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.

22.2 Convex Functions

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

Definition 22.1 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

$$\left\{ \begin{bmatrix} \boldsymbol{x} \\ f(\boldsymbol{x}) \end{bmatrix} : \boldsymbol{x} \in \Omega \right\}.$$

∎

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 22.4). We next define the epigraph of a real-valued function.

Definition 22.2 The *epigraph* of a function $f : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^n$, denoted $\operatorname{epi}(f)$, is the set of points in $\Omega \times \mathbb{R}$ given by

$$\operatorname{epi}(f) = \left\{ \begin{bmatrix} \boldsymbol{x} \\ \beta \end{bmatrix} : \ \boldsymbol{x} \in \Omega, \ \beta \in \mathbb{R}, \ \beta \geq f(\boldsymbol{x}) \right\}.$$

∎

The epigraph $\operatorname{epi}(f)$ of a function f is simply the set of points in $\Omega \times \mathbb{R}$ on or above the graph of f (see Figure 22.4). We can also think of $\operatorname{epi}(f)$ 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 22.3 A function $f : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^n$, is *convex* on Ω if its epigraph is a convex set. ∎

Theorem 22.1 *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

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

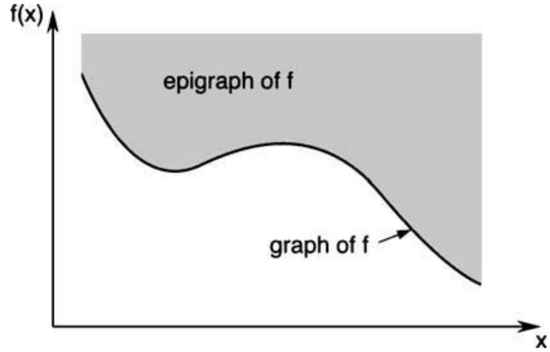

Figure 22.4 Graph and epigraph of a function $f : \mathbb{R} \to \mathbb{R}$.

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 $\boldsymbol{w} \notin \text{epi}(f)$, 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 22.2 *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 $[\boldsymbol{x}^\top, a]^\top$ and $[\boldsymbol{y}^\top, b]^\top$ 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

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

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

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

which implies that $\operatorname{epi}(f)$ 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,

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

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

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

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

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

This completes the proof. ∎

A geometric interpretation of Theorem 22.2 is given in Figure 22.5. 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 $[\boldsymbol{x}^\top, f(\boldsymbol{x})]^\top$ and $[\boldsymbol{y}^\top, f(\boldsymbol{y})]^\top$ must lie on or above the graph of f.

Using Theorem 22.2, it is straightforward to show that any nonnegative scaling of a convex function is convex, and that the sum of convex functions is convex.

Theorem 22.3 *Suppose that f, f_1, and f_2 are convex functions. Then, for any $a \geq 0$, the function af is convex. Moreover, $f_1 + f_2$ is convex.* □

Proof. Let $\boldsymbol{x}, \boldsymbol{y} \in \Omega$ and $\alpha \in (0,1)$. Fix $a \geq 0$. For convenience, write $\bar{f} = af$. We have

$$\begin{aligned} \bar{f}(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) &= af(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \\ &\leq a\,(\alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y})) \text{ because } f \text{ is convex and } a \geq 0 \\ &= \alpha(af(\boldsymbol{x})) + (1-\alpha)(af(\boldsymbol{y})) \\ &= \alpha \bar{f}(\boldsymbol{x}) + (1-\alpha)\bar{f}(\boldsymbol{y}), \end{aligned}$$

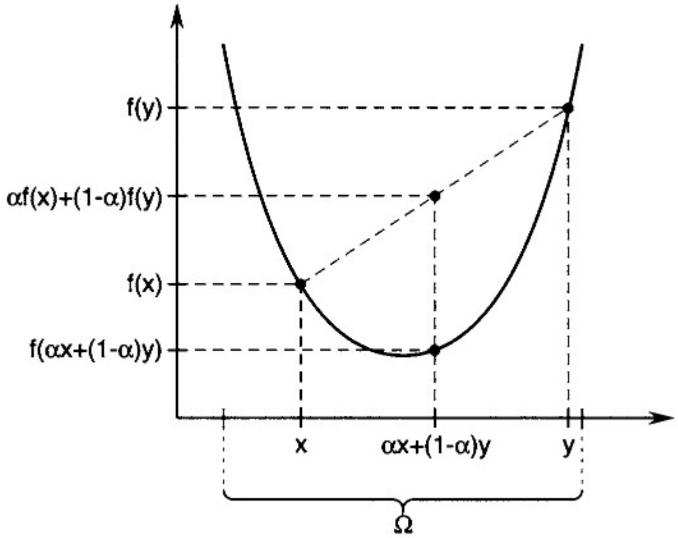

Figure 22.5 Geometric interpretation of Theorem 22.2.

which implies that $\bar{f}$ is convex.

