array}\right] B = \left[\begin{array}{c} I \\ 0 \end{array}\right].$$

Multiplying (2.8) on the left with A, we obtain

$$A_1 x = A_1 v_0 + y, \qquad A_2 x = A_2 v_0.$$

From this we see that $x \in C$ if and only if $A_2 x = A_2 v_0$, and the vector $y = A_1 x - A_1 v_0$ satisfies $y \succeq 0$ and $\mathbf{1}^T y \leq 1$. In other words we have $x \in C$ if and only if

$$A_2 x = A_2 v_0, \qquad A_1 x \succeq A_1 v_0, \qquad \mathbf{1}^T A_1 x \leq 1 + \mathbf{1}^T A_1 v_0,$$

which is a set of linear equalities and inequalities in x, and so describes a polyhedron.

Convex hull description of polyhedra

The convex hull of the finite set $\{v_1, \ldots, v_k\}$ is

$$\mathbf{conv}\{v_1, \ldots, v_k\} = \{\theta_1 v_1 + \cdots + \theta_k v_k \mid \theta \succeq 0,\ \mathbf{1}^T\theta = 1\}.$$

This set is a polyhedron, and bounded, but (except in special cases, *e.g.*, a simplex) it is not simple to express it in the form (2.5), *i.e.*, by a set of linear equalities and inequalities.

A generalization of this convex hull description is

$$\{\theta_1 v_1 + \cdots + \theta_k v_k \mid \theta_1 + \cdots + \theta_m = 1,\ \theta_i \geq 0,\ i = 1, \ldots, k\}, \tag{2.9}$$

where $m \leq k$. Here we consider nonnegative linear combinations of v_i, but only the first m coefficients are required to sum to one. Alternatively, we can interpret (2.9) as the convex hull of the points $v_1, \ldots, v_m$, plus the conic hull of the points $v_{m+1}, \ldots, v_k$. The set (2.9) defines a polyhedron, and conversely, every polyhedron can be represented in this form (although we will not show this).

The question of how a polyhedron is represented is subtle, and has very important practical consequences. As a simple example consider the unit ball in the ℓ_∞-norm in $\mathbf{R}^n$,

$$C = \{x \mid |x_i| \leq 1,\ i = 1, \ldots, n\}.$$

The set C can be described in the form (2.5) with $2n$ linear inequalities $\pm e_i^T x \leq 1$, where e_i is the ith unit vector. To describe it in the convex hull form (2.9) requires at least 2^n points:

$$C = \mathbf{conv}\{v_1, \ldots, v_{2^n}\},$$

where $v_1, \ldots, v_{2^n}$ are the 2^n vectors all of whose components are 1 or -1. Thus the size of the two descriptions differs greatly, for large n.

2.2.5 The positive semidefinite cone

We use the notation $\mathbf{S}^n$ to denote the set of symmetric $n \times n$ matrices,

$$\mathbf{S}^n = \{X \in \mathbf{R}^{n \times n} \mid X = X^T\},$$

which is a vector space with dimension $n(n+1)/2$. We use the notation $\mathbf{S}_+^n$ to denote the set of symmetric positive semidefinite matrices:

$$\mathbf{S}_+^n = \{X \in \mathbf{S}^n \mid X \succeq 0\},$$

and the notation $\mathbf{S}_{++}^n$ to denote the set of symmetric positive definite matrices:

$$\mathbf{S}_{++}^n = \{X \in \mathbf{S}^n \mid X \succ 0\}.$$

(This notation is meant to be analogous to $\mathbf{R}_+$, which denotes the nonnegative reals, and $\mathbf{R}_{++}$, which denotes the positive reals.)

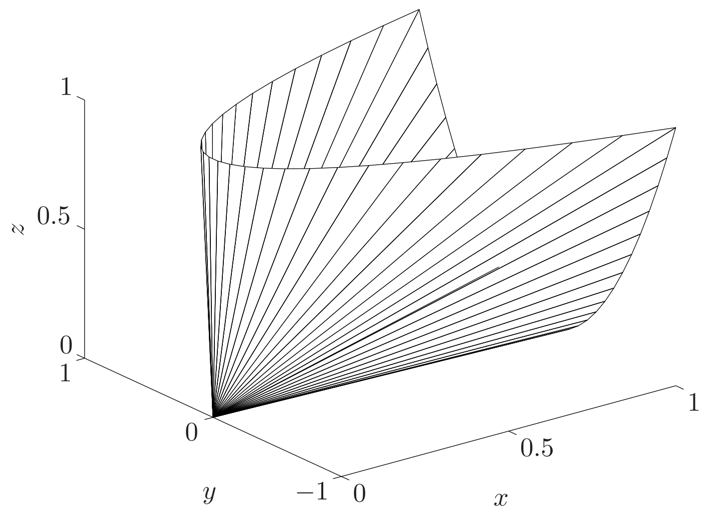

Figure 2.12 Boundary of positive semidefinite cone in $\mathbf{S}^2$.

The set $\mathbf{S}^n_+$ is a convex cone: if $\theta_1, \theta_2 \geq 0$ and $A,\ B \in \mathbf{S}^n_+$, then $\theta_1 A + \theta_2 B \in \mathbf{S}^n_+$. This can be seen directly from the definition of positive semidefiniteness: for any $x \in \mathbf{R}^n$, we have

$$x^T(\theta_1 A + \theta_2 B)x = \theta_1 x^T A x + \theta_2 x^T B x \geq 0,$$

if $A \succeq 0$, $B \succeq 0$ and $\theta_1,\ \theta_2 \geq 0$.

Example 2.6 *Positive semidefinite cone in* $\mathbf{S}^2$. We have

$$X = \begin{bmatrix} x & y \\ y & z \end{bmatrix} \in \mathbf{S}^2_+ \quad \Longleftrightarrow \quad x \geq 0, \quad z \geq 0, \quad xz \geq y^2.$$

The boundary of this cone is shown in figure 2.12, plotted in $\mathbf{R}^3$ as (x, y, z).

2.3 Operations that preserve convexity

In this section we describe some operations that preserve convexity of sets, or allow us to construct convex sets from others. These operations, together with the simple examples described in §2.2, form a calculus of convex sets that is useful for determining or establishing convexity of sets.

2.3.1 Intersection

Convexity is preserved under intersection: if S_1 and S_2 are convex, then $S_1 \cap S_2$ is convex. This property extends to the intersection of an infinite number of sets: if S_α is convex for every $\alpha \in \mathcal{A}$, then $\bigcap_{\alpha\in\mathcal{A}} S_\alpha$ is convex. (Subspaces, affine sets, and convex cones are also closed under arbitrary intersections.) As a simple example, a polyhedron is the intersection of halfspaces and hyperplanes (which are convex), and therefore is convex.

Example 2.7 The positive semidefinite cone $\mathbf{S}^n_+$ can be expressed as

$$\bigcap_{z\neq 0}\{X \in \mathbf{S}^n \mid z^T X z \geq 0\}.$$

For each $z \neq 0$, $z^T X z$ is a (not identically zero) linear function of X, so the sets

$$\{X \in \mathbf{S}^n \mid z^T X z \geq 0\}$$

are, in fact, halfspaces in $\mathbf{S}^n$. Thus the positive semidefinite cone is the intersection of an infinite number of halfspaces, and so is convex.

Example 2.8 We consider the set

$$S = \{x \in \mathbf{R}^m \mid |p(t)| \leq 1 \text{ for } |t| \leq \pi/3\}, \tag{2.10}$$

where $p(t) = \sum_{k=1}^m x_k \cos kt$. The set S can be expressed as the intersection of an infinite number of *slabs*: $S = \bigcap_{|t|\leq\pi/3} S_t$, where

$$S_t = \{x \mid -1 \leq (\cos t, \ldots, \cos mt)^T x \leq 1\},$$

and so is convex. The definition and the set are illustrated in figures 2.13 and 2.14, for $m = 2$.

In the examples above we establish convexity of a set by expressing it as a (possibly infinite) intersection of halfspaces. We will see in §2.5.1 that a converse holds: *every* closed convex set S is a (usually infinite) intersection of halfspaces. In fact, a closed convex set S is the intersection of all halfspaces that contain it:

$$S = \bigcap \{\mathcal{H} \mid \mathcal{H} \text{ halfspace}, \ S \subseteq \mathcal{H}\}.$$

2.3.2 Affine functions

Recall that a function $f : \mathbf{R}^n \to \mathbf{R}^m$ is *affine* if it is a sum of a linear function and a constant, *i.e.*, if it has the form $f(x) = Ax + b$, where $A \in \mathbf{R}^{m\times n}$ and $b \in \mathbf{R}^m$. Suppose $S \subseteq \mathbf{R}^n$ is convex and $f : \mathbf{R}^n \to \mathbf{R}^m$ is an affine function. Then the image of S under f,

$$f(S) = \{f(x) \mid x \in S\},$$

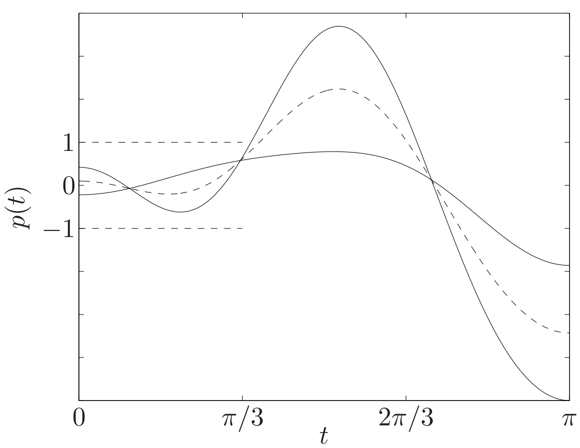

Figure 2.13 Three trigonometric polynomials associated with points in the set S defined in (2.10), for $m = 2$. The trigonometric polynomial plotted with dashed line type is the average of the other two.

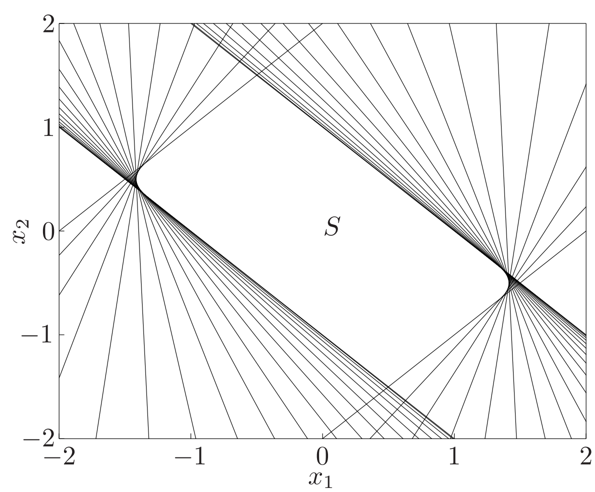

Figure 2.14 The set S defined in (2.10), for $m = 2$, is shown as the white area in the middle of the plot. The set is the intersection of an infinite number of slabs (20 of which are shown), hence convex.

is convex. Similarly, if $f : \mathbf{R}^k \to \mathbf{R}^n$ is an affine function, the *inverse image* of S under f,

$$f^{-1}(S) = \{x \mid f(x) \in S\},$$

is convex.

Two simple examples are *scaling* and *translation*. If $S \subseteq \mathbf{R}^n$ is convex, $\alpha \in \mathbf{R}$, and $a \in \mathbf{R}^n$, then the sets αS and $S + a$ are convex, where

$$\alpha S = \{\alpha x \mid x \in S\}, \qquad S + a = \{x + a \mid x \in S\}.$$

The *projection* of a convex set onto some of its coordinates is convex: if $S \subseteq \mathbf{R}^m \times \mathbf{R}^n$ is convex, then

$$T = \{x_1 \in \mathbf{R}^m \mid (x_1, x_2) \in S \text{ for some } x_2 \in \mathbf{R}^n\}$$

is convex.

The *sum* of two sets is defined as

$$S_1 + S_2 = \{x + y \mid x \in S_1,\ y \in S_2\}.$$

If S_1 and S_2 are convex, then $S_1 + S_2$ is convex. To see this, if S_1 and S_2 are convex, then so is the direct or Cartesian product

$$S_1 \times S_2 = \{(x_1, x_2) \mid x_1 \in S_1,\ x_2 \in S_2\}.$$

The image of this set under the linear function $f(x_1, x_2) = x_1 + x_2$ is the sum $S_1 + S_2$.

