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

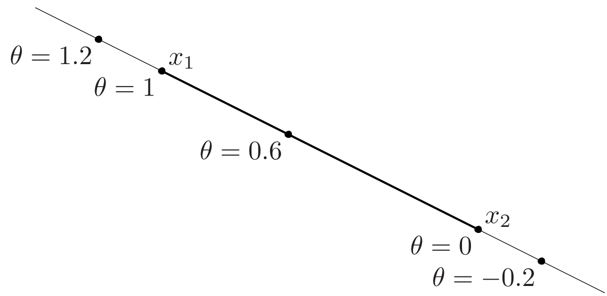


Figure 2.1 The line passing through x_1 and x_2 is described parametrically by $\theta x_1 + (1-\theta)x_2$, where θ varies over $\mathbf{R}$. The line segment between x_1 and x_2, which corresponds to θ between 0 and 1, is shown darker.

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

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

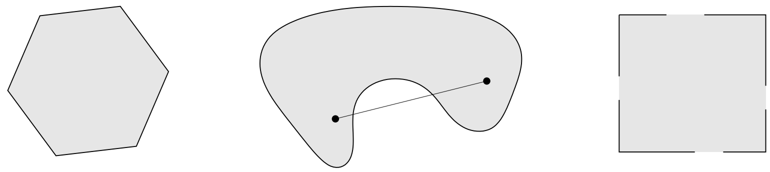


Figure 2.2 Some simple convex and nonconvex sets. *Left.* The hexagon, which includes its boundary (shown darker), is convex. *Middle.* The kidney shaped set is not convex, since the line segment between the two points in the set shown as dots is not contained in the set. *Right.* The square contains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

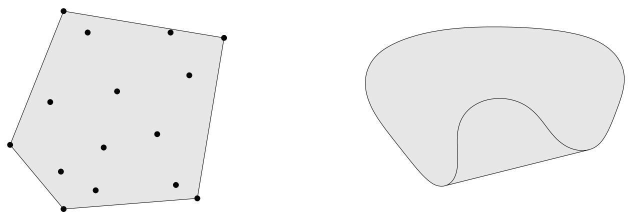


Figure 2.3 The convex hulls of two sets in $\mathbf{R}^2$. *Left.* The convex hull of a set of fifteen points (shown as dots) is the pentagon (shown shaded). *Right.* The convex hull of the kidney shaped set in figure 2.2 is the shaded set.

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).

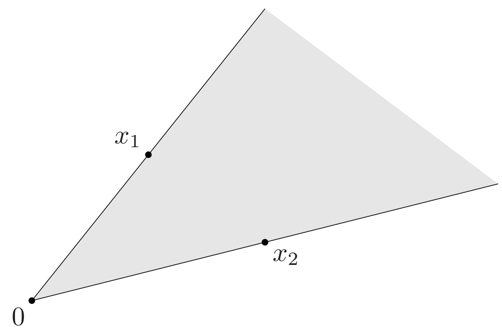


Figure 2.4 The pie slice shows all points of the form $\theta_1 x_1 + \theta_2 x_2$, where $\theta_1,\ \theta_2 \geq 0$. The apex of the slice (which corresponds to $\theta_1 = \theta_2 = 0$) is at 0; its edges (which correspond to $\theta_1 = 0$ or $\theta_2 = 0$) pass through the points x_1 and x_2.

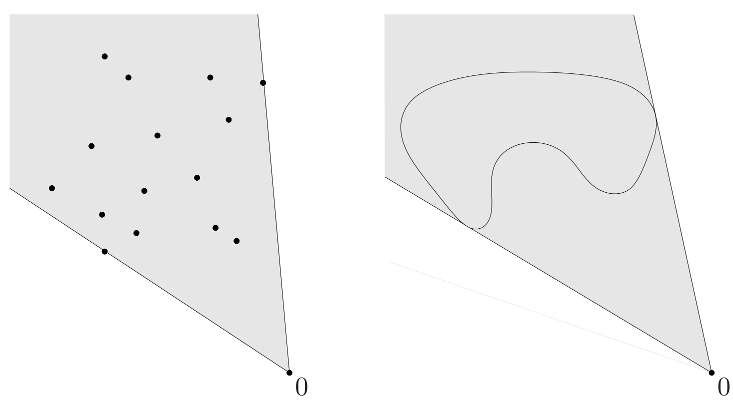


Figure 2.5 The conic hulls (shown shaded) of the two sets of figure 2.3.

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.

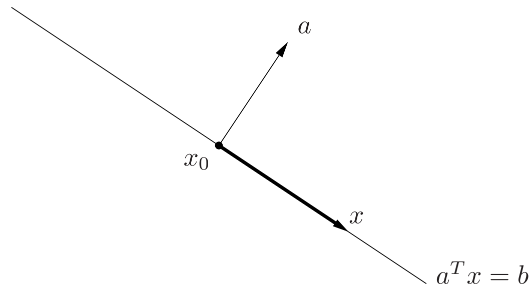


Figure 2.6 Hyperplane in $\mathbf{R}^2$, with normal vector a and a point x_0 in the hyperplane. For any point x in the hyperplane, $x - x_0$ (shown as the darker arrow) is orthogonal to a.

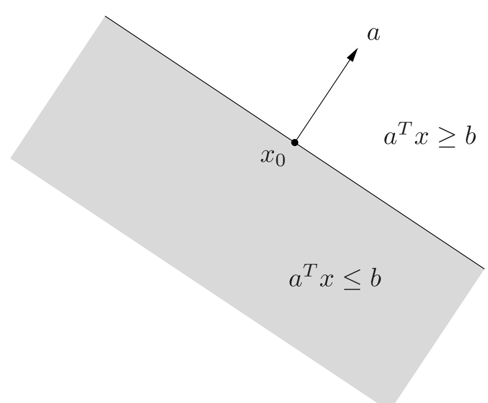


Figure 2.7 A hyperplane defined by $a^T x = b$ in $\mathbf{R}^2$ determines two halfspaces. The halfspace determined by $a^T x \geq b$ (not shaded) is the halfspace extending in the direction a. The halfspace determined by $a^T x \leq b$ (which is shown shaded) extends in the direction $-a$. The vector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

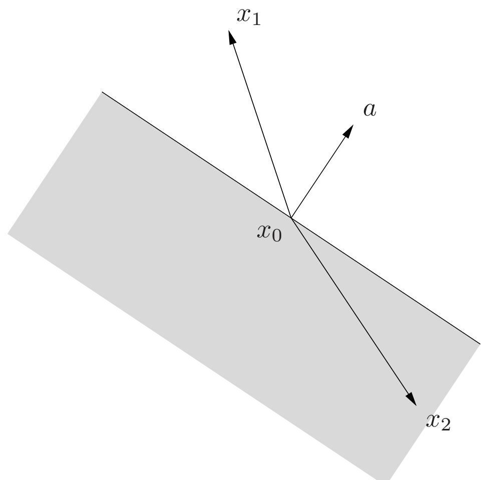


Figure 2.8 The shaded set is the halfspace determined by $a^T(x - x_0) \leq 0$. The vector $x_1 - x_0$ makes an acute angle with a, so x_1 is not in the halfspace. The vector $x_2 - x_0$ makes an obtuse angle with a, and so is in the halfspace.

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

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

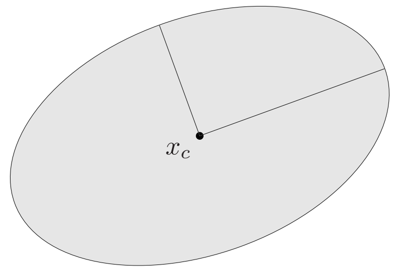


Figure 2.9 An ellipsoid in $\mathbf{R}^2$, shown shaded. The center x_c is shown as a dot, and the two semi-axes are shown as line segments.

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

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

(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

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

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

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

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 $\{x \mid \|x - x_c\| \le r\}$, is convex. The *norm cone* associated with the norm $\|\cdot\|$ is the set

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

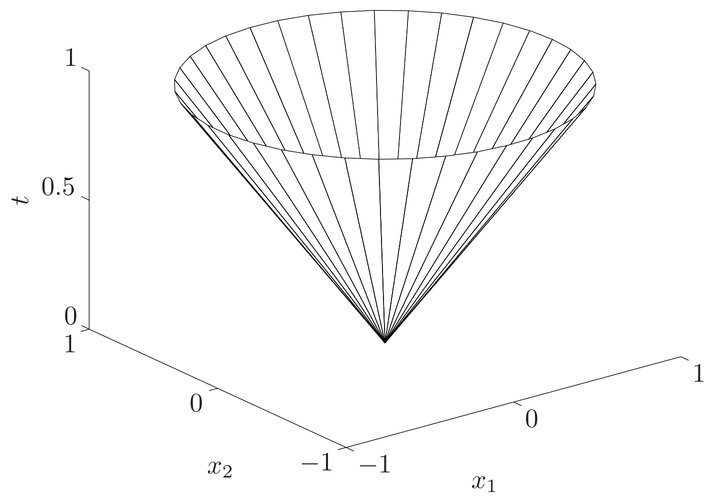


Figure 2.10 Boundary of second-order cone in $\mathbf{R}^3$, $\{(x_1, x_2, t) \mid (x_1^2+x_2^2)^{1/2} \leq t\}$.

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.*,

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

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

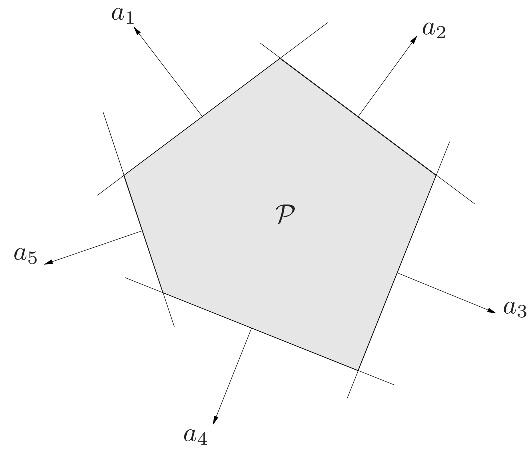


Figure 2.11 The polyhedron $\mathcal{P}$ (shown shaded) is the intersection of five halfspaces, with outward normal vectors $a_1, \ldots, a_5$.

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 $u_i \le v_i$ for $i = 1, \ldots, m$.

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

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

(Here $\mathbf{R}_+$ denotes the set of nonnegative numbers: $\mathbf{R}_+ = \{x \in \mathbf{R} \mid x \ge 0\}$.) 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

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

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.)

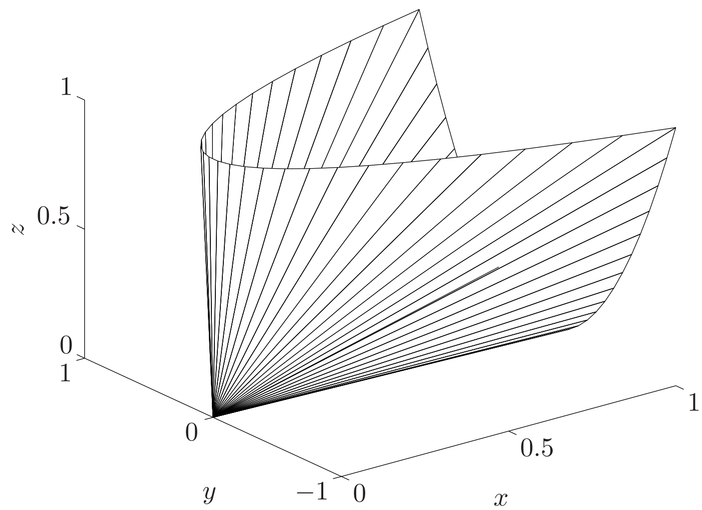


Figure 2.12 Boundary of positive semidefinite cone in $\mathbf{S}^2$.

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,

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

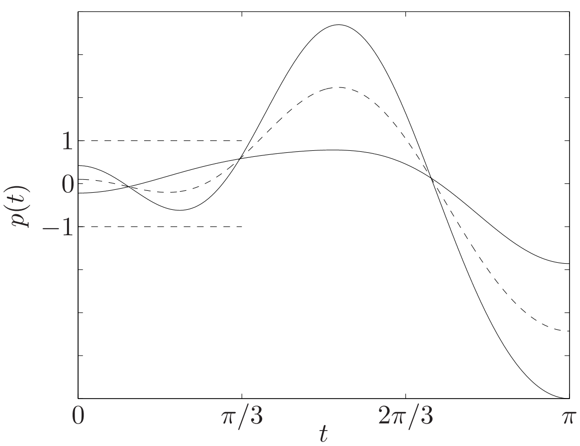


Figure 2.13 Three trigonometric polynomials associated with points in the set S defined in (2.10), for $m = 2$. The trigonometric polynomial plotted with dashed line type is the average of the other two.

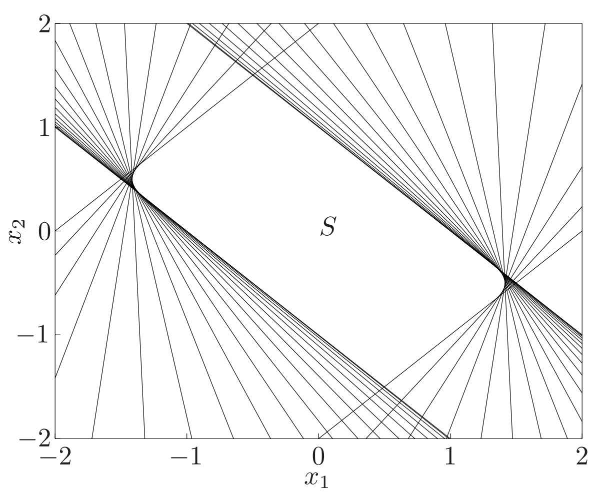


Figure 2.14 The set S defined in (2.10), for $m = 2$, is shown as the white area in the middle of the plot. The set is the intersection of an infinite number of slabs (20 of which are shown), hence convex.

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

$$\{x \mid x^T P x \leq (c^T x)^2,\ c^T x \geq 0\}$$

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,

$$\{(z,t) \mid z^T z \leq t^2,\ t \geq 0\},$$

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

Example 2.12 *Ellipsoid.* The ellipsoid

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

where $P \in \mathbf{S}^n_{++}$, is the image of the unit Euclidean ball $\{u \mid \|u\|_2 \leq 1\}$ 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

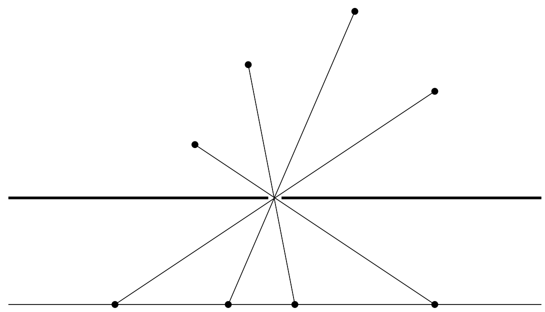


Figure 2.15 Pin-hole camera interpretation of perspective function. The dark horizontal line represents the plane $x_3 = 0$ in $\mathbf{R}^3$, which is opaque, except for a pin-hole at the origin. Objects or light sources above the plane appear on the image plane $x_3 = -1$, which is shown as the lighter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

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

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.*,

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

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

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

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 $f(x)$.

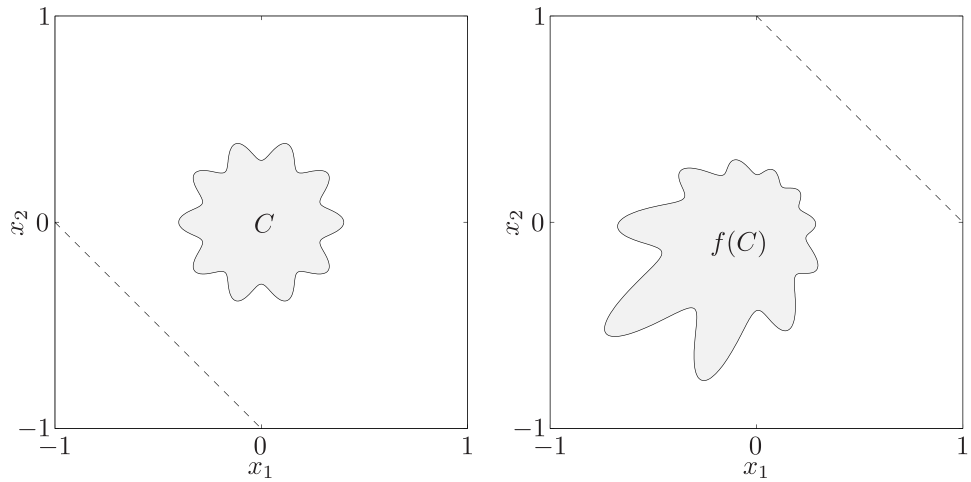


Figure 2.16 *Left.* A set $C \subseteq \mathbf{R}^2$. The dashed line shows the boundary of the domain of the linear-fractional function $f(x) = x/(x_1 + x_2 + 1)$ with $\mathbf{dom}\, f = \{(x_1, x_2) \mid x_1 + x_2 + 1 > 0\}$. *Right.* Image of C under f. The dashed line shows the boundary of the domain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

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

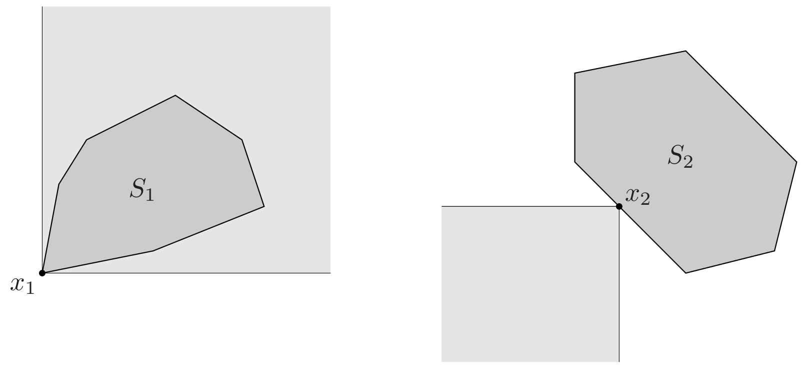


Figure 2.17 *Left.* The set S_1 has a minimum element x_1 with respect to componentwise inequality in $\mathbf{R}^2$. The set $x_1 + K$ is shaded lightly; x_1 is the minimum element of S_1 since $S_1 \subseteq x_1 + K$. *Right.* The point x_2 is a minimal point of S_2. The set $x_2 - K$ is shown lightly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

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

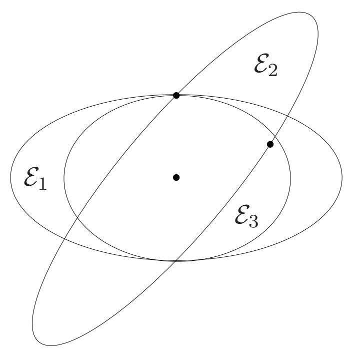


Figure 2.18 Three ellipsoids in $\mathbf{R}^2$, centered at the origin (shown as the lower dot), that contain the points shown as the upper dots. The ellipsoid $\mathcal{E}_1$ is not minimal, since there exist ellipsoids that contain the points, and are smaller (*e.g.*, $\mathcal{E}_3$). $\mathcal{E}_3$ 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 $\mathcal{E}_2$.

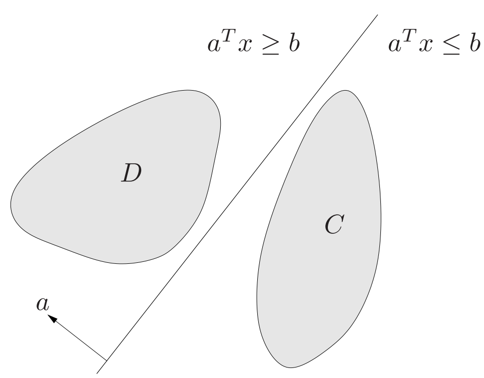


Figure 2.19 The hyperplane $\{x \mid a^T x = b\}$ separates the disjoint convex sets C and D. The affine function $a^T x - b$ is nonpositive on C and nonnegative on D.

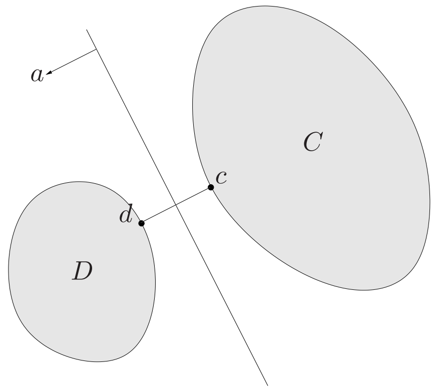


Figure 2.20 Construction of a separating hyperplane between two convex sets. The points $c \in C$ and $d \in D$ are the pair of points in the two sets that are closest to each other. The separating hyperplane is orthogonal to, and bisects, the line segment between c and d.

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

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

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 $t \leq 1$, 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

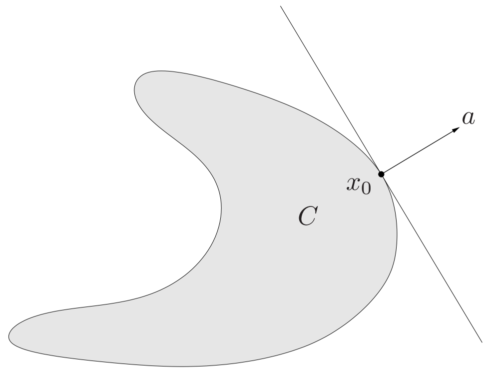


Figure 2.21 The hyperplane $\{x \mid a^T x = a^T x_0\}$ supports C at x_0.

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 $\{x \mid a^T x \leq a^T x_0\}$ 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

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

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 $V^\perp = \{y \mid v^T y = 0 \text{ for all } v \in V\}$.

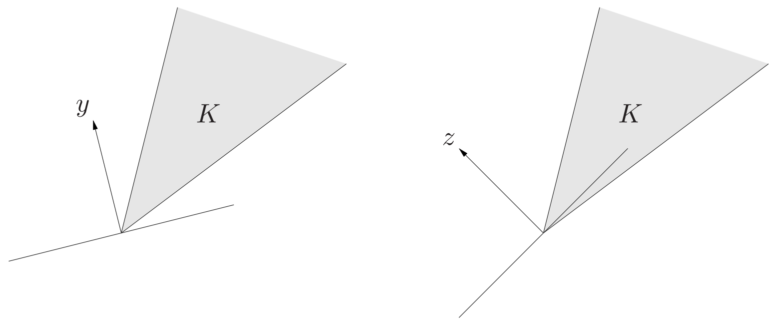


Figure 2.22 *Left.* The halfspace with inward normal y contains the cone K, so $y \in K^*$. *Right.* The halfspace with inward normal z does not contain K, so $z \notin K^*$.

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

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

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

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

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

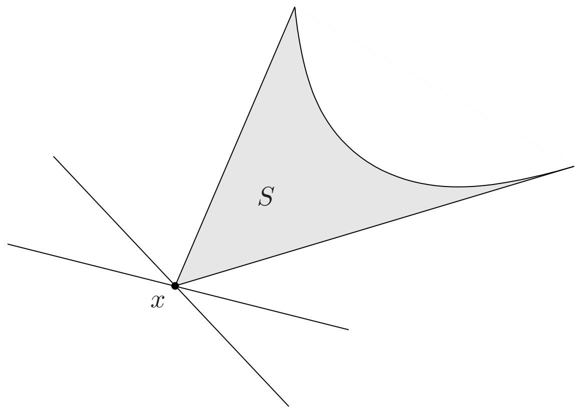


Figure 2.23 Dual characterization of minimum element. The point x is the minimum element of the set S with respect to $\mathbf{R}_+^2$. This is equivalent to: for every $\lambda \succ 0$, the hyperplane $\{z \mid \lambda^T(z-x) = 0\}$ strictly supports S at x, *i.e.*, contains S on one side, and touches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

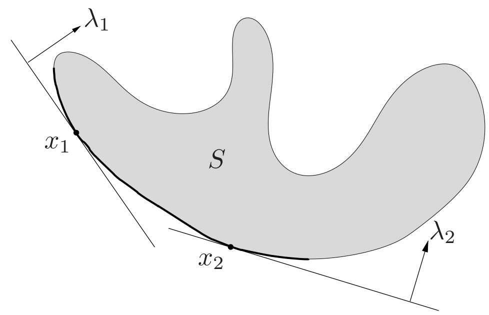


Figure 2.24 A set $S \subseteq \mathbf{R}^2$. Its set of minimal points, with respect to $\mathbf{R}^2_+$, is shown as the darker section of its (lower, left) boundary. The minimizer of $\lambda_1^T z$ over S is x_1, and is minimal since $\lambda_1 \succ 0$. The minimizer of $\lambda_2^T z$ over S is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

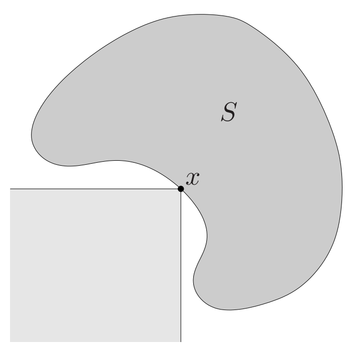


Figure 2.25 The point x is a minimal element of $S \subseteq \mathbf{R}^2$ with respect to $\mathbf{R}^2_+$. However there exists no λ for which x minimizes $\lambda^T z$ over $z \in S$.

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.

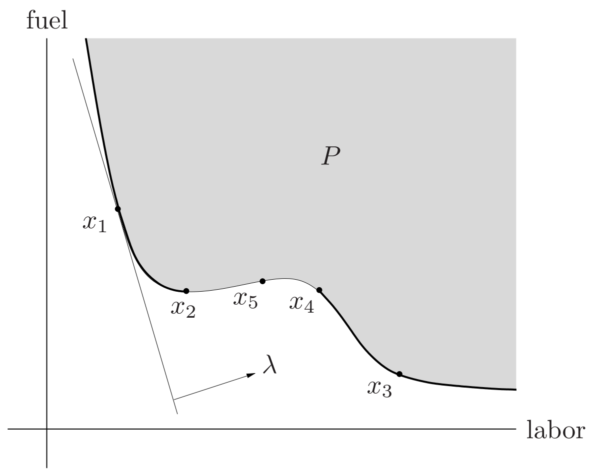


Figure 2.27 The production set P, for a product that requires labor and fuel to produce, is shown shaded. The two dark curves show the efficient production frontier. The points x_1, x_2 and x_3 are efficient. The points x_4 and x_5 are not (since in particular, x_2 corresponds to a production method 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 $\{X X^T \mid X \in \mathbf{R}^{n \times k},\ \mathbf{rank}\, X = k\}$. 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

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

Show that $P_{\text{inner}} \subseteq C \subseteq P_{\text{outer}}$. 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

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

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

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

(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 $y \in K^*$, $-y \in \tilde{K}^*$.

Chapter 3

Convex functions

3.1 Basic properties and examples

3.1.1 Definition

A function $f : \mathbf{R}^n \to \mathbf{R}$ is *convex* if $\mathbf{dom}\, f$ is a convex set and if for all x, $y \in \mathbf{dom}\, f$, and θ with $0 \le \theta \le 1$, we have

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y). \tag{3.1}$$

Geometrically, this inequality means that the line segment between $(x, f(x))$ and $(y, f(y))$, which is the *chord* from x to y, lies above the graph of f (figure 3.1). A function f is *strictly convex* if strict inequality holds in (3.1) whenever $x \ne y$ and $0 < \theta < 1$. We say f is *concave* if $-f$ is convex, and *strictly concave* if $-f$ is strictly convex.

For an affine function we always have equality in (3.1), so all affine (and therefore also linear) functions are both convex and concave. Conversely, any function that is convex and concave is affine.

A function is convex if and only if it is convex when restricted to any line that intersects its domain. In other words f is convex if and only if for all $x \in \mathbf{dom}\, f$ and

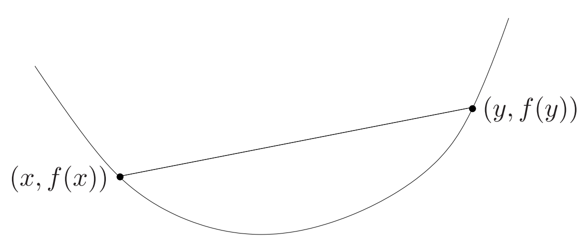


Figure 3.1 Graph of a convex function. The chord (*i.e.*, line segment) between any two points on the graph lies above the graph.

all v, the function $g(t) = f(x + tv)$ is convex (on its domain, $\{t \mid x + tv \in \mathbf{dom}\, f\}$). This property is very useful, since it allows us to check whether a function is convex by restricting it to a line.

The *analysis* of convex functions is a well developed field, which we will not pursue in any depth. One simple result, for example, is that a convex function is continuous on the relative interior of its domain; it can have discontinuities only on its relative boundary.

3.1.2 Extended-value extensions

It is often convenient to extend a convex function to all of $\mathbf{R}^n$ by defining its value to be ∞ outside its domain. If f is convex we define its *extended-value extension* $\tilde{f} : \mathbf{R}^n \to \mathbf{R} \cup \{\infty\}$ by

$$\tilde{f}(x) = \begin{cases} f(x) & x \in \mathbf{dom}\, f \\ \infty & x \notin \mathbf{dom}\, f. \end{cases}$$

The extension $\tilde{f}$ is defined on all $\mathbf{R}^n$, and takes values in $\mathbf{R} \cup \{\infty\}$. We can recover the domain of the original function f from the extension $\tilde{f}$ as $\mathbf{dom}\, f = \{x \mid \tilde{f}(x) < \infty\}$.

The extension can simplify notation, since we do not need to explicitly describe the domain, or add the qualifier 'for all $x \in \mathbf{dom}\, f$' every time we refer to $f(x)$. Consider, for example, the basic defining inequality (3.1). In terms of the extension $\tilde{f}$, we can express it as: for $0 < \theta < 1$,

$$\tilde{f}(\theta x + (1 - \theta)y) \le \theta \tilde{f}(x) + (1 - \theta)\tilde{f}(y)$$

for *any* x and y. (For $\theta = 0$ or $\theta = 1$ the inequality always holds.) Of course here we must interpret the inequality using extended arithmetic and ordering. For x and y both in $\mathbf{dom}\, f$, this inequality coincides with (3.1); if either is outside $\mathbf{dom}\, f$, then the righthand side is ∞, and the inequality therefore holds. As another example of this notational device, suppose f_1 and f_2 are two convex functions on $\mathbf{R}^n$. The pointwise sum $f = f_1 + f_2$ is the function with domain $\mathbf{dom}\, f = \mathbf{dom}\, f_1 \cap \mathbf{dom}\, f_2$, with $f(x) = f_1(x) + f_2(x)$ for any $x \in \mathbf{dom}\, f$. Using extended-value extensions we can simply say that for any x, $\tilde{f}(x) = \tilde{f}_1(x) + \tilde{f}_2(x)$. In this equation the domain of f has been automatically defined as $\mathbf{dom}\, f = \mathbf{dom}\, f_1 \cap \mathbf{dom}\, f_2$, since $\tilde{f}(x) = \infty$ whenever $x \notin \mathbf{dom}\, f_1$ or $x \notin \mathbf{dom}\, f_2$. In this example we are relying on extended arithmetic to automatically define the domain.

In this book we will use the same symbol to denote a convex function and its extension, whenever there is no harm from the ambiguity. This is the same as assuming that all convex functions are implicitly extended, *i.e.*, are defined as ∞ outside their domains.

Example 3.1 *Indicator function of a convex set.* Let $C \subseteq \mathbf{R}^n$ be a convex set, and consider the (convex) function I_C with domain C and $I_C(x) = 0$ for all $x \in C$. In other words, the function is identically zero on the set C. Its extended-value extension

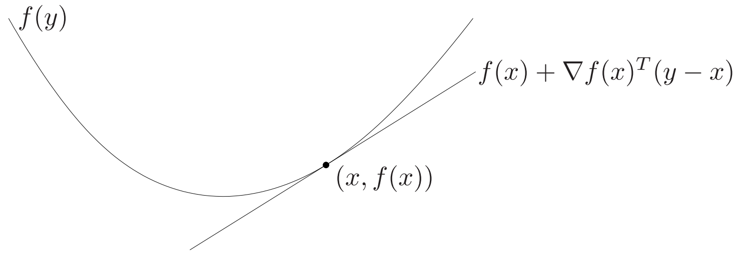


Figure 3.2 If f is convex and differentiable, then $f(x)+\nabla f(x)^T(y-x) \leq f(y)$ for all $x,\ y \in \mathbf{dom}\, f$.

is given by

$$\tilde{I}_C(x) = \begin{cases} 0 & x \in C \\ \infty & x \notin C. \end{cases}$$

The convex function $\tilde{I}_C$ is called the *indicator function* of the set C.

We can play several notational tricks with the indicator function $\tilde{I}_C$. For example the problem of minimizing a function f (defined on all of $\mathbf{R}^n$, say) on the set C is the same as minimizing the function $f + \tilde{I}_C$ over all of $\mathbf{R}^n$. Indeed, the function $f + \tilde{I}_C$ is (by our convention) f restricted to the set C.

In a similar way we can extend a concave function by defining it to be $-\infty$ outside its domain.

3.1.3 First-order conditions

Suppose f is differentiable (*i.e.*, its gradient ∇f exists at each point in $\mathbf{dom}\, f$, which is open). Then f is convex if and only if $\mathbf{dom}\, f$ is convex and

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \tag{3.2}$$

holds for all $x,\ y \in \mathbf{dom}\, f$. This inequality is illustrated in figure 3.2.

The affine function of y given by $f(x)+\nabla f(x)^T(y-x)$ is, of course, the first-order Taylor approximation of f near x. The inequality (3.2) states that for a convex function, the first-order Taylor approximation is in fact a *global underestimator* of the function. Conversely, if the first-order Taylor approximation of a function is always a global underestimator of the function, then the function is convex.

The inequality (3.2) shows that from *local information* about a convex function (*i.e.*, its value and derivative at a point) we can derive *global information* (*i.e.*, a global underestimator of it). This is perhaps the most important property of convex functions, and explains some of the remarkable properties of convex functions and convex optimization problems. As one simple example, the inequality (3.2) shows that if $\nabla f(x) = 0$, then for all $y \in \mathbf{dom}\, f$, $f(y) \geq f(x)$, *i.e.*, x is a global minimizer of the function f.

Strict convexity can also be characterized by a first-order condition: f is strictly convex if and only if $\mathbf{dom}\, f$ is convex and for $x,\ y \in \mathbf{dom}\, f$, $x \neq y$, we have

$$f(y) > f(x) + \nabla f(x)^T (y - x). \tag{3.3}$$

For concave functions we have the corresponding characterization: f is concave if and only if $\mathbf{dom}\, f$ is convex and

$$f(y) \leq f(x) + \nabla f(x)^T (y - x)$$

for all $x,\ y \in \mathbf{dom}\, f$.

Proof of first-order convexity condition

To prove (3.2), we first consider the case $n = 1$: We show that a differentiable function $f : \mathbf{R} \to \mathbf{R}$ is convex if and only if

$$f(y) \geq f(x) + f'(x)(y - x) \tag{3.4}$$

for all x and y in $\mathbf{dom}\, f$.

Assume first that f is convex and $x,\ y \in \mathbf{dom}\, f$. Since $\mathbf{dom}\, f$ is convex (*i.e.*, an interval), we conclude that for all $0 < t \leq 1$, $x + t(y - x) \in \mathbf{dom}\, f$, and by convexity of f,

$$f(x + t(y - x)) \leq (1 - t)f(x) + tf(y).$$

If we divide both sides by t, we obtain

$$f(y) \geq f(x) + \frac{f(x + t(y - x)) - f(x)}{t},$$

and taking the limit as $t \to 0$ yields (3.4).

To show sufficiency, assume the function satisfies (3.4) for all x and y in $\mathbf{dom}\, f$ (which is an interval). Choose any $x \neq y$, and $0 \leq \theta \leq 1$, and let $z = \theta x + (1 - \theta)y$. Applying (3.4) twice yields

$$f(x) \geq f(z) + f'(z)(x - z), \qquad f(y) \geq f(z) + f'(z)(y - z).$$

Multiplying the first inequality by θ, the second by $1 - \theta$, and adding them yields

$$\theta f(x) + (1 - \theta)f(y) \geq f(z),$$

which proves that f is convex.

Now we can prove the general case, with $f : \mathbf{R}^n \to \mathbf{R}$. Let $x,\ y \in \mathbf{R}^n$ and consider f restricted to the line passing through them, *i.e.*, the function defined by $g(t) = f(ty + (1 - t)x)$, so $g'(t) = \nabla f(ty + (1 - t)x)^T (y - x)$.

First assume f is convex, which implies g is convex, so by the argument above we have $g(1) \geq g(0) + g'(0)$, which means

$$f(y) \geq f(x) + \nabla f(x)^T (y - x).$$

Now assume that this inequality holds for any x and y, so if $ty + (1 - t)x \in \mathbf{dom}\, f$ and $\tilde{t}y + (1 - \tilde{t})x \in \mathbf{dom}\, f$, we have

$$f(ty + (1 - t)x) \geq f(\tilde{t}y + (1 - \tilde{t})x) + \nabla f(\tilde{t}y + (1 - \tilde{t})x)^T (y - x)(t - \tilde{t}),$$

i.e., $g(t) \geq g(\tilde{t}) + g'(\tilde{t})(t - \tilde{t})$. We have seen that this implies that g is convex.

3.1.4 Second-order conditions

We now assume that f is twice differentiable, that is, its *Hessian* or second derivative $\nabla^2 f$ exists at each point in $\mathbf{dom}\, f$, which is open. Then f is convex if and only if $\mathbf{dom}\, f$ is convex and its Hessian is positive semidefinite: for all $x \in \mathbf{dom}\, f$,

$$\nabla^2 f(x) \succeq 0.$$

For a function on $\mathbf{R}$, this reduces to the simple condition $f''(x) \geq 0$ (and $\mathbf{dom}\, f$ convex, *i.e.*, an interval), which means that the derivative is nondecreasing. The condition $\nabla^2 f(x) \succeq 0$ can be interpreted geometrically as the requirement that the graph of the function have positive (upward) curvature at x. We leave the proof of the second-order condition as an exercise (exercise 3.8).

Similarly, f is concave if and only if $\mathbf{dom}\, f$ is convex and $\nabla^2 f(x) \preceq 0$ for all $x \in \mathbf{dom}\, f$. Strict convexity can be partially characterized by second-order conditions. If $\nabla^2 f(x) \succ 0$ for all $x \in \mathbf{dom}\, f$, then f is strictly convex. The converse, however, is not true: for example, the function $f : \mathbf{R} \to \mathbf{R}$ given by $f(x) = x^4$ is strictly convex but has zero second derivative at $x = 0$.

Example 3.2 *Quadratic functions.* Consider the quadratic function $f : \mathbf{R}^n \to \mathbf{R}$, with $\mathbf{dom}\, f = \mathbf{R}^n$, given by

$$f(x) = (1/2)x^T P x + q^T x + r,$$

with $P \in \mathbf{S}^n$, $q \in \mathbf{R}^n$, and $r \in \mathbf{R}$. Since $\nabla^2 f(x) = P$ for all x, f is convex if and only if $P \succeq 0$ (and concave if and only if $P \preceq 0$).

For quadratic functions, strict convexity is easily characterized: f is strictly convex if and only if $P \succ 0$ (and strictly concave if and only if $P \prec 0$).

Remark 3.1 The separate requirement that $\mathbf{dom}\, f$ be convex cannot be dropped from the first- or second-order characterizations of convexity and concavity. For example, the function $f(x) = 1/x^2$, with $\mathbf{dom}\, f = \{x \in \mathbf{R} \mid x \neq 0\}$, satisfies $f''(x) > 0$ for all $x \in \mathbf{dom}\, f$, but is not a convex function.

3.1.5 Examples

We have already mentioned that all linear and affine functions are convex (and concave), and have described the convex and concave quadratic functions. In this section we give a few more examples of convex and concave functions. We start with some functions on $\mathbf{R}$, with variable x.

- *Exponential.* e^{ax} is convex on $\mathbf{R}$, for any $a \in \mathbf{R}$.
- *Powers.* x^a is convex on $\mathbf{R}_{++}$ when $a \geq 1$ or $a \leq 0$, and concave for $0 \leq a \leq 1$.
- *Powers of absolute value.* $|x|^p$, for $p \geq 1$, is convex on $\mathbf{R}$.
- *Logarithm.* $\log x$ is concave on $\mathbf{R}_{++}$.

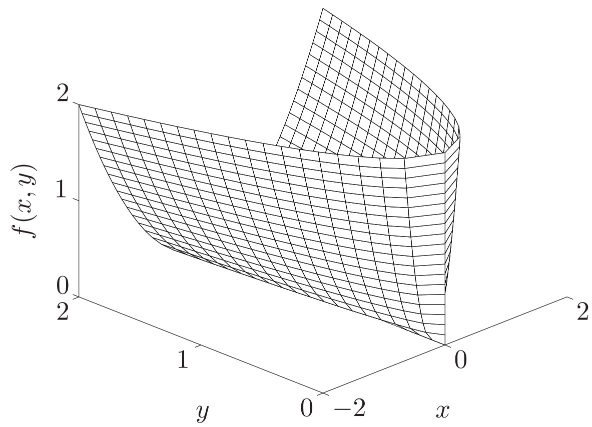


Figure 3.3 Graph of $f(x, y) = x^2/y$.

- *Negative entropy.* $x \log x$ (either on $\mathbf{R}_{++}$, or on $\mathbf{R}_+$, defined as 0 for $x = 0$) is convex.

Convexity or concavity of these examples can be shown by verifying the basic inequality (3.1), or by checking that the second derivative is nonnegative or nonpositive. For example, with $f(x) = x \log x$ we have

$$f'(x) = \log x + 1, \qquad f''(x) = 1/x,$$

so that $f''(x) > 0$ for $x > 0$. This shows that the negative entropy function is (strictly) convex.

We now give a few interesting examples of functions on $\mathbf{R}^n$.

- *Norms.* Every norm on $\mathbf{R}^n$ is convex.
- *Max function.* $f(x) = \max\{x_1, \ldots, x_n\}$ is convex on $\mathbf{R}^n$.
- *Quadratic-over-linear function.* The function $f(x, y) = x^2/y$, with

  $$\mathbf{dom}\, f = \mathbf{R} \times \mathbf{R}_{++} = \{(x, y) \in \mathbf{R}^2 \mid y > 0\},$$

  is convex (figure 3.3).
- *Log-sum-exp.* The function $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex on $\mathbf{R}^n$. This function can be interpreted as a differentiable (in fact, analytic) approximation of the max function, since

  $$\max\{x_1, \ldots, x_n\} \leq f(x) \leq \max\{x_1, \ldots, x_n\} + \log n$$

  for all x. (The second inequality is tight when all components of x are equal.) Figure 3.4 shows f for $n = 2$.

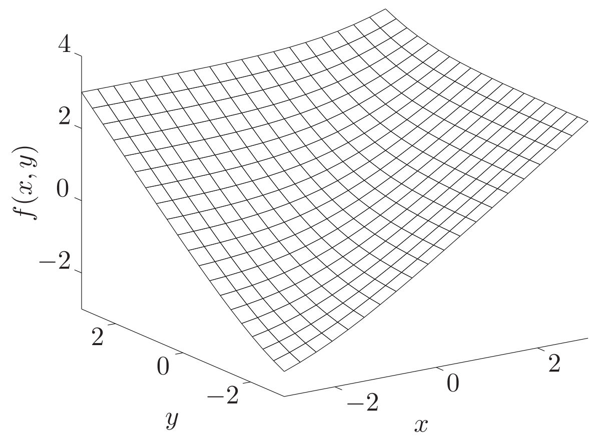


Figure 3.4 Graph of $f(x, y) = \log(e^x + e^y)$.

- *Geometric mean.* The geometric mean $f(x) = \left(\prod_{i=1}^n x_i\right)^{1/n}$ is concave on $\mathbf{dom}\, f = \mathbf{R}^n_{++}$.
- *Log-determinant.* The function $f(X) = \log\det X$ is concave on $\mathbf{dom}\, f = \mathbf{S}^n_{++}$.

Convexity (or concavity) of these examples can be verified in several ways, such as directly verifying the inequality (3.1), verifying that the Hessian is positive semidefinite, or restricting the function to an arbitrary line and verifying convexity of the resulting function of one variable.

Norms. If $f : \mathbf{R}^n \to \mathbf{R}$ is a norm, and $0 \leq \theta \leq 1$, then

$$f(\theta x + (1-\theta)y) \leq f(\theta x) + f((1-\theta)y) = \theta f(x) + (1-\theta)f(y).$$

The inequality follows from the triangle inequality, and the equality follows from homogeneity of a norm.

Max function. The function $f(x) = \max_i x_i$ satisfies, for $0 \leq \theta \leq 1$,

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \max_i (\theta x_i + (1-\theta)y_i) \\ &\leq \theta \max_i x_i + (1-\theta)\max_i y_i \\ &= \theta f(x) + (1-\theta)f(y). \end{aligned}$$

Quadratic-over-linear function. To show that the quadratic-over-linear function $f(x, y) = x^2/y$ is convex, we note that (for $y > 0$),

$$\nabla^2 f(x, y) = \frac{2}{y^3}\begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} = \frac{2}{y^3}\begin{bmatrix} y \\ -x \end{bmatrix}\begin{bmatrix} y \\ -x \end{bmatrix}^T \succeq 0.$$

Log-sum-exp. The Hessian of the log-sum-exp function is

$$\nabla^2 f(x) = \frac{1}{(\mathbf{1}^T z)^2}\left((\mathbf{1}^T z)\,\mathbf{diag}(z) - zz^T\right),$$

where $z = (e^{x_1}, \dots, e^{x_n})$. To verify that $\nabla^2 f(x) \succeq 0$ we must show that for all v, $v^T \nabla^2 f(x) v \geq 0$, *i.e.*,

$$v^T \nabla^2 f(x) v = \frac{1}{(\mathbf{1}^T z)^2}\left(\left(\sum_{i=1}^n z_i\right)\left(\sum_{i=1}^n v_i^2 z_i\right) - \left(\sum_{i=1}^n v_i z_i\right)^2\right) \geq 0.$$

But this follows from the Cauchy-Schwarz inequality $(a^T a)(b^T b) \geq (a^T b)^2$ applied to the vectors with components $a_i = v_i \sqrt{z_i}$, $b_i = \sqrt{z_i}$.

Geometric mean. In a similar way we can show that the geometric mean $f(x) = \left(\prod_{i=1}^n x_i\right)^{1/n}$ is concave on $\mathbf{dom}\, f = \mathbf{R}^n_{++}$. Its Hessian $\nabla^2 f(x)$ is given by

$$\frac{\partial^2 f(x)}{\partial x_k^2} = -(n-1)\frac{\left(\prod_{i=1}^n x_i\right)^{1/n}}{n^2 x_k^2}, \qquad \frac{\partial^2 f(x)}{\partial x_k \partial x_l} = \frac{\left(\prod_{i=1}^n x_i\right)^{1/n}}{n^2 x_k x_l} \quad \text{for } k \neq l,$$

and can be expressed as

$$\nabla^2 f(x) = -\frac{\prod_{i=1}^n x_i^{1/n}}{n^2}\left(n\,\mathbf{diag}(1/x_1^2, \dots, 1/x_n^2) - qq^T\right)$$

where $q_i = 1/x_i$. We must show that $\nabla^2 f(x) \preceq 0$, *i.e.*, that

$$v^T \nabla^2 f(x) v = -\frac{\prod_{i=1}^n x_i^{1/n}}{n^2}\left(n\sum_{i=1}^n v_i^2/x_i^2 - \left(\sum_{i=1}^n v_i/x_i\right)^2\right) \leq 0$$

for all v. Again this follows from the Cauchy-Schwarz inequality $(a^T a)(b^T b) \geq (a^T b)^2$, applied to the vectors $a = \mathbf{1}$ and $b_i = v_i/x_i$.

Log-determinant. For the function $f(X) = \log\det X$, we can verify concavity by considering an arbitrary line, given by $X = Z + tV$, where $Z,\ V \in \mathbf{S}^n$. We define $g(t) = f(Z + tV)$, and restrict g to the interval of values of t for which $Z + tV \succ 0$. Without loss of generality, we can assume that $t = 0$ is inside this interval, *i.e.*, $Z \succ 0$. We have

$$\begin{aligned} g(t) &= \log\det(Z + tV) \\ &= \log\det(Z^{1/2}(I + tZ^{-1/2}VZ^{-1/2})Z^{1/2}) \\ &= \sum_{i=1}^n \log(1 + t\lambda_i) + \log\det Z \end{aligned}$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $Z^{-1/2}VZ^{-1/2}$. Therefore we have

$$g'(t) = \sum_{i=1}^n \frac{\lambda_i}{1 + t\lambda_i}, \qquad g''(t) = -\sum_{i=1}^n \frac{\lambda_i^2}{(1 + t\lambda_i)^2}.$$

Since $g''(t) \leq 0$, we conclude that f is concave.

3.1.6 Sublevel sets

The α-*sublevel set* of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as

$$C_\alpha = \{x \in \mathbf{dom}\, f \mid f(x) \leq \alpha\}.$$

Sublevel sets of a convex function are convex, for any value of α. The proof is immediate from the definition of convexity: if $x,\ y \in C_\alpha$, then $f(x) \leq \alpha$ and $f(y) \leq \alpha$, and so $f(\theta x + (1-\theta)y) \leq \alpha$ for $0 \leq \theta \leq 1$, and hence $\theta x + (1-\theta)y \in C_\alpha$.

The converse is not true: a function can have all its sublevel sets convex, but not be a convex function. For example, $f(x) = -e^x$ is not convex on $\mathbf{R}$ (indeed, it is strictly concave) but all its sublevel sets are convex.

If f is concave, then its α-*superlevel set*, given by $\{x \in \mathbf{dom}\, f \mid f(x) \geq \alpha\}$, is a convex set. The sublevel set property is often a good way to establish convexity of a set, by expressing it as a sublevel set of a convex function, or as the superlevel set of a concave function.

Example 3.3 The geometric and arithmetic means of $x \in \mathbf{R}^n_+$ are, respectively,

$$G(x) = \left(\prod_{i=1}^n x_i\right)^{1/n}, \qquad A(x) = \frac{1}{n}\sum_{i=1}^n x_i,$$

(where we take $0^{1/n} = 0$ in our definition of G). The arithmetic-geometric mean inequality states that $G(x) \leq A(x)$.

Suppose $0 \leq \alpha \leq 1$, and consider the set

$$\{x \in \mathbf{R}^n_+ \mid G(x) \geq \alpha A(x)\},$$

i.e., the set of vectors with geometric mean at least as large as a factor α times the arithmetic mean. This set is convex, since it is the 0-superlevel set of the function $G(x) - \alpha A(x)$, which is concave. In fact, the set is positively homogeneous, so it is a convex cone.

3.1.7 Epigraph

The graph of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as

$$\{(x, f(x)) \mid x \in \mathbf{dom}\, f\},$$

which is a subset of $\mathbf{R}^{n+1}$. The *epigraph* of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as

$$\mathbf{epi}\, f = \{(x, t) \mid x \in \mathbf{dom}\, f,\ f(x) \leq t\},$$

which is a subset of $\mathbf{R}^{n+1}$. ('Epi' means 'above' so epigraph means 'above the graph'.) The definition is illustrated in figure 3.5.

The link between convex sets and convex functions is via the epigraph: A function is convex if and only if its epigraph is a convex set. A function is concave if and only if its *hypograph*, defined as

$$\mathbf{hypo}\, f = \{(x, t) \mid t \leq f(x)\},$$

is a convex set.

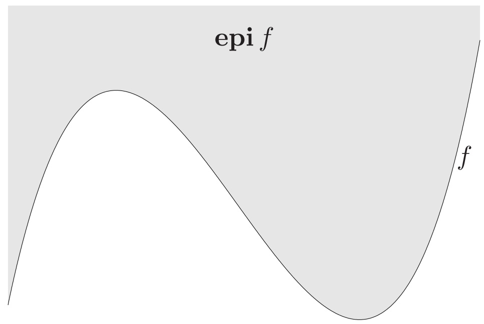


Figure 3.5 Epigraph of a function f, shown shaded. The lower boundary, shown darker, is the graph of f.

Example 3.4 *Matrix fractional function.* The function $f : \mathbf{R}^n \times \mathbf{S}^n \to \mathbf{R}$, defined as

$$f(x, Y) = x^T Y^{-1} x$$

is convex on $\mathbf{dom}\, f = \mathbf{R}^n \times \mathbf{S}^n_{++}$. (This generalizes the quadratic-over-linear function $f(x, y) = x^2/y$, with $\mathbf{dom}\, f = \mathbf{R} \times \mathbf{R}_{++}$.)

One easy way to establish convexity of f is via its epigraph:

$$\begin{aligned} \mathbf{epi}\, f &= \{(x, Y, t) \mid Y \succ 0,\ x^T Y^{-1} x \leq t\} \\ &= \left\{ (x, Y, t) \;\middle|\; \begin{bmatrix} Y & x \\ x^T & t \end{bmatrix} \succeq 0,\ Y \succ 0 \right\}, \end{aligned}$$

using the Schur complement condition for positive semidefiniteness of a block matrix (see §A.5.5). The last condition is a linear matrix inequality in (x, Y, t), and therefore $\mathbf{epi}\, f$ is convex.

For the special case $n = 1$, the matrix fractional function reduces to the quadratic-over-linear function x^2/y, and the associated LMI representation is

$$\begin{bmatrix} y & x \\ x & t \end{bmatrix} \succeq 0, \qquad y > 0$$

(the graph of which is shown in figure 3.3).

Many results for convex functions can be proved (or interpreted) geometrically using epigraphs, and applying results for convex sets. As an example, consider the first-order condition for convexity:

$$f(y) \geq f(x) + \nabla f(x)^T (y - x),$$

where f is convex and $x,\ y \in \mathbf{dom}\, f$. We can interpret this basic inequality geometrically in terms of $\mathbf{epi}\, f$. If $(y, t) \in \mathbf{epi}\, f$, then

$$t \geq f(y) \geq f(x) + \nabla f(x)^T (y - x).$$

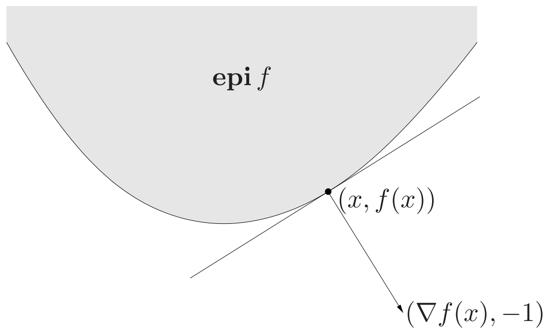


Figure 3.6 For a differentiable convex function f, the vector $(\nabla f(x), -1)$ defines a supporting hyperplane to the epigraph of f at x.

We can express this as:

$$(y,t) \in \textbf{epi}\, f \implies \begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix}^T \left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0.$$

This means that the hyperplane defined by $(\nabla f(x), -1)$ supports $\textbf{epi}\, f$ at the boundary point $(x, f(x))$; see figure 3.6.