Next, write $f_3 = f_1 + f_2$. We have

$$\begin{aligned} f_3(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) &= f_1(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) + f_2(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \\ &\leq (\alpha f_1(\boldsymbol{x}) + (1-\alpha) f_1(\boldsymbol{y})) + (\alpha f_2(\boldsymbol{x}) + (1-\alpha) f_2(\boldsymbol{y})) \\ &\quad \text{by convexity of } f_1 \text{ and } f_2 \\ &= \alpha(f_1(\boldsymbol{x}) + f_2(\boldsymbol{x})) + (1-\alpha)(f_1(\boldsymbol{y}) + f_2(\boldsymbol{y})) \\ &= \alpha f_3(\boldsymbol{x}) + (1-\alpha) f_3(\boldsymbol{y}), \end{aligned}$$

which implies that f_3 is convex. ∎

Theorem 22.3 implies that for any given collection of convex functions $f_1, \ldots, f_\ell$ and nonnegative numbers $c_1, \ldots, c_\ell$, the function $c_1 f_2 + \cdots + c_\ell f_\ell$ is convex. Using a method of proof similar to that used in Theorem 22.3, it is similarly straightforward to show that the function $\max\{f_1, \ldots, f_\ell\}$ is convex (see Exercise 22.6).

We now define the notion of strict convexity.

Definition 22.4 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 this definition, we see that for a strictly convex function, all points on the open line segment connecting the points $[\boldsymbol{x}^\top, f(\boldsymbol{x})]^\top$ and $[\boldsymbol{y}^\top, f(\boldsymbol{y})]^\top$ lie (strictly) above the graph of f.

Definition 22.5 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.

To show that a function is not convex, we need only produce a pair of points $\boldsymbol{x}, \boldsymbol{y} \in \Omega$ and an $\alpha \in (0,1)$ such that the inequality in Theorem 22.2 is violated.

Example 22.4 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, $\boldsymbol{x} = [1, 2]^\top \in \Omega$ and $\boldsymbol{y} = [2, 1]^\top \in \Omega$. 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. ∎

Example 22.4 is an illustration of the following general result.

Proposition 22.1 *A quadratic form $f : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^n$, given by $f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x}$, $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$, $\boldsymbol{Q} = \boldsymbol{Q}^\top$, is convex on Ω if and only if for all $\boldsymbol{x}, \boldsymbol{y} \in \Omega$, $(\boldsymbol{x} - \boldsymbol{y})^\top \boldsymbol{Q} (\boldsymbol{x} - \boldsymbol{y}) \geq 0$.* □

Proof. The result follows from Theorem 22.2. Indeed, the function $f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x}$ 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 this equation yields

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

Therefore, f is convex if and only if

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

which proves the result. ∎

Example 22.5 In Example 22.4, $f(\boldsymbol{x}) = x_1 x_2$, which can be written as $f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}$, where

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

Let $\Omega = \{\boldsymbol{x} : \boldsymbol{x} \geq \boldsymbol{0}\}$, and $\boldsymbol{x} = [2,2]^\top \in \Omega$, $\boldsymbol{y} = [1,3]^\top \in \Omega$. We have

$$\boldsymbol{y} - \boldsymbol{x} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

and

$$(\boldsymbol{y}-\boldsymbol{x})^\top \boldsymbol{Q}(\boldsymbol{y}-\boldsymbol{x}) = \frac{1}{2}[-1,1]\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = -1 < 0.$$

Hence, by Proposition 22.1, f is not convex on Ω. ∎

Differentiable convex functions can be characterized using the following theorem.

Theorem 22.4 *Let $f : \Omega \to \mathbb{R}$, $f \in \mathcal{C}^1$, be defined on an open convex set $\Omega \subset \mathbb{R}^n$. Then, f is convex on Ω if and only if for all $\boldsymbol{x}, \boldsymbol{y} \in \Omega$,*

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + Df(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x}).$$

□

Proof. $\Rightarrow$: Suppose that $f : \Omega \to \mathbb{R}$ is differentiable and convex. Then, by Theorem 22.2, for any $\boldsymbol{y}, \boldsymbol{x} \in \Omega$ and $\alpha \in (0,1)$ we have

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

Rearranging terms yields

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

Upon dividing both sides of this inequality by α, we get

$$\frac{f(\boldsymbol{x}+\alpha(\boldsymbol{y}-\boldsymbol{x}))-f(\boldsymbol{x})}{\alpha} \leq f(\boldsymbol{y})-f(\boldsymbol{x}).$$