We can also consider the *partial sum* of $S_1,\ S_2 \in \mathbf{R}^n \times \mathbf{R}^m$, defined as

$$S = \{(x, y_1 + y_2) \mid (x, y_1) \in S_1,\ (x, y_2) \in S_2\},$$

where $x \in \mathbf{R}^n$ and $y_i \in \mathbf{R}^m$. For $m = 0$, the partial sum gives the intersection of S_1 and S_2; for $n = 0$, it is set addition. Partial sums of convex sets are convex (see exercise 2.16).

Example 2.9 *Polyhedron.* The polyhedron $\{x \mid Ax \preceq b,\ Cx = d\}$ can be expressed as the inverse image of the Cartesian product of the nonnegative orthant and the origin under the affine function $f(x) = (b - Ax, d - Cx)$:

$$\{x \mid Ax \preceq b,\ Cx = d\} = \{x \mid f(x) \in \mathbf{R}^m_+ \times \{0\}\}.$$

Example 2.10 *Solution set of linear matrix inequality.* The condition

$$A(x) = x_1 A_1 + \cdots + x_n A_n \preceq B, \tag{2.11}$$

where $B,\ A_i \in \mathbf{S}^m$, is called a *linear matrix inequality* (LMI) in x. (Note the similarity to an ordinary linear inequality,

$$a^T x = x_1 a_1 + \cdots + x_n a_n \le b,$$

with $b,\ a_i \in \mathbf{R}$.)

The solution set of a linear matrix inequality, $\{x \mid A(x) \preceq B\}$, is convex. Indeed, it is the inverse image of the positive semidefinite cone under the affine function $f : \mathbf{R}^n \to \mathbf{S}^m$ given by $f(x) = B - A(x)$.

Example 2.11 *Hyperbolic cone.* The set

$$\{x \mid x^T P x \le (c^T x)^2,\ c^T x \ge 0\}$$

where $P \in \mathbf{S}^n_+$ and $c \in \mathbf{R}^n$, is convex, since it is the inverse image of the second-order cone,

$$\{(z,t) \mid z^T z \le t^2,\ t \ge 0\},$$

under the affine function $f(x) = (P^{1/2}x, c^T x)$.

Example 2.12 *Ellipsoid.* The ellipsoid

$$\mathcal{E} = \{x \mid (x - x_c)^T P^{-1}(x - x_c) \le 1\},$$

where $P \in \mathbf{S}^n_{++}$, is the image of the unit Euclidean ball $\{u \mid \|u\|_2 \le 1\}$ under the affine mapping $f(u) = P^{1/2}u + x_c$. (It is also the inverse image of the unit ball under the affine mapping $g(x) = P^{-1/2}(x - x_c)$.)

2.3.3 Linear-fractional and perspective functions

In this section we explore a class of functions, called *linear-fractional*, that is more general than affine but still preserves convexity.

The perspective function

We define the *perspective function* $P : \mathbf{R}^{n+1} \to \mathbf{R}^n$, with domain $\mathbf{dom}\, P = \mathbf{R}^n \times \mathbf{R}_{++}$, as $P(z,t) = z/t$. (Here $\mathbf{R}_{++}$ denotes the set of positive numbers: $\mathbf{R}_{++} = \{x \in \mathbf{R} \mid x > 0\}$.) The perspective function scales or normalizes vectors so the last component is one, and then drops the last component.

Remark 2.1 We can interpret the perspective function as the action of a *pin-hole camera*. A pin-hole camera (in $\mathbf{R}^3$) consists of an opaque horizontal plane $x_3 = 0$, with a single pin-hole at the origin, through which light can pass, and a horizontal image plane $x_3 = -1$. An object at x, above the camera (*i.e.*, with $x_3 > 0$), forms an image at the point $-(x_1/x_3, x_2/x_3, 1)$ on the image plane. Dropping the last component of the image point (since it is always -1), the image of a point at x appears at $y = -(x_1/x_3, x_2/x_3) = -P(x)$ on the image plane. This is illustrated in figure 2.15.

If $C \subseteq \mathbf{dom}\, P$ is convex, then its image

$$P(C) = \{P(x) \mid x \in C\}$$

is convex. This result is certainly intuitive: a convex object, viewed through a pin-hole camera, yields a convex image. To establish this fact we show that line segments are mapped to line segments under the perspective function. (This too

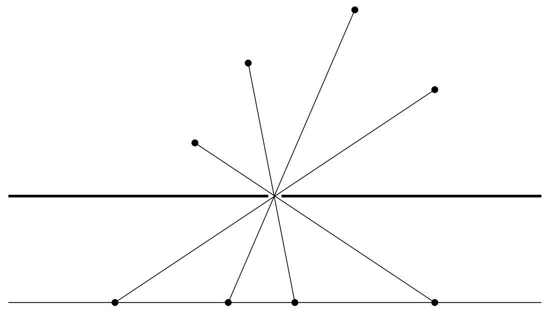

Figure 2.15 Pin-hole camera interpretation of perspective function. The dark horizontal line represents the plane $x_3 = 0$ in $\mathbf{R}^3$, which is opaque, except for a pin-hole at the origin. Objects or light sources above the plane appear on the image plane $x_3 = -1$, which is shown as the lighter horizontal line. The mapping of the position of a source to the position of its image is related to the perspective function.

makes sense: a line segment, viewed through a pin-hole camera, yields a line segment image.) Suppose that $x = (\tilde{x}, x_{n+1}),\ y = (\tilde{y}, y_{n+1}) \in \mathbf{R}^{n+1}$ with $x_{n+1} > 0$, $y_{n+1} > 0$. Then for $0 \leq \theta \leq 1$,

$$P(\theta x + (1-\theta)y) = \frac{\theta\tilde{x} + (1-\theta)\tilde{y}}{\theta x_{n+1} + (1-\theta)y_{n+1}} = \mu P(x) + (1-\mu)P(y),$$

where

$$\mu = \frac{\theta x_{n+1}}{\theta x_{n+1} + (1-\theta)y_{n+1}} \in [0,1].$$

This correspondence between θ and μ is monotonic: as θ varies between 0 and 1 (which sweeps out the line segment $[x, y]$), μ varies between 0 and 1 (which sweeps out the line segment $[P(x), P(y)]$). This shows that $P([x, y]) = [P(x), P(y)]$.

Now suppose C is convex with $C \subseteq \mathbf{dom}\, P$ (*i.e.*, $x_{n+1} > 0$ for all $x \in C$), and $x,\ y \in C$. To establish convexity of $P(C)$ we need to show that the line segment $[P(x), P(y)]$ is in $P(C)$. But this line segment is the image of the line segment $[x, y]$ under P, and so lies in $P(C)$.

The inverse image of a convex set under the perspective function is also convex: if $C \subseteq \mathbf{R}^n$ is convex, then

$$P^{-1}(C) = \{(x, t) \in \mathbf{R}^{n+1} \mid x/t \in C,\ t > 0\}$$

is convex. To show this, suppose $(x, t) \in P^{-1}(C)$, $(y, s) \in P^{-1}(C)$, and $0 \leq \theta \leq 1$. We need to show that

$$\theta(x, t) + (1-\theta)(y, s) \in P^{-1}(C),$$

i.e., that

$$\frac{\theta x + (1-\theta)y}{\theta t + (1-\theta)s} \in C$$

($\theta t + (1-\theta)s > 0$ is obvious). This follows from

$$\frac{\theta x + (1-\theta)y}{\theta t + (1-\theta)s} = \mu(x/t) + (1-\mu)(y/s),$$

where

$$\mu = \frac{\theta t}{\theta t + (1-\theta)s} \in [0,1].$$

Linear-fractional functions

A *linear-fractional function* is formed by composing the perspective function with an affine function. Suppose $g : \mathbf{R}^n \to \mathbf{R}^{m+1}$ is affine, *i.e.*,

$$g(x) = \begin{bmatrix} A \\ c^T \end{bmatrix} x + \begin{bmatrix} b \\ d \end{bmatrix}, \qquad (2.12)$$

where $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$, $c \in \mathbf{R}^n$, and $d \in \mathbf{R}$. The function $f : \mathbf{R}^n \to \mathbf{R}^m$ given by $f = P \circ g$, *i.e.*,

$$f(x) = (Ax+b)/(c^Tx+d), \qquad \mathbf{dom}\, f = \{x \mid c^Tx + d > 0\}, \qquad (2.13)$$

is called a *linear-fractional* (or *projective*) function. If $c = 0$ and $d > 0$, the domain of f is $\mathbf{R}^n$, and f is an affine function. So we can think of affine and linear functions as special cases of linear-fractional functions.

Remark 2.2 *Projective interpretation.* It is often convenient to represent a linear-fractional function as a matrix

$$Q = \begin{bmatrix} A & b \\ c^T & d \end{bmatrix} \in \mathbf{R}^{(m+1)\times(n+1)} \qquad (2.14)$$

that acts on (multiplies) points of form $(x, 1)$, which yields $(Ax + b, c^Tx + d)$. This result is then scaled or normalized so that its last component is one, which yields $(f(x), 1)$.

This representation can be interpreted geometrically by associating $\mathbf{R}^n$ with a set of rays in $\mathbf{R}^{n+1}$ as follows. With each point z in $\mathbf{R}^n$ we associate the (open) ray $\mathcal{P}(z) = \{t(z, 1) \mid t > 0\}$ in $\mathbf{R}^{n+1}$. The last component of this ray takes on positive values. Conversely any ray in $\mathbf{R}^{n+1}$, with base at the origin and last component which takes on positive values, can be written as $\mathcal{P}(v) = \{t(v, 1) \mid t \geq 0\}$ for some $v \in \mathbf{R}^n$. This (projective) correspondence $\mathcal{P}$ between $\mathbf{R}^n$ and the halfspace of rays with positive last component is one-to-one and onto.

The linear-fractional function (2.13) can be expressed as

$$f(x) = \mathcal{P}^{-1}(Q\mathcal{P}(x)).$$

Thus, we start with $x \in \mathbf{dom}\, f$, *i.e.*, $c^Tx + d > 0$. We then form the ray $\mathcal{P}(x)$ in $\mathbf{R}^{n+1}$. The linear transformation with matrix Q acts on this ray to produce another ray $Q\mathcal{P}(x)$. Since $x \in \mathbf{dom}\, f$, the last component of this ray assumes positive values. Finally we take the inverse projective transformation to recover $f(x)$.

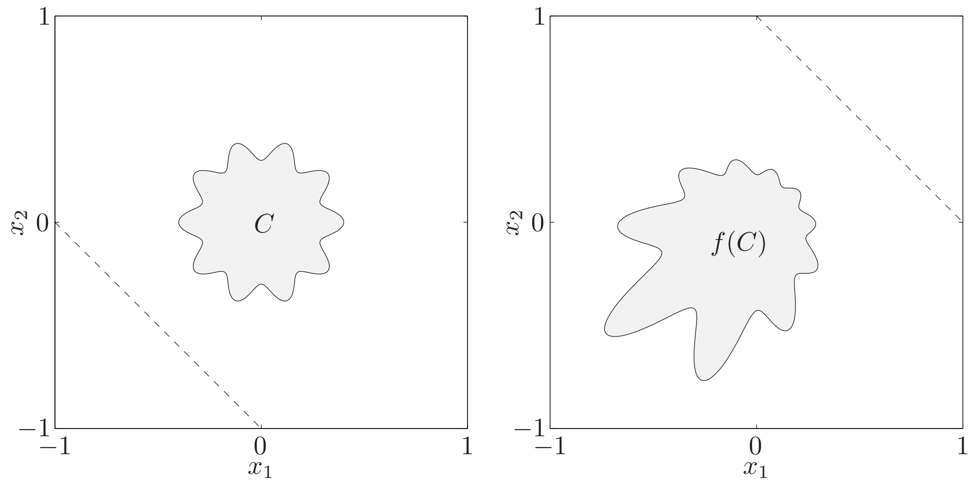

Figure 2.16 *Left.* A set $C \subseteq \mathbf{R}^2$. The dashed line shows the boundary of the domain of the linear-fractional function $f(x) = x/(x_1 + x_2 + 1)$ with $\mathbf{dom}\, f = \{(x_1, x_2) \mid x_1 + x_2 + 1 > 0\}$. *Right.* Image of C under f. The dashed line shows the boundary of the domain of f^{-1}.

