1 Introduction and Preliminaries

1.1 Constrained Optimization

We consider constrained optimization problems of the form

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = b \\ & x \in X. \end{array}$$

Such a problem is given by a vector $x \in \mathbb{R}^n$ of *decision variables*, an *objective function* $f : \mathbb{R}^n \to \mathbb{R}$, a *functional constraint* $h(x) = b$ where $h : \mathbb{R}^n \to \mathbb{R}^m$ and $b \in \mathbb{R}^m$, and a *regional constraint* $x \in X$ where $X \subseteq \mathbb{R}^n$. The set $X(b) = \{x \in X : h(x) = b\}$ is called the *feasible set*, and a problem is called *feasible* if $X(b)$ is non-empty. A vector x^* is called optimal if it is in the feasible set and minimizes f among all vectors in the feasible set. The assumption that the functional constraint holds with equality is without loss of generality: an inequality constraint like $g(x) \leqslant b$ can be re-written as $g(x) + z = b$, where z is a new *slack variable* with the additional regional constraint $z \geqslant 0$. Since minimization of $f(x)$ and maximization of $-f(x)$ are equivalent, we will often concentrate on one of the two.

1.2 Linear Programs

The special case where the objective function and constraints are linear is called a linear program (LP). In matrix-vector notation we can write an LP as

$$\begin{array}{lll} \text{minimize} & c^T x & \\ \text{subject to} & a_i^T x \geqslant b_i, & i \in M_1 \\ & a_i^T x \leqslant b_i, & i \in M_2 \\ & a_i^T x = b_i, & i \in M_3 \\ & x_j \geqslant 0, & j \in N_1 \\ & x_j \leqslant 0, & j \in N_2 \end{array}$$

where $c \in \mathbb{R}^n$ is a cost vector, $x \in \mathbb{R}^n$ is a vector of decision variables, and constraints are given by $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$ for $i \in \{1, \dots, m\}$. Index sets $M_1, M_2, M_3 \subseteq \{1, \dots, m\}$ and $N_1, N_2 \subseteq \{1, \dots, n\}$ are used to distinguish between different types of contraints.

An equality constraint $a_i^T x = b_i$ is equivalent to the pair of constraints $a_i^T \leqslant b_i$ and $a_i^T x \geqslant b_i$, and a constraint of the form $a_i^T x \leqslant b_i$ can be rewritten as $(-a_i)^T x \geqslant -b_i$. Each occurrence of an unconstrained variable x_j can be replaced by $x_j^+ + x_j^-$, where x_j^+

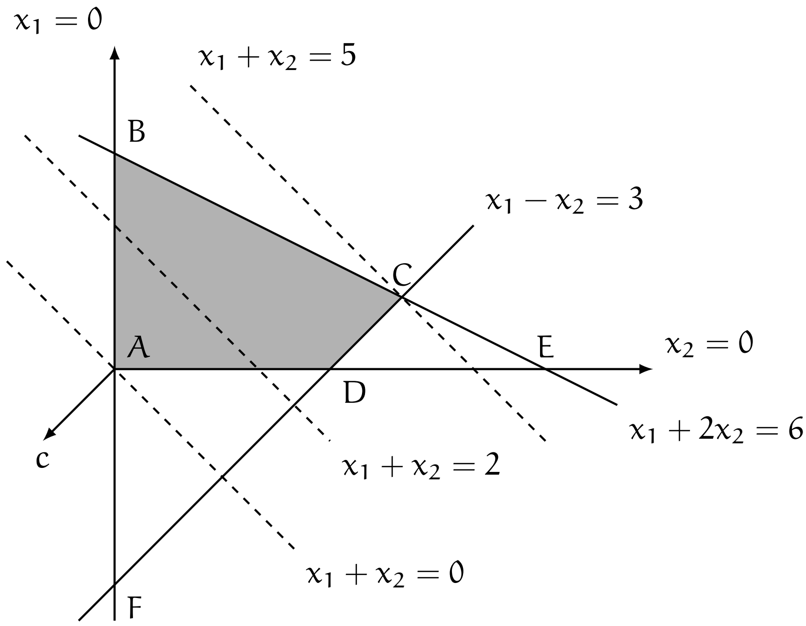

Figure 1.1: Geometric interpretation of the linear program in Example 1.1

and x_j^- are two new variables with $x_j^+ \geqslant 0$ and $x_j^- \leqslant 0$. We can thus write every linear program in the *general form*

$$\min\{\, c^T x : Ax \geqslant b, x \geqslant 0 \,\} \tag{1.1}$$

where $x, c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, and $A \in \mathbb{R}^{m \times n}$. Observe that constraints of the form $x_j \geqslant 0$ and $x_j \geqslant 0$ are just special cases of constraints of the form $a_i^T x \geqslant b_i$, but we often choose to make them explicit.