If we now take the limit as $\alpha \to 0$ and apply the definition of the directional derivative of f at $\boldsymbol{x}$ in the direction $\boldsymbol{y}-\boldsymbol{x}$ (see Section 6.2), we get

$$Df(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x}) \leq f(\boldsymbol{y})-f(\boldsymbol{x})$$

or

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x})+Df(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x}).$$

$\Leftarrow$: Assume that Ω is convex, $f : \Omega \to \mathbb{R}$ is differentiable, and for all $\boldsymbol{x}, \boldsymbol{y} \in \Omega$,

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x})+Df(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x}).$$

Let $\boldsymbol{u}, \boldsymbol{v} \in \Omega$ and $\alpha \in (0,1)$. Because Ω is convex,

$$\boldsymbol{w}=\alpha\boldsymbol{u}+(1-\alpha)\boldsymbol{v} \in \Omega.$$

We also have

$$f(\boldsymbol{u}) \geq f(\boldsymbol{w})+Df(\boldsymbol{w})(\boldsymbol{u}-\boldsymbol{w})$$

and

$$f(\boldsymbol{v}) \geq f(\boldsymbol{w})+Df(\boldsymbol{w})(\boldsymbol{v}-\boldsymbol{w}).$$

Multiplying the first of this inequalities by α and the second by $(1-\alpha)$ and then adding them together yields

$$\alpha f(\boldsymbol{u})+(1-\alpha)f(\boldsymbol{v}) \geq f(\boldsymbol{w})+Df(\boldsymbol{w})\left(\alpha\boldsymbol{u}+(1-\alpha)\boldsymbol{v}-\boldsymbol{w}\right).$$

But

$$\boldsymbol{w}=\alpha\boldsymbol{u}+(1-\alpha)\boldsymbol{v}.$$

Hence,

$$\alpha f(\boldsymbol{u})+(1-\alpha)f(\boldsymbol{v}) \geq f(\alpha\boldsymbol{u}+(1-\alpha)\boldsymbol{v}).$$

Hence, by Theorem 22.2, f is a convex function. ∎

In Theorem 22.4, 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 Theorem 22.4 is given in Figure 22.6. To explain the interpretation, let $\boldsymbol{x}_0 \in \Omega$. The function $\ell(\boldsymbol{x}) = f(\boldsymbol{x}_0)+Df(\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0)$ is the linear approximation to f at $\boldsymbol{x}_0$. The theorem says that the graph of f always lies above its linear approximation at any point. In other words,

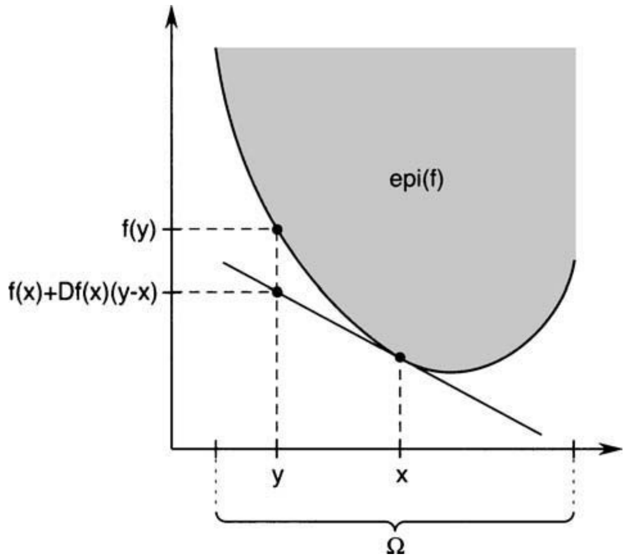

Figure 22.6 Geometric interpretation of Theorem 22.4.

the linear approximation to a convex function f at any point of its domain lies below epi(f).

This geometric idea leads to a generalization of the gradient to the case where f is not differentiable. Let $f : \Omega \to \mathbb{R}$ be defined on an open convex set $\Omega \subset \mathbb{R}^n$. A vector $\boldsymbol{g} \in \mathbb{R}^n$ is said to be a *subgradient* of f at a point $\boldsymbol{x} \in \Omega$ if for all $\boldsymbol{y} \in \Omega$,

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \boldsymbol{g}^\top(\boldsymbol{y} - \boldsymbol{x}).$$

As in the case of the standard gradient, if $\boldsymbol{g}$ is a subgradient, then for a given $\boldsymbol{x}_0 \in \Omega$, the function $\ell(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \boldsymbol{g}^\top(\boldsymbol{x} - \boldsymbol{x}_0)$ lies below epi(f).

For functions that are twice continuously differentiable, the following theorem gives another possible characterization of convexity.