Like the perspective function, linear-fractional functions preserve convexity. If C is convex and lies in the domain of f (*i.e.*, $c^T x + d > 0$ for $x \in C$), then its image $f(C)$ is convex. This follows immediately from results above: the image of C under the affine mapping (2.12) is convex, and the image of the resulting set under the perspective function P, which yields $f(C)$, is convex. Similarly, if $C \subseteq \mathbf{R}^m$ is convex, then the inverse image $f^{-1}(C)$ is convex.

Example 2.13 *Conditional probabilities.* Suppose u and v are random variables that take on values in $\{1, \ldots, n\}$ and $\{1, \ldots, m\}$, respectively, and let p_{ij} denote $\mathbf{prob}(u = i, v = j)$. Then the conditional probability $f_{ij} = \mathbf{prob}(u = i | v = j)$ is given by

$$f_{ij} = \frac{p_{ij}}{\sum_{k=1}^{n} p_{kj}}.$$

Thus f is obtained by a linear-fractional mapping from p.

It follows that if C is a convex set of joint probabilities for (u, v), then the associated set of conditional probabilities of u given v is also convex.

Figure 2.16 shows a set $C \subseteq \mathbf{R}^2$, and its image under the linear-fractional function

$$f(x) = \frac{1}{x_1 + x_2 + 1} x, \qquad \mathbf{dom}\, f = \{(x_1, x_2) \mid x_1 + x_2 + 1 > 0\}.$$

2.4 Generalized inequalities

2.4.1 Proper cones and generalized inequalities

A cone $K \subseteq \mathbf{R}^n$ is called a *proper cone* if it satisfies the following:

- K is convex.
- K is closed.
- K is *solid*, which means it has nonempty interior.
- K is *pointed*, which means that it contains no line (or equivalently, $x \in K, \ -x \in K \implies x = 0$).

A proper cone K can be used to define a *generalized inequality*, which is a partial ordering on $\mathbf{R}^n$ that has many of the properties of the standard ordering on $\mathbf{R}$. We associate with the proper cone K the partial ordering on $\mathbf{R}^n$ defined by

$$x \preceq_K y \iff y - x \in K.$$

We also write $x \succeq_K y$ for $y \preceq_K x$. Similarly, we define an associated strict partial ordering by

$$x \prec_K y \iff y - x \in \mathbf{int}\, K,$$

and write $x \succ_K y$ for $y \prec_K x$. (To distinguish the generalized inequality $\preceq_K$ from the strict generalized inequality, we sometimes refer to $\preceq_K$ as the nonstrict generalized inequality.)

When $K = \mathbf{R}_+$, the partial ordering $\preceq_K$ is the usual ordering $\leq$ on $\mathbf{R}$, and the strict partial ordering $\prec_K$ is the same as the usual strict ordering $<$ on $\mathbf{R}$. So generalized inequalities include as a special case ordinary (nonstrict and strict) inequality in $\mathbf{R}$.

Example 2.14 *Nonnegative orthant and componentwise inequality.* The nonnegative orthant $K = \mathbf{R}^n_+$ is a proper cone. The associated generalized inequality $\preceq_K$ corresponds to componentwise inequality between vectors: $x \preceq_K y$ means that $x_i \leq y_i$, $i = 1, \ldots, n$. The associated strict inequality corresponds to componentwise strict inequality: $x \prec_K y$ means that $x_i < y_i$, $i = 1, \ldots, n$.

The nonstrict and strict partial orderings associated with the nonnegative orthant arise so frequently that we drop the subscript $\mathbf{R}^n_+$; it is understood when the symbol $\preceq$ or $\prec$ appears between vectors.

Example 2.15 *Positive semidefinite cone and matrix inequality.* The positive semidefinite cone $\mathbf{S}^n_+$ is a proper cone in $\mathbf{S}^n$. The associated generalized inequality $\preceq_K$ is the usual matrix inequality: $X \preceq_K Y$ means $Y - X$ is positive semidefinite. The interior of $\mathbf{S}^n_+$ (in $\mathbf{S}^n$) consists of the positive definite matrices, so the strict generalized inequality also agrees with the usual strict inequality between symmetric matrices: $X \prec_K Y$ means $Y - X$ is positive definite.

Here, too, the partial ordering arises so frequently that we drop the subscript: for symmetric matrices we write simply $X \preceq Y$ or $X \prec Y$. It is understood that the generalized inequalities are with respect to the positive semidefinite cone.

Example 2.16 *Cone of polynomials nonnegative on* $[0,1]$. Let K be defined as

$$K = \{c \in \mathbf{R}^n \mid c_1 + c_2 t + \cdots + c_n t^{n-1} \geq 0 \text{ for } t \in [0,1]\}, \tag{2.15}$$

i.e., K is the cone of (coefficients of) polynomials of degree $n-1$ that are nonnegative on the interval $[0,1]$. It can be shown that K is a proper cone; its interior is the set of coefficients of polynomials that are positive on the interval $[0,1]$.

Two vectors $c,\ d \in \mathbf{R}^n$ satisfy $c \preceq_K d$ if and only if

$$c_1 + c_2 t + \cdots + c_n t^{n-1} \leq d_1 + d_2 t + \cdots + d_n t^{n-1}$$

for all $t \in [0,1]$.

Properties of generalized inequalities

A generalized inequality $\preceq_K$ satisfies many properties, such as

- $\preceq_K$ *is preserved under addition*: if $x \preceq_K y$ and $u \preceq_K v$, then $x+u \preceq_K y+v$.
- $\preceq_K$ *is transitive*: if $x \preceq_K y$ and $y \preceq_K z$ then $x \preceq_K z$.
- $\preceq_K$ *is preserved under nonnegative scaling*: if $x \preceq_K y$ and $\alpha \geq 0$ then $\alpha x \preceq_K \alpha y$.
- $\preceq_K$ *is reflexive:* $x \preceq_K x$.
- $\preceq_K$ *is antisymmetric:* if $x \preceq_K y$ and $y \preceq_K x$, then $x = y$.
- $\preceq_K$ *is preserved under limits:* if $x_i \preceq_K y_i$ for $i = 1,\ 2, \ldots$, $x_i \to x$ and $y_i \to y$ as $i \to \infty$, then $x \preceq_K y$.

The corresponding strict generalized inequality $\prec_K$ satisfies, for example,

- if $x \prec_K y$ then $x \preceq_K y$.
- if $x \prec_K y$ and $u \preceq_K v$ then $x + u \prec_K y + v$.
- if $x \prec_K y$ and $\alpha > 0$ then $\alpha x \prec_K \alpha y$.
- $x \not\prec_K x$.
- if $x \prec_K y$, then for u and v small enough, $x + u \prec_K y + v$.

These properties are inherited from the definitions of $\preceq_K$ and $\prec_K$, and the properties of proper cones; see exercise 2.30.

2.4.2 Minimum and minimal elements

The notation of generalized inequality (*i.e.*, $\preceq_K$, $\prec_K$) is meant to suggest the analogy to ordinary inequality on $\mathbf{R}$ (*i.e.*, $\leq$, $<$). While many properties of ordinary inequality do hold for generalized inequalities, some important ones do not. The most obvious difference is that $\leq$ on $\mathbf{R}$ is a *linear ordering*: any two points are *comparable*, meaning either $x \leq y$ or $y \leq x$. This property does not hold for other generalized inequalities. One implication is that concepts like minimum and maximum are more complicated in the context of generalized inequalities. We briefly discuss this in this section.

We say that $x \in S$ is the *minimum* element of S (with respect to the generalized inequality $\preceq_K$) if for every $y \in S$ we have $x \preceq_K y$. We define the *maximum* element of a set S, with respect to a generalized inequality, in a similar way. If a set has a minimum (maximum) element, then it is unique. A related concept is *minimal element*. We say that $x \in S$ is a *minimal* element of S (with respect to the generalized inequality $\preceq_K$) if $y \in S$, $y \preceq_K x$ only if $y = x$. We define *maximal* element in a similar way. A set can have many different minimal (maximal) elements.

We can describe minimum and minimal elements using simple set notation. A point $x \in S$ is the minimum element of S if and only if

$$S \subseteq x + K.$$

Here $x + K$ denotes all the points that are comparable to x and greater than or equal to x (according to $\preceq_K$). A point $x \in S$ is a minimal element if and only if

$$(x - K) \cap S = \{x\}.$$

Here $x - K$ denotes all the points that are comparable to x and less than or equal to x (according to $\preceq_K$); the only point in common with S is x.

For $K = \mathbf{R}_+$, which induces the usual ordering on $\mathbf{R}$, the concepts of minimal and minimum are the same, and agree with the usual definition of the minimum element of a set.

Example 2.17 Consider the cone $\mathbf{R}^2_+$, which induces componentwise inequality in $\mathbf{R}^2$. Here we can give some simple geometric descriptions of minimal and minimum elements. The inequality $x \preceq y$ means y is above and to the right of x. To say that $x \in S$ is the minimum element of a set S means that all other points of S lie above and to the right. To say that x is a minimal element of a set S means that no other point of S lies to the left and below x. This is illustrated in figure 2.17.

Example 2.18 *Minimum and minimal elements of a set of symmetric matrices.* We associate with each $A \in \mathbf{S}^n_{++}$ an ellipsoid centered at the origin, given by

$$\mathcal{E}_A = \{x \mid x^T A^{-1} x \leq 1\}.$$

We have $A \preceq B$ if and only if $\mathcal{E}_A \subseteq \mathcal{E}_B$.

Let $v_1, \ldots, v_k \in \mathbf{R}^n$ be given and define

$$S = \{P \in \mathbf{S}^n_{++} \mid v_i^T P^{-1} v_i \leq 1,\ i = 1, \ldots, k\},$$

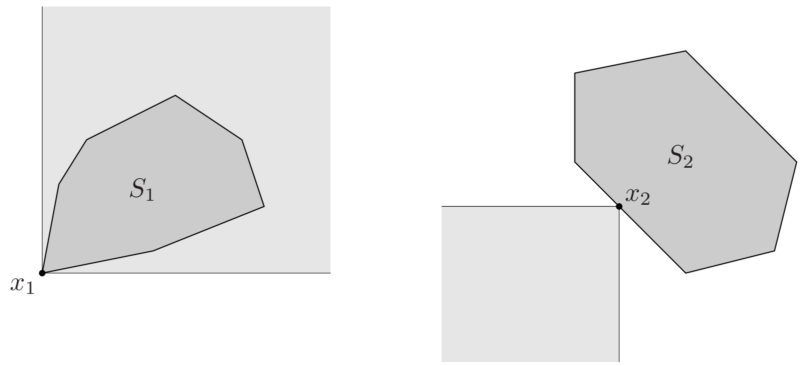

Figure 2.17 *Left.* The set S_1 has a minimum element x_1 with respect to componentwise inequality in $\mathbf{R}^2$. The set $x_1 + K$ is shaded lightly; x_1 is the minimum element of S_1 since $S_1 \subseteq x_1 + K$. *Right.* The point x_2 is a minimal point of S_2. The set $x_2 - K$ is shown lightly shaded. The point x_2 is minimal because $x_2 - K$ and S_2 intersect only at x_2.

which corresponds to the set of ellipsoids that contain the points $v_1, \ldots, v_k$. The set S does not have a minimum element: for any ellipsoid that contains the points $v_1, \ldots, v_k$ we can find another one that contains the points, and is not comparable to it. An ellipsoid is minimal if it contains the points, but no smaller ellipsoid does. Figure 2.18 shows an example in $\mathbf{R}^2$ with $k = 2$.

2.5 Separating and supporting hyperplanes

2.5.1 Separating hyperplane theorem

In this section we describe an idea that will be important later: the use of hyperplanes or affine functions to separate convex sets that do not intersect. The basic result is the *separating hyperplane theorem*: Suppose C and D are nonempty disjoint convex sets, *i.e.*, $C \cap D = \emptyset$. Then there exist $a \neq 0$ and b such that $a^T x \leq b$ for all $x \in C$ and $a^T x \geq b$ for all $x \in D$. In other words, the affine function $a^T x - b$ is nonpositive on C and nonnegative on D. The hyperplane $\{x \mid a^T x = b\}$ is called a *separating hyperplane* for the sets C and D, or is said to *separate* the sets C and D. This is illustrated in figure 2.19.