3.1.8 Jensen's inequality and extensions

The basic inequality (3.1), *i.e.*,

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

is sometimes called *Jensen's inequality*. It is easily extended to convex combinations of more than two points: If f is convex, $x_1, \ldots, x_k \in \textbf{dom}\, f$, and $\theta_1, \ldots, \theta_k \geq 0$ with $\theta_1 + \cdots + \theta_k = 1$, then

$$f(\theta_1 x_1 + \cdots + \theta_k x_k) \leq \theta_1 f(x_1) + \cdots + \theta_k f(x_k).$$

As in the case of convex sets, the inequality extends to infinite sums, integrals, and expected values. For example, if $p(x) \geq 0$ on $S \subseteq \textbf{dom}\, f$, $\int_S p(x)\, dx = 1$, then

$$f\left(\int_S p(x) x\, dx \right) \leq \int_S f(x) p(x)\, dx,$$

provided the integrals exist. In the most general case we can take any probability measure with support in $\textbf{dom}\, f$. If x is a random variable such that $x \in \textbf{dom}\, f$ with probability one, and f is convex, then we have

$$f(\mathbf{E}\, x) \leq \mathbf{E}\, f(x), \tag{3.5}$$

provided the expectations exist. We can recover the basic inequality (3.1) from this general form, by taking the random variable x to have support $\{x_1, x_2\}$, with

$\mathbf{prob}(x = x_1) = \theta$, $\mathbf{prob}(x = x_2) = 1 - \theta$. Thus the inequality (3.5) characterizes convexity: If f is not convex, there is a random variable x, with $x \in \mathbf{dom}\, f$ with probability one, such that $f(\mathbf{E}\, x) > \mathbf{E}\, f(x)$.

All of these inequalities are now called *Jensen's inequality*, even though the inequality studied by Jensen was the very simple one

$$f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2}.$$

Remark 3.2 We can interpret (3.5) as follows. Suppose $x \in \mathbf{dom}\, f \subseteq \mathbf{R}^n$ and z is any zero mean random vector in $\mathbf{R}^n$. Then we have

$$\mathbf{E}\, f(x+z) \geq f(x).$$

Thus, randomization or dithering (*i.e.*, adding a zero mean random vector to the argument) cannot decrease the value of a convex function on average.

3.1.9 Inequalities

Many famous inequalities can be derived by applying Jensen's inequality to some appropriate convex function. (Indeed, convexity and Jensen's inequality can be made the foundation of a theory of inequalities.) As a simple example, consider the arithmetic-geometric mean inequality:

$$\sqrt{ab} \leq (a+b)/2 \tag{3.6}$$

for $a, b \geq 0$. The function $-\log x$ is convex; Jensen's inequality with $\theta = 1/2$ yields

$$-\log\left(\frac{a+b}{2}\right) \leq \frac{-\log a - \log b}{2}.$$

Taking the exponential of both sides yields (3.6).

As a less trivial example we prove Hölder's inequality: for $p > 1$, $1/p + 1/q = 1$, and $x,\ y \in \mathbf{R}^n$,

$$\sum_{i=1}^n x_i y_i \leq \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \left(\sum_{i=1}^n |y_i|^q\right)^{1/q}.$$

By convexity of $-\log x$, and Jensen's inequality with general θ, we obtain the more general arithmetic-geometric mean inequality

$$a^\theta b^{1-\theta} \leq \theta a + (1-\theta) b,$$

valid for $a,\ b \geq 0$ and $0 \leq \theta \leq 1$. Applying this with

$$a = \frac{|x_i|^p}{\sum_{j=1}^n |x_j|^p}, \qquad b = \frac{|y_i|^q}{\sum_{j=1}^n |y_j|^q}, \qquad \theta = 1/p,$$

yields

$$\left(\frac{|x_i|^p}{\sum_{j=1}^n |x_j|^p}\right)^{1/p} \left(\frac{|y_i|^q}{\sum_{j=1}^n |y_j|^q}\right)^{1/q} \leq \frac{|x_i|^p}{p \sum_{j=1}^n |x_j|^p} + \frac{|y_i|^q}{q \sum_{j=1}^n |y_j|^q}.$$

Summing over i then yields Hölder's inequality.

3.2 Operations that preserve convexity

In this section we describe some operations that preserve convexity or concavity of functions, or allow us to construct new convex and concave functions. We start with some simple operations such as addition, scaling, and pointwise supremum, and then describe some more sophisticated operations (some of which include the simple operations as special cases).

3.2.1 Nonnegative weighted sums

Evidently if f is a convex function and $\alpha \geq 0$, then the function αf is convex. If f_1 and f_2 are both convex functions, then so is their sum $f_1 + f_2$. Combining nonnegative scaling and addition, we see that the set of convex functions is itself a convex cone: a nonnegative weighted sum of convex functions,

$$f = w_1 f_1 + \cdots + w_m f_m,$$

is convex. Similarly, a nonnegative weighted sum of concave functions is concave. A nonnegative, nonzero weighted sum of strictly convex (concave) functions is strictly convex (concave).

These properties extend to infinite sums and integrals. For example if $f(x, y)$ is convex in x for each $y \in \mathcal{A}$, and $w(y) \geq 0$ for each $y \in \mathcal{A}$, then the function g defined as

$$g(x) = \int_{\mathcal{A}} w(y) f(x, y)\, dy$$

is convex in x (provided the integral exists).

The fact that convexity is preserved under nonnegative scaling and addition is easily verified directly, or can be seen in terms of the associated epigraphs. For example, if $w \geq 0$ and f is convex, we have

$$\mathbf{epi}(wf) = \begin{bmatrix} I & 0 \\ 0 & w \end{bmatrix} \mathbf{epi}\, f,$$

which is convex because the image of a convex set under a linear mapping is convex.

3.2.2 Composition with an affine mapping

Suppose $f : \mathbf{R}^n \to \mathbf{R}$, $A \in \mathbf{R}^{n \times m}$, and $b \in \mathbf{R}^n$. Define $g : \mathbf{R}^m \to \mathbf{R}$ by

$$g(x) = f(Ax + b),$$

with $\mathbf{dom}\, g = \{x \mid Ax + b \in \mathbf{dom}\, f\}$. Then if f is convex, so is g; if f is concave, so is g.

3.2.3 Pointwise maximum and supremum

If f_1 and f_2 are convex functions then their *pointwise maximum* f, defined by

$$f(x) = \max\{f_1(x), f_2(x)\},$$

with $\mathbf{dom}\, f = \mathbf{dom}\, f_1 \cap \mathbf{dom}\, f_2$, is also convex. This property is easily verified: if $0 \le \theta \le 1$ and $x,\ y \in \mathbf{dom}\, f$, then

$$\begin{array}{rcl} f(\theta x + (1-\theta)y) & = & \max\{f_1(\theta x + (1-\theta)y), f_2(\theta x + (1-\theta)y)\} \\ & \le & \max\{\theta f_1(x) + (1-\theta)f_1(y), \theta f_2(x) + (1-\theta)f_2(y)\} \\ & \le & \theta \max\{f_1(x), f_2(x)\} + (1-\theta)\max\{f_1(y), f_2(y)\} \\ & = & \theta f(x) + (1-\theta)f(y), \end{array}$$

which establishes convexity of f. It is easily shown that if $f_1, \ldots, f_m$ are convex, then their pointwise maximum

$$f(x) = \max\{f_1(x), \ldots, f_m(x)\}$$

is also convex.

Example 3.5 *Piecewise-linear functions.* The function

$$f(x) = \max\{a_1^T x + b_1, \ldots, a_L^T x + b_L\}$$

defines a piecewise-linear (or really, affine) function (with L or fewer regions). It is convex since it is the pointwise maximum of affine functions.

The converse can also be shown: any piecewise-linear convex function with L or fewer regions can be expressed in this form. (See exercise 3.29.)

Example 3.6 *Sum of r largest components.* For $x \in \mathbf{R}^n$ we denote by $x_{[i]}$ the ith largest component of x, *i.e.*,

$$x_{[1]} \ge x_{[2]} \ge \cdots \ge x_{[n]}$$

are the components of x sorted in nonincreasing order. Then the function

$$f(x) = \sum_{i=1}^{r} x_{[i]},$$

i.e., the sum of the r largest elements of x, is a convex function. This can be seen by writing it as

$$f(x) = \sum_{i=1}^{r} x_{[i]} = \max\{x_{i_1} + \cdots + x_{i_r} \mid 1 \le i_1 < i_2 < \cdots < i_r \le n\},$$

i.e., the maximum of all possible sums of r different components of x. Since it is the pointwise maximum of $n!/(r!(n-r)!)$ linear functions, it is convex.

As an extension it can be shown that the function $\sum_{i=1}^{r} w_i x_{[i]}$ is convex, provided $w_1 \ge w_2 \ge \cdots \ge w_r \ge 0$. (See exercise 3.19.)

The pointwise maximum property extends to the pointwise supremum over an infinite set of convex functions. If for each $y \in \mathcal{A}$, $f(x, y)$ is convex in x, then the function g, defined as

$$g(x) = \sup_{y \in \mathcal{A}} f(x, y) \tag{3.7}$$

is convex in x. Here the domain of g is

$$\mathbf{dom}\, g = \{x \mid (x, y) \in \mathbf{dom}\, f \text{ for all } y \in \mathcal{A},\ \sup_{y \in \mathcal{A}} f(x, y) < \infty\}.$$

Similarly, the pointwise infimum of a set of concave functions is a concave function.

In terms of epigraphs, the pointwise supremum of functions corresponds to the intersection of epigraphs: with f, g, and $\mathcal{A}$ as defined in (3.7), we have

$$\mathbf{epi}\, g = \bigcap_{y \in \mathcal{A}} \mathbf{epi}\, f(\cdot, y).$$

Thus, the result follows from the fact that the intersection of a family of convex sets is convex.

Example 3.7 *Support function of a set.* Let $C \subseteq \mathbf{R}^n$, with $C \neq \emptyset$. The *support function* S_C associated with the set C is defined as

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

(and, naturally, $\mathbf{dom}\, S_C = \{x \mid \sup_{y \in C} x^T y < \infty\}$).

For each $y \in C$, $x^T y$ is a linear function of x, so S_C is the pointwise supremum of a family of linear functions, hence convex.

Example 3.8 *Distance to farthest point of a set.* Let $C \subseteq \mathbf{R}^n$. The distance (in any norm) to the farthest point of C,

$$f(x) = \sup_{y \in C} \|x - y\|,$$

is convex. To see this, note that for any y, the function $\|x - y\|$ is convex in x. Since f is the pointwise supremum of a family of convex functions (indexed by $y \in C$), it is a convex function of x.

Example 3.9 *Least-squares cost as a function of weights.* Let $a_1, \ldots, a_n \in \mathbf{R}^m$. In a weighted least-squares problem we minimize the objective function $\sum_{i=1}^n w_i(a_i^T x - b_i)^2$ over $x \in \mathbf{R}^m$. We refer to w_i as *weights*, and allow negative w_i (which opens the possibility that the objective function is unbounded below).

We define the (optimal) *weighted least-squares cost* as

$$g(w) = \inf_x \sum_{i=1}^n w_i(a_i^T x - b_i)^2,$$

with domain

$$\mathbf{dom}\, g = \left\{ w \;\middle|\; \inf_x \sum_{i=1}^n w_i(a_i^T x - b_i)^2 > -\infty \right\}.$$

Since g is the infimum of a family of linear functions of w (indexed by $x \in \mathbf{R}^m$), it is a concave function of w.

We can derive an explicit expression for g, at least on part of its domain. Let $W = \mathbf{diag}(w)$, the diagonal matrix with elements $w_1, \dots, w_n$, and let $A \in \mathbf{R}^{n \times m}$ have rows a_i^T, so we have

$$g(w) = \inf_x (Ax - b)^T W(Ax - b) = \inf_x (x^T A^T W Ax - 2b^T W Ax + b^T W b).$$

From this we see that if $A^T W A \not\succeq 0$, the quadratic function is unbounded below in x, so $g(w) = -\infty$, *i.e.*, $w \notin \mathbf{dom}\, g$. We can give a simple expression for g when $A^T W A \succ 0$ (which defines a strict linear matrix inequality), by analytically minimizing the quadratic function:

$$\begin{aligned} g(w) &= b^T W b - b^T W A(A^T W A)^{-1} A^T W b \\ &= \sum_{i=1}^n w_i b_i^2 - \sum_{i=1}^n w_i^2 b_i^2 a_i^T \left(\sum_{j=1}^n w_j a_j a_j^T \right)^{-1} a_i. \end{aligned}$$

Concavity of g from this expression is not immediately obvious (but does follow, for example, from convexity of the matrix fractional function; see example 3.4).

Example 3.10 *Maximum eigenvalue of a symmetric matrix.* The function $f(X) = \lambda_{\max}(X)$, with $\mathbf{dom}\, f = \mathbf{S}^m$, is convex. To see this, we express f as

$$f(X) = \sup\{y^T X y \mid \|y\|_2 = 1\},$$

i.e., as the pointwise supremum of a family of linear functions of X (*i.e.*, $y^T X y$) indexed by $y \in \mathbf{R}^m$.

Example 3.11 *Norm of a matrix.* Consider $f(X) = \|X\|_2$ with $\mathbf{dom}\, f = \mathbf{R}^{p \times q}$, where $\|\cdot\|_2$ denotes the spectral norm or maximum singular value. Convexity of f follows from

$$f(X) = \sup\{u^T X v \mid \|u\|_2 = 1,\ \|v\|_2 = 1\},$$

which shows it is the pointwise supremum of a family of linear functions of X.

As a generalization suppose $\|\cdot\|_a$ and $\|\cdot\|_b$ are norms on $\mathbf{R}^p$ and $\mathbf{R}^q$, respectively. The induced norm of a matrix $X \in \mathbf{R}^{p \times q}$ is defined as

$$\|X\|_{a,b} = \sup_{v \neq 0} \frac{\|Xv\|_a}{\|v\|_b}.$$

(This reduces to the spectral norm when both norms are Euclidean.) The induced norm can be expressed as

$$\begin{aligned} \|X\|_{a,b} &= \sup\{\|Xv\|_a \mid \|v\|_b = 1\} \\ &= \sup\{u^T X v \mid \|u\|_{a*} = 1,\ \|v\|_b = 1\}, \end{aligned}$$

where $\|\cdot\|_{a*}$ is the dual norm of $\|\cdot\|_a$, and we use the fact that

$$\|z\|_a = \sup\{u^T z \mid \|u\|_{a*} = 1\}.$$

Since we have expressed $\|X\|_{a,b}$ as a supremum of linear functions of X, it is a convex function.

Representation as pointwise supremum of affine functions

The examples above illustrate a good method for establishing convexity of a function: by expressing it as the pointwise supremum of a family of affine functions. Except for a technical condition, a converse holds: almost every convex function can be expressed as the pointwise supremum of a family of affine functions. For example, if $f : \mathbf{R}^n \to \mathbf{R}$ is convex, with $\mathbf{dom}\, f = \mathbf{R}^n$, then we have

$$f(x) = \sup\{g(x) \mid g \text{ affine},\ g(z) \leq f(z) \text{ for all } z\}.$$

In other words, f is the pointwise supremum of the set of all affine global underestimators of it. We give the proof of this result below, and leave the case where $\mathbf{dom}\, f \neq \mathbf{R}^n$ as an exercise (exercise 3.28).

Suppose f is convex with $\mathbf{dom}\, f = \mathbf{R}^n$. The inequality

$$f(x) \geq \sup\{g(x) \mid g \text{ affine},\ g(z) \leq f(z) \text{ for all } z\}$$

is clear, since if g is any affine underestimator of f, we have $g(x) \leq f(x)$. To establish equality, we will show that for each $x \in \mathbf{R}^n$, there is an affine function g, which is a global underestimator of f, and satisfies $g(x) = f(x)$.

The epigraph of f is, of course, a convex set. Hence we can find a supporting hyperplane to it at $(x, f(x))$, *i.e.*, $a \in \mathbf{R}^n$ and $b \in \mathbf{R}$ with $(a, b) \neq 0$ and

$$\begin{bmatrix} a \\ b \end{bmatrix}^T \begin{bmatrix} x - z \\ f(x) - t \end{bmatrix} \leq 0$$

for all $(z, t) \in \mathbf{epi}\, f$. This means that

$$a^T(x - z) + b(f(x) - f(z) - s) \leq 0 \tag{3.8}$$

for all $z \in \mathbf{dom}\, f = \mathbf{R}^n$ and all $s \geq 0$ (since $(z, t) \in \mathbf{epi}\, f$ means $t = f(z) + s$ for some $s \geq 0$). For the inequality (3.8) to hold for all $s \geq 0$, we must have $b \geq 0$. If $b = 0$, then the inequality (3.8) reduces to $a^T(x - z) \leq 0$ for all $z \in \mathbf{R}^n$, which implies $a = 0$ and contradicts $(a, b) \neq 0$. We conclude that $b > 0$, *i.e.*, that the supporting hyperplane is not vertical.

Using the fact that $b > 0$ we rewrite (3.8) for $s = 0$ as

$$g(z) = f(x) + (a/b)^T(x - z) \leq f(z)$$

for all z. The function g is an affine underestimator of f, and satisfies $g(x) = f(x)$.

3.2.4 Composition

In this section we examine conditions on $h : \mathbf{R}^k \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}^k$ that guarantee convexity or concavity of their composition $f = h \circ g : \mathbf{R}^n \to \mathbf{R}$, defined by

$$f(x) = h(g(x)), \qquad \mathbf{dom}\, f = \{x \in \mathbf{dom}\, g \mid g(x) \in \mathbf{dom}\, h\}.$$

Scalar composition

We first consider the case $k = 1$, so $h : \mathbf{R} \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}$. We can restrict ourselves to the case $n = 1$ (since convexity is determined by the behavior of a function on arbitrary lines that intersect its domain).

To discover the composition rules, we start by assuming that h and g are twice differentiable, with $\mathbf{dom}\, g = \mathbf{dom}\, h = \mathbf{R}$. In this case, convexity of f reduces to $f'' \geq 0$ (meaning, $f''(x) \geq 0$ for all $x \in \mathbf{R}$).

The second derivative of the composition function $f = h \circ g$ is given by

$$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x). \tag{3.9}$$

Now suppose, for example, that g is convex (so $g'' \geq 0$) and h is convex and nondecreasing (so $h'' \geq 0$ and $h' \geq 0$). It follows from (3.9) that $f'' \geq 0$, *i.e.*, f is convex. In a similar way, the expression (3.9) gives the results:

$$\begin{array}{l} f \text{ is convex if } h \text{ is convex and nondecreasing, and } g \text{ is convex,} \\ f \text{ is convex if } h \text{ is convex and nonincreasing, and } g \text{ is concave,} \\ f \text{ is concave if } h \text{ is concave and nondecreasing, and } g \text{ is concave,} \\ f \text{ is concave if } h \text{ is concave and nonincreasing, and } g \text{ is convex.} \end{array} \tag{3.10}$$

These statements are valid when the functions g and h are twice differentiable and have domains that are all of $\mathbf{R}$. It turns out that very similar composition rules hold in the general case $n > 1$, without assuming differentiability of h and g, or that $\mathbf{dom}\, g = \mathbf{R}^n$ and $\mathbf{dom}\, h = \mathbf{R}$:

$$\begin{array}{l} f \text{ is convex if } h \text{ is convex, } \tilde{h} \text{ is nondecreasing, and } g \text{ is convex,} \\ f \text{ is convex if } h \text{ is convex, } \tilde{h} \text{ is nonincreasing, and } g \text{ is concave,} \\ f \text{ is concave if } h \text{ is concave, } \tilde{h} \text{ is nondecreasing, and } g \text{ is concave,} \\ f \text{ is concave if } h \text{ is concave, } \tilde{h} \text{ is nonincreasing, and } g \text{ is convex.} \end{array} \tag{3.11}$$

Here $\tilde{h}$ denotes the extended-value extension of the function h, which assigns the value ∞ ($-\infty$) to points not in $\mathbf{dom}\, h$ for h convex (concave). The only difference between these results, and the results in (3.10), is that we require that the *extended-value extension* function $\tilde{h}$ be nonincreasing or nondecreasing, on all of $\mathbf{R}$.

To understand what this means, suppose h is convex, so $\tilde{h}$ takes on the value ∞ outside $\mathbf{dom}\, h$. To say that $\tilde{h}$ is nondecreasing means that for *any* x, $y \in \mathbf{R}$, with $x < y$, we have $\tilde{h}(x) \leq \tilde{h}(y)$. In particular, this means that if $y \in \mathbf{dom}\, h$, then $x \in \mathbf{dom}\, h$. In other words, the domain of h extends infinitely in the negative direction; it is either $\mathbf{R}$, or an interval of the form $(-\infty, a)$ or $(-\infty, a]$. In a similar way, to say that h is convex and $\tilde{h}$ is nonincreasing means that h is nonincreasing and $\mathbf{dom}\, h$ extends infinitely in the positive direction. This is illustrated in figure 3.7.

Example 3.12 Some simple examples will illustrate the conditions on h that appear in the composition theorems.

- The function $h(x) = \log x$, with $\mathbf{dom}\, h = \mathbf{R}_{++}$, is concave and satisfies $\tilde{h}$ nondecreasing.

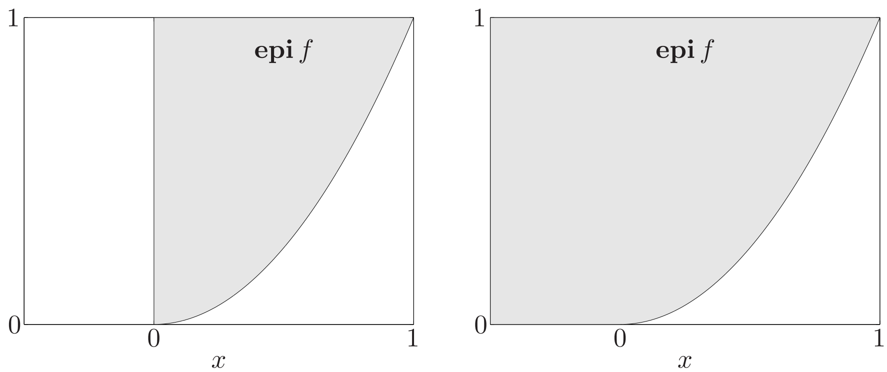


Figure 3.7 *Left.* The function x^2, with domain $\mathbf{R}_+$, is convex and nondecreasing on its domain, but its extended-value extension is *not* nondecreasing. *Right.* The function $\max\{x, 0\}^2$, with domain $\mathbf{R}$, is convex, and its extended-value extension is nondecreasing.

- The function $h(x) = x^{1/2}$, with $\mathbf{dom}\, h = \mathbf{R}_+$, is concave and satisfies the condition $\tilde{h}$ nondecreasing.
- The function $h(x) = x^{3/2}$, with $\mathbf{dom}\, h = \mathbf{R}_+$, is convex but *does not* satisfy the condition $\tilde{h}$ nondecreasing. For example, we have $\tilde{h}(-1) = \infty$, but $\tilde{h}(1) = 1$.
- The function $h(x) = x^{3/2}$ for $x \geq 0$, and $h(x) = 0$ for $x < 0$, with $\mathbf{dom}\, h = \mathbf{R}$, is convex and *does* satisfy the condition $\tilde{h}$ nondecreasing.

The composition results (3.11) can be proved directly, without assuming differentiability, or using the formula (3.9). As an example, we will prove the following composition theorem: if g is convex, h is convex, and $\tilde{h}$ is nondecreasing, then $f = h \circ g$ is convex. Assume that $x,\ y \in \mathbf{dom}\, f$, and $0 \leq \theta \leq 1$. Since $x,\ y \in \mathbf{dom}\, f$, we have that $x,\ y \in \mathbf{dom}\, g$ and $g(x),\ g(y) \in \mathbf{dom}\, h$. Since $\mathbf{dom}\, g$ is convex, we conclude that $\theta x + (1-\theta)y \in \mathbf{dom}\, g$, and from convexity of g, we have

$$g(\theta x + (1-\theta)y) \leq \theta g(x) + (1-\theta)g(y). \tag{3.12}$$

Since $g(x),\ g(y) \in \mathbf{dom}\, h$, we conclude that $\theta g(x) + (1-\theta)g(y) \in \mathbf{dom}\, h$, *i.e.*, the righthand side of (3.12) is in $\mathbf{dom}\, h$. Now we use the assumption that $\tilde{h}$ is nondecreasing, which means that its domain extends infinitely in the negative direction. Since the righthand side of (3.12) is in $\mathbf{dom}\, h$, we conclude that the lefthand side, *i.e.*, $g(\theta x+(1-\theta)y) \in \mathbf{dom}\, h$. This means that $\theta x+(1-\theta)y \in \mathbf{dom}\, f$. At this point, we have shown that $\mathbf{dom}\, f$ is convex.

Now using the fact that $\tilde{h}$ is nondecreasing and the inequality (3.12), we get

$$h(g(\theta x + (1-\theta)y)) \leq h(\theta g(x) + (1-\theta)g(y)). \tag{3.13}$$

From convexity of h, we have

$$h(\theta g(x) + (1-\theta)g(y)) \leq \theta h(g(x)) + (1-\theta)h(g(y)). \tag{3.14}$$

Putting (3.13) and (3.14) together, we have

$$h(g(\theta x + (1-\theta)y)) \le \theta h(g(x)) + (1-\theta)h(g(y)).$$

which proves the composition theorem.

Example 3.13 *Simple composition results.*

- If g is convex then $\exp g(x)$ is convex.
- If g is concave and positive, then $\log g(x)$ is concave.
- If g is concave and positive, then $1/g(x)$ is convex.
- If g is convex and nonnegative and $p \ge 1$, then $g(x)^p$ is convex.
- If g is convex then $-\log(-g(x))$ is convex on $\{x \mid g(x) < 0\}$.

Remark 3.3 The requirement that monotonicity hold for the extended-value extension $\tilde{h}$, and not just the function h, cannot be removed. For example, consider the function $g(x) = x^2$, with $\textbf{dom}\, g = \mathbf{R}$, and $h(x) = 0$, with $\textbf{dom}\, h = [1, 2]$. Here g is convex, and h is convex and nondecreasing. But the function $f = h \circ g$, given by

$$f(x) = 0, \qquad \textbf{dom}\, f = [-\sqrt{2}, -1] \cup [1, \sqrt{2}],$$

is not convex, since its domain is not convex. Here, of course, the function $\tilde{h}$ is *not* nondecreasing.

Vector composition

We now turn to the more complicated case when $k \ge 1$. Suppose

$$f(x) = h(g(x)) = h(g_1(x), \ldots, g_k(x)),$$

with $h : \mathbf{R}^k \to \mathbf{R}$, $g_i : \mathbf{R}^n \to \mathbf{R}$. Again without loss of generality we can assume $n = 1$. As in the case $k = 1$, we start by assuming the functions are twice differentiable, with $\textbf{dom}\, g = \mathbf{R}$ and $\textbf{dom}\, h = \mathbf{R}^k$, in order to discover the composition rules. We have

$$f''(x) = g'(x)^T \nabla^2 h(g(x)) g'(x) + \nabla h(g(x))^T g''(x), \tag{3.15}$$

which is the vector analog of (3.9). Again the issue is to determine conditions under which $f''(x) \ge 0$ for all x (or $f''(x) \le 0$ for all x for concavity). From (3.15) we can derive many rules, for example:

f is convex if h is convex, h is nondecreasing in each argument, and g_i are convex,

f is convex if h is convex, h is nonincreasing in each argument, and g_i are concave,

f is concave if h is concave, h is nondecreasing in each argument, and g_i are concave.

As in the scalar case, similar composition results hold in general, with $n > 1$, no assumption of differentiability of h or g, and general domains. For the general results, the monotonicity condition on h must hold for the extended-value extension $\tilde{h}$.

To understand the meaning of the condition that the extended-value extension $\tilde{h}$ be monotonic, we consider the case where $h : \mathbf{R}^k \to \mathbf{R}$ is convex, and $\tilde{h}$ nondecreasing, *i.e.*, whenever $u \preceq v$, we have $\tilde{h}(u) \leq \tilde{h}(v)$. This implies that if $v \in \mathbf{dom}\, h$, then so is u: the domain of h must extend infinitely in the $-\mathbf{R}^k_+$ directions. We can express this compactly as $\mathbf{dom}\, h - \mathbf{R}^k_+ = \mathbf{dom}\, h$.

Example 3.14 *Vector composition examples.*

- Let $h(z) = z_{[1]} + \cdots + z_{[r]}$, the sum of the r largest components of $z \in \mathbf{R}^k$. Then h is convex and nondecreasing in each argument. Suppose $g_1, \ldots, g_k$ are convex functions on $\mathbf{R}^n$. Then the composition function $f = h \circ g$, *i.e.*, the pointwise sum of the r largest g_i's, is convex.
- The function $h(z) = \log(\sum_{i=1}^k e^{z_i})$ is convex and nondecreasing in each argument, so $\log(\sum_{i=1}^k e^{g_i})$ is convex whenever g_i are.
- For $0 < p \leq 1$, the function $h(z) = (\sum_{i=1}^k z_i^p)^{1/p}$ on $\mathbf{R}^k_+$ is concave, and its extension (which has the value $-\infty$ for $z \not\succeq 0$) is nondecreasing in each component. So if g_i are concave and nonnegative, we conclude that $f(x) = (\sum_{i=1}^k g_i(x)^p)^{1/p}$ is concave.
- Suppose $p \geq 1$, and $g_1, \ldots, g_k$ are convex and nonnegative. Then the function $(\sum_{i=1}^k g_i(x)^p)^{1/p}$ is convex.

  To show this, we consider the function $h : \mathbf{R}^k \to \mathbf{R}$ defined as
  $$h(z) = \left(\sum_{i=1}^k \max\{z_i, 0\}^p \right)^{1/p},$$
  with $\mathbf{dom}\, h = \mathbf{R}^k$, so $h = \tilde{h}$. This function is convex, and nondecreasing, so we conclude $h(g(x))$ is a convex function of x. For $z \succeq 0$, we have $h(z) = (\sum_{i=1}^k z_i^p)^{1/p}$, so our conclusion is that $(\sum_{i=1}^k g_i(x)^p)^{1/p}$ is convex.
- The geometric mean $h(z) = (\prod_{i=1}^k z_i)^{1/k}$ on $\mathbf{R}^k_+$ is concave and its extension is nondecreasing in each argument. It follows that if $g_1, \ldots, g_k$ are nonnegative concave functions, then so is their geometric mean, $(\prod_{i=1}^k g_i)^{1/k}$.

3.2.5 Minimization

We have seen that the maximum or supremum of an arbitrary family of convex functions is convex. It turns out that some special forms of minimization also yield convex functions. If f is convex in (x, y), and C is a convex nonempty set, then the function
$$g(x) = \inf_{y \in C} f(x, y) \qquad (3.16)$$

is convex in x, provided $g(x) > -\infty$ for all x. The domain of g is the projection of $\mathbf{dom}\, f$ on its x-coordinates, *i.e.*,

$$\mathbf{dom}\, g = \{x \mid (x, y) \in \mathbf{dom}\, f \text{ for some } y \in C\}.$$

We prove this by verifying Jensen's inequality for $x_1,\ x_2 \in \mathbf{dom}\, g$. Let $\epsilon > 0$. Then there are $y_1,\ y_2 \in C$ such that $f(x_i, y_i) \leq g(x_i) + \epsilon$ for $i = 1,\ 2$. Now let $\theta \in [0, 1]$. We have

$$\begin{aligned} g(\theta x_1 + (1-\theta)x_2) &= \inf_{y \in C} f(\theta x_1 + (1-\theta)x_2, y) \\ &\leq f(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2) \\ &\leq \theta f(x_1, y_1) + (1-\theta) f(x_2, y_2) \\ &\leq \theta g(x_1) + (1-\theta) g(x_2) + \epsilon. \end{aligned}$$

Since this holds for any $\epsilon > 0$, we have

$$g(\theta x_1 + (1-\theta)x_2) \leq \theta g(x_1) + (1-\theta) g(x_2).$$

The result can also be seen in terms of epigraphs. With f, g, and C defined as in (3.16), and assuming the infimum over $y \in C$ is attained for each x, we have

$$\mathbf{epi}\, g = \{(x, t) \mid (x, y, t) \in \mathbf{epi}\, f \text{ for some } y \in C\}.$$

Thus $\mathbf{epi}\, g$ is convex, since it is the projection of a convex set on some of its components.

Example 3.15 *Schur complement.* Suppose the quadratic function

$$f(x, y) = x^T Ax + 2x^T By + y^T Cy,$$

(where A and C are symmetric) is convex in (x, y), which means

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0.$$

We can express $g(x) = \inf_y f(x, y)$ as

$$g(x) = x^T (A - BC^\dagger B^T) x,$$

where $C^\dagger$ is the pseudo-inverse of C (see §A.5.4). By the minimization rule, g is convex, so we conclude that $A - BC^\dagger B^T \succeq 0$.

If C is invertible, *i.e.*, $C \succ 0$, then the matrix $A - BC^{-1}B^T$ is called the *Schur complement* of C in the matrix

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

(see §A.5.5).

Example 3.16 *Distance to a set.* The distance of a point x to a set $S \subseteq \mathbf{R}^n$, in the norm $\|\cdot\|$, is defined as

$$\mathbf{dist}(x, S) = \inf_{y \in S} \|x - y\|.$$

The function $\|x-y\|$ is convex in (x, y), so if the set S is convex, the distance function $\mathbf{dist}(x, S)$ is a convex function of x.

Example 3.17 Suppose h is convex. Then the function g defined as

$$g(x) = \inf\{h(y) \mid Ay = x\}$$

is convex. To see this, we define f by

$$f(x, y) = \begin{cases} h(y) & \text{if } Ay = x \\ \infty & \text{otherwise,} \end{cases}$$

which is convex in (x, y). Then g is the minimum of f over y, and hence is convex. (It is not hard to show directly that g is convex.)

3.2.6 Perspective of a function

If $f : \mathbf{R}^n \to \mathbf{R}$, then the *perspective* of f is the function $g : \mathbf{R}^{n+1} \to \mathbf{R}$ defined by

$$g(x, t) = tf(x/t),$$

with domain

$$\mathbf{dom}\, g = \{(x, t) \mid x/t \in \mathbf{dom}\, f,\ t > 0\}.$$

The perspective operation preserves convexity: If f is a convex function, then so is its perspective function g. Similarly, if f is concave, then so is g.

This can be proved several ways, for example, direct verification of the defining inequality (see exercise 3.33). We give a short proof here using epigraphs and the perspective mapping on $\mathbf{R}^{n+1}$ described in §2.3.3 (which will also explain the name 'perspective'). For $t > 0$ we have

$$\begin{aligned} (x, t, s) \in \mathbf{epi}\, g &\iff tf(x/t) \le s \\ &\iff f(x/t) \le s/t \\ &\iff (x/t, s/t) \in \mathbf{epi}\, f. \end{aligned}$$

Therefore $\mathbf{epi}\, g$ is the inverse image of $\mathbf{epi}\, f$ under the perspective mapping that takes (u, v, w) to $(u, w)/v$. It follows (see §2.3.3) that $\mathbf{epi}\, g$ is convex, so the function g is convex.

Example 3.18 *Euclidean norm squared.* The perspective of the convex function $f(x) = x^T x$ on $\mathbf{R}^n$ is

$$g(x, t) = t(x/t)^T(x/t) = \frac{x^T x}{t},$$

which is convex in (x, t) for $t > 0$.

We can deduce convexity of g using several other methods. First, we can express g as the sum of the quadratic-over-linear functions x_i^2/t, which were shown to be convex in §3.1.5. We can also express g as a special case of the matrix fractional function $x^T(tI)^{-1}x$ (see example 3.4).

Example 3.19 *Negative logarithm.* Consider the convex function $f(x) = -\log x$ on $\mathbf{R}_{++}$. Its perspective is

$$g(x,t) = -t\log(x/t) = t\log(t/x) = t\log t - t\log x,$$

and is convex on $\mathbf{R}^2_{++}$. The function g is called the *relative entropy* of t and x. For $x = 1$, g reduces to the negative entropy function.

From convexity of g we can establish convexity or concavity of several interesting related functions. First, the relative entropy of two vectors $u,\ v \in \mathbf{R}^n_{++}$, defined as

$$\sum_{i=1}^n u_i \log(u_i/v_i),$$

is convex in (u,v), since it is a sum of relative entropies of $u_i,\ v_i$.

A closely related function is the *Kullback-Leibler divergence* between $u,\ v \in \mathbf{R}^n_{++}$, given by

$$D_{\mathrm{kl}}(u,v) = \sum_{i=1}^n \left(u_i \log(u_i/v_i) - u_i + v_i\right), \tag{3.17}$$

which is convex, since it is the relative entropy plus a linear function of (u,v). The Kullback-Leibler divergence satisfies $D_{\mathrm{kl}}(u,v) \geq 0$, and $D_{\mathrm{kl}}(u,v) = 0$ if and only if $u = v$, and so can be used as a measure of deviation between two positive vectors; see exercise 3.13. (Note that the relative entropy and the Kullback-Leibler divergence are the same when u and v are probability vectors, *i.e.*, satisfy $\mathbf{1}^T u = \mathbf{1}^T v = 1$.)

If we take $v_i = \mathbf{1}^T u$ in the relative entropy function, we obtain the concave (and homogeneous) function of $u \in \mathbf{R}^n_{++}$ given by

$$\sum_{i=1}^n u_i \log(\mathbf{1}^T u/u_i) = (\mathbf{1}^T u)\sum_{i=1}^n z_i \log(1/z_i),$$

where $z = u/(\mathbf{1}^T u)$, which is called the *normalized entropy* function. The vector $z = u/\mathbf{1}^T u$ is a normalized vector or probability distribution, since its components sum to one; the normalized entropy of u is $\mathbf{1}^T u$ times the entropy of this normalized distribution.

Example 3.20 Suppose $f : \mathbf{R}^m \to \mathbf{R}$ is convex, and $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$, $c \in \mathbf{R}^n$, and $d \in \mathbf{R}$. We define

$$g(x) = (c^T x + d) f\left((Ax+b)/(c^T x + d)\right),$$

with

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

Then g is convex.

3.3 The conjugate function

In this section we introduce an operation that will play an important role in later chapters.

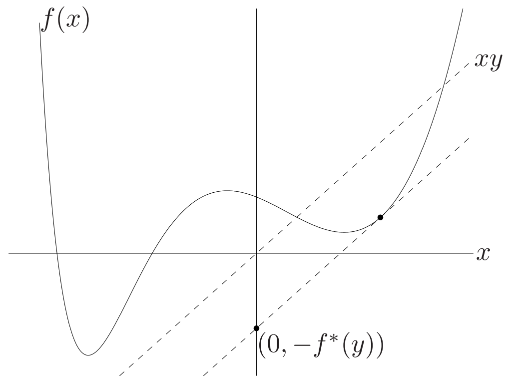


Figure 3.8 A function $f : \mathbf{R} \to \mathbf{R}$, and a value $y \in \mathbf{R}$. The conjugate function $f^*(y)$ is the maximum gap between the linear function yx and $f(x)$, as shown by the dashed line in the figure. If f is differentiable, this occurs at a point x where $f'(x) = y$.

3.3.1 Definition and examples

Let $f : \mathbf{R}^n \to \mathbf{R}$. The function $f^* : \mathbf{R}^n \to \mathbf{R}$, defined as

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} \left(y^T x - f(x) \right), \tag{3.18}$$

is called the *conjugate* of the function f. The domain of the conjugate function consists of $y \in \mathbf{R}^n$ for which the supremum is finite, *i.e.*, for which the difference $y^T x - f(x)$ is bounded above on $\mathbf{dom}\, f$. This definition is illustrated in figure 3.8.

We see immediately that f^* is a convex function, since it is the pointwise supremum of a family of convex (indeed, affine) functions of y. This is true whether or not f is convex. (Note that when f is convex, the subscript $x \in \mathbf{dom}\, f$ is not necessary since, by convention, $y^T x - f(x) = -\infty$ for $x \notin \mathbf{dom}\, f$.)

We start with some simple examples, and then describe some rules for conjugating functions. This allows us to derive an analytical expression for the conjugate of many common convex functions.

Example 3.21 We derive the conjugates of some convex functions on $\mathbf{R}$.

- *Affine function.* $f(x) = ax + b$. As a function of x, $yx - ax - b$ is bounded if and only if $y = a$, in which case it is constant. Therefore the domain of the conjugate function f^* is the singleton $\{a\}$, and $f^*(a) = -b$.
- *Negative logarithm.* $f(x) = -\log x$, with $\mathbf{dom}\, f = \mathbf{R}_{++}$. The function $xy + \log x$ is unbounded above if $y \geq 0$ and reaches its maximum at $x = -1/y$ otherwise. Therefore, $\mathbf{dom}\, f^* = \{y \mid y < 0\} = -\mathbf{R}_{++}$ and $f^*(y) = -\log(-y) - 1$ for $y < 0$.
- *Exponential.* $f(x) = e^x$. $xy - e^x$ is unbounded if $y < 0$. For $y > 0$, $xy - e^x$ reaches its maximum at $x = \log y$, so we have $f^*(y) = y \log y - y$. For $y = 0$,

$f^*(y) = \sup_x -e^x = 0$. In summary, $\mathbf{dom}\, f^* = \mathbf{R}_+$ and $f^*(y) = y \log y - y$ (with the interpretation $0 \log 0 = 0$).

- *Negative entropy.* $f(x) = x \log x$, with $\mathbf{dom}\, f = \mathbf{R}_+$ (and $f(0) = 0$). The function $xy - x \log x$ is bounded above on $\mathbf{R}_+$ for all y, hence $\mathbf{dom}\, f^* = \mathbf{R}$. It attains its maximum at $x = e^{y-1}$, and substituting we find $f^*(y) = e^{y-1}$.
- *Inverse.* $f(x) = 1/x$ on $\mathbf{R}_{++}$. For $y > 0$, $yx - 1/x$ is unbounded above. For $y = 0$ this function has supremum 0; for $y < 0$ the supremum is attained at $x = (-y)^{-1/2}$. Therefore we have $f^*(y) = -2(-y)^{1/2}$, with $\mathbf{dom}\, f^* = -\mathbf{R}_+$.

Example 3.22 *Strictly convex quadratic function.* Consider $f(x) = \frac{1}{2}x^T Q x$, with $Q \in \mathbf{S}^n_{++}$. The function $y^T x - \frac{1}{2}x^T Q x$ is bounded above as a function of x for all y. It attains its maximum at $x = Q^{-1}y$, so

$$f^*(y) = \frac{1}{2} y^T Q^{-1} y.$$

Example 3.23 *Log-determinant.* We consider $f(X) = \log\det X^{-1}$ on $\mathbf{S}^n_{++}$. The conjugate function is defined as

$$f^*(Y) = \sup_{X \succ 0} \left(\mathbf{tr}(YX) + \log\det X\right),$$

since $\mathbf{tr}(YX)$ is the standard inner product on $\mathbf{S}^n$. We first show that $\mathbf{tr}(YX) + \log\det X$ is unbounded above unless $Y \prec 0$. If $Y \not\prec 0$, then Y has an eigenvector v, with $\|v\|_2 = 1$, and eigenvalue $\lambda \geq 0$. Taking $X = I + tvv^T$ we find that

$$\mathbf{tr}(YX) + \log\det X = \mathbf{tr}\, Y + t\lambda + \log\det(I + tvv^T) = \mathbf{tr}\, Y + t\lambda + \log(1+t),$$

which is unbounded above as $t \to \infty$.

Now consider the case $Y \prec 0$. We can find the maximizing X by setting the gradient with respect to X equal to zero:

$$\nabla_X \left(\mathbf{tr}(YX) + \log\det X\right) = Y + X^{-1} = 0$$

(see §A.4.1), which yields $X = -Y^{-1}$ (which is, indeed, positive definite). Therefore we have

$$f^*(Y) = \log\det(-Y)^{-1} - n,$$

with $\mathbf{dom}\, f^* = -\mathbf{S}^n_{++}$.

Example 3.24 *Indicator function.* Let I_S be the indicator function of a (not necessarily convex) set $S \subseteq \mathbf{R}^n$, *i.e.*, $I_S(x) = 0$ on $\mathbf{dom}\, I_S = S$. Its conjugate is

$$I_S^*(y) = \sup_{x \in S} y^T x,$$

which is the support function of the set S.

Example 3.25 *Log-sum-exp function.* To derive the conjugate of the log-sum-exp function $f(x) = \log(\sum_{i=1}^n e^{x_i})$, we first determine the values of y for which the maximum over x of $y^T x - f(x)$ is attained. By setting the gradient with respect to x equal to zero, we obtain the condition

$$y_i = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}, \quad i = 1, \ldots, n.$$

These equations are solvable for x if and only if $y \succ 0$ and $\mathbf{1}^T y = 1$. By substituting the expression for y_i into $y^T x - f(x)$ we obtain $f^*(y) = \sum_{i=1}^n y_i \log y_i$. This expression for f^* is still correct if some components of y are zero, as long as $y \succeq 0$ and $\mathbf{1}^T y = 1$, and we interpret $0 \log 0$ as 0.

In fact the domain of f^* is exactly given by $\mathbf{1}^T y = 1$, $y \succeq 0$. To show this, suppose that a component of y is negative, say, $y_k < 0$. Then we can show that $y^T x - f(x)$ is unbounded above by choosing $x_k = -t$, and $x_i = 0$, $i \neq k$, and letting t go to infinity.

If $y \succeq 0$ but $\mathbf{1}^T y \neq 1$, we choose $x = t\mathbf{1}$, so that

$$y^T x - f(x) = t\mathbf{1}^T y - t - \log n.$$

If $\mathbf{1}^T y > 1$, this grows unboundedly as $t \to \infty$; if $\mathbf{1}^T y < 1$, it grows unboundedly as $t \to -\infty$.

In summary,

$$f^*(y) = \begin{cases} \sum_{i=1}^n y_i \log y_i & \text{if } y \succeq 0 \text{ and } \mathbf{1}^T y = 1 \\ \infty & \text{otherwise.} \end{cases}$$

In other words, the conjugate of the log-sum-exp function is the negative entropy function, restricted to the probability simplex.

Example 3.26 *Norm.* Let $\|\cdot\|$ be a norm on $\mathbf{R}^n$, with dual norm $\|\cdot\|_*$. We will show that the conjugate of $f(x) = \|x\|$ is

$$f^*(y) = \begin{cases} 0 & \|y\|_* \leq 1 \\ \infty & \text{otherwise,} \end{cases}$$

i.e., the conjugate of a norm is the indicator function of the dual norm unit ball.

If $\|y\|_* > 1$, then by definition of the dual norm, there is a $z \in \mathbf{R}^n$ with $\|z\| \leq 1$ and $y^T z > 1$. Taking $x = tz$ and letting $t \to \infty$, we have

$$y^T x - \|x\| = t(y^T z - \|z\|) \to \infty,$$

which shows that $f^*(y) = \infty$. Conversely, if $\|y\|_* \leq 1$, then we have $y^T x \leq \|x\| \|y\|_*$ for all x, which implies for all x, $y^T x - \|x\| \leq 0$. Therefore $x = 0$ is the value that maximizes $y^T x - \|x\|$, with maximum value 0.

Example 3.27 *Norm squared.* Now consider the function $f(x) = (1/2)\|x\|^2$, where $\|\cdot\|$ is a norm, with dual norm $\|\cdot\|_*$. We will show that its conjugate is $f^*(y) = (1/2)\|y\|_*^2$. From $y^T x \leq \|y\|_* \|x\|$, we conclude

$$y^T x - (1/2)\|x\|^2 \leq \|y\|_* \|x\| - (1/2)\|x\|^2$$

for all x. The righthand side is a quadratic function of $\|x\|$, which has maximum value $(1/2)\|y\|_*^2$. Therefore for all x, we have

$$y^T x - (1/2)\|x\|^2 \leq (1/2)\|y\|_*^2,$$

which shows that $f^*(y) \leq (1/2)\|y\|_*^2$.

To show the other inequality, let x be any vector with $y^T x = \|y\|_* \|x\|$, scaled so that $\|x\| = \|y\|_*$. Then we have, for this x,

$$y^T x - (1/2)\|x\|^2 = (1/2)\|y\|_*^2,$$

which shows that $f^*(y) \geq (1/2)\|y\|_*^2$.

Example 3.28 *Revenue and profit functions.* We consider a business or enterprise that consumes n resources and produces a product that can be sold. We let $r = (r_1, \ldots, r_n)$ denote the vector of resource quantities consumed, and $S(r)$ denote the sales revenue derived from the product produced (as a function of the resources consumed). Now let p_i denote the price (per unit) of resource i, so the total amount paid for resources by the enterprise is $p^T r$. The profit derived by the firm is then $S(r) - p^T r$. Let us fix the prices of the resources, and ask what is the maximum profit that can be made, by wisely choosing the quantities of resources consumed. This maximum profit is given by

$$M(p) = \sup_r \left(S(r) - p^T r \right).$$

The function $M(p)$ gives the maximum profit attainable, as a function of the resource prices. In terms of conjugate functions, we can express M as

$$M(p) = (-S)^*(-p).$$

Thus the maximum profit (as a function of resource prices) is closely related to the conjugate of gross sales (as a function of resources consumed).

3.3.2 Basic properties

Fenchel's inequality

From the definition of conjugate function, we immediately obtain the inequality

$$f(x) + f^*(y) \geq x^T y$$

for all x, y. This is called *Fenchel's inequality* (or *Young's inequality* when f is differentiable).

For example with $f(x) = (1/2)x^T Q x$, where $Q \in \mathbf{S}_{++}^n$, we obtain the inequality

$$x^T y \leq (1/2)x^T Q x + (1/2)y^T Q^{-1} y.$$

Conjugate of the conjugate

The examples above, and the name 'conjugate', suggest that the conjugate of the conjugate of a convex function is the original function. This is the case provided a technical condition holds: if f is convex, and f is closed (*i.e.*, $\mathbf{epi}\, f$ is a closed set; see §A.3.3), then $f^{**} = f$. For example, if $\mathbf{dom}\, f = \mathbf{R}^n$, then we have $f^{**} = f$, *i.e.*, the conjugate of the conjugate of f is f again (see exercise 3.39).

Differentiable functions

The conjugate of a differentiable function f is also called the *Legendre transform* of f. (To distinguish the general definition from the differentiable case, the term *Fenchel conjugate* is sometimes used instead of conjugate.)

Suppose f is convex and differentiable, with $\mathbf{dom}\, f = \mathbf{R}^n$. Any maximizer x^* of $y^T x - f(x)$ satisfies $y = \nabla f(x^*)$, and conversely, if x^* satisfies $y = \nabla f(x^*)$, then x^* maximizes $y^T x - f(x)$. Therefore, if $y = \nabla f(x^*)$, we have

$$f^*(y) = x^{*T} \nabla f(x^*) - f(x^*).$$

This allows us to determine $f^*(y)$ for any y for which we can solve the gradient equation $y = \nabla f(z)$ for z.

We can express this another way. Let $z \in \mathbf{R}^n$ be arbitrary and define $y = \nabla f(z)$. Then we have

$$f^*(y) = z^T \nabla f(z) - f(z).$$

Scaling and composition with affine transformation

For $a > 0$ and $b \in \mathbf{R}$, the conjugate of $g(x) = af(x) + b$ is $g^*(y) = af^*(y/a) - b$.

Suppose $A \in \mathbf{R}^{n \times n}$ is nonsingular and $b \in \mathbf{R}^n$. Then the conjugate of $g(x) = f(Ax + b)$ is

$$g^*(y) = f^*(A^{-T} y) - b^T A^{-T} y,$$

with $\mathbf{dom}\, g^* = A^T \mathbf{dom}\, f^*$.

Sums of independent functions

If $f(u, v) = f_1(u) + f_2(v)$, where f_1 and f_2 are convex functions with conjugates f_1^* and f_2^*, respectively, then

$$f^*(w, z) = f_1^*(w) + f_2^*(z).$$

In other words, the conjugate of the sum of *independent* convex functions is the sum of the conjugates. ('Independent' means they are functions of different variables.)

3.4 Quasiconvex functions

3.4.1 Definition and examples

A function $f : \mathbf{R}^n \to \mathbf{R}$ is called *quasiconvex* (or *unimodal*) if its domain and all its sublevel sets

$$S_\alpha = \{x \in \mathbf{dom}\, f \mid f(x) \leq \alpha\},$$

for $\alpha \in \mathbf{R}$, are convex. A function is *quasiconcave* if $-f$ is quasiconvex, *i.e.*, every superlevel set $\{x \mid f(x) \geq \alpha\}$ is convex. A function that is both quasiconvex and quasiconcave is called *quasilinear*. If a function f is quasilinear, then its domain, and every level set $\{x \mid f(x) = \alpha\}$ is convex.

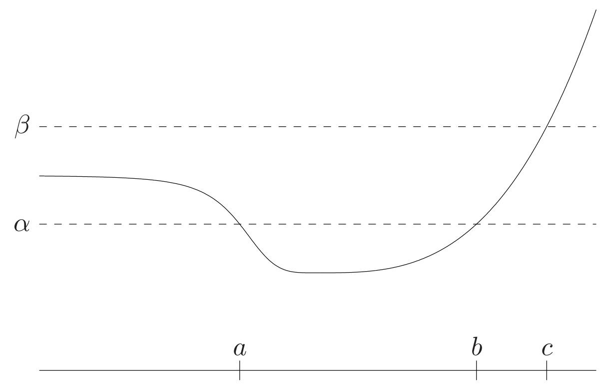


Figure 3.9 A quasiconvex function on $\mathbf{R}$. For each α, the α-sublevel set S_α is convex, *i.e.*, an interval. The sublevel set S_α is the interval $[a, b]$. The sublevel set S_β is the interval $(-\infty, c]$.

For a function on $\mathbf{R}$, quasiconvexity requires that each sublevel set be an interval (including, possibly, an infinite interval). An example of a quasiconvex function on $\mathbf{R}$ is shown in figure 3.9.

Convex functions have convex sublevel sets, and so are quasiconvex. But simple examples, such as the one shown in figure 3.9, show that the converse is not true.

Example 3.29 Some examples on $\mathbf{R}$:

- *Logarithm.* $\log x$ on $\mathbf{R}_{++}$ is quasiconvex (and quasiconcave, hence quasilinear).
- *Ceiling function.* $\operatorname{ceil}(x) = \inf\{z \in \mathbf{Z} \mid z \geq x\}$ is quasiconvex (and quasiconcave).

These examples show that quasiconvex functions can be concave, or discontinuous. We now give some examples on $\mathbf{R}^n$.