Theorem 22.5 *Let $f : \Omega \to \mathbb{R}$, $f \in \mathcal{C}^2$, be defined on an open convex set $\Omega \subset \mathbb{R}^n$. Then, f is convex on Ω if and only if for each $\boldsymbol{x} \in \Omega$, the Hessian $\boldsymbol{F}(\boldsymbol{x})$ of f at $\boldsymbol{x}$ is a positive semidefinite matrix.* □

Proof. $\Leftarrow$: Let $\boldsymbol{x}, \boldsymbol{y} \in \Omega$. Because $f \in \mathcal{C}^2$, by Taylor's theorem there exists $\alpha \in (0, 1)$ such that

$$f(\boldsymbol{y}) = f(\boldsymbol{x}) + Df(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x}) + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{x})^\top \boldsymbol{F}(\boldsymbol{x} + \alpha(\boldsymbol{y} - \boldsymbol{x}))(\boldsymbol{y} - \boldsymbol{x}).$$

Because $\boldsymbol{F}(\boldsymbol{x} + \alpha(\boldsymbol{y} - \boldsymbol{x}))$ is positive semidefinite,

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

Therefore, we have

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + Df(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x}),$$

which implies that f is convex, by Theorem 22.4.

$\Rightarrow$: We use contraposition. Assume that there exists $\boldsymbol{x} \in \Omega$ such that $\boldsymbol{F}(\boldsymbol{x})$ is not positive semidefinite. Therefore, there exists $\boldsymbol{d} \in \mathbb{R}^n$ such that $\boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x})\boldsymbol{d} < 0$. By assumption, Ω is open; thus, the point $\boldsymbol{x}$ is an interior point. By the continuity of the Hessian matrix, there exists a nonzero $s \in \mathbb{R}$ such that $\boldsymbol{x} + s\boldsymbol{d} \in \Omega$, and if we write $\boldsymbol{y} = \boldsymbol{x} + s\boldsymbol{d}$, then for all points $\boldsymbol{z}$ on the line segment joining $\boldsymbol{x}$ and $\boldsymbol{y}$, we have $\boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{z})\boldsymbol{d} < 0$. By Taylor's theorem there exists $\alpha \in (0,1)$ such that

$$\begin{aligned} f(\boldsymbol{y}) &= f(\boldsymbol{x}) + Df(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x}) + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{x})^\top \boldsymbol{F}(\boldsymbol{x}+\alpha(\boldsymbol{y}-\boldsymbol{x}))(\boldsymbol{y}-\boldsymbol{x}) \\ &= f(\boldsymbol{x}) + Df(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x}) + \frac{1}{2}s^2\boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}+\alpha s\boldsymbol{d})\boldsymbol{d}. \end{aligned}$$

Because $\alpha \in (0,1)$, the point $\boldsymbol{x} + \alpha s\boldsymbol{d}$ is on the line segment joining $\boldsymbol{x}$ and $\boldsymbol{y}$, and therefore

$$\boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}+\alpha s\boldsymbol{d})\boldsymbol{d} < 0.$$

Because $s \neq 0$, we have $s^2 > 0$, and hence

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

Therefore, by Theorem 22.4, f is not a convex function. ∎

Theorem 22.5 can be strengthened to include nonopen sets by modifying the condition to be $(\boldsymbol{y}-\boldsymbol{x})^\top \boldsymbol{F}(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x}, \boldsymbol{y} \in \Omega$ (and assuming that $f \in \mathcal{C}^2$ on some open set that contains Ω; for example, $f \in \mathcal{C}^2$ on $\mathbb{R}^n$). A proof similar to that above can be used in this case.

Note that by definition of concavity, a function $f : \Omega \to \mathbb{R}$, $f \in \mathcal{C}^2$, is concave over the convex set $\Omega \subset \mathbb{R}^n$ if and only if for all $\boldsymbol{x} \in \Omega$, the Hessian $\boldsymbol{F}(\boldsymbol{x})$ of f is negative semidefinite.

Example 22.6 Determine whether the following functions are convex, concave, or neither:

1. $f : \mathbb{R} \to \mathbb{R}$, $f(x) = -8x^2$.
2. $f : \mathbb{R}^3 \to \mathbb{R}$, $f(\boldsymbol{x}) = 4x_1^2 + 3x_2^2 + 5x_3^2 + 6x_1x_2 + x_1x_3 - 3x_1 - 2x_2 + 15$.
3. $f : \mathbb{R}^2 \to \mathbb{R}$, $f(\boldsymbol{x}) = 2x_1x_2 - x_1^2 - x_2^2$.

Solution:

1. We use Theorem 22.5. We first compute the Hessian, which in this case is just the second derivative: $(d^2f/dx^2)(x) = -16 < 0$ for all $x \in \mathbb{R}$. Hence, f is concave over $\mathbb{R}$.