Proof of separating hyperplane theorem

Here we consider a special case, and leave the extension of the proof to the general case as an exercise (exercise 2.22). We assume that the (Euclidean) *distance* between C and D, defined as

$$\mathbf{dist}(C, D) = \inf\{\|u - v\|_2 \mid u \in C,\ v \in D\},$$

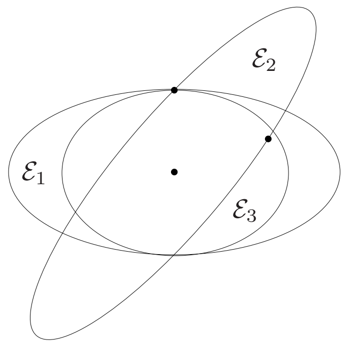

Figure 2.18 Three ellipsoids in $\mathbf{R}^2$, centered at the origin (shown as the lower dot), that contain the points shown as the upper dots. The ellipsoid $\mathcal{E}_1$ is not minimal, since there exist ellipsoids that contain the points, and are smaller (*e.g.*, $\mathcal{E}_3$). $\mathcal{E}_3$ is not minimal for the same reason. The ellipsoid $\mathcal{E}_2$ is minimal, since no other ellipsoid (centered at the origin) contains the points and is contained in $\mathcal{E}_2$.

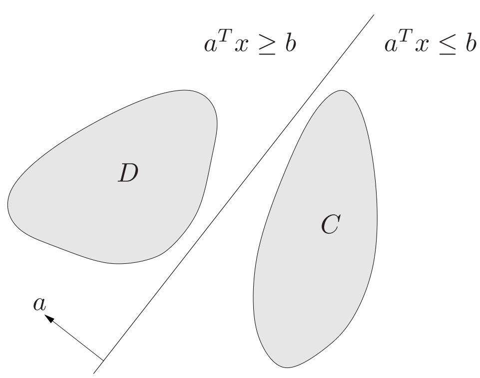

Figure 2.19 The hyperplane $\{x \mid a^T x = b\}$ separates the disjoint convex sets C and D. The affine function $a^T x - b$ is nonpositive on C and nonnegative on D.

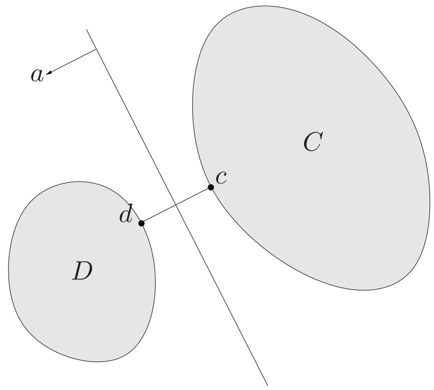

Figure 2.20 Construction of a separating hyperplane between two convex sets. The points $c \in C$ and $d \in D$ are the pair of points in the two sets that are closest to each other. The separating hyperplane is orthogonal to, and bisects, the line segment between c and d.

is positive, and that there exist points $c \in C$ and $d \in D$ that achieve the minimum distance, *i.e.*, $\|c - d\|_2 = \mathbf{dist}(C, D)$. (These conditions are satisfied, for example, when C and D are closed and one set is bounded.)

Define

$$a = d - c, \qquad b = \frac{\|d\|_2^2 - \|c\|_2^2}{2}.$$

We will show that the affine function

$$f(x) = a^T x - b = (d - c)^T (x - (1/2)(d + c))$$

is nonpositive on C and nonnegative on D, *i.e.*, that the hyperplane $\{x \mid a^T x = b\}$ separates C and D. This hyperplane is perpendicular to the line segment between c and d, and passes through its midpoint, as shown in figure 2.20.

We first show that f is nonnegative on D. The proof that f is nonpositive on C is similar (or follows by swapping C and D and considering $-f$). Suppose there were a point $u \in D$ for which

$$f(u) = (d - c)^T (u - (1/2)(d + c)) < 0. \qquad (2.16)$$

We can express $f(u)$ as

$$f(u) = (d - c)^T (u - d + (1/2)(d - c)) = (d - c)^T (u - d) + (1/2)\|d - c\|_2^2.$$

We see that (2.16) implies $(d - c)^T (u - d) < 0$. Now we observe that

$$\left.\frac{d}{dt}\|d + t(u - d) - c\|_2^2\right|_{t=0} = 2(d - c)^T (u - d) < 0,$$

so for some small $t > 0$, with $t \le 1$, we have

$$\|d + t(u - d) - c\|_2 < \|d - c\|_2,$$

i.e., the point $d + t(u - d)$ is closer to c than d is. Since D is convex and contains d and u, we have $d + t(u - d) \in D$. But this is impossible, since d is assumed to be the point in D that is closest to C.

Example 2.19 *Separation of an affine and a convex set.* Suppose C is convex and D is affine, *i.e.*, $D = \{Fu + g \mid u \in \mathbf{R}^m\}$, where $F \in \mathbf{R}^{n \times m}$. Suppose C and D are disjoint, so by the separating hyperplane theorem there are $a \neq 0$ and b such that $a^T x \leq b$ for all $x \in C$ and $a^T x \geq b$ for all $x \in D$.

Now $a^T x \geq b$ for all $x \in D$ means $a^T F u \geq b - a^T g$ for all $u \in \mathbf{R}^m$. But a linear function is bounded below on $\mathbf{R}^m$ only when it is zero, so we conclude $a^T F = 0$ (and hence, $b \leq a^T g$).

Thus we conclude that there exists $a \neq 0$ such that $F^T a = 0$ and $a^T x \leq a^T g$ for all $x \in C$.

Strict separation

The separating hyperplane we constructed above satisfies the stronger condition that $a^T x < b$ for all $x \in C$ and $a^T x > b$ for all $x \in D$. This is called *strict separation* of the sets C and D. Simple examples show that in general, disjoint convex sets need not be strictly separable by a hyperplane (even when the sets are closed; see exercise 2.23). In many special cases, however, strict separation can be established.

Example 2.20 *Strict separation of a point and a closed convex set.* Let C be a closed convex set and $x_0 \notin C$. Then there exists a hyperplane that strictly separates x_0 from C.

To see this, note that the two sets C and $B(x_0, \epsilon)$ do not intersect for some $\epsilon > 0$. By the separating hyperplane theorem, there exist $a \neq 0$ and b such that $a^T x \leq b$ for $x \in C$ and $a^T x \geq b$ for $x \in B(x_0, \epsilon)$.

Using $B(x_0, \epsilon) = \{x_0 + u \mid \|u\|_2 \leq \epsilon\}$, the second condition can be expressed as

$$a^T (x_0 + u) \geq b \quad \text{for all} \quad \|u\|_2 \leq \epsilon.$$

The u that minimizes the lefthand side is $u = -\epsilon a / \|a\|_2$; using this value we have

$$a^T x_0 - \epsilon \|a\|_2 \geq b.$$

Therefore the affine function

$$f(x) = a^T x - b - \epsilon \|a\|_2 / 2$$

is negative on C and positive at x_0.

As an immediate consequence we can establish a fact that we already mentioned above: a closed convex set is the intersection of all halfspaces that contain it. Indeed, let C be closed and convex, and let S be the intersection of all halfspaces containing C. Obviously $x \in C \Rightarrow x \in S$. To show the converse, suppose there exists $x \in S$, $x \notin C$. By the strict separation result there exists a hyperplane that strictly separates x from C, *i.e.*, there is a halfspace containing C but not x. In other words, $x \notin S$.

Converse separating hyperplane theorems

The converse of the separating hyperplane theorem (*i.e.*, existence of a separating hyperplane implies that C and D do not intersect) is not true, unless one imposes additional constraints on C or D, even beyond convexity. As a simple counterexample, consider $C = D = \{0\} \subseteq \mathbf{R}$. Here the hyperplane $x = 0$ separates C and D.

By adding conditions on C and D various converse separation theorems can be derived. As a very simple example, suppose C and D are convex sets, with C open, and there exists an affine function f that is nonpositive on C and nonnegative on D. Then C and D are disjoint. (To see this we first note that f must be negative on C; for if f were zero at a point of C then f would take on positive values near the point, which is a contradiction. But then C and D must be disjoint since f is negative on C and nonnegative on D.) Putting this converse together with the separating hyperplane theorem, we have the following result: any two convex sets C and D, at least one of which is open, are disjoint if and only if there exists a separating hyperplane.

Example 2.21 *Theorem of alternatives for strict linear inequalities.* We derive the necessary and sufficient conditions for solvability of a system of strict linear inequalities

$$Ax \prec b. \tag{2.17}$$

These inequalities are infeasible if and only if the (convex) sets

$$C = \{b - Ax \mid x \in \mathbf{R}^n\}, \qquad D = \mathbf{R}^m_{++} = \{y \in \mathbf{R}^m \mid y \succ 0\}$$

do not intersect. The set D is open; C is an affine set. Hence by the result above, C and D are disjoint if and only if there exists a separating hyperplane, *i.e.*, a nonzero $\lambda \in \mathbf{R}^m$ and $\mu \in \mathbf{R}$ such that $\lambda^T y \leq \mu$ on C and $\lambda^T y \geq \mu$ on D.

Each of these conditions can be simplified. The first means $\lambda^T(b - Ax) \leq \mu$ for all x. This implies (as in example 2.19) that $A^T\lambda = 0$ and $\lambda^T b \leq \mu$. The second inequality means $\lambda^T y \geq \mu$ for all $y \succ 0$. This implies $\mu \leq 0$ and $\lambda \succeq 0$, $\lambda \neq 0$.

Putting it all together, we find that the set of strict inequalities (2.17) is infeasible if and only if there exists $\lambda \in \mathbf{R}^m$ such that

$$\lambda \neq 0, \qquad \lambda \succeq 0, \qquad A^T\lambda = 0, \qquad \lambda^T b \leq 0. \tag{2.18}$$

This is also a system of linear inequalities and linear equations in the variable $\lambda \in \mathbf{R}^m$. We say that (2.17) and (2.18) form a pair of *alternatives*: for any data A and b, exactly one of them is solvable.

2.5.2 Supporting hyperplanes

Suppose $C \subseteq \mathbf{R}^n$, and x_0 is a point in its boundary $\mathbf{bd}\, C$, *i.e.*,

$$x_0 \in \mathbf{bd}\, C = \mathbf{cl}\, C \setminus \mathbf{int}\, C.$$

If $a \neq 0$ satisfies $a^T x \leq a^T x_0$ for all $x \in C$, then the hyperplane $\{x \mid a^T x = a^T x_0\}$ is called a *supporting hyperplane* to C at the point x_0. This is equivalent to saying

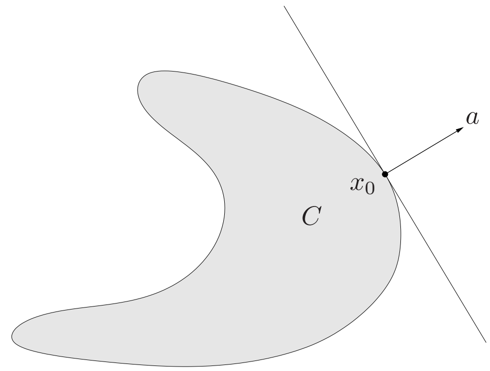

Figure 2.21 The hyperplane $\{x \mid a^T x = a^T x_0\}$ supports C at x_0.

that the point x_0 and the set C are separated by the hyperplane $\{x \mid a^T x = a^T x_0\}$. The geometric interpretation is that the hyperplane $\{x \mid a^T x = a^T x_0\}$ is tangent to C at x_0, and the halfspace $\{x \mid a^T x \le a^T x_0\}$ contains C. This is illustrated in figure 2.21.

A basic result, called the *supporting hyperplane theorem*, states that for any nonempty convex set C, and any $x_0 \in \mathbf{bd}\, C$, there exists a supporting hyperplane to C at x_0. The supporting hyperplane theorem is readily proved from the separating hyperplane theorem. We distinguish two cases. If the interior of C is nonempty, the result follows immediately by applying the separating hyperplane theorem to the sets $\{x_0\}$ and $\mathbf{int}\, C$. If the interior of C is empty, then C must lie in an affine set of dimension less than n, and any hyperplane containing that affine set contains C and x_0, and is a (trivial) supporting hyperplane.