Example 3.30 *Length of a vector.* We define the *length* of $x \in \mathbf{R}^n$ as the largest index of a nonzero component, *i.e.*,

$$f(x) = \max\{i \mid x_i \neq 0\}.$$

(We define the length of the zero vector to be zero.) This function is quasiconvex on $\mathbf{R}^n$, since its sublevel sets are subspaces:

$$f(x) \leq \alpha \iff x_i = 0 \text{ for } i = \lfloor \alpha \rfloor + 1, \ldots, n.$$

Example 3.31 Consider $f : \mathbf{R}^2 \to \mathbf{R}$, with $\mathbf{dom}\, f = \mathbf{R}^2_+$ and $f(x_1, x_2) = x_1 x_2$. This function is neither convex nor concave since its Hessian

$$\nabla^2 f(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

is indefinite; it has one positive and one negative eigenvalue. The function f is quasiconcave, however, since the superlevel sets

$$\{x \in \mathbf{R}_+^2 \mid x_1 x_2 \geq \alpha\}$$

are convex sets for all α. (Note, however, that f is *not* quasiconcave on $\mathbf{R}^2$.)

Example 3.32 *Linear-fractional function.* The function

$$f(x) = \frac{a^T x + b}{c^T x + d},$$

with $\mathbf{dom}\, f = \{x \mid c^T x + d > 0\}$, is quasiconvex, and quasiconcave, *i.e.*, quasilinear. Its α-sublevel set is

$$\begin{aligned} S_\alpha &= \{x \mid c^T x + d > 0,\ (a^T x + b)/(c^T x + d) \leq \alpha\} \\ &= \{x \mid c^T x + d > 0,\ a^T x + b \leq \alpha(c^T x + d)\}, \end{aligned}$$

which is convex, since it is the intersection of an open halfspace and a closed halfspace. (The same method can be used to show its superlevel sets are convex.)

Example 3.33 *Distance ratio function.* Suppose $a, b \in \mathbf{R}^n$, and define

$$f(x) = \frac{\|x - a\|_2}{\|x - b\|_2},$$

i.e., the ratio of the Euclidean distance to a to the distance to b. Then f is quasiconvex on the halfspace $\{x \mid \|x - a\|_2 \leq \|x - b\|_2\}$. To see this, we consider the α-sublevel set of f, with $\alpha \leq 1$ since $f(x) \leq 1$ on the halfspace $\{x \mid \|x - a\|_2 \leq \|x - b\|_2\}$. This sublevel set is the set of points satisfying

$$\|x - a\|_2 \leq \alpha \|x - b\|_2.$$

Squaring both sides, and rearranging terms, we see that this is equivalent to

$$(1 - \alpha^2) x^T x - 2(a - \alpha^2 b)^T x + a^T a - \alpha^2 b^T b \leq 0.$$

This describes a convex set (in fact a Euclidean ball) if $\alpha \leq 1$.

Example 3.34 *Internal rate of return.* Let $x = (x_0, x_1, \ldots, x_n)$ denote a cash flow sequence over n periods, where $x_i > 0$ means a payment to us in period i, and $x_i < 0$ means a payment by us in period i. We define the *present value* of a cash flow, with interest rate $r \geq 0$, to be

$$\mathrm{PV}(x, r) = \sum_{i=0}^{n} (1 + r)^{-i} x_i.$$

(The factor $(1 + r)^{-i}$ is a *discount factor* for a payment by or to us in period i.)

Now we consider cash flows for which $x_0 < 0$ and $x_0 + x_1 + \cdots + x_n > 0$. This means that we start with an investment of $|x_0|$ in period 0, and that the total of the remaining cash flow, $x_1 + \cdots + x_n$, (not taking any discount factors into account) exceeds our initial investment.

For such a cash flow, $\mathrm{PV}(x, 0) > 0$ and $\mathrm{PV}(x, r) \to x_0 < 0$ as $r \to \infty$, so it follows that for at least one $r \geq 0$, we have $\mathrm{PV}(x, r) = 0$. We define the *internal rate of return* of the cash flow as the smallest interest rate $r \geq 0$ for which the present value is zero:

$$\mathrm{IRR}(x) = \inf\{r \geq 0 \mid \mathrm{PV}(x, r) = 0\}.$$

Internal rate of return is a quasiconcave function of x (restricted to $x_0 < 0$, $x_1 + \cdots + x_n > 0$). To see this, we note that

$$\mathrm{IRR}(x) \geq R \iff \mathrm{PV}(x, r) > 0 \text{ for } 0 \leq r < R.$$

The lefthand side defines the R-superlevel set of IRR. The righthand side is the intersection of the sets $\{x \mid \mathrm{PV}(x, r) > 0\}$, indexed by r, over the range $0 \leq r < R$. For each r, $\mathrm{PV}(x, r) > 0$ defines an open halfspace, so the righthand side defines a convex set.

3.4.2 Basic properties

The examples above show that quasiconvexity is a considerable generalization of convexity. Still, many of the properties of convex functions hold, or have analogs, for quasiconvex functions. For example, there is a variation on Jensen's inequality that characterizes quasiconvexity: A function f is quasiconvex if and only if $\mathbf{dom}\, f$ is convex and for any $x,\ y \in \mathbf{dom}\, f$ and $0 \leq \theta \leq 1$,

$$f(\theta x + (1 - \theta)y) \leq \max\{f(x), f(y)\}, \tag{3.19}$$

i.e., the value of the function on a segment does not exceed the maximum of its values at the endpoints. The inequality (3.19) is sometimes called Jensen's inequality for quasiconvex functions, and is illustrated in figure 3.10.

Example 3.35 *Cardinality of a nonnegative vector.* The *cardinality* or *size* of a vector $x \in \mathbf{R}^n$ is the number of nonzero components, and denoted $\mathbf{card}(x)$. The function $\mathbf{card}$ is quasiconcave on $\mathbf{R}^n_+$ (but not $\mathbf{R}^n$). This follows immediately from the modified Jensen inequality

$$\mathbf{card}(x + y) \geq \min\{\mathbf{card}(x), \mathbf{card}(y)\},$$

which holds for $x,\ y \succeq 0$.

Example 3.36 *Rank of positive semidefinite matrix.* The function $\mathbf{rank}\, X$ is quasiconcave on $\mathbf{S}^n_+$. This follows from the modified Jensen inequality (3.19),

$$\mathbf{rank}(X + Y) \geq \min\{\mathbf{rank}\, X, \mathbf{rank}\, Y\}$$

which holds for $X,\ Y \in \mathbf{S}^n_+$. (This can be considered an extension of the previous example, since $\mathbf{rank}(\mathbf{diag}(x)) = \mathbf{card}(x)$ for $x \succeq 0$.)

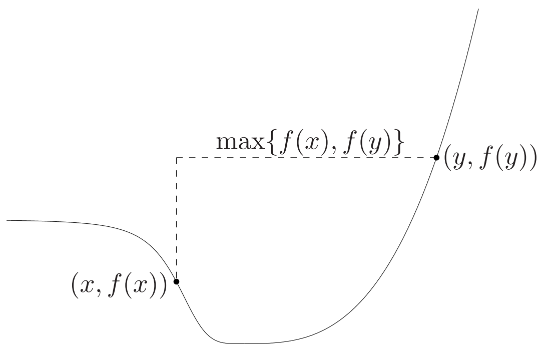


Figure 3.10 A quasiconvex function on $\mathbf{R}$. The value of f between x and y is no more than $\max\{f(x), f(y)\}$.

Like convexity, quasiconvexity is characterized by the behavior of a function f on lines: f is quasiconvex if and only if its restriction to any line intersecting its domain is quasiconvex. In particular, quasiconvexity of a function can be verified by restricting it to an arbitrary line, and then checking quasiconvexity of the resulting function on $\mathbf{R}$.

Quasiconvex functions on R

We can give a simple characterization of quasiconvex functions on $\mathbf{R}$. We consider continuous functions, since stating the conditions in the general case is cumbersome. A continuous function $f : \mathbf{R} \to \mathbf{R}$ is quasiconvex if and only if at least one of the following conditions holds:

- f is nondecreasing
- f is nonincreasing
- there is a point $c \in \mathbf{dom}\, f$ such that for $t \leq c$ (and $t \in \mathbf{dom}\, f$), f is nonincreasing, and for $t \geq c$ (and $t \in \mathbf{dom}\, f$), f is nondecreasing.

The point c can be chosen as any point which is a global minimizer of f. Figure 3.11 illustrates this.

3.4.3 Differentiable quasiconvex functions

First-order conditions

Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is differentiable. Then f is quasiconvex if and only if $\mathbf{dom}\, f$ is convex and for all x, $y \in \mathbf{dom}\, f$

$$f(y) \leq f(x) \Longrightarrow \nabla f(x)^T (y - x) \leq 0. \tag{3.20}$$

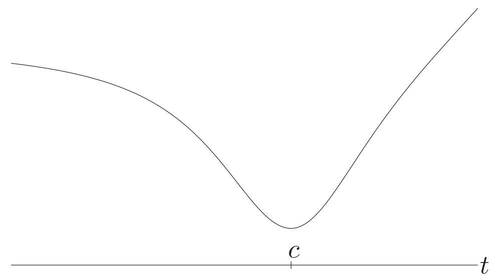


Figure 3.11 A quasiconvex function on $\mathbf{R}$. The function is nonincreasing for $t \leq c$ and nondecreasing for $t \geq c$.

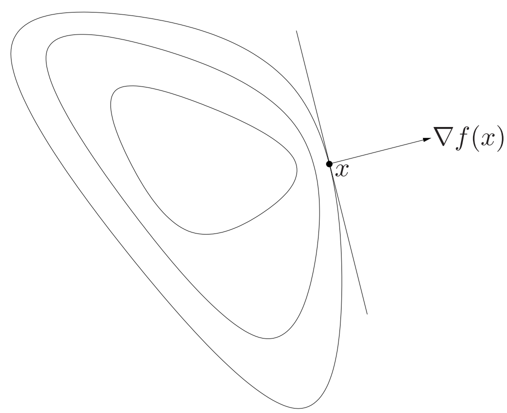


Figure 3.12 Three level curves of a quasiconvex function f are shown. The vector $\nabla f(x)$ defines a supporting hyperplane to the sublevel set $\{z \mid f(z) \leq f(x)\}$ at x.

This is the analog of inequality (3.2), for quasiconvex functions. We leave the proof as an exercise (exercise 3.43).

The condition (3.20) has a simple geometric interpretation when $\nabla f(x) \neq 0$. It states that $\nabla f(x)$ defines a supporting hyperplane to the sublevel set $\{y \mid f(y) \leq f(x)\}$, at the point x, as illustrated in figure 3.12.

While the first-order condition for convexity (3.2), and the first-order condition for quasiconvexity (3.20) are similar, there are some important differences. For example, if f is convex and $\nabla f(x) = 0$, then x is a global minimizer of f. But this statement is *false* for quasiconvex functions: it is possible that $\nabla f(x) = 0$, but x is not a global minimizer of f.

Second-order conditions

Now suppose f is twice differentiable. If f is quasiconvex, then for all $x \in \mathbf{dom}\, f$, and all $y \in \mathbf{R}^n$, we have

$$y^T \nabla f(x) = 0 \Longrightarrow y^T \nabla^2 f(x) y \geq 0. \tag{3.21}$$

For a quasiconvex function on $\mathbf{R}$, this reduces to the simple condition

$$f'(x) = 0 \Longrightarrow f''(x) \geq 0,$$

i.e., at any point with zero slope, the second derivative is nonnegative. For a quasiconvex function on $\mathbf{R}^n$, the interpretation of the condition (3.21) is a bit more complicated. As in the case $n = 1$, we conclude that whenever $\nabla f(x) = 0$, we must have $\nabla^2 f(x) \succeq 0$. When $\nabla f(x) \neq 0$, the condition (3.21) means that $\nabla^2 f(x)$ is positive semidefinite on the $(n-1)$-dimensional subspace $\nabla f(x)^\perp$. This implies that $\nabla^2 f(x)$ can have at most one negative eigenvalue.

As a (partial) converse, if f satisfies

$$y^T \nabla f(x) = 0 \Longrightarrow y^T \nabla^2 f(x) y > 0 \tag{3.22}$$

for all $x \in \mathbf{dom}\, f$ and all $y \in \mathbf{R}^n$, $y \neq 0$, then f is quasiconvex. This condition is the same as requiring $\nabla^2 f(x)$ to be positive definite for any point with $\nabla f(x) = 0$, and for all other points, requiring $\nabla^2 f(x)$ to be positive definite on the $(n-1)$-dimensional subspace $\nabla f(x)^\perp$.

Proof of second-order conditions for quasiconvexity

By restricting the function to an arbitrary line, it suffices to consider the case in which $f : \mathbf{R} \to \mathbf{R}$.

We first show that if $f : \mathbf{R} \to \mathbf{R}$ is quasiconvex on an interval (a, b), then it must satisfy (3.21), *i.e.*, if $f'(c) = 0$ with $c \in (a, b)$, then we must have $f''(c) \geq 0$. If $f'(c) = 0$ with $c \in (a, b)$, $f''(c) < 0$, then for small positive ϵ we have $f(c - \epsilon) < f(c)$ and $f(c + \epsilon) < f(c)$. It follows that the sublevel set $\{x \mid f(x) \leq f(c) - \epsilon\}$ is disconnected for small positive ϵ, and therefore not convex, which contradicts our assumption that f is quasiconvex.

Now we show that if the condition (3.22) holds, then f is quasiconvex. Assume that (3.22) holds, *i.e.*, for each $c \in (a, b)$ with $f'(c) = 0$, we have $f''(c) > 0$. This means that whenever the function f' crosses the value 0, it is strictly increasing. Therefore it can cross the value 0 at most once. If f' does not cross the value 0 at all, then f is either nonincreasing or nondecreasing on (a, b), and therefore quasiconvex. Otherwise it must cross the value 0 exactly once, say at $c \in (a, b)$. Since $f''(c) > 0$, it follows that $f'(t) \leq 0$ for $a < t \leq c$, and $f'(t) \geq 0$ for $c \leq t < b$. This shows that f is quasiconvex.

3.4.4 Operations that preserve quasiconvexity

Nonnegative weighted maximum

A nonnegative weighted maximum of quasiconvex functions, *i.e.*,

$$f = \max\{w_1 f_1, \ldots, w_m f_m\},$$

with $w_i \geq 0$ and f_i quasiconvex, is quasiconvex. The property extends to the general pointwise supremum

$$f(x) = \sup_{y \in C} (w(y)g(x, y))$$

where $w(y) \geq 0$ and $g(x, y)$ is quasiconvex in x for each y. This fact can be easily verified: $f(x) \leq \alpha$ if and only if

$$w(y)g(x, y) \leq \alpha \text{ for all } y \in C,$$

i.e., the α-sublevel set of f is the intersection of the α-sublevel sets of the functions $w(y)g(x, y)$ in the variable x.

Example 3.37 *Generalized eigenvalue.* The *maximum generalized eigenvalue* of a pair of symmetric matrices (X, Y), with $Y \succ 0$, is defined as

$$\lambda_{\max}(X, Y) = \sup_{u \neq 0} \frac{u^T X u}{u^T Y u} = \sup\{\lambda \mid \det(\lambda Y - X) = 0\}.$$

(See §A.5.3). This function is quasiconvex on $\mathbf{dom}\, f = \mathbf{S}^n \times \mathbf{S}^n_{++}$.

To see this we consider the expression

$$\lambda_{\max}(X, Y) = \sup_{u \neq 0} \frac{u^T X u}{u^T Y u}.$$

For each $u \neq 0$, the function $u^T X u / u^T Y u$ is linear-fractional in (X, Y), hence a quasiconvex function of (X, Y). We conclude that $\lambda_{\max}$ is quasiconvex, since it is the supremum of a family of quasiconvex functions.

Composition

If $g : \mathbf{R}^n \to \mathbf{R}$ is quasiconvex and $h : \mathbf{R} \to \mathbf{R}$ is nondecreasing, then $f = h \circ g$ is quasiconvex.

The composition of a quasiconvex function with an affine or linear-fractional transformation yields a quasiconvex function. If f is quasiconvex, then $g(x) = f(Ax + b)$ is quasiconvex, and $\tilde{g}(x) = f((Ax + b)/(c^T x + d))$ is quasiconvex on the set

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

Minimization

If $f(x, y)$ is quasiconvex jointly in x and y and C is a convex set, then the function

$$g(x) = \inf_{y \in C} f(x, y)$$

is quasiconvex.

To show this, we need to show that $\{x \mid g(x) \leq \alpha\}$ is convex, where $\alpha \in \mathbf{R}$ is arbitrary. From the definition of g, $g(x) \leq \alpha$ if and only if for any $\epsilon > 0$ there exists a $y \in C$ with $f(x, y) \leq \alpha + \epsilon$. Now let x_1 and x_2 be two points in the α-sublevel set of g. Then for any $\epsilon > 0$, there exists y_1, $y_2 \in C$ with

$$f(x_1, y_1) \leq \alpha + \epsilon, \qquad f(x_2, y_2) \leq \alpha + \epsilon,$$

and since f is quasiconvex in x and y, we also have

$$f(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2) \leq \alpha + \epsilon,$$

for $0 \leq \theta \leq 1$. Hence $g(\theta x_1 + (1-\theta)x_2) \leq \alpha$, which proves that $\{x \mid g(x) \leq \alpha\}$ is convex.

3.4.5 Representation via family of convex functions

In the sequel, it will be convenient to represent the sublevel sets of a quasiconvex function f (which are convex) via inequalities of convex functions. We seek a family of convex functions $\phi_t : \mathbf{R}^n \to \mathbf{R}$, indexed by $t \in \mathbf{R}$, with

$$f(x) \leq t \iff \phi_t(x) \leq 0, \tag{3.23}$$

i.e., the t-sublevel set of the quasiconvex function f is the 0-sublevel set of the convex function ϕ_t. Evidently ϕ_t must satisfy the property that for all $x \in \mathbf{R}^n$, $\phi_t(x) \leq 0 \implies \phi_s(x) \leq 0$ for $s \geq t$. This is satisfied if for each x, $\phi_t(x)$ is a nonincreasing function of t, *i.e.*, $\phi_s(x) \leq \phi_t(x)$ whenever $s \geq t$.

To see that such a representation always exists, we can take

$$\phi_t(x) = \begin{cases} 0 & f(x) \leq t \\ \infty & \text{otherwise,} \end{cases}$$

i.e., ϕ_t is the indicator function of the t-sublevel of f. Obviously this representation is not unique; for example if the sublevel sets of f are closed, we can take

$$\phi_t(x) = \mathbf{dist}\left(x, \{z \mid f(z) \leq t\}\right).$$

We are usually interested in a family ϕ_t with nice properties, such as differentiability.

Example 3.38 *Convex over concave function.* Suppose p is a convex function, q is a concave function, with $p(x) \geq 0$ and $q(x) > 0$ on a convex set C. Then the function f defined by $f(x) = p(x)/q(x)$, on C, is quasiconvex.

Here we have

$$f(x) \leq t \iff p(x) - tq(x) \leq 0,$$

so we can take $\phi_t(x) = p(x) - tq(x)$ for $t \geq 0$. For each t, ϕ_t is convex and for each x, $\phi_t(x)$ is decreasing in t.

3.5 Log-concave and log-convex functions

3.5.1 Definition

A function $f : \mathbf{R}^n \to \mathbf{R}$ is *logarithmically concave* or *log-concave* if $f(x) > 0$ for all $x \in \mathbf{dom}\, f$ and $\log f$ is concave. It is said to be *logarithmically convex* or *log-convex* if $\log f$ is convex. Thus f is log-convex if and only if $1/f$ is log-concave. It is convenient to allow f to take on the value zero, in which case we take $\log f(x) = -\infty$. In this case we say f is log-concave if the extended-value function $\log f$ is concave.

We can express log-concavity directly, without logarithms: a function $f : \mathbf{R}^n \to \mathbf{R}$, with convex domain and $f(x) > 0$ for all $x \in \mathbf{dom}\, f$, is log-concave if and only if for all $x,\ y \in \mathbf{dom}\, f$ and $0 \le \theta \le 1$, we have

$$f(\theta x + (1-\theta)y) \ge f(x)^\theta f(y)^{1-\theta}.$$

In particular, the value of a log-concave function at the average of two points is at least the *geometric mean* of the values at the two points.

From the composition rules we know that e^h is convex if h is convex, so a log-convex function is convex. Similarly, a nonnegative concave function is log-concave. It is also clear that a log-convex function is quasiconvex and a log-concave function is quasiconcave, since the logarithm is monotone increasing.

Example 3.39 *Some simple examples of log-concave and log-convex functions.*

- *Affine function.* $f(x) = a^T x + b$ is log-concave on $\{x \mid a^T x + b > 0\}$.
- *Powers.* $f(x) = x^a$, on $\mathbf{R}_{++}$, is log-convex for $a \le 0$, and log-concave for $a \ge 0$.
- *Exponentials.* $f(x) = e^{ax}$ is log-convex and log-concave.
- The cumulative distribution function of a Gaussian density,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du,$$

  is log-concave (see exercise 3.54).
- *Gamma function.* The Gamma function,

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u}\, du,$$

  is log-convex for $x \ge 1$ (see exercise 3.52).
- *Determinant.* $\det X$ is log concave on $\mathbf{S}^n_{++}$.
- *Determinant over trace.* $\det X / \mathbf{tr}\, X$ is log concave on $\mathbf{S}^n_{++}$ (see exercise 3.49).

Example 3.40 *Log-concave density functions.* Many common probability density functions are log-concave. Two examples are the multivariate normal distribution,

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-\bar{x})^T \Sigma^{-1} (x-\bar{x})}$$

(where $\bar{x} \in \mathbf{R}^n$ and $\Sigma \in \mathbf{S}^n_{++}$), and the exponential distribution on $\mathbf{R}^n_+$,

$$f(x) = \left(\prod_{i=1}^n \lambda_i \right) e^{-\lambda^T x}$$

(where $\lambda \succ 0$). Another example is the uniform distribution over a convex set C,

$$f(x) = \begin{cases} 1/\alpha & x \in C \\ 0 & x \notin C \end{cases}$$

where $\alpha = \mathbf{vol}(C)$ is the volume (Lebesgue measure) of C. In this case $\log f$ takes on the value $-\infty$ outside C, and $-\log \alpha$ on C, hence is concave.

As a more exotic example consider the Wishart distribution, defined as follows. Let $x_1, \ldots, x_p \in \mathbf{R}^n$ be independent Gaussian random vectors with zero mean and covariance $\Sigma \in \mathbf{S}^n$, with $p > n$. The random matrix $X = \sum_{i=1}^p x_i x_i^T$ has the Wishart density

$$f(X) = a \, (\det X)^{(p-n-1)/2} \, e^{-\frac{1}{2} \mathbf{tr}(\Sigma^{-1} X)},$$

with $\mathbf{dom}\, f = \mathbf{S}^n_{++}$, and a is a positive constant. The Wishart density is log-concave, since

$$\log f(X) = \log a + \frac{p-n-1}{2} \log\det X - \frac{1}{2} \mathbf{tr}(\Sigma^{-1} X),$$

which is a concave function of X.

3.5.2 Properties

Twice differentiable log-convex/concave functions

Suppose f is twice differentiable, with $\mathbf{dom}\, f$ convex, so

$$\nabla^2 \log f(x) = \frac{1}{f(x)} \nabla^2 f(x) - \frac{1}{f(x)^2} \nabla f(x) \nabla f(x)^T.$$

We conclude that f is log-convex if and only if for all $x \in \mathbf{dom}\, f$,

$$f(x) \nabla^2 f(x) \succeq \nabla f(x) \nabla f(x)^T,$$

and log-concave if and only if for all $x \in \mathbf{dom}\, f$,

$$f(x) \nabla^2 f(x) \preceq \nabla f(x) \nabla f(x)^T.$$

Multiplication, addition, and integration

Log-convexity and log-concavity are closed under multiplication and positive scaling. For example, if f and g are log-concave, then so is the pointwise product $h(x) = f(x)g(x)$, since $\log h(x) = \log f(x) + \log g(x)$, and $\log f(x)$ and $\log g(x)$ are concave functions of x.

Simple examples show that the sum of log-concave functions is not, in general, log-concave. Log-convexity, however, is preserved under sums. Let f and g be log-convex functions, *i.e.*, $F = \log f$ and $G = \log g$ are convex. From the composition rules for convex functions, it follows that

$$\log\left(\exp F + \exp G\right) = \log(f + g)$$

is convex. Therefore the sum of two log-convex functions is log-convex.

More generally, if $f(x,y)$ is log-convex in x for each $y \in C$ then

$$g(x) = \int_C f(x,y)\, dy$$

is log-convex.

Example 3.41 *Laplace transform of a nonnegative function and the moment and cumulant generating functions.* Suppose $p : \mathbf{R}^n \to \mathbf{R}$ satisfies $p(x) \geq 0$ for all x. The Laplace transform of p,

$$P(z) = \int p(x) e^{-z^T x}\, dx,$$

is log-convex on $\mathbf{R}^n$. (Here $\mathbf{dom}\, P$ is, naturally, $\{z \mid P(z) < \infty\}$.)

Now suppose p is a density, *i.e.*, satisfies $\int p(x)\, dx = 1$. The function $M(z) = P(-z)$ is called the *moment generating function* of the density. It gets its name from the fact that the moments of the density can be found from the derivatives of the moment generating function, evaluated at $z = 0$, *e.g.*,

$$\nabla M(0) = \mathbf{E}\, v, \qquad \nabla^2 M(0) = \mathbf{E}\, vv^T,$$

where v is a random variable with density p.

The function $\log M(z)$, which is convex, is called the *cumulant generating function* for p, since its derivatives give the cumulants of the density. For example, the first and second derivatives of the cumulant generating function, evaluated at zero, are the mean and covariance of the associated random variable:

$$\nabla \log M(0) = \mathbf{E}\, v, \qquad \nabla^2 \log M(0) = \mathbf{E}(v - \mathbf{E}\, v)(v - \mathbf{E}\, v)^T.$$

Integration of log-concave functions

In some special cases log-concavity is preserved by integration. If $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ is log-concave, then

$$g(x) = \int f(x,y)\, dy$$

is a log-concave function of x (on $\mathbf{R}^n$). (The integration here is over $\mathbf{R}^m$.) A proof of this result is not simple; see the references.

This result has many important consequences, some of which we describe in the rest of this section. It implies, for example, that marginal distributions of log-concave probability densities are log-concave. It also implies that log-concavity is closed under convolution, *i.e.*, if f and g are log-concave on $\mathbf{R}^n$, then so is the convolution

$$(f * g)(x) = \int f(x-y) g(y)\, dy.$$

(To see this, note that $g(y)$ and $f(x-y)$ are log-concave in (x,y), hence the product $f(x-y)g(y)$ is; then the integration result applies.)

Suppose $C \subseteq \mathbf{R}^n$ is a convex set and w is a random vector in $\mathbf{R}^n$ with log-concave probability density p. Then the function

$$f(x) = \mathbf{prob}(x + w \in C)$$

is log-concave in x. To see this, express f as

$$f(x) = \int g(x + w)p(w)\, dw,$$

where g is defined as

$$g(u) = \begin{cases} 1 & u \in C \\ 0 & u \notin C, \end{cases}$$

(which is log-concave) and apply the integration result.

Example 3.42 The *cumulative distribution function* of a probability density function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as

$$F(x) = \mathbf{prob}(w \preceq x) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(z)\, dz_1 \cdots dz_n,$$

where w is a random variable with density f. If f is log-concave, then F is log-concave. We have already encountered a special case: the cumulative distribution function of a Gaussian random variable,

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt,$$

is log-concave. (See example 3.39 and exercise 3.54.)

Example 3.43 *Yield function.* Let $x \in \mathbf{R}^n$ denote the nominal or target value of a set of parameters of a product that is manufactured. Variation in the manufacturing process causes the parameters of the product, when manufactured, to have the value $x + w$, where $w \in \mathbf{R}^n$ is a random vector that represents manufacturing variation, and is usually assumed to have zero mean. The *yield* of the manufacturing process, as a function of the nominal parameter values, is given by

$$Y(x) = \mathbf{prob}(x + w \in S),$$

where $S \subseteq \mathbf{R}^n$ denotes the set of acceptable parameter values for the product, *i.e.*, the product *specifications*.

If the density of the manufacturing error w is log-concave (for example, Gaussian) and the set S of product specifications is convex, then the yield function Y is log-concave. This implies that the α*-yield region*, defined as the set of nominal parameters for which the yield exceeds α, is convex. For example, the 95% yield region

$$\{x \mid Y(x) \geq 0.95\} = \{x \mid \log Y(x) \geq \log 0.95\}$$

is convex, since it is a superlevel set of the concave function $\log Y$.

Example 3.44 *Volume of polyhedron.* Let $A \in \mathbf{R}^{m\times n}$. Define

$$P_u = \{x \in \mathbf{R}^n \mid Ax \preceq u\}.$$

Then its volume $\mathbf{vol}\, P_u$ is a log-concave function of u.

To prove this, note that the function

$$\Psi(x,u) = \begin{cases} 1 & Ax \preceq u \\ 0 & \text{otherwise,} \end{cases}$$

is log-concave. By the integration result, we conclude that

$$\int \Psi(x,u)\, dx = \mathbf{vol}\, P_u$$

is log-concave.

3.6 Convexity with respect to generalized inequalities

We now consider generalizations of the notions of monotonicity and convexity, using generalized inequalities instead of the usual ordering on $\mathbf{R}$.

3.6.1 Monotonicity with respect to a generalized inequality

Suppose $K \subseteq \mathbf{R}^n$ is a proper cone with associated generalized inequality $\preceq_K$. A function $f : \mathbf{R}^n \to \mathbf{R}$ is called K-*nondecreasing* if

$$x \preceq_K y \Longrightarrow f(x) \le f(y),$$

and K-*increasing* if

$$x \preceq_K y,\ x \ne y \Longrightarrow f(x) < f(y).$$

We define K-*nonincreasing* and K-*decreasing* functions in a similar way.

Example 3.45 *Monotone vector functions.* A function $f : \mathbf{R}^n \to \mathbf{R}$ is nondecreasing with respect to $\mathbf{R}^n_+$ if and only if

$$x_1 \le y_1, \ldots, x_n \le y_n \implies f(x) \le f(y)$$

for all x, y. This is the same as saying that f, when restricted to any component x_i (*i.e.*, x_i is considered the variable while x_j for $j \ne i$ are fixed), is nondecreasing.

Example 3.46 *Matrix monotone functions.* A function $f : \mathbf{S}^n \to \mathbf{R}$ is called *matrix monotone* (increasing, decreasing) if it is monotone with respect to the positive semidefinite cone. Some examples of matrix monotone functions of the variable $X \in \mathbf{S}^n$:

- $\mathbf{tr}(WX)$, where $W \in \mathbf{S}^n$, is matrix nondecreasing if $W \succeq 0$, and matrix increasing if $W \succ 0$ (it is matrix nonincreasing if $W \preceq 0$, and matrix decreasing if $W \prec 0$).
- $\mathbf{tr}(X^{-1})$ is matrix decreasing on $\mathbf{S}^n_{++}$.
- $\det X$ is matrix increasing on $\mathbf{S}^n_{++}$, and matrix nondecreasing on $\mathbf{S}^n_{+}$.

Gradient conditions for monotonicity

Recall that a differentiable function $f : \mathbf{R} \to \mathbf{R}$, with convex (*i.e.*, interval) domain, is nondecreasing if and only if $f'(x) \geq 0$ for all $x \in \mathbf{dom}\, f$, and increasing if $f'(x) > 0$ for all $x \in \mathbf{dom}\, f$ (but the converse is not true). These conditions are readily extended to the case of monotonicity with respect to a generalized inequality. A differentiable function f, with convex domain, is K-nondecreasing if and only if

$$\nabla f(x) \succeq_{K^*} 0 \tag{3.24}$$

for all $x \in \mathbf{dom}\, f$. Note the difference with the simple scalar case: the gradient must be nonnegative in the *dual* inequality. For the strict case, we have the following: If

$$\nabla f(x) \succ_{K^*} 0 \tag{3.25}$$

for all $x \in \mathbf{dom}\, f$, then f is K-increasing. As in the scalar case, the converse is not true.

Let us prove these first-order conditions for monotonicity. First, assume that f satisfies (3.24) for all x, but is not K-nondecreasing, *i.e.*, there exist x, y with $x \preceq_K y$ and $f(y) < f(x)$. By differentiability of f there exists a $t \in [0, 1]$ with

$$\frac{d}{dt} f(x + t(y - x)) = \nabla f(x + t(y - x))^T (y - x) < 0.$$

Since $y - x \in K$ this means

$$\nabla f(x + t(y - x)) \notin K^*,$$

which contradicts our assumption that (3.24) is satisfied everywhere. In a similar way it can be shown that (3.25) implies f is K-increasing.

It is also straightforward to see that it is necessary that (3.24) hold everywhere. Assume (3.24) does not hold for $x = z$. By the definition of dual cone this means there exists a $v \in K$ with

$$\nabla f(z)^T v < 0.$$

Now consider $h(t) = f(z + tv)$ as a function of t. We have $h'(0) = \nabla f(z)^T v < 0$, and therefore there exists $t > 0$ with $h(t) = f(z + tv) < h(0) = f(z)$, which means f is not K-nondecreasing.

3.6.2 Convexity with respect to a generalized inequality

Suppose $K \subseteq \mathbf{R}^m$ is a proper cone with associated generalized inequality $\preceq_K$. We say $f : \mathbf{R}^n \to \mathbf{R}^m$ is *K-convex* if for all x, y, and $0 \leq \theta \leq 1$,

$$f(\theta x + (1 - \theta)y) \ \preceq_K \ \theta f(x) + (1 - \theta) f(y).$$

The function is *strictly K-convex* if

$$f(\theta x + (1-\theta)y) \prec_K \theta f(x) + (1-\theta)f(y)$$

for all $x \neq y$ and $0 < \theta < 1$. These definitions reduce to ordinary convexity and strict convexity when $m = 1$ (and $K = \mathbf{R}_+$).

Example 3.47 *Convexity with respect to componentwise inequality.* A function $f : \mathbf{R}^n \to \mathbf{R}^m$ is convex with respect to componentwise inequality (*i.e.*, the generalized inequality induced by $\mathbf{R}^m_+$) if and only if for all x, y and $0 \leq \theta \leq 1$,

$$f(\theta x + (1-\theta)y) \preceq \theta f(x) + (1-\theta)f(y),$$

i.e., each component f_i is a convex function. The function f is strictly convex with respect to componentwise inequality if and only if each component f_i is strictly convex.

Example 3.48 *Matrix convexity.* Suppose f is a symmetric matrix valued function, *i.e.*, $f : \mathbf{R}^n \to \mathbf{S}^m$. The function f is convex with respect to matrix inequality if

$$f(\theta x + (1-\theta)y) \preceq \theta f(x) + (1-\theta)f(y)$$

for any x and y, and for $\theta \in [0, 1]$. This is sometimes called *matrix convexity.* An equivalent definition is that the scalar function $z^T f(x) z$ is convex for all vectors z. (This is often a good way to prove matrix convexity). A matrix function is strictly matrix convex if

$$f(\theta x + (1-\theta)y) \prec \theta f(x) + (1-\theta)f(y)$$

when $x \neq y$ and $0 < \theta < 1$, or, equivalently, if $z^T f z$ is strictly convex for every $z \neq 0$.

Some examples:

- The function $f(X) = XX^T$ where $X \in \mathbf{R}^{n\times m}$ is matrix convex, since for fixed z the function $z^T XX^T z = \|X^T z\|_2^2$ is a convex quadratic function of (the components of) X. For the same reason, $f(X) = X^2$ is matrix convex on $\mathbf{S}^n$.
- The function X^p is matrix convex on $\mathbf{S}^n_{++}$ for $1 \leq p \leq 2$ or $-1 \leq p \leq 0$, and matrix concave for $0 \leq p \leq 1$.
- The function $f(X) = e^X$ is *not* matrix convex on $\mathbf{S}^n$, for $n \geq 2$.

Many of the results for convex functions have extensions to K-convex functions. As a simple example, a function is K-convex if and only if its restriction to any line in its domain is K-convex. In the rest of this section we list a few results for K-convexity that we will use later; more results are explored in the exercises.

Dual characterization of K-convexity

A function f is K-convex if and only if for every $w \succeq_{K^*} 0$, the (real-valued) function $w^T f$ is convex (in the ordinary sense); f is strictly K-convex if and only if for every nonzero $w \succeq_{K^*} 0$ the function $w^T f$ is strictly convex. (These follow directly from the definitions and properties of dual inequality.)

Differentiable K-convex functions

A differentiable function f is K-convex if and only if its domain is convex, and for all $x,\ y \in \mathbf{dom}\, f$,

$$f(y) \succeq_K f(x) + Df(x)(y - x).$$

(Here $Df(x) \in \mathbf{R}^{m\times n}$ is the derivative or Jacobian matrix of f at x; see §A.4.1.) The function f is strictly K-convex if and only if for all $x,\ y \in \mathbf{dom}\, f$ with $x \neq y$,

$$f(y) \succ_K f(x) + Df(x)(y - x).$$

Composition theorem

Many of the results on composition can be generalized to K-convexity. For example, if $g : \mathbf{R}^n \to \mathbf{R}^p$ is K-convex, $h : \mathbf{R}^p \to \mathbf{R}$ is convex, and $\tilde{h}$ (the extended-value extension of h) is K-nondecreasing, then $h \circ g$ is convex. This generalizes the fact that a nondecreasing convex function of a convex function is convex. The condition that $\tilde{h}$ be K-nondecreasing implies that $\mathbf{dom}\, h - K = \mathbf{dom}\, h$.

Example 3.49 The quadratic matrix function $g : \mathbf{R}^{m\times n} \to \mathbf{S}^n$ defined by

$$g(X) = X^T AX + B^T X + X^T B + C,$$

where $A \in \mathbf{S}^m$, $B \in \mathbf{R}^{m\times n}$, and $C \in \mathbf{S}^n$, is convex when $A \succeq 0$.

The function $h : \mathbf{S}^n \to \mathbf{R}$ defined by $h(Y) = -\log\det(-Y)$ is convex and increasing on $\mathbf{dom}\, h = -\mathbf{S}^n_{++}$.

By the composition theorem, we conclude that

$$f(X) = -\log\det(-(X^T AX + B^T X + X^T B + C))$$

is convex on

$$\mathbf{dom}\, f = \{X \in \mathbf{R}^{m\times n} \mid X^T AX + B^T X + X^T B + C \prec 0\}.$$

This generalizes the fact that

$$-\log(-(ax^2 + bx + c))$$

is convex on

$$\{x \in \mathbf{R} \mid ax^2 + bx + c < 0\},$$

provided $a \geq 0$.

Bibliography

The standard reference on convex analysis is Rockafellar [Roc70]. Other books on convex functions are Stoer and Witzgall [SW70], Roberts and Varberg [RV73], Van Tiel [vT84], Hiriart-Urruty and Lemaréchal [HUL93], Ekeland and Témam [ET99], Borwein and Lewis [BL00], Florenzano and Le Van [FL01], Barvinok [Bar02], and Bertsekas, Nedić, and Ozdaglar [Ber03]. Most nonlinear programming texts also include chapters on convex functions (see, for example, Mangasarian [Man94], Bazaraa, Sherali, and Shetty [BSS93], Bertsekas [Ber99], Polyak [Pol87], and Peressini, Sullivan, and Uhl [PSU88]).

Jensen's inequality appears in [Jen06]. A general study of inequalities, in which Jensen's inequality plays a central role, is presented by Hardy, Littlewood, and Pólya [HLP52], and Beckenbach and Bellman [BB65].

The term *perspective function* is from Hiriart-Urruty and Lemaréchal [HUL93, volume 1, page 100]. For the definitions in example 3.19 (relative entropy and Kullback-Leibler divergence), and the related exercise 3.13, see Cover and Thomas [CT91].

Some important early references on quasiconvex functions (as well as other extensions of convexity) are Nikaidô [Nik54], Mangasarian [Man94, chapter 9], Arrow and Enthoven [AE61], Ponstein [Pon67], and Luenberger [Lue68]. For a more comprehensive reference list, we refer to Bazaraa, Sherali, and Shetty [BSS93, page 126].

Prékopa [Pré80] gives a survey of log-concave functions. Log-convexity of the Laplace transform is mentioned in Barndorff-Nielsen [BN78, §7]. For a proof of the integration result of log-concave functions, see Prékopa [Pré71, Pré73].

Generalized inequalities are used extensively in the recent literature on cone programming, starting with Nesterov and Nemirovski [NN94, page 156]; see also Ben-Tal and Nemirovski [BTN01] and the references at the end of chapter 4. Convexity with respect to generalized inequalities also appears in the work of Luenberger [Lue69, §8.2] and Isii [Isi64]. Matrix monotonicity and matrix convexity are attributed to Löwner [Löw34], and are discussed in detail by Davis [Dav63], Roberts and Varberg [RV73, page 216] and Marshall and Olkin [MO79, §16E]. For the result on convexity and concavity of the function X^p in example 3.48, see Bondar [Bon94, theorem 16.1]. For a simple example that demonstrates that e^X is not matrix convex, see Marshall and Olkin [MO79, page 474].

Exercises

Definition of convexity

3.1 Suppose $f : \mathbf{R} \to \mathbf{R}$ is convex, and $a,\ b \in \mathbf{dom}\, f$ with $a < b$.

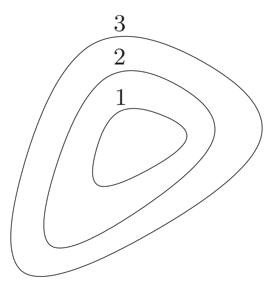
(a) Show that
$$f(x) \leq \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b)$$
for all $x \in [a, b]$.

(b) Show that
$$\frac{f(x) - f(a)}{x - a} \leq \frac{f(b) - f(a)}{b - a} \leq \frac{f(b) - f(x)}{b - x}$$
for all $x \in (a, b)$. Draw a sketch that illustrates this inequality.

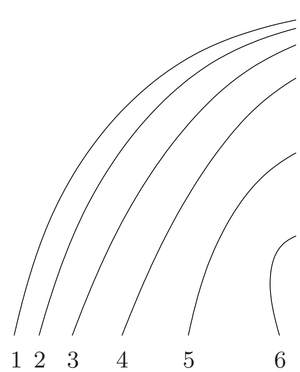
(c) Suppose f is differentiable. Use the result in (b) to show that
$$f'(a) \leq \frac{f(b) - f(a)}{b - a} \leq f'(b).$$
Note that these inequalities also follow from (3.2):
$$f(b) \geq f(a) + f'(a)(b - a), \qquad f(a) \geq f(b) + f'(b)(a - b).$$

(d) Suppose f is twice differentiable. Use the result in (c) to show that $f''(a) \geq 0$ and $f''(b) \geq 0$.

3.2 *Level sets of convex, concave, quasiconvex, and quasiconcave functions.* Some level sets of a function f are shown below. The curve labeled 1 shows $\{x \mid f(x) = 1\}$, etc.



Could f be convex (concave, quasiconvex, quasiconcave)? Explain your answer. Repeat for the level curves shown below.



3.3 *Inverse of an increasing convex function.* Suppose $f : \mathbf{R} \to \mathbf{R}$ is increasing and convex on its domain (a, b). Let g denote its inverse, *i.e.*, the function with domain $(f(a), f(b))$ and $g(f(x)) = x$ for $a < x < b$. What can you say about convexity or concavity of g?

3.4 [RV73, page 15] Show that a continuous function $f : \mathbf{R}^n \to \mathbf{R}$ is convex if and only if for every line segment, its average value on the segment is less than or equal to the average of its values at the endpoints of the segment: For every $x,\ y \in \mathbf{R}^n$,

$$\int_0^1 f(x + \lambda(y - x))\, d\lambda \leq \frac{f(x) + f(y)}{2}.$$

3.5 [RV73, page 22] *Running average of a convex function.* Suppose $f : \mathbf{R} \to \mathbf{R}$ is convex, with $\mathbf{R}_+ \subseteq \mathbf{dom}\, f$. Show that its *running average* F, defined as

$$F(x) = \frac{1}{x} \int_0^x f(t)\, dt, \qquad \mathbf{dom}\, F = \mathbf{R}_{++},$$

is convex. *Hint.* For each s, $f(sx)$ is convex in x, so $\int_0^1 f(sx)\, ds$ is convex.

3.6 *Functions and epigraphs.* When is the epigraph of a function a halfspace? When is the epigraph of a function a convex cone? When is the epigraph of a function a polyhedron?

3.7 Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is convex with $\mathbf{dom}\, f = \mathbf{R}^n$, and bounded above on $\mathbf{R}^n$. Show that f is constant.

3.8 *Second-order condition for convexity.* Prove that a twice differentiable function f is convex if and only if its domain is convex and $\nabla^2 f(x) \succeq 0$ for all $x \in \mathbf{dom}\, f$. *Hint.* First consider the case $f : \mathbf{R} \to \mathbf{R}$. You can use the first-order condition for convexity (which was proved on page 70).

3.9 *Second-order conditions for convexity on an affine set.* Let $F \in \mathbf{R}^{n \times m}$, $\hat{x} \in \mathbf{R}^n$. The *restriction* of $f : \mathbf{R}^n \to \mathbf{R}$ to the affine set $\{Fz + \hat{x} \mid z \in \mathbf{R}^m\}$ is defined as the function $\tilde{f} : \mathbf{R}^m \to \mathbf{R}$ with

$$\tilde{f}(z) = f(Fz + \hat{x}), \qquad \mathbf{dom}\, \tilde{f} = \{z \mid Fz + \hat{x} \in \mathbf{dom}\, f\}.$$

Suppose f is twice differentiable with a convex domain.

(a) Show that $\tilde{f}$ is convex if and only if for all $z \in \mathbf{dom}\, \tilde{f}$

$$F^T \nabla^2 f(Fz + \hat{x}) F \succeq 0.$$

(b) Suppose $A \in \mathbf{R}^{p \times n}$ is a matrix whose nullspace is equal to the range of F, *i.e.*, $AF = 0$ and $\mathbf{rank}\, A = n - \mathbf{rank}\, F$. Show that $\tilde{f}$ is convex if for all $z \in \mathbf{dom}\, \tilde{f}$ there exists a $\lambda \in \mathbf{R}$ such that

$$\nabla^2 f(Fz + \hat{x}) + \lambda A^T A \succeq 0.$$

Hint. Use the following result: If $B \in \mathbf{S}^n$ and $A \in \mathbf{R}^{p \times n}$, then $x^T B x \geq 0$ for all $x \in \mathcal{N}(A)$ if there exists a λ such that $B + \lambda A^T A \succeq 0$.

3.10 *An extension of Jensen's inequality.* One interpretation of Jensen's inequality is that randomization or dithering hurts, *i.e.*, raises the average value of a convex function: For f convex and v a zero mean random variable, we have $\mathbf{E}\, f(x_0 + v) \geq f(x_0)$. This leads to the following conjecture. If f is convex, then the larger the variance of v, the larger $\mathbf{E}\, f(x_0 + v)$.

(a) Give a counterexample that shows that this conjecture is false. Find zero mean random variables v and w, with $\mathbf{var}(v) > \mathbf{var}(w)$, a convex function f, and a point x_0, such that $\mathbf{E}\, f(x_0 + v) < \mathbf{E}\, f(x_0 + w)$.

(b) The conjecture is true when v and w are scaled versions of each other. Show that $\mathbf{E}\, f(x_0 + tv)$ is monotone increasing in $t \geq 0$, when f is convex and v is zero mean.

3.11 *Monotone mappings.* A function $\psi : \mathbf{R}^n \to \mathbf{R}^n$ is called *monotone* if for all $x,\ y \in \mathbf{dom}\, \psi$,

$$(\psi(x) - \psi(y))^T (x - y) \geq 0.$$

(Note that 'monotone' as defined here is not the same as the definition given in §3.6.1. Both definitions are widely used.) Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is a differentiable convex function. Show that its gradient ∇f is monotone. Is the converse true, *i.e.*, is every monotone mapping the gradient of a convex function?

3.12 Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is convex, $g : \mathbf{R}^n \to \mathbf{R}$ is concave, $\mathbf{dom}\, f = \mathbf{dom}\, g = \mathbf{R}^n$, and for all x, $g(x) \leq f(x)$. Show that there exists an affine function h such that for all x, $g(x) \leq h(x) \leq f(x)$. In other words, if a concave function g is an underestimator of a convex function f, then we can fit an affine function between f and g.

3.13 *Kullback-Leibler divergence and the information inequality.* Let D_{kl} be the Kullback-Leibler divergence, as defined in (3.17). Prove the *information inequality*: $D_{\mathrm{kl}}(u, v) \geq 0$ for all $u,\ v \in \mathbf{R}^n_{++}$. Also show that $D_{\mathrm{kl}}(u, v) = 0$ if and only if $u = v$.

Hint. The Kullback-Leibler divergence can be expressed as

$$D_{\mathrm{kl}}(u, v) = f(u) - f(v) - \nabla f(v)^T (u - v),$$

where $f(v) = \sum_{i=1}^n v_i \log v_i$ is the negative entropy of v.

3.14 *Convex-concave functions and saddle-points.* We say the function $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ is *convex-concave* if $f(x, z)$ is a concave function of z, for each fixed x, and a convex function of x, for each fixed z. We also require its domain to have the product form $\mathbf{dom}\, f = A \times B$, where $A \subseteq \mathbf{R}^n$ and $B \subseteq \mathbf{R}^m$ are convex.

(a) Give a second-order condition for a twice differentiable function $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ to be convex-concave, in terms of its Hessian $\nabla^2 f(x, z)$.

(b) Suppose that $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ is convex-concave and differentiable, with $\nabla f(\tilde{x}, \tilde{z}) = 0$. Show that the *saddle-point property* holds: for all x, z, we have

$$f(\tilde{x}, z) \leq f(\tilde{x}, \tilde{z}) \leq f(x, \tilde{z}).$$

Show that this implies that f satisfies the *strong max-min property*:

$$\sup_z \inf_x f(x, z) = \inf_x \sup_z f(x, z)$$

(and their common value is $f(\tilde{x}, \tilde{z})$).

(c) Now suppose that $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ is differentiable, but not necessarily convex-concave, and the saddle-point property holds at $\tilde{x}, \tilde{z}$:

$$f(\tilde{x}, z) \leq f(\tilde{x}, \tilde{z}) \leq f(x, \tilde{z})$$

for all x, z. Show that $\nabla f(\tilde{x}, \tilde{z}) = 0$.

Examples

3.15 *A family of concave utility functions.* For $0 < \alpha \leq 1$ let

$$u_\alpha(x) = \frac{x^\alpha - 1}{\alpha},$$

with $\mathbf{dom}\, u_\alpha = \mathbf{R}_+$. We also define $u_0(x) = \log x$ (with $\mathbf{dom}\, u_0 = \mathbf{R}_{++}$).

(a) Show that for $x > 0$, $u_0(x) = \lim_{\alpha \to 0} u_\alpha(x)$.

(b) Show that u_α are concave, monotone increasing, and all satisfy $u_\alpha(1) = 0$.

These functions are often used in economics to model the benefit or utility of some quantity of goods or money. Concavity of u_α means that the marginal utility (*i.e.*, the increase in utility obtained for a fixed increase in the goods) decreases as the amount of goods increases. In other words, concavity models the effect of *satiation.*

3.16 For each of the following functions determine whether it is convex, concave, quasiconvex, or quasiconcave.

(a) $f(x) = e^x - 1$ on $\mathbf{R}$.

(b) $f(x_1, x_2) = x_1 x_2$ on $\mathbf{R}^2_{++}$.

(c) $f(x_1, x_2) = 1/(x_1 x_2)$ on $\mathbf{R}^2_{++}$.

(d) $f(x_1, x_2) = x_1/x_2$ on $\mathbf{R}^2_{++}$.

(e) $f(x_1, x_2) = x_1^2/x_2$ on $\mathbf{R} \times \mathbf{R}_{++}$.

(f) $f(x_1, x_2) = x_1^\alpha x_2^{1-\alpha}$, where $0 \le \alpha \le 1$, on $\mathbf{R}^2_{++}$.

3.17 Suppose $p < 1$, $p \ne 0$. Show that the function

$$f(x) = \left(\sum_{i=1}^n x_i^p \right)^{1/p}$$

with $\mathbf{dom}\, f = \mathbf{R}^n_{++}$ is concave. This includes as special cases $f(x) = (\sum_{i=1}^n x_i^{1/2})^2$ and the *harmonic mean* $f(x) = (\sum_{i=1}^n 1/x_i)^{-1}$. *Hint.* Adapt the proofs for the log-sum-exp function and the geometric mean in §3.1.5.

3.18 Adapt the proof of concavity of the log-determinant function in §3.1.5 to show the following.

(a) $f(X) = \mathbf{tr}\left(X^{-1}\right)$ is convex on $\mathbf{dom}\, f = \mathbf{S}^n_{++}$.

(b) $f(X) = (\det X)^{1/n}$ is concave on $\mathbf{dom}\, f = \mathbf{S}^n_{++}$.

3.19 *Nonnegative weighted sums and integrals.*

(a) Show that $f(x) = \sum_{i=1}^r \alpha_i x_{[i]}$ is a convex function of x, where $\alpha_1 \ge \alpha_2 \ge \cdots \ge \alpha_r \ge 0$, and $x_{[i]}$ denotes the ith largest component of x. (You can use the fact that $f(x) = \sum_{i=1}^k x_{[i]}$ is convex on $\mathbf{R}^n$.)

(b) Let $T(x, \omega)$ denote the trigonometric polynomial

$$T(x, \omega) = x_1 + x_2 \cos \omega + x_3 \cos 2\omega + \cdots + x_n \cos(n-1)\omega.$$

Show that the function

$$f(x) = -\int_0^{2\pi} \log T(x, \omega)\, d\omega$$

is convex on $\{x \in \mathbf{R}^n \mid T(x, \omega) > 0,\ 0 \le \omega \le 2\pi\}$.

3.20 *Composition with an affine function.* Show that the following functions $f : \mathbf{R}^n \to \mathbf{R}$ are convex.

(a) $f(x) = \|Ax - b\|$, where $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, and $\|\cdot\|$ is a norm on $\mathbf{R}^m$.

(b) $f(x) = -\left(\det(A_0 + x_1 A_1 + \cdots + x_n A_n)\right)^{1/m}$, on $\{x \mid A_0 + x_1 A_1 + \cdots + x_n A_n \succ 0\}$, where $A_i \in \mathbf{S}^m$.

(c) $f(X) = \mathbf{tr}\left(A_0 + x_1 A_1 + \cdots + x_n A_n\right)^{-1}$, on $\{x \mid A_0 + x_1 A_1 + \cdots + x_n A_n \succ 0\}$, where $A_i \in \mathbf{S}^m$. (Use the fact that $\mathbf{tr}(X^{-1})$ is convex on $\mathbf{S}^m_{++}$; see exercise 3.18.)

3.21 *Pointwise maximum and supremum.* Show that the following functions $f : \mathbf{R}^n \to \mathbf{R}$ are convex.

(a) $f(x) = \max_{i=1,\ldots,k} \|A^{(i)}x - b^{(i)}\|$, where $A^{(i)} \in \mathbf{R}^{m\times n}$, $b^{(i)} \in \mathbf{R}^m$ and $\|\cdot\|$ is a norm on $\mathbf{R}^m$.

(b) $f(x) = \sum_{i=1}^{r} |x|_{[i]}$ on $\mathbf{R}^n$, where $|x|$ denotes the vector with $|x|_i = |x_i|$ (*i.e.*, $|x|$ is the absolute value of x, componentwise), and $|x|_{[i]}$ is the ith largest component of $|x|$. In other words, $|x|_{[1]}, |x|_{[2]}, \ldots, |x|_{[n]}$ are the absolute values of the components of x, sorted in nonincreasing order.