2. The Hessian matrix of f is

$$F(x) = \begin{bmatrix} 8 & 6 & 1 \\ 6 & 6 & 0 \\ 1 & 0 & 10 \end{bmatrix}.$$

The leading principal minors of $F(x)$ are

$$\Delta_1 = 8 > 0,$$
$$\Delta_2 = \det \begin{bmatrix} 8 & 6 \\ 6 & 6 \end{bmatrix} = 12 > 0,$$
$$\Delta_3 = \det F(x) = 114 > 0.$$

Hence, $F(x)$ is positive definite for all $x \in \mathbb{R}^3$. Therefore, f is a convex function over $\mathbb{R}^3$.

3. The Hessian of f is

$$F(x) = \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix},$$

which is negative semidefinite for all $x \in \mathbb{R}^2$. Hence, f is concave on $\mathbb{R}^2$. ∎

22.3 Convex Optimization Problems

In this section we consider optimization problems where the objective function is a convex function and the constraint set is a convex set. We refer to such problems as *convex optimization problems* or *convex programming problems*. Optimization problems that can be classified as convex programming problems include linear programs and optimization problems with quadratic objective function and linear constraints. Convex programming problems are interesting for several reasons. Specifically, as we shall see, local minimizers are global for such problems. Furthermore, first-order necessary conditions become sufficient conditions for minimization.

Our first theorem below states that in convex programming problems, local minimizers are also global.

Theorem 22.6 *Let $f : \Omega \to \mathbb{R}$ be a convex function defined on a convex set $\Omega \subset \mathbb{R}^n$. Then, a point is a global minimizer of f over Ω if and only if it is a local minimizer of f.* □

Proof. $\Rightarrow$: This is obvious.

$\Leftarrow$: We prove this by contraposition. Suppose that $\boldsymbol{x}^*$ is not a global minimizer of f over Ω. Then, for some $\boldsymbol{y} \in \Omega$, we have $f(\boldsymbol{y}) < f(\boldsymbol{x}^*)$. By assumption, the function f is convex, and hence for all $\alpha \in (0, 1)$,

$$f(\alpha \boldsymbol{y} + (1 - \alpha)\boldsymbol{x}^*) \leq \alpha f(\boldsymbol{y}) + (1 - \alpha) f(\boldsymbol{x}^*).$$

Because $f(\boldsymbol{y}) < f(\boldsymbol{x}^*)$, we have

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

Thus, for all $\alpha \in (0, 1)$,

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

Hence, there exist points that are arbitrarily close to $\boldsymbol{x}^*$ and have lower objective function value. For example, the sequence $\{\boldsymbol{y}_n\}$ of points given by

$$\boldsymbol{y}_n = \frac{1}{n}\boldsymbol{y} + \left(1 - \frac{1}{n}\right)\boldsymbol{x}^*$$

converges to $\boldsymbol{x}^*$, and $f(\boldsymbol{y}_n) < f(\boldsymbol{x}^*)$. Hence, $\boldsymbol{x}^*$ is not a local minimizer, which completes the proof. ∎

We now show that the set of global optimizers is convex. For this, we need the following lemma.

Lemma 22.1 *Let $g : \Omega \to \mathbb{R}$ be a convex function defined on a convex set $\Omega \subset \mathbb{R}^n$. Then, for each $c \in \mathbb{R}$, the set*

$$\Gamma_c = \{\boldsymbol{x} \in \Omega : g(\boldsymbol{x}) \leq c\}$$

is a convex set. □

Proof. Let $\boldsymbol{x}, \boldsymbol{y} \in \Gamma_c$. Then, $g(\boldsymbol{x}) \leq c$ and $g(\boldsymbol{y}) \leq c$. Because g is convex, for all $\alpha \in (0, 1)$,

$$g(\alpha \boldsymbol{x} + (1 - \alpha)\boldsymbol{y}) \leq \alpha g(\boldsymbol{x}) + (1 - \alpha) g(\boldsymbol{y}) \leq c.$$

Hence, $\alpha \boldsymbol{x} + (1 - \alpha)\boldsymbol{y} \in \Gamma_c$, which implies that Γ_c is convex. ∎

Corollary 22.1 *Let $f : \Omega \to \mathbb{R}$ be a convex function defined on a convex set $\Omega \subset \mathbb{R}^n$. Then, the set of all global minimizers of f over Ω is a convex set.* □

Proof. The result follows immediately from Lemma 22.1 by setting

$$c = \min_{\boldsymbol{x} \in \Omega} f(\boldsymbol{x}).$$

∎

We now show that if the objective function is continuously differentiable and convex, then the first-order necessary condition (see Theorem 6.1) for a point to be a minimizer is also sufficient. We use the following lemma.