There is also a partial converse of the supporting hyperplane theorem: If a set is closed, has nonempty interior, and has a supporting hyperplane at every point in its boundary, then it is convex. (See exercise 2.27.)

2.6 Dual cones and generalized inequalities

2.6.1 Dual cones

Let K be a cone. The set

$$K^* = \{y \mid x^T y \ge 0 \text{ for all } x \in K\} \tag{2.19}$$

is called the *dual cone* of K. As the name suggests, K^* is a cone, and is always convex, even when the original cone K is not (see exercise 2.31).

Geometrically, $y \in K^*$ if and only if $-y$ is the normal of a hyperplane that supports K at the origin. This is illustrated in figure 2.22.

Example 2.22 *Subspace.* The dual cone of a subspace $V \subseteq \mathbf{R}^n$ (which is a cone) is its orthogonal complement $V^\perp = \{y \mid v^T y = 0 \text{ for all } v \in V\}$.

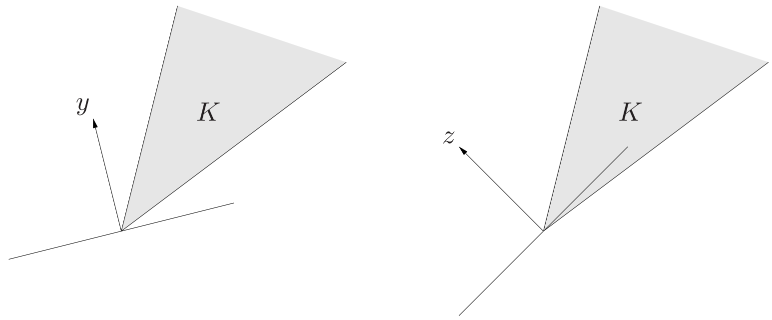

Figure 2.22 *Left.* The halfspace with inward normal y contains the cone K, so $y \in K^*$. *Right.* The halfspace with inward normal z does not contain K, so $z \notin K^*$.

Example 2.23 *Nonnegative orthant.* The cone $\mathbf{R}^n_+$ is its own dual:

$$x^T y \geq 0 \text{ for all } x \succeq 0 \iff y \succeq 0.$$

We call such a cone *self-dual*.

Example 2.24 *Positive semidefinite cone.* On the set of symmetric $n \times n$ matrices $\mathbf{S}^n$, we use the standard inner product $\mathbf{tr}(XY) = \sum_{i,j=1}^n X_{ij} Y_{ij}$ (see §A.1.1). The positive semidefinite cone $\mathbf{S}^n_+$ is self-dual, *i.e.*, for $X,\ Y \in \mathbf{S}^n$,

$$\mathbf{tr}(XY) \geq 0 \text{ for all } X \succeq 0 \iff Y \succeq 0.$$

We will establish this fact.

Suppose $Y \notin \mathbf{S}^n_+$. Then there exists $q \in \mathbf{R}^n$ with

$$q^T Y q = \mathbf{tr}(qq^T Y) < 0.$$

Hence the positive semidefinite matrix $X = qq^T$ satisfies $\mathbf{tr}(XY) < 0$; it follows that $Y \notin (\mathbf{S}^n_+)^*$.

Now suppose $X,\ Y \in \mathbf{S}^n_+$. We can express X in terms of its eigenvalue decomposition as $X = \sum_{i=1}^n \lambda_i q_i q_i^T$, where (the eigenvalues) $\lambda_i \geq 0$, $i = 1, \ldots, n$. Then we have

$$\mathbf{tr}(YX) = \mathbf{tr}\left(Y \sum_{i=1}^n \lambda_i q_i q_i^T \right) = \sum_{i=1}^n \lambda_i q_i^T Y q_i \geq 0.$$

This shows that $Y \in (\mathbf{S}^n_+)^*$.

Example 2.25 *Dual of a norm cone.* Let $\|\cdot\|$ be a norm on $\mathbf{R}^n$. The dual of the associated cone $K = \{(x,t) \in \mathbf{R}^{n+1} \mid \|x\| \leq t\}$ is the cone defined by the dual norm, *i.e.*,

$$K^* = \{(u,v) \in \mathbf{R}^{n+1} \mid \|u\|_* \leq v\},$$

where the dual norm is given by $\|u\|_* = \sup\{u^T x \mid \|x\| \leq 1\}$ (see (A.1.6)).

To prove the result we have to show that

$$x^T u + tv \geq 0 \text{ whenever } \|x\| \leq t \iff \|u\|_* \leq v. \tag{2.20}$$

Let us start by showing that the righthand condition on (u, v) implies the lefthand condition. Suppose $\|u\|_* \leq v$, and $\|x\| \leq t$ for some $t > 0$. (If $t = 0$, x must be zero, so obviously $u^T x + vt \geq 0$.) Applying the definition of the dual norm, and the fact that $\|-x/t\| \leq 1$, we have

$$u^T(-x/t) \leq \|u\|_* \leq v,$$

and therefore $u^T x + vt \geq 0$.

Next we show that the lefthand condition in (2.20) implies the righthand condition in (2.20). Suppose $\|u\|_* > v$, *i.e.*, that the righthand condition does not hold. Then by the definition of the dual norm, there exists an x with $\|x\| \leq 1$ and $x^T u > v$. Taking $t = 1$, we have

$$u^T(-x) + v < 0,$$

which contradicts the lefthand condition in (2.20).

Dual cones satisfy several properties, such as:

- K^* is closed and convex.
- $K_1 \subseteq K_2$ implies $K_2^* \subseteq K_1^*$.
- If K has nonempty interior, then K^* is pointed.
- If the closure of K is pointed then K^* has nonempty interior.
- K^{**} is the closure of the convex hull of K. (Hence if K is convex and closed, $K^{**} = K$.)

(See exercise 2.31.) These properties show that if K is a proper cone, then so is its dual K^*, and moreover, that $K^{**} = K$.

2.6.2 Dual generalized inequalities

Now suppose that the convex cone K is proper, so it induces a generalized inequality $\preceq_K$. Then its dual cone K^* is also proper, and therefore induces a generalized inequality. We refer to the generalized inequality $\preceq_{K^*}$ as the *dual* of the generalized inequality $\preceq_K$.

Some important properties relating a generalized inequality and its dual are:

- $x \preceq_K y$ if and only if $\lambda^T x \leq \lambda^T y$ for all $\lambda \succeq_{K^*} 0$.
- $x \prec_K y$ if and only if $\lambda^T x < \lambda^T y$ for all $\lambda \succeq_{K^*} 0$, $\lambda \neq 0$.

Since $K = K^{**}$, the dual generalized inequality associated with $\preceq_{K^*}$ is $\preceq_K$, so these properties hold if the generalized inequality and its dual are swapped. As a specific example, we have $\lambda \preceq_{K^*} \mu$ if and only if $\lambda^T x \leq \mu^T x$ for all $x \succeq_K 0$.

Example 2.26 *Theorem of alternatives for linear strict generalized inequalities.* Suppose $K \subseteq \mathbf{R}^m$ is a proper cone. Consider the strict generalized inequality

$$Ax \prec_K b, \tag{2.21}$$

where $x \in \mathbf{R}^n$.

We will derive a theorem of alternatives for this inequality. Suppose it is infeasible, *i.e.*, the affine set $\{b - Ax \mid x \in \mathbf{R}^n\}$ does not intersect the open convex set $\mathbf{int}\, K$. Then there is a separating hyperplane, *i.e.*, a nonzero $\lambda \in \mathbf{R}^m$ and $\mu \in \mathbf{R}$ such that $\lambda^T(b - Ax) \leq \mu$ for all x, and $\lambda^T y \geq \mu$ for all $y \in \mathbf{int}\, K$. The first condition implies $A^T\lambda = 0$ and $\lambda^T b \leq \mu$. The second condition implies $\lambda^T y \geq \mu$ for all $y \in K$, which can only happen if $\lambda \in K^*$ and $\mu \leq 0$.

Putting it all together we find that if (2.21) is infeasible, then there exists λ such that

$$\lambda \neq 0, \qquad \lambda \succeq_{K^*} 0, \qquad A^T\lambda = 0, \qquad \lambda^T b \leq 0. \tag{2.22}$$

Now we show the converse: if (2.22) holds, then the inequality system (2.21) cannot be feasible. Suppose that both inequality systems hold. Then we have $\lambda^T(b - Ax) > 0$, since $\lambda \neq 0$, $\lambda \succeq_{K^*} 0$, and $b - Ax \succ_K 0$. But using $A^T\lambda = 0$ we find that $\lambda^T(b - Ax) = \lambda^T b \leq 0$, which is a contradiction.

Thus, the inequality systems (2.21) and (2.22) are alternatives: for any data A, b, exactly one of them is feasible. (This generalizes the alternatives (2.17), (2.18) for the special case $K = \mathbf{R}^m_+$.)

2.6.3 Minimum and minimal elements via dual inequalities

We can use dual generalized inequalities to characterize minimum and minimal elements of a (possibly nonconvex) set $S \subseteq \mathbf{R}^m$ with respect to the generalized inequality induced by a proper cone K.

Dual characterization of minimum element

We first consider a characterization of the *minimum* element: x is the minimum element of S, with respect to the generalized inequality $\preceq_K$, if and only if for all $\lambda \succ_{K^*} 0$, x is the unique minimizer of $\lambda^T z$ over $z \in S$. Geometrically, this means that for any $\lambda \succ_{K^*} 0$, the hyperplane

$$\{z \mid \lambda^T(z - x) = 0\}$$

is a strict supporting hyperplane to S at x. (By strict supporting hyperplane, we mean that the hyperplane intersects S only at the point x.) Note that convexity of the set S is *not* required. This is illustrated in figure 2.23.

To show this result, suppose x is the minimum element of S, *i.e.*, $x \preceq_K z$ for all $z \in S$, and let $\lambda \succ_{K^*} 0$. Let $z \in S$, $z \neq x$. Since x is the minimum element of S, we have $z - x \succeq_K 0$. From $\lambda \succ_{K^*} 0$ and $z - x \succeq_K 0$, $z - x \neq 0$, we conclude $\lambda^T(z - x) > 0$. Since z is an arbitrary element of S, not equal to x, this shows that x is the unique minimizer of $\lambda^T z$ over $z \in S$. Conversely, suppose that for all $\lambda \succ_{K^*} 0$, x is the unique minimizer of $\lambda^T z$ over $z \in S$, but x is not the minimum

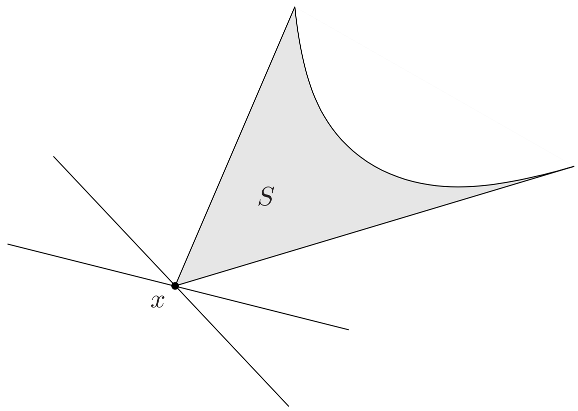

Figure 2.23 Dual characterization of minimum element. The point x is the minimum element of the set S with respect to $\mathbf{R}^2_+$. This is equivalent to: for every $\lambda \succ 0$, the hyperplane $\{z \mid \lambda^T(z-x) = 0\}$ strictly supports S at x, *i.e.*, contains S on one side, and touches it only at x.

element of S. Then there exists $z \in S$ with $z \not\succeq_K x$. Since $z - x \not\succeq_K 0$, there exists $\tilde{\lambda} \succeq_{K^*} 0$ with $\tilde{\lambda}^T(z-x) < 0$. Hence $\lambda^T(z-x) < 0$ for $\lambda \succ_{K^*} 0$ in the neighborhood of $\tilde{\lambda}$. This contradicts the assumption that x is the unique minimizer of $\lambda^T z$ over S.

Dual characterization of minimal elements

We now turn to a similar characterization of *minimal elements*. Here there is a gap between the necessary and sufficient conditions. If $\lambda \succ_{K^*} 0$ and x minimizes $\lambda^T z$ over $z \in S$, then x is minimal. This is illustrated in figure 2.24.