3.22 *Composition rules.* Show that the following functions are convex.

(a) $f(x) = -\log(-\log(\sum_{i=1}^{m} e^{a_i^Tx+b_i}))$ on $\mathbf{dom}\, f = \{x \mid \sum_{i=1}^{m} e^{a_i^Tx+b_i} < 1\}$. You can use the fact that $\log(\sum_{i=1}^{n} e^{y_i})$ is convex.

(b) $f(x,u,v) = -\sqrt{uv - x^Tx}$ on $\mathbf{dom}\, f = \{(x,u,v) \mid uv > x^Tx,\ u,\ v > 0\}$. Use the fact that x^Tx/u is convex in (x,u) for $u > 0$, and that $-\sqrt{x_1x_2}$ is convex on $\mathbf{R}^2_{++}$.

(c) $f(x,u,v) = -\log(uv - x^Tx)$ on $\mathbf{dom}\, f = \{(x,u,v) \mid uv > x^Tx,\ u,\ v > 0\}$.

(d) $f(x,t) = -(t^p - \|x\|_p^p)^{1/p}$ where $p > 1$ and $\mathbf{dom}\, f = \{(x,t) \mid t \geq \|x\|_p\}$. You can use the fact that $\|x\|_p^p/u^{p-1}$ is convex in (x,u) for $u > 0$ (see exercise 3.23), and that $-x^{1/p}y^{1-1/p}$ is convex on $\mathbf{R}^2_+$ (see exercise 3.16).

(e) $f(x,t) = -\log(t^p - \|x\|_p^p)$ where $p > 1$ and $\mathbf{dom}\, f = \{(x,t) \mid t > \|x\|_p\}$. You can use the fact that $\|x\|_p^p/u^{p-1}$ is convex in (x,u) for $u > 0$ (see exercise 3.23).

3.23 *Perspective of a function.*

(a) Show that for $p > 1$,

$$f(x,t) = \frac{|x_1|^p + \cdots + |x_n|^p}{t^{p-1}} = \frac{\|x\|_p^p}{t^{p-1}}$$

is convex on $\{(x,t) \mid t > 0\}$.

(b) Show that

$$f(x) = \frac{\|Ax+b\|_2^2}{c^Tx+d}$$

is convex on $\{x \mid c^Tx + d > 0\}$, where $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$, $c \in \mathbf{R}^n$ and $d \in \mathbf{R}$.

3.24 *Some functions on the probability simplex.* Let x be a real-valued random variable which takes values in $\{a_1, \ldots, a_n\}$ where $a_1 < a_2 < \cdots < a_n$, with $\mathbf{prob}(x = a_i) = p_i$, $i = 1, \ldots, n$. For each of the following functions of p (on the probability simplex $\{p \in \mathbf{R}^n_+ \mid \mathbf{1}^Tp = 1\}$), determine if the function is convex, concave, quasiconvex, or quasiconcave.

(a) $\mathbf{E}\, x$.

(b) $\mathbf{prob}(x \geq \alpha)$.

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

(d) $\sum_{i=1}^{n} p_i \log p_i$, the negative entropy of the distribution.

(e) $\mathbf{var}\, x = \mathbf{E}(x - \mathbf{E}\, x)^2$.

(f) $\mathbf{quartile}(x) = \inf\{\beta \mid \mathbf{prob}(x \leq \beta) \geq 0.25\}$.

(g) The cardinality of the smallest set $\mathcal{A} \subseteq \{a_1, \ldots, a_n\}$ with probability $\geq 90\%$. (By cardinality we mean the number of elements in $\mathcal{A}$.)

(h) The minimum width interval that contains 90% of the probability, *i.e.*,

$$\inf\{\beta - \alpha \mid \mathbf{prob}(\alpha \leq x \leq \beta) \geq 0.9\}.$$

3.25 *Maximum probability distance between distributions.* Let $p,\ q \in \mathbf{R}^n$ represent two probability distributions on $\{1,\dots,n\}$ (so $p,\ q \succeq 0$, $\mathbf{1}^T p = \mathbf{1}^T q = 1$). We define the *maximum probability distance* $d_{\mathrm{mp}}(p,q)$ between p and q as the maximum difference in probability assigned by p and q, over all events:

$$d_{\mathrm{mp}}(p,q) = \max\{|\,\mathbf{prob}(p,C) - \mathbf{prob}(q,C)| \mid C \subseteq \{1,\dots,n\}\}.$$

Here $\mathbf{prob}(p,C)$ is the probability of C, under the distribution p, *i.e.*, $\mathbf{prob}(p,C) = \sum_{i\in C} p_i$.

Find a simple expression for d_{mp}, involving $\|p-q\|_1 = \sum_{i=1}^n |p_i - q_i|$, and show that d_{mp} is a convex function on $\mathbf{R}^n \times \mathbf{R}^n$. (Its domain is $\{(p,q) \mid p,\ q \succeq 0,\ \mathbf{1}^T p = \mathbf{1}^T q = 1\}$, but it has a natural extension to all of $\mathbf{R}^n \times \mathbf{R}^n$.)

3.26 *More functions of eigenvalues.* Let $\lambda_1(X) \geq \lambda_2(X) \geq \cdots \geq \lambda_n(X)$ denote the eigenvalues of a matrix $X \in \mathbf{S}^n$. We have already seen several functions of the eigenvalues that are convex or concave functions of X.

- The maximum eigenvalue $\lambda_1(X)$ is convex (example 3.10). The minimum eigenvalue $\lambda_n(X)$ is concave.
- The sum of the eigenvalues (or trace), $\mathbf{tr}\, X = \lambda_1(X) + \cdots + \lambda_n(X)$, is linear.
- The sum of the inverses of the eigenvalues (or trace of the inverse), $\mathbf{tr}(X^{-1}) = \sum_{i=1}^n 1/\lambda_i(X)$, is convex on $\mathbf{S}^n_{++}$ (exercise 3.18).
- The geometric mean of the eigenvalues, $(\det X)^{1/n} = (\prod_{i=1}^n \lambda_i(X))^{1/n}$, and the logarithm of the product of the eigenvalues, $\log\det X = \sum_{i=1}^n \log\lambda_i(X)$, are concave on $X \in \mathbf{S}^n_{++}$ (exercise 3.18 and page 74).

In this problem we explore some more functions of eigenvalues, by exploiting variational characterizations.

(a) *Sum of k largest eigenvalues.* Show that $\sum_{i=1}^k \lambda_i(X)$ is convex on $\mathbf{S}^n$. *Hint.* [HJ85, page 191] Use the variational characterization

$$\sum_{i=1}^k \lambda_i(X) = \sup\{\mathbf{tr}(V^T X V) \mid V \in \mathbf{R}^{n\times k},\ V^T V = I\}.$$

(b) *Geometric mean of k smallest eigenvalues.* Show that $(\prod_{i=n-k+1}^n \lambda_i(X))^{1/k}$ is concave on $\mathbf{S}^n_{++}$. *Hint.* [MO79, page 513] For $X \succ 0$, we have

$$\left(\prod_{i=n-k+1}^n \lambda_i(X)\right)^{1/k} = \frac{1}{k}\inf\{\mathbf{tr}(V^T X V) \mid V \in \mathbf{R}^{n\times k},\ \det V^T V = 1\}.$$

(c) *Log of product of k smallest eigenvalues.* Show that $\sum_{i=n-k+1}^n \log\lambda_i(X)$ is concave on $\mathbf{S}^n_{++}$. *Hint.* [MO79, page 513] For $X \succ 0$,

$$\prod_{i=n-k+1}^n \lambda_i(X) = \inf\left\{\prod_{i=1}^k (V^T X V)_{ii} \;\middle|\; V \in \mathbf{R}^{n\times k},\ V^T V = I\right\}.$$

3.27 *Diagonal elements of Cholesky factor.* Each $X \in \mathbf{S}^n_{++}$ has a unique *Cholesky factorization* $X = LL^T$, where L is lower triangular, with $L_{ii} > 0$. Show that L_{ii} is a concave function of X (with domain $\mathbf{S}^n_{++}$).

Hint. L_{ii} can be expressed as $L_{ii} = (w - z^T Y^{-1} z)^{1/2}$, where

$$\begin{bmatrix} Y & z \\ z^T & w \end{bmatrix}$$

is the leading $i \times i$ submatrix of X.

Operations that preserve convexity

3.28 *Expressing a convex function as the pointwise supremum of a family of affine functions.* In this problem we extend the result proved on page 83 to the case where $\mathbf{dom}\, f \neq \mathbf{R}^n$. Let $f : \mathbf{R}^n \to \mathbf{R}$ be a convex function. Define $\tilde{f} : \mathbf{R}^n \to \mathbf{R}$ as the pointwise supremum of all affine functions that are global underestimators of f:

$$\tilde{f}(x) = \sup\{g(x) \mid g \text{ affine}, \ g(z) \leq f(z) \text{ for all } z\}.$$

(a) Show that $f(x) = \tilde{f}(x)$ for $x \in \mathbf{int}\,\mathbf{dom}\, f$.

(b) Show that $f = \tilde{f}$ if f is closed (*i.e.*, $\mathbf{epi}\, f$ is a closed set; see §A.3.3).

3.29 *Representation of piecewise-linear convex functions.* A convex function $f : \mathbf{R}^n \to \mathbf{R}$, with $\mathbf{dom}\, f = \mathbf{R}^n$, is called *piecewise-linear* if there exists a partition of $\mathbf{R}^n$ as

$$\mathbf{R}^n = X_1 \cup X_2 \cup \cdots \cup X_L,$$

where $\mathbf{int}\, X_i \neq \emptyset$ and $\mathbf{int}\, X_i \cap \mathbf{int}\, X_j = \emptyset$ for $i \neq j$, and a family of affine functions $a_1^T x + b_1, \ldots, a_L^T x + b_L$ such that $f(x) = a_i^T x + b_i$ for $x \in X_i$.

Show that such a function has the form $f(x) = \max\{a_1^T x + b_1, \ldots, a_L^T x + b_L\}$.

3.30 *Convex hull or envelope of a function.* The *convex hull* or *convex envelope* of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as

$$g(x) = \inf\{t \mid (x, t) \in \mathbf{conv}\,\mathbf{epi}\, f\}.$$

Geometrically, the epigraph of g is the convex hull of the epigraph of f.

Show that g is the largest convex underestimator of f. In other words, show that if h is convex and satisfies $h(x) \leq f(x)$ for all x, then $h(x) \leq g(x)$ for all x.

3.31 [Roc70, page 35] *Largest homogeneous underestimator.* Let f be a convex function. Define the function g as

$$g(x) = \inf_{\alpha > 0} \frac{f(\alpha x)}{\alpha}.$$

(a) Show that g is homogeneous ($g(tx) = tg(x)$ for all $t \geq 0$).

(b) Show that g is the largest homogeneous underestimator of f: If h is homogeneous and $h(x) \leq f(x)$ for all x, then we have $h(x) \leq g(x)$ for all x.

(c) Show that g is convex.

3.32 *Products and ratios of convex functions.* In general the product or ratio of two convex functions is not convex. However, there are some results that apply to functions on $\mathbf{R}$. Prove the following.

(a) If f and g are convex, both nondecreasing (or nonincreasing), and positive functions on an interval, then fg is convex.

(b) If f, g are concave, positive, with one nondecreasing and the other nonincreasing, then fg is concave.

(c) If f is convex, nondecreasing, and positive, and g is concave, nonincreasing, and positive, then f/g is convex.

3.33 *Direct proof of perspective theorem.* Give a direct proof that the perspective function g, as defined in §3.2.6, of a convex function f is convex: Show that $\mathbf{dom}\, g$ is a convex set, and that for (x, t), $(y, s) \in \mathbf{dom}\, g$, and $0 \leq \theta \leq 1$, we have

$$g(\theta x + (1-\theta)y, \theta t + (1-\theta)s) \leq \theta g(x, t) + (1-\theta) g(y, s).$$

3.34 *The Minkowski function.* The *Minkowski function* of a convex set C is defined as

$$M_C(x) = \inf\{t > 0 \mid t^{-1} x \in C\}.$$

(a) Draw a picture giving a geometric interpretation of how to find $M_C(x)$.

(b) Show that M_C is homogeneous, *i.e.*, $M_C(\alpha x) = \alpha M_C(x)$ for $\alpha \geq 0$.

(c) What is $\mathbf{dom}\, M_C$?

(d) Show that M_C is a convex function.

(e) Suppose C is also closed, bounded, symmetric (if $x \in C$ then $-x \in C$), and has nonempty interior. Show that M_C is a norm. What is the corresponding unit ball?

3.35 *Support function calculus.* Recall that the support function of a set $C \subseteq \mathbf{R}^n$ is defined as $S_C(y) = \sup\{y^T x \mid x \in C\}$. On page 81 we showed that S_C is a convex function.

(a) Show that $S_B = S_{\mathbf{conv}\, B}$.

(b) Show that $S_{A+B} = S_A + S_B$.

(c) Show that $S_{A \cup B} = \max\{S_A, S_B\}$.

(d) Let B be closed and convex. Show that $A \subseteq B$ if and only if $S_A(y) \leq S_B(y)$ for all y.

Conjugate functions

3.36 Derive the conjugates of the following functions.

(a) *Max function.* $f(x) = \max_{i=1,\ldots,n} x_i$ on $\mathbf{R}^n$.

(b) *Sum of largest elements.* $f(x) = \sum_{i=1}^r x_{[i]}$ on $\mathbf{R}^n$.

(c) *Piecewise-linear function on* $\mathbf{R}$. $f(x) = \max_{i=1,\ldots,m}(a_i x + b_i)$ on $\mathbf{R}$. You can assume that the a_i are sorted in increasing order, *i.e.*, $a_1 \leq \cdots \leq a_m$, and that none of the functions $a_i x + b_i$ is redundant, *i.e.*, for each k there is at least one x with $f(x) = a_k x + b_k$.

(d) *Power function.* $f(x) = x^p$ on $\mathbf{R}_{++}$, where $p > 1$. Repeat for $p < 0$.

(e) *Negative geometric mean.* $f(x) = -(\prod x_i)^{1/n}$ on $\mathbf{R}^n_{++}$.

(f) *Negative generalized logarithm for second-order cone.* $f(x,t) = -\log(t^2 - x^T x)$ on $\{(x,t) \in \mathbf{R}^n \times \mathbf{R} \mid \|x\|_2 < t\}$.

3.37 Show that the conjugate of $f(X) = \mathbf{tr}(X^{-1})$ with $\mathbf{dom}\, f = \mathbf{S}^n_{++}$ is given by

$$f^*(Y) = -2\,\mathbf{tr}(-Y)^{1/2}, \qquad \mathbf{dom}\, f^* = -\mathbf{S}^n_+.$$

Hint. The gradient of f is $\nabla f(X) = -X^{-2}$.

3.38 *Young's inequality.* Let $f : \mathbf{R} \to \mathbf{R}$ be an increasing function, with $f(0) = 0$, and let g be its inverse. Define F and G as

$$F(x) = \int_0^x f(a)\, da, \qquad G(y) = \int_0^y g(a)\, da.$$

Show that F and G are conjugates. Give a simple graphical interpretation of Young's inequality,

$$xy \leq F(x) + G(y).$$

3.39 *Properties of conjugate functions.*

(a) *Conjugate of convex plus affine function.* Define $g(x) = f(x) + c^T x + d$, where f is convex. Express g^* in terms of f^* (and c, d).

(b) *Conjugate of perspective.* Express the conjugate of the perspective of a convex function f in terms of f^*.

(c) *Conjugate and minimization.* Let $f(x, z)$ be convex in (x, z) and define $g(x) = \inf_z f(x, z)$. Express the conjugate g^* in terms of f^*.

As an application, express the conjugate of $g(x) = \inf_z\{h(z) \mid Az + b = x\}$, where h is convex, in terms of h^*, A, and b.

(d) *Conjugate of conjugate.* Show that the conjugate of the conjugate of a closed convex function is itself: $f = f^{**}$ if f is closed and convex. (A function is closed if its epigraph is closed; see §A.3.3.) *Hint.* Show that f^{**} is the pointwise supremum of all affine global underestimators of f. Then apply the result of exercise 3.28.

3.40 *Gradient and Hessian of conjugate function.* Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is convex and twice continuously differentiable. Suppose $\bar{y}$ and $\bar{x}$ are related by $\bar{y} = \nabla f(\bar{x})$, and that $\nabla^2 f(\bar{x}) \succ 0$.

(a) Show that $\nabla f^*(\bar{y}) = \bar{x}$.

(b) Show that $\nabla^2 f^*(\bar{y}) = \nabla^2 f(\bar{x})^{-1}$.

3.41 *Conjugate of negative normalized entropy.* Show that the conjugate of the negative normalized entropy

$$f(x) = \sum_{i=1}^n x_i \log(x_i/\mathbf{1}^T x),$$

with $\mathbf{dom}\, f = \mathbf{R}^n_{++}$, is given by

$$f^*(y) = \begin{cases} 0 & \sum_{i=1}^n e^{y_i} \leq 1 \\ +\infty & \text{otherwise.} \end{cases}$$

Quasiconvex functions

3.42 *Approximation width.* Let $f_0, \ldots, f_n : \mathbf{R} \to \mathbf{R}$ be given continuous functions. We consider the problem of approximating f_0 as a linear combination of $f_1, \ldots, f_n$. For $x \in \mathbf{R}^n$, we say that $f = x_1 f_1 + \cdots + x_n f_n$ approximates f_0 with tolerance $\epsilon > 0$ over the interval $[0, T]$ if $|f(t) - f_0(t)| \leq \epsilon$ for $0 \leq t \leq T$. Now we choose a fixed tolerance $\epsilon > 0$ and define the *approximation width* as the largest T such that f approximates f_0 over the interval $[0, T]$:

$$W(x) = \sup\{T \mid |x_1 f_1(t) + \cdots + x_n f_n(t) - f_0(t)| \leq \epsilon \text{ for } 0 \leq t \leq T\}.$$

Show that W is quasiconcave.

3.43 *First-order condition for quasiconvexity.* Prove the first-order condition for quasiconvexity given in §3.4.3: A differentiable function $f : \mathbf{R}^n \to \mathbf{R}$, with $\mathbf{dom}\, f$ convex, is quasiconvex if and only if for all $x, y \in \mathbf{dom}\, f$,

$$f(y) \leq f(x) \Longrightarrow \nabla f(x)^T (y - x) \leq 0.$$

Hint. It suffices to prove the result for a function on $\mathbf{R}$; the general result follows by restriction to an arbitrary line.

3.44 *Second-order conditions for quasiconvexity.* In this problem we derive alternate representations of the second-order conditions for quasiconvexity given in §3.4.3. Prove the following.

(a) A point $x \in \mathbf{dom}\, f$ satisfies (3.21) if there exists a σ such that

$$\nabla^2 f(x) + \sigma \nabla f(x) \nabla f(x)^T \succeq 0. \tag{3.26}$$

It satisfies (3.22) for all $y \neq 0$ if and only if there exists a σ such

$$\nabla^2 f(x) + \sigma \nabla f(x) \nabla f(x)^T \succ 0. \tag{3.27}$$

Hint. We can assume without loss of generality that $\nabla^2 f(x)$ is diagonal.

(b) A point $x \in \mathbf{dom}\, f$ satisfies (3.21) if and only if either $\nabla f(x) = 0$ and $\nabla^2 f(x) \succeq 0$, or $\nabla f(x) \neq 0$ and the matrix

$$H(x) = \begin{bmatrix} \nabla^2 f(x) & \nabla f(x) \\ \nabla f(x)^T & 0 \end{bmatrix}$$

has exactly one negative eigenvalue. It satisfies (3.22) for all $y \neq 0$ if and only if $H(x)$ has exactly one nonpositive eigenvalue.

Hint. You can use the result of part (a). The following result, which follows from the eigenvalue interlacing theorem in linear algebra, may also be useful: If $B \in \mathbf{S}^n$ and $a \in \mathbf{R}^n$, then

$$\lambda_n \left(\begin{bmatrix} B & a \\ a^T & 0 \end{bmatrix} \right) \geq \lambda_n(B).$$

3.45 Use the first and second-order conditions for quasiconvexity given in §3.4.3 to verify quasiconvexity of the function $f(x) = -x_1 x_2$, with $\mathbf{dom}\, f = \mathbf{R}^2_{++}$.

3.46 *Quasilinear functions with domain $\mathbf{R}^n$.* A function on $\mathbf{R}$ that is quasilinear (*i.e.*, quasiconvex and quasiconcave) is monotone, *i.e.*, either nondecreasing or nonincreasing. In this problem we consider a generalization of this result to functions on $\mathbf{R}^n$.

Suppose the function $f : \mathbf{R}^n \to \mathbf{R}$ is quasilinear and continuous with $\mathbf{dom}\, f = \mathbf{R}^n$. Show that it can be expressed as $f(x) = g(a^T x)$, where $g : \mathbf{R} \to \mathbf{R}$ is monotone and $a \in \mathbf{R}^n$. In other words, a quasilinear function with domain $\mathbf{R}^n$ must be a monotone function of a linear function. (The converse is also true.)

Log-concave and log-convex functions

3.47 Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is differentiable, $\mathbf{dom}\, f$ is convex, and $f(x) > 0$ for all $x \in \mathbf{dom}\, f$. Show that f is log-concave if and only if for all $x, y \in \mathbf{dom}\, f$,

$$\frac{f(y)}{f(x)} \leq \exp\left(\frac{\nabla f(x)^T (y - x)}{f(x)} \right).$$

3.48 Show that if $f : \mathbf{R}^n \to \mathbf{R}$ is log-concave and $a \geq 0$, then the function $g = f - a$ is log-concave, where $\mathbf{dom}\, g = \{x \in \mathbf{dom}\, f \mid f(x) > a\}$.

3.49 Show that the following functions are log-concave.

(a) *Logistic function:* $f(x) = e^x/(1 + e^x)$ with $\mathbf{dom}\, f = \mathbf{R}$.

(b) *Harmonic mean:*

$$f(x) = \frac{1}{1/x_1 + \cdots + 1/x_n}, \qquad \mathbf{dom}\, f = \mathbf{R}^n_{++}.$$

(c) *Product over sum:*

$$f(x) = \frac{\prod_{i=1}^n x_i}{\sum_{i=1}^n x_i}, \qquad \mathbf{dom}\, f = \mathbf{R}^n_{++}.$$

(d) *Determinant over trace:*

$$f(X) = \frac{\det X}{\mathbf{tr}\, X}, \qquad \mathbf{dom}\, f = \mathbf{S}^n_{++}.$$

3.50 *Coefficients of a polynomial as a function of the roots.* Show that the coefficients of a polynomial with real negative roots are log-concave functions of the roots. In other words, the functions $a_i : \mathbf{R}^n \to \mathbf{R}$, defined by the identity

$$s^n + a_1(\lambda)s^{n-1} + \cdots + a_{n-1}(\lambda)s + a_n(\lambda) = (s - \lambda_1)(s - \lambda_2)\cdots(s - \lambda_n),$$

are log-concave on $-\mathbf{R}^n_{++}$.

Hint. The function

$$S_k(x) = \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} x_{i_1} x_{i_2} \cdots x_{i_k},$$

with $\mathbf{dom}\, S_k \in \mathbf{R}^n_+$ and $1 \le k \le n$, is called the kth elementary symmetric function on $\mathbf{R}^n$. It can be shown that $S_k^{1/k}$ is concave (see [ML57]).

3.51 [BL00, page 41] Let p be a polynomial on $\mathbf{R}$, with all its roots real. Show that it is log-concave on any interval on which it is positive.

3.52 [MO79, §3.E.2] *Log-convexity of moment functions.* Suppose $f : \mathbf{R} \to \mathbf{R}$ is nonnegative with $\mathbf{R}_+ \subseteq \mathbf{dom}\, f$. For $x \ge 0$ define

$$\phi(x) = \int_0^\infty u^x f(u)\, du.$$

Show that ϕ is a log-convex function. (If x is a positive integer, and f is a probability density function, then $\phi(x)$ is the xth moment of the distribution.)

Use this to show that the Gamma function,

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u}\, du,$$

is log-convex for $x \ge 1$.

3.53 Suppose x and y are independent random vectors in $\mathbf{R}^n$, with log-concave probability density functions f and g, respectively. Show that the probability density function of the sum $z = x + y$ is log-concave.

3.54 *Log-concavity of Gaussian cumulative distribution function.* The cumulative distribution function of a Gaussian random variable,

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2}\, dt,$$

is log-concave. This follows from the general result that the convolution of two log-concave functions is log-concave. In this problem we guide you through a simple self-contained proof that f is log-concave. Recall that f is log-concave if and only if $f''(x)f(x) \le f'(x)^2$ for all x.

(a) Verify that $f''(x)f(x) \le f'(x)^2$ for $x \ge 0$. That leaves us the hard part, which is to show the inequality for $x < 0$.

(b) Verify that for any t and x we have $t^2/2 \ge -x^2/2 + xt$.

(c) Using part (b) show that $e^{-t^2/2} \le e^{x^2/2 - xt}$. Conclude that, for $x < 0$,

$$\int_{-\infty}^x e^{-t^2/2}\, dt \le e^{x^2/2} \int_{-\infty}^x e^{-xt}\, dt.$$

(d) Use part (c) to verify that $f''(x)f(x) \le f'(x)^2$ for $x \le 0$.

3.55 *Log-concavity of the cumulative distribution function of a log-concave probability density.* In this problem we extend the result of exercise 3.54. Let $g(t) = \exp(-h(t))$ be a differentiable log-concave probability density function, and let

$$f(x) = \int_{-\infty}^{x} g(t)\, dt = \int_{-\infty}^{x} e^{-h(t)}\, dt$$

be its cumulative distribution. We will show that f is log-concave, *i.e.*, it satisfies $f''(x)f(x) \leq (f'(x))^2$ for all x.

(a) Express the derivatives of f in terms of the function h. Verify that $f''(x)f(x) \leq (f'(x))^2$ if $h'(x) \geq 0$.

(b) Assume that $h'(x) < 0$. Use the inequality

$$h(t) \geq h(x) + h'(x)(t - x)$$

(which follows from convexity of h), to show that

$$\int_{-\infty}^{x} e^{-h(t)}\, dt \leq \frac{e^{-h(x)}}{-h'(x)}.$$

Use this inequality to verify that $f''(x)f(x) \leq (f'(x))^2$ if $h'(x) < 0$.

3.56 *More log-concave densities.* Show that the following densities are log-concave.

(a) [MO79, page 493] The *gamma density*, defined by

$$f(x) = \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x},$$

with $\mathbf{dom}\, f = \mathbf{R}_+$. The parameters λ and α satisfy $\lambda \geq 1$, $\alpha > 0$.

(b) [MO79, page 306] The *Dirichlet density*

$$f(x) = \frac{\Gamma(\mathbf{1}^T\lambda)}{\Gamma(\lambda_1)\cdots\Gamma(\lambda_{n+1})} x_1^{\lambda_1-1} \cdots x_n^{\lambda_n-1} \left(1 - \sum_{i=1}^n x_i\right)^{\lambda_{n+1}-1}$$

with $\mathbf{dom}\, f = \{x \in \mathbf{R}^n_{++} \mid \mathbf{1}^T x < 1\}$. The parameter λ satisfies $\lambda \succeq \mathbf{1}$.

Convexity with respect to a generalized inequality

3.57 Show that the function $f(X) = X^{-1}$ is matrix convex on $\mathbf{S}^n_{++}$.

3.58 *Schur complement.* Suppose $X \in \mathbf{S}^n$ partitioned as

$$X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix},$$

where $A \in \mathbf{S}^k$. The *Schur complement* of X (with respect to A) is $S = C - B^T A^{-1} B$ (see §A.5.5). Show that the Schur complement, viewed as a function from $\mathbf{S}^n$ into $\mathbf{S}^{n-k}$, is matrix concave on $\mathbf{S}^n_{++}$.

3.59 *Second-order conditions for K-convexity.* Let $K \subseteq \mathbf{R}^m$ be a proper convex cone, with associated generalized inequality $\preceq_K$. Show that a twice differentiable function $f : \mathbf{R}^n \to \mathbf{R}^m$, with convex domain, is K-convex if and only if for all $x \in \mathbf{dom}\, f$ and all $y \in \mathbf{R}^n$,

$$\sum_{i,j=1}^n \frac{\partial^2 f(x)}{\partial x_i \partial x_j} y_i y_j \succeq_K 0,$$

i.e., the second derivative is a K-nonnegative bilinear form. (Here $\partial^2 f/\partial x_i \partial x_j \in \mathbf{R}^m$, with components $\partial^2 f_k/\partial x_i \partial x_j$, for $k = 1, \ldots, m$; see §A.4.1.)

3.60 *Sublevel sets and epigraph of K-convex functions.* Let $K \subseteq \mathbf{R}^m$ be a proper convex cone with associated generalized inequality $\preceq_K$, and let $f : \mathbf{R}^n \to \mathbf{R}^m$. For $\alpha \in \mathbf{R}^m$, the α-sublevel set of f (with respect to $\preceq_K$) is defined as

$$C_\alpha = \{x \in \mathbf{R}^n \mid f(x) \preceq_K \alpha\}.$$

The epigraph of f, with respect to $\preceq_K$, is defined as the set

$$\mathbf{epi}_K f = \{(x, t) \in \mathbf{R}^{n+m} \mid f(x) \preceq_K t\}.$$

Show the following:

(a) If f is K-convex, then its sublevel sets C_α are convex for all α.

(b) f is K-convex if and only if $\mathbf{epi}_K f$ is a convex set.

Chapter 4

Convex optimization problems

4.1 Optimization problems

4.1.1 Basic terminology

We use the notation

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array} \tag{4.1}$$

to describe the problem of finding an x that minimizes $f_0(x)$ among all x that satisfy the conditions $f_i(x) \leq 0$, $i = 1, \ldots, m$, and $h_i(x) = 0$, $i = 1, \ldots, p$. We call $x \in \mathbf{R}^n$ the *optimization variable* and the function $f_0 : \mathbf{R}^n \to \mathbf{R}$ the *objective function* or *cost function*. The inequalities $f_i(x) \leq 0$ are called *inequality constraints*, and the corresponding functions $f_i : \mathbf{R}^n \to \mathbf{R}$ are called the *inequality constraint functions*. The equations $h_i(x) = 0$ are called the *equality constraints*, and the functions $h_i : \mathbf{R}^n \to \mathbf{R}$ are the *equality constraint functions*. If there are no constraints (*i.e.*, $m = p = 0$) we say the problem (4.1) is *unconstrained*.

The set of points for which the objective and all constraint functions are defined,

$$\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \ \cap \ \bigcap_{i=1}^{p} \mathbf{dom}\, h_i,$$

is called the *domain* of the optimization problem (4.1). A point $x \in \mathcal{D}$ is *feasible* if it satisfies the constraints $f_i(x) \leq 0$, $i = 1, \ldots, m$, and $h_i(x) = 0$, $i = 1, \ldots, p$. The problem (4.1) is said to be feasible if there exists at least one feasible point, and *infeasible* otherwise. The set of all feasible points is called the *feasible set* or the *constraint set*.

The *optimal value* $p^\star$ of the problem (4.1) is defined as

$$p^\star = \inf \left\{ f_0(x) \mid f_i(x) \leq 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, p \right\}.$$

We allow $p^\star$ to take on the extended values $\pm\infty$. If the problem is infeasible, we have $p^\star = \infty$ (following the standard convention that the infimum of the empty set is ∞). If there are feasible points x_k with $f_0(x_k) \to -\infty$ as $k \to \infty$, then $p^\star = -\infty$, and we say the problem (4.1) is *unbounded below.*

Optimal and locally optimal points

We say $x^\star$ is an *optimal point*, or solves the problem (4.1), if $x^\star$ is feasible and $f_0(x^\star) = p^\star$. The set of all optimal points is the *optimal set*, denoted

$$X_{\text{opt}} = \{x \mid f_i(x) \le 0,\ i = 1,\dots,m,\ h_i(x) = 0,\ i = 1,\dots,p,\ f_0(x) = p^\star\}.$$

If there exists an optimal point for the problem (4.1), we say the optimal value is *attained* or *achieved*, and the problem is *solvable.* If X_{opt} is empty, we say the optimal value is not attained or not achieved. (This always occurs when the problem is unbounded below.) A feasible point x with $f_0(x) \le p^\star + \epsilon$ (where $\epsilon > 0$) is called ϵ*-suboptimal*, and the set of all ϵ-suboptimal points is called the ϵ*-suboptimal set* for the problem (4.1).

We say a feasible point x is *locally optimal* if there is an $R > 0$ such that

$$\begin{aligned} f_0(x) = \inf\{f_0(z) \mid\ & f_i(z) \le 0,\ i = 1,\dots,m, \\ & h_i(z) = 0,\ i = 1,\dots,p,\ \|z - x\|_2 \le R\}, \end{aligned}$$

or, in other words, x solves the optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(z) \\ \text{subject to} & f_i(z) \le 0, \quad i = 1,\dots,m \\ & h_i(z) = 0, \quad i = 1,\dots,p \\ & \|z - x\|_2 \le R \end{array}$$

with variable z. Roughly speaking, this means x minimizes f_0 over nearby points in the feasible set. The term 'globally optimal' is sometimes used for 'optimal' to distinguish between 'locally optimal' and 'optimal'. Throughout this book, however, optimal will mean globally optimal.

If x is feasible and $f_i(x) = 0$, we say the ith inequality constraint $f_i(x) \le 0$ is *active* at x. If $f_i(x) < 0$, we say the constraint $f_i(x) \le 0$ is *inactive.* (The equality constraints are active at all feasible points.) We say that a constraint is *redundant* if deleting it does not change the feasible set.

Example 4.1 We illustrate these definitions with a few simple unconstrained optimization problems with variable $x \in \mathbf{R}$, and $\mathbf{dom}\, f_0 = \mathbf{R}_{++}$.

- $f_0(x) = 1/x$: $p^\star = 0$, but the optimal value is not achieved.
- $f_0(x) = -\log x$: $p^\star = -\infty$, so this problem is unbounded below.
- $f_0(x) = x \log x$: $p^\star = -1/e$, achieved at the (unique) optimal point $x^\star = 1/e$.

Feasibility problems

If the objective function is identically zero, the optimal value is either zero (if the feasible set is nonempty) or ∞ (if the feasible set is empty). We call this the *feasibility problem*, and will sometimes write it as

$$\begin{array}{ll} \text{find} & x \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p. \end{array}$$

The feasibility problem is thus to determine whether the constraints are consistent, and if so, find a point that satisfies them.

4.1.2 Expressing problems in standard form

We refer to (4.1) as an optimization problem in *standard form*. In the standard form problem we adopt the convention that the righthand side of the inequality and equality constraints are zero. This can always be arranged by subtracting any nonzero righthand side: we represent the equality constraint $g_i(x) = \tilde{g}_i(x)$, for example, as $h_i(x) = 0$, where $h_i(x) = g_i(x) - \tilde{g}_i(x)$. In a similar way we express inequalities of the form $f_i(x) \ge 0$ as $-f_i(x) \le 0$.

Example 4.2 *Box constraints.* Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & l_i \le x_i \le u_i, \quad i = 1, \ldots, n, \end{array}$$

where $x \in \mathbf{R}^n$ is the variable. The constraints are called *variable bounds* (since they give lower and upper bounds for each x_i) or *box constraints* (since the feasible set is a box).

We can express this problem in standard form as

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & l_i - x_i \le 0, \quad i = 1, \ldots, n \\ & x_i - u_i \le 0, \quad i = 1, \ldots, n. \end{array}$$

There are $2n$ inequality constraint functions:

$$f_i(x) = l_i - x_i, \quad i = 1, \ldots, n,$$

and

$$f_i(x) = x_{i-n} - u_{i-n}, \quad i = n+1, \ldots, 2n.$$

Maximization problems

We concentrate on the minimization problem by convention. We can solve the *maximization* problem

$$\begin{array}{ll} \text{maximize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array} \tag{4.2}$$

by minimizing the function $-f_0$ subject to the constraints. By this correspondence we can define all the terms above for the maximization problem (4.2). For example the optimal value of (4.2) is defined as

$$p^\star = \sup\{f_0(x) \mid f_i(x) \le 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, p\},$$

and a feasible point x is ϵ-suboptimal if $f_0(x) \ge p^\star - \epsilon$. When the maximization problem is considered, the objective is sometimes called the *utility* or *satisfaction level* instead of the cost.

4.1.3 Equivalent problems

In this book we will use the notion of equivalence of optimization problems in an informal way. We call two problems *equivalent* if from a solution of one, a solution of the other is readily found, and vice versa. (It is possible, but complicated, to give a formal definition of equivalence.)

As a simple example, consider the problem

$$\begin{array}{ll} \text{minimize} & \tilde{f}(x) = \alpha_0 f_0(x) \\ \text{subject to} & \tilde{f}_i(x) = \alpha_i f_i(x) \le 0, \quad i = 1, \ldots, m \\ & \tilde{h}_i(x) = \beta_i h_i(x) = 0, \quad i = 1, \ldots, p, \end{array} \tag{4.3}$$

where $\alpha_i > 0$, $i = 0, \ldots, m$, and $\beta_i \ne 0$, $i = 1, \ldots, p$. This problem is obtained from the standard form problem (4.1) by scaling the objective and inequality constraint functions by positive constants, and scaling the equality constraint functions by nonzero constants. As a result, the feasible sets of the problem (4.3) and the original problem (4.1) are identical. A point x is optimal for the original problem (4.1) if and only if it is optimal for the scaled problem (4.3), so we say the two problems are equivalent. The two problems (4.1) and (4.3) are not, however, the same (unless α_i and β_i are all equal to one), since the objective and constraint functions differ.

We now describe some general transformations that yield equivalent problems.

Change of variables

Suppose $\phi : \mathbf{R}^n \to \mathbf{R}^n$ is one-to-one, with image covering the problem domain $\mathcal{D}$, *i.e.*, $\phi(\mathbf{dom}\, \phi) \supseteq \mathcal{D}$. We define functions $\tilde{f}_i$ and $\tilde{h}_i$ as

$$\tilde{f}_i(z) = f_i(\phi(z)), \quad i = 0, \ldots, m, \qquad \tilde{h}_i(z) = h_i(\phi(z)), \quad i = 1, \ldots, p.$$

Now consider the problem

$$\begin{array}{ll} \text{minimize} & \tilde{f}_0(z) \\ \text{subject to} & \tilde{f}_i(z) \le 0, \quad i = 1, \ldots, m \\ & \tilde{h}_i(z) = 0, \quad i = 1, \ldots, p, \end{array} \tag{4.4}$$

with variable z. We say that the standard form problem (4.1) and the problem (4.4) are related by the *change of variable* or *substitution of variable* $x = \phi(z)$.

The two problems are clearly equivalent: if x solves the problem (4.1), then $z = \phi^{-1}(x)$ solves the problem (4.4); if z solves the problem (4.4), then $x = \phi(z)$ solves the problem (4.1).

Transformation of objective and constraint functions

Suppose that $\psi_0 : \mathbf{R} \to \mathbf{R}$ is monotone increasing, $\psi_1, \ldots, \psi_m : \mathbf{R} \to \mathbf{R}$ satisfy $\psi_i(u) \leq 0$ if and only if $u \leq 0$, and $\psi_{m+1}, \ldots, \psi_{m+p} : \mathbf{R} \to \mathbf{R}$ satisfy $\psi_i(u) = 0$ if and only if $u = 0$. We define functions $\tilde{f}_i$ and $\tilde{h}_i$ as the compositions

$$\tilde{f}_i(x) = \psi_i(f_i(x)), \quad i = 0, \ldots, m, \qquad \tilde{h}_i(x) = \psi_{m+i}(h_i(x)), \quad i = 1, \ldots, p.$$

Evidently the associated problem

$$\begin{array}{ll} \text{minimize} & \tilde{f}_0(x) \\ \text{subject to} & \tilde{f}_i(x) \leq 0, \quad i = 1, \ldots, m \\ & \tilde{h}_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

and the standard form problem (4.1) are equivalent; indeed, the feasible sets are identical, and the optimal points are identical. (The example (4.3) above, in which the objective and constraint functions are scaled by appropriate constants, is the special case when all ψ_i are linear.)

Example 4.3 *Least-norm and least-norm-squared problems.* As a simple example consider the unconstrained Euclidean norm minimization problem

$$\text{minimize} \quad \|Ax - b\|_2, \tag{4.5}$$

with variable $x \in \mathbf{R}^n$. Since the norm is always nonnegative, we can just as well solve the problem

$$\text{minimize} \quad \|Ax - b\|_2^2 = (Ax - b)^T (Ax - b), \tag{4.6}$$

in which we minimize the square of the Euclidean norm. The problems (4.5) and (4.6) are clearly equivalent; the optimal points are the same. The two problems are not the same, however. For example, the objective in (4.5) is not differentiable at any x with $Ax - b = 0$, whereas the objective in (4.6) is differentiable for all x (in fact, quadratic).

Slack variables

One simple transformation is based on the observation that $f_i(x) \leq 0$ if and only if there is an $s_i \geq 0$ that satisfies $f_i(x) + s_i = 0$. Using this transformation we obtain the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & s_i \geq 0, \quad i = 1, \ldots, m \\ & f_i(x) + s_i = 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p, \end{array} \tag{4.7}$$

where the variables are $x \in \mathbf{R}^n$ and $s \in \mathbf{R}^m$. This problem has $n + m$ variables, m inequality constraints (the nonnegativity constraints on s_i), and $m + p$ equality constraints. The new variable s_i is called the *slack variable* associated with the original inequality constraint $f_i(x) \leq 0$. Introducing slack variables replaces each inequality constraint with an equality constraint, and a nonnegativity constraint.

The problem (4.7) is equivalent to the original standard form problem (4.1). Indeed, if (x, s) is feasible for the problem (4.7), then x is feasible for the original problem, since $s_i = -f_i(x) \geq 0$. Conversely, if x is feasible for the original problem, then (x, s) is feasible for the problem (4.7), where we take $s_i = -f_i(x)$. Similarly, x is optimal for the original problem (4.1) if and only if (x, s) is optimal for the problem (4.7), where $s_i = -f_i(x)$.

Eliminating equality constraints

If we can explicitly parametrize all solutions of the equality constraints

$$h_i(x) = 0, \quad i = 1, \ldots, p, \tag{4.8}$$

using some parameter $z \in \mathbf{R}^k$, then we can *eliminate* the equality constraints from the problem, as follows. Suppose the function $\phi : \mathbf{R}^k \to \mathbf{R}^n$ is such that x satisfies (4.8) if and only if there is some $z \in \mathbf{R}^k$ such that $x = \phi(z)$. The optimization problem

$$\begin{array}{ll} \text{minimize} & \tilde{f}_0(z) = f_0(\phi(z)) \\ \text{subject to} & \tilde{f}_i(z) = f_i(\phi(z)) \leq 0, \quad i = 1, \ldots, m \end{array}$$

is then equivalent to the original problem (4.1). This transformed problem has variable $z \in \mathbf{R}^k$, m inequality constraints, and no equality constraints. If z is optimal for the transformed problem, then $x = \phi(z)$ is optimal for the original problem. Conversely, if x is optimal for the original problem, then (since x is feasible) there is at least one z such that $x = \phi(z)$. Any such z is optimal for the transformed problem.

Eliminating linear equality constraints

The process of eliminating variables can be described more explicitly, and easily carried out numerically, when the equality constraints are all linear, *i.e.*, have the form $Ax = b$. If $Ax = b$ is inconsistent, *i.e.*, $b \notin \mathcal{R}(A)$, then the original problem is infeasible. Assuming this is not the case, let x_0 denote any solution of the equality constraints. Let $F \in \mathbf{R}^{n \times k}$ be any matrix with $\mathcal{R}(F) = \mathcal{N}(A)$, so the general solution of the linear equations $Ax = b$ is given by $Fz + x_0$, where $z \in \mathbf{R}^k$. (We can choose F to be full rank, in which case we have $k = n - \mathbf{rank}\, A$.)

Substituting $x = Fz + x_0$ into the original problem yields the problem

$$\begin{array}{ll} \text{minimize} & f_0(Fz + x_0) \\ \text{subject to} & f_i(Fz + x_0) \leq 0, \quad i = 1, \ldots, m, \end{array}$$

with variable z, which is equivalent to the original problem, has no equality constraints, and $\mathbf{rank}\, A$ fewer variables.

Introducing equality constraints

We can also *introduce* equality constraints and new variables into a problem. Instead of describing the general case, which is complicated and not very illuminating, we give a typical example that will be useful later. Consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(A_0 x + b_0) \\ \text{subject to} & f_i(A_i x + b_i) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p, \end{array}$$

where $x \in \mathbf{R}^n$, $A_i \in \mathbf{R}^{k_i \times n}$, and $f_i : \mathbf{R}^{k_i} \to \mathbf{R}$. In this problem the objective and constraint functions are given as compositions of the functions f_i with affine transformations defined by $A_i x + b_i$.

We introduce new variables $y_i \in \mathbf{R}^{k_i}$, as well as new equality constraints $y_i = A_i x + b_i$, for $i = 0, \ldots, m$, and form the equivalent problem

$$\begin{array}{ll} \text{minimize} & f_0(y_0) \\ \text{subject to} & f_i(y_i) \le 0, \quad i = 1, \ldots, m \\ & y_i = A_i x + b_i, \quad i = 0, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p. \end{array}$$

This problem has $k_0 + \cdots + k_m$ new variables,

$$y_0 \in \mathbf{R}^{k_0}, \quad \ldots, \quad y_m \in \mathbf{R}^{k_m},$$

and $k_0 + \cdots + k_m$ new equality constraints,

$$y_0 = A_0 x + b_0, \quad \ldots, \quad y_m = A_m x + b_m.$$

The objective and inequality constraints in this problem are *independent*, *i.e.*, involve different optimization variables.

Optimizing over some variables

We always have

$$\inf_{x,y} f(x,y) = \inf_x \tilde{f}(x)$$

where $\tilde{f}(x) = \inf_y f(x,y)$. In other words, we can always minimize a function by first minimizing over some of the variables, and then minimizing over the remaining ones. This simple and general principle can be used to transform problems into equivalent forms. The general case is cumbersome to describe and not illuminating, so we describe instead an example.

Suppose the variable $x \in \mathbf{R}^n$ is partitioned as $x = (x_1, x_2)$, with $x_1 \in \mathbf{R}^{n_1}$, $x_2 \in \mathbf{R}^{n_2}$, and $n_1 + n_2 = n$. We consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x_1, x_2) \\ \text{subject to} & f_i(x_1) \le 0, \quad i = 1, \ldots, m_1 \\ & \tilde{f}_i(x_2) \le 0, \quad i = 1, \ldots, m_2, \end{array} \tag{4.9}$$

in which the constraints are independent, in the sense that each constraint function depends on x_1 or x_2. We first minimize over x_2. Define the function $\tilde{f}_0$ of x_1 by

$$\tilde{f}_0(x_1) = \inf\{f_0(x_1, z) \mid \tilde{f}_i(z) \le 0,\ i = 1, \ldots, m_2\}.$$

The problem (4.9) is then equivalent to

$$\begin{array}{ll} \text{minimize} & \tilde{f}_0(x_1) \\ \text{subject to} & f_i(x_1) \le 0, \quad i = 1, \ldots, m_1. \end{array} \tag{4.10}$$

Example 4.4 *Minimizing a quadratic function with constraints on some variables.* Consider a problem with strictly convex quadratic objective, with some of the variables unconstrained:

$$\begin{array}{ll} \text{minimize} & x_1^T P_{11} x_1 + 2x_1^T P_{12} x_2 + x_2^T P_{22} x_2 \\ \text{subject to} & f_i(x_1) \le 0, \quad i = 1, \ldots, m, \end{array}$$

where P_{11} and P_{22} are symmetric. Here we can analytically minimize over x_2:

$$\inf_{x_2} \left(x_1^T P_{11} x_1 + 2x_1^T P_{12} x_2 + x_2^T P_{22} x_2 \right) = x_1^T \left(P_{11} - P_{12} P_{22}^{-1} P_{12}^T \right) x_1$$

(see §A.5.5). Therefore the original problem is equivalent to

$$\begin{array}{ll} \text{minimize} & x_1^T \left(P_{11} - P_{12} P_{22}^{-1} P_{12}^T \right) x_1 \\ \text{subject to} & f_i(x_1) \le 0, \quad i = 1, \ldots, m. \end{array}$$

Epigraph problem form

The *epigraph form* of the standard problem (4.1) is the problem

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & f_0(x) - t \le 0 \\ & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p, \end{array} \tag{4.11}$$

with variables $x \in \mathbf{R}^n$ and $t \in \mathbf{R}$. We can easily see that it is equivalent to the original problem: (x, t) is optimal for (4.11) if and only if x is optimal for (4.1) and $t = f_0(x)$. Note that the objective function of the epigraph form problem is a *linear* function of the variables x, t.

The epigraph form problem (4.11) can be interpreted geometrically as an optimization problem in the 'graph space' (x, t): we minimize t over the epigraph of f_0, subject to the constraints on x. This is illustrated in figure 4.1.

Implicit and explicit constraints

By a simple trick already mentioned in §3.1.2, we can include any of the constraints *implicitly* in the objective function, by redefining its domain. As an extreme example, the standard form problem can be expressed as the *unconstrained* problem

$$\text{minimize} \quad F(x), \tag{4.12}$$

where we define the function F as f_0, but with domain restricted to the feasible set:

$$\mathbf{dom}\, F = \{x \in \mathbf{dom}\, f_0 \mid f_i(x) \le 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, p\},$$

and $F(x) = f_0(x)$ for $x \in \mathbf{dom}\, F$. (Equivalently, we can define $F(x)$ to have value ∞ for x not feasible.) The problems (4.1) and (4.12) are clearly equivalent: they have the same feasible set, optimal points, and optimal value.

Of course this transformation is nothing more than a notational trick. Making the constraints implicit has not made the problem any easier to analyze or solve,

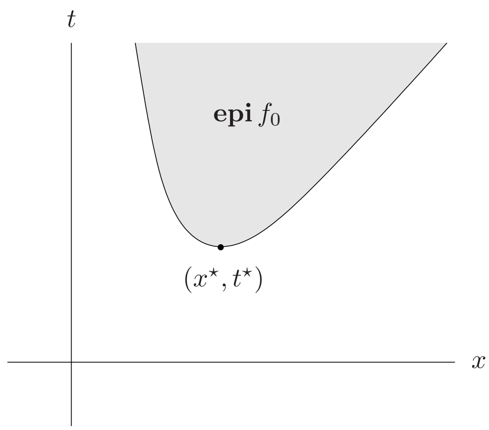


Figure 4.1 Geometric interpretation of epigraph form problem, for a problem with no constraints. The problem is to find the point in the epigraph (shown shaded) that minimizes t, *i.e.*, the 'lowest' point in the epigraph. The optimal point is $(x^\star, t^\star)$.

even though the problem (4.12) is, at least nominally, unconstrained. In some ways the transformation makes the problem more difficult. Suppose, for example, that the objective f_0 in the original problem is differentiable, so in particular its domain is open. The restricted objective function F is probably not differentiable, since its domain is likely not to be open.

Conversely, we will encounter problems with implicit constraints, which we can then make explicit. As a simple example, consider the unconstrained problem

$$\text{minimize} \quad f(x) \tag{4.13}$$

where the function f is given by

$$f(x) = \begin{cases} x^T x & Ax = b \\ \infty & \text{otherwise.} \end{cases}$$

Thus, the objective function is equal to the quadratic form $x^T x$ on the affine set defined by $Ax = b$, and ∞ off the affine set. Since we can clearly restrict our attention to points that satisfy $Ax = b$, we say that the problem (4.13) has an *implicit equality constraint* $Ax = b$ hidden in the objective. We can make the implicit equality constraint explicit, by forming the equivalent problem

$$\begin{array}{ll} \text{minimize} & x^T x \\ \text{subject to} & Ax = b. \end{array} \tag{4.14}$$

While the problems (4.13) and (4.14) are clearly equivalent, they are not the same. The problem (4.13) is unconstrained, but its objective function is not differentiable. The problem (4.14), however, has an equality constraint, but its objective and constraint functions are differentiable.

4.1.4 Parameter and oracle problem descriptions

For a problem in the standard form (4.1), there is still the question of how the objective and constraint functions are specified. In many cases these functions have some analytical or closed form, *i.e.*, are given by a formula or expression that involves the variable x as well as some parameters. Suppose, for example, the objective is quadratic, so it has the form $f_0(x) = (1/2)x^T Px + q^T x + r$. To specify the objective function we give the coefficients (also called *problem parameters* or *problem data*) $P \in \mathbf{S}^n$, $q \in \mathbf{R}^n$, and $r \in \mathbf{R}$. We call this a *parameter problem description*, since the specific problem to be solved (*i.e.*, the problem instance) is specified by giving the values of the parameters that appear in the expressions for the objective and constraint functions.

In other cases the objective and constraint functions are described by *oracle* models (which are also called *black box* or *subroutine* models). In an oracle model, we do not know f explicitly, but can evaluate $f(x)$ (and usually also some derivatives) at any $x \in \mathbf{dom}\, f$. This is referred to as *querying the oracle*, and is usually associated with some cost, such as time. We are also given some prior information about the function, such as convexity and a bound on its values. As a concrete example of an oracle model, consider an unconstrained problem, in which we are to minimize the function f. The function value $f(x)$ and its gradient $\nabla f(x)$ are evaluated in a subroutine. We can call the subroutine at any $x \in \mathbf{dom}\, f$, but do not have access to its source code. Calling the subroutine with argument x yields (when the subroutine returns) $f(x)$ and $\nabla f(x)$. Note that in the oracle model, we never really know the function; we only know the function value (and some derivatives) at the points where we have queried the oracle. (We also know some given prior information about the function, such as differentiability and convexity.)

In practice the distinction between a parameter and oracle problem description is not so sharp. If we are given a parameter problem description, we can construct an oracle for it, which simply evaluates the required functions and derivatives when queried. Most of the algorithms we study in part III work with an oracle model, but can be made more efficient when they are restricted to solve a specific parametrized family of problems.

4.2 Convex optimization

4.2.1 Convex optimization problems in standard form

A *convex optimization problem* is one of the form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \dots, m \\ & a_i^T x = b_i, \quad i = 1, \dots, p, \end{array} \tag{4.15}$$

where $f_0, \dots, f_m$ are convex functions. Comparing (4.15) with the general standard form problem (4.1), the convex problem has three additional requirements:

- the objective function must be convex,
- the inequality constraint functions must be convex,
- the equality constraint functions $h_i(x) = a_i^T x - b_i$ must be affine.

We immediately note an important property: The feasible set of a convex optimization problem is convex, since it is the intersection of the domain of the problem

$$\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i,$$

which is a convex set, with m (convex) sublevel sets $\{x \mid f_i(x) \leq 0\}$ and p hyperplanes $\{x \mid a_i^T x = b_i\}$. (We can assume without loss of generality that $a_i \neq 0$: if $a_i = 0$ and $b_i = 0$ for some i, then the ith equality constraint can be deleted; if $a_i = 0$ and $b_i \neq 0$, the ith equality constraint is inconsistent, and the problem is infeasible.) Thus, in a convex optimization problem, we minimize a convex objective function over a convex set.