Lemma 22.2 *Let $f : \Omega \to \mathbb{R}$ be a convex function defined on the convex set $\Omega \subset \mathbb{R}^n$ and $f \in \mathcal{C}^1$ on an open convex set containing Ω. Suppose that the point $\boldsymbol{x}^* \in \Omega$ is such that for all $\boldsymbol{x} \in \Omega$, $\boldsymbol{x} \neq \boldsymbol{x}^*$, we have*

$$Df(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0.$$

Then, $\boldsymbol{x}^*$ is a global minimizer of f over Ω. □

Proof. Because the function f is convex, by Theorem 22.4, for all $\boldsymbol{x} \in \Omega$, we have

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + Df(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*).$$

Hence, the condition $Df(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0$ implies that $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$. ∎

Observe that for any $\boldsymbol{x} \in \Omega$, the vector $\boldsymbol{x} - \boldsymbol{x}^*$ can be interpreted as a feasible direction at $\boldsymbol{x}^*$ (see Definition 6.2). Using Lemma 22.2, we have the following theorem (cf. Theorem 6.1).

Theorem 22.7 *Let $f : \Omega \to \mathbb{R}$ be a convex function defined on the convex set $\Omega \subset \mathbb{R}^n$, and $f \in \mathcal{C}^1$ on an open convex set containing Ω. Suppose that the point $\boldsymbol{x}^* \in \Omega$ is such that for any feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$, we have*

$$\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) \geq 0.$$

Then, $\boldsymbol{x}^*$ is a global minimizer of f over Ω. □

Proof. Let $\boldsymbol{x} \in \Omega$, $\boldsymbol{x} \neq \boldsymbol{x}^*$. By convexity of Ω,

$$\boldsymbol{x}^* + \alpha(\boldsymbol{x} - \boldsymbol{x}^*) = \alpha\boldsymbol{x} + (1 - \alpha)\boldsymbol{x}^* \in \Omega$$

for all $\alpha \in (0, 1)$. Hence, the vector $\boldsymbol{d} = \boldsymbol{x} - \boldsymbol{x}^*$ is a feasible direction at $\boldsymbol{x}^*$ (see Definition 6.2). By assumption,

$$Df(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*) = \boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) \geq 0.$$

Hence, by Lemma 22.2, $\boldsymbol{x}^*$ is a global minimizer of f over Ω. ∎

From Theorem 22.7, we easily deduce the following corollary (compare this with Corollary 6.1).

Corollary 22.2 *Let $f : \Omega \to \mathbb{R}$, $f \in \mathcal{C}^1$, be a convex function defined on the convex set $\Omega \subset \mathbb{R}^n$. Suppose that the point $\boldsymbol{x}^* \in \Omega$ is such that*

$$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}.$$

Then, $\boldsymbol{x}^*$ is a global minimizer of f over Ω. □

We now consider the constrained optimization problem

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}. \end{array}$$

We assume that the feasible set is convex. An example where this is the case is when

$$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}.$$

The following theorem states that provided the feasible set is convex, the Lagrange condition is sufficient for a point to be a minimizer.

Theorem 22.8 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $f \in \mathcal{C}^1$, be a convex function on the set of feasible points*

$$\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}\},$$

where $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{h} \in \mathcal{C}^1$, and Ω is convex. Suppose that there exist $\boldsymbol{x}^* \in \Omega$ and $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that

$$Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*) = \mathbf{0}^\top.$$

Then, $\boldsymbol{x}^*$ is a global minimizer of f over Ω. □

Proof. By Theorem 22.4, for all $\boldsymbol{x} \in \Omega$, we have

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + Df(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*).$$

Substituting $Df(\boldsymbol{x}^*) = -\boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*)$ into the inequality above yields

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) - \boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*).$$

Because Ω is convex, $(1-\alpha)\boldsymbol{x}^* + \alpha\boldsymbol{x} \in \Omega$ for all $\alpha \in (0,1)$. Thus,

$$\boldsymbol{h}(\boldsymbol{x}^* + \alpha(\boldsymbol{x} - \boldsymbol{x}^*)) = \boldsymbol{h}((1-\alpha)\boldsymbol{x}^* + \alpha\boldsymbol{x}) = \mathbf{0}$$

for all $\alpha \in (0,1)$. Premultiplying by $\boldsymbol{\lambda}^{*\top}$, subtracting $\boldsymbol{\lambda}^{*\top}\boldsymbol{h}(\boldsymbol{x}^*) = 0$, and dividing by α, we get

