21

Convex Optimization Problems

21.1 INTRODUCTION

The optimization problems posed at the beginning of this part are, in general, very difficult to solve. The source of these difficulties may be in the objective function or the constraints. Even if the objective function is simple and "well behaved," the nature of the constraints may make the problem difficult to solve. We illustrate some of these difficulties in the following examples.

Example 21.1 Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 \\ \text{subject to} & x_2 - x_1 - 2 \leq 0 \\ & x_1^2 - x_2 - 4 \leq 0. \end{array}$$

The problem is depicted in Figure 21.1. As we can see in Figure 21.1, the constrained minimizer is the same as the unconstrained minimizer. At the minimizer, all the constraints are inactive. If we had only known about this fact we could have approached this problem as an unconstrained optimization problem using techniques from Part II. ∎

Example 21.2 Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 \\ \text{subject to} & x_1 - 10 \leq 0 \\ & x_1 - x_2^2 - 4 \geq 0. \end{array}$$

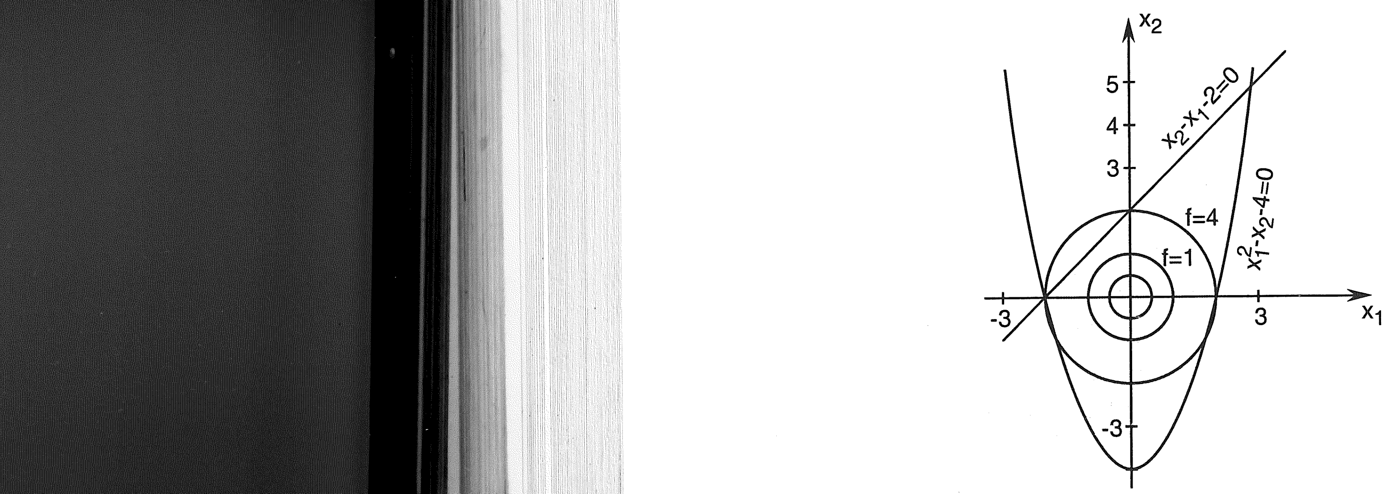

Figure 21.1 Situation where the constrained and the unconstrained minimizers are the same

Figure 21.2 Situation where only one constraint is active

The problem is depicted in Figure 21.2. At the solution, only one constraint is active. If we had only known about this we could have handled this problem as a constrained optimization problem using the Lagrange multiplier method. ∎

Example 21.3 Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 \\ \text{subject to} & 4 - x_1 - x_2^2 \leq 0 \end{array}$$

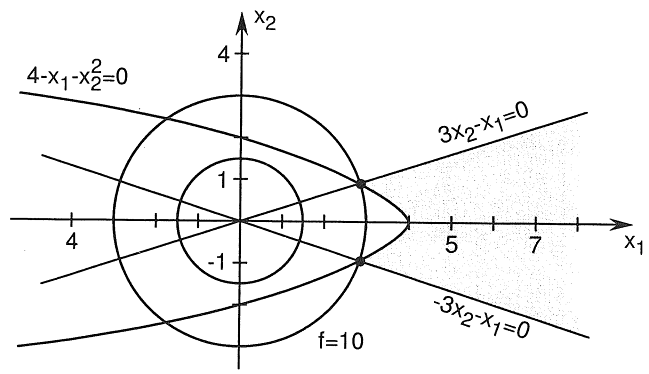

Figure 21.3 Situation where the constraints introduce local minimizers

$$\begin{aligned} 3x_2 - x_1 &\leq 0 \\ -3x_2 - x_1 &\leq 0. \end{aligned}$$

The problem is depicted in Figure 21.3. This example illustrates the situation where the constraints introduce local minimizers, even though the objective function itself has only one unconstrained global minimizer. ∎

Some of the difficulties illustrated in the above examples can be eliminated if we restrict our problems to convex feasible regions. Admittedly, some important real-life problems do not fit into this framework. On the other hand, it is possible to give results of a *global* nature for this class of optimization problems. In the next section, we introduce the notion of a *convex function*, which plays an important role in our subsequent treatment of such problems.