If f_0 is quasiconvex instead of convex, we say the problem (4.15) is a (standard form) *quasiconvex optimization problem*. Since the sublevel sets of a convex or quasiconvex function are convex, we conclude that for a convex or quasiconvex optimization problem the ϵ-suboptimal sets are convex. In particular, the optimal set is convex. If the objective is strictly convex, then the optimal set contains at most one point.

Concave maximization problems

With a slight abuse of notation, we will also refer to

$$\begin{array}{ll} \text{maximize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & a_i^T x = b_i, \quad i = 1, \ldots, p, \end{array} \tag{4.16}$$

as a convex optimization problem if the objective function f_0 is concave, and the inequality constraint functions $f_1, \ldots, f_m$ are convex. This *concave maximization problem* is readily solved by minimizing the convex objective function $-f_0$. All of the results, conclusions, and algorithms that we describe for the minimization problem are easily transposed to the maximization case. In a similar way the maximization problem (4.16) is called *quasiconvex* if f_0 is quasiconcave.

Abstract form convex optimization problem

It is important to note a subtlety in our definition of convex optimization problem. Consider the example with $x \in \mathbf{R}^2$,

$$\begin{array}{ll} \text{minimize} & f_0(x) = x_1^2 + x_2^2 \\ \text{subject to} & f_1(x) = x_1/(1 + x_2^2) \leq 0 \\ & h_1(x) = (x_1 + x_2)^2 = 0, \end{array} \tag{4.17}$$

which is in the standard form (4.1). This problem is *not* a convex optimization problem in standard form since the equality constraint function h_1 is not affine, and the inequality constraint function f_1 is not convex. Nevertheless the feasible set, which is $\{x \mid x_1 \le 0,\ x_1 + x_2 = 0\}$, is convex. So although in this problem we are minimizing a convex function f_0 over a convex set, it is not a convex optimization problem by our definition.

Of course, the problem is readily reformulated as

$$\begin{array}{ll} \text{minimize} & f_0(x) = x_1^2 + x_2^2 \\ \text{subject to} & \tilde{f}_1(x) = x_1 \le 0 \\ & \tilde{h}_1(x) = x_1 + x_2 = 0, \end{array} \tag{4.18}$$

which is in standard convex optimization form, since f_0 and $\tilde{f}_1$ are convex, and $\tilde{h}_1$ is affine.

Some authors use the term *abstract convex optimization problem* to describe the (abstract) problem of minimizing a convex function over a convex set. Using this terminology, the problem (4.17) is an abstract convex optimization problem. *We will not use this terminology in this book.* For us, a convex optimization problem is not just one of minimizing a convex function over a convex set; it is also required that the feasible set be described specifically by a set of inequalities involving convex functions, and a set of linear equality constraints. The problem (4.17) is *not* a convex optimization problem, but the problem (4.18) *is* a convex optimization problem. (The two problems are, however, equivalent.)

Our adoption of the stricter definition of convex optimization problem does not matter much in practice. To solve the abstract problem of minimizing a convex function over a convex set, we need to find a description of the set in terms of convex inequalities and linear equality constraints. As the example above suggests, this is usually straightforward.

4.2.2 Local and global optima

A fundamental property of convex optimization problems is that any locally optimal point is also (globally) optimal. To see this, suppose that x is locally optimal for a convex optimization problem, *i.e.*, x is feasible and

$$f_0(x) = \inf\{f_0(z) \mid z \text{ feasible},\ \|z - x\|_2 \le R\}, \tag{4.19}$$

for some $R > 0$. Now suppose that x is *not* globally optimal, *i.e.*, there is a feasible y such that $f_0(y) < f_0(x)$. Evidently $\|y - x\|_2 > R$, since otherwise $f_0(x) \le f_0(y)$. Consider the point z given by

$$z = (1 - \theta)x + \theta y, \qquad \theta = \frac{R}{2\|y - x\|_2}.$$

Then we have $\|z - x\|_2 = R/2 < R$, and by convexity of the feasible set, z is feasible. By convexity of f_0 we have

$$f_0(z) \le (1 - \theta) f_0(x) + \theta f_0(y) < f_0(x),$$

which contradicts (4.19). Hence there exists no feasible y with $f_0(y) < f_0(x)$, *i.e.*, x is globally optimal.

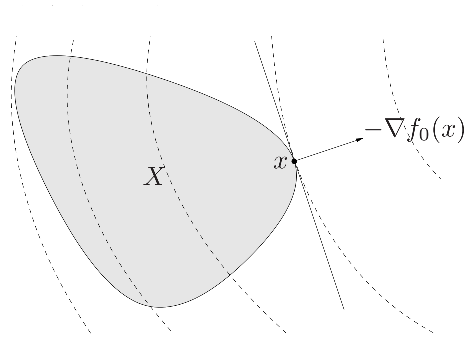


Figure 4.2 Geometric interpretation of the optimality condition (4.21). The feasible set X is shown shaded. Some level curves of f_0 are shown as dashed lines. The point x is optimal: $-\nabla f_0(x)$ defines a supporting hyperplane (shown as a solid line) to X at x.

It is not true that locally optimal points of quasiconvex optimization problems are globally optimal; see §4.2.5.

4.2.3 An optimality criterion for differentiable f_0

Suppose that the objective f_0 in a convex optimization problem is differentiable, so that for all $x, y \in \textbf{dom}\, f_0$,

$$f_0(y) \geq f_0(x) + \nabla f_0(x)^T (y - x) \tag{4.20}$$

(see §3.1.3). Let X denote the feasible set, *i.e.*,

$$X = \{x \mid f_i(x) \leq 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, p\}.$$

Then x is optimal if and only if $x \in X$ and

$$\nabla f_0(x)^T (y - x) \geq 0 \text{ for all } y \in X. \tag{4.21}$$

This optimality criterion can be understood geometrically: If $\nabla f_0(x) \neq 0$, it means that $-\nabla f_0(x)$ defines a supporting hyperplane to the feasible set at x (see figure 4.2).

Proof of optimality condition

First suppose $x \in X$ and satisfies (4.21). Then if $y \in X$ we have, by (4.20), $f_0(y) \geq f_0(x)$. This shows x is an optimal point for (4.1).

Conversely, suppose x is optimal, but the condition (4.21) does not hold, *i.e.*, for some $y \in X$ we have

$$\nabla f_0(x)^T (y - x) < 0.$$

Consider the point $z(t) = ty + (1-t)x$, where $t \in [0,1]$ is a parameter. Since $z(t)$ is on the line segment between x and y, and the feasible set is convex, $z(t)$ is feasible. We claim that for small positive t we have $f_0(z(t)) < f_0(x)$, which will prove that x is not optimal. To show this, note that

$$\frac{d}{dt} f_0(z(t))\bigg|_{t=0} = \nabla f_0(x)^T(y-x) < 0,$$

so for small positive t, we have $f_0(z(t)) < f_0(x)$.

We will pursue the topic of optimality conditions in much more depth in chapter 5, but here we examine a few simple examples.

Unconstrained problems

For an unconstrained problem (*i.e.*, $m = p = 0$), the condition (4.21) reduces to the well known necessary and sufficient condition

$$\nabla f_0(x) = 0 \tag{4.22}$$

for x to be optimal. While we have already seen this optimality condition, it is useful to see how it follows from (4.21). Suppose x is optimal, which means here that $x \in \mathbf{dom}\, f_0$, and for all feasible y we have $\nabla f_0(x)^T(y-x) \geq 0$. Since f_0 is differentiable, its domain is (by definition) open, so all y sufficiently close to x are feasible. Let us take $y = x - t\nabla f_0(x)$, where $t \in \mathbf{R}$ is a parameter. For t small and positive, y is feasible, and so

$$\nabla f_0(x)^T(y-x) = -t\|\nabla f_0(x)\|_2^2 \geq 0,$$

from which we conclude $\nabla f_0(x) = 0$.

There are several possible situations, depending on the number of solutions of (4.22). If there are no solutions of (4.22), then there are no optimal points; the optimal value of the problem is not attained. Here we can distinguish between two cases: the problem is unbounded below, or the optimal value is finite, but not attained. On the other hand we can have multiple solutions of the equation (4.22), in which case each such solution is a minimizer of f_0.

Example 4.5 *Unconstrained quadratic optimization.* Consider the problem of minimizing the quadratic function

$$f_0(x) = (1/2)x^T Px + q^T x + r,$$

where $P \in \mathbf{S}_+^n$ (which makes f_0 convex). The necessary and sufficient condition for x to be a minimizer of f_0 is

$$\nabla f_0(x) = Px + q = 0.$$

Several cases can occur, depending on whether this (linear) equation has no solutions, one solution, or many solutions.

- If $q \notin \mathcal{R}(P)$, then there is no solution. In this case f_0 is unbounded below.
- If $P \succ 0$ (which is the condition for f_0 to be strictly convex), then there is a unique minimizer, $x^\star = -P^{-1}q$.

- If P is singular, but $q \in \mathcal{R}(P)$, then the set of optimal points is the (affine) set $X_{\rm opt} = -P^\dagger q + \mathcal{N}(P)$, where $P^\dagger$ denotes the pseudo-inverse of P (see §A.5.4).

Example 4.6 *Analytic centering.* Consider the (unconstrained) problem of minimizing the (convex) function $f_0 : \mathbf{R}^n \to \mathbf{R}$, defined as

$$f_0(x) = -\sum_{i=1}^m \log(b_i - a_i^T x), \qquad \mathbf{dom}\, f_0 = \{x \mid Ax \prec b\},$$

where $a_1^T, \ldots, a_m^T$ are the rows of A. The function f_0 is differentiable, so the necessary and sufficient conditions for x to be optimal are

$$Ax \prec b, \qquad \nabla f_0(x) = \sum_{i=1}^m \frac{1}{b_i - a_i^T x} a_i = 0. \tag{4.23}$$

(The condition $Ax \prec b$ is just $x \in \mathbf{dom}\, f_0$.) If $Ax \prec b$ is infeasible, then the domain of f_0 is empty. Assuming $Ax \prec b$ is feasible, there are still several possible cases (see exercise 4.2):

- There are no solutions of (4.23), and hence no optimal points for the problem. This occurs if and only if f_0 is unbounded below.
- There are many solutions of (4.23). In this case it can be shown that the solutions form an affine set.
- There is a unique solution of (4.23), *i.e.*, a unique minimizer of f_0. This occurs if and only if the open polyhedron $\{x \mid Ax \prec b\}$ is nonempty and bounded.

Problems with equality constraints only

Consider the case where there are equality constraints but no inequality constraints, *i.e.*,

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax = b. \end{array}$$

Here the feasible set is affine. We assume that it is nonempty; otherwise the problem is infeasible. The optimality condition for a feasible x is that

$$\nabla f_0(x)^T (y - x) \geq 0$$

must hold for all y satisfying $Ay = b$. Since x is feasible, every feasible y has the form $y = x + v$ for some $v \in \mathcal{N}(A)$. The optimality condition can therefore be expressed as:

$$\nabla f_0(x)^T v \geq 0 \text{ for all } v \in \mathcal{N}(A).$$

If a linear function is nonnegative on a subspace, then it must be zero on the subspace, so it follows that $\nabla f_0(x)^T v = 0$ for all $v \in \mathcal{N}(A)$. In other words,

$$\nabla f_0(x) \perp \mathcal{N}(A).$$

Using the fact that $\mathcal{N}(A)^\perp = \mathcal{R}(A^T)$, this optimality condition can be expressed as $\nabla f_0(x) \in \mathcal{R}(A^T)$, *i.e.*, there exists a $\nu \in \mathbf{R}^p$ such that

$$\nabla f_0(x) + A^T \nu = 0.$$

Together with the requirement $Ax = b$ (*i.e.*, that x is feasible), this is the classical Lagrange multiplier optimality condition, which we will study in greater detail in chapter 5.

Minimization over the nonnegative orthant

As another example we consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \succeq 0, \end{array}$$

where the only inequality constraints are nonnegativity constraints on the variables.

The optimality condition (4.21) is then

$$x \succeq 0, \qquad \nabla f_0(x)^T (y - x) \geq 0 \text{ for all } y \succeq 0.$$

The term $\nabla f_0(x)^T y$, which is a linear function of y, is unbounded below on $y \succeq 0$, unless we have $\nabla f_0(x) \succeq 0$. The condition then reduces to $-\nabla f_0(x)^T x \geq 0$. But $x \succeq 0$ and $\nabla f_0(x) \succeq 0$, so we must have $\nabla f_0(x)^T x = 0$, *i.e.*,

$$\sum_{i=1}^n (\nabla f_0(x))_i x_i = 0.$$

Now each of the terms in this sum is the product of two nonnegative numbers, so we conclude that each term must be zero, *i.e.*, $(\nabla f_0(x))_i \, x_i = 0$ for $i = 1, \ldots, n$.

The optimality condition can therefore be expressed as

$$x \succeq 0, \qquad \nabla f_0(x) \succeq 0, \qquad x_i \, (\nabla f_0(x))_i = 0, \quad i = 1, \ldots, n.$$

The last condition is called *complementarity*, since it means that the sparsity patterns (*i.e.*, the set of indices corresponding to nonzero components) of the vectors x and $\nabla f_0(x)$ are complementary (*i.e.*, have empty intersection). We will encounter complementarity conditions again in chapter 5.

4.2.4 Equivalent convex problems

It is useful to see which of the transformations described in §4.1.3 preserve convexity.

Eliminating equality constraints

For a convex problem the equality constraints must be linear, *i.e.*, of the form $Ax = b$. In this case they can be eliminated by finding a particular solution x_0 of

$Ax = b$, and a matrix F whose range is the nullspace of A, which results in the problem

$$\begin{array}{ll} \text{minimize} & f_0(Fz + x_0) \\ \text{subject to} & f_i(Fz + x_0) \le 0, \quad i = 1, \ldots, m, \end{array}$$

with variable z. Since the composition of a convex function with an affine function is convex, eliminating equality constraints preserves convexity of a problem. Moreover, the process of eliminating equality constraints (and reconstructing the solution of the original problem from the solution of the transformed problem) involves standard linear algebra operations.

At least in principle, this means we can restrict our attention to convex optimization problems which have no equality constraints. In many cases, however, it is better to retain the equality constraints, since eliminating them can make the problem harder to understand and analyze, or ruin the efficiency of an algorithm that solves it. This is true, for example, when the variable x has very large dimension, and eliminating the equality constraints would destroy sparsity or some other useful structure of the problem.

Introducing equality constraints

We can introduce new variables and equality constraints into a convex optimization problem, provided the equality constraints are linear, and the resulting problem will also be convex. For example, if an objective or constraint function has the form $f_i(A_i x + b_i)$, where $A_i \in \mathbf{R}^{k_i \times n}$, we can introduce a new variable $y_i \in \mathbf{R}^{k_i}$, replace $f_i(A_i x + b_i)$ with $f_i(y_i)$, and add the linear equality constraint $y_i = A_i x + b_i$.

Slack variables

By introducing slack variables we have the new constraints $f_i(x) + s_i = 0$. Since equality constraint functions must be affine in a convex problem, we must have f_i affine. In other words: introducing slack variables for *linear inequalities* preserves convexity of a problem.

Epigraph problem form

The epigraph form of the convex optimization problem (4.15) is

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & f_0(x) - t \le 0 \\ & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & a_i^T x = b_i, \quad i = 1, \ldots, p. \end{array}$$

The objective is linear (hence convex) and the new constraint function $f_0(x) - t$ is also convex in (x, t), so the epigraph form problem is convex as well.

It is sometimes said that a linear objective is *universal* for convex optimization, since any convex optimization problem is readily transformed to one with linear objective. The epigraph form of a convex problem has several practical uses. By assuming the objective of a convex optimization problem is linear, we can simplify theoretical analysis. It can also simplify algorithm development, since an algorithm that solves convex optimization problems with linear objective can, using the transformation above, solve any convex optimization problem (provided it can handle the constraint $f_0(x) - t \le 0$).

Minimizing over some variables

Minimizing a convex function over some variables preserves convexity. Therefore, if f_0 in (4.9) is jointly convex in x_1 and x_2, and f_i, $i = 1, \ldots, m_1$, and $\tilde{f}_i$, $i = 1, \ldots, m_2$, are convex, then the equivalent problem (4.10) is convex.

4.2.5 Quasiconvex optimization

Recall that a quasiconvex optimization problem has the standard form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & Ax = b, \end{array} \tag{4.24}$$

where the inequality constraint functions $f_1, \ldots, f_m$ are convex, and the objective f_0 is quasiconvex (instead of convex, as in a convex optimization problem). (Quasiconvex constraint functions can be replaced with equivalent convex constraint functions, *i.e.*, constraint functions that are convex and have the same 0-sublevel set, as in §3.4.5.)

In this section we point out some basic differences between convex and quasiconvex optimization problems, and also show how solving a quasiconvex optimization problem can be reduced to solving a sequence of convex optimization problems.

Locally optimal solutions and optimality conditions

The most important difference between convex and quasiconvex optimization is that a quasiconvex optimization problem can have locally optimal solutions that are not (globally) optimal. This phenomenon can be seen even in the simple case of unconstrained minimization of a quasiconvex function on $\mathbf{R}$, such as the one shown in figure 4.3.

Nevertheless, a variation of the optimality condition (4.21) given in §4.2.3 does hold for quasiconvex optimization problems with differentiable objective function. Let X denote the feasible set for the quasiconvex optimization problem (4.24). It follows from the first-order condition for quasiconvexity (3.20) that x is optimal if

$$x \in X, \qquad \nabla f_0(x)^T (y - x) > 0 \text{ for all } y \in X \setminus \{x\}. \tag{4.25}$$

There are two important differences between this criterion and the analogous one (4.21) for convex optimization:

- The condition (4.25) is only *sufficient* for optimality; simple examples show that it need not hold for an optimal point. In contrast, the condition (4.21) is necessary and sufficient for x to solve the convex problem.
- The condition (4.25) requires the gradient of f_0 to be nonzero, whereas the condition (4.21) does not. Indeed, when $\nabla f_0(x) = 0$ in the convex case, the condition (4.21) is satisfied, and x is optimal.

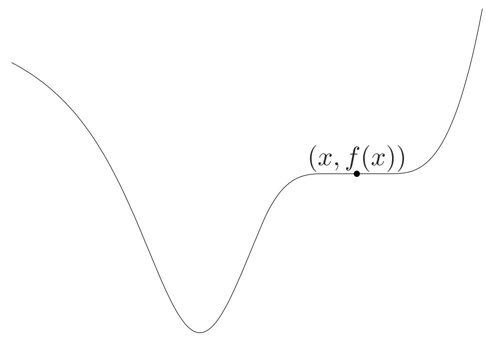


Figure 4.3 A quasiconvex function f on $\mathbf{R}$, with a locally optimal point x that is not globally optimal. This example shows that the simple optimality condition $f'(x) = 0$, valid for convex functions, does not hold for quasiconvex functions.

Quasiconvex optimization via convex feasibility problems

One general approach to quasiconvex optimization relies on the representation of the sublevel sets of a quasiconvex function via a family of convex inequalities, as described in §3.4.5. Let $\phi_t : \mathbf{R}^n \to \mathbf{R}$, $t \in \mathbf{R}$, be a family of convex functions that satisfy

$$f_0(x) \le t \iff \phi_t(x) \le 0,$$

and also, for each x, $\phi_t(x)$ is a nonincreasing function of t, *i.e.*, $\phi_s(x) \le \phi_t(x)$ whenever $s \ge t$.

Let $p^\star$ denote the optimal value of the quasiconvex optimization problem (4.24). If the feasibility problem

$$\begin{array}{ll} \text{find} & x \\ \text{subject to} & \phi_t(x) \le 0 \\ & f_i(x) \le 0, \quad i = 1, \dots, m \\ & Ax = b, \end{array} \tag{4.26}$$

is feasible, then we have $p^\star \le t$. Conversely, if the problem (4.26) is infeasible, then we can conclude $p^\star \ge t$. The problem (4.26) is a convex feasibility problem, since the inequality constraint functions are all convex, and the equality constraints are linear. Thus, we can check whether the optimal value $p^\star$ of a quasiconvex optimization problem is less than or more than a given value t by solving the convex feasibility problem (4.26). If the convex feasibility problem is feasible then we have $p^\star \le t$, and any feasible point x is feasible for the quasiconvex problem and satisfies $f_0(x) \le t$. If the convex feasibility problem is infeasible, then we know that $p^\star \ge t$.

This observation can be used as the basis of a simple algorithm for solving the quasiconvex optimization problem (4.24) using *bisection*, solving a convex feasibility problem at each step. We assume that the problem is feasible, and start with an interval $[l, u]$ known to contain the optimal value $p^\star$. We then solve the convex feasibility problem at its midpoint $t = (l + u)/2$, to determine whether the optimal value is in the lower or upper half of the interval, and update the interval accordingly. This produces a new interval, which also contains the optimal value, but has half the width of the initial interval. This is repeated until the width of the interval is small enough:

Algorithm 4.1 *Bisection method for quasiconvex optimization.*

given $l \leq p^\star$, $u \geq p^\star$, tolerance $\epsilon > 0$.

repeat
1. $t := (l + u)/2$.
2. Solve the convex feasibility problem (4.26).
3. **if** (4.26) is feasible, $u := t$; **else** $l := t$.

until $u - l \leq \epsilon$.

The interval $[l, u]$ is guaranteed to contain $p^\star$, *i.e.*, we have $l \leq p^\star \leq u$ at each step. In each iteration the interval is divided in two, *i.e.*, bisected, so the length of the interval after k iterations is $2^{-k}(u - l)$, where $u - l$ is the length of the initial interval. It follows that exactly $\lceil \log_2((u - l)/\epsilon) \rceil$ iterations are required before the algorithm terminates. Each step involves solving the convex feasibility problem (4.26).

4.3 Linear optimization problems

When the objective and constraint functions are all affine, the problem is called a *linear program* (LP). A general linear program has the form

$$\begin{array}{ll} \text{minimize} & c^T x + d \\ \text{subject to} & Gx \preceq h \\ & Ax = b, \end{array} \tag{4.27}$$

where $G \in \mathbf{R}^{m \times n}$ and $A \in \mathbf{R}^{p \times n}$. Linear programs are, of course, convex optimization problems.

It is common to omit the constant d in the objective function, since it does not affect the optimal (or feasible) set. Since we can maximize an affine objective $c^T x + d$, by minimizing $-c^T x - d$ (which is still convex), we also refer to a maximization problem with affine objective and constraint functions as an LP.

The geometric interpretation of an LP is illustrated in figure 4.4. The feasible set of the LP (4.27) is a polyhedron $\mathcal{P}$; the problem is to minimize the affine function $c^T x + d$ (or, equivalently, the linear function $c^T x$) over $\mathcal{P}$.

Standard and inequality form linear programs

Two special cases of the LP (4.27) are so widely encountered that they have been given separate names. In a *standard form LP* the only inequalities are componen-

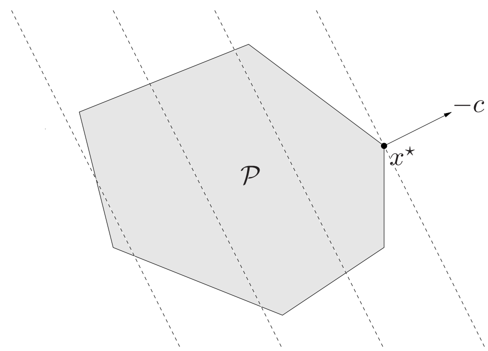


Figure 4.4 Geometric interpretation of an LP. The feasible set $\mathcal{P}$, which is a polyhedron, is shaded. The objective $c^T x$ is linear, so its level curves are hyperplanes orthogonal to c (shown as dashed lines). The point $x^\star$ is optimal; it is the point in $\mathcal{P}$ as far as possible in the direction $-c$.

twise nonnegativity constraints $x \succeq 0$:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0. \end{array} \tag{4.28}$$

If the LP has no equality constraints, it is called an *inequality form LP*, usually written as

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b. \end{array} \tag{4.29}$$

Converting LPs to standard form

It is sometimes useful to transform a general LP (4.27) to one in standard form (4.28) (for example in order to use an algorithm for standard form LPs). The first step is to introduce slack variables s_i for the inequalities, which results in

$$\begin{array}{ll} \text{minimize} & c^T x + d \\ \text{subject to} & Gx + s = h \\ & Ax = b \\ & s \succeq 0. \end{array}$$

The second step is to express the variable x as the difference of two nonnegative variables x^+ and x^-, *i.e.*, $x = x^+ - x^-$, $x^+, x^- \succeq 0$. This yields the problem

$$\begin{array}{ll} \text{minimize} & c^T x^+ - c^T x^- + d \\ \text{subject to} & Gx^+ - Gx^- + s = h \\ & Ax^+ - Ax^- = b \\ & x^+ \succeq 0, \quad x^- \succeq 0, \quad s \succeq 0, \end{array}$$

which is an LP in standard form, with variables x^+, x^-, and s. (For equivalence of this problem and the original one (4.27), see exercise 4.10.)

These techniques for manipulating problems (along with many others we will see in the examples and exercises) can be used to formulate many problems as linear programs. With some abuse of terminology, it is common to refer to a problem that can be formulated as an LP as an LP, even if it does not have the form (4.27).

4.3.1 Examples

LPs arise in a vast number of fields and applications; here we give a few typical examples.

Diet problem

A healthy diet contains m different nutrients in quantities at least equal to $b_1, \ldots, b_m$. We can compose such a diet by choosing nonnegative quantities $x_1, \ldots, x_n$ of n different foods. One unit quantity of food j contains an amount a_{ij} of nutrient i, and has a cost of c_j. We want to determine the cheapest diet that satisfies the nutritional requirements. This problem can be formulated as the LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \succeq b \\ & x \succeq 0. \end{array}$$

Several variations on this problem can also be formulated as LPs. For example, we can insist on an exact amount of a nutrient in the diet (which gives a linear equality constraint), or we can impose an upper bound on the amount of a nutrient, in addition to the lower bound as above.

Chebyshev center of a polyhedron

We consider the problem of finding the largest Euclidean ball that lies in a polyhedron described by linear inequalities,

$$\mathcal{P} = \{x \in \mathbf{R}^n \mid a_i^T x \le b_i,\ i = 1, \ldots, m\}.$$

(The center of the optimal ball is called the *Chebyshev center* of the polyhedron; it is the point deepest inside the polyhedron, *i.e.*, farthest from the boundary; see §8.5.1.) We represent the ball as

$$\mathcal{B} = \{x_c + u \mid \|u\|_2 \le r\}.$$

The variables in the problem are the center $x_c \in \mathbf{R}^n$ and the radius r; we wish to maximize r subject to the constraint $\mathcal{B} \subseteq \mathcal{P}$.

We start by considering the simpler constraint that $\mathcal{B}$ lies in one halfspace $a_i^T x \le b_i$, *i.e.*,

$$\|u\|_2 \le r \implies a_i^T(x_c + u) \le b_i. \tag{4.30}$$

Since

$$\sup\{a_i^T u \mid \|u\|_2 \le r\} = r\|a_i\|_2$$

we can write (4.30) as

$$a_i^T x_c + r\|a_i\|_2 \le b_i, \tag{4.31}$$

which is a linear inequality in x_c and r. In other words, the constraint that the ball lies in the halfspace determined by the inequality $a_i^T x \le b_i$ can be written as a linear inequality.

Therefore $\mathcal{B} \subseteq \mathcal{P}$ if and only if (4.31) holds for all $i = 1, \dots, m$. Hence the Chebyshev center can be determined by solving the LP

$$\begin{array}{ll} \text{maximize} & r \\ \text{subject to} & a_i^T x_c + r\|a_i\|_2 \le b_i, \quad i = 1, \dots, m, \end{array}$$

with variables r and x_c. (For more on the Chebyshev center, see §8.5.1.)

Dynamic activity planning

We consider the problem of choosing, or planning, the activity levels of n activities, or sectors of an economy, over N time periods. We let $x_j(t) \ge 0$, $t = 1, \dots, N$, denote the activity level of sector j, in period t. The activities both consume and produce products or goods in proportion to their activity levels. The amount of good i produced per unit of activity j is given by a_{ij}. Similarly, the amount of good i consumed per unit of activity j is b_{ij}. The total amount of goods produced in period t is given by $Ax(t) \in \mathbf{R}^m$, and the amount of goods consumed is $Bx(t) \in \mathbf{R}^m$. (Although we refer to these products as 'goods', they can also include unwanted products such as pollutants.)

The goods consumed in a period cannot exceed those produced in the previous period: we must have $Bx(t+1) \preceq Ax(t)$ for $t = 1, \dots, N$. A vector $g_0 \in \mathbf{R}^m$ of initial goods is given, which constrains the first period activity levels: $Bx(1) \preceq g_0$. The (vectors of) excess goods not consumed by the activities are given by

$$\begin{array}{rcl} s(0) & = & g_0 - Bx(1) \\ s(t) & = & Ax(t) - Bx(t+1), \quad t = 1, \dots, N-1 \\ s(N) & = & Ax(N). \end{array}$$

The objective is to maximize a discounted total value of excess goods:

$$c^T s(0) + \gamma c^T s(1) + \cdots + \gamma^N c^T s(N),$$

where $c \in \mathbf{R}^m$ gives the values of the goods, and $\gamma > 0$ is a discount factor. (The value c_i is negative if the ith product is unwanted, *e.g.*, a pollutant; $|c_i|$ is then the cost of disposal per unit.)

Putting it all together we arrive at the LP

$$\begin{array}{ll} \text{maximize} & c^T s(0) + \gamma c^T s(1) + \cdots + \gamma^N c^T s(N) \\ \text{subject to} & x(t) \succeq 0, \quad t = 1, \dots, N \\ & s(t) \succeq 0, \quad t = 0, \dots, N \\ & s(0) = g_0 - Bx(1) \\ & s(t) = Ax(t) - Bx(t+1), \quad t = 1, \dots, N-1 \\ & s(N) = Ax(N), \end{array}$$

with variables $x(1), \dots, x(N)$, $s(0), \dots, s(N)$. This problem is a standard form LP; the variables $s(t)$ are the slack variables associated with the constraints $Bx(t+1) \preceq Ax(t)$.

Chebyshev inequalities

We consider a probability distribution for a discrete random variable x on a set $\{u_1, \ldots, u_n\} \subseteq \mathbf{R}$ with n elements. We describe the distribution of x by a vector $p \in \mathbf{R}^n$, where

$$p_i = \mathbf{prob}(x = u_i),$$

so p satisfies $p \succeq 0$ and $\mathbf{1}^T p = 1$. Conversely, if p satisfies $p \succeq 0$ and $\mathbf{1}^T p = 1$, then it defines a probability distribution for x. We assume that u_i are known and fixed, but the distribution p is not known.

If f is any function of x, then

$$\mathbf{E}\, f = \sum_{i=1}^{n} p_i f(u_i)$$

is a linear function of p. If S is any subset of $\mathbf{R}$, then

$$\mathbf{prob}(x \in S) = \sum_{u_i \in S} p_i$$

is a linear function of p.

Although we do not know p, we are given prior knowledge of the following form: We know upper and lower bounds on expected values of some functions of x, and probabilities of some subsets of $\mathbf{R}$. This prior knowledge can be expressed as linear inequality constraints on p,

$$\alpha_i \leq a_i^T p \leq \beta_i, \quad i = 1, \ldots, m.$$

The problem is to give lower and upper bounds on $\mathbf{E}\, f_0(x) = a_0^T p$, where f_0 is some function of x.

To find a lower bound we solve the LP

$$\begin{array}{ll} \text{minimize} & a_0^T p \\ \text{subject to} & p \succeq 0, \quad \mathbf{1}^T p = 1 \\ & \alpha_i \leq a_i^T p \leq \beta_i, \quad i = 1, \ldots, m, \end{array}$$

with variable p. The optimal value of this LP gives the lowest possible value of $\mathbf{E}\, f_0(X)$ for any distribution that is consistent with the prior information. Moreover, the bound is sharp: the optimal solution gives a distribution that is consistent with the prior information and achieves the lower bound. In a similar way, we can find the best upper bound by maximizing $a_0^T p$ subject to the same constraints. (We will consider Chebyshev inequalities in more detail in §7.4.1.)

Piecewise-linear minimization

Consider the (unconstrained) problem of minimizing the piecewise-linear, convex function

$$f(x) = \max_{i=1,\ldots,m} (a_i^T x + b_i).$$

This problem can be transformed to an equivalent LP by first forming the epigraph problem,

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \max_{i=1,\ldots,m} (a_i^T x + b_i) \leq t, \end{array}$$

and then expressing the inequality as a set of m separate inequalities:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & a_i^T x + b_i \leq t, \quad i = 1, \ldots, m. \end{array}$$

This is an LP (in inequality form), with variables x and t.

4.3.2 Linear-fractional programming

The problem of minimizing a ratio of affine functions over a polyhedron is called a *linear-fractional program*:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{array} \tag{4.32}$$

where the objective function is given by

$$f_0(x) = \frac{c^T x + d}{e^T x + f}, \qquad \textbf{dom}\, f_0 = \{x \mid e^T x + f > 0\}.$$

The objective function is quasiconvex (in fact, quasilinear) so linear-fractional programs are quasiconvex optimization problems.

Transforming to a linear program

If the feasible set

$$\{x \mid Gx \preceq h,\ Ax = b,\ e^T x + f > 0\}$$

is nonempty, the linear-fractional program (4.32) can be transformed to an equivalent linear program

$$\begin{array}{ll} \text{minimize} & c^T y + dz \\ \text{subject to} & Gy - hz \preceq 0 \\ & Ay - bz = 0 \\ & e^T y + fz = 1 \\ & z \geq 0 \end{array} \tag{4.33}$$

with variables y, z.

To show the equivalence, we first note that if x is feasible in (4.32) then the pair

$$y = \frac{x}{e^T x + f}, \qquad z = \frac{1}{e^T x + f}$$

is feasible in (4.33), with the same objective value $c^T y + dz = f_0(x)$. It follows that the optimal value of (4.32) is greater than or equal to the optimal value of (4.33).

Conversely, if (y, z) is feasible in (4.33), with $z \neq 0$, then $x = y/z$ is feasible in (4.32), with the same objective value $f_0(x) = c^T y + dz$. If (y, z) is feasible in (4.33) with $z = 0$, and x_0 is feasible for (4.32), then $x = x_0 + ty$ is feasible in (4.32) for all $t \geq 0$. Moreover, $\lim_{t \to \infty} f_0(x_0 + ty) = c^T y + dz$, so we can find feasible points in (4.32) with objective values arbitrarily close to the objective value of (y, z). We conclude that the optimal value of (4.32) is less than or equal to the optimal value of (4.33).

Generalized linear-fractional programming

A generalization of the linear-fractional program (4.32) is the *generalized linear-fractional program* in which

$$f_0(x) = \max_{i=1,\ldots,r} \frac{c_i^T x + d_i}{e_i^T x + f_i}, \qquad \mathbf{dom}\, f_0 = \{x \mid e_i^T x + f_i > 0,\ i = 1, \ldots, r\}.$$

The objective function is the pointwise maximum of r quasiconvex functions, and therefore quasiconvex, so this problem is quasiconvex. When $r = 1$ it reduces to the standard linear-fractional program.

Example 4.7 *Von Neumann growth problem.* We consider an economy with n sectors, and activity levels $x_i > 0$ in the current period, and activity levels $x_i^+ > 0$ in the next period. (In this problem we only consider one period.) There are m goods which are consumed, and also produced, by the activity: An activity level x consumes goods $Bx \in \mathbf{R}^m$, and produces goods Ax. The goods consumed in the next period cannot exceed the goods produced in the current period, *i.e.*, $Bx^+ \preceq Ax$. The *growth rate* in sector i, over the period, is given by x_i^+/x_i.

Von Neumann's growth problem is to find an activity level vector x that maximizes the minimum growth rate across all sectors of the economy. This problem can be expressed as a generalized linear-fractional problem

$$\begin{array}{ll} \text{maximize} & \min_{i=1,\ldots,n} x_i^+/x_i \\ \text{subject to} & x^+ \succeq 0 \\ & Bx^+ \preceq Ax \end{array}$$

with domain $\{(x, x^+) \mid x \succ 0\}$. Note that this problem is homogeneous in x and x^+, so we can replace the implicit constraint $x \succ 0$ by the explicit constraint $x \succeq \mathbf{1}$.

4.4 Quadratic optimization problems

The convex optimization problem (4.15) is called a *quadratic program* (QP) if the objective function is (convex) quadratic, and the constraint functions are affine. A quadratic program can be expressed in the form

$$\begin{array}{ll} \text{minimize} & (1/2)x^T Px + q^T x + r \\ \text{subject to} & Gx \preceq h \\ & Ax = b, \end{array} \tag{4.34}$$

where $P \in \mathbf{S}_+^n$, $G \in \mathbf{R}^{m\times n}$, and $A \in \mathbf{R}^{p\times n}$. In a quadratic program, we minimize a convex quadratic function over a polyhedron, as illustrated in figure 4.5.

If the objective in (4.15) as well as the inequality constraint functions are (convex) quadratic, as in

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P_0 x + q_0^T x + r_0 \\ \text{subject to} & (1/2)x^T P_i x + q_i^T x + r_i \le 0, \quad i = 1, \ldots, m \\ & Ax = b, \end{array} \tag{4.35}$$

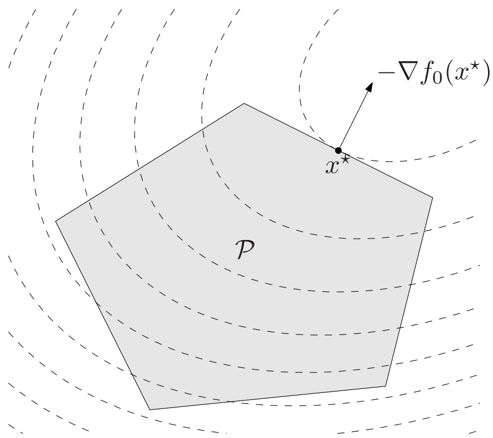


Figure 4.5 Geometric illustration of QP. The feasible set $\mathcal{P}$, which is a polyhedron, is shown shaded. The contour lines of the objective function, which is convex quadratic, are shown as dashed curves. The point $x^\star$ is optimal.

where $P_i \in \mathbf{S}^n_+$, $i = 0, 1 \ldots, m$, the problem is called a *quadratically constrained quadratic program* (QCQP). In a QCQP, we minimize a convex quadratic function over a feasible region that is the intersection of ellipsoids (when $P_i \succ 0$).

Quadratic programs include linear programs as a special case, by taking $P = 0$ in (4.34). Quadratically constrained quadratic programs include quadratic programs (and therefore also linear programs) as a special case, by taking $P_i = 0$ in (4.35), for $i = 1, \ldots, m$.

4.4.1 Examples

Least-squares and regression

The problem of minimizing the convex quadratic function

$$\|Ax - b\|_2^2 = x^T A^T A x - 2b^T A x + b^T b$$

is an (unconstrained) QP. It arises in many fields and has many names, *e.g.*, *regression analysis* or *least-squares approximation*. This problem is simple enough to have the well known analytical solution $x = A^\dagger b$, where $A^\dagger$ is the pseudo-inverse of A (see §A.5.4).

When linear inequality constraints are added, the problem is called *constrained regression* or *constrained least-squares*, and there is no longer a simple analytical solution. As an example we can consider regression with lower and upper bounds on the variables, *i.e.*,

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|_2^2 \\ \text{subject to} & l_i \leq x_i \leq u_i, \quad i = 1, \ldots, n, \end{array}$$

which is a QP. (We will study least-squares and regression problems in far more depth in chapters 6 and 7.)

Distance between polyhedra

The (Euclidean) distance between the polyhedra $\mathcal{P}_1 = \{x \mid A_1x \preceq b_1\}$ and $\mathcal{P}_2 = \{x \mid A_2x \preceq b_2\}$ in $\mathbf{R}^n$ is defined as

$$\mathbf{dist}(\mathcal{P}_1, \mathcal{P}_2) = \inf\{\|x_1 - x_2\|_2 \mid x_1 \in \mathcal{P}_1,\ x_2 \in \mathcal{P}_2\}.$$

If the polyhedra intersect, the distance is zero.

To find the distance between $\mathcal{P}_1$ and $\mathcal{P}_2$, we can solve the QP

$$\begin{array}{ll} \text{minimize} & \|x_1 - x_2\|_2^2 \\ \text{subject to} & A_1x_1 \preceq b_1, \quad A_2x_2 \preceq b_2, \end{array}$$

with variables $x_1,\ x_2 \in \mathbf{R}^n$. This problem is infeasible if and only if one of the polyhedra is empty. The optimal value is zero if and only if the polyhedra intersect, in which case the optimal x_1 and x_2 are equal (and is a point in the intersection $\mathcal{P}_1 \cap \mathcal{P}_2$). Otherwise the optimal x_1 and x_2 are the points in $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively, that are closest to each other. (We will study geometric problems involving distance in more detail in chapter 8.)

Bounding variance

We consider again the Chebyshev inequalities example (page 150), where the variable is an unknown probability distribution given by $p \in \mathbf{R}^n$, about which we have some prior information. The variance of a random variable $f(x)$ is given by

$$\mathbf{E}\, f^2 - (\mathbf{E}\, f)^2 = \sum_{i=1}^n f_i^2 p_i - \left(\sum_{i=1}^n f_i p_i\right)^2,$$

(where $f_i = f(u_i)$), which is a concave quadratic function of p.

It follows that we can maximize the variance of $f(x)$, subject to the given prior information, by solving the QP

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^n f_i^2 p_i - \left(\sum_{i=1}^n f_i p_i\right)^2 \\ \text{subject to} & p \succeq 0, \quad \mathbf{1}^T p = 1 \\ & \alpha_i \leq a_i^T p \leq \beta_i, \quad i = 1, \ldots, m. \end{array}$$

The optimal value gives the maximum possible variance of $f(x)$, over all distributions that are consistent with the prior information; the optimal p gives a distribution that achieves this maximum variance.

Linear program with random cost

We consider an LP,

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Gx \preceq h \\ & Ax = b, \end{array}$$

with variable $x \in \mathbf{R}^n$. We suppose that the cost function (vector) $c \in \mathbf{R}^n$ is *random*, with mean value $\bar{c}$ and covariance $\mathbf{E}(c - \bar{c})(c - \bar{c})^T = \Sigma$. (We assume for simplicity that the other problem parameters are deterministic.) For a given $x \in \mathbf{R}^n$, the cost $c^T x$ is a (scalar) random variable with mean $\mathbf{E}\, c^T x = \bar{c}^T x$ and variance

$$\mathbf{var}(c^T x) = \mathbf{E}(c^T x - \mathbf{E}\, c^T x)^2 = x^T \Sigma x.$$

In general there is a trade-off between small expected cost and small cost variance. One way to take variance into account is to minimize a linear combination of the expected value and the variance of the cost, *i.e.*,

$$\mathbf{E}\, c^T x + \gamma\, \mathbf{var}(c^T x),$$

which is called the *risk-sensitive cost*. The parameter $\gamma \geq 0$ is called the *risk-aversion parameter*, since it sets the relative values of cost variance and expected value. (For $\gamma > 0$, we are willing to trade off an increase in expected cost for a sufficiently large decrease in cost variance).

To minimize the risk-sensitive cost we solve the QP

$$\begin{array}{ll} \text{minimize} & \bar{c}^T x + \gamma x^T \Sigma x \\ \text{subject to} & Gx \preceq h \\ & Ax = b. \end{array}$$

Markowitz portfolio optimization

We consider a classical portfolio problem with n assets or stocks held over a period of time. We let x_i denote the amount of asset i held throughout the period, with x_i in dollars, at the price at the beginning of the period. A normal long position in asset i corresponds to $x_i > 0$; a short position in asset i (*i.e.*, the obligation to buy the asset at the end of the period) corresponds to $x_i < 0$. We let p_i denote the relative price change of asset i over the period, *i.e.*, its change in price over the period divided by its price at the beginning of the period. The overall return on the portfolio is $r = p^T x$ (given in dollars). The optimization variable is the portfolio vector $x \in \mathbf{R}^n$.

A wide variety of constraints on the portfolio can be considered. The simplest set of constraints is that $x_i \geq 0$ (*i.e.*, no short positions) and $\mathbf{1}^T x = B$ (*i.e.*, the total budget to be invested is B, which is often taken to be one).

We take a stochastic model for price changes: $p \in \mathbf{R}^n$ is a random vector, with known mean $\bar{p}$ and covariance Σ. Therefore with portfolio $x \in \mathbf{R}^n$, the return r is a (scalar) random variable with mean $\bar{p}^T x$ and variance $x^T \Sigma x$. The choice of portfolio x involves a trade-off between the mean of the return, and its variance.

The classical portfolio optimization problem, introduced by Markowitz, is the QP

$$\begin{array}{ll} \text{minimize} & x^T \Sigma x \\ \text{subject to} & \bar{p}^T x \geq r_{\min} \\ & \mathbf{1}^T x = 1, \quad x \succeq 0, \end{array}$$

where x, the portfolio, is the variable. Here we find the portfolio that minimizes the return variance (which is associated with the *risk* of the portfolio) subject to achieving a minimum acceptable mean return $r_{\min}$, and satisfying the portfolio budget and no-shorting constraints.

Many extensions are possible. One standard extension, for example, is to allow short positions, *i.e.*, $x_i < 0$. To do this we introduce variables x_{long} and x_{short}, with

$$x_{\text{long}} \succeq 0, \qquad x_{\text{short}} \succeq 0, \qquad x = x_{\text{long}} - x_{\text{short}}, \qquad \mathbf{1}^T x_{\text{short}} \leq \eta \mathbf{1}^T x_{\text{long}}.$$

The last constraint limits the total short position at the beginning of the period to some fraction η of the total long position at the beginning of the period.

As another extension we can include linear transaction costs in the portfolio optimization problem. Starting from a given initial portfolio x_{init} we buy and sell assets to achieve the portfolio x, which we then hold over the period as described above. We are charged a transaction fee for buying and selling assets, which is proportional to the amount bought or sold. To handle this, we introduce variables u_{buy} and u_{sell}, which determine the amount of each asset we buy and sell before the holding period. We have the constraints

$$x = x_{\text{init}} + u_{\text{buy}} - u_{\text{sell}}, \qquad u_{\text{buy}} \succeq 0, \qquad u_{\text{sell}} \succeq 0.$$

We replace the simple budget constraint $\mathbf{1}^T x = 1$ with the condition that the initial buying and selling, including transaction fees, involves zero net cash:

$$(1 - f_{\text{sell}})\mathbf{1}^T u_{\text{sell}} = (1 + f_{\text{buy}})\mathbf{1}^T u_{\text{buy}}$$

Here the lefthand side is the total proceeds from selling assets, less the selling transaction fee, and the righthand side is the total cost, including transaction fee, of buying assets. The constants $f_{\text{buy}} \geq 0$ and $f_{\text{sell}} \geq 0$ are the transaction fee rates for buying and selling (assumed the same across assets, for simplicity).

The problem of minimizing return variance, subject to a minimum mean return, and the budget and trading constraints, is a QP with variables x, u_{buy}, u_{sell}.

4.4.2 Second-order cone programming

A problem that is closely related to quadratic programming is the *second-order cone program* (SOCP):

$$\begin{array}{ll} \text{minimize} & f^T x \\ \text{subject to} & \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, m \\ & Fx = g, \end{array} \tag{4.36}$$

where $x \in \mathbf{R}^n$ is the optimization variable, $A_i \in \mathbf{R}^{n_i \times n}$, and $F \in \mathbf{R}^{p \times n}$. We call a constraint of the form

$$\|Ax + b\|_2 \leq c^T x + d,$$

where $A \in \mathbf{R}^{k \times n}$, a *second-order cone constraint*, since it is the same as requiring the affine function $(Ax + b, c^T x + d)$ to lie in the second-order cone in $\mathbf{R}^{k+1}$.

When $c_i = 0$, $i = 1, \ldots, m$, the SOCP (4.36) is equivalent to a QCQP (which is obtained by squaring each of the constraints). Similarly, if $A_i = 0$, $i = 1, \ldots, m$, then the SOCP (4.36) reduces to a (general) LP. Second-order cone programs are, however, more general than QCQPs (and of course, LPs).

Robust linear programming

We consider a linear program in inequality form,

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \quad i = 1, \ldots, m, \end{array}$$

in which there is some uncertainty or variation in the parameters c, a_i, b_i. To simplify the exposition we assume that c and b_i are fixed, and that a_i are known to lie in given ellipsoids:

$$a_i \in \mathcal{E}_i = \{\overline{a}_i + P_i u \mid \|u\|_2 \leq 1\},$$

where $P_i \in \mathbf{R}^{n \times n}$. (If P_i is singular we obtain 'flat' ellipsoids, of dimension $\mathbf{rank}\, P_i$; $P_i = 0$ means that a_i is known perfectly.)

We will require that the constraints be satisfied for all possible values of the parameters a_i, which leads us to the *robust linear program*

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i \text{ for all } a_i \in \mathcal{E}_i, \quad i = 1, \ldots, m. \end{array} \tag{4.37}$$

The robust linear constraint, $a_i^T x \leq b_i$ for all $a_i \in \mathcal{E}_i$, can be expressed as

$$\sup\{a_i^T x \mid a_i \in \mathcal{E}_i\} \leq b_i,$$

the lefthand side of which can be expressed as

$$\begin{array}{rcl} \sup\{a_i^T x \mid a_i \in \mathcal{E}_i\} & = & \overline{a}_i^T x + \sup\{u^T P_i^T x \mid \|u\|_2 \leq 1\} \\ & = & \overline{a}_i^T x + \|P_i^T x\|_2. \end{array}$$

Thus, the robust linear constraint can be expressed as

$$\overline{a}_i^T x + \|P_i^T x\|_2 \leq b_i,$$

which is evidently a second-order cone constraint. Hence the robust LP (4.37) can be expressed as the SOCP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \overline{a}_i^T x + \|P_i^T x\|_2 \leq b_i, \quad i = 1, \ldots, m. \end{array}$$

Note that the additional norm terms act as *regularization terms*; they prevent x from being large in directions with considerable uncertainty in the parameters a_i.

Linear programming with random constraints

The robust LP described above can also be considered in a statistical framework. Here we suppose that the parameters a_i are independent Gaussian random vectors, with mean $\overline{a}_i$ and covariance Σ_i. We require that each constraint $a_i^T x \leq b_i$ should hold with a probability (or confidence) exceeding η, where $\eta \geq 0.5$, *i.e.*,

$$\mathbf{prob}(a_i^T x \leq b_i) \geq \eta. \tag{4.38}$$

We will show that this probability constraint can be expressed as a second-order cone constraint.

Letting $u = a_i^T x$, with σ^2 denoting its variance, this constraint can be written as

$$\mathbf{prob}\left(\frac{u-\overline{u}}{\sigma} \leq \frac{b_i - \overline{u}}{\sigma}\right) \geq \eta.$$

Since $(u-\overline{u})/\sigma$ is a zero mean unit variance Gaussian variable, the probability above is simply $\Phi((b_i-\overline{u})/\sigma)$, where

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt$$

is the cumulative distribution function of a zero mean unit variance Gaussian random variable. Thus the probability constraint (4.38) can be expressed as

$$\frac{b_i - \overline{u}}{\sigma} \geq \Phi^{-1}(\eta),$$

or, equivalently,

$$\overline{u} + \Phi^{-1}(\eta)\sigma \leq b_i.$$

From $\overline{u} = \overline{a}_i^T x$ and $\sigma = (x^T \Sigma_i x)^{1/2}$ we obtain

$$\overline{a}_i^T x + \Phi^{-1}(\eta)\|\Sigma_i^{1/2} x\|_2 \leq b_i.$$

By our assumption that $\eta \geq 1/2$, we have $\Phi^{-1}(\eta) \geq 0$, so this constraint is a second-order cone constraint.

In summary, the problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \mathbf{prob}(a_i^T x \leq b_i) \geq \eta, \quad i = 1, \ldots, m \end{array}$$

can be expressed as the SOCP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \overline{a}_i^T x + \Phi^{-1}(\eta)\|\Sigma_i^{1/2} x\|_2 \leq b_i, \quad i = 1, \ldots, m. \end{array}$$

(We will consider robust convex optimization problems in more depth in chapter 6. See also exercises 4.13, 4.28, and 4.59.)

Example 4.8 *Portfolio optimization with loss risk constraints.* We consider again the classical Markowitz portfolio problem described above (page 155). We assume here that the price change vector $p \in \mathbf{R}^n$ is a Gaussian random variable, with mean $\overline{p}$ and covariance Σ. Therefore the return r is a Gaussian random variable with mean $\overline{r} = \overline{p}^T x$ and variance $\sigma_r^2 = x^T \Sigma x$.

Consider a *loss risk constraint* of the form

$$\mathbf{prob}(r \leq \alpha) \leq \beta, \tag{4.39}$$

where α is a given unwanted return level (*e.g.*, a large loss) and β is a given maximum probability.

As in the stochastic interpretation of the robust LP given above, we can express this constraint using the cumulative distribution function Φ of a unit Gaussian random variable. The inequality (4.39) is equivalent to

$$\bar{p}^T x + \Phi^{-1}(\beta)\,\|\Sigma^{1/2}x\|_2 \geq \alpha.$$

Provided $\beta \leq 1/2$ (*i.e.*, $\Phi^{-1}(\beta) \leq 0$), this loss risk constraint is a second-order cone constraint. (If $\beta > 1/2$, the loss risk constraint becomes nonconvex in x.)

The problem of maximizing the expected return subject to a bound on the loss risk (with $\beta \leq 1/2$), can therefore be cast as an SOCP with one second-order cone constraint:

$$\begin{array}{ll} \text{maximize} & \bar{p}^T x \\ \text{subject to} & \bar{p}^T x + \Phi^{-1}(\beta)\,\|\Sigma^{1/2}x\|_2 \geq \alpha \\ & x \succeq 0, \quad \mathbf{1}^T x = 1. \end{array}$$

There are many extensions of this problem. For example, we can impose several loss risk constraints, *i.e.*,

$$\mathbf{prob}(r \leq \alpha_i) \leq \beta_i, \quad i = 1, \ldots, k,$$

(where $\beta_i \leq 1/2$), which expresses the risks (β_i) we are willing to accept for various levels of loss (α_i).

Minimal surface

Consider a differentiable function $f : \mathbf{R}^2 \to \mathbf{R}$ with $\mathbf{dom}\, f = C$. The surface area of its graph is given by

$$A = \int_C \sqrt{1 + \|\nabla f(x)\|_2^2}\, dx = \int_C \|(\nabla f(x), 1)\|_2\, dx,$$

which is a convex functional of f. The *minimal surface problem* is to find the function f that minimizes A subject to some constraints, for example, some given values of f on the boundary of C.