$$\frac{\boldsymbol{\lambda}^{*\top}\boldsymbol{h}(\boldsymbol{x}^* + \alpha(\boldsymbol{x} - \boldsymbol{x}^*)) - \boldsymbol{\lambda}^{*\top}\boldsymbol{h}(\boldsymbol{x}^*)}{\alpha} = 0$$

for all $\alpha \in (0,1)$. If we now take the limit as $\alpha \to 0$ and apply the definition of the directional derivative of $\boldsymbol{\lambda}^{*\top}\boldsymbol{h}$ at $\boldsymbol{x}^*$ in the direction $\boldsymbol{x} - \boldsymbol{x}^*$ (see Section 6.2), we get

$$\boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*) = 0.$$

Hence,

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*),$$

which implies that x^* is a global minimizer of f over Ω. ∎

Consider the general constrained optimization problem

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \\ & \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}. \end{array}$$

As before, we assume that the feasible set is convex. This is the case if, for example, the two sets $\{\boldsymbol{x} : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$ and $\{\boldsymbol{x} : \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\}$ are convex, because the feasible set is the intersection of these two sets (see also Theorem 4.1). We have already seen an example where the set $\{\boldsymbol{x} : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$ is convex. On the other hand, an example where the set $\{\boldsymbol{x} : \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\}$ is convex is when the components of $\boldsymbol{g} = [g_1, \ldots, g_p]^\top$ are all convex functions. Indeed, the set $\{\boldsymbol{x} : \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\}$ is the intersection of the sets $\{\boldsymbol{x} : g_i(\boldsymbol{x}) \leq 0\}$, $i = 1, \ldots, p$. Because each of these sets is convex (by Lemma 22.1), their intersection is also convex.

We now prove that the Karush-Kuhn-Tucker (KKT) condition is sufficient for a point to be a minimizer to the problem above.

Theorem 22.9 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $f \in \mathcal{C}^1$, be a convex function on the set of feasible points*

$$\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\},$$

where $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^p$, $\boldsymbol{h}, \boldsymbol{g} \in \mathcal{C}^1$, and Ω is convex. Suppose that there exist $\boldsymbol{x}^* \in \Omega$, $\boldsymbol{\lambda}^* \in \mathbb{R}^m$, and $\boldsymbol{\mu}^* \in \mathbb{R}^p$, such that

1. $\boldsymbol{\mu}^* \geq \boldsymbol{0}$.
2. $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*) + \boldsymbol{\mu}^{*\top} D\boldsymbol{g}(\boldsymbol{x}^*) = \boldsymbol{0}^\top$.
3. $\boldsymbol{\mu}^{*\top} \boldsymbol{g}(\boldsymbol{x}^*) = 0$.

Then, $\boldsymbol{x}^*$ is a global minimizer of f over Ω. □

Proof. Suppose that $\boldsymbol{x} \in \Omega$. By convexity of f and Theorem 22.4,

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + Df(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*).$$

Using condition 2, we get

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) - \boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*) - \boldsymbol{\mu}^{*\top} D\boldsymbol{g}(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*).$$

As in the proof of Theorem 22.8, we can show that $\boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*) = 0$. We now claim that $\boldsymbol{\mu}^{*\top} D\boldsymbol{g}(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*) \leq 0$. To see this, note that because Ω is convex, $(1 - \alpha)\boldsymbol{x}^* + \alpha\boldsymbol{x} \in \Omega$ for all $\alpha \in (0, 1)$. Thus,

$$\boldsymbol{g}(\boldsymbol{x}^* + \alpha(\boldsymbol{x} - \boldsymbol{x}^*)) = \boldsymbol{g}((1 - \alpha)\boldsymbol{x}^* + \alpha\boldsymbol{x}) \leq \boldsymbol{0}$$

for all $\alpha \in (0,1)$. Premultiplying by $\boldsymbol{\mu}^{*\top} \geq \mathbf{0}^\top$ (by condition 1), subtracting $\boldsymbol{\mu}^{*\top}\boldsymbol{g}(\boldsymbol{x}^*) = 0$ (by condition 3), and dividing by α, we get

$$\frac{\boldsymbol{\mu}^{*\top}\boldsymbol{g}(\boldsymbol{x}^* + \alpha(\boldsymbol{x} - \boldsymbol{x}^*)) - \boldsymbol{\mu}^{*\top}\boldsymbol{g}(\boldsymbol{x}^*)}{\alpha} \leq 0.$$

We now take the limit as $\alpha \to 0$ to obtain $\boldsymbol{\mu}^{*\top} D\boldsymbol{g}(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*) \leq 0$.