To show this, suppose that $\lambda \succ_{K^*} 0$, and x minimizes $\lambda^T z$ over S, but x is not minimal, *i.e.*, there exists a $z \in S$, $z \neq x$, and $z \preceq_K x$. Then $\lambda^T(x-z) > 0$, which contradicts our assumption that x is the minimizer of $\lambda^T z$ over S.

The converse is in general false: a point x can be minimal in S, but not a minimizer of $\lambda^T z$ over $z \in S$, for any λ, as shown in figure 2.25. This figure suggests that convexity plays an important role in the converse, which is correct. Provided the set S is convex, we can say that for any minimal element x there exists a nonzero $\lambda \succeq_{K^*} 0$ such that x minimizes $\lambda^T z$ over $z \in S$.

To show this, suppose x is minimal, which means that $((x-K) \setminus \{x\}) \cap S = \emptyset$. Applying the separating hyperplane theorem to the convex sets $(x-K) \setminus \{x\}$ and S, we conclude that there is a $\lambda \neq 0$ and μ such that $\lambda^T(x-y) \leq \mu$ for all $y \in K$, and $\lambda^T z \geq \mu$ for all $z \in S$. From the first inequality we conclude $\lambda \succeq_{K^*} 0$. Since $x \in S$ and $x \in x - K$, we have $\lambda^T x = \mu$, so the second inequality implies that μ is the minimum value of $\lambda^T z$ over S. Therefore, x is a minimizer of $\lambda^T z$ over S, where $\lambda \neq 0$, $\lambda \succeq_{K^*} 0$.

This converse theorem cannot be strengthened to $\lambda \succ_{K^*} 0$. Examples show that a point x can be a minimal point of a convex set S, but not a minimizer of

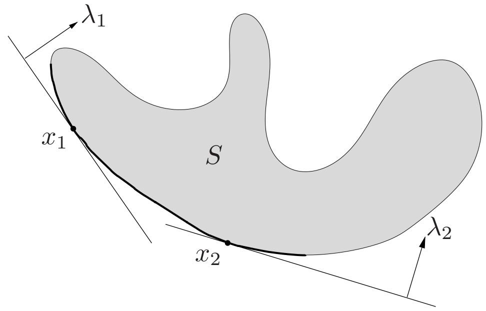

Figure 2.24 A set $S \subseteq \mathbf{R}^2$. Its set of minimal points, with respect to $\mathbf{R}^2_+$, is shown as the darker section of its (lower, left) boundary. The minimizer of $\lambda_1^T z$ over S is x_1, and is minimal since $\lambda_1 \succ 0$. The minimizer of $\lambda_2^T z$ over S is x_2, which is another minimal point of S, since $\lambda_2 \succ 0$.

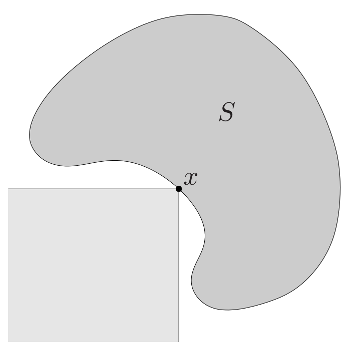

Figure 2.25 The point x is a minimal element of $S \subseteq \mathbf{R}^2$ with respect to $\mathbf{R}^2_+$. However there exists no λ for which x minimizes $\lambda^T z$ over $z \in S$.

Figure 2.26 *Left.* The point $x_1 \in S_1$ is minimal, but is not a minimizer of $\lambda^T z$ over S_1 for any $\lambda \succ 0$. (It does, however, minimize $\lambda^T z$ over $z \in S_1$ for $\lambda = (1,0)$.) *Right.* The point $x_2 \in S_2$ is *not* minimal, but it does minimize $\lambda^T z$ over $z \in S_2$ for $\lambda = (0,1) \succeq 0$.

$\lambda^T z$ over $z \in S$ for any $\lambda \succ_{K^*} 0$. (See figure 2.26, left.) Nor is it true that any minimizer of $\lambda^T z$ over $z \in S$, with $\lambda \succeq_{K^*} 0$, is minimal (see figure 2.26, right.)

Example 2.27 *Pareto optimal production frontier.* We consider a product which requires n resources (such as labor, electricity, natural gas, water) to manufacture. The product can be manufactured or produced in many ways. With each production method, we associate a *resource vector* $x \in \mathbf{R}^n$, where x_i denotes the amount of resource i consumed by the method to manufacture the product. We assume that $x_i \geq 0$ (*i.e.*, resources are consumed by the production methods) and that the resources are valuable (so using less of any resource is preferred).

The *production set* $P \subseteq \mathbf{R}^n$ is defined as the set of all resource vectors x that correspond to some production method.

Production methods with resource vectors that are minimal elements of P, with respect to componentwise inequality, are called *Pareto optimal* or *efficient*. The set of minimal elements of P is called the *efficient production frontier*.

We can give a simple interpretation of Pareto optimality. We say that one production method, with resource vector x, is *better* than another, with resource vector y, if $x_i \leq y_i$ for all i, and for some i, $x_i < y_i$. In other words, one production method is better than another if it uses no more of each resource than another method, and for at least one resource, actually uses less. This corresponds to $x \preceq y$, $x \neq y$. Then we can say: A production method is Pareto optimal or efficient if there is no better production method.

We can find Pareto optimal production methods (*i.e.*, minimal resource vectors) by minimizing

$$\lambda^T x = \lambda_1 x_1 + \cdots + \lambda_n x_n$$

over the set P of production vectors, using any λ that satisfies $\lambda \succ 0$.

Here the vector λ has a simple interpretation: λ_i is the *price* of resource i. By minimizing $\lambda^T x$ over P we are finding the overall cheapest production method (for the resource prices λ_i). As long as the prices are positive, the resulting production method is guaranteed to be efficient.

These ideas are illustrated in figure 2.27.

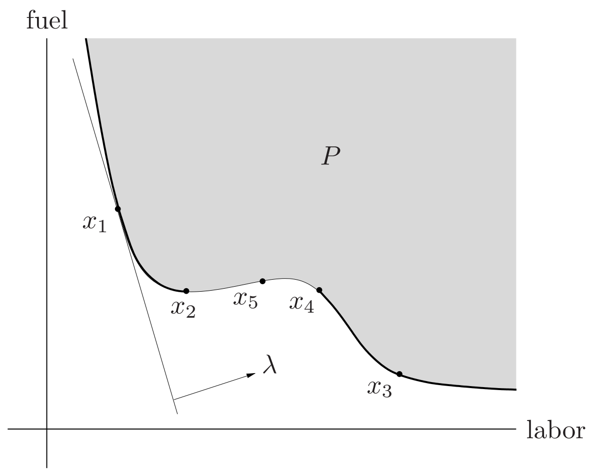

Figure 2.27 The production set P, for a product that requires labor and fuel to produce, is shown shaded. The two dark curves show the efficient production frontier. The points x_1, x_2 and x_3 are efficient. The points x_4 and x_5 are not (since in particular, x_2 corresponds to a production method that uses no more fuel, and less labor). The point x_1 is also the minimum cost production method for the price vector λ (which is positive). The point x_2 is efficient, but cannot be found by minimizing the total cost $\lambda^T x$ for any price vector $\lambda \succeq 0$.

Bibliography

Minkowski is generally credited with the first systematic study of convex sets, and the introduction of fundamental concepts such as supporting hyperplanes and the supporting hyperplane theorem, the Minkowski distance function (exercise 3.34), extreme points of a convex set, and many others.

Some well known early surveys are Bonnesen and Fenchel [BF48], Eggleston [Egg58], Klee [Kle63], and Valentine [Val64]. More recent books devoted to the geometry of convex sets include Lay [Lay82] and Webster [Web94]. Klee [Kle71], Fenchel [Fen83], Tikhomorov [Tik90], and Berger [Ber90] give very readable overviews of the history of convexity and its applications throughout mathematics.

Linear inequalities and polyhedral sets are studied extensively in connection with the linear programming problem, for which we give references at the end of chapter 4. Some landmark publications in the history of linear inequalities and linear programming are Motzkin [Mot33], von Neumann and Morgenstern [vNM53], Kantorovich [Kan60], Koopmans [Koo51], and Dantzig [Dan63]. Dantzig [Dan63, Chapter 2] includes an historical survey of linear inequalities, up to around 1963.

Generalized inequalities were introduced in nonlinear optimization during the 1960s (see Luenberger [Lue69, §8.2] and Isii [Isi64]), and are used extensively in cone programming (see the references in chapter 4). Bellman and Fan [BF63] is an early paper on sets of generalized linear inequalities (with respect to the positive semidefinite cone).

For extensions and a proof of the separating hyperplane theorem we refer the reader to Rockafellar [Roc70, part III], and Hiriart-Urruty and Lemaréchal [HUL93, volume 1, §III4]. Dantzig [Dan63, page 21] attributes the term *theorem of the alternative* to von Neumann and Morgenstern [vNM53, page 138]. For more references on theorems of alternatives, see chapter 5.

The terminology of example 2.27 (including Pareto optimality, efficient production, and the price interpretation of λ) is discussed in detail by Luenberger [Lue95].

Convex geometry plays a prominent role in the classical theory of moments (Krein and Nudelman [KN77], Karlin and Studden [KS66]). A famous example is the duality between the cone of nonnegative polynomials and the cone of power moments; see exercise 2.37.

Exercises

Definition of convexity

2.1 Let $C \subseteq \mathbf{R}^n$ be a convex set, with $x_1, \ldots, x_k \in C$, and let $\theta_1, \ldots, \theta_k \in \mathbf{R}$ satisfy $\theta_i \geq 0$, $\theta_1 + \cdots + \theta_k = 1$. Show that $\theta_1 x_1 + \cdots + \theta_k x_k \in C$. (The definition of convexity is that this holds for $k = 2$; you must show it for arbitrary k.) *Hint.* Use induction on k.

2.2 Show that a set is convex if and only if its intersection with any line is convex. Show that a set is affine if and only if its intersection with any line is affine.

2.3 *Midpoint convexity.* A set C is *midpoint convex* if whenever two points a, b are in C, the average or midpoint $(a+b)/2$ is in C. Obviously a convex set is midpoint convex. It can be proved that under mild conditions midpoint convexity implies convexity. As a simple case, prove that if C is closed and midpoint convex, then C is convex.

2.4 Show that the convex hull of a set S is the intersection of all convex sets that contain S. (The same method can be used to show that the conic, or affine, or linear hull of a set S is the intersection of all conic sets, or affine sets, or subspaces that contain S.)

Examples

2.5 What is the distance between two parallel hyperplanes $\{x \in \mathbf{R}^n \mid a^T x = b_1\}$ and $\{x \in \mathbf{R}^n \mid a^T x = b_2\}$?

2.6 *When does one halfspace contain another?* Give conditions under which

$$\{x \mid a^T x \leq b\} \subseteq \{x \mid \tilde{a}^T x \leq \tilde{b}\}$$

(where $a \neq 0$, $\tilde{a} \neq 0$). Also find the conditions under which the two halfspaces are equal.

2.7 *Voronoi description of halfspace.* Let a and b be distinct points in $\mathbf{R}^n$. Show that the set of all points that are closer (in Euclidean norm) to a than b, *i.e.*, $\{x \mid \|x-a\|_2 \leq \|x-b\|_2\}$, is a halfspace. Describe it explicitly as an inequality of the form $c^T x \leq d$. Draw a picture.

2.8 Which of the following sets S are polyhedra? If possible, express S in the form $S = \{x \mid Ax \preceq b,\ Fx = g\}$.

(a) $S = \{y_1 a_1 + y_2 a_2 \mid -1 \leq y_1 \leq 1,\ -1 \leq y_2 \leq 1\}$, where $a_1, a_2 \in \mathbf{R}^n$.

(b) $S = \{x \in \mathbf{R}^n \mid x \succeq 0,\ \mathbf{1}^T x = 1,\ \sum_{i=1}^n x_i a_i = b_1,\ \sum_{i=1}^n x_i a_i^2 = b_2\}$, where $a_1, \ldots, a_n \in \mathbf{R}$ and $b_1, b_2 \in \mathbf{R}$.

(c) $S = \{x \in \mathbf{R}^n \mid x \succeq 0,\ x^T y \leq 1 \text{ for all } y \text{ with } \|y\|_2 = 1\}$.