We will approximate this problem by discretizing the function f. Let $C = [0,1] \times [0,1]$, and let f_{ij} denote the value of f at the point $(i/K, j/K)$, for $i,\ j = 0, \ldots, K$. An approximate expression for the gradient of f at the point $x = (i/K, j/K)$ can be found using forward differences:

$$\nabla f(x) \approx K \begin{bmatrix} f_{i+1,j} - f_{i,j} \\ f_{i,j+1} - f_{i,j} \end{bmatrix}.$$

Substituting this into the expression for the area of the graph, and approximating the integral as a sum, we obtain an approximation for the area of the graph:

$$A \approx A_{\text{disc}} = \frac{1}{K^2} \sum_{i,j=0}^{K-1} \left\| \begin{bmatrix} K(f_{i+1,j} - f_{i,j}) \\ K(f_{i,j+1} - f_{i,j}) \\ 1 \end{bmatrix} \right\|_2$$

The discretized area approximation A_{disc} is a convex function of f_{ij}.

We can consider a wide variety of constraints on f_{ij}, such as equality or inequality constraints on any of its entries (for example, on the boundary values), or on its moments. As an example, we consider the problem of finding the minimal area surface with fixed boundary values on the left and right edges of the square:

$$\begin{array}{ll} \text{minimize} & A_{\text{disc}} \\ \text{subject to} & f_{0j} = l_j, \quad j = 0, \ldots, K \\ & f_{Kj} = r_j, \quad j = 0, \ldots, K \end{array} \tag{4.40}$$

where f_{ij}, $i, j = 0, \ldots, K$, are the variables, and l_j, r_j are the given boundary values on the left and right sides of the square.

We can transform the problem (4.40) into an SOCP by introducing new variables t_{ij}, $i,\ j = 0, \ldots, K-1$:

$$\begin{array}{ll} \text{minimize} & (1/K^2) \sum_{i,j=0}^{K-1} t_{ij} \\ \text{subject to} & \left\| \begin{bmatrix} K(f_{i+1,j} - f_{i,j}) \\ K(f_{i,j+1} - f_{i,j}) \\ 1 \end{bmatrix} \right\|_2 \leq t_{ij}, \quad i,\ j = 0, \ldots, K-1 \\ & f_{0j} = l_j, \quad j = 0, \ldots, K \\ & f_{Kj} = r_j, \quad j = 0, \ldots, K. \end{array}$$

4.5 Geometric programming

In this section we describe a family of optimization problems that are *not* convex in their natural form. These problems can, however, be transformed to convex optimization problems, by a change of variables and a transformation of the objective and constraint functions.

4.5.1 Monomials and posynomials

A function $f : \mathbf{R}^n \to \mathbf{R}$ with $\mathbf{dom}\, f = \mathbf{R}^n_{++}$, defined as

$$f(x) = c x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}, \tag{4.41}$$

where $c > 0$ and $a_i \in \mathbf{R}$, is called a *monomial function*, or simply, a *monomial*. The exponents a_i of a monomial can be any real numbers, including fractional or negative, but the coefficient c can only be positive. (The term 'monomial' conflicts with the standard definition from algebra, in which the exponents must be nonnegative integers, but this should not cause any confusion.) A sum of monomials, *i.e.*, a function of the form

$$f(x) = \sum_{k=1}^{K} c_k x_1^{a_{1k}} x_2^{a_{2k}} \cdots x_n^{a_{nk}}, \tag{4.42}$$

where $c_k > 0$, is called a *posynomial function* (with K terms), or simply, a *posynomial*.

Posynomials are closed under addition, multiplication, and nonnegative scaling. Monomials are closed under multiplication and division. If a posynomial is multiplied by a monomial, the result is a posynomial; similarly, a posynomial can be divided by a monomial, with the result a posynomial.

4.5.2 Geometric programming

An optimization problem of the form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 1, \quad i = 1, \ldots, m \\ & h_i(x) = 1, \quad i = 1, \ldots, p \end{array} \tag{4.43}$$

where $f_0, \ldots, f_m$ are posynomials and $h_1, \ldots, h_p$ are monomials, is called a *geometric program* (GP). The domain of this problem is $\mathcal{D} = \mathbf{R}^n_{++}$; the constraint $x \succ 0$ is implicit.

Extensions of geometric programming

Several extensions are readily handled. If f is a posynomial and h is a monomial, then the constraint $f(x) \leq h(x)$ can be handled by expressing it as $f(x)/h(x) \leq 1$ (since f/h is posynomial). This includes as a special case a constraint of the form $f(x) \leq a$, where f is posynomial and $a > 0$. In a similar way if h_1 and h_2 are both nonzero monomial functions, then we can handle the equality constraint $h_1(x) = h_2(x)$ by expressing it as $h_1(x)/h_2(x) = 1$ (since h_1/h_2 is monomial). We can maximize a nonzero monomial objective function, by minimizing its inverse (which is also a monomial).

For example, consider the problem

$$\begin{array}{ll} \text{maximize} & x/y \\ \text{subject to} & 2 \leq x \leq 3 \\ & x^2 + 3y/z \leq \sqrt{y} \\ & x/y = z^2, \end{array}$$

with variables $x,\ y,\ z \in \mathbf{R}$ (and the implicit constraint $x,\ y,\ z > 0$). Using the simple transformations described above, we obtain the equivalent standard form GP

$$\begin{array}{ll} \text{minimize} & x^{-1}y \\ \text{subject to} & 2x^{-1} \leq 1, \quad (1/3)x \leq 1 \\ & x^2y^{-1/2} + 3y^{1/2}z^{-1} \leq 1 \\ & xy^{-1}z^{-2} = 1. \end{array}$$

We will refer to a problem like this one, that is easily transformed to an equivalent GP in the standard form (4.43), also as a GP. (In the same way that we refer to a problem easily transformed to an LP as an LP.)

4.5.3 Geometric program in convex form

Geometric programs are not (in general) convex optimization problems, but they can be transformed to convex problems by a change of variables and a transformation of the objective and constraint functions.

We will use the variables defined as $y_i = \log x_i$, so $x_i = e^{y_i}$. If f is the monomial function of x given in (4.41), *i.e.*,

$$f(x) = c x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n},$$

then

$$\begin{aligned} f(x) &= f(e^{y_1}, \ldots, e^{y_n}) \\ &= c(e^{y_1})^{a_1} \cdots (e^{y_n})^{a_n} \\ &= e^{a^T y + b}, \end{aligned}$$

where $b = \log c$. The change of variables $y_i = \log x_i$ turns a monomial function into the exponential of an affine function.

Similarly, if f is the posynomial given by (4.42), *i.e.*,

$$f(x) = \sum_{k=1}^{K} c_k x_1^{a_{1k}} x_2^{a_{2k}} \cdots x_n^{a_{nk}},$$

then

$$f(x) = \sum_{k=1}^{K} e^{a_k^T y + b_k},$$

where $a_k = (a_{1k}, \ldots, a_{nk})$ and $b_k = \log c_k$. After the change of variables, a posynomial becomes a sum of exponentials of affine functions.

The geometric program (4.43) can be expressed in terms of the new variable y as

$$\begin{array}{ll} \text{minimize} & \sum_{k=1}^{K_0} e^{a_{0k}^T y + b_{0k}} \\ \text{subject to} & \sum_{k=1}^{K_i} e^{a_{ik}^T y + b_{ik}} \leq 1, \quad i = 1, \ldots, m \\ & e^{g_i^T y + h_i} = 1, \quad i = 1, \ldots, p, \end{array}$$

where $a_{ik} \in \mathbf{R}^n$, $i = 0, \ldots, m$, contain the exponents of the posynomial inequality constraints, and $g_i \in \mathbf{R}^n$, $i = 1, \ldots, p$, contain the exponents of the monomial equality constraints of the original geometric program.

Now we transform the objective and constraint functions, by taking the logarithm. This results in the problem

$$\begin{array}{ll} \text{minimize} & \tilde{f}_0(y) = \log \left(\sum_{k=1}^{K_0} e^{a_{0k}^T y + b_{0k}} \right) \\ \text{subject to} & \tilde{f}_i(y) = \log \left(\sum_{k=1}^{K_i} e^{a_{ik}^T y + b_{ik}} \right) \leq 0, \quad i = 1, \ldots, m \\ & \tilde{h}_i(y) = g_i^T y + h_i = 0, \quad i = 1, \ldots, p. \end{array} \tag{4.44}$$

Since the functions $\tilde{f}_i$ are convex, and $\tilde{h}_i$ are affine, this problem is a convex optimization problem. We refer to it as a *geometric program in convex form*. To distinguish it from the original geometric program, we refer to (4.43) as a *geometric program in posynomial form*.

Note that the transformation between the posynomial form geometric program (4.43) and the convex form geometric program (4.44) does not involve any computation; the problem data for the two problems are the same. It simply changes the form of the objective and constraint functions.

If the posynomial objective and constraint functions all have only one term, *i.e.*, are monomials, then the convex form geometric program (4.44) reduces to a (general) linear program. We can therefore consider geometric programming to be a generalization, or extension, of linear programming.

4.5.4 Examples

Frobenius norm diagonal scaling

Consider a matrix $M \in \mathbf{R}^{n\times n}$, and the associated linear function that maps u into $y = Mu$. Suppose we scale the coordinates, *i.e.*, change variables to $\tilde{u} = Du$, $\tilde{y} = Dy$, where D is diagonal, with $D_{ii} > 0$. In the new coordinates the linear function is given by $\tilde{y} = DMD^{-1}\tilde{u}$.

Now suppose we want to choose the scaling in such a way that the resulting matrix, DMD^{-1}, is small. We will use the Frobenius norm (squared) to measure the size of the matrix:

$$\begin{aligned} \|DMD^{-1}\|_F^2 &= \mathbf{tr}\left(\left(DMD^{-1}\right)^T \left(DMD^{-1}\right)\right) \\ &= \sum_{i,j=1}^{n} \left(DMD^{-1}\right)_{ij}^2 \\ &= \sum_{i,j=1}^{n} M_{ij}^2 d_i^2/d_j^2, \end{aligned}$$

where $D = \mathbf{diag}(d)$. Since this is a posynomial in d, the problem of choosing the scaling d to minimize the Frobenius norm is an unconstrained geometric program,

$$\text{minimize} \quad \textstyle\sum_{i,j=1}^{n} M_{ij}^2 d_i^2/d_j^2,$$

with variable d. The only exponents in this geometric program are 0, 2, and -2.

Design of a cantilever beam

We consider the design of a cantilever beam, which consists of N segments, numbered from right to left as $1, \ldots, N$, as shown in figure 4.6. Each segment has unit length and a uniform rectangular cross-section with width w_i and height h_i. A vertical load (force) F is applied at the right end of the beam. This load causes the beam to deflect (downward), and induces stress in each segment of the beam. We assume that the deflections are small, and that the material is linearly elastic, with Young's modulus E.

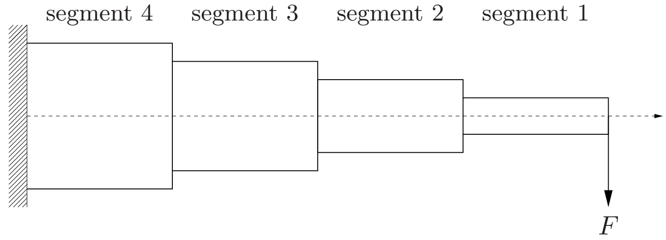


Figure 4.6 Segmented cantilever beam with 4 segments. Each segment has unit length and a rectangular profile. A vertical force F is applied at the right end of the beam.

The design variables in the problem are the widths w_i and heights h_i of the N segments. We seek to minimize the total volume of the beam (which is proportional to its weight),

$$w_1 h_1 + \cdots + w_N h_N,$$

subject to some design constraints. We impose upper and lower bounds on width and height of the segments,

$$w_{\min} \leq w_i \leq w_{\max}, \quad h_{\min} \leq h_i \leq h_{\max}, \quad i = 1, \ldots, N,$$

as well as the aspect ratios,

$$S_{\min} \leq h_i / w_i \leq S_{\max}.$$

In addition, we have a limit on the maximum allowable stress in the material, and on the vertical deflection at the end of the beam.

We first consider the maximum stress constraint. The maximum stress in segment i, which we denote σ_i, is given by $\sigma_i = 6iF/(w_i h_i^2)$. We impose the constraints

$$\frac{6iF}{w_i h_i^2} \leq \sigma_{\max}, \quad i = 1, \ldots, N,$$

to ensure that the stress does not exceed the maximum allowable value $\sigma_{\max}$ anywhere in the beam.

The last constraint is a limit on the vertical deflection at the end of the beam, which we will denote y_1:

$$y_1 \leq y_{\max}.$$

The deflection y_1 can be found by a recursion that involves the deflection and slope of the beam segments:

$$v_i = 12(i - 1/2) \frac{F}{E w_i h_i^3} + v_{i+1}, \qquad y_i = 6(i - 1/3) \frac{F}{E w_i h_i^3} + v_{i+1} + y_{i+1}, \quad (4.45)$$

for $i = N, N-1, \ldots, 1$, with starting values $v_{N+1} = y_{N+1} = 0$. In this recursion, y_i is the deflection at the right end of segment i, and v_i is the slope at that point. We can use the recursion (4.45) to show that these deflection and slope quantities are in fact posynomial functions of the variables w and h. We first note that v_{N+1} and y_{N+1} are zero, and therefore posynomials. Now assume that v_{i+1} and y_{i+1} are posynomial functions of w and h. The lefthand equation in (4.45) shows that v_i is the sum of a monomial and a posynomial (*i.e.*, v_{i+1}), and therefore is a posynomial. From the righthand equation in (4.45), we see that the deflection y_i is the sum of a monomial and two posynomials (v_{i+1} and y_{i+1}), and so is a posynomial. In particular, the deflection at the end of the beam, y_1, is a posynomial.

The problem is then

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^N w_i h_i \\ \text{subject to} & w_{\min} \le w_i \le w_{\max}, \quad i=1,\dots,N \\ & h_{\min} \le h_i \le h_{\max}, \quad i=1,\dots,N \\ & S_{\min} \le h_i/w_i \le S_{\max}, \quad i=1,\dots,N \\ & 6iF/(w_i h_i^2) \le \sigma_{\max}, \quad i=1,\dots,N \\ & y_1 \le y_{\max}, \end{array} \tag{4.46}$$

with variables w and h. This is a GP, since the objective is a posynomial, and the constraints can all be expressed as posynomial inequalities. (In fact, the constraints can be all be expressed as monomial inequalities, with the exception of the deflection limit, which is a complicated posynomial inequality.)

When the number of segments N is large, the number of monomial terms appearing in the posynomial y_1 grows approximately as N^2. Another formulation of this problem, explored in exercise 4.31, is obtained by introducing $v_1,\dots,v_N$ and $y_1,\dots,y_N$ as variables, and including a modified version of the recursion as a set of constraints. This formulation avoids this growth in the number of monomial terms.

Minimizing spectral radius via Perron-Frobenius theory

Suppose the matrix $A \in \mathbf{R}^{n\times n}$ is elementwise nonnegative, *i.e.*, $A_{ij} \ge 0$ for $i,j = 1,\dots,n$, and irreducible, which means that the matrix $(I+A)^{n-1}$ is elementwise positive. The Perron-Frobenius theorem states that A has a positive real eigenvalue $\lambda_{\rm pf}$ equal to its spectral radius, *i.e.*, the largest magnitude of its eigenvalues. The Perron-Frobenius eigenvalue $\lambda_{\rm pf}$ determines the asymptotic rate of growth or decay of A^k, as $k \to \infty$; in fact, the matrix $((1/\lambda_{\rm pf})A)^k$ converges. Roughly speaking, this means that as $k\to\infty$, A^k grows like $\lambda_{\rm pf}^k$, if $\lambda_{\rm pf} > 1$, or decays like $\lambda_{\rm pf}^k$, if $\lambda_{\rm pf} < 1$.

A basic result in the theory of nonnegative matrices states that the Perron-Frobenius eigenvalue is given by

$$\lambda_{\rm pf} = \inf\{\lambda \mid Av \preceq \lambda v \text{ for some } v \succ 0\}$$

(and moreover, that the infimum is achieved). The inequality $Av \preceq \lambda v$ can be expressed as

$$\sum_{j=1}^n A_{ij} v_j/(\lambda v_i) \le 1, \quad i=1,\dots,n, \tag{4.47}$$

which is a set of posynomial inequalities in the variables A_{ij}, v_i, and λ. Thus, the condition that $\lambda_{\rm pf} \le \lambda$ can be expressed as a set of posynomial inequalities in A, v, and λ. This allows us to solve some optimization problems involving the Perron-Frobenius eigenvalue using geometric programming.

Suppose that the entries of the matrix A are posynomial functions of some underlying variable $x \in \mathbf{R}^k$. In this case the inequalities (4.47) are posynomial inequalities in the variables $x \in \mathbf{R}^k$, $v \in \mathbf{R}^n$, and $\lambda \in \mathbf{R}$. We consider the problem of choosing x to minimize the Perron-Frobenius eigenvalue (or spectral radius) of A, possibly subject to posynomial inequalities on x,

$$\begin{array}{ll} \text{minimize} & \lambda_{\text{pf}}(A(x)) \\ \text{subject to} & f_i(x) \leq 1, \quad i = 1, \ldots, p, \end{array}$$

where f_i are posynomials. Using the characterization above, we can express this problem as the GP

$$\begin{array}{ll} \text{minimize} & \lambda \\ \text{subject to} & \sum_{j=1}^n A_{ij} v_j / (\lambda v_i) \leq 1, \quad i = 1, \ldots, n \\ & f_i(x) \leq 1, \quad i = 1, \ldots, p, \end{array}$$

where the variables are x, v, and λ.

As a specific example, we consider a simple model for the population dynamics for a bacterium, with time or period denoted by $t = 0, 1, 2, \ldots$, in hours. The vector $p(t) \in \mathbf{R}^4_+$ characterizes the population age distribution at period t: $p_1(t)$ is the total population between 0 and 1 hours old; $p_2(t)$ is the total population between 1 and 2 hours old; and so on. We (arbitrarily) assume that no bacteria live more than 4 hours. The population propagates in time as $p(t+1) = Ap(t)$, where

$$A = \begin{bmatrix} b_1 & b_2 & b_3 & b_4 \\ s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & s_3 & 0 \end{bmatrix}.$$

Here b_i is the birth rate among bacteria in age group i, and s_i is the survival rate from age group i into age group $i+1$. We assume that $b_i > 0$ and $0 < s_i < 1$, which implies that the matrix A is irreducible.

The Perron-Frobenius eigenvalue of A determines the asymptotic growth or decay rate of the population. If $\lambda_{\text{pf}} < 1$, the population converges to zero like λ_{pf}^t, and so has a half-life of $-1/\log_2 \lambda_{\text{pf}}$ hours. If $\lambda_{\text{pf}} > 1$ the population grows geometrically like λ_{pf}^t, with a doubling time of $1/\log_2 \lambda_{\text{pf}}$ hours. Minimizing the spectral radius of A corresponds to finding the fastest decay rate, or slowest growth rate, for the population.

As our underlying variables, on which the matrix A depends, we take c_1 and c_2, the concentrations of two chemicals in the environment that affect the birth and survival rates of the bacteria. We model the birth and survival rates as monomial functions of the two concentrations:

$$\begin{array}{rcl} b_i & = & b_i^{\text{nom}} (c_1/c_1^{\text{nom}})^{\alpha_i} (c_2/c_2^{\text{nom}})^{\beta_i}, \quad i = 1, \ldots, 4 \\ s_i & = & s_i^{\text{nom}} (c_1/c_1^{\text{nom}})^{\gamma_i} (c_2/c_2^{\text{nom}})^{\delta_i}, \quad i = 1, \ldots, 3. \end{array}$$

Here, b_i^{nom} is nominal birth rate, s_i^{nom} is nominal survival rate, and c_i^{nom} is nominal concentration of chemical i. The constants α_i, β_i, γ_i, and δ_i give the effect on the birth and survival rates due to changes in the concentrations of the chemicals away from the nominal values. For example $\alpha_2 = -0.3$ and $\gamma_1 = 0.5$ means that an increase in concentration of chemical 1, over the nominal concentration, causes a decrease in the birth rate of bacteria that are between 1 and 2 hours old, and an increase in the survival rate of bacteria from 0 to 1 hours old.

We assume that the concentrations c_1 and c_2 can be independently increased or decreased (say, within a factor of 2), by administering drugs, and pose the problem of finding the drug mix that maximizes the population decay rate (*i.e.*, minimizes $\lambda_{\mathrm{pf}}(A)$). Using the approach described above, this problem can be posed as the GP

$$\begin{array}{ll} \text{minimize} & \lambda \\ \text{subject to} & b_1 v_1 + b_2 v_2 + b_3 v_3 + b_4 v_4 \leq \lambda v_1 \\ & s_1 v_1 \leq \lambda v_2 \\ & s_2 v_2 \leq \lambda v_3 \\ & s_3 v_3 \leq \lambda v_4 \\ & 1/2 \leq c_i / c_i^{\mathrm{nom}} \leq 2, \quad i = 1, 2 \\ & b_i = b_i^{\mathrm{nom}} (c_1/c_1^{\mathrm{nom}})^{\alpha_i} (c_2/c_2^{\mathrm{nom}})^{\beta_i}, \quad i = 1, \ldots, 4 \\ & s_i = s_i^{\mathrm{nom}} (c_1/c_1^{\mathrm{nom}})^{\gamma_i} (c_2/c_2^{\mathrm{nom}})^{\delta_i}, \quad i = 1, \ldots, 3, \end{array}$$

with variables b_i, s_i, c_i, v_i, and λ.

4.6 Generalized inequality constraints

One very useful generalization of the standard form convex optimization problem (4.15) is obtained by allowing the inequality constraint functions to be vector valued, and using generalized inequalities in the constraints:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \preceq_{K_i} 0, \quad i = 1, \ldots, m \\ & Ax = b, \end{array} \tag{4.48}$$

where $f_0 : \mathbf{R}^n \to \mathbf{R}$, $K_i \subseteq \mathbf{R}^{k_i}$ are proper cones, and $f_i : \mathbf{R}^n \to \mathbf{R}^{k_i}$ are K_i-convex. We refer to this problem as a (standard form) *convex optimization problem with generalized inequality constraints*. Problem (4.15) is a special case with $K_i = \mathbf{R}_+$, $i = 1, \ldots, m$.

Many of the results for ordinary convex optimization problems hold for problems with generalized inequalities. Some examples are:

- The feasible set, any sublevel set, and the optimal set are convex.
- Any point that is locally optimal for the problem (4.48) is globally optimal.
- The optimality condition for differentiable f_0, given in §4.2.3, holds without any change.

We will also see (in chapter 11) that convex optimization problems with generalized inequality constraints can often be solved as easily as ordinary convex optimization problems.

4.6.1 Conic form problems

Among the simplest convex optimization problems with generalized inequalities are the *conic form problems* (or *cone programs*), which have a linear objective and one inequality constraint function, which is affine (and therefore K-convex):

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Fx + g \preceq_K 0 \\ & Ax = b. \end{array} \tag{4.49}$$

When K is the nonnegative orthant, the conic form problem reduces to a linear program. We can view conic form problems as a generalization of linear programs in which componentwise inequality is replaced with a generalized linear inequality.

Continuing the analogy to linear programming, we refer to the conic form problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x \succeq_K 0 \\ & Ax = b \end{array}$$

as a *conic form problem in standard form.* Similarly, the problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Fx + g \preceq_K 0 \end{array}$$

is called a *conic form problem in inequality form.*

4.6.2 Semidefinite programming

When K is $\mathbf{S}_+^k$, the cone of positive semidefinite $k \times k$ matrices, the associated conic form problem is called a *semidefinite program* (SDP), and has the form

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + \cdots + x_n F_n + G \preceq 0 \\ & Ax = b, \end{array} \tag{4.50}$$

where $G,\ F_1, \ldots, F_n \in \mathbf{S}^k$, and $A \in \mathbf{R}^{p \times n}$. The inequality here is a linear matrix inequality (see example 2.10).

If the matrices $G,\ F_1, \ldots, F_n$ are all diagonal, then the LMI in (4.50) is equivalent to a set of n linear inequalities, and the SDP (4.50) reduces to a linear program.

Standard and inequality form semidefinite programs

Following the analogy to LP, a *standard form SDP* has linear equality constraints, and a (matrix) nonnegativity constraint on the variable $X \in \mathbf{S}^n$:

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(A_i X) = b_i, \quad i = 1, \ldots, p \\ & X \succeq 0, \end{array} \tag{4.51}$$

where $C,\ A_1,\ldots,A_p \in \mathbf{S}^n$. (Recall that $\mathbf{tr}(CX) = \sum_{i,j=1}^n C_{ij}X_{ij}$ is the form of a general real-valued linear function on $\mathbf{S}^n$.) This form should be compared to the standard form linear program (4.28). In LP and SDP standard forms, we minimize a linear function of the variable, subject to p linear equality constraints on the variable, and a nonnegativity constraint on the variable.

An *inequality form SDP*, analogous to an inequality form LP (4.29), has no equality constraints, and one LMI:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 A_1 + \cdots + x_n A_n \preceq B, \end{array}$$

with variable $x \in \mathbf{R}^n$, and parameters $B,\ A_1,\ldots,A_n \in \mathbf{S}^k$, $c \in \mathbf{R}^n$.

Multiple LMIs and linear inequalities

It is common to refer to a problem with linear objective, linear equality and inequality constraints, and several LMI constraints, *i.e.*,

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & F^{(i)}(x) = x_1 F_1^{(i)} + \cdots + x_n F_n^{(i)} + G^{(i)} \preceq 0, \quad i = 1,\ldots,K \\ & Gx \preceq h, \qquad Ax = b, \end{array}$$

as an SDP as well. Such problems are readily transformed to an SDP, by forming a large block diagonal LMI from the individual LMIs and linear inequalities:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \mathbf{diag}(Gx - h, F^{(1)}(x),\ldots,F^{(K)}(x)) \preceq 0 \\ & Ax = b. \end{array}$$

4.6.3 Examples

Second-order cone programming

The SOCP (4.36) can be expressed as a conic form problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & -(A_i x + b_i, c_i^T x + d_i) \preceq_{K_i} 0, \quad i = 1,\ldots,m \\ & Fx = g, \end{array}$$

in which

$$K_i = \{(y,t) \in \mathbf{R}^{n_i+1} \mid \|y\|_2 \le t\},$$

i.e., the second-order cone in $\mathbf{R}^{n_i+1}$. This explains the name *second-order cone program* for the optimization problem (4.36).

Matrix norm minimization

Let $A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n$, where $A_i \in \mathbf{R}^{p \times q}$. We consider the unconstrained problem

$$\text{minimize} \quad \|A(x)\|_2,$$

where $\|\cdot\|_2$ denotes the spectral norm (maximum singular value), and $x \in \mathbf{R}^n$ is the variable. This is a convex problem since $\|A(x)\|_2$ is a convex function of x.

Using the fact that $\|A\|_2 \leq s$ if and only if $A^T A \preceq s^2 I$ (and $s \geq 0$), we can express the problem in the form

$$\begin{array}{ll} \text{minimize} & s \\ \text{subject to} & A(x)^T A(x) \preceq sI, \end{array}$$

with variables x and s. Since the function $A(x)^T A(x) - sI$ is matrix convex in (x, s), this is a convex optimization problem with a single $q \times q$ matrix inequality constraint.

We can also formulate the problem using a single linear matrix inequality of size $(p+q) \times (p+q)$, using the fact that

$$A^T A \preceq t^2 I \text{ (and } t \geq 0) \iff \left[\begin{array}{cc} tI & A \\ A^T & tI \end{array}\right] \succeq 0.$$

(see §A.5.5). This results in the SDP

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \left[\begin{array}{cc} tI & A(x) \\ A(x)^T & tI \end{array}\right] \succeq 0 \end{array}$$

in the variables x and t.

Moment problems

Let t be a random variable in $\mathbf{R}$. The expected values $\mathbf{E}\, t^k$ (assuming they exist) are called the (power) *moments* of the distribution of t. The following classical results give a characterization of a moment sequence.

If there is a probability distribution on $\mathbf{R}$ such that $x_k = \mathbf{E}\, t^k$, $k = 0, \ldots, 2n$, then $x_0 = 1$ and

$$H(x_0, \ldots, x_{2n}) = \left[\begin{array}{cccccc} x_0 & x_1 & x_2 & \cdots & x_{n-1} & x_n \\ x_1 & x_2 & x_3 & \cdots & x_n & x_{n+1} \\ x_2 & x_3 & x_4 & \cdots & x_{n+1} & x_{n+2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ x_{n-1} & x_n & x_{n+1} & \cdots & x_{2n-2} & x_{2n-1} \\ x_n & x_{n+1} & x_{n+2} & \cdots & x_{2n-1} & x_{2n} \end{array}\right] \succeq 0. \qquad (4.52)$$

(The matrix H is called the *Hankel matrix* associated with $x_0, \ldots, x_{2n}$.) This is easy to see: Let $x_i = \mathbf{E}\, t^i$, $i = 0, \ldots, 2n$ be the moments of some distribution, and let $y = (y_0, y_1, \ldots y_n) \in \mathbf{R}^{n+1}$. Then we have

$$y^T H(x_0, \ldots, x_{2n}) y = \sum_{i,j=0}^{n} y_i y_j \,\mathbf{E}\, t^{i+j} = \mathbf{E}(y_0 + y_1 t^1 + \cdots + y_n t^n)^2 \geq 0.$$

The following partial converse is less obvious: If $x_0 = 1$ and $H(x) \succ 0$, then there exists a probability distribution on $\mathbf{R}$ such that $x_i = \mathbf{E}\, t^i$, $i = 0, \ldots, 2n$. (For a proof, see exercise 2.37.) Now suppose that $x_0 = 1$, and $H(x) \succeq 0$ (but possibly $H(x) \not\succ 0$), *i.e.*, the linear matrix inequality (4.52) holds, but possibly not strictly. In this case, there is a sequence of distributions on $\mathbf{R}$, whose moments converge to x. In summary: the condition that $x_0, \ldots, x_{2n}$ be the moments of some distribution on $\mathbf{R}$ (or the limit of the moments of a sequence of distributions) can be expressed as the linear matrix inequality (4.52) in the variable x, together with the linear equality $x_0 = 1$. Using this fact, we can cast some interesting problems involving moments as SDPs.

Suppose t is a random variable on $\mathbf{R}$. We do not know its distribution, but we do know some bounds on the moments, *i.e.*,

$$\underline{\mu}_k \le \mathbf{E}\, t^k \le \overline{\mu}_k, \quad k = 1, \ldots, 2n$$

(which includes, as a special case, knowing exact values of some of the moments). Let $p(t) = c_0 + c_1 t + \cdots + c_{2n} t^{2n}$ be a given polynomial in t. The expected value of $p(t)$ is linear in the moments $\mathbf{E}\, t^i$:

$$\mathbf{E}\, p(t) = \sum_{i=0}^{2n} c_i \,\mathbf{E}\, t^i = \sum_{i=0}^{2n} c_i x_i.$$

We can compute upper and lower bounds for $\mathbf{E}\, p(t)$,

$$\begin{array}{ll} \text{minimize (maximize)} & \mathbf{E}\, p(t) \\ \text{subject to} & \underline{\mu}_k \le \mathbf{E}\, t^k \le \overline{\mu}_k, \quad k = 1, \ldots, 2n, \end{array}$$

over all probability distributions that satisfy the given moment bounds, by solving the SDP

$$\begin{array}{ll} \text{minimize (maximize)} & c_1 x_1 + \cdots + c_{2n} x_{2n} \\ \text{subject to} & \underline{\mu}_k \le x_k \le \overline{\mu}_k, \quad k = 1, \ldots, 2n \\ & H(1, x_1, \ldots, x_{2n}) \succeq 0 \end{array}$$

with variables $x_1, \ldots, x_{2n}$. This gives bounds on $\mathbf{E}\, p(t)$, over all probability distributions that satisfy the known moment constraints. The bounds are sharp in the sense that there exists a sequence of distributions, whose moments satisfy the given moment bounds, for which $\mathbf{E}\, p(t)$ converges to the upper and lower bounds found by these SDPs.

Bounding portfolio risk with incomplete covariance information

We consider once again the setup for the classical Markowitz portfolio problem (see page 155). We have a portfolio of n assets or stocks, with x_i denoting the amount of asset i that is held over some investment period, and p_i denoting the relative price change of asset i over the period. The change in total value of the portfolio is $p^T x$. The price change vector p is modeled as a random vector, with mean and covariance

$$\overline{p} = \mathbf{E}\, p, \qquad \Sigma = \mathbf{E}(p - \overline{p})(p - \overline{p})^T.$$

The change in value of the portfolio is therefore a random variable with mean $\overline{p}^T x$ and standard deviation $\sigma = (x^T \Sigma x)^{1/2}$. The risk of a large loss, *i.e.*, a change in portfolio value that is substantially below its expected value, is directly related to the standard deviation σ, and increases with it. For this reason the standard deviation σ (or the variance σ^2) is used as a measure of the risk associated with the portfolio.

In the classical portfolio optimization problem, the portfolio x is the optimization variable, and we minimize the risk subject to a minimum mean return and other constraints. The price change statistics $\overline{p}$ and Σ are known problem parameters. In the risk bounding problem considered here, we turn the problem around: we assume the portfolio x is known, but only partial information is available about the covariance matrix Σ. We might have, for example, an upper and lower bound on each entry:

$$L_{ij} \leq \Sigma_{ij} \leq U_{ij}, \quad i,\ j = 1, \ldots, n,$$

where L and U are given. We now pose the question: what is the maximum risk for our portfolio, over all covariance matrices consistent with the given bounds? We define the *worst-case variance* of the portfolio as

$$\sigma_{\mathrm{wc}}^2 = \sup\{x^T \Sigma x \mid L_{ij} \leq \Sigma_{ij} \leq U_{ij},\ i, j = 1, \ldots, n,\ \Sigma \succeq 0\}.$$

We have added the condition $\Sigma \succeq 0$, which the covariance matrix must, of course, satisfy.

We can find σ_{wc} by solving the SDP

$$\begin{array}{ll} \text{maximize} & x^T \Sigma x \\ \text{subject to} & L_{ij} \leq \Sigma_{ij} \leq U_{ij}, \quad i,\ j = 1, \ldots, n \\ & \Sigma \succeq 0 \end{array}$$

with variable $\Sigma \in \mathbf{S}^n$ (and problem parameters x, L, and U). The optimal Σ is the worst covariance matrix consistent with our given bounds on the entries, where 'worst' means largest risk with the (given) portfolio x. We can easily construct a distribution for p that is consistent with the given bounds, and achieves the worst-case variance, from an optimal Σ for the SDP. For example, we can take $p = \overline{p} + \Sigma^{1/2} v$, where v is any random vector with $\mathbf{E}\, v = 0$ and $\mathbf{E}\, vv^T = I$.

Evidently we can use the same method to determine σ_{wc} for any prior information about Σ that is convex. We list here some examples.

- *Known variance of certain portfolios.* We might have equality constraints such as

  $$u_k^T \Sigma u_k = \sigma_k^2,$$

  where u_k and σ_k are given. This corresponds to prior knowledge that certain known portfolios (given by u_k) have known (or very accurately estimated) variance.

- *Including effects of estimation error.* If the covariance Σ is estimated from empirical data, the estimation method will give an estimate $\hat{\Sigma}$, and some information about the reliability of the estimate, such as a confidence ellipsoid. This can be expressed as

  $$C(\Sigma - \hat{\Sigma}) \leq \alpha,$$

  where C is a positive definite quadratic form on $\mathbf{S}^n$, and the constant α determines the confidence level.

- *Factor models.* The covariance might have the form

$$\Sigma = F\Sigma_{\text{factor}}F^T + D,$$

where $F \in \mathbf{R}^{n\times k}$, $\Sigma_{\text{factor}} \in \mathbf{S}^k$, and D is diagonal. This corresponds to a model of the price changes of the form

$$p = Fz + d,$$

where z is a random variable (the underlying *factors* that affect the price changes) and d_i are independent (additional volatility of each asset price). We assume that the factors are known. Since Σ is linearly related to Σ_{factor} and D, we can impose any convex constraint on them (representing prior information) and still compute σ_{wc} using convex optimization.

- *Information about correlation coefficients.* In the simplest case, the diagonal entries of Σ (*i.e.*, the volatilities of each asset price) are known, and bounds on correlation coefficients between price changes are known:

$$l_{ij} \le \rho_{ij} = \frac{\Sigma_{ij}}{\Sigma_{ii}^{1/2}\Sigma_{jj}^{1/2}} \le u_{ij}, \quad i,\ j = 1,\ldots,n.$$

Since Σ_{ii} are known, but Σ_{ij} for $i \ne j$ are not, these are linear inequalities.

Fastest mixing Markov chain on a graph

We consider an undirected graph, with nodes $1,\ldots,n$, and a set of edges

$$\mathcal{E} \subseteq \{1,\ldots,n\} \times \{1,\ldots,n\}.$$

Here $(i,j) \in \mathcal{E}$ means that nodes i and j are connected by an edge. Since the graph is undirected, $\mathcal{E}$ is symmetric: $(i,j) \in \mathcal{E}$ if and only if $(j,i) \in \mathcal{E}$. We allow the possibility of self-loops, *i.e.*, we can have $(i,i) \in \mathcal{E}$.

We define a Markov chain, with state $X(t) \in \{1,\ldots,n\}$, for $t \in \mathbf{Z}_+$ (the set of nonnegative integers), as follows. With each edge $(i,j) \in \mathcal{E}$ we associate a probability P_{ij}, which is the probability that X makes a transition between nodes i and j. State transitions can only occur across edges; we have $P_{ij} = 0$ for $(i,j) \notin \mathcal{E}$. The probabilities associated with the edges must be nonnegative, and for each node, the sum of the probabilities of links connected to the node (including a self-loop, if there is one) must equal one.

The Markov chain has transition probability matrix

$$P_{ij} = \mathbf{prob}(X(t+1) = i \mid X(t) = j), \quad i,j = 1,\ldots,n.$$

This matrix must satisfy

$$P_{ij} \ge 0, \quad i,\ j = 1,\ldots,n, \qquad \mathbf{1}^T P = \mathbf{1}^T, \qquad P = P^T, \tag{4.53}$$

and also

$$P_{ij} = 0 \quad \text{for } (i,j) \notin \mathcal{E}. \tag{4.54}$$

Since P is symmetric and $\mathbf{1}^T P = \mathbf{1}^T$, we conclude $P\mathbf{1} = \mathbf{1}$, so the uniform distribution $(1/n)\mathbf{1}$ is an equilibrium distribution for the Markov chain. Convergence of the distribution of $X(t)$ to $(1/n)\mathbf{1}$ is determined by the second largest (in magnitude) eigenvalue of P, *i.e.*, by $r = \max\{\lambda_2, -\lambda_n\}$, where

$$1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

are the eigenvalues of P. We refer to r as the *mixing rate* of the Markov chain. If $r = 1$, then the distribution of $X(t)$ need not converge to $(1/n)\mathbf{1}$ (which means the Markov chain does not mix). When $r < 1$, the distribution of $X(t)$ approaches $(1/n)\mathbf{1}$ asymptotically as r^t, as $t \to \infty$. Thus, the smaller r is, the faster the Markov chain mixes.

The *fastest mixing Markov chain problem* is to find P, subject to the constraints (4.53) and (4.54), that minimizes r. (The problem data is the graph, *i.e.*, $\mathcal{E}$.) We will show that this problem can be formulated as an SDP.

Since the eigenvalue $\lambda_1 = 1$ is associated with the eigenvector $\mathbf{1}$, we can express the mixing rate as the norm of the matrix P, restricted to the subspace $\mathbf{1}^\perp$: $r = \|QPQ\|_2$, where $Q = I - (1/n)\mathbf{1}\mathbf{1}^T$ is the matrix representing orthogonal projection on $\mathbf{1}^\perp$. Using the property $P\mathbf{1} = \mathbf{1}$, we have

$$\begin{aligned} r &= \|QPQ\|_2 \\ &= \|(I - (1/n)\mathbf{1}\mathbf{1}^T)P(I - (1/n)\mathbf{1}\mathbf{1}^T)\|_2 \\ &= \|P - (1/n)\mathbf{1}\mathbf{1}^T\|_2. \end{aligned}$$

This shows that the mixing rate r is a convex function of P, so the fastest mixing Markov chain problem can be cast as the convex optimization problem

$$\begin{array}{ll} \text{minimize} & \|P - (1/n)\mathbf{1}\mathbf{1}^T\|_2 \\ \text{subject to} & P\mathbf{1} = \mathbf{1} \\ & P_{ij} \geq 0, \quad i,j = 1,\ldots,n \\ & P_{ij} = 0 \text{ for } (i,j) \notin \mathcal{E}, \end{array}$$

with variable $P \in \mathbf{S}^n$. We can express the problem as an SDP by introducing a scalar variable t to bound the norm of $P - (1/n)\mathbf{1}\mathbf{1}^T$:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & -tI \preceq P - (1/n)\mathbf{1}\mathbf{1}^T \preceq tI \\ & P\mathbf{1} = \mathbf{1} \\ & P_{ij} \geq 0, \quad i,j = 1,\ldots,n \\ & P_{ij} = 0 \text{ for } (i,j) \notin \mathcal{E}. \end{array} \tag{4.55}$$

4.7 Vector optimization

4.7.1 General and convex vector optimization problems

In §4.6 we extended the standard form problem (4.1) to include vector-valued constraint functions. In this section we investigate the meaning of a vector-valued *objective function*. We denote a general *vector optimization problem* as

$$\begin{array}{lll} \text{minimize (with respect to } K) & f_0(x) & \\ \text{subject to} & f_i(x) \leq 0, & i = 1, \ldots, m \\ & h_i(x) = 0, & i = 1, \ldots, p. \end{array} \tag{4.56}$$

Here $x \in \mathbf{R}^n$ is the optimization variable, $K \subseteq \mathbf{R}^q$ is a proper cone, $f_0 : \mathbf{R}^n \to \mathbf{R}^q$ is the objective function, $f_i : \mathbf{R}^n \to \mathbf{R}$ are the inequality constraint functions, and $h_i : \mathbf{R}^n \to \mathbf{R}$ are the equality constraint functions. The only difference between this problem and the standard optimization problem (4.1) is that here, the objective function takes values in $\mathbf{R}^q$, and the problem specification includes a proper cone K, which is used to compare objective values. In the context of vector optimization, the standard optimization problem (4.1) is sometimes called a *scalar optimization problem*.

We say the vector optimization problem (4.56) is a *convex vector optimization problem* if the objective function f_0 is K-convex, the inequality constraint functions $f_1, \ldots, f_m$ are convex, and the equality constraint functions $h_1, \ldots, h_p$ are affine. (As in the scalar case, we usually express the equality constraints as $Ax = b$, where $A \in \mathbf{R}^{p \times n}$.)

What meaning can we give to the vector optimization problem (4.56)? Suppose x and y are two feasible points (*i.e.*, they satisfy the constraints). Their associated objective values, $f_0(x)$ and $f_0(y)$, are to be compared using the generalized inequality $\preceq_K$. We interpret $f_0(x) \preceq_K f_0(y)$ as meaning that x is 'better than or equal' in value to y (as judged by the objective f_0, with respect to K). The confusing aspect of vector optimization is that the two objective values $f_0(x)$ and $f_0(y)$ need not be comparable; we can have neither $f_0(x) \preceq_K f_0(y)$ nor $f_0(y) \preceq_K f_0(x)$, *i.e.*, neither is better than the other. This cannot happen in a scalar objective optimization problem.

4.7.2 Optimal points and values

We first consider a special case, in which the meaning of the vector optimization problem is clear. Consider the set of objective values of feasible points,

$$\mathcal{O} = \{f_0(x) \mid \exists x \in \mathcal{D},\ f_i(x) \leq 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, p\} \subseteq \mathbf{R}^q,$$

which is called the set of *achievable objective values*. If this set has a minimum element (see §2.4.2), *i.e.*, there is a feasible x such that $f_0(x) \preceq_K f_0(y)$ for all feasible y, then we say x is *optimal* for the problem (4.56), and refer to $f_0(x)$ as the *optimal value* of the problem. (When a vector optimization problem has an optimal value, it is unique.) If $x^\star$ is an optimal point, then $f_0(x^\star)$, the objective at $x^\star$, can be compared to the objective at every other feasible point, and is better than or equal to it. Roughly speaking, $x^\star$ is unambiguously a best choice for x, among feasible points.

A point $x^\star$ is optimal if and only if it is feasible and

$$\mathcal{O} \subseteq f_0(x^\star) + K \tag{4.57}$$

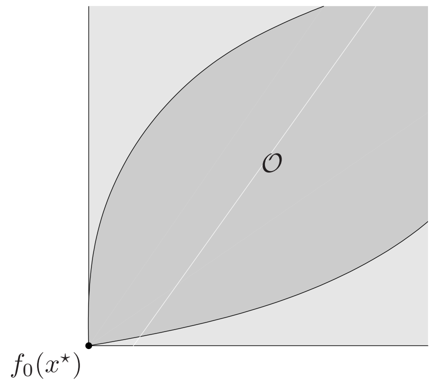


Figure 4.7 The set $\mathcal{O}$ of achievable values for a vector optimization with objective values in $\mathbf{R}^2$, with cone $K = \mathbf{R}^2_+$, is shown shaded. In this case, the point labeled $f_0(x^\star)$ is the optimal value of the problem, and $x^\star$ is an optimal point. The objective value $f_0(x^\star)$ can be compared to every other achievable value $f_0(y)$, and is better than or equal to $f_0(y)$. (Here, 'better than or equal to' means 'is below and to the left of'.) The lightly shaded region is $f_0(x^\star)+K$, which is the set of all $z \in \mathbf{R}^2$ corresponding to objective values worse than (or equal to) $f_0(x^\star)$.

(see §2.4.2). The set $f_0(x^\star) + K$ can be interpreted as the set of values that are worse than, or equal to, $f_0(x^\star)$, so the condition (4.57) states that every achievable value falls in this set. This is illustrated in figure 4.7. Most vector optimization problems do not have an optimal point and an optimal value, but this does occur in some special cases.

Example 4.9 *Best linear unbiased estimator.* Suppose $y = Ax + v$, where $v \in \mathbf{R}^m$ is a measurement noise, $y \in \mathbf{R}^m$ is a vector of measurements, and $x \in \mathbf{R}^n$ is a vector to be estimated, given the measurement y. We assume that A has rank n, and that the measurement noise satisfies $\mathbf{E}\, v = 0$, $\mathbf{E}\, vv^T = I$, *i.e.*, its components are zero mean and uncorrelated.

A *linear estimator* of x has the form $\widehat{x} = Fy$. The estimator is called *unbiased* if for all x we have $\mathbf{E}\, \widehat{x} = x$, *i.e.*, if $FA = I$. The error covariance of an unbiased estimator is

$$\mathbf{E}(\widehat{x} - x)(\widehat{x} - x)^T = \mathbf{E}\, Fvv^T F^T = FF^T.$$

Our goal is to find an unbiased estimator that has a 'small' error covariance matrix. We can compare error covariances using matrix inequality, *i.e.*, with respect to $\mathbf{S}^n_+$. This has the following interpretation: Suppose $\widehat{x}_1 = F_1 y$, $\widehat{x}_2 = F_2 y$ are two unbiased estimators. Then the first estimator is at least as good as the second, *i.e.*, $F_1 F_1^T \preceq F_2 F_2^T$, if and only if for all c,

$$\mathbf{E}(c^T \widehat{x}_1 - c^T x)^2 \le \mathbf{E}(c^T \widehat{x}_2 - c^T x)^2.$$

In other words, for any linear function of x, the estimator F_1 yields at least as good an estimate as does F_2.

We can express the problem of finding an unbiased estimator for x as the vector optimization problem

$$\begin{array}{ll} \text{minimize (w.r.t. } \mathbf{S}_+^n) & FF^T \\ \text{subject to} & FA = I, \end{array} \qquad (4.58)$$

with variable $F \in \mathbf{R}^{n \times m}$. The objective FF^T is convex with respect to $\mathbf{S}_+^n$, so the problem (4.58) is a convex vector optimization problem. An easy way to see this is to observe that $v^T FF^T v = \|F^T v\|_2^2$ is a convex function of F for any fixed v.

It is a famous result that the problem (4.58) has an optimal solution, the least-squares estimator, or pseudo-inverse,

$$F^\star = A^\dagger = (A^T A)^{-1} A^T.$$

For any F with $FA = I$, we have $FF^T \succeq F^\star F^{\star T}$. The matrix

$$F^\star F^{\star T} = A^\dagger A^{\dagger T} = (A^T A)^{-1}$$

is the optimal value of the problem (4.58).

4.7.3 Pareto optimal points and values

We now consider the case (which occurs in most vector optimization problems of interest) in which the set of achievable objective values does not have a minimum element, so the problem does not have an optimal point or optimal value. In these cases *minimal* elements of the set of achievable values play an important role. We say that a feasible point x is *Pareto optimal* (or *efficient*) if $f_0(x)$ is a minimal element of the set of achievable values $\mathcal{O}$. In this case we say that $f_0(x)$ is a *Pareto optimal value* for the vector optimization problem (4.56). Thus, a point x is Pareto optimal if it is feasible and, for any feasible y, $f_0(y) \preceq_K f_0(x)$ implies $f_0(y) = f_0(x)$. In other words: any feasible point y that is better than or equal to x (*i.e.*, $f_0(y) \preceq_K f_0(x)$) has exactly the same objective value as x.

A point x is Pareto optimal if and only if it is feasible and

$$(f_0(x) - K) \cap \mathcal{O} = \{f_0(x)\} \qquad (4.59)$$

(see §2.4.2). The set $f_0(x) - K$ can be interpreted as the set of values that are better than or equal to $f_0(x)$, so the condition (4.59) states that the only achievable value better than or equal to $f_0(x)$ is $f_0(x)$ itself. This is illustrated in figure 4.8.

A vector optimization problem can have many Pareto optimal values (and points). The set of Pareto optimal values, denoted $\mathcal{P}$, satisfies

$$\mathcal{P} \subseteq \mathcal{O} \cap \mathbf{bd}\,\mathcal{O},$$

i.e., every Pareto optimal value is an achievable objective value that lies in the boundary of the set of achievable objective values (see exercise 4.52).

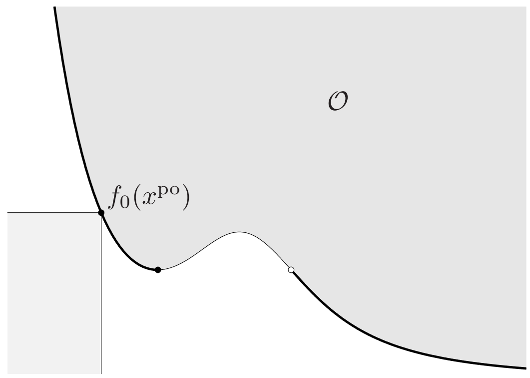


Figure 4.8 The set $\mathcal{O}$ of achievable values for a vector optimization problem with objective values in $\mathbf{R}^2$, with cone $K = \mathbf{R}^2_+$, is shown shaded. This problem does not have an optimal point or value, but it does have a set of Pareto optimal points, whose corresponding values are shown as the darkened curve on the lower left boundary of $\mathcal{O}$. The point labeled $f_0(x^{\text{po}})$ is a Pareto optimal value, and x^{po} is a Pareto optimal point. The lightly shaded region is $f_0(x^{\text{po}}) - K$, which is the set of all $z \in \mathbf{R}^2$ corresponding to objective values better than (or equal to) $f_0(x^{\text{po}})$.

4.7.4 Scalarization

Scalarization is a standard technique for finding Pareto optimal (or optimal) points for a vector optimization problem, based on the characterization of minimum and minimal points via dual generalized inequalities given in §2.6.3. Choose any $\lambda \succ_{K^*} 0$, *i.e.*, any vector that is positive in the dual generalized inequality. Now consider the *scalar* optimization problem

$$\begin{array}{ll} \text{minimize} & \lambda^T f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p, \end{array} \tag{4.60}$$

and let x be an optimal point. Then x is Pareto optimal for the vector optimization problem (4.56). This follows from the dual inequality characterization of minimal points given in §2.6.3, and is also easily shown directly. If x were not Pareto optimal, then there is a y that is feasible, satisfies $f_0(y) \preceq_K f_0(x)$, and $f_0(x) \neq f_0(y)$. Since $f_0(x) - f_0(y) \succeq_K 0$ and is nonzero, we have $\lambda^T(f_0(x) - f_0(y)) > 0$, *i.e.*, $\lambda^T f_0(x) > \lambda^T f_0(y)$. This contradicts the assumption that x is optimal for the scalar problem (4.60).

Using scalarization, we can find Pareto optimal points for *any* vector optimization problem by solving the ordinary scalar optimization problem (4.60). The vector λ, which is sometimes called the *weight vector*, must satisfy $\lambda \succ_{K^*} 0$. The weight vector is a free parameter; by varying it we obtain (possibly) different Pareto optimal solutions of the vector optimization problem (4.56). This is illustrated in figure 4.9. The figure also shows an example of a Pareto optimal point that cannot

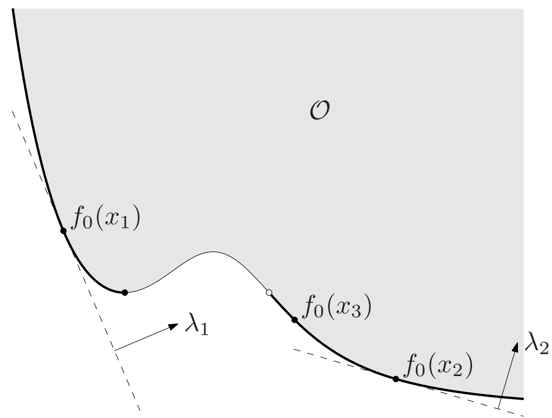


Figure 4.9 *Scalarization.* The set $\mathcal{O}$ of achievable values for a vector optimization problem with cone $K = \mathbf{R}^2_+$. Three Pareto optimal values $f_0(x_1)$, $f_0(x_2)$, $f_0(x_3)$ are shown. The first two values can be obtained by scalarization: $f_0(x_1)$ minimizes $\lambda_1^T u$ over all $u \in \mathcal{O}$ and $f_0(x_2)$ minimizes $\lambda_2^T u$, where $\lambda_1, \lambda_2 \succ 0$. The value $f_0(x_3)$ is Pareto optimal, but cannot be found by scalarization.

be obtained via scalarization, for any value of the weight vector $\lambda \succ_{K^*} 0$.

The method of scalarization can be interpreted geometrically. A point x is optimal for the scalarized problem, *i.e.*, minimizes $\lambda^T f_0$ over the feasible set, if and only if $\lambda^T(f_0(y) - f_0(x)) \geq 0$ for all feasible y. But this is the same as saying that $\{u \mid -\lambda^T(u - f_0(x)) = 0\}$ is a supporting hyperplane to the set of achievable objective values $\mathcal{O}$ at the point $f_0(x)$; in particular

$$\{u \mid \lambda^T(u - f_0(x)) < 0\} \cap \mathcal{O} = \emptyset. \tag{4.61}$$

(See figure 4.9.) Thus, when we find an optimal point for the scalarized problem, we not only find a Pareto optimal point for the original vector optimization problem; we also find an entire halfspace in $\mathbf{R}^q$, given by (4.61), of objective values that cannot be achieved.