A linear program of the form

$$\min\{\, c^T x : Ax = b, x \geqslant 0 \,\} \tag{1.2}$$

is said to be in *standard form*. The standard form is of course a special case of the general form. On the other hand, we can also bring every general form problem into the standard form by replacing each inequality constraint of the form $a_i^T x \leqslant b_i$ or $a_i^T x \geqslant b_i$ by a constraint $a_i^T x + s_i = b_i$ or $a_i^T x - s_i = b_i$, where s_i is a new slack variable, and an additional constraint $s_i \geqslant 0$.

The general form is typically used to discuss the theory of linear programming, while the standard form is often more convenient when designing algorithms.

Example 1.1. Consider the linear following program, which is illustrated in Figure 1.1:

$$\begin{array}{ll} \text{minimize} & -(x_1 + x_2) \\ \text{subject to} & x_1 + 2x_2 \leqslant 6 \\ & x_1 - x_2 \quad \leqslant 3 \\ & x_1, x_2 \qquad \geqslant 0 \end{array}$$

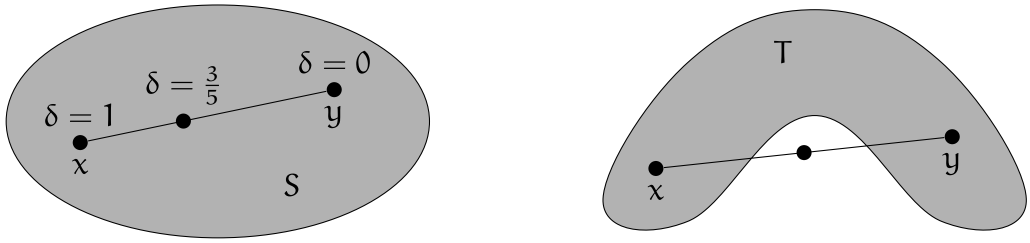

Figure 1.2: A convex set S and a non-convex set T

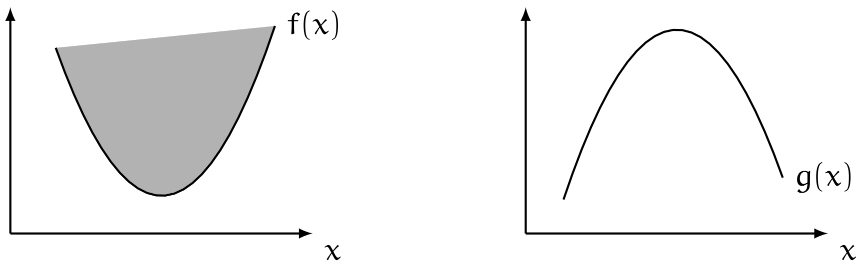

Figure 1.3: A convex function f and a concave function g

Solid lines indicate sets of points for which one of the constraints is satisfied with equality. The feasible set is shaded. Dashed lines, orthogonal to the cost vector c, indicate sets of points for which the value of the objective function is constant.

1.3 Review: Unconstrained Optimization and Convexity

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$, and let $x \in \mathbb{R}^n$. A necessary condition for x to minimize f over $\mathbb{R}^n$ is that $\nabla f(x) = 0$, where

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right)^T$$

is the *gradient* of f. A general function f may have many local minima on the feasible set X, which makes it difficult to find a global minimum. However, if X is convex, and f is convex on X, then any local minimum of f on X is also a global minimum on X.

Let $S \subseteq \mathbb{R}^n$. S is called a *convex set* if for all $\delta \in [0, 1]$, $x, y \in S$ implies that $\delta x + (1 - \delta)y \in S$. An illustration in show in Figure 1.2. A function $f : S \to \mathbb{R}$ is called *convex function* if the set of points above its graph is convex, i.e., if for all $x, y \in S$ and $\delta \in [0, 1]$, $\delta f(x) + (1 - \delta)f(y) \geqslant f(\delta x + (1 - \delta)y)$. Function f is *concave* if $-f$ is convex. An illustration is shown in Figure 1.3.

If f is twice differentiable, it is convex on a convex set S if its *Hessian*

$$\mathcal{H}f = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{ij}$$

is positive semidefinite on S. A symmetric $n \times n$ matrix A is called positive semidefinite if $v^T A v \geqslant 0$ for all $v \in \mathbb{R}^n$, or equivalently, if all eigenvalues of A are non-negative.

THEOREM 1.2. *Let* $X \subseteq \mathbb{R}^n$ *be convex,* $f : \mathbb{R}^n \to \mathbb{R}$ *twice differentiable on* X*. Let* $\nabla f(x^*) = 0$ *for* $x^* \in X$ *and* $\mathcal{H}f(x)$ *positive semidefinite for all* $x \in X$*. Then* x^* *minimizes* $f(x)$ *subject to* $x \in X$*.*

It is easy to see that in the case of LPs, the feasible set is convex and the objective function is both convex and concave. But even when these two conditions are satisfied, the above theorem cannot generally be used to solve constrained optimization problems, because the gradient might not be zero anywhere on the feasible set.