21.2 CONVEX FUNCTIONS

We begin with a definition of the graph of a real-valued function.

Definition 21.1 The *graph* of $f : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^n$, is the set of points in $\Omega \times \mathbb{R} \subset \mathbb{R}^{n+1}$ given by

$$\{[\boldsymbol{x}, f(\boldsymbol{x})]^T : \boldsymbol{x} \in \Omega\}.$$

∎

We can visualize the graph of f as simply the set of points on a "plot" of $f(\boldsymbol{x})$ versus $\boldsymbol{x}$ (see Figure 21.4). We next define the "epigraph" of a real-valued function.

Definition 21.2 The *epigraph* of a function $f : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^n$, denoted $\text{epi}(f)$, is the set of points in $\Omega \times \mathbb{R}$ given by

$$\text{epi}(f) = \{[\boldsymbol{x}, \beta]^T : \boldsymbol{x} \in \Omega, \beta \in \mathbb{R}, \beta \geq f(\boldsymbol{x})\}.$$

∎

Figure 21.4 The graph and epigraph of a function $f : \mathbb{R} \to \mathbb{R}$

The epigraph epi(f) of a function f is simply the set of points in $\Omega \times \mathbb{R}$ on or above the graph of f (see Figure 21.4). We can also think of epi(f) as a subset of $\mathbb{R}^{n+1}$.

Recall that a set $\Omega \subset \mathbb{R}^n$ is convex if for every $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \Omega$ and $\alpha \in (0,1)$, $\alpha\boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2 \in \Omega$ (see Section 4.3). We now introduce the notion of a "convex function."

Definition 21.3 A function $f : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^n$, is *convex* on Ω if its epigraph is a convex set. ∎

Theorem 21.1 *If a function $f : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^n$, is convex on Ω, then Ω is a convex set.* □

Proof. We prove this theorem by contraposition. Suppose that Ω is not a convex set. Then, there exist two points $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ such that for some $\alpha \in (0,1)$,

$$\boldsymbol{z} = \alpha\boldsymbol{y}_1 + (1-\alpha)\boldsymbol{y}_2 \notin \Omega.$$

Let

$$\beta_1 = f(\boldsymbol{y}_1), \quad \beta_2 = f(\boldsymbol{y}_2).$$

Then, the pairs

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \beta_1 \end{bmatrix}, \quad \begin{bmatrix} \boldsymbol{y}_2 \\ \beta_2 \end{bmatrix}$$

belong to the graph of f, and hence also the epigraph of f. Let

$$\boldsymbol{w} = \alpha \begin{bmatrix} \boldsymbol{y}_1 \\ \beta_1 \end{bmatrix} + (1-\alpha) \begin{bmatrix} \boldsymbol{y}_2 \\ \beta_2 \end{bmatrix}.$$

We have

$$\boldsymbol{w} = \begin{bmatrix} \boldsymbol{z} \\ \alpha\beta_1 + (1-\alpha)\beta_2 \end{bmatrix}.$$

But note that $\boldsymbol{w} \notin$ epi(f), because $\boldsymbol{z} \notin \Omega$. Therefore, epi$(f)$ is not convex, and hence f is not a convex function. ∎

The next theorem gives a very useful characterization of convex functions. This characterization is often used as a definition for a convex function.

Theorem 21.2 *A function $f : \Omega \to \mathbb{R}$ defined on a convex set $\Omega \subset \mathbb{R}^n$ is convex if and only if for all $\boldsymbol{x}, \boldsymbol{y} \in \Omega$ and all $\alpha \in (0,1)$, we have*

$$f(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \le \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}).$$

□

Proof. $\Leftarrow$: Assume that for all $\boldsymbol{x}, \boldsymbol{y} \in \Omega$ and $\alpha \in (0,1)$,

$$f(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \le \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}).$$

Let $[\boldsymbol{x}^T, a]^T$ and $[\boldsymbol{y}^T, b]^T$ be two points in epi(f), where $a, b \in \mathbb{R}$. From the definition of epi(f) it follows that

$$f(\boldsymbol{x}) \le a, \quad f(\boldsymbol{y}) \le b.$$

Therefore, using the first inequality above, we have

$$f(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \le \alpha a + (1-\alpha) b.$$

Because Ω is convex, $\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} \in \Omega$. Hence,

$$\begin{bmatrix} \alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} \\ \alpha a + (1-\alpha) b \end{bmatrix} \in \text{epi}(f),$$

which implies that epi(f) is a convex set, and hence f is a convex function.