Scalarization of convex vector optimization problems

Now suppose the vector optimization problem (4.56) is convex. Then the scalarized problem (4.60) is also convex, since $\lambda^T f_0$ is a (scalar-valued) convex function (by the results in §3.6). This means that we can find Pareto optimal points of a convex vector optimization problem by solving a convex scalar optimization problem. For each choice of the weight vector $\lambda \succ_{K^*} 0$ we get a (usually different) Pareto optimal point.

For convex vector optimization problems we have a partial converse: For every Pareto optimal point x^{po}, there is some nonzero $\lambda \succeq_{K^*} 0$ such that x^{po} is a solution of the scalarized problem (4.60). So, roughly speaking, for convex problems the method of scalarization yields all Pareto optimal points, as the weight vector λ

varies over the K^*-nonnegative, nonzero values. We have to be careful here, because it is *not* true that every solution of the scalarized problem, with $\lambda \succeq_{K^*} 0$ and $\lambda \neq 0$, is a Pareto optimal point for the vector problem. (In contrast, *every* solution of the scalarized problem with $\lambda \succ_{K^*} 0$ is Pareto optimal.)

In some cases we can use this partial converse to find *all* Pareto optimal points of a convex vector optimization problem. Scalarization with $\lambda \succ_{K^*} 0$ gives a set of Pareto optimal points (as it would in a nonconvex vector optimization problem as well). To find the remaining Pareto optimal solutions, we have to consider nonzero weight vectors λ that satisfy $\lambda \succeq_{K^*} 0$. For each such weight vector, we first identify all solutions of the scalarized problem. Then among these solutions we must check which are, in fact, Pareto optimal for the vector optimization problem. These 'extreme' Pareto optimal points can also be found as the limits of the Pareto optimal points obtained from positive weight vectors.

To establish this partial converse, we consider the set

$$\mathcal{A} = \mathcal{O} + K = \{t \in \mathbf{R}^q \mid f_0(x) \preceq_K t \text{ for some feasible } x\}, \tag{4.62}$$

which consists of all values that are worse than or equal to (with respect to $\preceq_K$) some achievable objective value. While the set $\mathcal{O}$ of achievable objective values need not be convex, the set $\mathcal{A}$ is convex, when the problem is convex. Moreover, the minimal elements of $\mathcal{A}$ are exactly the same as the minimal elements of the set $\mathcal{O}$ of achievable values, *i.e.*, they are the same as the Pareto optimal values. (See exercise 4.53.) Now we use the results of §2.6.3 to conclude that any minimal element of $\mathcal{A}$ minimizes $\lambda^T z$ over $\mathcal{A}$ for some nonzero $\lambda \succeq_{K^*} 0$. This means that every Pareto optimal point for the vector optimization problem is optimal for the scalarized problem, for some nonzero weight $\lambda \succeq_{K^*} 0$.

Example 4.10 *Minimal upper bound on a set of matrices.* We consider the (convex) vector optimization problem, with respect to the positive semidefinite cone,

$$\begin{array}{ll} \text{minimize (w.r.t. } \mathbf{S}^n_+) & X \\ \text{subject to} & X \succeq A_i, \quad i = 1, \dots, m, \end{array} \tag{4.63}$$

where $A_i \in \mathbf{S}^n$, $i = 1, \dots, m$, are given. The constraints mean that X is an upper bound on the given matrices $A_1, \dots, A_m$; a Pareto optimal solution of (4.63) is a *minimal upper bound* on the matrices.

To find a Pareto optimal point, we apply scalarization: we choose any $W \in \mathbf{S}^n_{++}$ and form the problem

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(WX) \\ \text{subject to} & X \succeq A_i, \quad i = 1, \dots, m, \end{array} \tag{4.64}$$

which is an SDP. Different choices for W will, in general, give different minimal solutions.

The partial converse tells us that if X is Pareto optimal for the vector problem (4.63) then it is optimal for the SDP (4.64), for some nonzero weight matrix $W \succeq 0$. (In this case, however, not every solution of (4.64) is Pareto optimal for the vector optimization problem.)

We can give a simple geometric interpretation for this problem. We associate with each $A \in \mathbf{S}^n_{++}$ an ellipsoid centered at the origin, given by

$$\mathcal{E}_A = \{u \mid u^T A^{-1} u \leq 1\},$$

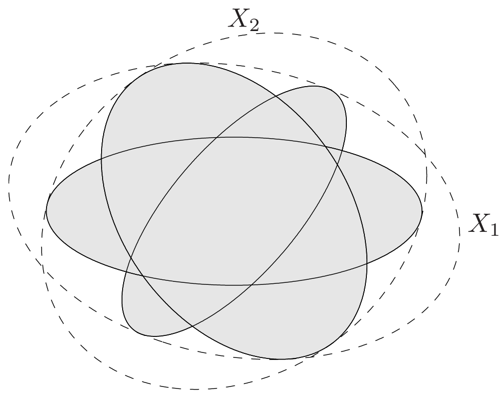


Figure 4.10 Geometric interpretation of the problem (4.63). The three shaded ellipsoids correspond to the data A_1, A_2, $A_3 \in \mathbf{S}^2_{++}$; the Pareto optimal points correspond to minimal ellipsoids that contain them. The two ellipsoids, with boundaries labeled X_1 and X_2, show two minimal ellipsoids obtained by solving the SDP (4.64) for two different weight matrices W_1 and W_2.

so that $A \preceq B$ if and only if $\mathcal{E}_A \subseteq \mathcal{E}_B$. A Pareto optimal point X for the problem (4.63) corresponds to a minimal ellipsoid that contains the ellipsoids associated with $A_1, \ldots, A_m$. An example is shown in figure 4.10.

4.7.5 Multicriterion optimization

When a vector optimization problem involves the cone $K = \mathbf{R}^q_+$, it is called a *multicriterion* or *multi-objective* optimization problem. The components of f_0, say, $F_1, \ldots, F_q$, can be interpreted as q different scalar objectives, each of which we would like to minimize. We refer to F_i as the *ith objective* of the problem. A multicriterion optimization problem is convex if $f_1, \ldots, f_m$ are convex, $h_1, \ldots, h_p$ are affine, and the objectives $F_1, \ldots, F_q$ are convex.

Since multicriterion problems are vector optimization problems, all of the material of §4.7.1–§4.7.4 applies. For multicriterion problems, though, we can be a bit more specific in the interpretations. If x is feasible, we can think of $F_i(x)$ as its score or value, according to the ith objective. If x and y are both feasible, $F_i(x) \leq F_i(y)$ means that x is at least as good as y, according to the ith objective; $F_i(x) < F_i(y)$ means that x is better than y, or x beats y, according to the ith objective. If x and y are both feasible, we say that x is *better than* y, or x *dominates* y, if $F_i(x) \leq F_i(y)$ for $i = 1, \ldots, q$, and for at least one j, $F_j(x) < F_j(y)$. Roughly speaking, x is better than y if x meets or beats y on all objectives, and beats it in at least one objective.

In a multicriterion problem, an optimal point $x^\star$ satisfies

$$F_i(x^\star) \leq F_i(y), \quad i = 1, \ldots, q,$$

for every feasible y. In other words, $x^\star$ is simultaneously optimal for each of the scalar problems

$$\begin{array}{ll} \text{minimize} & F_j(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p, \end{array}$$

for $j = 1, \ldots, q$. When there is an optimal point, we say that the objectives are *noncompeting*, since no compromises have to be made among the objectives; each objective is as small as it could be made, even if the others were ignored.

A Pareto optimal point x^{po} satisfies the following: if y is feasible and $F_i(y) \leq F_i(x^{\text{po}})$ for $i = 1, \ldots, q$, then $F_i(x^{\text{po}}) = F_i(y)$, $i = 1, \ldots, q$. This can be restated as: a point is Pareto optimal if and only if it is feasible and there is no better feasible point. In particular, if a feasible point is not Pareto optimal, there is at least one other feasible point that is better. In searching for good points, then, we can clearly limit our search to Pareto optimal points.

Trade-off analysis

Now suppose that x and y are Pareto optimal points with, say,

$$\begin{array}{ll} F_i(x) < F_i(y), & i \in A \\ F_i(x) = F_i(y), & i \in B \\ F_i(x) > F_i(y), & i \in C, \end{array}$$

where $A \cup B \cup C = \{1, \ldots, q\}$. In other words, A is the set of (indices of) objectives for which x beats y, B is the set of objectives for which the points x and y are tied, and C is the set of objectives for which y beats x. If A and C are empty, then the two points x and y have exactly the same objective values. If this is not the case, then both A and C must be nonempty. In other words, when comparing two Pareto optimal points, they either obtain the same performance (*i.e.*, all objectives equal), or, each beats the other in at least one objective.

In comparing the point x to y, we say that we have *traded* or *traded off* better objective values for $i \in A$ for worse objective values for $i \in C$. *Optimal trade-off analysis* (or just trade-off analysis) is the study of how much worse we must do in one or more objectives in order to do better in some other objectives, or more generally, the study of what sets of objective values are achievable.

As an example, consider a bi-criterion (*i.e.*, two criterion) problem. Suppose x is a Pareto optimal point, with objectives $F_1(x)$ and $F_2(x)$. We might ask how much larger $F_2(z)$ would have to be, in order to obtain a feasible point z with $F_1(z) \leq F_1(x) - a$, where $a > 0$ is some constant. Roughly speaking, we are asking how much we must pay in the second objective to obtain an improvement of a in the first objective. If a large increase in F_2 must be accepted to realize a small decrease in F_1, we say that there is a *strong trade-off* between the objectives, near the Pareto optimal value $(F_1(x), F_2(x))$. If, on the other hand, a large decrease in F_1 can be obtained with only a small increase in F_2, we say that the trade-off between the objectives is *weak* (near the Pareto optimal value $(F_1(x), F_2(x))$).

We can also consider the case in which we trade worse performance in the first objective for an improvement in the second. Here we find how much smaller $F_2(z)$

can be made, to obtain a feasible point z with $F_1(z) \leq F_1(x) + a$, where $a > 0$ is some constant. In this case we receive a benefit in the second objective, *i.e.*, a reduction in F_2 compared to $F_2(x)$. If this benefit is large (*i.e.*, by increasing F_1 a small amount we obtain a large reduction in F_2), we say the objectives exhibit a strong trade-off. If it is small, we say the objectives trade off weakly (near the Pareto optimal value $(F_1(x), F_2(x))$).

Optimal trade-off surface

The set of Pareto optimal values for a multicriterion problem is called the *optimal trade-off surface* (in general, when $q > 2$) or the *optimal trade-off curve* (when $q = 2$). (Since it would be foolish to accept any point that is not Pareto optimal, we can restrict our trade-off analysis to Pareto optimal points.) Trade-off analysis is also sometimes called *exploring the optimal trade-off surface*. (The optimal trade-off surface is usually, but not always, a surface in the usual sense. If the problem has an optimal point, for example, the optimal trade-off surface consists of a single point, the optimal value.)

An optimal trade-off curve is readily interpreted. An example is shown in figure 4.11, on page 185, for a (convex) bi-criterion problem. From this curve we can easily visualize and understand the trade-offs between the two objectives.

- The endpoint at the right shows the smallest possible value of F_2, without any consideration of F_1.
- The endpoint at the left shows the smallest possible value of F_1, without any consideration of F_2.
- By finding the intersection of the curve with a vertical line at $F_1 = \alpha$, we can see how large F_2 must be to achieve $F_1 \leq \alpha$.
- By finding the intersection of the curve with a horizontal line at $F_2 = \beta$, we can see how large F_1 must be to achieve $F_2 \leq \beta$.
- The slope of the optimal trade-off curve at a point on the curve (*i.e.*, a Pareto optimal value) shows the *local* optimal trade-off between the two objectives. Where the slope is steep, small changes in F_1 are accompanied by large changes in F_2.
- A point of large curvature is one where small decreases in one objective can only be accomplished by a large increase in the other. This is the proverbial *knee of the trade-off curve*, and in many applications represents a good compromise solution.

All of these have simple extensions to a trade-off surface, although visualizing a surface with more than three objectives is difficult.

Scalarizing multicriterion problems

When we scalarize a multicriterion problem by forming the weighted sum objective

$$\lambda^T f_0(x) = \sum_{i=1}^{q} \lambda_i F_i(x),$$

where $\lambda \succ 0$, we can interpret λ_i as the *weight* we attach to the ith objective. The weight λ_i can be thought of as quantifying our desire to make F_i small (or our objection to having F_i large). In particular, we should take λ_i large if we want F_i to be small; if we care much less about F_i, we can take λ_i small. We can interpret the ratio λ_i/λ_j as the *relative weight* or relative importance of the ith objective compared to the jth objective. Alternatively, we can think of λ_i/λ_j as *exchange rate* between the two objectives, since in the weighted sum objective a decrease (say) in F_i by α is considered the same as an increase in F_j in the amount $(\lambda_i/\lambda_j)\alpha$.

These interpretations give us some intuition about how to set or change the weights while exploring the optimal trade-off surface. Suppose, for example, that the weight vector $\lambda \succ 0$ yields the Pareto optimal point $x^{\rm po}$, with objective values $F_1(x^{\rm po}), \ldots, F_q(x^{\rm po})$. To find a (possibly) new Pareto optimal point which trades off a better kth objective value (say), for (possibly) worse objective values for the other objectives, we form a new weight vector $\tilde{\lambda}$ with

$$\tilde{\lambda}_k > \lambda_k, \qquad \tilde{\lambda}_j = \lambda_j, \quad j \neq k, \quad j = 1, \ldots, q,$$

i.e., we increase the weight on the kth objective. This yields a new Pareto optimal point $\tilde{x}^{\rm po}$ with $F_k(\tilde{x}^{\rm po}) \leq F_k(x^{\rm po})$ (and usually, $F_k(\tilde{x}^{\rm po}) < F_k(x^{\rm po})$), *i.e.*, a new Pareto optimal point with an improved kth objective.

We can also see that at any point where the optimal trade-off surface is smooth, λ gives the inward normal to the surface at the associated Pareto optimal point. In particular, when we choose a weight vector λ and apply scalarization, we obtain a Pareto optimal point where λ gives the local trade-offs among objectives.

In practice, optimal trade-off surfaces are explored by ad hoc adjustment of the weights, based on the intuitive ideas above. We will see later (in chapter 5) that the basic idea of scalarization, *i.e.*, minimizing a weighted sum of objectives, and then adjusting the weights to obtain a suitable solution, is the essence of duality.

4.7.6 Examples

Regularized least-squares

We are given $A \in \mathbf{R}^{m\times n}$ and $b \in \mathbf{R}^m$, and want to choose $x \in \mathbf{R}^n$ taking into account two quadratic objectives:

- $F_1(x) = \|Ax - b\|_2^2 = x^TA^TAx - 2b^TAx + b^Tb$ is a measure of the misfit between Ax and b,
- $F_2(x) = \|x\|_2^2 = x^Tx$ is a measure of the size of x.

Our goal is to find x that gives a good fit (*i.e.*, small F_1) and that is not large (*i.e.*, small F_2). We can formulate this problem as a vector optimization problem with respect to the cone $\mathbf{R}^2_+$, *i.e.*, a bi-criterion problem (with no constraints):

$$\text{minimize (w.r.t. } \mathbf{R}^2_+) \quad f_0(x) = (F_1(x), F_2(x)).$$

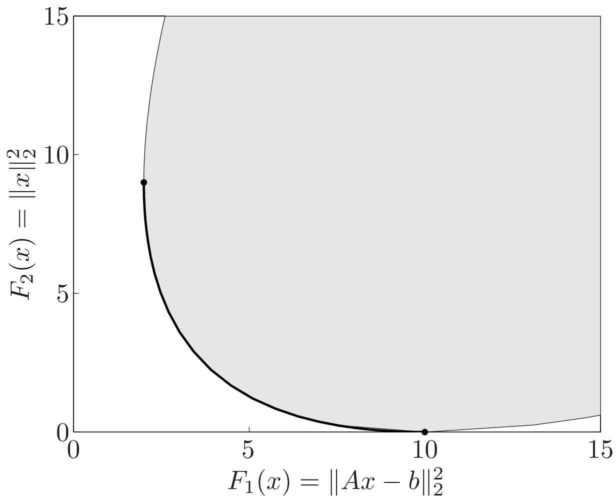


Figure 4.11 Optimal trade-off curve for a regularized least-squares problem. The shaded set is the set of achievable values $(\|Ax-b\|_2^2, \|x\|_2^2)$. The optimal trade-off curve, shown darker, is the lower left part of the boundary.

We can scalarize this problem by taking $\lambda_1 > 0$ and $\lambda_2 > 0$ and minimizing the scalar weighted sum objective

$$\begin{aligned}\lambda^T f_0(x) &= \lambda_1 F_1(x) + \lambda_2 F_2(x)\\ &= x^T(\lambda_1 A^T A + \lambda_2 I)x - 2\lambda_1 b^T Ax + \lambda_1 b^T b,\end{aligned}$$

which yields

$$x(\mu) = (\lambda_1 A^T A + \lambda_2 I)^{-1}\lambda_1 A^T b = (A^T A + \mu I)^{-1} A^T b,$$

where $\mu = \lambda_2/\lambda_1$. For any $\mu > 0$, this point is Pareto optimal for the bi-criterion problem. We can interpret $\mu = \lambda_2/\lambda_1$ as the relative weight we assign F_2 compared to F_1.

This method produces all Pareto optimal points, except two, associated with the extremes $\mu \to \infty$ and $\mu \to 0$. In the first case we have the Pareto optimal solution $x = 0$, which would be obtained by scalarization with $\lambda = (0, 1)$. At the other extreme we have the Pareto optimal solution $A^\dagger b$, where $A^\dagger$ is the pseudo-inverse of A. This Pareto optimal solution is obtained as the limit of the optimal solution of the scalarized problem as $\mu \to 0$, *i.e.*, as $\lambda \to (1, 0)$. (We will encounter the regularized least-squares problem again in §6.3.2.)

Figure 4.11 shows the optimal trade-off curve and the set of achievable values for a regularized least-squares problem with problem data $A \in \mathbf{R}^{100\times 10}$, $b \in \mathbf{R}^{100}$. (See exercise 4.50 for more discussion.)

Risk-return trade-off in portfolio optimization

The classical Markowitz portfolio optimization problem described on page 155 is naturally expressed as a bi-criterion problem, where the objectives are the negative mean return (since we wish to *maximize* mean return) and the variance of the return:

$$\begin{array}{ll} \text{minimize (w.r.t. } \mathbf{R}^2_+) & (F_1(x), F_2(x)) = (-\overline{p}^T x, x^T \Sigma x) \\ \text{subject to} & \mathbf{1}^T x = 1, \quad x \succeq 0. \end{array}$$

In forming the associated scalarized problem, we can (without loss of generality) take $\lambda_1 = 1$ and $\lambda_2 = \mu > 0$:

$$\begin{array}{ll} \text{minimize} & -\overline{p}^T x + \mu x^T \Sigma x \\ \text{subject to} & \mathbf{1}^T x = 1, \quad x \succeq 0, \end{array}$$

which is a QP. In this example too, we get all Pareto optimal portfolios except for the two limiting cases corresponding to $\mu \to 0$ and $\mu \to \infty$. Roughly speaking, in the first case we get a maximum mean return, without regard for return variance; in the second case we form a minimum variance return, without regard for mean return. Assuming that $\overline{p}_k > \overline{p}_i$ for $i \neq k$, *i.e.*, that asset k is the unique asset with maximum mean return, the portfolio allocation $x = e_k$ is the only one corresponding to $\mu \to 0$. (In other words, we concentrate the portfolio entirely in the asset that has maximum mean return.) In many portfolio problems asset n corresponds to a *risk-free* investment, with (deterministic) return $r_{\rm rf}$. Assuming that Σ, with its last row and column (which are zero) removed, is full rank, then the other extreme Pareto optimal portfolio is $x = e_n$, *i.e.*, the portfolio is concentrated entirely in the risk-free asset.

As a specific example, we consider a simple portfolio optimization problem with 4 assets, with price change mean and standard deviations given in the following table.

Asset	$\overline{p}_i$	$\Sigma_{ii}^{1/2}$
1	12%	20%
2	10%	10%
3	7%	5%
4	3%	0%

Asset 4 is a risk-free asset, with a (certain) 3% return. Assets 3, 2, and 1 have increasing mean returns, ranging from 7% to 12%, as well as increasing standard deviations, which range from 5% to 20%. The correlation coefficients between the assets are $\rho_{12} = 30\%$, $\rho_{13} = -40\%$, and $\rho_{23} = 0\%$.

Figure 4.12 shows the optimal trade-off curve for this portfolio optimization problem. The plot is given in the conventional way, with the horizontal axis showing standard deviation (*i.e.*, squareroot of variance) and the vertical axis showing expected return. The lower plot shows the optimal asset allocation vector x for each Pareto optimal point.

The results in this simple example agree with our intuition. For small risk, the optimal allocation consists mostly of the risk-free asset, with a mixture of the other assets in smaller quantities. Note that a mixture of asset 3 and asset 1, which are negatively correlated, gives some hedging, *i.e.*, lowers variance for a given level of mean return. At the other end of the trade-off curve, we see that aggressive growth portfolios (*i.e.*, those with large mean returns) concentrate the allocation in assets 1 and 2, the ones with the largest mean returns (and variances).

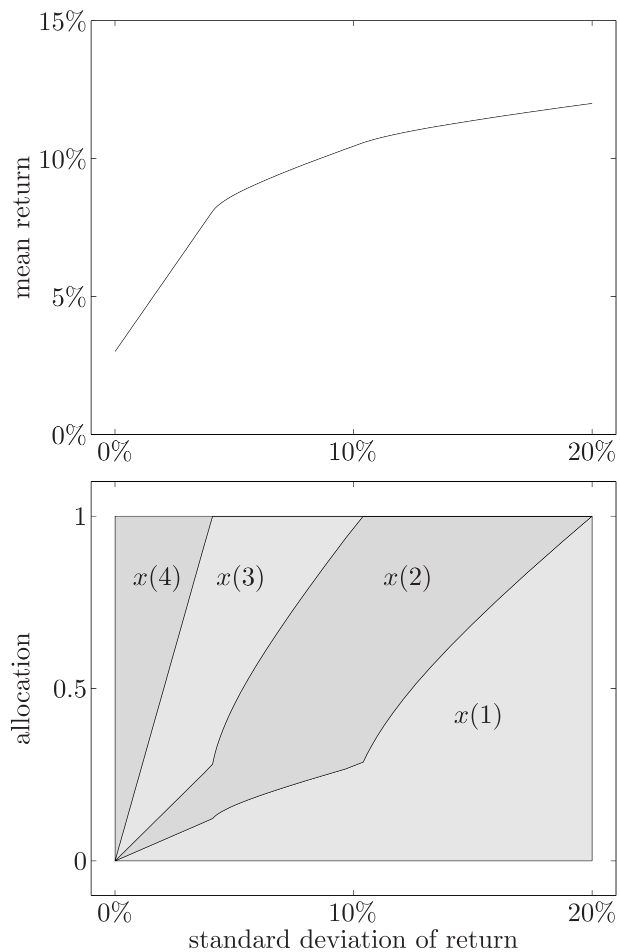


Figure 4.12 *Top.* Optimal risk-return trade-off curve for a simple portfolio optimization problem. The lefthand endpoint corresponds to putting all resources in the risk-free asset, and so has zero standard deviation. The righthand endpoint corresponds to putting all resources in asset 1, which has highest mean return. *Bottom.* Corresponding optimal allocations.

Bibliography

Linear programming has been studied extensively since the 1940s, and is the subject of many excellent books, including Dantzig [Dan63], Luenberger [Lue84], Schrijver [Sch86], Papadimitriou and Steiglitz [PS98], Bertsimas and Tsitsiklis [BT97], Vanderbei [Van96], and Roos, Terlaky, and Vial [RTV97]. Dantzig and Schrijver also provide detailed accounts of the history of linear programming. For a recent survey, see Todd [Tod02].

Schaible [Sch82, Sch83] gives an overview of fractional programming, which includes linear-fractional problems and extensions such as convex-concave fractional problems (see exercise 4.7). The model of a growing economy in example 4.7 appears in von Neumann [vN46].

Research on quadratic programming began in the 1950s (see, *e.g.*, Frank and Wolfe [FW56], Markowitz [Mar56], Hildreth [Hil57]), and was in part motivated by the portfolio optimization problem discussed on page 155 (Markowitz [Mar52]), and the LP with random cost discussed on page 154 (see Freund [Fre56]).

Interest in second-order cone programming is more recent, and started with Nesterov and Nemirovski [NN94, §6.2.3]. The theory and applications of SOCPs are surveyed by Alizadeh and Goldfarb [AG03], Ben-Tal and Nemirovski [BTN01, lecture 3] (where the problem is referred to as *conic quadratic programming*), and Lobo, Vandenberghe, Boyd, and Lebret [LVBL98].

Robust linear programming, and robust convex optimization in general, originated with Ben-Tal and Nemirovski [BTN98, BTN99] and El Ghaoui and Lebret [EL97]. Goldfarb and Iyengar [GI03a, GI03b] discuss robust QCQPs and applications in portfolio optimization. El Ghaoui, Oustry, and Lebret [EOL98] focus on robust semidefinite programming.

Geometric programming has been known since the 1960s. Its use in engineering design was first advocated by Duffin, Peterson, and Zener [DPZ67] and Zener [Zen71]. Peterson [Pet76] and Ecker [Eck80] describe the progress made during the 1970s. These articles and books also include examples of engineering applications, in particular in chemical and civil engineering. Fishburn and Dunlop [FD85], Sapatnekar, Rao, Vaidya, and Kang [SRVK93], and Hershenson, Boyd, and Lee [HBL01]) apply geometric programming to problems in integrated circuit design. The cantilever beam design example (page 163) is from Vanderplaats [Van84, page 147]. The variational characterization of the Perron-Frobenius eigenvalue (page 165) is proved in Berman and Plemmons [BP94, page 31].

Nesterov and Nemirovski [NN94, chapter 4] introduced the conic form problem (4.49) as a standard problem format in nonlinear convex optimization. The cone programming approach is further developed in Ben-Tal and Nemirovski [BTN01], who also describe numerous applications.

Alizadeh [Ali91] and Nesterov and Nemirovski [NN94, §6.4] were the first to make a systematic study of semidefinite programming, and to point out the wide variety of applications in convex optimization. Subsequent research in semidefinite programming during the 1990s was driven by applications in combinatorial optimization (Goemans and Williamson [GW95]), control (Boyd, El Ghaoui, Feron, and Balakrishnan [BEFB94], Scherer, Gahinet, and Chilali [SGC97], Dullerud and Paganini [DP00]), communications and signal processing (Luo [Luo03], Davidson, Luo, Wong, and Ma [DLW00, MDW+02]), and other areas of engineering. The book edited by Wolkowicz, Saigal, and Vandenberghe [WSV00] and the articles by Todd [Tod01], Lewis and Overton [LO96], and Vandenberghe and Boyd [VB95] provide overviews and extensive bibliographies. Connections between SDP and moment problems, of which we give a simple example on page 170, are explored in detail by Bertsimas and Sethuraman [BS00], Nesterov [Nes00], and Lasserre [Las02]. The fastest mixing Markov chain problem is from Boyd, Diaconis, and Xiao [BDX04].

Multicriterion optimization and Pareto optimality are fundamental tools in economics; see Pareto [Par71], Debreu [Deb59] and Luenberger [Lue95]. The result in example 4.9 is known as the Gauss-Markov theorem (Kailath, Sayed, and Hassibi [KSH00, page 97]).

Exercises

Basic terminology and optimality conditions

4.1 Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x_1, x_2) \\ \text{subject to} & 2x_1 + x_2 \geq 1 \\ & x_1 + 3x_2 \geq 1 \\ & x_1 \geq 0, \quad x_2 \geq 0. \end{array}$$

Make a sketch of the feasible set. For each of the following objective functions, give the optimal set and the optimal value.

(a) $f_0(x_1, x_2) = x_1 + x_2$.

(b) $f_0(x_1, x_2) = -x_1 - x_2$.

(c) $f_0(x_1, x_2) = x_1$.

(d) $f_0(x_1, x_2) = \max\{x_1, x_2\}$.

(e) $f_0(x_1, x_2) = x_1^2 + 9x_2^2$.

4.2 Consider the optimization problem

$$\text{minimize} \quad f_0(x) = -\sum_{i=1}^m \log(b_i - a_i^T x)$$

with domain $\mathbf{dom}\, f_0 = \{x \mid Ax \prec b\}$, where $A \in \mathbf{R}^{m \times n}$ (with rows a_i^T). We assume that $\mathbf{dom}\, f_0$ is nonempty.

Prove the following facts (which include the results quoted without proof on page 141).

(a) $\mathbf{dom}\, f_0$ is unbounded if and only if there exists a $v \neq 0$ with $Av \preceq 0$.

(b) f_0 is unbounded below if and only if there exists a v with $Av \preceq 0$, $Av \neq 0$. *Hint.* There exists a v such that $Av \preceq 0$, $Av \neq 0$ if and only if there exists no $z \succ 0$ such that $A^T z = 0$. This follows from the theorem of alternatives in example 2.21, page 50.

(c) If f_0 is bounded below then its minimum is attained, *i.e.*, there exists an x that satisfies the optimality condition (4.23).

(d) The optimal set is affine: $X_{\text{opt}} = \{x^\star + v \mid Av = 0\}$, where $x^\star$ is any optimal point.

4.3 Prove that $x^\star = (1, 1/2, -1)$ is optimal for the optimization problem

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & -1 \leq x_i \leq 1, \quad i = 1, 2, 3, \end{array}$$

where

$$P = \begin{bmatrix} 13 & 12 & -2 \\ 12 & 17 & 6 \\ -2 & 6 & 12 \end{bmatrix}, \qquad q = \begin{bmatrix} -22.0 \\ -14.5 \\ 13.0 \end{bmatrix}, \qquad r = 1.$$

4.4 [P. Parrilo] *Symmetries and convex optimization.* Suppose $\mathcal{G} = \{Q_1, \ldots, Q_k\} \subseteq \mathbf{R}^{n \times n}$ is a group, *i.e.*, closed under products and inverse. We say that the function $f : \mathbf{R}^n \to \mathbf{R}$ is $\mathcal{G}$-*invariant*, or *symmetric with respect to* $\mathcal{G}$, if $f(Q_i x) = f(x)$ holds for all x and $i = 1, \ldots, k$. We define $\overline{x} = (1/k)\sum_{i=1}^k Q_i x$, which is the average of x over its $\mathcal{G}$-orbit. We define the *fixed subspace* of $\mathcal{G}$ as

$$\mathcal{F} = \{x \mid Q_i x = x,\ i = 1, \ldots, k\}.$$

(a) Show that for any $x \in \mathbf{R}^n$, we have $\overline{x} \in \mathcal{F}$.

(b) Show that if $f : \mathbf{R}^n \to \mathbf{R}$ is convex and $\mathcal{G}$-invariant, then $f(\overline{x}) \leq f(x)$.

(c) We say the optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \end{array}$$

is $\mathcal{G}$-invariant if the objective f_0 is $\mathcal{G}$-invariant, and the feasible set is $\mathcal{G}$-invariant, which means

$$f_1(x) \leq 0, \ldots, f_m(x) \leq 0 \implies f_1(Q_i x) \leq 0, \ldots, f_m(Q_i x) \leq 0,$$

for $i = 1, \ldots, k$. Show that if the problem is convex and $\mathcal{G}$-invariant, and there exists an optimal point, then there exists an optimal point in $\mathcal{F}$. In other words, we can adjoin the equality constraints $x \in \mathcal{F}$ to the problem, without loss of generality.

(d) As an example, suppose f is convex and symmetric, *i.e.*, $f(Px) = f(x)$ for every permutation P. Show that if f has a minimizer, then it has a minimizer of the form $\alpha\mathbf{1}$. (This means to minimize f over $x \in \mathbf{R}^n$, we can just as well minimize $f(t\mathbf{1})$ over $t \in \mathbf{R}$.)

4.5 *Equivalent convex problems.* Show that the following three convex problems are equivalent. Carefully explain how the solution of each problem is obtained from the solution of the other problems. The problem data are the matrix $A \in \mathbf{R}^{m \times n}$ (with rows a_i^T), the vector $b \in \mathbf{R}^m$, and the constant $M > 0$.

(a) The *robust least-squares problem*

$$\text{minimize} \quad \textstyle\sum_{i=1}^m \phi(a_i^T x - b_i),$$

with variable $x \in \mathbf{R}^n$, where $\phi : \mathbf{R} \to \mathbf{R}$ is defined as

$$\phi(u) = \begin{cases} u^2 & |u| \leq M \\ M(2|u| - M) & |u| > M. \end{cases}$$

(This function is known as the *Huber penalty function*; see §6.1.2.)

(b) The *least-squares problem with variable weights*

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m (a_i^T x - b_i)^2/(w_i + 1) + M^2 \mathbf{1}^T w \\ \text{subject to} & w \succeq 0, \end{array}$$

with variables $x \in \mathbf{R}^n$ and $w \in \mathbf{R}^m$, and domain $\mathcal{D} = \{(x, w) \in \mathbf{R}^n \times \mathbf{R}^m \mid w \succ -\mathbf{1}\}$. *Hint.* Optimize over w assuming x is fixed, to establish a relation with the problem in part (a).

(This problem can be interpreted as a weighted least-squares problem in which we are allowed to adjust the weight of the ith residual. The weight is one if $w_i = 0$, and decreases if we increase w_i. The second term in the objective penalizes large values of w, *i.e.*, large adjustments of the weights.)

(c) The *quadratic program*

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m (u_i^2 + 2Mv_i) \\ \text{subject to} & -u - v \preceq Ax - b \preceq u + v \\ & 0 \preceq u \preceq M\mathbf{1} \\ & v \succeq 0. \end{array}$$

4.6 *Handling convex equality constraints.* A convex optimization problem can have only *linear* equality constraint functions. In some special cases, however, it is possible to handle convex equality constraint functions, *i.e.*, constraints of the form $h(x) = 0$, where h is convex. We explore this idea in this problem.

Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h(x) = 0, \end{array} \tag{4.65}$$

where f_i and h are convex functions with domain $\mathbf{R}^n$. Unless h is affine, this is *not* a convex optimization problem. Consider the related problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m, \\ & h(x) \leq 0, \end{array} \tag{4.66}$$

where the convex equality constraint has been relaxed to a convex inequality. This problem is, of course, convex.

Now suppose we can guarantee that at any optimal solution $x^\star$ of the convex problem (4.66), we have $h(x^\star) = 0$, *i.e.*, the inequality $h(x) \leq 0$ is always active at the solution. Then we can solve the (nonconvex) problem (4.65) by solving the convex problem (4.66).

Show that this is the case if there is an index r such that

- f_0 is monotonically increasing in x_r
- $f_1, \ldots, f_m$ are nondecreasing in x_r
- h is monotonically decreasing in x_r.

We will see specific examples in exercises 4.31 and 4.58.

4.7 *Convex-concave fractional problems.* Consider a problem of the form

$$\begin{array}{ll} \text{minimize} & f_0(x)/(c^T x + d) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

where $f_0, f_1, \ldots, f_m$ are convex, and the domain of the objective function is defined as $\{x \in \mathbf{dom}\, f_0 \mid c^T x + d > 0\}$.

(a) Show that this is a quasiconvex optimization problem.

(b) Show that the problem is equivalent to

$$\begin{array}{ll} \text{minimize} & g_0(y, t) \\ \text{subject to} & g_i(y, t) \leq 0, \quad i = 1, \ldots, m \\ & Ay = bt \\ & c^T y + dt = 1, \end{array}$$

where g_i is the perspective of f_i (see §3.2.6). The variables are $y \in \mathbf{R}^n$ and $t \in \mathbf{R}$. Show that this problem is convex.

(c) Following a similar argument, derive a convex formulation for the *convex-concave* fractional problem

$$\begin{array}{ll} \text{minimize} & f_0(x)/h(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

where $f_0, f_1, \ldots, f_m$ are convex, h is concave, the domain of the objective function is defined as $\{x \in \mathbf{dom}\, f_0 \cap \mathbf{dom}\, h \mid h(x) > 0\}$ and $f_0(x) \geq 0$ everywhere.

As an example, apply your technique to the (unconstrained) problem with

$$f_0(x) = (\mathbf{tr}\, F(x))/m, \qquad h(x) = (\det(F(x))^{1/m},$$

with $\mathbf{dom}(f_0/h) = \{x \mid F(x) \succ 0\}$, where $F(x) = F_0 + x_1 F_1 + \cdots + x_n F_n$ for given $F_i \in \mathbf{S}^m$. In this problem, we minimize the ratio of the arithmetic mean over the geometric mean of the eigenvalues of an affine matrix function $F(x)$.

Linear optimization problems

4.8 *Some simple LPs.* Give an explicit solution of each of the following LPs.

(a) *Minimizing a linear function over an affine set.*

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b. \end{array}$$

(b) *Minimizing a linear function over a halfspace.*

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a^T x \leq b, \end{array}$$

where $a \neq 0$.

(c) *Minimizing a linear function over a rectangle.*

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & l \preceq x \preceq u, \end{array}$$

where l and u satisfy $l \preceq u$.

(d) *Minimizing a linear function over the probability simplex.*

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \mathbf{1}^T x = 1, \quad x \succeq 0. \end{array}$$

What happens if the equality constraint is replaced by an inequality $\mathbf{1}^T x \leq 1$?

We can interpret this LP as a simple portfolio optimization problem. The vector x represents the allocation of our total budget over different assets, with x_i the fraction invested in asset i. The return of each investment is fixed and given by $-c_i$, so our total return (which we want to maximize) is $-c^T x$. If we replace the budget constraint $\mathbf{1}^T x = 1$ with an inequality $\mathbf{1}^T x \leq 1$, we have the option of not investing a portion of the total budget.

(e) *Minimizing a linear function over a unit box with a total budget constraint.*

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \mathbf{1}^T x = \alpha, \quad 0 \preceq x \preceq \mathbf{1}, \end{array}$$

where α is an integer between 0 and n. What happens if α is not an integer (but satisfies $0 \leq \alpha \leq n$)? What if we change the equality to an inequality $\mathbf{1}^T x \leq \alpha$?

(f) *Minimizing a linear function over a unit box with a weighted budget constraint.*

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & d^T x = \alpha, \quad 0 \preceq x \preceq \mathbf{1}, \end{array}$$

with $d \succ 0$, and $0 \leq \alpha \leq \mathbf{1}^T d$.

4.9 *Square LP.* Consider the LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \end{array}$$

with A square and nonsingular. Show that the optimal value is given by

$$p^\star = \begin{cases} c^T A^{-1} b & A^{-T} c \preceq 0 \\ -\infty & \text{otherwise.} \end{cases}$$

4.10 *Converting general LP to standard form.* Work out the details on page 147 of §4.3. Explain in detail the relation between the feasible sets, the optimal solutions, and the optimal values of the standard form LP and the original LP.

4.11 *Problems involving ℓ_1- and ℓ_∞-norms.* Formulate the following problems as LPs. Explain in detail the relation between the optimal solution of each problem and the solution of its equivalent LP.

(a) Minimize $\|Ax - b\|_\infty$ (ℓ_∞-norm approximation).

(b) Minimize $\|Ax - b\|_1$ (ℓ_1-norm approximation).

(c) Minimize $\|Ax - b\|_1$ subject to $\|x\|_\infty \leq 1$.

(d) Minimize $\|x\|_1$ subject to $\|Ax - b\|_\infty \leq 1$.

(e) Minimize $\|Ax - b\|_1 + \|x\|_\infty$.

In each problem, $A \in \mathbf{R}^{m \times n}$ and $b \in \mathbf{R}^m$ are given. (See §6.1 for more problems involving approximation and constrained approximation.)

4.12 *Network flow problem.* Consider a network of n nodes, with directed links connecting each pair of nodes. The variables in the problem are the flows on each link: x_{ij} will denote the flow from node i to node j. The cost of the flow along the link from node i to node j is given by $c_{ij} x_{ij}$, where c_{ij} are given constants. The total cost across the network is

$$C = \sum_{i,j=1}^{n} c_{ij} x_{ij}.$$

Each link flow x_{ij} is also subject to a given lower bound l_{ij} (usually assumed to be nonnegative) and an upper bound u_{ij}.

The external supply at node i is given by b_i, where $b_i > 0$ means an external flow enters the network at node i, and $b_i < 0$ means that at node i, an amount $|b_i|$ flows out of the network. We assume that $\mathbf{1}^T b = 0$, *i.e.*, the total external supply equals total external demand. At each node we have conservation of flow: the total flow into node i along links and the external supply, minus the total flow out along the links, equals zero.

The problem is to minimize the total cost of flow through the network, subject to the constraints described above. Formulate this problem as an LP.

4.13 *Robust LP with interval coefficients.* Consider the problem, with variable $x \in \mathbf{R}^n$,

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \text{ for all } A \in \mathcal{A}, \end{array}$$

where $\mathcal{A} \subseteq \mathbf{R}^{m \times n}$ is the set

$$\mathcal{A} = \{A \in \mathbf{R}^{m \times n} \mid \bar{A}_{ij} - V_{ij} \leq A_{ij} \leq \bar{A}_{ij} + V_{ij},\ i = 1, \ldots, m,\ j = 1, \ldots, n\}.$$

(The matrices $\bar{A}$ and V are given.) This problem can be interpreted as an LP where each coefficient of A is only known to lie in an interval, and we require that x must satisfy the constraints for all possible values of the coefficients.

Express this problem as an LP. The LP you construct should be efficient, *i.e.*, it should not have dimensions that grow exponentially with n or m.

4.14 *Approximating a matrix in infinity norm.* The ℓ_∞-norm induced norm of a matrix $A \in \mathbf{R}^{m\times n}$, denoted $\|A\|_\infty$, is given by

$$\|A\|_\infty = \sup_{x\neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty} = \max_{i=1,\ldots,m} \sum_{j=1}^{n} |a_{ij}|.$$

This norm is sometimes called the max-row-sum norm, for obvious reasons (see §A.1.5).

Consider the problem of approximating a matrix, in the max-row-sum norm, by a linear combination of other matrices. That is, we are given $k+1$ matrices $A_0, \ldots, A_k \in \mathbf{R}^{m\times n}$, and need to find $x \in \mathbf{R}^k$ that minimizes

$$\|A_0 + x_1 A_1 + \cdots + x_k A_k\|_\infty.$$

Express this problem as a linear program. Explain the significance of any extra variables in your LP. Carefully explain how your LP formulation solves this problem, *e.g.*, what is the relation between the feasible set for your LP and this problem?

4.15 *Relaxation of Boolean LP.* In a *Boolean linear program*, the variable x is constrained to have components equal to zero or one:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & x_i \in \{0,1\}, \quad i = 1,\ldots,n. \end{array} \tag{4.67}$$

In general, such problems are very difficult to solve, even though the feasible set is finite (containing at most 2^n points).

In a general method called *relaxation*, the constraint that x_i be zero or one is replaced with the linear inequalities $0 \le x_i \le 1$:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & 0 \le x_i \le 1, \quad i = 1,\ldots,n. \end{array} \tag{4.68}$$

We refer to this problem as the *LP relaxation* of the Boolean LP (4.67). The LP relaxation is far easier to solve than the original Boolean LP.

(a) Show that the optimal value of the LP relaxation (4.68) is a lower bound on the optimal value of the Boolean LP (4.67). What can you say about the Boolean LP if the LP relaxation is infeasible?

(b) It sometimes happens that the LP relaxation has a solution with $x_i \in \{0,1\}$. What can you say in this case?

4.16 *Minimum fuel optimal control.* We consider a linear dynamical system with state $x(t) \in \mathbf{R}^n$, $t = 0,\ldots,N$, and actuator or input signal $u(t) \in \mathbf{R}$, for $t = 0,\ldots,N-1$. The dynamics of the system is given by the linear recurrence

$$x(t+1) = Ax(t) + bu(t), \quad t = 0,\ldots,N-1,$$

where $A \in \mathbf{R}^{n\times n}$ and $b \in \mathbf{R}^n$ are given. We assume that the initial state is zero, *i.e.*, $x(0) = 0$.

The *minimum fuel optimal control problem* is to choose the inputs $u(0),\ldots,u(N-1)$ so as to minimize the total fuel consumed, which is given by

$$F = \sum_{t=0}^{N-1} f(u(t)),$$

subject to the constraint that $x(N) = x_{\mathrm{des}}$, where N is the (given) time horizon, and $x_{\mathrm{des}} \in \mathbf{R}^n$ is the (given) desired final or target state. The function $f : \mathbf{R} \to \mathbf{R}$ is the *fuel use map* for the actuator, and gives the amount of fuel used as a function of the actuator signal amplitude. In this problem we use

$$f(a) = \begin{cases} |a| & |a| \le 1 \\ 2|a| - 1 & |a| > 1. \end{cases}$$

This means that fuel use is proportional to the absolute value of the actuator signal, for actuator signals between -1 and 1; for larger actuator signals the marginal fuel efficiency is half.

Formulate the minimum fuel optimal control problem as an LP.

4.17 *Optimal activity levels.* We consider the selection of n nonnegative activity levels, denoted $x_1, \ldots, x_n$. These activities consume m resources, which are limited. Activity j consumes $A_{ij}x_j$ of resource i, where A_{ij} are given. The total resource consumption is additive, so the total of resource i consumed is $c_i = \sum_{j=1}^n A_{ij}x_j$. (Ordinarily we have $A_{ij} \ge 0$, *i.e.*, activity j consumes resource i. But we allow the possibility that $A_{ij} < 0$, which means that activity j actually *generates* resource i as a by-product.) Each resource consumption is limited: we must have $c_i \le c_i^{\max}$, where $c_i^{\max}$ are given. Each activity generates revenue, which is a piecewise-linear concave function of the activity level:

$$r_j(x_j) = \begin{cases} p_j x_j & 0 \le x_j \le q_j \\ p_j q_j + p_j^{\mathrm{disc}}(x_j - q_j) & x_j \ge q_j. \end{cases}$$

Here $p_j > 0$ is the basic price, $q_j > 0$ is the quantity discount level, and p_j^{disc} is the quantity discount price, for (the product of) activity j. (We have $0 < p_j^{\mathrm{disc}} < p_j$.) The total revenue is the sum of the revenues associated with each activity, *i.e.*, $\sum_{j=1}^n r_j(x_j)$. The goal is to choose activity levels that maximize the total revenue while respecting the resource limits. Show how to formulate this problem as an LP.

4.18 *Separating hyperplanes and spheres.* Suppose you are given two sets of points in $\mathbf{R}^n$, $\{v^1, v^2, \ldots, v^K\}$ and $\{w^1, w^2, \ldots, w^L\}$. Formulate the following two problems as LP feasibility problems.

(a) Determine a hyperplane that separates the two sets, *i.e.*, find $a \in \mathbf{R}^n$ and $b \in \mathbf{R}$ with $a \ne 0$ such that

$$a^T v^i \le b, \quad i = 1, \ldots, K, \qquad a^T w^i \ge b, \quad i = 1, \ldots, L.$$

Note that we require $a \ne 0$, so you have to make sure that your formulation excludes the trivial solution $a = 0$, $b = 0$. You can assume that

$$\mathbf{rank} \begin{bmatrix} v^1 & v^2 & \cdots & v^K & w^1 & w^2 & \cdots & w^L \\ 1 & 1 & \cdots & 1 & 1 & 1 & \cdots & 1 \end{bmatrix} = n + 1$$

(*i.e.*, the affine hull of the $K + L$ points has dimension n).

(b) Determine a sphere separating the two sets of points, *i.e.*, find $x_c \in \mathbf{R}^n$ and $R \ge 0$ such that

$$\|v^i - x_c\|_2 \le R, \quad i = 1, \ldots, K, \qquad \|w^i - x_c\|_2 \ge R, \quad i = 1, \ldots, L.$$

(Here x_c is the center of the sphere; R is its radius.)

(See chapter 8 for more on separating hyperplanes, separating spheres, and related topics.)

4.19 Consider the problem

$$\begin{array}{ll} \text{minimize} & \|Ax-b\|_1/(c^Tx+d) \\ \text{subject to} & \|x\|_\infty \le 1, \end{array}$$

where $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$, $c \in \mathbf{R}^n$, and $d \in \mathbf{R}$. We assume that $d > \|c\|_1$, which implies that $c^Tx + d > 0$ for all feasible x.

(a) Show that this is a quasiconvex optimization problem.

(b) Show that it is equivalent to the convex optimization problem

$$\begin{array}{ll} \text{minimize} & \|Ay-bt\|_1 \\ \text{subject to} & \|y\|_\infty \le t \\ & c^Ty + dt = 1, \end{array}$$

with variables $y \in \mathbf{R}^n$, $t \in \mathbf{R}$.

4.20 *Power assignment in a wireless communication system.* We consider n transmitters with powers $p_1, \ldots, p_n \ge 0$, transmitting to n receivers. These powers are the optimization variables in the problem. We let $G \in \mathbf{R}^{n\times n}$ denote the matrix of *path gains* from the transmitters to the receivers; $G_{ij} \ge 0$ is the path gain from transmitter j to receiver i. The *signal power* at receiver i is then $S_i = G_{ii}p_i$, and the *interference power* at receiver i is $I_i = \sum_{k\ne i} G_{ik}p_k$. The *signal to interference plus noise ratio*, denoted SINR, at receiver i, is given by $S_i/(I_i + \sigma_i)$, where $\sigma_i > 0$ is the (self-) noise power in receiver i. The objective in the problem is to maximize the minimum SINR ratio, over all receivers, *i.e.*, to maximize

$$\min_{i=1,\ldots,n} \frac{S_i}{I_i + \sigma_i}.$$

There are a number of constraints on the powers that must be satisfied, in addition to the obvious one $p_i \ge 0$. The first is a maximum allowable power for each transmitter, *i.e.*, $p_i \le P_i^{\max}$, where $P_i^{\max} > 0$ is given. In addition, the transmitters are partitioned into groups, with each group sharing the same power supply, so there is a total power constraint for each group of transmitter powers. More precisely, we have subsets $K_1, \ldots, K_m$ of $\{1, \ldots, n\}$ with $K_1 \cup \cdots \cup K_m = \{1, \ldots, n\}$, and $K_j \cap K_l = 0$ if $j \ne l$. For each group K_l, the total associated transmitter power cannot exceed $P_l^{\mathrm{gp}} > 0$:

$$\sum_{k\in K_l} p_k \le P_l^{\mathrm{gp}}, \quad l = 1, \ldots, m.$$

Finally, we have a limit $P_k^{\mathrm{rc}} > 0$ on the total received power at each receiver:

$$\sum_{k=1}^n G_{ik}p_k \le P_i^{\mathrm{rc}}, \quad i = 1, \ldots, n.$$

(This constraint reflects the fact that the receivers will saturate if the total received power is too large.)

Formulate the SINR maximization problem as a generalized linear-fractional program.

Quadratic optimization problems

4.21 *Some simple QCQPs.* Give an explicit solution of each of the following QCQPs.

(a) *Minimizing a linear function over an ellipsoid centered at the origin.*

$$\begin{array}{ll} \text{minimize} & c^Tx \\ \text{subject to} & x^TAx \le 1, \end{array}$$

where $A \in \mathbf{S}^n_{++}$ and $c \ne 0$. What is the solution if the problem is not convex ($A \notin \mathbf{S}^n_+$)?

(b) *Minimizing a linear function over an ellipsoid.*

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & (x - x_c)^T A (x - x_c) \leq 1, \end{array}$$

where $A \in \mathbf{S}^n_{++}$ and $c \neq 0$.

(c) *Minimizing a quadratic form over an ellipsoid centered at the origin.*

$$\begin{array}{ll} \text{minimize} & x^T B x \\ \text{subject to} & x^T A x \leq 1, \end{array}$$

where $A \in \mathbf{S}^n_{++}$ and $B \in \mathbf{S}^n_{+}$. Also consider the nonconvex extension with $B \notin \mathbf{S}^n_{+}$. (See §B.1.)

4.22 Consider the QCQP

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & x^T x \leq 1, \end{array}$$

with $P \in \mathbf{S}^n_{++}$. Show that $x^\star = -(P + \lambda I)^{-1} q$ where $\lambda = \max\{0, \bar{\lambda}\}$ and $\bar{\lambda}$ is the largest solution of the nonlinear equation

$$q^T (P + \lambda I)^{-2} q = 1.$$

4.23 *ℓ_4-norm approximation via QCQP.* Formulate the ℓ_4-norm approximation problem

$$\text{minimize} \quad \|Ax - b\|_4 = (\textstyle\sum_{i=1}^m (a_i^T x - b_i)^4)^{1/4}$$

as a QCQP. The matrix $A \in \mathbf{R}^{m \times n}$ (with rows a_i^T) and the vector $b \in \mathbf{R}^m$ are given.

4.24 *Complex ℓ_1-, ℓ_2- and ℓ_∞-norm approximation.* Consider the problem

$$\text{minimize} \quad \|Ax - b\|_p,$$

where $A \in \mathbf{C}^{m \times n}$, $b \in \mathbf{C}^m$, and the variable is $x \in \mathbf{C}^n$. The complex ℓ_p-norm is defined by

$$\|y\|_p = \left(\sum_{i=1}^m |y_i|^p \right)^{1/p}$$

for $p \geq 1$, and $\|y\|_\infty = \max_{i=1,\ldots,m} |y_i|$. For $p = 1$, 2, and ∞, express the complex ℓ_p-norm approximation problem as a QCQP or SOCP with real variables and data.

4.25 *Linear separation of two sets of ellipsoids.* Suppose we are given $K + L$ ellipsoids

$$\mathcal{E}_i = \{P_i u + q_i \mid \|u\|_2 \leq 1\}, \quad i = 1, \ldots, K + L,$$

where $P_i \in \mathbf{S}^n$. We are interested in finding a hyperplane that strictly separates $\mathcal{E}_1, \ldots, \mathcal{E}_K$ from $\mathcal{E}_{K+1}, \ldots, \mathcal{E}_{K+L}$, *i.e.*, we want to compute $a \in \mathbf{R}^n$, $b \in \mathbf{R}$ such that

$$a^T x + b > 0 \text{ for } x \in \mathcal{E}_1 \cup \cdots \cup \mathcal{E}_K, \qquad a^T x + b < 0 \text{ for } x \in \mathcal{E}_{K+1} \cup \cdots \cup \mathcal{E}_{K+L},$$

or prove that no such hyperplane exists. Express this problem as an SOCP feasibility problem.

4.26 *Hyperbolic constraints as SOC constraints.* Verify that $x \in \mathbf{R}^n$, $y, z \in \mathbf{R}$ satisfy

$$x^T x \leq yz, \qquad y \geq 0, \qquad z \geq 0$$

if and only if

$$\left\| \begin{bmatrix} 2x \\ y - z \end{bmatrix} \right\|_2 \leq y + z, \qquad y \geq 0, \qquad z \geq 0.$$

Use this observation to cast the following problems as SOCPs.

(a) *Maximizing harmonic mean.*

$$\text{maximize} \quad \left(\sum_{i=1}^m 1/(a_i^T x - b_i)\right)^{-1},$$

with domain $\{x \mid Ax \succ b\}$, where a_i^T is the ith row of A.