(d) $S = \{x \in \mathbf{R}^n \mid x \succeq 0,\ x^T y \leq 1 \text{ for all } y \text{ with } \sum_{i=1}^n |y_i| = 1\}$.

2.9 *Voronoi sets and polyhedral decomposition.* Let $x_0, \ldots, x_K \in \mathbf{R}^n$ be distinct. Consider the set of points that are closer (in Euclidean norm) to x_0 than the other x_i, *i.e.*,

$$V = \{x \in \mathbf{R}^n \mid \|x - x_0\|_2 \leq \|x - x_i\|_2,\ i = 1, \ldots, K\}.$$

V is called the *Voronoi region* around x_0 with respect to $x_1, \ldots, x_K$.

(a) Show that V is a polyhedron. Express V in the form $V = \{x \mid Ax \preceq b\}$.

(b) Conversely, given a polyhedron P with nonempty interior, show how to find $x_0, \ldots, x_K$ so that the polyhedron is the Voronoi region of x_0 with respect to $x_1, \ldots, x_K$.

(c) We can also consider the sets

$$V_k = \{x \in \mathbf{R}^n \mid \|x - x_k\|_2 \leq \|x - x_i\|_2,\ i \neq k\}.$$

The set V_k consists of points in $\mathbf{R}^n$ for which the closest point in the set $\{x_0, \ldots, x_K\}$ is x_k.

The sets $V_0, \ldots, V_K$ give a polyhedral decomposition of $\mathbf{R}^n$. More precisely, the sets V_k are polyhedra with nonempty interior, $\bigcup_{k=0}^K V_k = \mathbf{R}^n$, and $\mathbf{int}\, V_i \cap \mathbf{int}\, V_j = \emptyset$ for $i \neq j$, *i.e.*, V_i and V_j intersect at most along a boundary.

Suppose that $P_1, \ldots, P_m$ are polyhedra with nonempty interior such that $\bigcup_{i=1}^m P_i = \mathbf{R}^n$, $\mathbf{int}\, P_i \cap \mathbf{int}\, P_j = \emptyset$ for $i \neq j$. Can this polyhedral decomposition of $\mathbf{R}^n$ be described as the Voronoi regions generated by an appropriate set of points?

2.10 *Solution set of a quadratic inequality.* Let $C \subseteq \mathbf{R}^n$ be the solution set of a quadratic inequality,

$$C = \{x \in \mathbf{R}^n \mid x^T A x + b^T x + c \leq 0\},$$

with $A \in \mathbf{S}^n$, $b \in \mathbf{R}^n$, and $c \in \mathbf{R}$.

(a) Show that C is convex if $A \succeq 0$.

(b) Show that the intersection of C and the hyperplane defined by $g^T x + h = 0$ (where $g \neq 0$) is convex if $A + \lambda g g^T \succeq 0$ for some $\lambda \in \mathbf{R}$.

Are the converses of these statements true?

2.11 *Hyperbolic sets.* Show that the *hyperbolic set* $\{x \in \mathbf{R}_+^2 \mid x_1 x_2 \geq 1\}$ is convex. As a generalization, show that $\{x \in \mathbf{R}_+^n \mid \prod_{i=1}^n x_i \geq 1\}$ is convex. *Hint.* If $a, b \geq 0$ and $0 \leq \theta \leq 1$, then $a^\theta b^{1-\theta} \leq \theta a + (1-\theta) b$; see §3.1.9.

2.12 Which of the following sets are convex?

(a) A *slab*, *i.e.*, a set of the form $\{x \in \mathbf{R}^n \mid \alpha \leq a^T x \leq \beta\}$.

(b) A *rectangle*, *i.e.*, a set of the form $\{x \in \mathbf{R}^n \mid \alpha_i \leq x_i \leq \beta_i,\ i = 1, \ldots, n\}$. A rectangle is sometimes called a *hyperrectangle* when $n > 2$.

(c) A *wedge*, *i.e.*, $\{x \in \mathbf{R}^n \mid a_1^T x \leq b_1,\ a_2^T x \leq b_2\}$.

(d) The set of points closer to a given point than a given set, *i.e.*,

$$\{x \mid \|x - x_0\|_2 \leq \|x - y\|_2 \text{ for all } y \in S\}$$

where $S \subseteq \mathbf{R}^n$.

(e) The set of points closer to one set than another, *i.e.*,

$$\{x \mid \mathbf{dist}(x, S) \leq \mathbf{dist}(x, T)\},$$

where $S, T \subseteq \mathbf{R}^n$, and

$$\mathbf{dist}(x, S) = \inf\{\|x - z\|_2 \mid z \in S\}.$$

(f) [HUL93, volume 1, page 93] The set $\{x \mid x + S_2 \subseteq S_1\}$, where $S_1, S_2 \subseteq \mathbf{R}^n$ with S_1 convex.

(g) The set of points whose distance to a does not exceed a fixed fraction θ of the distance to b, *i.e.*, the set $\{x \mid \|x - a\|_2 \leq \theta \|x - b\|_2\}$. You can assume $a \neq b$ and $0 \leq \theta \leq 1$.

2.13 *Conic hull of outer products.* Consider the set of rank-k *outer products*, defined as $\{XX^T \mid X \in \mathbf{R}^{n \times k},\ \mathbf{rank}\, X = k\}$. Describe its conic hull in simple terms.

2.14 *Expanded and restricted sets.* Let $S \subseteq \mathbf{R}^n$, and let $\|\cdot\|$ be a norm on $\mathbf{R}^n$.

(a) For $a \geq 0$ we define S_a as $\{x \mid \mathbf{dist}(x, S) \leq a\}$, where $\mathbf{dist}(x, S) = \inf_{y \in S} \|x - y\|$. We refer to S_a as S *expanded* or *extended* by a. Show that if S is convex, then S_a is convex.

(b) For $a \geq 0$ we define $S_{-a} = \{x \mid B(x, a) \subseteq S\}$, where $B(x, a)$ is the ball (in the norm $\|\cdot\|$), centered at x, with radius a. We refer to S_{-a} as S *shrunk* or *restricted* by a, since S_{-a} consists of all points that are at least a distance a from $\mathbf{R}^n \backslash S$. Show that if S is convex, then S_{-a} is convex.

2.15 *Some sets of probability distributions.* Let x be a real-valued random variable with $\mathbf{prob}(x = a_i) = p_i$, $i = 1, \ldots, n$, where $a_1 < a_2 < \cdots < a_n$. Of course $p \in \mathbf{R}^n$ lies in the standard probability simplex $P = \{p \mid \mathbf{1}^T p = 1,\ p \succeq 0\}$. Which of the following conditions are convex in p? (That is, for which of the following conditions is the set of $p \in P$ that satisfy the condition convex?)

(a) $\alpha \leq \mathbf{E} f(x) \leq \beta$, where $\mathbf{E} f(x)$ is the expected value of $f(x)$, *i.e.*, $\mathbf{E} f(x) = \sum_{i=1}^n p_i f(a_i)$. (The function $f : \mathbf{R} \to \mathbf{R}$ is given.)

(b) $\mathbf{prob}(x > \alpha) \leq \beta$.

(c) $\mathbf{E}\, |x^3| \leq \alpha\, \mathbf{E}\, |x|$.

(d) $\mathbf{E}\, x^2 \leq \alpha$.

(e) $\mathbf{E}\, x^2 \geq \alpha$.

(f) $\mathbf{var}(x) \leq \alpha$, where $\mathbf{var}(x) = \mathbf{E}(x - \mathbf{E}\, x)^2$ is the variance of x.

(g) $\mathbf{var}(x) \geq \alpha$.

(h) $\mathbf{quartile}(x) \geq \alpha$, where $\mathbf{quartile}(x) = \inf\{\beta \mid \mathbf{prob}(x \leq \beta) \geq 0.25\}$.

(i) $\mathbf{quartile}(x) \leq \alpha$.

Operations that preserve convexity

2.16 Show that if S_1 and S_2 are convex sets in $\mathbf{R}^{m+n}$, then so is their partial sum

$$S = \{(x, y_1 + y_2) \mid x \in \mathbf{R}^m,\ y_1,\ y_2 \in \mathbf{R}^n, (x, y_1) \in S_1,\ (x, y_2) \in S_2\}.$$

2.17 *Image of polyhedral sets under perspective function.* In this problem we study the image of hyperplanes, halfspaces, and polyhedra under the perspective function $P(x,t) = x/t$, with $\mathbf{dom}\, P = \mathbf{R}^n \times \mathbf{R}_{++}$. For each of the following sets C, give a simple description of

$$P(C) = \{v/t \mid (v,t) \in C,\ t > 0\}.$$

(a) The polyhedron $C = \mathbf{conv}\{(v_1, t_1), \ldots, (v_K, t_K)\}$ where $v_i \in \mathbf{R}^n$ and $t_i > 0$.

(b) The hyperplane $C = \{(v,t) \mid f^T v + gt = h\}$ (with f and g not both zero).

(c) The halfspace $C = \{(v,t) \mid f^T v + gt \leq h\}$ (with f and g not both zero).

(d) The polyhedron $C = \{(v,t) \mid Fv + gt \preceq h\}$.

2.18 *Invertible linear-fractional functions.* Let $f : \mathbf{R}^n \to \mathbf{R}^n$ be the linear-fractional function

$$f(x) = (Ax + b)/(c^T x + d), \qquad \mathbf{dom}\, f = \{x \mid c^T x + d > 0\}.$$

Suppose the matrix

$$Q = \begin{bmatrix} A & b \\ c^T & d \end{bmatrix}$$

is nonsingular. Show that f is invertible and that f^{-1} is a linear-fractional mapping. Give an explicit expression for f^{-1} and its domain in terms of A, b, c, and d. *Hint.* It may be easier to express f^{-1} in terms of Q.

2.19 *Linear-fractional functions and convex sets.* Let $f : \mathbf{R}^m \to \mathbf{R}^n$ be the linear-fractional function

$$f(x) = (Ax + b)/(c^T x + d), \qquad \mathbf{dom}\, f = \{x \mid c^T x + d > 0\}.$$

In this problem we study the inverse image of a convex set C under f, *i.e.*,

$$f^{-1}(C) = \{x \in \mathbf{dom}\, f \mid f(x) \in C\}.$$

For each of the following sets $C \subseteq \mathbf{R}^n$, give a simple description of $f^{-1}(C)$.

(a) The halfspace $C = \{y \mid g^T y \leq h\}$ (with $g \neq 0$).

(b) The polyhedron $C = \{y \mid Gy \preceq h\}$.

(c) The ellipsoid $\{y \mid y^T P^{-1} y \leq 1\}$ (where $P \in \mathbf{S}^n_{++}$).

(d) The solution set of a linear matrix inequality, $C = \{y \mid y_1 A_1 + \cdots + y_n A_n \preceq B\}$, where $A_1, \ldots, A_n, B \in \mathbf{S}^p$.

Separation theorems and supporting hyperplanes

2.20 *Strictly positive solution of linear equations.* Suppose $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$, with $b \in \mathcal{R}(A)$. Show that there exists an x satisfying

$$x \succ 0, \qquad Ax = b$$

if and only if there exists no λ with

$$A^T\lambda \succeq 0, \qquad A^T\lambda \neq 0, \qquad b^T\lambda \leq 0.$$

Hint. First prove the following fact from linear algebra: $c^Tx = d$ for all x satisfying $Ax = b$ if and only if there is a vector λ such that $c = A^T\lambda$, $d = b^T\lambda$.

2.21 *The set of separating hyperplanes.* Suppose that C and D are disjoint subsets of $\mathbf{R}^n$. Consider the set of $(a, b) \in \mathbf{R}^{n+1}$ for which $a^Tx \leq b$ for all $x \in C$, and $a^Tx \geq b$ for all $x \in D$. Show that this set is a convex cone (which is the singleton $\{0\}$ if there is no hyperplane that separates C and D).

2.22 Finish the proof of the separating hyperplane theorem in §2.5.1: Show that a separating hyperplane exists for two disjoint convex sets C and D. You can use the result proved in §2.5.1, *i.e.*, that a separating hyperplane exists when there exist points in the two sets whose distance is equal to the distance between the two sets.

Hint. If C and D are disjoint convex sets, then the set $\{x - y \mid x \in C,\ y \in D\}$ is convex and does not contain the origin.