From the above, we have

$$\begin{aligned} f(\boldsymbol{x}) &\geq f(\boldsymbol{x}^*) - \boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*) - \boldsymbol{\mu}^{*\top} D\boldsymbol{g}(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*) \\ &\geq f(\boldsymbol{x}^*) \end{aligned}$$

for all $\boldsymbol{x} \in \Omega$, which completes the proof. ∎

Example 22.7 A bank account starts out with 0 dollars. At the beginning of each month, we deposit some money into the bank account. Denote by x_k the amount deposited in the kth month, $k = 1, 2, \ldots$. Suppose that the monthly interest rate is $r > 0$ and the interest is paid into the account at the end of each month (and compounded). We wish to maximize the total amount of money accumulated at the end of n months, such that the total money deposited during the n months does not exceed D dollars (where $D > 0$).

To solve this problem we first show that the problem can be posed as a linear program, and is therefore a convex optimization problem. Let y_k be the total amount in the bank at the end of the kth month. Then, $y_k = (1+r)(y_{k-1} + x_k)$, $k \geq 1$, with $y_0 = 0$. Therefore, we want to maximize y_n subject to the constraint that $x_k \geq 0$, $k = 1, \ldots, n$, and $x_1 + \cdots + x_n \leq D$. It is easy to deduce that

$$y_n = (1+r)^n x_1 + (1+r)^{n-1} x_2 + \cdots + (1+r)x_n.$$

Let $\boldsymbol{c}^\top = [(1+r)^n, (1+r)^{n-1}, \ldots, (1+r)]$, $\boldsymbol{e}^\top = [1, \ldots, 1]$, and $\boldsymbol{x} = [x_1, \ldots, x_n]^\top$. Then, we can write the problem as

$$\begin{aligned} \text{maximize} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{e}^\top \boldsymbol{x} \leq D \\ & \boldsymbol{x} \geq \mathbf{0}, \end{aligned}$$

which is a linear program.

It is intuitively clear that the optimal strategy is to deposit D dollars in the first month. To show that this strategy is indeed optimal, we use Theorem 22.9. Let $\boldsymbol{x}^* = [D, 0, \ldots, 0]^\top \in \mathbb{R}^n$. Because the problem is a convex programming problem, it suffices to show that $\boldsymbol{x}^*$ satisfies the KKT condition (see Theorem 22.9). The KKT condition for this problem is

$$
\begin{aligned}
-\boldsymbol{c}^\top + \mu^{(1)}\boldsymbol{e}^\top - \boldsymbol{\mu}^{(2)\top} &= 0, \\
\mu^{(1)}(\boldsymbol{e}^\top \boldsymbol{x}^* - D) &= 0, \\
\boldsymbol{\mu}^{(2)\top}\boldsymbol{x}^* &= 0, \\
\boldsymbol{e}^\top \boldsymbol{x}^* - D &\le 0, \\
-\boldsymbol{x}^* &\le \boldsymbol{0}, \\
\mu^{(1)} &\ge 0, \\
\boldsymbol{\mu}^{(2)} &\ge \boldsymbol{0}, \\
\boldsymbol{e}^\top \boldsymbol{x} &\le D, \\
\boldsymbol{x} &\ge \boldsymbol{0},
\end{aligned}
$$

where $\mu^{(1)} \in \mathbb{R}$ and $\boldsymbol{\mu}^{(2)} \in \mathbb{R}^n$. Let $\mu^{(1)} = (1+r)^n$ and $\boldsymbol{\mu}^{(2)} = (1+r)^n\boldsymbol{e} - \boldsymbol{c}$. Then, it is clear that the KKT condition is satisfied. Therefore, $\boldsymbol{x}^*$ is a global minimizer. ∎

An entire book devoted to the vast topic of convexity and optimization is [7]. For extensions of the theory of convex optimization, we refer the reader to [136, Chapter 10]. The study of convex programming problems also serves as a prerequisite to *nondifferentiable optimization* (see, e.g., [38]).

22.4 Semidefinite Programming

Semidefinite programming is a subfield of convex optimization concerned with minimizing a linear objective function subject to a linear matrix inequality. The linear matrix inequality constraint defines a convex feasible set over which the linear objective function is to be minimized. Semidefinite programming can be viewed as an extension of linear programming, where the componentwise inequalities on vectors are replaced by matrix inequalities (see Exercise 22.20). For further reading on the subject of semidefinite programming, we recommend an excellent survey paper by Vandenberghe and Boyd [128].

Linear Matrix Inequalities and Their Properties

Consider $n+1$ real symmetric matrices

$$\boldsymbol{F}_i = \boldsymbol{F}_i^\top \in \mathbb{R}^{m\times m}, \quad i = 0, 1, \ldots, n$$

and a vector

$$\boldsymbol{x} = [x_1, \ldots, x_n]^\top \in \mathbb{R}^n.$$