(b) *Maximizing geometric mean.*

$$\text{maximize} \quad \left(\prod_{i=1}^m (a_i^T x - b_i)\right)^{1/m},$$

with domain $\{x \mid Ax \succeq b\}$, where a_i^T is the ith row of A.

4.27 *Matrix fractional minimization via SOCP.* Express the following problem as an SOCP:

$$\begin{array}{ll} \text{minimize} & (Ax+b)^T (I + B\,\mathbf{diag}(x) B^T)^{-1}(Ax+b) \\ \text{subject to} & x \succeq 0, \end{array}$$

with $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$, $B \in \mathbf{R}^{m\times n}$. The variable is $x \in \mathbf{R}^n$.

Hint. First show that the problem is equivalent to

$$\begin{array}{ll} \text{minimize} & v^T v + w^T \mathbf{diag}(x)^{-1} w \\ \text{subject to} & v + Bw = Ax + b \\ & x \succeq 0, \end{array}$$

with variables $v \in \mathbf{R}^m$, $w, x \in \mathbf{R}^n$. (If $x_i = 0$ we interpret w_i^2/x_i as zero if $w_i = 0$ and as ∞ otherwise.) Then use the results of exercise 4.26.

4.28 *Robust quadratic programming.* In §4.4.2 we discussed robust linear programming as an application of second-order cone programming. In this problem we consider a similar robust variation of the (convex) *quadratic* program

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Ax \preceq b. \end{array}$$

For simplicity we assume that only the matrix P is subject to errors, and the other parameters (q, r, A, b) are exactly known. The robust quadratic program is defined as

$$\begin{array}{ll} \text{minimize} & \sup_{P\in\mathcal{E}}((1/2)x^T P x + q^T x + r) \\ \text{subject to} & Ax \preceq b \end{array}$$

where $\mathcal{E}$ is the set of possible matrices P.

For each of the following sets $\mathcal{E}$, express the robust QP as a convex problem. Be as specific as you can. If the problem can be expressed in a standard form (*e.g.*, QP, QCQP, SOCP, SDP), say so.

(a) A finite set of matrices: $\mathcal{E} = \{P_1, \ldots, P_K\}$, where $P_i \in \mathbf{S}_+^n$, $i = 1, \ldots, K$.

(b) A set specified by a nominal value $P_0 \in \mathbf{S}_+^n$ plus a bound on the eigenvalues of the deviation $P - P_0$:

$$\mathcal{E} = \{P \in \mathbf{S}^n \mid -\gamma I \preceq P - P_0 \preceq \gamma I\}$$

where $\gamma \in \mathbf{R}$ and $P_0 \in \mathbf{S}_+^n$,

(c) An ellipsoid of matrices:

$$\mathcal{E} = \left\{ P_0 + \sum_{i=1}^K P_i u_i \;\middle|\; \|u\|_2 \le 1 \right\}.$$

You can assume $P_i \in \mathbf{S}_+^n$, $i = 0, \ldots, K$.

4.29 *Maximizing probability of satisfying a linear inequality.* Let c be a random variable in $\mathbf{R}^n$, normally distributed with mean $\bar{c}$ and covariance matrix R. Consider the problem

$$\begin{array}{ll} \text{maximize} & \mathbf{prob}(c^T x \geq \alpha) \\ \text{subject to} & Fx \preceq g, \quad Ax = b. \end{array}$$

Assuming there exists a feasible point $\tilde{x}$ for which $\bar{c}^T \tilde{x} \geq \alpha$, show that this problem is equivalent to a convex or quasiconvex optimization problem. Formulate the problem as a QP, QCQP, or SOCP (if the problem is convex), or explain how you can solve it by solving a sequence of QP, QCQP, or SOCP feasibility problems (if the problem is quasiconvex).

Geometric programming

4.30 A heated fluid at temperature T (degrees above ambient temperature) flows in a pipe with fixed length and circular cross section with radius r. A layer of insulation, with thickness $w \ll r$, surrounds the pipe to reduce heat loss through the pipe walls. The design variables in this problem are T, r, and w.

The heat loss is (approximately) proportional to Tr/w, so over a fixed lifetime, the energy cost due to heat loss is given by $\alpha_1 Tr/w$. The cost of the pipe, which has a fixed wall thickness, is approximately proportional to the total material, *i.e.*, it is given by $\alpha_2 r$. The cost of the insulation is also approximately proportional to the total insulation material, *i.e.*, $\alpha_3 rw$ (using $w \ll r$). The total cost is the sum of these three costs.

The heat flow down the pipe is entirely due to the flow of the fluid, which has a fixed velocity, *i.e.*, it is given by $\alpha_4 Tr^2$. The constants α_i are all positive, as are the variables T, r, and w.

Now the problem: maximize the total heat flow down the pipe, subject to an upper limit $C_{\max}$ on total cost, and the constraints

$$T_{\min} \leq T \leq T_{\max}, \qquad r_{\min} \leq r \leq r_{\max}, \qquad w_{\min} \leq w \leq w_{\max}, \quad w \leq 0.1r.$$

Express this problem as a geometric program.

4.31 *Recursive formulation of optimal beam design problem.* Show that the GP (4.46) is equivalent to the GP

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^N w_i h_i \\ \text{subject to} & w_i/w_{\max} \leq 1, \quad w_{\min}/w_i \leq 1, \quad i = 1, \ldots, N \\ & h_i/h_{\max} \leq 1, \quad h_{\min}/h_i \leq 1, \quad i = 1, \ldots, N \\ & h_i/(w_i S_{\max}) \leq 1, \quad S_{\min} w_i/h_i \leq 1, \quad i = 1, \ldots, N \\ & 6iF/(\sigma_{\max} w_i h_i^2) \leq 1, \quad i = 1, \ldots, N \\ & (2i-1)d_i/v_i + v_{i+1}/v_i \leq 1, \quad i = 1, \ldots, N \\ & (i - 1/3)d_i/y_i + v_{i+1}/y_i + y_{i+1}/y_i \leq 1, \quad i = 1, \ldots, N \\ & y_1/y_{\max} \leq 1 \\ & E w_i h_i^3 d_i/(6F) = 1, \quad i = 1, \ldots, N. \end{array}$$

The variables are w_i, h_i, v_i, d_i, y_i for $i = 1, \ldots, N$.

4.32 *Approximating a function as a monomial.* Suppose the function $f : \mathbf{R}^n \to \mathbf{R}$ is differentiable at a point $x_0 \succ 0$, with $f(x_0) > 0$. How would you find a monomial function $\hat{f} : \mathbf{R}^n \to \mathbf{R}$ such that $f(x_0) = \hat{f}(x_0)$ and for x near x_0, $\hat{f}(x)$ is very near $f(x)$?

4.33 Express the following problems as convex optimization problems.

(a) Minimize $\max\{p(x), q(x)\}$, where p and q are posynomials.

(b) Minimize $\exp(p(x)) + \exp(q(x))$, where p and q are posynomials.

(c) Minimize $p(x)/(r(x) - q(x))$, subject to $r(x) > q(x)$, where p, q are posynomials, and r is a monomial.

4.34 *Log-convexity of Perron-Frobenius eigenvalue.* Let $A \in \mathbf{R}^{n \times n}$ be an elementwise positive matrix, *i.e.*, $A_{ij} > 0$. (The results of this problem hold for irreducible nonnegative matrices as well.) Let $\lambda_{\mathrm{pf}}(A)$ denotes its Perron-Frobenius eigenvalue, *i.e.*, its eigenvalue of largest magnitude. (See the definition and the example on page 165.) Show that $\log \lambda_{\mathrm{pf}}(A)$ is a convex function of $\log A_{ij}$. This means, for example, that we have the inequality

$$\lambda_{\mathrm{pf}}(C) \leq (\lambda_{\mathrm{pf}}(A)\lambda_{\mathrm{pf}}(B))^{1/2},$$

where $C_{ij} = (A_{ij}B_{ij})^{1/2}$, and A and B are elementwise positive matrices.

Hint. Use the characterization of the Perron-Frobenius eigenvalue given in (4.47), or, alternatively, use the characterization

$$\log \lambda_{\mathrm{pf}}(A) = \lim_{k \to \infty} (1/k) \log(\mathbf{1}^T A^k \mathbf{1}).$$

4.35 *Signomial and geometric programs.* A *signomial* is a linear combination of monomials of some positive variables $x_1, \dots, x_n$. Signomials are more general than posynomials, which are signomials with all positive coefficients. A *signomial program* is an optimization problem of the form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p, \end{array}$$

where $f_0, \dots, f_m$ and $h_1, \dots, h_p$ are signomials. In general, signomial programs are very difficult to solve.

Some signomial programs can be transformed to GPs, and therefore solved efficiently. Show how to do this for a signomial program of the following form:

- The objective signomial f_0 is a posynomial, *i.e.*, its terms have only positive coefficients.
- Each inequality constraint signomial $f_1, \dots, f_m$ has exactly one term with a negative coefficient: $f_i = p_i - q_i$ where p_i is posynomial, and q_i is monomial.
- Each equality constraint signomial $h_1, \dots, h_p$ has exactly one term with a positive coefficient and one term with a negative coefficient: $h_i = r_i - s_i$ where r_i and s_i are monomials.

4.36 Explain how to reformulate a general GP as an equivalent GP in which every posynomial (in the objective and constraints) has at most two monomial terms. *Hint.* Express each sum (of monomials) as a sum of sums, each with two terms.

4.37 *Generalized posynomials and geometric programming.* Let $x_1, \dots, x_n$ be positive variables, and suppose the functions $f_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \dots, k$, are posynomials of $x_1, \dots, x_n$. If $\phi : \mathbf{R}^k \to \mathbf{R}$ is a polynomial with nonnegative coefficients, then the composition

$$h(x) = \phi(f_1(x), \dots, f_k(x)) \tag{4.69}$$

is a posynomial, since posynomials are closed under products, sums, and multiplication by nonnegative scalars. For example, suppose f_1 and f_2 are posynomials, and consider the polynomial $\phi(z_1, z_2) = 3z_1^2 z_2 + 2z_1 + 3z_2^3$ (which has nonnegative coefficients). Then $h = 3f_1^2 f_2 + 2f_1 + f_2^3$ is a posynomial.

In this problem we consider a generalization of this idea, in which ϕ is allowed to be a posynomial, *i.e.*, can have fractional exponents. Specifically, assume that $\phi : \mathbf{R}^k \to \mathbf{R}$ is a posynomial, with all its exponents nonnegative. In this case we will call the function h defined in (4.69) a *generalized posynomial*. As an example, suppose f_1 and f_2 are posynomials, and consider the posynomial (with nonnegative exponents) $\phi(z_1, z_2) = 2z_1^{0.3} z_2^{1.2} + z_1 z_2^{0.5} + 2$. Then the function

$$h(x) = 2f_1(x)^{0.3} f_2(x)^{1.2} + f_1(x) f_2(x)^{0.5} + 2$$

is a generalized posynomial. Note that it is *not* a posynomial, however (unless f_1 and f_2 are monomials or constants).

A *generalized geometric program* (GGP) is an optimization problem of the form

$$\begin{array}{ll} \text{minimize} & h_0(x) \\ \text{subject to} & h_i(x) \le 1, \quad i = 1, \dots, m \\ & g_i(x) = 1, \quad i = 1, \dots, p, \end{array} \tag{4.70}$$

where $g_1, \dots, g_p$ are monomials, and $h_0, \dots, h_m$ are generalized posynomials.

Show how to express this generalized geometric program as an equivalent geometric program. Explain any new variables you introduce, and explain how your GP is equivalent to the GGP (4.70).

Semidefinite programming and conic form problems

4.38 *LMIs and SDPs with one variable.* The *generalized eigenvalues* of a matrix pair (A, B), where $A, B \in \mathbf{S}^n$, are defined as the roots of the polynomial $\det(\lambda B - A)$ (see §A.5.3).

Suppose B is nonsingular, and that A and B can be simultaneously diagonalized by a congruence, *i.e.*, there exists a nonsingular $R \in \mathbf{R}^{n \times n}$ such that

$$R^T A R = \mathbf{diag}(a), \qquad R^T B R = \mathbf{diag}(b),$$

where $a, b \in \mathbf{R}^n$. (A sufficient condition for this to hold is that there exists t_1, t_2 such that $t_1 A + t_2 B \succ 0$.)

(a) Show that the generalized eigenvalues of (A, B) are real, and given by $\lambda_i = a_i / b_i$, $i = 1, \dots, n$.

(b) Express the solution of the SDP

$$\begin{array}{ll} \text{minimize} & ct \\ \text{subject to} & tB \preceq A, \end{array}$$

with variable $t \in \mathbf{R}$, in terms of a and b.

4.39 *SDPs and congruence transformations.* Consider the SDP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + x_2 F_2 + \cdots + x_n F_n + G \preceq 0, \end{array}$$

with $F_i, G \in \mathbf{S}^k$, $c \in \mathbf{R}^n$.

(a) Suppose $R \in \mathbf{R}^{k \times k}$ is nonsingular. Show that the SDP is equivalent to the SDP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 \tilde{F}_1 + x_2 \tilde{F}_2 + \cdots + x_n \tilde{F}_n + \tilde{G} \preceq 0, \end{array}$$

where $\tilde{F}_i = R^T F_i R$, $\tilde{G} = R^T G R$.

(b) Suppose there exists a nonsingular R such that $\tilde{F}_i$ and $\tilde{G}$ are diagonal. Show that the SDP is equivalent to an LP.

(c) Suppose there exists a nonsingular R such that $\tilde{F}_i$ and $\tilde{G}$ have the form

$$\tilde{F}_i = \left[\begin{array}{cc} \alpha_i I & a_i \\ a_i^T & \alpha_i \end{array} \right], \quad i = 1, \dots, n, \qquad \tilde{G} = \left[\begin{array}{cc} \beta I & b \\ b^T & \beta \end{array} \right],$$

where $\alpha_i, \beta \in \mathbf{R}$, $a_i, b \in \mathbf{R}^{k-1}$. Show that the SDP is equivalent to an SOCP with a single second-order cone constraint.

4.40 *LPs, QPs, QCQPs, and SOCPs as SDPs.* Express the following problems as SDPs.

(a) The LP (4.27).

(b) The QP (4.34), the QCQP (4.35) and the SOCP (4.36). *Hint.* Suppose $A \in \mathbf{S}^r_{++}$, $C \in \mathbf{S}^s$, and $B \in \mathbf{R}^{r\times s}$. Then

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0 \iff C - B^T A^{-1} B \succeq 0.$$

For a more complete statement, which applies also to singular A, and a proof, see §A.5.5.

(c) The matrix fractional optimization problem

$$\text{minimize} \quad (Ax+b)^T F(x)^{-1}(Ax+b)$$

where $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$,

$$F(x) = F_0 + x_1 F_1 + \cdots + x_n F_n,$$

with $F_i \in \mathbf{S}^m$, and we take the domain of the objective to be $\{x \mid F(x) \succ 0\}$. You can assume the problem is feasible (there exists at least one x with $F(x) \succ 0$).

4.41 *LMI tests for copositive matrices and P_0-matrices.* A matrix $A \in \mathbf{S}^n$ is said to be *copositive* if $x^T Ax \geq 0$ for all $x \succeq 0$ (see exercise 2.35). A matrix $A \in \mathbf{R}^{n\times n}$ is said to be a *P_0-matrix* if $\max_{i=1,\ldots,n} x_i(Ax)_i \geq 0$ for all x. Checking whether a matrix is copositive or a P_0-matrix is very difficult in general. However, there exist useful sufficient conditions that can be verified using semidefinite programming.

(a) Show that A is copositive if it can be decomposed as a sum of a positive semidefinite and an elementwise nonnegative matrix:

$$A = B + C, \qquad B \succeq 0, \qquad C_{ij} \geq 0, \quad i,j = 1,\ldots,n. \tag{4.71}$$

Express the problem of finding B and C that satisfy (4.71) as an SDP feasibility problem.

(b) Show that A is a P_0-matrix if there exists a positive diagonal matrix D such that

$$DA + A^T D \succeq 0. \tag{4.72}$$

Express the problem of finding a D that satisfies (4.72) as an SDP feasibility problem.

4.42 *Complex LMIs and SDPs.* A complex LMI has the form

$$x_1 F_1 + \cdots + x_n F_n + G \preceq 0$$

where $F_1, \ldots, F_n$, G are complex $n \times n$ Hermitian matrices, *i.e.*, $F_i^H = F_i$, $G^H = G$, and $x \in \mathbf{R}^n$ is a real variable. A complex SDP is the problem of minimizing a (real) linear function of x subject to a complex LMI constraint.

Complex LMIs and SDPs can be transformed to real LMIs and SDPs, using the fact that

$$X \succeq 0 \iff \begin{bmatrix} \Re X & -\Im X \\ \Im X & \Re X \end{bmatrix} \succeq 0,$$

where $\Re X \in \mathbf{R}^{n\times n}$ is the real part of the complex Hermitian matrix X, and $\Im X \in \mathbf{R}^{n\times n}$ is the imaginary part of X.

Verify this result, and show how to pose a complex SDP as a real SDP.

4.43 *Eigenvalue optimization via SDP.* Suppose $A : \mathbf{R}^n \to \mathbf{S}^m$ is affine, *i.e.*,

$$A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n$$

where $A_i \in \mathbf{S}^m$. Let $\lambda_1(x) \geq \lambda_2(x) \geq \cdots \geq \lambda_m(x)$ denote the eigenvalues of $A(x)$. Show how to pose the following problems as SDPs.

(a) Minimize the maximum eigenvalue $\lambda_1(x)$.

(b) Minimize the spread of the eigenvalues, $\lambda_1(x) - \lambda_m(x)$.

(c) Minimize the condition number of $A(x)$, subject to $A(x) \succ 0$. The condition number is defined as $\kappa(A(x)) = \lambda_1(x)/\lambda_m(x)$, with domain $\{x \mid A(x) \succ 0\}$. You may assume that $A(x) \succ 0$ for at least one x.
Hint. You need to minimize λ/γ, subject to

$$0 \prec \gamma I \preceq A(x) \preceq \lambda I.$$

Change variables to $y = x/\gamma$, $t = \lambda/\gamma$, $s = 1/\gamma$.

(d) Minimize the sum of the absolute values of the eigenvalues, $|\lambda_1(x)| + \cdots + |\lambda_m(x)|$.
Hint. Express $A(x)$ as $A(x) = A_+ - A_-$, where $A_+ \succeq 0$, $A_- \succeq 0$.

4.44 *Optimization over polynomials.* Pose the following problem as an SDP. Find the polynomial $p : \mathbf{R} \to \mathbf{R}$,

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

that satisfies given bounds $l_i \leq p(t_i) \leq u_i$, at m specified points t_i, and, of all the polynomials that satisfy these bounds, has the greatest minimum value:

$$\begin{array}{ll} \text{maximize} & \inf_t p(t) \\ \text{subject to} & l_i \leq p(t_i) \leq u_i, \quad i = 1, \ldots, m. \end{array}$$

The variables are $x \in \mathbf{R}^{2k+1}$.

Hint. Use the LMI characterization of nonnegative polynomials derived in exercise 2.37, part (b).

4.45 [Nes00, Par00] *Sum-of-squares representation via LMIs.* Consider a polynomial $p : \mathbf{R}^n \to \mathbf{R}$ of degree $2k$. The polynomial is said to be positive semidefinite (PSD) if $p(x) \geq 0$ for all $x \in \mathbf{R}^n$. Except for special cases (*e.g.*, $n = 1$ or $k = 1$), it is extremely difficult to determine whether or not a given polynomial is PSD, let alone solve an optimization problem, with the coefficients of p as variables, with the constraint that p be PSD.

A famous sufficient condition for a polynomial to be PSD is that it have the form

$$p(x) = \sum_{i=1}^{r} q_i(x)^2,$$

for some polynomials q_i, with degree no more than k. A polynomial p that has this sum-of-squares form is called SOS.

The condition that a polynomial p be SOS (viewed as a constraint on its coefficients) turns out to be equivalent to an LMI, and therefore a variety of optimization problems, with SOS constraints, can be posed as SDPs. You will explore these ideas in this problem.

(a) Let $f_1, \ldots, f_s$ be all monomials of degree k or less. (Here we mean monomial in the standard sense, *i.e.*, $x_1^{m_1} \cdots x_n^{m_n}$, where $m_i \in \mathbf{Z}_+$, and not in the sense used in geometric programming.) Show that if p can be expressed as a positive semidefinite quadratic form $p = f^T V f$, with $V \in \mathbf{S}_+^s$, then p is SOS. Conversely, show that if p is SOS, then it can be expressed as a positive semidefinite quadratic form in the monomials, *i.e.*, $p = f^T V f$, for some $V \in \mathbf{S}_+^s$.

(b) Show that the condition $p = f^T V f$ is a set of linear equality constraints relating the coefficients of p and the matrix V. Combined with part (a) above, this shows that the condition that p be SOS is equivalent to a set of linear equalities relating V and the coefficients of p, and the matrix inequality $V \succeq 0$.

(c) Work out the LMI conditions for SOS explicitly for the case where p is polynomial of degree four in two variables.

4.46 *Multidimensional moments.* The moments of a random variable t on $\mathbf{R}^2$ are defined as $\mu_{ij} = \mathbf{E}\, t_1^i t_2^j$, where i, j are nonnegative integers. In this problem we derive necessary conditions for a set of numbers μ_{ij}, $0 \leq i, j \leq 2k$, $i + j \leq 2k$, to be the moments of a distribution on $\mathbf{R}^2$.

Let $p : \mathbf{R}^2 \to \mathbf{R}$ be a polynomial of degree k with coefficients c_{ij},

$$p(t) = \sum_{i=0}^{k} \sum_{j=0}^{k-i} c_{ij} t_1^i t_2^j,$$

and let t be a random variable with moments μ_{ij}. Suppose $c \in \mathbf{R}^{(k+1)(k+2)/2}$ contains the coefficients c_{ij} in some specific order, and $\mu \in \mathbf{R}^{(k+1)(2k+1)}$ contains the moments μ_{ij} in the same order. Show that $\mathbf{E}\, p(t)^2$ can be expressed as a quadratic form in c:

$$\mathbf{E}\, p(t)^2 = c^T H(\mu) c,$$

where $H : \mathbf{R}^{(k+1)(2k+1)} \to \mathbf{S}^{(k+1)(k+2)/2}$ is a linear function of μ. From this, conclude that μ must satisfy the LMI $H(\mu) \succeq 0$.

Remark: For random variables on $\mathbf{R}$, the matrix H can be taken as the Hankel matrix defined in (4.52). In this case, $H(\mu) \succeq 0$ is a necessary and sufficient condition for μ to be the moments of a distribution, or the limit of a sequence of moments. On $\mathbf{R}^2$, however, the LMI is only a necessary condition.

4.47 *Maximum determinant positive semidefinite matrix completion.* We consider a matrix $A \in \mathbf{S}^n$, with some entries specified, and the others not specified. The *positive semidefinite matrix completion problem* is to determine values of the unspecified entries of the matrix so that $A \succeq 0$ (or to determine that such a completion does not exist).

(a) Explain why we can assume without loss of generality that the diagonal entries of A are specified.

(b) Show how to formulate the positive semidefinite completion problem as an SDP feasibility problem.

(c) Assume that A has at least one completion that is positive definite, and the diagonal entries of A are specified (*i.e.*, fixed). The positive definite completion with largest determinant is called the *maximum determinant completion*. Show that the maximum determinant completion is unique. Show that if $A^\star$ is the maximum determinant completion, then $(A^\star)^{-1}$ has zeros in all the entries of the original matrix that were not specified. *Hint.* The gradient of the function $f(X) = \log\det X$ is $\nabla f(X) = X^{-1}$ (see §A.4.1).

(d) Suppose A is specified on its tridiagonal part, *i.e.*, we are given $A_{11}, \ldots, A_{nn}$ and $A_{12}, \ldots, A_{n-1,n}$. Show that if there exists a positive definite completion of A, then there is a positive definite completion whose inverse is tridiagonal.

4.48 *Generalized eigenvalue minimization.* Recall (from example 3.37, or §A.5.3) that the largest generalized eigenvalue of a pair of matrices $(A, B) \in \mathbf{S}^k \times \mathbf{S}^k_{++}$ is given by

$$\lambda_{\max}(A, B) = \sup_{u \neq 0} \frac{u^T A u}{u^T B u} = \max\{\lambda \mid \det(\lambda B - A) = 0\}.$$

As we have seen, this function is quasiconvex (if we take $\mathbf{S}^k \times \mathbf{S}^k_{++}$ as its domain).

We consider the problem

$$\text{minimize} \quad \lambda_{\max}(A(x), B(x)) \tag{4.73}$$

where $A, B : \mathbf{R}^n \to \mathbf{S}^k$ are affine functions, defined as

$$A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n, \qquad B(x) = B_0 + x_1 B_1 + \cdots + x_n B_n.$$

with $A_i, B_i \in \mathbf{S}^k$.

(a) Give a family of convex functions $\phi_t : \mathbf{S}^k \times \mathbf{S}^k \to \mathbf{R}$, that satisfy

$$\lambda_{\max}(A, B) \le t \iff \phi_t(A, B) \le 0$$

for all $(A, B) \in \mathbf{S}^k \times \mathbf{S}^k_{++}$. Show that this allows us to solve (4.73) by solving a sequence of convex feasibility problems.

(b) Give a family of matrix-convex functions $\Phi_t : \mathbf{S}^k \times \mathbf{S}^k \to \mathbf{S}^k$ that satisfy

$$\lambda_{\max}(A, B) \le t \iff \Phi_t(A, B) \preceq 0$$

for all $(A, B) \in \mathbf{S}^k \times \mathbf{S}^k_{++}$. Show that this allows us to solve (4.73) by solving a sequence of convex feasibility problems with LMI constraints.

(c) Suppose $B(x) = (a^T x + b)I$, with $a \ne 0$. Show that (4.73) is equivalent to the convex problem

$$\begin{array}{ll} \text{minimize} & \lambda_{\max}(sA_0 + y_1 A_1 + \cdots + y_n A_n) \\ \text{subject to} & a^T y + bs = 1 \\ & s \ge 0, \end{array}$$

with variables $y \in \mathbf{R}^n$, $s \in \mathbf{R}$.

4.49 *Generalized fractional programming.* Let $K \in \mathbf{R}^m$ be a proper cone. Show that the function $f_0 : \mathbf{R}^n \to \mathbf{R}^m$, defined by

$$f_0(x) = \inf\{t \mid Cx + d \preceq_K t(Fx + g)\}, \qquad \mathbf{dom}\, f_0 = \{x \mid Fx + g \succ_K 0\},$$

with $C, F \in \mathbf{R}^{m \times n}$, $d, g \in \mathbf{R}^m$, is quasiconvex.

A quasiconvex optimization problem with objective function of this form is called a *generalized fractional program.* Express the generalized linear-fractional program of page 152 and the generalized eigenvalue minimization problem (4.73) as generalized fractional programs.

Vector and multicriterion optimization

4.50 *Bi-criterion optimization.* Figure 4.11 shows the optimal trade-off curve and the set of achievable values for the bi-criterion optimization problem

$$\text{minimize (w.r.t. } \mathbf{R}^2_+) \quad (\|Ax - b\|^2, \|x\|_2^2),$$

for some $A \in \mathbf{R}^{100 \times 10}$, $b \in \mathbf{R}^{100}$. Answer the following questions using information from the plot. We denote by $x_{\rm ls}$ the solution of the least-squares problem

$$\text{minimize} \quad \|Ax - b\|_2^2.$$

(a) What is $\|x_{\rm ls}\|_2$?

(b) What is $\|Ax_{\rm ls} - b\|_2$?

(c) What is $\|b\|_2$?

(d) Give the optimal value of the problem

$$\begin{array}{ll} \text{minimize} & \|Ax-b\|_2^2 \\ \text{subject to} & \|x\|_2^2 = 1. \end{array}$$

(e) Give the optimal value of the problem

$$\begin{array}{ll} \text{minimize} & \|Ax-b\|_2^2 \\ \text{subject to} & \|x\|_2^2 \leq 1. \end{array}$$

(f) Give the optimal value of the problem

$$\text{minimize } \|Ax-b\|_2^2 + \|x\|_2^2.$$

(g) What is the rank of A?

4.51 *Monotone transformation of objective in vector optimization.* Consider the vector optimization problem (4.56). Suppose we form a new vector optimization problem by replacing the objective f_0 with $\phi \circ f_0$, where $\phi : \mathbf{R}^q \to \mathbf{R}^q$ satisfies

$$u \preceq_K v, \; u \neq v \Longrightarrow \phi(u) \preceq_K \phi(v), \; \phi(u) \neq \phi(v).$$

Show that a point x is Pareto optimal (or optimal) for one problem if and only if it is Pareto optimal (optimal) for the other, so the two problems are equivalent. In particular, composing each objective in a multicriterion problem with an increasing function does not affect the Pareto optimal points.

4.52 *Pareto optimal points and the boundary of the set of achievable values.* Consider a vector optimization problem with cone K. Let $\mathcal{P}$ denote the set of Pareto optimal values, and let $\mathcal{O}$ denote the set of achievable objective values. Show that $\mathcal{P} \subseteq \mathcal{O} \cap \mathbf{bd}\, \mathcal{O}$, *i.e.*, every Pareto optimal value is an achievable objective value that lies in the boundary of the set of achievable objective values.

4.53 Suppose the vector optimization problem (4.56) is convex. Show that the set

$$\mathcal{A} = \mathcal{O} + K = \{t \in \mathbf{R}^q \mid f_0(x) \preceq_K t \text{ for some feasible } x\},$$

is convex. Also show that the minimal elements of $\mathcal{A}$ are the same as the minimal points of $\mathcal{O}$.

4.54 *Scalarization and optimal points.* Suppose a (not necessarily convex) vector optimization problem has an optimal point $x^\star$. Show that $x^\star$ is a solution of the associated scalarized problem for any choice of $\lambda \succ_{K^*} 0$. Also show the converse: If a point x is a solution of the scalarized problem for any choice of $\lambda \succ_{K^*} 0$, then it is an optimal point for the (not necessarily convex) vector optimization problem.

4.55 *Generalization of weighted-sum scalarization.* In §4.7.4 we showed how to obtain Pareto optimal solutions of a vector optimization problem by replacing the vector objective $f_0 : \mathbf{R}^n \to \mathbf{R}^q$ with the scalar objective $\lambda^T f_0$, where $\lambda \succ_{K^*} 0$. Let $\psi : \mathbf{R}^q \to \mathbf{R}$ be a K-increasing function, *i.e.*, satisfying

$$u \preceq_K v, \; u \neq v \Longrightarrow \psi(u) < \psi(v).$$

Show that any solution of the problem

$$\begin{array}{lll} \text{minimize} & \psi(f_0(x)) & \\ \text{subject to} & f_i(x) \leq 0, & i = 1, \ldots, m \\ & h_i(x) = 0, & i = 1, \ldots, p \end{array}$$

is Pareto optimal for the vector optimization problem

$$\begin{array}{ll} \text{minimize (w.r.t. } K) & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p. \end{array}$$

Note that $\psi(u) = \lambda^T u$, where $\lambda \succ_{K^*} 0$, is a special case.

As a related example, show that in a multicriterion optimization problem (*i.e.*, a vector optimization problem with $f_0 = F : \mathbf{R}^n \to \mathbf{R}^q$, and $K = \mathbf{R}^q_+$), a *unique* solution of the scalar optimization problem

$$\begin{array}{ll} \text{minimize} & \max_{i=1,\ldots,q} F_i(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p, \end{array}$$

is Pareto optimal.

Miscellaneous problems

4.56 [P. Parrilo] We consider the problem of minimizing the convex function $f_0 : \mathbf{R}^n \to \mathbf{R}$ over the convex hull of the union of some convex sets, $\mathbf{conv}\left(\bigcup_{i=1}^q C_i\right)$. These sets are described via convex inequalities,

$$C_i = \{x \mid f_{ij}(x) \le 0, \ j = 1, \ldots, k_i\},$$

where $f_{ij} : \mathbf{R}^n \to \mathbf{R}$ are convex. Our goal is to formulate this problem as a convex optimization problem.

The obvious approach is to introduce variables $x_1, \ldots, x_q \in \mathbf{R}^n$, with $x_i \in C_i$, $\theta \in \mathbf{R}^q$ with $\theta \succeq 0$, $\mathbf{1}^T\theta = 1$, and a variable $x \in \mathbf{R}^n$, with $x = \theta_1 x_1 + \cdots + \theta_q x_q$. This equality constraint is not affine in the variables, so this approach does not yield a convex problem.

A more sophisticated formulation is given by

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & s_i f_{ij}(z_i/s_i) \le 0, \quad i = 1, \ldots, q, \quad j = 1, \ldots, k_i \\ & \mathbf{1}^T s = 1, \quad s \succeq 0 \\ & x = z_1 + \cdots + z_q, \end{array}$$

with variables $z_1, \ldots, z_q \in \mathbf{R}^n$, $x \in \mathbf{R}^n$, and $s_1, \ldots, s_q \in \mathbf{R}$. (When $s_i = 0$, we take $s_i f_{ij}(z_i/s_i)$ to be 0 if $z_i = 0$ and ∞ if $z_i \ne 0$.) Explain why this problem is convex, and equivalent to the original problem.

4.57 *Capacity of a communication channel.* We consider a communication channel, with input $X(t) \in \{1, \ldots, n\}$, and output $Y(t) \in \{1, \ldots, m\}$, for $t = 1, 2, \ldots$ (in seconds, say). The relation between the input and the output is given statistically:

$$p_{ij} = \mathbf{prob}(Y(t) = i | X(t) = j), \quad i = 1, \ldots, m, \quad j = 1, \ldots, n.$$

The matrix $P \in \mathbf{R}^{m \times n}$ is called the *channel transition matrix*, and the channel is called a *discrete memoryless channel*.

A famous result of Shannon states that information can be sent over the communication channel, with arbitrarily small probability of error, at any rate less than a number C, called the *channel capacity*, in bits per second. Shannon also showed that the capacity of a discrete memoryless channel can be found by solving an optimization problem. Assume that X has a probability distribution denoted $x \in \mathbf{R}^n$, *i.e.*,

$$x_j = \mathbf{prob}(X = j), \quad j = 1, \ldots, n.$$

The *mutual information* between X and Y is given by

$$I(X;Y) = \sum_{i=1}^{m} \sum_{j=1}^{n} x_j p_{ij} \log_2 \frac{p_{ij}}{\sum_{k=1}^{n} x_k p_{ik}}.$$

Then the channel capacity C is given by

$$C = \sup_x I(X;Y),$$

where the supremum is over all possible probability distributions for the input X, *i.e.*, over $x \succeq 0$, $\mathbf{1}^T x = 1$.

Show how the channel capacity can be computed using convex optimization.

Hint. Introduce the variable $y = Px$, which gives the probability distribution of the output Y, and show that the mutual information can be expressed as

$$I(X;Y) = c^T x - \sum_{i=1}^{m} y_i \log_2 y_i,$$

where $c_j = \sum_{i=1}^{m} p_{ij} \log_2 p_{ij}$, $j = 1, \ldots, n$.

4.58 *Optimal consumption.* In this problem we consider the optimal way to consume (or spend) an initial amount of money (or other asset) k_0 over time. The variables are $c_0, \ldots, c_T$, where $c_t \geq 0$ denotes the *consumption* in period t. The utility derived from a consumption level c is given by $u(c)$, where $u : \mathbf{R} \to \mathbf{R}$ is an increasing concave function. The present value of the utility derived from the consumption is given by

$$U = \sum_{t=0}^{T} \beta^t u(c_t),$$

where $0 < \beta < 1$ is a *discount factor*.

Let k_t denote the amount of money available for investment in period t. We assume that it earns an investment return given by $f(k_t)$, where $f : \mathbf{R} \to \mathbf{R}$ is an increasing, concave *investment return function*, which satisfies $f(0) = 0$. For example if the funds earn simple interest at rate R percent per period, we have $f(a) = (R/100)a$. The amount to be consumed, *i.e.*, c_t, is withdrawn at the end of the period, so we have the recursion

$$k_{t+1} = k_t + f(k_t) - c_t, \quad t = 0, \ldots, T.$$

The initial sum $k_0 > 0$ is given. We require $k_t \geq 0$, $t = 1, \ldots, T+1$ (but more sophisticated models, which allow $k_t < 0$, can be considered).

Show how to formulate the problem of maximizing U as a convex optimization problem. Explain how the problem you formulate is equivalent to this one, and exactly how the two are related.

Hint. Show that we can replace the recursion for k_t given above with the inequalities

$$k_{t+1} \leq k_t + f(k_t) - c_t, \quad t = 0, \ldots, T.$$

(Interpretation: the inequalities give you the option of throwing money away in each period.) For a more general version of this trick, see exercise 4.6.

4.59 *Robust optimization.* In some optimization problems there is uncertainty or variation in the objective and constraint functions, due to parameters or factors that are either beyond our control or unknown. We can model this situation by making the objective and constraint functions $f_0, \ldots, f_m$ functions of the optimization variable $x \in \mathbf{R}^n$ *and* a parameter vector $u \in \mathbf{R}^k$ that is unknown, or varies. In the *stochastic optimization* approach, the parameter vector u is modeled as a random variable with a known distribution, and we work with the expected values $\mathbf{E}_u f_i(x, u)$. In the *worst-case analysis* approach, we are given a set U that u is known to lie in, and we work with the maximum or worst-case values $\sup_{u \in U} f_i(x, u)$. To simplify the discussion, we assume there are no equality constraints.

(a) *Stochastic optimization.* We consider the problem

$$\begin{array}{ll} \text{minimize} & \mathbf{E}\, f_0(x, u) \\ \text{subject to} & \mathbf{E}\, f_i(x, u) \leq 0, \quad i = 1, \ldots, m, \end{array}$$

where the expectation is with respect to u. Show that if f_i are convex in x for each u, then this stochastic optimization problem is convex.

(b) *Worst-case optimization.* We consider the problem

$$\begin{array}{ll} \text{minimize} & \sup_{u \in U} f_0(x, u) \\ \text{subject to} & \sup_{u \in U} f_i(x, u) \leq 0, \quad i = 1, \ldots, m. \end{array}$$

Show that if f_i are convex in x for each u, then this worst-case optimization problem is convex.

(c) *Finite set of possible parameter values.* The observations made in parts (a) and (b) are most useful when we have analytical or easily evaluated expressions for the expected values $\mathbf{E}\, f_i(x, u)$ or the worst-case values $\sup_{u \in U} f_i(x, u)$.

Suppose we are given the set of possible values of the parameter is finite, *i.e.*, we have $u \in \{u_1, \ldots, u_N\}$. For the stochastic case, we are also given the probabilities of each value: $\mathbf{prob}(u = u_i) = p_i$, where $p \in \mathbf{R}^N$, $p \succeq 0$, $\mathbf{1}^T p = 1$. In the worst-case formulation, we simply take $U \in \{u_1, \ldots, u_N\}$.

Show how to set up the worst-case and stochastic optimization problems explicitly (*i.e.*, give explicit expressions for $\sup_{u \in U} f_i$ and $\mathbf{E}_u f_i$).

4.60 *Log-optimal investment strategy.* We consider a portfolio problem with n assets held over N periods. At the beginning of each period, we re-invest our total wealth, redistributing it over the n assets using a fixed, constant, allocation strategy $x \in \mathbf{R}^n$, where $x \succeq 0$, $\mathbf{1}^T x = 1$. In other words, if $W(t-1)$ is our wealth at the beginning of period t, then during period t we invest $x_i W(t-1)$ in asset i. We denote by $\lambda(t)$ the total return during period t, *i.e.*, $\lambda(t) = W(t)/W(t-1)$. At the end of the N periods our wealth has been multiplied by the factor $\prod_{t=1}^N \lambda(t)$. We call

$$\frac{1}{N} \sum_{t=1}^{N} \log \lambda(t)$$

the *growth rate* of the investment over the N periods. We are interested in determining an allocation strategy x that maximizes growth of our total wealth for large N.

We use a discrete stochastic model to account for the uncertainty in the returns. We assume that during each period there are m possible scenarios, with probabilities π_j, $j = 1, \ldots, m$. In scenario j, the return for asset i over one period is given by p_{ij}. Therefore, the return $\lambda(t)$ of our portfolio during period t is a random variable, with m possible values $p_1^T x, \ldots, p_m^T x$, and distribution

$$\pi_j = \mathbf{prob}(\lambda(t) = p_j^T x), \quad j = 1, \ldots, m.$$

We assume the same scenarios for each period, with (identical) independent distributions. Using the law of large numbers, we have

$$\lim_{N \to \infty} \frac{1}{N} \log \left(\frac{W(N)}{W(0)} \right) = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \log \lambda(t) = \mathbf{E} \log \lambda(t) = \sum_{j=1}^{m} \pi_j \log(p_j^T x).$$

In other words, with investment strategy x, the long term growth rate is given by

$$R_{\mathrm{lt}} = \sum_{j=1}^{m} \pi_j \log(p_j^T x).$$

The investment strategy x that maximizes this quantity is called the *log-optimal investment strategy*, and can be found by solving the optimization problem

$$\begin{array}{ll} \text{maximize} & \sum_{j=1}^{m} \pi_j \log(p_j^T x) \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1, \end{array}$$

with variable $x \in \mathbf{R}^n$.

Show that this is a convex optimization problem.

4.61 *Optimization with logistic model.* A random variable $X \in \{0, 1\}$ satisfies

$$\mathbf{prob}(X = 1) = p = \frac{\exp(a^T x + b)}{1 + \exp(a^T x + b)},$$

where $x \in \mathbf{R}^n$ is a vector of variables that affect the probability, and a and b are known parameters. We can think of $X = 1$ as the event that a consumer buys a product, and x as a vector of variables that affect the probability, *e.g.*, advertising effort, retail price, discounted price, packaging expense, and other factors. The variable x, which we are to optimize over, is subject to a set of linear constraints, $Fx \preceq g$.

Formulate the following problems as convex optimization problems.

(a) *Maximizing buying probability.* The goal is to choose x to maximize p.

(b) *Maximizing expected profit.* Let $c^T x + d$ be the profit derived from selling the product, which we assume is positive for all feasible x. The goal is to maximize the expected profit, which is $p(c^T x + d)$.

4.62 *Optimal power and bandwidth allocation in a Gaussian broadcast channel.* We consider a communication system in which a central node transmits messages to n receivers. ('Gaussian' refers to the type of noise that corrupts the transmissions.) Each receiver channel is characterized by its (transmit) power level $P_i \geq 0$ and its bandwidth $W_i \geq 0$. The power and bandwidth of a receiver channel determine its *bit rate* R_i (the rate at which information can be sent) via

$$R_i = \alpha_i W_i \log(1 + \beta_i P_i / W_i),$$

where α_i and β_i are known positive constants. For $W_i = 0$, we take $R_i = 0$ (which is what you get if you take the limit as $W_i \to 0$).

The powers must satisfy a total power constraint, which has the form

$$P_1 + \cdots + P_n = P_{\mathrm{tot}},$$

where $P_{\mathrm{tot}} > 0$ is a given total power available to allocate among the channels. Similarly, the bandwidths must satisfy

$$W_1 + \cdots + W_n = W_{\mathrm{tot}},$$

where $W_{\mathrm{tot}} > 0$ is the (given) total available bandwidth. The optimization variables in this problem are the powers and bandwidths, *i.e.*, $P_1, \ldots, P_n$, $W_1, \ldots, W_n$.

The objective is to maximize the total utility,

$$\sum_{i=1}^{n} u_i(R_i),$$

where $u_i : \mathbf{R} \to \mathbf{R}$ is the utility function associated with the ith receiver. (You can think of $u_i(R_i)$ as the revenue obtained for providing a bit rate R_i to receiver i, so the objective is to maximize the total revenue.) You can assume that the utility functions u_i are nondecreasing and concave.

Pose this problem as a convex optimization problem.

4.63 *Optimally balancing manufacturing cost and yield.* The vector $x \in \mathbf{R}^n$ denotes the nominal parameters in a manufacturing process. The yield of the process, *i.e.*, the fraction of manufactured goods that is acceptable, is given by $Y(x)$. We assume that Y is log-concave (which is often the case; see example 3.43). The cost per unit to manufacture the product is given by $c^T x$, where $c \in \mathbf{R}^n$. The cost per acceptable unit is $c^T x / Y(x)$. We want to minimize $c^T x / Y(x)$, subject to some convex constraints on x such as a linear inequalities $Ax \preceq b$. (You can assume that over the feasible set we have $c^T x > 0$ and $Y(x) > 0$.)

This problem is *not* a convex or quasiconvex optimization problem, but it can be solved using convex optimization and a one-dimensional search. The basic ideas are given below; you must supply all details and justification.

(a) Show that the function $f : \mathbf{R} \to \mathbf{R}$ given by

$$f(a) = \sup\{Y(x) \mid Ax \preceq b,\ c^T x = a\},$$

which gives the maximum yield versus cost, is log-concave. This means that by solving a convex optimization problem (in x) we can evaluate the function f.

(b) Suppose that we evaluate the function f for enough values of a to give a good approximation over the range of interest. Explain how to use these data to (approximately) solve the problem of minimizing cost per good product.

4.64 *Optimization with recourse.* In an optimization problem with recourse, also called *two-stage optimization*, the cost function and constraints depend not only on our choice of variables, but also on a discrete random variable $s \in \{1, \ldots, S\}$, which is interpreted as specifying which of S *scenarios* occurred. The scenario random variable s has known probability distribution π, with $\pi_i = \mathbf{prob}(s = i)$, $i = 1, \ldots, S$.

In two-stage optimization, we are to choose the values of two variables, $x \in \mathbf{R}^n$ and $z \in \mathbf{R}^q$. The variable x must be chosen *before* the particular scenario s is known; the variable z, however, is chosen *after* the value of the scenario random variable is known. In other words, z is a function of the scenario random variable s. To describe our choice z, we list the values we would choose under the different scenarios, *i.e.*, we list the vectors

$$z_1, \ldots, z_S \in \mathbf{R}^q.$$

Here z_3 is our choice of z when $s = 3$ occurs, and so on. The set of values

$$x \in \mathbf{R}^n, \qquad z_1, \ldots, z_S \in \mathbf{R}^q$$

is called the *policy*, since it tells us what choice to make for x (independent of which scenario occurs), and also, what choice to make for z in each possible scenario.

The variable z is called the *recourse variable* (or *second-stage variable*), since it allows us to take some action or make a choice after we know which scenario occurred. In contrast, our choice of x (which is called the *first-stage variable*) must be made without any knowledge of the scenario.

For simplicity we will consider the case with no constraints. The cost function is given by

$$f : \mathbf{R}^n \times \mathbf{R}^q \times \{1, \ldots, S\} \to \mathbf{R},$$

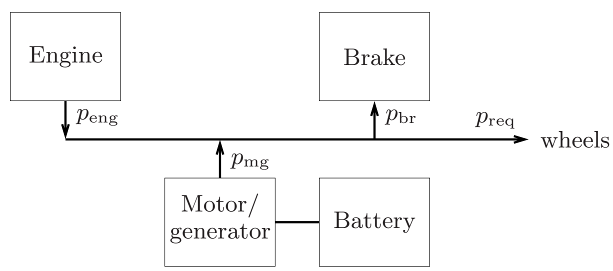
where $f(x, z, i)$ gives the cost when the first-stage choice x is made, second-stage choice z is made, and scenario i occurs. We will take as the overall objective, to be minimized over all policies, the expected cost

$$\mathbf{E}\, f(x, z_s, s) = \sum_{i=1}^{S} \pi_i f(x, z_i, i).$$

Suppose that f is a convex function of (x, z), for each scenario $i = 1, \ldots, S$. Explain how to find an optimal policy, *i.e.*, one that minimizes the expected cost over all possible policies, using convex optimization.

4.65 *Optimal operation of a hybrid vehicle.* A hybrid vehicle has an internal combustion engine, a motor/generator connected to a storage battery, and a conventional (friction) brake. In this exercise we consider a (highly simplified) model of a *parallel hybrid vehicle*, in which both the motor/generator and the engine are directly connected to the drive wheels. The engine can provide power to the wheels, and the brake can take power from the wheels, turning it into heat. The motor/generator can act as a motor, when it uses energy stored in the battery to deliver power to the wheels, or as a generator, when it takes power from the wheels or engine, and uses the power to charge the battery. When the generator takes power from the wheels and charges the battery, it is called *regenerative braking*; unlike ordinary friction braking, the energy taken from the wheels is *stored*, and can be used later. The vehicle is judged by driving it over a known, fixed test track to evaluate its fuel efficiency.

A diagram illustrating the power flow in the hybrid vehicle is shown below. The arrows indicate the direction in which the power flow is considered positive. The engine power p_{eng}, for example, is positive when it is delivering power; the brake power p_{br} is positive when it is taking power from the wheels. The power p_{req} is the required power at the wheels. It is positive when the wheels require power (*e.g.*, when the vehicle accelerates, climbs a hill, or cruises on level terrain). The required wheel power is negative when the vehicle must decelerate rapidly, or descend a hill.



All of these powers are functions of time, which we discretize in one second intervals, with $t = 1, 2, \ldots, T$. The required wheel power $p_{\mathrm{req}}(1), \ldots, p_{\mathrm{req}}(T)$ is given. (The speed of the vehicle on the track is specified, so together with known road slope information, and known aerodynamic and other losses, the power required at the wheels can be calculated.)

Power is conserved, which means we have

$$p_{\mathrm{req}}(t) = p_{\mathrm{eng}}(t) + p_{\mathrm{mg}}(t) - p_{\mathrm{br}}(t), \quad t = 1, \ldots, T.$$

The brake can only dissipate power, so we have $p_{\mathrm{br}}(t) \geq 0$ for each t. The engine can only provide power, and only up to a given limit $P_{\mathrm{eng}}^{\max}$, *i.e.*, we have

$$0 \leq p_{\mathrm{eng}}(t) \leq P_{\mathrm{eng}}^{\max}, \quad t = 1, \ldots, T.$$

The motor/generator power is also limited: p_{mg} must satisfy

$$P_{\mathrm{mg}}^{\min} \leq p_{\mathrm{mg}}(t) \leq P_{\mathrm{mg}}^{\max}, \quad t = 1, \ldots, T.$$

Here $P_{\mathrm{mg}}^{\max} > 0$ is the maximum motor power, and $-P_{\mathrm{mg}}^{\min} > 0$ is the maximum generator power.

The battery charge or energy at time t is denoted $E(t)$, $t = 1, \ldots, T + 1$. The battery energy satisfies

$$E(t + 1) = E(t) - p_{\mathrm{mg}}(t) - \eta |p_{\mathrm{mg}}(t)|, \quad t = 1, \ldots, T,$$

where $\eta > 0$ is a known parameter. (The term $-p_{\rm mg}(t)$ represents the energy removed or added the battery by the motor/generator, ignoring any losses. The term $-\eta|p_{\rm mg}(t)|$ represents energy lost through inefficiencies in the battery or motor/generator.)

The battery charge must be between 0 (empty) and its limit $E_{\rm batt}^{\rm max}$ (full), at all times. (If $E(t) = 0$, the battery is fully discharged, and no more energy can be extracted from it; when $E(t) = E_{\rm batt}^{\rm max}$, the battery is full and cannot be charged.) To make the comparison with non-hybrid vehicles fair, we fix the initial battery charge to equal the final battery charge, so the net energy change is zero over the track: $E(1) = E(T+1)$. We do not specify the value of the initial (and final) energy.

The objective in the problem (to be minimized) is the total fuel consumed by the engine, which is

$$F_{\rm total} = \sum_{t=1}^{T} F(p_{\rm eng}(t)),$$

where $F : \mathbf{R} \to \mathbf{R}$ is the *fuel use characteristic* of the engine. We assume that F is positive, increasing, and convex.

Formulate this problem as a convex optimization problem, with variables $p_{\rm eng}(t)$, $p_{\rm mg}(t)$, and $p_{\rm br}(t)$ for $t = 1, \ldots, T$, and $E(t)$ for $t = 1, \ldots, T+1$. Explain why your formulation is equivalent to the problem described above.

Chapter 5

Duality

5.1 The Lagrange dual function

5.1.1 The Lagrangian

We consider an optimization problem in the standard form (4.1):

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p, \end{array} \tag{5.1}$$

with variable $x \in \mathbf{R}^n$. We assume its domain $\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \;\cap\; \bigcap_{i=1}^{p} \mathbf{dom}\, h_i$ is nonempty, and denote the optimal value of (5.1) by $p^\star$. We do not assume the problem (5.1) is convex.

The basic idea in Lagrangian duality is to take the constraints in (5.1) into account by augmenting the objective function with a weighted sum of the constraint functions. We define the *Lagrangian* $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ associated with the problem (5.1) as

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x),$$

with $\mathbf{dom}\, L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$. We refer to λ_i as the *Lagrange multiplier* associated with the ith inequality constraint $f_i(x) \leq 0$; similarly we refer to ν_i as the Lagrange multiplier associated with the ith equality constraint $h_i(x) = 0$. The vectors λ and ν are called the *dual variables* or *Lagrange multiplier vectors* associated with the problem (5.1).

5.1.2 The Lagrange dual function

We define the *Lagrange dual function* (or just *dual function*) $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ as the minimum value of the Lagrangian over x: for $\lambda \in \mathbf{R}^m$, $\nu \in \mathbf{R}^p$,

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right).$$

When the Lagrangian is unbounded below in x, the dual function takes on the value $-\infty$. Since the dual function is the pointwise infimum of a family of affine functions of (λ, ν), it is concave, even when the problem (5.1) is not convex.

5.1.3 Lower bounds on optimal value

The dual function yields lower bounds on the optimal value $p^\star$ of the problem (5.1): For any $\lambda \succeq 0$ and any ν we have

$$g(\lambda, \nu) \le p^\star. \tag{5.2}$$

This important property is easily verified. Suppose $\tilde{x}$ is a feasible point for the problem (5.1), *i.e.*, $f_i(\tilde{x}) \le 0$ and $h_i(\tilde{x}) = 0$, and $\lambda \succeq 0$. Then we have

$$\sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \le 0,$$

since each term in the first sum is nonpositive, and each term in the second sum is zero, and therefore

$$L(\tilde{x}, \lambda, \nu) = f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \le f_0(\tilde{x}).$$

Hence

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \le L(\tilde{x}, \lambda, \nu) \le f_0(\tilde{x}).$$

Since $g(\lambda, \nu) \le f_0(\tilde{x})$ holds for every feasible point $\tilde{x}$, the inequality (5.2) follows. The lower bound (5.2) is illustrated in figure 5.1, for a simple problem with $x \in \mathbf{R}$ and one inequality constraint.

The inequality (5.2) holds, but is vacuous, when $g(\lambda, \nu) = -\infty$. The dual function gives a nontrivial lower bound on $p^\star$ only when $\lambda \succeq 0$ and $(\lambda, \nu) \in \mathbf{dom}\, g$, *i.e.*, $g(\lambda, \nu) > -\infty$. We refer to a pair (λ, ν) with $\lambda \succeq 0$ and $(\lambda, \nu) \in \mathbf{dom}\, g$ as *dual feasible*, for reasons that will become clear later.

5.1.4 Linear approximation interpretation

The Lagrangian and lower bound property can be given a simple interpretation, based on a linear approximation of the indicator functions of the sets $\{0\}$ and $-\mathbf{R}_+$.