2.23 Give an example of two closed convex sets that are disjoint but cannot be strictly separated.

2.24 *Supporting hyperplanes.*

(a) Express the closed convex set $\{x \in \mathbf{R}^2_+ \mid x_1x_2 \geq 1\}$ as an intersection of halfspaces.

(b) Let $C = \{x \in \mathbf{R}^n \mid \|x\|_\infty \leq 1\}$, the ℓ_∞-norm unit ball in $\mathbf{R}^n$, and let $\hat{x}$ be a point in the boundary of C. Identify the supporting hyperplanes of C at $\hat{x}$ explicitly.

2.25 *Inner and outer polyhedral approximations.* Let $C \subseteq \mathbf{R}^n$ be a closed convex set, and suppose that $x_1, \ldots, x_K$ are on the boundary of C. Suppose that for each i, $a_i^T(x - x_i) = 0$ defines a supporting hyperplane for C at x_i, *i.e.*, $C \subseteq \{x \mid a_i^T(x - x_i) \leq 0\}$. Consider the two polyhedra

$$P_{\mathrm{inner}} = \mathbf{conv}\{x_1, \ldots, x_K\}, \qquad P_{\mathrm{outer}} = \{x \mid a_i^T(x - x_i) \leq 0,\ i = 1, \ldots, K\}.$$

Show that $P_{\mathrm{inner}} \subseteq C \subseteq P_{\mathrm{outer}}$. Draw a picture illustrating this.

2.26 *Support function.* The support function of a set $C \subseteq \mathbf{R}^n$ is defined as

$$S_C(y) = \sup\{y^Tx \mid x \in C\}.$$

(We allow $S_C(y)$ to take on the value $+\infty$.) Suppose that C and D are closed convex sets in $\mathbf{R}^n$. Show that $C = D$ if and only if their support functions are equal.

2.27 *Converse supporting hyperplane theorem.* Suppose the set C is closed, has nonempty interior, and has a supporting hyperplane at every point in its boundary. Show that C is convex.

Convex cones and generalized inequalities

2.28 *Positive semidefinite cone for $n = 1,\ 2,\ 3$.* Give an explicit description of the positive semidefinite cone $\mathbf{S}^n_+$, in terms of the matrix coefficients and ordinary inequalities, for $n = 1,\ 2,\ 3$. To describe a general element of $\mathbf{S}^n$, for $n = 1,\ 2,\ 3$, use the notation

$$x_1, \qquad \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix}, \qquad \begin{bmatrix} x_1 & x_2 & x_3 \\ x_2 & x_4 & x_5 \\ x_3 & x_5 & x_6 \end{bmatrix}.$$

2.29 *Cones in* $\mathbf{R}^2$. Suppose $K \subseteq \mathbf{R}^2$ is a closed convex cone.

(a) Give a simple description of K in terms of the polar coordinates of its elements ($x = r(\cos\phi, \sin\phi)$ with $r \geq 0$).

(b) Give a simple description of K^*, and draw a plot illustrating the relation between K and K^*.

(c) When is K pointed?

(d) When is K proper (hence, defines a generalized inequality)? Draw a plot illustrating what $x \preceq_K y$ means when K is proper.

2.30 *Properties of generalized inequalities.* Prove the properties of (nonstrict and strict) generalized inequalities listed in §2.4.1.

2.31 *Properties of dual cones.* Let K^* be the dual cone of a convex cone K, as defined in (2.19). Prove the following.

(a) K^* is indeed a convex cone.

(b) $K_1 \subseteq K_2$ implies $K_2^* \subseteq K_1^*$.

(c) K^* is closed.

(d) The interior of K^* is given by $\mathbf{int}\, K^* = \{y \mid y^T x > 0 \text{ for all } x \in \mathbf{cl}\, K\}$.

(e) If K has nonempty interior then K^* is pointed.

(f) K^{**} is the closure of K. (Hence if K is closed, $K^{**} = K$.)

(g) If the closure of K is pointed then K^* has nonempty interior.

2.32 Find the dual cone of $\{Ax \mid x \succeq 0\}$, where $A \in \mathbf{R}^{m\times n}$.

2.33 *The monotone nonnegative cone.* We define the *monotone nonnegative cone* as

$$K_{\mathrm{m+}} = \{x \in \mathbf{R}^n \mid x_1 \geq x_2 \geq \cdots \geq x_n \geq 0\}.$$

i.e., all nonnegative vectors with components sorted in nonincreasing order.

(a) Show that $K_{\mathrm{m+}}$ is a proper cone.

(b) Find the dual cone $K_{\mathrm{m+}}^*$. *Hint.* Use the identity

$$\begin{aligned}\sum_{i=1}^n x_i y_i \quad = \quad & (x_1 - x_2)y_1 + (x_2 - x_3)(y_1 + y_2) + (x_3 - x_4)(y_1 + y_2 + y_3) + \cdots \\ & + (x_{n-1} - x_n)(y_1 + \cdots + y_{n-1}) + x_n(y_1 + \cdots + y_n).\end{aligned}$$

2.34 *The lexicographic cone and ordering.* The *lexicographic cone* is defined as

$$K_{\mathrm{lex}} = \{0\} \cup \{x \in \mathbf{R}^n \mid x_1 = \cdots = x_k = 0,\ x_{k+1} > 0,\ \text{for some } k,\ 0 \leq k < n\},$$

i.e., all vectors whose first nonzero coefficient (if any) is positive.

(a) Verify that K_{lex} is a cone, but *not* a proper cone.

(b) We define the *lexicographic ordering* on $\mathbf{R}^n$ as follows: $x \leq_{\mathrm{lex}} y$ if and only if $y - x \in K_{\mathrm{lex}}$. (Since K_{lex} is not a proper cone, the lexicographic ordering is not a generalized inequality.) Show that the lexicographic ordering is a *linear ordering*: for any $x,\ y \in \mathbf{R}^n$, either $x \leq_{\mathrm{lex}} y$ or $y \leq_{\mathrm{lex}} x$. Therefore any set of vectors can be sorted with respect to the lexicographic cone, which yields the familiar sorting used in dictionaries.

(c) Find K_{lex}^*.

2.35 *Copositive matrices.* A matrix $X \in \mathbf{S}^n$ is called *copositive* if $z^T X z \geq 0$ for all $z \succeq 0$. Verify that the set of copositive matrices is a proper cone. Find its dual cone.

2.36 *Euclidean distance matrices.* Let $x_1, \ldots, x_n \in \mathbf{R}^k$. The matrix $D \in \mathbf{S}^n$ defined by $D_{ij} = \|x_i - x_j\|_2^2$ is called a *Euclidean distance matrix*. It satisfies some obvious properties such as $D_{ij} = D_{ji}$, $D_{ii} = 0$, $D_{ij} \geq 0$, and (from the triangle inequality) $D_{ik}^{1/2} \leq D_{ij}^{1/2} + D_{jk}^{1/2}$. We now pose the question: When is a matrix $D \in \mathbf{S}^n$ a Euclidean distance matrix (for some points in $\mathbf{R}^k$, for some k)? A famous result answers this question: $D \in \mathbf{S}^n$ is a Euclidean distance matrix if and only if $D_{ii} = 0$ and $x^T D x \leq 0$ for all x with $\mathbf{1}^T x = 0$. (See §8.3.3.)

Show that the set of Euclidean distance matrices is a convex cone.

2.37 *Nonnegative polynomials and Hankel LMIs.* Let K_{pol} be the set of (coefficients of) nonnegative polynomials of degree $2k$ on $\mathbf{R}$:

$$K_{\mathrm{pol}} = \{x \in \mathbf{R}^{2k+1} \mid x_1 + x_2 t + x_3 t^2 + \cdots + x_{2k+1} t^{2k} \geq 0 \text{ for all } t \in \mathbf{R}\}.$$

(a) Show that K_{pol} is a proper cone.

(b) A basic result states that a polynomial of degree $2k$ is nonnegative on $\mathbf{R}$ if and only if it can be expressed as the sum of squares of two polynomials of degree k or less. In other words, $x \in K_{\mathrm{pol}}$ if and only if the polynomial

$$p(t) = x_1 + x_2 t + x_3 t^2 + \cdots + x_{2k+1} t^{2k}$$

can be expressed as

$$p(t) = r(t)^2 + s(t)^2,$$

where r and s are polynomials of degree k.

Use this result to show that

$$K_{\mathrm{pol}} = \left\{ x \in \mathbf{R}^{2k+1} \;\middle|\; x_i = \sum_{m+n=i+1} Y_{mn} \text{ for some } Y \in \mathbf{S}_+^{k+1} \right\}.$$

In other words, $p(t) = x_1 + x_2 t + x_3 t^2 + \cdots + x_{2k+1} t^{2k}$ is nonnegative if and only if there exists a matrix $Y \in \mathbf{S}_+^{k+1}$ such that

$$\begin{array}{rcl} x_1 & = & Y_{11} \\ x_2 & = & Y_{12} + Y_{21} \\ x_3 & = & Y_{13} + Y_{22} + Y_{31} \\ & \vdots & \\ x_{2k+1} & = & Y_{k+1,k+1}. \end{array}$$

(c) Show that $K_{\mathrm{pol}}^* = K_{\mathrm{han}}$ where

$$K_{\mathrm{han}} = \{z \in \mathbf{R}^{2k+1} \mid H(z) \succeq 0\}$$

and

$$H(z) = \begin{bmatrix} z_1 & z_2 & z_3 & \cdots & z_k & z_{k+1} \\ z_2 & z_3 & z_4 & \cdots & z_{k+1} & z_{k+2} \\ z_3 & z_4 & z_5 & \cdots & z_{k+2} & z_{k+4} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ z_k & z_{k+1} & z_{k+2} & \cdots & z_{2k-1} & z_{2k} \\ z_{k+1} & z_{k+2} & z_{k+3} & \cdots & z_{2k} & z_{2k+1} \end{bmatrix}.$$

(This is the *Hankel matrix* with coefficients $z_1, \ldots, z_{2k+1}$.)

(d) Let $K_{\rm mom}$ be the conic hull of the set of all vectors of the form $(1, t, t^2, \ldots, t^{2k})$, where $t \in \mathbf{R}$. Show that $y \in K_{\rm mom}$ if and only if $y_1 \geq 0$ and

$$y = y_1(1, \mathbf{E}\, u, \mathbf{E}\, u^2, \ldots, \mathbf{E}\, u^{2k})$$

for some random variable u. In other words, the elements of $K_{\rm mom}$ are nonnegative multiples of the moment vectors of all possible distributions on $\mathbf{R}$. Show that $K_{\rm pol} = K_{\rm mom}^*$.

(e) Combining the results of (c) and (d), conclude that $K_{\rm han} = \mathbf{cl}\, K_{\rm mom}$.

As an example illustrating the relation between $K_{\rm mom}$ and $K_{\rm han}$, take $k = 2$ and $z = (1, 0, 0, 0, 1)$. Show that $z \in K_{\rm han}$, $z \notin K_{\rm mom}$. Find an explicit sequence of points in $K_{\rm mom}$ which converge to z.

2.38 [Roc70, pages 15, 61] *Convex cones constructed from sets.*

(a) The *barrier cone* of a set C is defined as the set of all vectors y such that $y^T x$ is bounded above over $x \in C$. In other words, a nonzero vector y is in the barrier cone if and only if it is the normal vector of a halfspace $\{x \mid y^T x \leq \alpha\}$ that contains C. Verify that the barrier cone is a convex cone (with no assumptions on C).

(b) The *recession cone* (also called *asymptotic cone*) of a set C is defined as the set of all vectors y such that for each $x \in C$, $x - ty \in C$ for all $t \geq 0$. Show that the recession cone of a convex set is a convex cone. Show that if C is nonempty, closed, and convex, then the recession cone of C is the dual of the barrier cone.

(c) The *normal cone* of a set C at a boundary point x_0 is the set of all vectors y such that $y^T(x - x_0) \leq 0$ for all $x \in C$ (*i.e.*, the set of vectors that define a supporting hyperplane to C at x_0). Show that the normal cone is a convex cone (with no assumptions on C). Give a simple description of the normal cone of a polyhedron $\{x \mid Ax \preceq b\}$ at a point in its boundary.

2.39 *Separation of cones.* Let K and $\tilde{K}$ be two convex cones whose interiors are nonempty and disjoint. Show that there is a nonzero y such that $y \in K^*$, $-y \in \tilde{K}^*$.