$\Rightarrow$: Assume that $f : \Omega \to \mathbb{R}$ is a convex function. Let $\boldsymbol{x}, \boldsymbol{y} \in \Omega$ and

$$f(\boldsymbol{x}) = a, \quad f(\boldsymbol{y}) = b.$$

Thus,

$$\begin{bmatrix} \boldsymbol{x} \\ a \end{bmatrix}, \begin{bmatrix} \boldsymbol{y} \\ b \end{bmatrix} \in \text{epi}(f).$$

Because f is a convex function, its epigraph is a convex subset of $\mathbb{R}^{n+1}$. Therefore, for all $\alpha \in (0,1)$, we have

$$\alpha \begin{bmatrix} \boldsymbol{x} \\ a \end{bmatrix} + (1-\alpha) \begin{bmatrix} \boldsymbol{y} \\ b \end{bmatrix} = \begin{bmatrix} \alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} \\ \alpha a + (1-\alpha) b \end{bmatrix} \in \text{epi}(f).$$

The above implies that for all $\alpha \in (0,1)$,

$$f(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \le \alpha a + (1-\alpha) b = \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}).$$

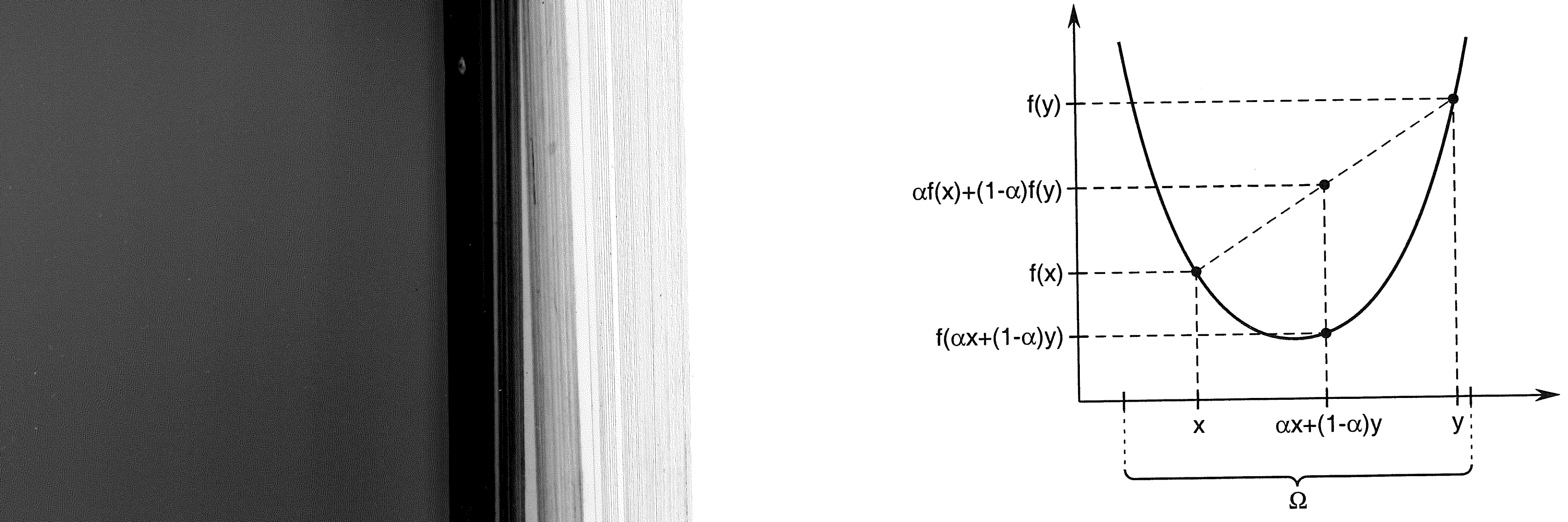

Figure 21.5 Geometric interpretation of Theorem 21.2

Thus, the proof is completed. ∎

In the above theorem, the assumption that Ω be open is not necessary, as long as $f \in \mathcal{C}^1$ on some open set that contains Ω (e.g., $f \in \mathcal{C}^1$ on $\mathbb{R}^n$).

A geometric interpretation of the above theorem is given in Figure 21.5. The theorem states that if $f : \Omega \to \mathbb{R}$ is a convex function over a convex set Ω, then for all $\boldsymbol{x}, \boldsymbol{y} \in \Omega$, the points on the line segment in $\mathbb{R}^{n+1}$ connecting $[\boldsymbol{x}^T, f(\boldsymbol{x})]^T$ and $[\boldsymbol{y}^T, f(\boldsymbol{y})]^T$ must lie on or above the graph of f.

Definition 21.4 A function $f : \Omega \to \mathbb{R}$ on a convex set $\Omega \subset \mathbb{R}^n$ is *strictly convex* if for all $\boldsymbol{x}, \boldsymbol{y} \in \Omega$, $\boldsymbol{x} \neq \boldsymbol{y}$, and $\alpha \in (0, 1)$, we have

$$f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) < \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}).$$

∎

From the above definition, we see that for a strictly convex function, all points on the open line segment connecting the points $[\boldsymbol{x}^T, f(\boldsymbol{x})]^T$ and $[\boldsymbol{y}^T, f(\boldsymbol{y})]^T$ lie (strictly) above the graph of f.

Definition 21.5 A function $f : \Omega \to \mathbb{R}$ on a convex set $\Omega \subset \mathbb{R}^n$ is (strictly) *concave* if $-f$ is (strictly) convex. ∎

Note that the graph of a strictly concave function always lies above the line segment connecting any two points on its graph.

Example 21.4 Let $f(\boldsymbol{x}) = x_1 x_2$. Is f convex over $\Omega = \{\boldsymbol{x} : x_1 \geq 0,\ x_2 \geq 0\}$?

The answer is no. Take, for example, $\boldsymbol{x} = [1, 2]^T \in \Omega$ and $\boldsymbol{y} = [2, 1]^T \in \Omega$. Then,

$$\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y} = \begin{bmatrix} 2-\alpha \\ 1+\alpha \end{bmatrix}.$$

Hence,

$$f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) = (2-\alpha)(1+\alpha) = 2 + \alpha - \alpha^2,$$

and

$$\alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}) = 2.$$

If, for example, $\alpha = 1/2 \in (0, 1)$, then

$$f\left(\frac{1}{2}\boldsymbol{x} + \frac{1}{2}\boldsymbol{y}\right) = \frac{9}{4} > \frac{1}{2} f(\boldsymbol{x}) + \frac{1}{2} f(\boldsymbol{y}),$$

which shows that f is not convex over Ω. ∎

The above numerical example is an illustration of the following general result.

Proposition 21.1 *A quadratic form $f : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^n$, given by $f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}$, $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$, $\boldsymbol{Q} = \boldsymbol{Q}^T$, is convex on Ω if and only if for all $\boldsymbol{x}, \boldsymbol{y} \in \Omega$, $(\boldsymbol{x}-\boldsymbol{y})^T \boldsymbol{Q} (\boldsymbol{x}-\boldsymbol{y}) \geq 0$.* □

Proof. The result follows from Theorem 21.2. Indeed, the function $f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}$ is convex if and only if for every $\alpha \in (0, 1)$, and every $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ we have

$$f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}),$$

or equivalently

$$\alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}) - f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \geq 0.$$

Substituting for f into the left-hand side of the above equation yields

$$\begin{aligned}
&\alpha \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} + (1-\alpha) \boldsymbol{y}^T \boldsymbol{Q} \boldsymbol{y} - (\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y})^T \boldsymbol{Q} (\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \\
&\quad = \alpha \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} + \boldsymbol{y}^T \boldsymbol{Q} \boldsymbol{y} - \alpha \boldsymbol{y}^T \boldsymbol{Q} \boldsymbol{y} - \alpha^2 \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} \\
&\qquad - (2\alpha - 2\alpha^2) \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{y} - (1 - 2\alpha + \alpha^2) \boldsymbol{y}^T \boldsymbol{Q} \boldsymbol{y} \\
&\quad = \alpha(1-\alpha) \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} - 2\alpha(1-\alpha) \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{y} + \alpha(1-\alpha) \boldsymbol{y}^T \boldsymbol{Q} \boldsymbol{y} \\
&\quad = \alpha(1-\alpha)(\boldsymbol{x}-\boldsymbol{y})^T \boldsymbol{Q} (\boldsymbol{x}-\boldsymbol{y}).
\end{aligned}$$

Therefore, f is convex if and only if

$$\alpha(1-\alpha)(\boldsymbol{x}-\boldsymbol{y})^T \boldsymbol{Q} (\boldsymbol{x}-\boldsymbol{y}) \geq 0,$$

which proves the result. ∎

Example 21.5 In the previous example, $f(\boldsymbol{x}) = x_1 x_2$, which can be written as $f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}$, where

$$\boldsymbol{Q} = \frac{1}{2}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$